Table of Contents

Effect of Temperature Compensator in Thermal Efficiency Management in Data Centre

C.T. Ikwuazom[1], D.O. Njoku[2], E.C. Nwokorie[2], I. A. Amaefule[3], O.C. Nwokonkwo[4], J.E. Jibiri[4],T.C. Okeahialam[1] F.U. Madu[5]

[1]*Department of Information Media Technology, Federal University of Technology, Minna, Niger State*
[2] *Department of Computer Science, Federal University of Technology, Owerri, Imo State*
[3] *Department of Computer Science, Imo State University, Owerri*
[4] *Department of Information Technology, Federal University of Technology, Owerri, Imo State*
[5]*Department of Computer Science, Federal Polytechnic, Nekede, Owerri*

Abstract: ***This paper has examined the effect of temperature compensator in thermal efficiency management in data centre. The use of computer room air control (CRAC) unit in ensuring thermal efficiency facilitate thermal process stability such that variation of temperature does not exceed predetermined value expected for effective working of data centre infrastructures such as computer hardware and server racks. A proportional-integral-derivative (PID) tuned compensator was designed in MATLAB environment and introduced into feedback network of thermal process in a data centre. Simulation result showed that predetermined temperature of 1^{o}C was maintained when the designed compensator was introduced. Hence, the compensator provided thermal efficiency and stability by ensuring that the expected temperature in the data centre was not exceed and no deviation in the ideal temperature of the computer room air temperature (CRAT) and actual value obtained as step response.***

Keyword: CRAC, Compensator, Data centre, Efficiency, Thermal process

1. Introduction

The use of electronic equipment in homes, offices and other enclosed spaces has rapidly increased in recent times. One of the most widely used electronics is computer and it is important electronic equipment in indoor environment whose use has dramatically increased over the past few decades. Computers are penetrating every work place and home, and are used for a vast range of tasks including as data storage facilities in data centre.
In many organisations and institutions such as Information and Communication Technology (ICT) companies and school computer laboratories/classrooms, a number of computers are introduced in offices and other indoor environments to facilitate data processing, storage, sharing of information, file management, and learning. The use of computers as essential electronic component of ICT is changing the way businesses are conducted by organisations including educational institutions. These changes have impacted on data/information sharing or storage practices among institutions globally.

Also, the demand for data processing is on the rise in recent times because of the technological advancements in computer and electronic systems, which has caused a rapid increase in data centre sector[1]. The advances in Information and Communication Technology (ICT) devices have invariably resulted in huge data that must be processed and kept in data centre. In fact, data centre is a key component of ICT that is used in the collection, storage, processing, and distribution of big amount of data for application such as business enterprise, cyber-physical system, and social networking. Hence, data centre workload and its energy consumption are rapidly increasing due to the continuous rising demand of remote data services [11].

With data centre rapidly becoming a critical asset for companies because valuable and sensitive data that are essential to the sustainability of business activities are stored in it, there is need to ensure that the computers in data centres are operationally reliable

One of the environment factors that can adversely affect the safety and reliability of the data centre is the thermal energy within the computer room. Thermal energy in a data centre can be described as a measure of air volume temperature of the room. Thus, temperature profile is critical to energy efficiency and effective working of a data centre. This is because computers and network devices in data centre are susceptible to high temperature. In some countries, standard temperature in data centre or server room is established. For instance, in Indonesia, the standard temperature for server room is 21 – 23^{o}C [9]. The allowable temperature range given by American Society of Heating, refrigerating, and Air-Conditioning Engineers standard is 18 – 27^{o}C [3]. Hence, regulating temperature in data centre and achieving optimal uptime and efficiency is critical [8].
Several techniques have been implemented to facilitate computer room air temperature control that ensures that temperature in a data centre is kept at predetermined value or range. In this paper, the objective is to study the effect of temperature on the effectiveness and efficiency of data centre and how the implementation of temperature monitoring and regulating algorithm helps in ensuring that desired or predetermined temperature are maintained irrespective of disturbance caused by ingress of external hot air.

2. CRAC in Data Centre Architecture

The importance of data centre makes IT companies and other businesses that heavily depend on technologies to invest huge amount of money on top-quality equipment [5]. Nevertheless, high level of thermal energy is generated by these equipment that can be harmful to the operation of the in data centre in various ways. Thus, the effect of temperature in the efficient performance of data centre is considered in this section. A typical data centre structure is shown in Figure 1, and mainly comprising Information Technology (IT) system and computer room air conditioner (CRAC).

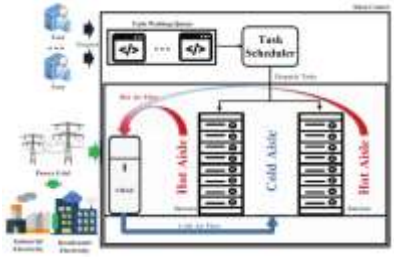


Figure 1: Data Centre Structure[2]

The purpose of CRAC in data centre is basically to ensure that a suitable environmental condition needed for the operation of IT equipment is created and maintained. That is the CRAC is the technology for controlling and maintaining optimal temperature, distribution and air flow within data centre. The CRAC technology ensures that heat conditions are regulated to provide the desired temperature for effective operation of data centre when the room is exposed to high-level temperatures as a result of the computer hardware capacity in data centres and network areas (Data Centre Cooling Explained). This way, the breakdown and permanent damage to hardware because of overheating that can be experienced from time is prevented. Generally, the operation of CRAC in data center is to guarantee precise humidity and stability of temperature suitable for high loading areas such as data centres, network areas and server rooms.

In the absence of CRAC in data centre, mainframes and racks of servers can become overheated with time. The heat rejection of an average rack is 3 kW but there are servers with up to 20 kW of heat rejection, which makes sizing of the CRAC critical for regulating the temperature of computer hardware in making sure it runs smoothly.
Recently, several techniques have been developed to regulate temperature in data centres to improve the performance of CRAC. Some of the techniques involve the use of classical concepts such as proportional and integral (PI) feedback control, and proportional-integral-derivative (PID) control algorithm. Also, intelligent control

systems such as fuzzy logic controller (FLC), and adaptive self-tuning PID-type fuzzy have been proposed in literature.

2. System Design

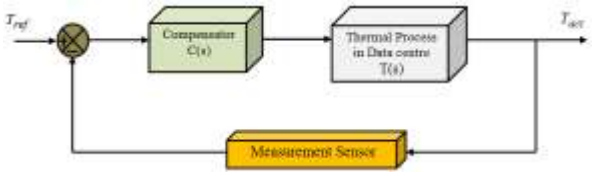
Thermal Model of Data Centre

The mathematical description of thermal process in data-centre is presented in this section. The model is a second order inertial and net delay process representing the transfer function of open volume to air temperature of a data centre given by[4]:

$$T(s) = \frac{10}{(20s+1)(30s+1)} e^{-12s} \quad (1)$$

Equation (1) represents the mathematical model of thermal process in terms of temperature in a data centre with a time delay of 12 minutes or 0.2 second. The mathematical derivation and operation of the fan speed in variable air volume (VAV) that was employed to determine the air flux in the data-centre considered in this study is in available in [4]. The model is used in this paper to study the effect of temperature in data centre via computer simulation using the MATLAB tool. Feedback network of temperature regulating closed loop system subject to unit step input in a data centre is shown in Figures 2. T_{ref} and T_{act} are the desired (or reference) temperature and actual temperature of the system.

System condition	Rise time (s)	Peak time (s)	Peak overshoot (%)	Settling time (s)	Final value to unit step input (^oC)	Remark on step response performance to unit step input
System without PID tuned compensator	1.41	> 250	0	2.68	10	Unsatisfactory
System with	1.13	2.51	7.17	3.44	1	Satisfactory



compensator

Figure 2: Feedback Network of Temperature Regulating Loop in Data Centre

3 Design of PID Tuned Compensator

The addition of compensator in feedback control system is an important technique that is commonly used in process industries [10]. The main advantage offers by compensated system is that corrective action takes place as soon as the actual output deviates from referenced input irrespective of the source and type of disturbance. Thus, compensators are sub-system that provides corrective action when introduced into system to compensate for performance deficiency of the plant or process[10].

PID tuned compensator has been proposed in [7] and [6] to achieved very impressive robust and desired input tracking performance. In this paper, a compensator is developed by employing PID tuning based on robust response time using Control and Estimation Tools Manager (CETM) of MATLAB computer simulation software. The tuning procedure is shown in Figure 3, which is single input single output (SISO) design tool graphical user interface (GUI). The PID tuned compensator incorporates first order derivative filter.

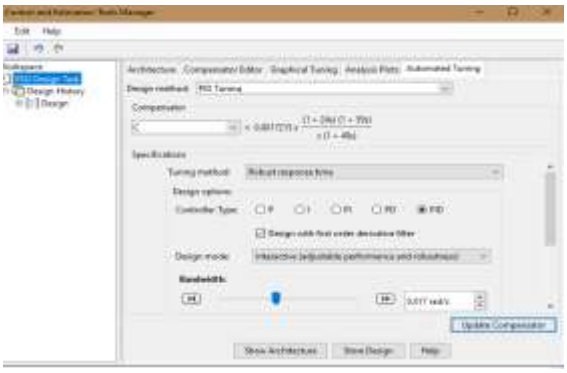


Figure 3 GUI of PID tuned compensator design

Thus from Fig. 3, the designed temperature PID tuned compensator for data centre is given by:

$$C(s) = 0.0017215\frac{(1+24s)(1+59s)}{s(1+49s)} \tag{2}$$

3. Simulation Result and Analysis

Using Equation (1) as a model of thermal process in data centre, the simulation of volume to air temperature was conducted in MATLAB R2015a environment and the system step response to unit step temperature input in degree is shown in Figure 4.

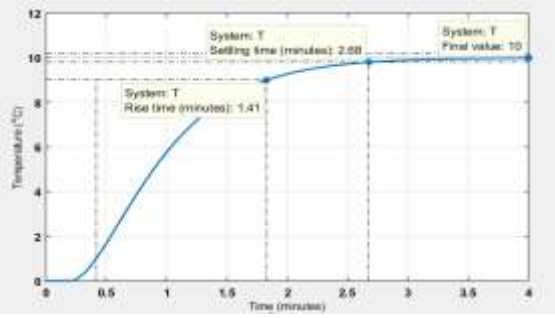


Figure 4 Temperature response of CRAC to unit step input (without compensator)

Further simulation was carried out to determine the step response of system to unit temperature input with the designed compensator introduced into the feedback network of the temperature regulator assuming the measurement (or temperature) sensor has a unit gain as shown in Figure 2. The result of the simulation is shown in Figure 5. The performance parameters of the system obtained in time domain from the simulation performed are given in Table 1.

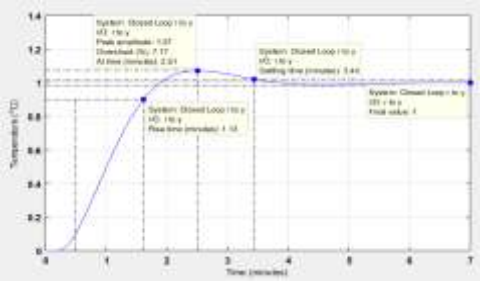


Figure 5 Temperature response of CRAC to unit step input (compensator)

Table 1: Time domain performance parameters of the system

As it can be seen from Table 1, when the thermal process performance has not been compensated in the data centre to ensure a desired temperature is maintained in the room, the actual temperature in the room seems to be higher than the desired value. That is with expected (or set point) temperature in the room at $1^{o}C$, the actual temperature maintained in the room was $10^{o}C$, a value very much higher than the desired temperature as shown in Figure 4. In a situation like this in which desired level of temperature is not achieved but rather a higher value, the data centre can be overheated and thereby leading to deficiency in data centre operation and damage to or malfunctioning of IT infrastructures such as computers. With the introduction of compensator in the feedback network of the thermal process of data centre, which forms a CRAC, the desired temperature was achieved at the output as can be seen of the final value to unit step input in Table1, which shows that the system maintained steady and precise temperature corresponding to the temperature expected of data centre to operate smoothly. Thus as shown in Figure 5, with actual temperature (step response) value equal to desired value (unit step input), the deviation or error is zero and the as such a steady state temperature is maintained in the data centre. It suffices to say that the use of temperature regulating system in data centre will certainly improve system efficiency and stability.

4. Conclusion

The effect introducing temperature compensator in ensuring thermal process efficiency in data centre has been examined via computer simulation in MATLAB 2015a environment. The system performance was analyzed in terms of transient and steady state step response to unit step input temperature (expressed as $1^{o}C$). The results obtained from the computer simulation tests conducted showed that remarkable performance in terms effectiveness

in maintaining an ideal predetermined temperature was achieved using a compensator in the feedback temperature control loop of thermal process in data centre.

References

[1] Capozzoli, A. & Primiceri, G. (2015). Cooling systems in data centers: state of art and emerging technologies. *7th International Conference on Sustainability in Energy and Building, Energy Procedia 83*(2015), 484-493, 2015. doi:10.1016/j.egypro.2015.12.168

[2] Chi, C., Ji, K., Song, P., Marahatta, A., Zhang, S., Zhang, F., Qiu, D., & Liu, Z. (2021). Cooperatively improving data center energy efficiency based on multi-agent deep reinforcement learning. *Energies, 14*(2071), 1-32. https://doi.org/10.3390/en14082071

[3] Cho, J., Park, B., & Jeong, Y. (2019). Thermal performance evaluation of a data center cooling system under fault conditions. *Energies*, 12 (2996), 1-16. https://doi.org/10.3390/en12152996
Data Centre Cooling Explained (n.d.). https://itair.co.nz/data-centre-cooling-explained/

[4] Deng, J., Yang, L., Cheng, X., & Liu, W. (2013). Self-tuning PID-type fuzzy adaptive control for CRAC in datacenters. *7th International Conference on Computer and Computing Technologies Agriculture (CCTA)*, Beijing, China, 215-225, 2013. https://doi.org/10.1007/978-3-642-54344-9_27

[5] El Helou, A. (2021). Server chiller & cooling to maintain ideal server room temperature. https://waterchillers.com/blog/post/server-chiller-and-cooling-to-maintain-ideal-server-room-temperature

[6]Ekengwu, B. O., Eze, P. C., Nwawelu, U. N., & Udechukwu, F. C. (2021). Effect of PID tuned digital compensator on servo-based ground station satellite antenna positioning control system. *2nd International Conference on Electrical Power Engineering (ICEPENG 2021)*, 97 – 101.

[7] Eze, P. C., Ugoh, C. A., & Inaibo, D. S. (2021). Positioning control of DC servomotor-based antenna using PID tuned compensator. *Journal of Engineering Sciences*, *8*(1), E9 – E16. https://doi.org/10.21272/jes.2021.8(1).e2

[8] King, D. (2010). The benefits of supply air temperature control in the datacenter. Future Facilities Limited, Aug., 2010.
[9]Purwanto, F. H., Utami, E., & Pramono, E. (2018). Design of server room temperature and humidity control system using fuzzy logic based on microcontroller. *2018 International Conference on Information and Communications Technology (ICOIACT)*, 390 – 395.

[10] Singh, A. K., Das, M., Basumatary, D., & Roy, G. (2014). A review note on compensator design for control education and engineering. *International Journal of Engineering Research & Technology*, *3*(2), 2493 – 2498.
[11]Zhang, Q., Meng, Z., Hong, X., Zhan, Y., Liu, J., Dong, J., Bai, T., Niu, J., & Deen, M. J. (2021). A survey on data center cooling systems: Technology, power consumption modeling and control strategy optimization. *Journal of System Architecture, 119*, 102253. https://doi.org/10.1016/j.sysarc.2021.1

SOFTWARE APPLICATIONS DEVELOPMENT FOR NATIONAL ECONOMIC GROWTH AND DEVELOPMENT: RELEVANCE OF METRICS IN SOFTWARE EVALUATION PROCESSES

Ikenna Caesar Nwandu[1], **Juliet N. Odii**[2], **Euphemia C. Nwokorie**[3], **Stanley A. Okolie**[4]
[1]Department of Software Engineering, Federal University of Technology, Owerri Nigeria.
[2,3,4]Department of Computer Science, Federal University of Technology, Owerri Nigeria.
ikenna.nwandu@futo.edu.ng, juliet.odii@futo.edu.ng, euphemia.nwokorie@futo.edu.ng*
stanley.okolie@futo.edu.ng

Abstract

There are quite a number of software metrics utilized by software developers in varying ways in their quest to quantifying their software products and the development methodologies in use. The use of these metrics often give different rank orderings of the same software attribute. This is why it is always a difficult task to make comparisons across software products. However, the use of metric in test-driven software development processes is essential to the software developer and other stakeholders owing to the fact that metrics increase the chances of discovering illuming problems regarding the software design and system performance. Unfortunately, the analysis of the relevance of metrics in a real-time software measurement and evaluation processes is often omitted. Such an evaluation should not be overlooked because it is an essential avenue to understanding the conditions at which metrication can be done, identifying the aspects of the software that require possible improvements, and exploring potential problems that would be detected by the metrics. In that regard, this paper makes an exploratory study of the relevance of metrics in test-driven software evaluation processes. The study formulated and used a six-step evaluation model to carry out the analysis with the aim of divulging the role of metrication in the quantification of software attributes.

Keywords: Software metrics, Test-driven development, Metrication, Measurement, Evaluation

1. INTRODUCTION

The development of good software products via competent software process or processes is surely meant to be coordinated via the quantification of the methods employed in the software development (Dumke & Foltin, 1996) [1]. The principal ideas of quantification is derived from the measurement and metrication of the development methodologies. However, there is no single metric that can give a universally valuable quantification. This is because different metrics may give different rank orderings of the same software attribute, thereby making comparisons across software products difficult and uncertain. This scenario calls for an objective definition and analysis of metrics. Discouragingly, the analyzed metrics are not easily validated. This is why test-driven developers are usually faced with the challenge of choosing and applying proper metrics to measure software attributes. However, evaluation and validation of metrics is very important to enable the software developer gain some insights as regards when a particular metric can be applied. The software developer also uses the process to identify any illuming problem(s) detected by the metric and find possible ways of defining potential solutions. To achieve this aim, it is important for the software developer to employ the metrics in a manner that they express the software attributes in numbers, a phenomenon known as measurement.

Measurement, as it were, assesses situations, tracks progress, evaluates effectiveness, and ultimately gives a pointer to the efficiency of software product. Software measures do not only play beneficial role in controlling the process of development but also maintains an excellent presentation of the ultimate product (Tutorialpoint, 2020) [2]. The measurement and metrication

of a software methodology is achieved through software testing. Software testing serves as a mutual support to guaranteeing the efficiency of the software product in a test-driven development environment. Software testing does not only focus on identifying the defects inserted in the earlier software developing phases. Rather, it is essential in demonstrating, validating, and certifying the absence of the defects and this purpose is usually achieved by employing the role of metrics. In software testing, the employed metrics increase the chances of discovering the problems as they directly affect software design and functionality, thereby giving a satisfactory condition which indicates that the system functions in conformation to all requirements (Nwandu & Asagba, 2017) [3]. The results of the testing process is essential to the software developer. The essentiality anchors on the fact that they are considered as fundamental factors in taking further decisions to be implemented during the course of completing the software development process. This is a pointer towards the assertion given by Meneely et al. (2012) [4] which describes a metric as an action-filled entity if it is able to guide the software manager in making empirical decisions with respect to the software product's status". In other words, a key essence of testing is to assess the functionality of software products. Hence, it is necessary to identify and apply appropriate metrics that will enable the measurement and evaluation of the assessed attributes of software products as well as promoting their development processes and tools.

2. DEFINING SOFTWARE METRICS

It is often said that the ability to measure something spoken about is an indication of an in-depth knowledge about it. However, being able to measure it and give a quantifying expression of it, is more satisfactory. A test-driven software development process is fashioned in such a manner as to measure certain attributes of interest vis-à-vis the development tools in use. This process is made possible via some units of measurement commonly known as metrics. Metrics enable the expression of software properties in numbers. They provide basis of measuring various aspects that describes software processes as well as the products that results from the processes. Different schools of thought have given several definitions for "Software Metrics". Some of them are presented in this section:

"A software metric is a measure of software characteristics which are quantifiable or countable." (stackify, 2020) [5]

"Software metric is defined as a quantitative measure that helps to estimate the progress, quality, and health of a software testing effort." (Guru99, 2020) [6]

"A software metric is a standard of measure of a degree to which a software system or process possesses some property." (Wikipedia, 2020) [7]
"Software metrics are a way of putting a value/measure on certain aspects of development allowing it to be compared to other projects." (Wikiversity, 2020) [8]
"The software Metrics can be defined as the continuous application of measurement-based techniques to the software development process and its products to supply meaningful and timely management information, together with the use of those techniques to improve that process and its products." (Spec-india, 2020) [9]
"Software Measures can be understood as a process of quantifying and symbolizing various attributes and aspects of software." (Tutorialpoint, 2020) [2]
Guru99.com provides a concise but more formal definition:

> "Software metrics is the quantitative indication of extent, capacity, dimension, amount or size of some attribute of a process or product."

3. RELATED WORK

Software metrics is a terminology that involves numerous activities, whose overall aim is to provide software measurement at some degree (Fentom and Bieman, 2015) [10]. There are three main components identified as driving forces on the measurement phase of software development (Dumke and Foltin, 1996) [1]. These components are the *process measurement* for understanding, evaluation and improvement of the development method; the *product measurement* for the quantification of the product (quality) characteristics and validation of these measures; and the *resource measurement* for the evaluation of the supports (CASE tools, measurement tools etc.) and the chosen implementation system. Dumke and Foltin gave a general class hierarchy for choice of metrics which included the process, product, and resources measurement with the following contents:

i. Process metrics – maturity, life cycle, and management.
ii. Product metrics – size, structure, architecture, complexity, and quality.
iii. Resource metrics – personnel, software, and hardware.

Bouwers et al. (2013) [11] carried out an analytical study for two architecture level metrics, Component Balance and Dependency Profiles. The study was centered on analyzing the challenges involved in applying Component Balance and Dependency Profiles in an industrial setting. These metrics inspired them to explore the usefulness of the metrics via the view-point of some experienced quality assessors. Bouwers et al. introduced a methodology for evaluating software metrics in industry with much reflection on its usefulness and further opined that the

relative importance of challenges involved in applying specific metrics cannot be determined in a purely academic setting. Bouwers et al. outlined four-step methodology to understudy the usefulness of the Component Balance and Dependency Profiles metrics as shown in figure 1. In the first stage, records of real-world experiences are taken by some designated observers in the form of memos. In the second stage, more data are gathered via interviewing assessors with a view to determine the perceived usefulness of the metric as observed by the assessors. The third stage concludes the procedure whereby the data extracted by both methods is analyzed and condensed separately.

An evaluating framework for proposed metrics was designed by Kaner and Bond (2004) [12] with major emphasis on analysis of "direct" measurement of an attribute. Kaner and Bond applied the framework to bug count as a code metric, underlying its uses and its ability to capture only a small part of the meaning of the attributes it is being used to measure.

An exploratory evaluation of measurement and metrics and their role in quantifying software quality was carried out by Ikerionwu and Nwandu (2021) [13]. The study aimed at identifying relevant software attributes and following a systematic procedure to measure the appropriate metrics. The study concluded with the assertion that metrication establishes relationships among attributes with an added advantage of simplification of users' understanding about software attributes and quality.

In this paper, we present an exploratory study of the relevance of metrication in test-driven software development in the context of metric-attribute relationship for proper quantification.

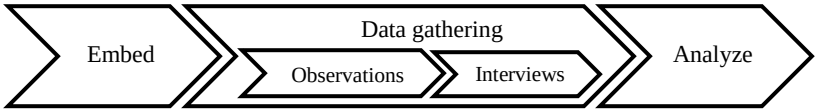


Figure 1: A four-step process for evaluating software metrics in practice (Bouwers et al., 2013).

4. FACTORS THAT INDICATE RELEVANCE OF SOFTWARE METRICS FOR QUALITY EVALUATION

To explore the relevance of metrics in test-driven software evaluation processes, six factors were identified – ***purpose, scope, target-attribute, methodology, metric-attribute relationship, and bottleneck*** **–** as basis for the evaluation. These factors are modelled as a surrounding-ring that revolves around software metrics with the view to emphasizing their relevance in software engineering (see figure 2). Key underlying queries were also identified to provide more expositions that would buttress the evaluating factors as shown in table 1.

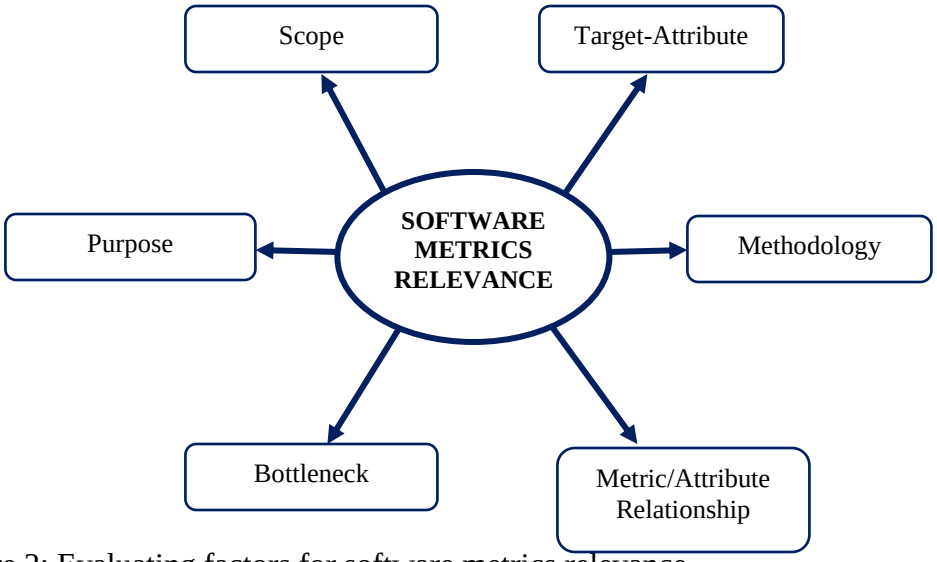


Figure 2: Evaluating factors for software metrics relevance.

The six factors for evaluation as identified are indicators that software metrics play the leading role of determining the state of software systems. The state would embrace the software's scope of functionality, efficiency of purpose, reliability of the attribute measures, usability of the target system, maintainability of system viability, and portability assurance. This would provide satisfaction that the system performs according to requirements and specifications and further gives the courage that the system would not cause havocs during its implementation and deployment.

5. METHODOLOGY

This study follows an exploratory methodology in the course of revealing the role of metrics in rolling out quality software products. The following six questions provide useful insight to the actual relevance of metrics in software measurement and evaluation process:

What is the purpose of metrication?

The purpose for metrication may vary in different quarters depending on the project and methodologies in use. We generally observe the purpose of metrication in software evaluation process as to:

i. propose an efficient method that ensures quick approach to measurement of software development.

ii. map out goals and put structures in place to measure performance with a view to achieving the goals.
iii. Gather good idea of the current product or process quality.
iv. determine the modules in the product/process that require improvement.
v. implement improvement plans in the quality of the product or process.
vi. make predictions on the quality of software at the completion of its development.
vii. make managerial and technical decisions for subsequent phase of activities.
viii. make managerial and technical decisions on whether to adopt a change in the process or the technology utilized.
ix. increase software usage and applications as well as user experience.
x. measure and evaluate the impact of decisions made during software development.

What is the scope of metrication?

The deployment of software metrics usually spans through various software development activities. These activities are meant to collectively contribute to software quality, hence their relevance in evaluation processes. Such activities include but may not be limited to the following:

i. Estimation of cost and effort.
ii. Measurement and modelling of productivity level.
iii. Collection of appropriate data.
iv. Modelling and measurement of data quantity.
v. Modelling of system reliability.
vi. Modelling and evaluation of system performance.
vii. Application of metrics that has inclination to system structure and complexity.
viii. System assessment to assure maturity and other capabilities.
ix. Applying metrics to ensure proper management of the system.
x. Evaluation of methods and tools to ensure their usability.

In view of the numerous potential areas of deployment of software metrics, the scope of metrication of software evaluation process can generally indicate that:

i. Some property of software or process can be accurately measured.
ii. There is a relationship between what can be measured and what is intended to know.
iii. This relationship is subject to validation, and can be expressed as a model or formula.

What are the target-attributes for metrication?

In order to determine software attributes that can hold metric value, the fact that attributes are given either number or symbol value would be considered. It is also important to uphold the fact that while a metric may rate at the top for both correlation with quality and potential benefit of an attribute, it may be time-consuming or expensive to determine its numerical value. However, from engineering perspective, studies give the idea that:

i. The attribute of every interesting thing is continually sought to be measured.
ii. All world entities are intended to be measured by software engineering practice
iii. Entities that can be already measured are intended to be improved upon.

What methodology do we use in metrication?

The method of applying metrics is essential in order to achieve its intended objectives. The following procedure was identified for achieving metrication in software evaluation processes:

i. Identify the metrics
ii. Define the identified metrics
iii. Introduce all the relevant metrics based on project needs
iv. Analyse the cost benefits aspect of each metrics
v. Define the goal for metrics
vi. Explain the need for metrics to stakeholders and testing team
vii. Educate the testing team about the data-points needed to be captured for processing the metrics
viii. Capture and verify the data
ix. Calculate metrics value using captured data
x. Develop an effective report and draw reasonable conclusion
xi. Release the report to the stakeholders
xii. Analyse feedback from stakeholders

What is the relationship of the metric to the target-attribute?

It is essential that software metrics should have relationship with what is to be measured. The ability of metric to divulge an attribute depends on the following:

i. Metric value should represent an attribute.
ii. When manipulated, metric value should preserve the relationship among entities of the attribute.
iii. The closer the relationship between entities, the closer the metric value to perfection.

What are the possible bottlenecks to metrication?

Calculating and understanding the value of a single, overall metric for a software product may be more trouble than it sounds. The underlisted assertions are possible challenges to metrication in software measurement and evaluation:

i. There are multiple definitions to software metric terms and they also differ in ways of counting or measuring attributes/characteristics
ii. The methods in which software metrics are used can be leveraged to deliberately avoiding the treatment of some problems that would have exposed interesting shortcomings.
iii. Software developers can be distracted from interesting goals (such as increasing customer satisfaction) due to oversimplification of software metrics.
iv. Software developers may run the risk of having too much data which may lead to a deviation from emphasising on the software metrics that help deliver useful software to customers.
v. The definition and derivation of software metrics are usually based on non-standardized assumptions.
vi. Terms that describe software metrics may depend upon tools available and working environment, paying less attention to the software nature and need for evaluation.
vii. Justification of most software metrics are usually based on historical data whose validity poses a lot of difficulty in its verification.

Table 1: Evaluating factors and corresponding underlying questions.

EVALUATING FACTOR	UNDERLYING QUESTION
Purpose	Why is a metric chosen?
Scope	What range of software activities can a metric assess?
Target-Attribute	What attributes are measurable?
Methodology	How is the metric applied?
Metric-Attribute Relationship	How efficient is a metric to revealing an attribute?
Bottleneck	What are the challenges to metric deployment?

6. CONCLUSION

This paper designed and described a six-step model for the evaluation of the relevance of metrics, in the context of metric-attribute relationship. The primary aim of this evaluation is to ensure a more proper metrication and quantification of software characteristics or processes.

Using an exploratory method, an overview of how, why, when, what, and which metric should be used, as well as the challenges of metrication are analyzed. This study is meant to provide a guide to software developers in deciding which metrics can be used in their software development, measurement and evaluation processes. This study also exposes developers and testers to determining when and how to use the metrics as appropriately identified. This study can also serve as a basis for conducting further research for scholars who may intend to perform similar types of studies. Future work can also gear towards improving the research method.

References

[1] Dumke, R. R., & Foltin, E. (1996) Metrics-based Evaluation of Object-Oriented Software Development Methods. University of Magdeburg, Germany

[2] Tutorialspoint. (2020). Software Requirements. https://www.tutorialspoint.com/software_engineering/software_requirements.htm

[3] Nwandu, I. C., & Asagba, P. O. (2017). Automated Software Testing For Reliable System Development. The International Journal of Science and Technoledge, 5(12), 44-49.

[4] Meneely, A., Smith, B., & Williams, L. (2012). Validating software metrics: A spectrum of philosophies. ACM Transactions on Software Engineering and Methodology (TOSEM), 21.

[5] Stackify. (2020). What Are Software Metrics and How Can You Track Them? https://stackify.com/track-software-metrics/

[6] Guru99. (2020). Software Testing Metrics: What is, Types & Example. https://www.guru99.com/software-testing-metrics-complete-tutorial.html

[7] Wikipedia. (2020). Software Metric. https://en.wikipedia.org/wiki/Software_metric

[8] Wikiversity (2020). Software Metrics and Measurement. https://en.wikiversity.org/wiki/Software_metrics_and_measurement

[9] Spec-india. (2020). Software Metrics and its Applications. https://www.spec-india.com/blog/software-metrics-and-its-applications

[10] Fentom, N., & Bieman, J. (2015). Software Metrics: A Rigorous and Practical Approach – Third Edition. CRC Press, Taylor & Francis Group, LLC

[11] Bouwers, E., Deursen, A., & Visser, J. (2013). Evaluating Usefulness of Software Metrics - an Industrial Experience Report. Software Engineering Research Group, Department of Software Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands.

[12] Kaner, C., & Bond, W. P. (2004). Software Engineering Metrics: What Do They Measure and How Do We Know? Florida Institute of Technology, Melbourne. Proceedings from 10th INTERNATIONAL SOFTWARE METRICS SYMPOSIUM (METRICS 2004), 1-12.

[13] Ikerionwu, C., & Nwandu, I. C. (2021). Quantifying Software Quality in Agile Development Environment. Software Engineering, 9(2), 36-44. doi: 10.11648/j.se.20210902.11

OPTIMIZED MOBILE PHONE VIRTUALIZATION USING HYPERVISOR TECHNIQUE

***** Iwuchukwu V. C, Nwokorie E. C., Okpalla C. L., Obi U. M., Egwu S. A.
Department of Computer Science
Federal University of Technology Owerri,
Imo State, Nigeria
vitalis.iwuchukwu@futo.edu.ng, euphemia.nwokorie@futo.edu.ng, lilymmao@yahoo.com., martinsuche.obi@futo.edu.ng, stellaegwu313@gmail.com

ABSTRACT
Mobile virtualization is an approach to mobile device management in which more than one virtual operating systems (platforms) are installed on a single wireless device. With the recent increase in mobile phone usage and over dependence of people on their phones these days to do virtually everything, mobile phone virtualization is now hugely seen as an alluring and attractive option because it gives both the users and manufacturers flexibility to address both privacy and security issues. To obtain an increment in sustainability and adoption in "Bring Your Own Device" (BYOD) methods within an enterprise, virtualization provides a viable option. In this paper, Component-based methodology was adopted in the design and implementation of mobile phone management and security using hypervisor virtualization. Each platform on the mobile device uses a hypervisor to provide access to a particular user with the use of some security authentication method. With this hypervisor virtualization system, two or more operating systems such as Android and Blackberry can run on the same phone.
Keywords: Virtualization, Hypervisor, Operating System, Virtual Machines, Bring-Your-own-Device.

1. Introduction

A mobile platform is an interface on a mobile phone with a dedicated operating system used for some specific purposes. It is an underlying computer system on which application programs can run. Most smartphones are not secured mainly because they have a single platform for every activity; and when an attacker strikes, important documents are affected. Consider a business mobile platform for example; segmenting mobile devices into personal and business realms will let employees do whatever they want on their side of the border while the employer retains control on the other. Applications and data in the two environments would remain isolated and protected. Mobile virtualization is an approach to mobile device management in which two or more virtual platforms are installed on a single wireless device (Jaramillo et al., 2014). It allows

one device to run two or more different operating systems, allowing the same phone to run both Android and Blackberry for example. Mobile virtualization is still at its infancy and being driven by security, cost reduction and end-user experience (Asokan et al., 2014). This paper undertakes the design of a secured mobile phone management system using hypervisor virtualization.

2.1 Literature Review

The number of mobile devices in the market is rapidly increasing, with the sale of tablets rising by 23.3% and smartphone sales rising by 39.2% in 2013 (Gartner & Meulen, 2014). In the near future, growth forecasts are very positively directed. In recent years, mobile devices are becoming more and more intelligent as applications and computing functions are moving from the personal computers and laptops to smaller and smarter devices such as the phones and tablets. As smart devices become the most important entry point to a wide variety of data ranging from private photos to large data sets required for different applications, both the end user and the providers of services that can be accessed via these devices have an increasing interest in protecting this data.

2.2 Mobile Device Platforms

The Android is growing heavily in recent years. Even with this growing Android trend, the mobile community can hardly be defined by one particular company. Other brands like Apple changed the name of its iPhone Operating System (OS) to iOS, RIM released a new version of its Blackberry OS and Microsoft had to also go back and release Windows Phone 7 (Warren, 2010). Former Apple Executive, Jean–Louis Gassee said that operating systems do not matter anymore: that what matters now is the user experience and development tools, stressing that what is described as OS now is really simply more of a platform. Different people have sufficient background about different mobile devices and their manufacturers, but a very few of them know much about the operating systems they use. It is important to learn and know about the different mobile OSs used by different companies so as to know the idea behind their smooth platforms (Guru, 2012).

a. **Symbian:** The Symbian OS is now officially owned by Nokia. This only implies that any other company will have to acquire permission from Nokia before using it. Before now, a couple of years ago, Symbian OS was the most used in the mobile phones and although it is still used in low-end phones, the demand rate has gone down drastically.
b. **Android:** In September 2008, Google introduced the first Android OS by the name of ‘Astro’. Later on, other upgrades versions ‘Bender’ and ‘Cupcake’ were also released. Typically, Google adopted the trend of naming Android versions based on a dessert or sweet in alphabetical order. For example, Donut, Éclair, Froyo, Gingerbread,

Honeycomb, Ice Cream, Sandwich, Jelly Bean, Marshmallow (Android 6.0) right to the latest which is called Snow Cone (Android 12.0).

c. **Apple iOS:** The Apple iOS was first introduced in 29th June, 2007 when the first iPhone was developed. Since then iOS has been undergoing many upgrades and currently the latest one is the iOS 15. Apple has still not allowed any other manufacturer to lay hands on its operating system. Unlike Android, Apple has concentrated more on the performance along with appearance.
d. **Blackberry OS:** This operating system is the property of RIM (Research In Motion) and was first introduced in 1999. RIM developed Blackberry operating system for its Blackberry smartphones. Also, Blackberry is very much different from other operating systems in its interface style as well as the smartphone design which has a trackball for movement or directions. RIM has not produced any new versions of Blackberry since 2013 as the most current version of Blackberry operating system is version 10 which was released in January 30, 2013.
e. **Windows Phone OS:** The Windows operating system is very widely known because it is also used on computers all over the world. The Windows Phone OS was introduced by Microsoft in collaboration with Nokia for use on its Windows phones. Many users of the devices found it a little bit difficult to operate it but then, it was very popular with people who were used to it.

2.3 Mobile Phone Architecture

No specific standards exist to define what makes a phone a smartphone. While some phones are GSM / CDMA (Global System for Mobile Communication / Code Division Multiple Access) mobile phones, others run high – level operating systems (Fallows & Ganson, 2009). Fallows and his companion opined that mobile architecture is based on the ARM (Advanced RISC Machine) and Snapdragon processors while the system's chip and multi - core architectures are relatively new developments. On the other hand, mobile devices contain a fairly large amount of circuitry, with each diligently designed to maximize its performance. The mobile phone has both analogue and digital circuits within it ranging from processors through keypad to display electronics. A typical mobile phone is made up of a single board, though there are a number of distinct functional areas within it. (Gurupadappa, 2015)

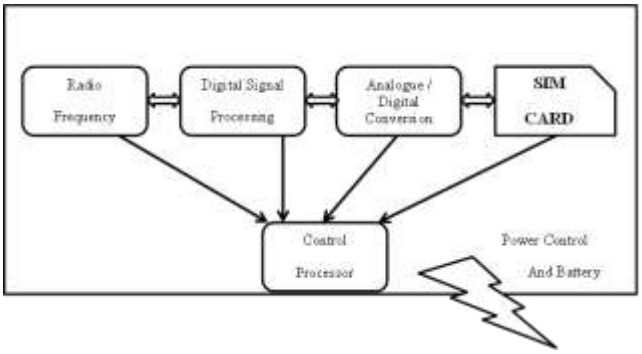


Figure 2.1: Typical Mobile Phone Architecture and Circuitry

2.4 The BYOD Idea

The phrase 'Bring Your Own Device' is closely tied to virtualization in many ways. As Information Technology units in different companies try to manage to handle incremental technological variations, workers in different spheres of life struggle to use their own mobile devices to gain entry into corporate information. Similar to Bring Your Own Network (BYON) and Bring Your own Computer (BYOC), BYOD has changed over time to enhance the workforce through a term known as "Consumerization" of IT. It is a phrase that is used when employees are allowed to bring their own personal devices to the workplace and use them to gain access to corporate information. This came about because almost company employees own a device like mobile phones, tablets, laptops, etc.

2.4.1 BYOD and Security

BYOD institutes a major IT security concern to employers since employees are given access to important data using their personal devices. Mostly, BYOD security policies are put in place to guide employees so as to streamline what they do with their devices in the workplace and to make sure that the security of the network is not compromised (Dern, 2014). Some of the security measures that should be put in place when implementing Bring Your Own Device include;

a. Using a virtual environment to separate personal operating system from work operating system.
b. Instituting clear and detailed security authentication methods (like passwords for every device that is used in the place of work and also having access to the network.

c. Security policies should be put in place to manage how these personal devices are used in the place of work like checking the types of applications installed, limiting use of email to corporate ones only and scheduled auditing of such devices to ensure there are no security breaches.

In all these, the virtualization method remains the key as it gives the employee a means of effectively managing the device to suit personal and corporate lifestyle.

2.5 The Concept of Virtualization

Virtualization can be defined as an aspect of mobile device management through which two or more virtual mobile platforms are installed on a single mobile device (Rouse, 2011). It was specified that the device should be wireless to run a virtual machine. A tablet or smartphone may have two virtual platforms for different purposes like business and personal use. Application on both environments will be isolated, inducing high level of security and management of data. Also, mobile virtualization can allow a single mobile device to run two or more operating systems such as Blackberry and iOS applications.

This blend of virtualization technologies - or virtual infrastructure - provides a layer of abstraction between computing, storage and networking hardware, and the applications running on it (see Figures 2.2 and 2.3). The deployment of virtual infrastructure is non-disruptive, since the user experiences are largely unchanged.

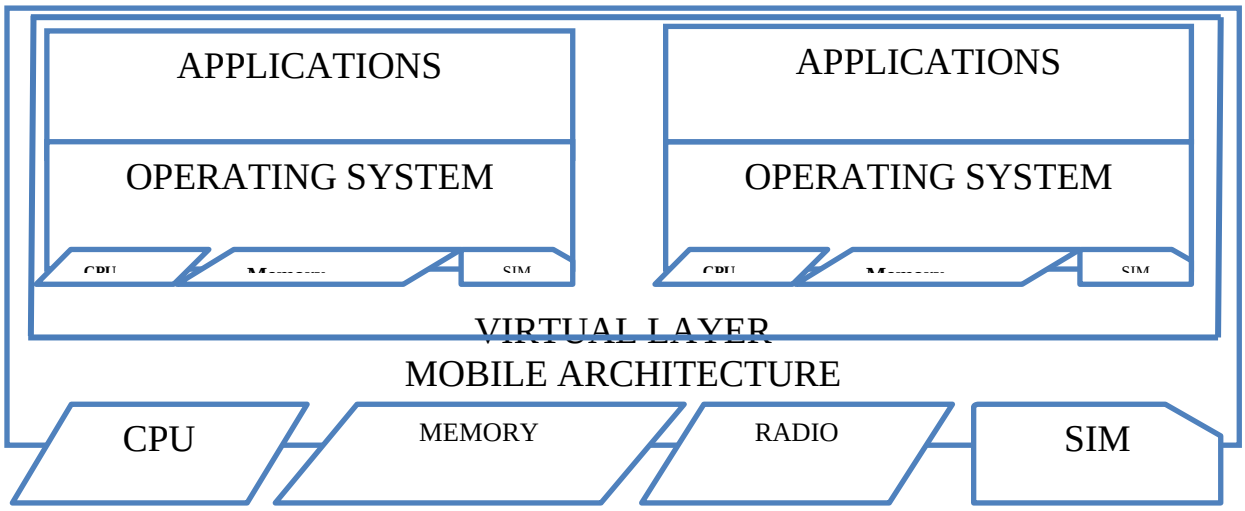


Figure 2.2: Before Virtualization

Figure 2.2 shows that there is a single operating system per device with the hardware and software tied together. On the other hand, figure 2.3 shows a device after virtualization whereby a virtual machine (or layer) installed in a single mobile device which creates a virtual platform with two operating systems, different applications on each platform.

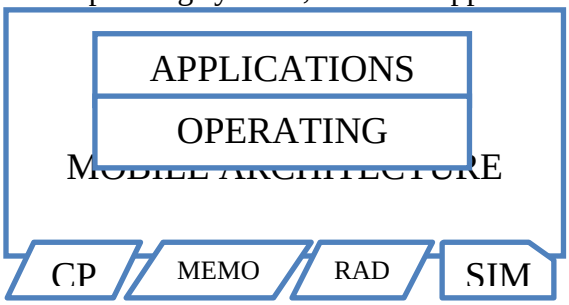


Figure 2.3: After Virtualization

2.5.1 Virtualization Methods

Marko (2013) described three methods of virtualization which are hypervisors, sandbox containers and encryption whereas Brockmeier (2010) only mentioned the hypervisors and containers which are the most popular virtualization methods.

a. Hypervisor Virtualization

A hypervisor (which is also called a virtual machine monitor) is an entity or a function that differentiates operating systems and applications from an underlying device hardware. This differentiation ensures that the underlying machine or device (called the host) operates one or more independent virtual machines as guests, ensuring that a number of guest virtual machines do share effectively the physical resources of the mobile device such as memory space, network bandwidth, SIMs and even processor cycles (Rouse & Bigelow, 2016).

A hypervisor can also be defined as a firmware, hardware or software on a device that develops and runs virtual machines. The term 'Hypervisor' was coined from 'Supervisor', which was initially used to describe the kernel of an operating system: As such the hypervisor acts as the supervisor of the kernel of an operating system (Golden, 2011).

i. **Bare – metal Hypervisor (Type 1):** This type is installed right behind the device hardware, underneath the operating system. As such, it requires extensive device hardware support which is a major problem on mobile devices that do have an equivalent to the standard x86 hardware platform. Bare metal hypervisors manages the guest operating systems and controls the mobile hardware.
ii. **Hosted Hypervisors (Type 2):** Hosted hypervisors run on a conventional operating system just as other computer programs do. A guest operating system runs as a process on the host. This is a difficult task to accomplish because mobile device operating systems are closely tied and controlled (like Apple products) and also hugely manufactured to handle some specific hardware configurations (Android is notorious here).

b. Application Sandbox Virtualization

This type of virtualization is also called "container-based" virtualization because it is like a container to hold and manage user programs. A sandbox, in Computer Security terms, is a mechanism for security put in place to separate running applications or programs. It is mostly used to run third party or untrusted applications from unverified sources or websites to avoid the risk of harming the device or operating system (Goldberg et al., 2011). It helps to provide a virtual space to control how guest programs run and how resources are used, as well as restrict access to certain guest programs or applications.

A sandbox is a tightly controlled platform where programs can be run. It helps to control what an application can do, letting it have just as many permissions as it requires avoiding extra permissions that can be misused (Hoffman, 2013).

2.5.2 Implementation of Sandbox

Implementing the Security Policy using a System Sandbox Access to the individual system resources is implemented by means of a unified interface provided by the operating system. Therefore, the system sandbox must provide access control of the individual OS components within this security implementation (Vokorokos et al., 2015). The most important OS components here are the files, the registry, network interfaces, the CPU, Input / Output, locks and processes.

c. Encryption Method

This method of virtualization offers encryption of certain software, hardware and applications of a device. Encryption can be defined as the process of providing security to the user's data in such way that only authorized users can access it (Akkinapalli & Rao, 2014). The several possible encryption techniques used for securing user's data in the cloud are given as follows

1. **Attribute Based Encryption (ABE):** This is a technique of cryptography in which encryption and decryption of a device is hinged on the user's characteristics and assess policies are defined based on these features. These policies can be classified in two ways which are the process of the attribute based encryption and the key policy ABE.
2. **Homomorphic Encryption:** This cryptographic strategy is more complex with mathematical computations performed on encrypted data without decrypting them using the user's private key.
3. **Cloud Computing Confidentiality Framework:** This is another technique that has to do with laying out clear objectives and goals of set up. A detailed analysis of impact is performed to identify the processes within the system. Also, the framework covers the data sets that need special protection and valuable information that are placed under the classified list.

2.5.3 Advantages of using Hypervisor

With these different virtualization methods, it is important to choose any that best suits the purpose of use of the device. It is worthy to note that encryption method is rarely used as it does not provide enough security to the users of the device. With sandapps, installing a new OS is not really too straightforward compared to the hypervisor method, like installing directly from the device memory. Sandapps will require a container to be created first before installation of a new operating system, and they can be bypassed easily since every file related to the container is resident on the device's RAM whereas hypervisors are closely tied to the hardware. So, it suffices to say that the hypervisor method provides more security and makes installation of a new operating system easy to execute.

3.1 Current Systems of Hypervisor Virtualization

Most times, a number of end-users are entitled to use a single device especially in a public environment and this can increase vulnerabilities. Different reasons have been highlighted for the loopholes in using a device collectively.

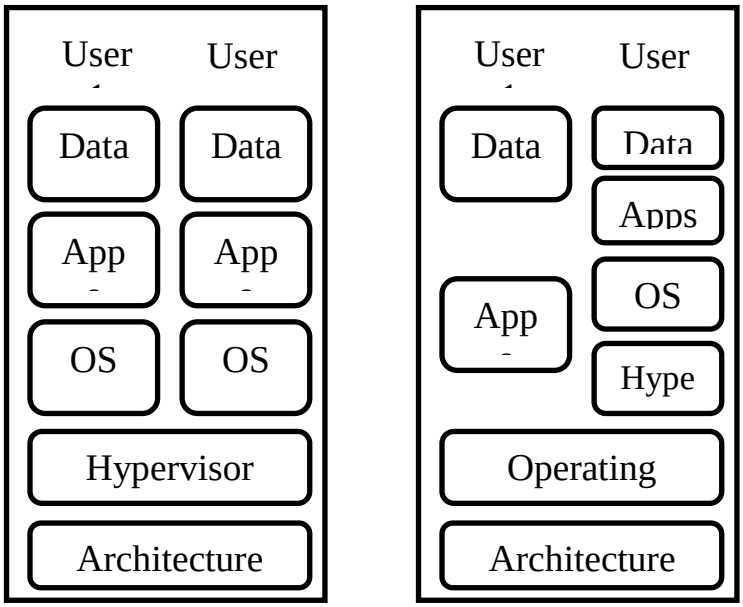


Figure 3.1: a) Type 1 – Hypervisor BYOD **b) Type 2: BYOD - Hypervisor**

Figure 3.1 shows two types of typical models of hypervisor Bring Your Own Device (BYOD), showing the arrangement of the device resources in the presence of a hypervisor. Type 1(Figure 3.1: a) is a hypervisor BYOD where the hypervisor is used to virtualized or separate the two different OSs and other resources on each operating system while the device architecture is the same. Type 2 (Figure 3.1: b) is another current model called BYOD – hypervisor. Here, a resident operating system pre-installed in the device bears the hypervisor which makes it possible for a fresh operating system to be installed with its data and applications. The resident OS has its own data and applications that may be the same or different from the newly installed ones. Since the same device is being used for all these, the architecture is still the same.

3.2 The Proposed System Model

The model of the proposed system (Figure 3.2) is structured from the existing system with some adjustments. Some of the peculiarities of the proposed system are highlighted below;

a. **Security Authentication:** This model will be designed in such a way that each operating system will have security level where a user must log in before access is granted.
b. **More Users:** In this model, more than two users can be allocated different operating systems on the device.
c. **Dedicated Applications:** Different applications run on different operating systems and as such, a dedicated application that does not run on one OS can be installed on the platform that accepts it.

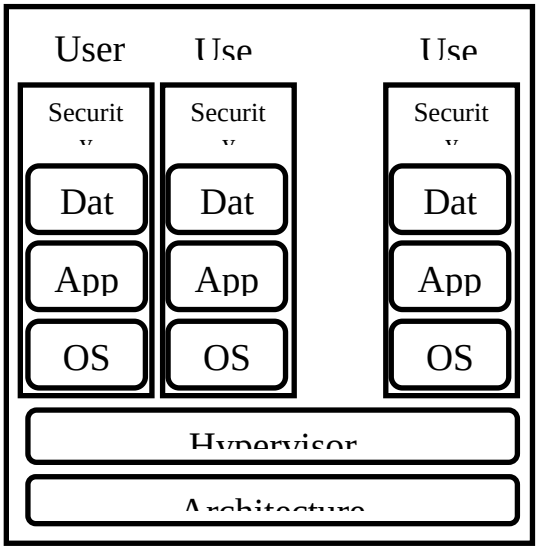


Figure 3.2: Virtualization Model of the Proposed System

3.3 The Designed System's Operational Methodology

The mobile device is booted to initiate the installation of the Basic Input and Output System (BIOS) of the device. A vital process in the design of the system is choosing the operating system to use by the user.

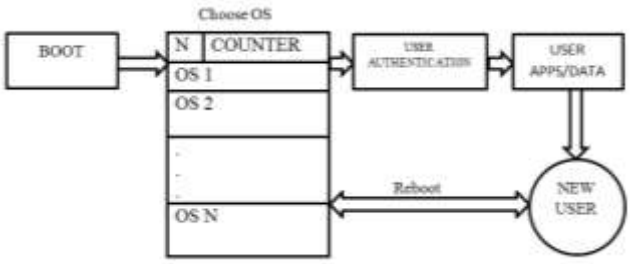


Figure 3.3: The proposed system's mode of Operation

The operating systems are listed and must have been preinstalled as specified by the customer or user. Then to authenticate the user for each operating platform, a security authentication method is used to ensure that the right person is logging into the system. At this point, the user can access the applications and data on that platform. If a new user is needed, the device is rebooted to create an OS for the person.

4.0 Conclusion

The idea of virtualization, though still at its infancy is gaining a lot of grounds in the IT area now. This work is just one of the ways of improving security and management of multi-user mobile devices using the virtualization technology and a security authentication method. The issue of data security has always been a major concern to mobile users especially where there is more than one user as obtainable mostly in corporate environments and workplaces. Also, the recent idea of BYOD policies allow employees to use their personal devices in the workplace thereby exposing important corporate information to security risks. Virtualization will help curb some of these risks to a large extent by segmenting the same device into separate platforms for different applications and uses.

REFERENCES

Asokan, N., Ekberg, J. and Kostiainen, T.K. (2014). Mobile Platform Security: Trusted Execution Environments. Summer School, Aalto University and University of Helsinki. 1 – 15.

Akkinapalli, K. and Rao, R. (2014). A Survey on Encryption and Improved Virtualization Security Techniques for Cloud Infrastructure. Global Journal of Computer Science and Technology: B Cloud and Distributed: Global Journals Inc (USA); Vol 14 Issue 2 Ver 1.0.

Brockmeier, J. (2010). Containers vs Hypervisors: Choosing The Best Virtualization Technology. Retrieved from www.linux.com/Containers vs. Hypervisors Choosing the Best Virtualization Technology.html.

Dern, D. P. (2014). How to keep your smartphone (and its data Secure. Retrieved from www.computerworld.com/How to keep your smartphone (and its data) secure.html.

Fallows, A. and Ganson, P. (2009). Smartphone Hardware Architecture. PDF: Fall site. 2 – 10.

Gartner, J. R. and Meulen R.v.d. (2014). Worldwide Traditional PC, Tablet, Ultramobile and Mobile Phone Shipments to Grow 4.2 Percent in 2014. Retrieved from http://www.gartner.com/newsroom/id/2791017

Goldberg I., Wagner D., Thomas R. and Brewer E. (1996). A Secure Environment for Untrusted Helper Applications (Confining the Wily Hacker) *(PDF). Proceedings of the Sixth USENIX UNIX Security Symposium.*

Golden, B. (2011). Virtualization For Dummies; Wiley, 3rd Edition. ISSN-13: 978-0470148310, ISSN-10: 0470148314. 20 – 50.

Guru, L. (2012). Top 10 Mobile Phones Operating Systems. Retrieved from http://www.shoutmeloud.com/top-10-mobile-phones-operating-systems

Gurupadappa, G. (2015). What is Mobile Phone Architecture? Retrieved from https://www.quora.com/What-is-mobile-phone-architecture

Hoffman, C. (2013). Sandboxes Explained. Retrieved from www.howtogeek.com/Sandboxes Explained How They're Already Protecting You and How to Sandbox Any Program.html.

Jaramillo, D., Furht, B., and Agarwal, A. (2014). Virtualization Techniques for Mobile Systems. Springer Journal, XIV, 73 p. 38 illus., Hardcover. ISBN 978-3-319-05740-8. 3 – 40.

Vokorokos, L., Anton, B., and Branislav, M. (2015). Application Security through Sandbox Virtualization. Acta Polytechnica Hungarica, Vol 12, No. 1. 84 – 100.

Marko, K. (2013). 3 Ways to Virtualize Mobile Devices, and Why You Should Do So. Retrieved from https://darkreading.com/risk-management/3-ways-to-virtualize-mobile-device---and-why-you-should-do-so/d/d-id/1110613

Rouse, M. (2011). Mobile Virtualization. Retrieved from www.whatis.com/What is mobile virtualization.html.

Rouse, M. and Bigelow, S.J. (2016). Hypervisor. Retrieved from www.whatis.com/What is hypervisor.html.

Warren, C. (2010). 5 Platforms that Defined the Mobile Space in 2010. Retrieved from www.mashable.com/5 Platforms that Defined the Mobile Space in 2010 [Mashable Awards].html.

CYBERSECURITY AND ITS AWARENESS FOR SUSTAINABLE AND CONDUCIVE ATMOSPHERE FOR NIGERIA ECONOMY

Oladimeji S. A[1]., Madu F.U[2]., Emeagi I.O[3]., Ezurike O[4] & Luke-Odoemena I.
Department of Computer Science, Federal Polytechnic Nekede, Owerri[1,2,3,4]
donsedof@gmail.com,

Abstract
The wide use of digital media is making attackers smarter by the day. It is not new to hear that despite the tremendous benefits of cyberspace, some criminally-minded individuals are taking undue advantage of cyberspace to perpetrate evils. The risk and severity of cyber attacks have clearly grown over the past few years. Since 2018, we have witnessed the most horrific cases of cybercrimes related to massive data breaches, flaws in microchips, cryptojacking, spamming and many others. Governments are responsible for protecting national security and public welfare. As Nigeria is tapping into the potentials of digital revolution, Nigeria will have to attach high level of seriousness to the established laws that address cyber threats and hold perpetrators of cyber attacks accountable, establish organizations and programs that help with Cybersecurity, and allocate money for cyber-public awareness, defence research, and education. The federal government should start expending more funds on evolving research on Cybersecurity, Blockchain technology, IoT, etc through FG-sponsored programmes. Any country that pays less heed to Cybersecurity will severely pay for the disasters that follow. This paper discussed cybersecurity awareness, cyber-attacks in Nigeria and provided counter-measures to mitigate cyber-attacks.
Keywords: Cybersecurity, IoT, Blockchain, Cyberspace, & Cyber-attacks etc.

1.0 Introduction

The driving force behind cybersecurity is the threat of cyber attacks. Each level of a cyber-physical Infrastructure which consists of operational software, information, and people is susceptible to security breakdown, whether through attack, data breach, infiltration, or accident. Cyber threats are asymmetric because they allow few individuals to perpetrate attacks upon the masses. Through Inter-connected computer, a belligerent cyber actor may conduct a cyber attack with minimal technical and operational resources. With a minimal chance of failure, cyber attacks offer a high return for a low financial investment. Because of the permeable nature of sophisticated networks, a cyber actor may infiltrate an adversary's network with minimal risk of discovery (Ajiji 2017) The increasing trend of ubiquitous computing with cyber threats is characterized by an attacker, a target system, a set of actions against the target, and the consequences resulting from the attack. Consequences include damages to the target, direct and indirect losses to victims, and variable impact on third parties. As cyberspace becomes increasingly pervasive and entrenched in society, it spawns the availability of more targets to attack, and an increase in the population of skilled attackers (Ajiji 2017) Defenders must familiarize themselves with the environment by understanding not only the cyber domain but also the human element, the attacker, their motives and goals. Consideration of the identified key components will provide greater fidelity to the orientation phase of the decision-making process.

Cyberspace has become the cornerstone of United States communication, commerce, military command and control, emergency services, mass transit, power plant distribution, and numerous other critical infrastructures essential to enabling and sustaining twenty-first century society. Cyberspace is a global domain within the information environment consisting of the interdependent network of information technology (IT) infrastructures, including the Internet, telecommunication networks, computer systems, and embedded processors and controllers. As societal dependency on information technology grows, so do cyber threats. A diverse group of Nation-states, non-state actors, state-sponsored groups, and individuals may wage malicious cyber-attacks on a target (Alkali 2010) Cyber and sabotage attacks on critical United States economic, energy, and transportation infrastructures may be viewed by invested adversaries as a way to circumvent United States strengths on the battlefield and attack United States interests directly at home. In support of the national security strategy, the nation must institute a multilateral strategic framework that focuses on the dynamic challenges of cyber in the Information Age.

2.0 Cyberspace

Physical space is the dimension most often associated with security. Physical space whether land, sea, or atmosphere, is demarcated into territories under the jurisdiction of sovereign state law. Throughout history, armies have been deployed across territories and bodies of water whether they were provinces, kingdoms, countries, or whole empires-in order to defend their own land or lay claim to other lands (in the name of security or national aggrandizement) Conversely; cyberspace is unconfined to a spatial dimension or effectively sanctioned by sovereign states or international law. Globally intercom1ected, cyberspace is a realm of digital information and communication that consists of decentralized computer networks with no single authority to supervise or regulate operation (Antonucci, 2017) In the past several years, cybersecurity has transitioned from an esoteric concept only comprehended by computer scientists and information system managers to a national security threat requiring the attention of the public and policy makers. President Barack Obama has declared America's digital infrastructure to be a "strategic national asset.

Through the course of developing national policies and strategies, cyberspace has become the fifth domain of warfare, after land, sea, air, and space. Cyberspace is a realm that is constantly growing worldwide and unlike the conventional domains, it cannot be rigidly demarcated into national boundaries or other territorial units (Azeez, 2019) Because of this idiosyncrasy, *The National Strategy to Secure Cyberspace* has emphasized that securing cyberspace is a global matter due to the interconnectedness of the world's computer systems. Conventional based policies, strategies, and initiatives from years past do- not directly address the new challenges and issues independently unique to the cyberspace domain, nor do they completely coincide with the legislation and agenda of foreign nation-states. Securing national cyberspace will require national cooperation to raise awareness, share information, promote security standards, and investigate and prosecute cybercrime (Aniekan, 2017).

3.0 The Cyber Attacker (The Human Element)

Technology is normally associated with cybersecurity; however the human element cannot be disregarded or ignored. United States Air Force colonel John Boyd argued that "Machines don't fight wars, Humans fight wars." A cyber threat is always given its existence from a human element. A wide spectrum of malicious cyber attackers exists from individual hackers, to criminal enterprises, to terrorist groups, to corporations, to nation-states. (Buchanan, 2016) Fundamentally, each attacker can be classified into two, a sovereign state or non-state actor. A non-state actor whose purposes are criminal and who is subject to the jurisdiction of one or more sovereign states includes hackers, criminal enterprises, terrorist groups, and corporations. Terrorists constitute a more serious set of non-state actors and are of concern both to law enforcement agencies and national security agencies. According to some analysts, as many as twenty countries have cyber-warfare capabilities, including China, Russia and North Korea. State actors normally target other sovereign states, although specific targets may be identical to those of non-state attackers (Buchanan, 2016). Unfortunately, these actors are not mutually exclusive and could amalgamate and create a customized threat. The attackers do not need to amass great arms; it can all be done covertly and cheaply, by hiring outside expertise.

4.0 Motives of Cyber Attackers

In general, an active or high-profile cyber attacker will have a motive and goal to attack a target. A motive for an attacker would be to conduct espionage, obtain monetary gains, and inflict malicious harm or further national or ideological interests. When a cyber attacker acts based upon a motive, at least one of the following goals are attempted (Cavelty, 2021) Knowing what motivates hackers is a key part of keeping them out of your business IT systems.

Financial Gain

The primary motivation of a hacker is money, and getting it can be done with a variety of methods. They could directly gain entry to a bank or investment account; steal a password to your financial sites and then transfer the assets over to one of their own; swindle an employee into completing a money transfer through a complicated spear phishing technique, or conduct a ransomware attack on your entire organization. The possibilities are endless, but most hackers are out to make a profit.

Recognition & Achievement

Some hackers are motivated by the sense of achievement that comes with cracking open a major system. Some may work in groups or independently, but, on some scale, they would like to be

recognized. This also simplifies the fact that cyber criminals are competitive by nature, and they love the challenge their actions bring. In fact, they often drive one another to complete more complicated hacks (Buchanan, 2016).

Insider Threats

Individuals who have access to critical information or systems can easily choose to misuse that access to the detriment of their organization. These threats can come from internal employees, vendors, a contractor or a partner and are viewed as some of the greatest cyber security threats to organizations. However, not all insider threats are intentional, according to an Insider Threat Report from Crowd Research Partners (Buchanan, 2016) Most (51%) are due to carelessness, negligence, or compromised credentials, but the potential impact is still present even in an unintentional scenario.

Political Motivation (Hacktivism)

Some cybercriminal groups use their hacking skills to go after large organizations. They are usually motivated by a cause of some sort, such as highlighting human rights or alerting a large corporation to their system vulnerabilities. Or, they may go up against groups whose ideologies do not align with their own (Buchanan, 2016) These groups can steal information and argue that they are practicing free speech, but more often than not, these groups will employ a DDoS (Distributed Denial of Service) attack to overload a website with too much traffic and cause it to crash.

State Actors

State-sponsored actors receive funding and assistance from a nation-state. They are specifically engaged in cybercrime to further their nation's own interests. Typically, they steal information, including intellectual property, personally identifying information, and money to fund or further espionage and exploitation causes. However, some state-sponsored actors do conduct damaging cyberattacks and claim that their cyber-espionage actions are legitimate activity on behalf of the state (Cavelty, 2019).

6. Corporate Espionage

This is a form of cyber-attack used to gain an advantage over a competing organization. It is conducted for commercial or financial purposes like:

- Acquiring property like processes or techniques, locations, customer data, pricing, sales, research, bids, or strategies.
- Theft of trade secrets, bribery, blackmail, or surveillance.

5.0 Cybercrimes and Cyber Laws in Nigeria

Nigerians have become cyber-creatures, spending a significant amount of time online. As the digital world expands, so does cybercrime in Nigeria. The necessity to combat these seemingly uncontrollable phenomena gave rise to Cyber Laws in Nigeria. Cyber law acts as a shield over cyberspace, preventing cybercrime from occurring. The government is committed to developing and enforcing regulations to combat illicit online activities. The "Cybercrimes (Prohibition and Prevention) Act, 2015" has a significant impact on cyber law in Nigeria. This Act creates a comprehensive legal, regulatory, and institutional framework in Nigeria to prohibit, prevent, detect, prosecute, and punish cybercrime. The Act also encourages cybersecurity and protection of computer systems and networks, electronic communications, data and computer programs, intellectual property, and privacy rights, as well as the protection of important national information infrastructure (Mordi, 2019)

Cybercrime

Cybercrime is a type of crime that takes place in cyberspace, or in the realm of computers and the Internet. Because our society is evolving towards an information society where communication occurs in cyberspace, cybercrime is now a global phenomenon. Cybercrime has the potential to significantly influence our lives, society, and economy.

Cyber Law

Any law that deals with the internet and similar technology is known as cyber law. Cyber Law is frequently referred to as "Law of the Internet" or "IT Law." It's a legal framework for dealing with issues relating to the Internet, computing, Cyberspace, and other associated matters. One of the newest aspects of the legal system is cyber law. This is due to the rapid advancement of internet technology (Mohamed, 2019). People who use the internet have legal safeguards under cyber law. This applies to both business and common citizens. Anyone who uses the internet should be familiar with cyber laws.

Intellectual property, contract, jurisdiction, data protection laws, privacy, and freedom of expression are all covered by cyber law. It oversees the distribution of software, information, online security, and e-commerce via the internet. E-documents are given legal validity in the field of Cyber Law. It also establishes a framework for e-commerce and e-filling. To put it another way, Cyber law is a legal framework for dealing with cybercrime. Due to the increased use of E-commerce, it is critical that suitable regulatory practices are in place to ensure that no malpractices occur (Mohamed, 2019). Cybersecurity laws vary a lot from country to country and jurisdiction to jurisdiction. Penalties depend on the nature of offence, and will range from a fine to imprisonment. It is critical for citizens to understand their particular countries' cyber laws in order to ensure that they are fully informed about all cybersecurity issues:

- Identity theft and impersonation
- Child pornography and related offences
- Cyberstalking
- Cybersquatting
- Racist and xenophobic offences
- Attempt, conspiracy, aiding and abetting
- Importation and fabrication of e-tools
- Breach of Confidentiality and Privacy
- Manipulation of ATM/POS Terminals
- Phishing, spamming, spreading of computer virus
- Electronic cards related fraud
- Use of fraudulent device or attached e-mails and website. (Nigerian Cybersecurity Act 2015)

6.0 Cases of Cybercrime in Nigeria

Cybersecurity threats and attacks have become rampant due to technology changes, social economic factors and inadequate criminal justice. Social media has a challenge to reporting cyber incidents as many people do not check the authenticity of news posted on social media. Cybersafe Foundation, a cyber-security awareness creator, has called for more awareness in tackling the growing menace of cybercrime in Nigeria. Speaking at a forum for cyber security experts, mostly from the financial sector, and IT journalists, in Lagos, Cybersafe urged more collaboration among various security entities in the country (Ameh, 2021) While it was earlier revealed at another forum that Nigeria had lost about N5.5 trillion to fraud and cybercrimes in 10 years, the experts at the Cybersafe forum, in their various presentations warned that cyber security threats and attacks are not going away, as the phenomenon could constitute the next pandemic, spelling out dangers to corporate bodies, government and individuals refusing to create barricades and walls for their platforms, digital tools and applications against cyber-attacks. Some of those, who spoke at the event include Mrs Favour Femi-Oyewole, Group Chief Information Security Officer, Access Bank, Abumere Igboa, Chief Information Security Officer, Stanbic IBTC, Dr. Obadare Peter Adewale, Chief Visioner at Digital Encode, Confidence Staveley, Cybersafe Foundation and Bharat Soni, Chief Information and Security officer at GTB Limited (Ameh, 2021)

Femi-Oyewole, while warning that cyber-attacks could be the next post-COVID pandemic, said it was important for organisations and individuals to begin to build resilience and back-ups for their systems, platforms and applications. She urged organizations to check their ability to bounce back, should they suffer any attack. “If anything happens to you, how quickly can you bounce back? Have you checked your resilience, do you have a backup?” she asked. Saying that

integrity, confidentiality and availability of a good cybersecurity system matters. "You need to put necessary measures in place to quickly detect breaches and remedy. Vulnerability is any flaw or weakness that can be exploited (Ameh, 2021)

"There is a need to put in counter-measures to prevent, minimize or report any breaches on time so that corrective measures can be taken immediately." According to her, the most important and first level of shield and line of defense against cyber-attacks is the human beings who should ensure that they do not open their systems and media platforms vulnerable. From his perspective, Soni, who listed the most recent cyber security breaches to include Twitter compromise of 2020, Colonial Ransomware attack 2021 and cyber breach of an undisclosed Nigerian Bank 2021, said organisations should work to mitigate cybersecurity challenges such as insider fraud, business email compromises, ransom ware and phishing (Ameh, 2021) According to him, cybersecurity threats and attacks have become rampant due to technology changes, social economic factors and inadequate criminal justice, adding social media has a challenge to reporting cyber incidents as many people do not check the authenticity of news posted on social media. Although we are highly regulated, we still need to know how to protect ourselves, while we enjoined IT journalists to adequately equip themselves with knowledge of trends in the cybersecurity ecosystem so they could help inform the public more accurately and actively.

Corps member bags two months for cybercrime

Justice Inyang Ekwo of the Federal High Court, Abuja, recently sentenced a member of the National Youth Service Corps (NYSC), Ajayi Temitope Ayokunle to two months imprisonment without option of fine. The 30-year old corps member, who posed as Dr. Joshua to defraud an American citizen, Maria, of $1,000 will in addition to his imprisonment, forfeit a cash sum of $500 found in his bank account and a sophisticated android telephone to the Federal Government. Justice Ekwo, who expressed utter disgust at the embarrassing rate of cybercrime among Nigerian youths, turned down the plea bargain entered by the convict with the Economic and Financial Crimes Commission (EFCC), wherein, he confessed to committing the crime and requested for a soft landing (Ameh, 2021)

Lagos Prince, one other jailed in Ilorin for Cybercrime

Justice Sikiru Oyinloye of the Kwara State High Court sitting in Ilorin has slammed a 6-month jail term on one Oyekan Abdulbaqqi Adedoyin, a self-acclaimed prince from Kosofe Local Government Area of Lagos State, for offences bordering on cybercrime. Prince Oyekan, 25, was jailed alongside one Oni Stephen Oluwaferanmi from Ilesha, Osun State. The duo of Prince Oyekan and Oni were prosecuted on separate charges by the Ilorin Zonal Command of the Economic and Financial Crimes Commission (EFCC). They pleaded guilty when the charges were read to them. Upon their pleas, counsel to the EFCC, Andrew Akoja, led witnesses to review the facts of the two cases. The witnesses who are operatives of the Commission narrated how the defendants were arrested based on credible intelligence (Ameh, 2021)

Cyber attack Trends

Cybercrime is the broad umbrella under which actions that are performed using cyberplatforms are considered criminal under justice systems. This may include online child sexual exploitation, money laundering through cryptocurrencies, funding and promoting terrorist organizations, and using the dark web for selling drugs and related criminal services. For the purpose of this paper, "cybercrime" is defined as, "cyberactions performed by non-state actors that violate criminal law, and may or may not have a political or national security purpose". Cyberattacks can be further categorized by state and origin as active or passive. An "active" attack aims to alter system resources or affect their operation. Conversely, a "passive" attack seeks to use information from a system but does not affect system resources. Instead, passive attacks aim to obtain data for an offline attack. The term "data breach" is used interchangeably with "cyberattack". Figure 1 illustrates the relationship between cybercrime, cyberattack and cyberwarfare (Frank et al, 2019)

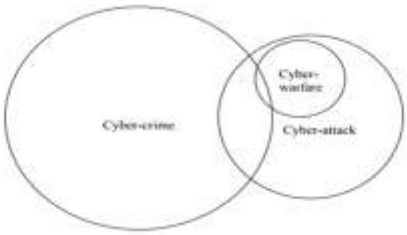


Figure 1: Relationship between Cyber-actions

Cyberattacks are on an upward trend. The Breach Level Index developed by a private company, Gemalto, shows that the number of data records compromised in data breaches by hacktivists, malicious insiders, malicious outsiders and state sponsors, and through accidental loss increased by 86 per cent since 2015, globally. Carbon Black, a cybersecurity company, reported that the dark web marketplace for ransomware has grown at a rate of over 2,500 per cent each year. Among web application attacks, Akamai reported that the United States is the largest target. In the Asia-Pacific region, Japan followed by Singapore and India have suffered the highest number of web application attacks during the second quarter of 2017, according to their live monitor of web-based attacks. Organizations in certain countries are more likely to experience data breaches. Based on a four-year study by IBM and the Ponemon Institute, South Africa and India have the highest estimated probability of data breach occurrences in the next 24 months, while Germany and Canada have the lowest. Arbor Networks analysed that enterprises, governments and educational institutes are most concerned with distributed denial-of-service (DDoS) attacks through social engineering and APTs on corporate networks (Global Cybersecurity Index 2019).

According to IBM, the biggest threats in cybersecurity over the past few years have been from phishing attacks, DDoS attacks and malware (especially ransomware). However, CompTia's *2016 International Trends in CyberSecurity* report states that 58 per cent of global firms struggle

more with security threats caused by human errors than technology risks, an issue that 61 per cent say has become more of a risk over the past two years. The motivations behind these attacks are diverse. Table 1 aims to summarize some of the cyber attack motivations and examples, while Figure 2 illustrates the trends by year and type of security incidents, developed by IBM. Figure 3 shows the number of cyberattacks in Asia-Pacific countries that have been verified by media report between 2013 and 2017, based on data from Gemalto's Breach Level Index, and Figure 4 shows the number of cyberthreats logged in the third quarter of 2017 in Asia-Pacific countries, based on data from Akamai's Real-Time Web Monitor. (Global Cybersecurity Index, 2019)

Table 1: Examples of Cyberattack motivation categorization

	Nuisance	Data Theft	Cybercrime	Hacktivism
Objective	Access and propagation	Economic and political advantage	Financial gain	Defamation, press and policy
Example	Botnets and Spam	Advanced persistent threats	Credit card theft	Website defacement

Figure 2: Sampling of security incidents by attack type, time and impact (2015-2020)

9.0 Nigeria's Cyber Vulnerability to Cybercrime

As government services go digital, criminals are spotting new opportunities for fraudulent claims and theft.

Digital revolution presents new risks: Around the world, the digitization of government is gathering pace, with a host of interactions now carried out online. In some countries, you can vote, pay bills and taxes, and get medical prescriptions – often using a single, digital citizen ID that's stored centrally. This has not escaped the attention of criminals that once focused primarily on retail banking and e-commerce. We're seeing a rise in fraudulent personal and corporate tax and VAT returns and associated rebates, along with bogus welfare claims (Koops, 2019)

Data leakage to Fraudsters: Data is leaking from both public and private sector organizations, either due to malicious hacking or rogue employees. Globally over 700 million personal data records were compromised in 2015, with the largest single breach exceeding information on 70 million individuals. Cyber criminals are also getting better at 'social engineering,' in the form of subtle emails or phone calls from apparently legitimate sources such as banks, financial advisers or even lawyers. In some cases these emails are even sent from the IT systems of those trusted advisers once the cyber-criminal has broken into their email system – the so called business email compromise fraud (Koops, 2019)

Cloud and BYOD: there is the need to continually monitor your cloud suppliers, carrying out audits to ensure ongoing compliance. In an increasing number of cases suppliers recognize this requirement and will offer a range of independently assured security certifications giving customers confidence in their ability. Governments may of course implement their own secure private cloud solutions, offering the ability to impose stricter standards and separation, albeit at increased cost and loss of economy of scale. The BYOD phenomenon impacts all organizations. Government and supplier employees are more frequently using their own smartphones, tablets and laptops for work (Zafar, 2014) Management should ideally establish an organization-wide BYOD policy, as well as taking steps to bring mobile devices under management and ensure that sensitive data can only be accessed and processed by secure applications over secure encrypted channels. All employees should be familiar with the acceptable use rules for BYOD on corporate networks, and these rules should also make clear the rights of management to delete data from personal devices in the event of theft or the owner leaving the organization.

IT outsourcing: Shared services, outsourcing and cloud are shifting provision outside of government, and one big challenge is to retain a core of in-house security expertise which ensures government remains an intelligent customer of such services, as well as having ready access to key skills. After all, people skilled in incident management, analytics, detection, monitoring and response services are scarce and in high demand. Many public sector agencies are living with financial constraints, and need to find creative ways to attract and develop cyber security professionals (Koops, 2019) On the other side, outsourcing could actually improve security, as cloud service providers tend to have relatively advanced cyber security, when compared to the legacy systems and outdated software which can be prevalent in many government IT infrastructures.

10. Most common types of Cyber Attacks

Cyber-attacks can take many forms, and the sophisticated methods used by hackers and criminals are constantly evolving. A cyber-attack will usually take place in one of the following ways:

- **Denial-of-service (DoS) or distributed denial-of-service (DDoS) attack.** This type of attack floods network servers or systems, using bandwidth and rendering them unusable.

- **Malware.** This attack occurs when a system user clicks a link or opens an email attachment, which can then install software on the machine to block access (ransomware) or obtain information (spyware).

- **Phishing.** This occurs when a cyber attacker attempts to steal sensitive information, such as a credit card number or login information, by posing as a trustworthy source.

- **Man-in-the-middle (MitM) attack.** These take place when an attacker inserts themselves into a two-party transaction, such as obtaining information from a device connected to an unsecure public Wi-Fi network.

As technology has progressed in recent years, the opportunities for cyber criminals have increased. Large organizations - such as government agencies - are prone to lapses in security procedures which make them prime candidates for hackers.

11.0 What Nigeria must do

As technologies mature and evolve, additional perspectives or new issues are expected to emerge. Governments can best ensure the protection of critical assets in cyberspace by ensuring the following principles for authentication policy:

Ensure the privacy of individuals: In the age of Big Data, the massive amounts of data generated and collected, sold and traded by third parties are a worrying trend globally. With AI, businesses can analyse more complex data and get more accurate results. Today, online services and smart devices are constantly collecting users' data, including sites visited, purchases made, geolocation, Wi-Fi network information, voice and image recordings, and other personal details, thus potentially reducing users' privacy and safety. Similarly, using tracking AI in business networks capable of monitoring emails, documents and photographs of employees and their activities in the network is a sensitive topic. Employees should be trained and informed about company practices collecting personal data and the use of such data. To track the extent to which countries are safeguarding their citizens' privacy rights, a global cyberprivacy index could be developed.

Establish an incident reporting mechanism: This includes monitoring and assessing the occurrences of cyberattacks and data breaches, including their nature, scope and impact, as well as details of the responses to incidents. One of the challenges government officials may encounter is obtaining an accurate picture of the cyberrisks without an incident reporting mechanism. When Australia and New Zealand established a government audit process, there

were 44 and 16 voluntarily reported data breaches in the respective countries. With the new Privacy Amendment (Notifiable Data Breaches) Act 2017 in Australia, the numbers are expected to increase dramatically as organizations are required to declare any "eligible data breaches".

Strengthen laws and legislations, and increase penalties for cyberattackers and hackers: There is no international framework that binds countries in terms of offensive cyberoperations, but some countries have started initiating the establishment of national laws and legislations to define responsibilities, increase penalties for various cybercrimes, and ensure citizens' safety and security.

Plan and implement digital safety and digital literacy initiatives: These initiatives could be supported by developing and updating cybersecurity policies in the private sector and government organizations, organizing training and awareness campaigns, and creating methods for measuring employees' compliance to cybersecurity standards and policies. It would be important to develop digital safety and digital literacy indicators to better gauge the current state of cybersecurity awareness among individuals and organizations worldwide.

Invest in cybersecurity research and initiatives:- One of the common challenges faced by governments in the region is securing sufficient funding and investments to address cybersecurity. The increasing level of sophistication in cyberattacks would require government officials to upgrade their cybersecurity knowledge and skills on a regular basis. Regional cooperation and knowledge sharing would also be crucial for addressing the wide range of cyberthreats and risks.

Promote cybersecurity best practices for individuals and organizations:- Although not exhaustive, some of the cybersecurity best practices for individual users include the following:

- **Clear cache in browsers and devices**:– This involves clearing browsing history, and removing stored passwords and related information. Clearing a browser's cache makes it more difficult for attackers to access personal information such as email passwords and bank account information. It is also important to change passwords regularly.
- **Update software regularly**:– Several hacks have been carried out by exploiting software vulnerabilities. Attackers exploit this weakness by writing codes to target a specific vulnerability. Software and system updates generally involve patching vulnerabilities, and improving operation system's functionality and performance.
- **Enable at least two-factor authentication for account log in**:- To deter password-guessing attacks, two-factor authentication can be helpful. With two-factor authentication, attackers will have to either acquire the physical component of the log in, or gain access to the cookies or tokens placed on the device by the authentication mechanism. Various account services such as Facebook, Google and online banking services offer two-factor authentication, and it is highly encouraged.

- **Know your right to privacy**: – The Cookie Law is a privacy legislation in Europe that requires websites to obtain consent from visitors to store or retrieve any information on a computer, smartphone or tablet. Individuals have the right to refuse the use of cookies to track browsing history. In addition, review privacy policies and adjust privacy settings of the sites used, particularly, social media sites.
- **Do not automatically connect to Wi-Fi networks and store Wi-Fi passwords**: – As discussed above, public Wi-Fi networks are vulnerable to hacks and pose security risks

12.0 Conclusion

The frequency of cyber-attacks is increasing, and government and private digital entities in Nigeria are the most vulnerable. Staying one step ahead of cyber criminals is not an easy task, but implementing robust, proactive security processes is the most effective way to deal with this dangerous threat. In creating a safe, digital environment for citizens and companies, government can embrace leading practices from the private sector, and encourage employees to be more cyber-aware

References

Ajiji, Y. M. (2017). "Cybersecurity Issues in Nigeria and Challenges." International Journal of Advanced Research in Computer Science and Software Engineering 7(4): 315–321.

Alagappa, M. (1987). The National Security of Developing States: Lessons from Thailand. Dover, MA: Auburn House.

Alkali, R. A. (2010). Issues in Nigerian Foreign Policy and International Relations. Kaduna, Nigeria: Media Press.

Ameh O.(2021, November, 24) 'Corps member bags two months for cybercrime. The Guardian,Nigeria.https://guardian.ng/news/corps-member-bags-two-months-for- cybercrime/

Aniekan, M. N., & Afolabi, M. B. (2017). "Introduction to Cybersecurity and Cybercrime." In Aniekan & Afolabi (Eds.) Intelligence and Security Studies Programme. Lagos, Nigeria: Spectrum.

Antonucci, D. (2017). The Cyber Risk Handbook: Creating and Measuring Effective Cybersecurity Capabilities. Hoboken, NJ: John Wiley & Sons.

Azeez, O. (2019). "Cybercrime Cost Nigeria N288 bn in 2018." Business a.m. https://www.businessamlive.com/cyber-crime-cost-nigeria-n288bn-in-2018/, accessed November, 25, 2021.

Buchanan, B. (2016). The Cybersecurity Dilemma: Hacking, Trust and Fear between Nations. New York: Oxford University Press.

Buzan, B., & Hansen, L. (2009). The Evolution of International Security Studies. New York: Cambridge University Press.

Cavelty, M. D., & Wenger, A. (2019). "Cyber Security Meets Security Politics: Complex Technology, Fragmented Politics, and Networked Science." Contemporary Security Policy 41(1): 1–28.

Federal Government of Nigeria (2015). Nigerian Cybersecurity Act 2015. http://lawnigeria.com/LawsoftheFederation/Cyber-Crime-Act,-2015.html. Accessed January 25, 2022.
Frank, I., & Odunayo, E. (2019). "Approach to Cybersecurity Issues in Nigeria: Challenges and Solutions." (IJCRSEE) International Journal of Cognitive Research in Science, Engineering and Education 1(1): 1–11.
Global Cybersecurity Index (2019). 2019 Global Cybersecurity Index. https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx. Accessed December 20, 2021.
Mohamed A.H, & Solanke, A. A. (2019). "Cybercrime and Digital Forensics: Bridging the Gap in Legislation, Investigation and Prosecution of Cybercrime in Nigeria." International Journal of Cybersecurity Intelligence and Cybercrime 2 (1): 56- 63.
Mordi, M. (2019). "Is Nigeria Really the Headquarters of Cybercrime in the World?" Guardian. https://guardian.ng/news/is-nigeria-really-the-headquarters-of-cybercrime-inthe-world/. Accessed October,18, 2021.
Wada F. and Odulaja G. O. (2014), "Electronic Banking and Cyber Crime In Nigeria - A Theoretical Policy Perspective on Causation," Afr J Comp & ICT, Vol 4(3), no. Issue 2.

E-COMMERCE: SOFTWARE STRATEGY FOR SUSTAINING BUSINESS TRANSACTIONS IN THE POST COVID-19 ERA

Ike U. Kingsley[1], Okere, R. Chinedum[2], Nlemadim, A. Lynda[2], Onyeike, G. Obinna[2]

[1]Department of Computer Science, Imo State Polytechnic Umuagwo
[2]Department of Computer Science, Federal Polytechnic Nekede
uchebenzene@yahoo.com

ABSTRACT

The paper examines e-commerce and determines its suitability for the conduct of business in the post Covid-19 era. The Covid-19 pandemic brought to light a number of challenges to the economy and trade worldwide. Overcrowding and frequent physical interactions between individuals have been identified as major causes of the spread of the disease. Although cities around the world have relaxed restrictions, the future is still uncertain. Events surrounding the Covid-19 pandemic required that alternative modes for business transactions are adopted and this made the already exploding e-commerce trend even more important. As traditional shopping became more difficult and scary, people sort alternative ways to respond to the

situation. The paper thus, presents e-commerce as the software strategy for business transactions and examines the effects of the covid-19 on online shopping behavior.As an alternative mode to do business, e-commerce has the capacity to cause less physical interactions between persons and avert overcrowding in the post covid-19 era thereby curtailing further spread or a resurgence of the dreaded corona virus.

Keywords: E-commerce, software, ICT, COVID-19

1. INTRODUCTION

Commerce has been a major impetus for human survival since the beginning of recorded history and beyond. The mass adoption of the internet has created a paradigm shift in the way businesses are conducted today. The growth of the internet continues to influence our lives and businesses. Irrespective of their type and size, firms and organizations are rethinking their strategies and operations. The past decade has seen a new kind of commerce; e-commerce; the buying and selling of products through human-computer interaction over the internet. An increasing number of businesses are using e-commerce to gain competitive advantage(Ike, 2021).

The COVID-19 pandemic brought challenges which affected every facet of live including commercial activities, thereby causing the need for more individuals and organizations to think towards adopting e-commerce, an already exploding trend. With traditional commerce, consumers have the option to come into market stores personally to buy products and this mostly causes overcrowding. In March 2020, much of the world went into lockdown, forcing many businesses to temporarily shut down and although today, cities have relaxed restrictions, the future is still uncertain.

Overcrowding which leads to crowd congestion and increased physical interactions with others has been identified as the major causes of the spread of the Covid-19. E-commerce takes care of this as it offers an alternative channel for maintaining business activities, social interactions and consumption in the post Covid-19 era, invariably preventing physical interactions thereby averting overcrowding common with traditional market places.

2. THE E-COMMERCE CONCEPT

Electronic commerce (e-commerce) has become a popular topic for business and academic research since the early 1990's. Different researches focus on different areas; applications, services, marketing, strategy, the internet, extranets and technologies in e-commerce but to name a few (Chan & Swatman, 2012). E-commerce has changed the way organizations perform their activities. Starting in the 1970's, three quite separate trends (Business Document Exchange, Logistics Management and Global Networking) came together to provide the infrastructure and techniques for what we have today as electronic commerce (Swatman (1996). This historical view of e-commerce has been further corroborated and extended by Zwass (1996) who stated that; traditional e-commerce, conducted with the use of information technologies centering on electronic data interchange over proprietary valued added networks is rapidly moving to the internet. The internet's www has become the prime driver of contemporary e-commerce.

Various definitions have been offered to explain the concept of e-commerce and all share certain similarities. E-commerce (electronic commerce) is the activity of electronically buying or selling of products on online services or over the internet.

Goel (2007) defines e-commerce as a modern business methodology that addresses the need of organizations, merchant and consumers to cut costs while improving the quality of goods and services and increasing the speed of service delivery, by using the internet. E-commerce differs from the traditional commerce in the way it enables the trading of goods, money and information electronically and there is no longer need for physical currency or goods to conduct business.

Sridhar (2017): electronic commerce is the modern business methodology to address needs of an organization, merchants, commerce to cut cost and to do the following; to improve quality//speed of delivery, more commonly associated with buying and selling of information, products and services via computer networks today;electronic data interchange, latest and dependable way to deliver electronic transactions by computer to computer communication combined with (JIT); Just in Time manufacturing methods; EDI and e-mail used for many years, e-commerce is a transaction of buying and selling online.

According to Meiryani (2019), e-commerce is the use of communicating networks and computers to carry out business processes. E-commerce thus, is the distribution, purchase, sale, marketing of goods and services through electronic systems such as the internet or television, www or other computer networks. E-commerce can involve electronic funds transfer; electronic data interchange, automated inventory management systems and automated data collection system. The information technology industry sees e-commerce activities as the application of e-business related to commercial transactions such as transfer of funds electronically, supply chain management, e-marketing or online marketing, online transaction processing, electronic data interchange and so on.

Electronic commerce draws on technologies such as mobile-commerce, electronic funds transfer, supply chain management, internet marketing, online transaction processing, electronic data interchange, inventory management system and automated data collection (Sridhar, 2017).

3. UNDERSTANDING THE COVID-19 EFFECT ON CONSUMER ONLINE SHOPPING BEHAVIOUR

The Covid-19 pandemic can be described as one of the defining events of the year 2020 and its implications may as well last decades. The situation has rapidly changed. The covid-19 pandemic has impacted our daily behaviour, from interacting with friends, colleagues, and family, to safety measures and working.Additionally, the pandemic has dramatically changed consumer's shopping behavior (Vinerean, 2020). Despite the introduction of vaccines and mandatory wearing of facemasks, the number of people deemed safe to gather in a single place has dwindled from thousands to hundreds to ten. Essentially, people have come to terms with the realities of our interconnected world and how difficult it is to temporarily separate connections to others. During this difficult period characterized by infections, lockdown and economic uncertainty, consumers have changed their purchasing behavior. Nonetheless as consumers' behavior changes, this provides new opportunities for organizations to adapt and tailor the experiences of targeted audience. These new buyer practices have significant ramifications for retailers and consumer-packaged-goods companies (Mckingsley & Company 2020).

The period of increased isolation and uncertainty caused by the Covid-19 pandemic requires new approaches and one of the responses that are observed is huge overnight changes to shopping behaviors.In examining, the impact of Covid-19 on consumer behavior, Sheth (2020) analyzed four contexts that transform consumer behavioral patterns such as (a) social contexts (major life events), (b)technology and its effects on disrupting old habits, (c) rules and regulations that arise based on public policies (d) natural disasters and pandemics. In the current context of the covid-19 pandemic, all these contextual aspects have to a certain degree affected consumers and their purchasing behavior. Since the beginning of the covid-19 outbreak, consumers have displayed stockpiling behaviour that significantly deviates from their usual shopping behaviour (Eger et al., 2021). The global spread of Covid-19 has been accompanied by a lot of uncertainty and at times contradictory information. When people are hearing different advice from multiple sources, they have a greater instinct to over-prepare. As news of Covid-19 spread and as it was officially declared a pandemic by the World Health Organization, people responded by stocking up. The pronouncement of lockdowns and restrictions led to panic buying. From bulk-buying to online shopping, people are changing what they are buying, when and how.

An unprecedented growth of e-commerce was witnessed during the lockdown in the year 2020. Studies show that in the United States for instance, as many as 29% of surveyed consumers insisted that they will never return to shopping in person again. In the UK, 43% of consumer's state that they expect to keep on shopping the same way even after lockdown is over. As part of the findings established by the United Nations conference on trade and development (2020), the pandemic has accentuated the trend towards greater adoption of social media and growth in sales through e-commerce websites. Shifts in consumption habits have also been observed, driven by the need for essential items.

4. TECHNOLOGY COMPONENTS OF E-COMMERCE

The technology and infrastructure used to develop the e-commerce application is the key to its success. Thus, the components of any e-commerce platform basically comprise of hardware and

software and must be selected in such a way that they can fulfill the needs of the e-commerce application. The figure 1 shows the components involved in e-commerce infrastructure;

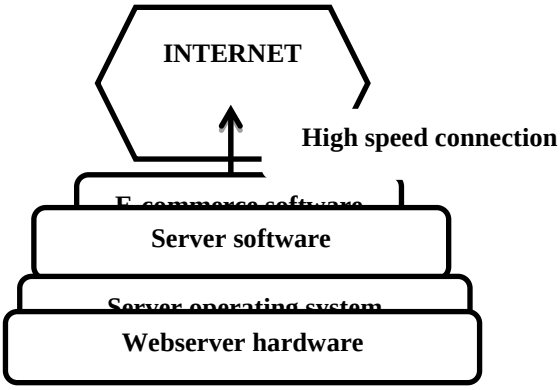


Figure:1 shows the components involved in e-commerce infrastructure;

4.1. Hardware

Hardware is any tangible physical component of a computing system (Ike, 2021). A web server hardware platform is one of the major components of the e-commerce infrastructure on which the performance of the whole e-commerce application depends. While selecting webserver hardware, the software that will run on the server of the e-commerce transaction to be processed must be considered. The amount of storage capacity and the computing power required depends on the volume of the e-commerce transaction to be processed. If the exact requirements are not known in advance, they can be upgraded to meet the requirement. The e-commerce applications rest on the hardware and together rely on a network to function. This is depicted in the figure 2.

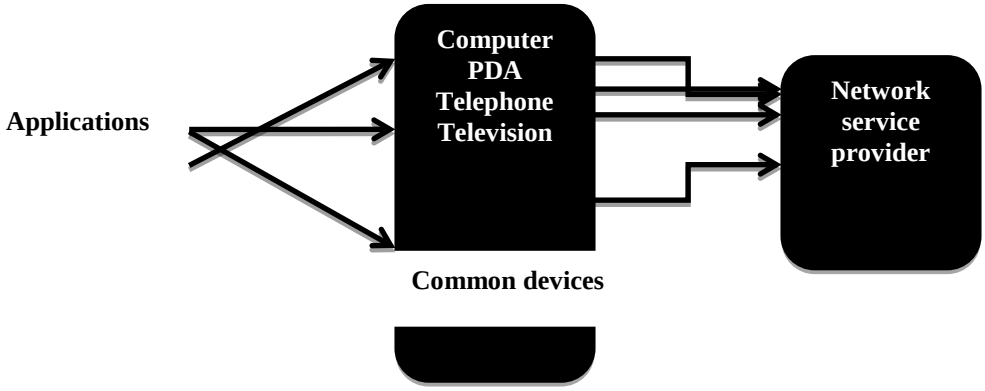


Figure 2: e-commerce Application framework

4.2. Software

Software is the main component that implements the e-commerce services and functionality. Software for e-commerce can be categorized intotwo namely; webserver software and e-commerce software.

4.2.1. Webserver software:

Webserver software is required in addition to the webserver operating system. An Operating System (OS) controls a computer and is responsible for coordinating and controlling the usage of computer hardware resources (Ike, 2022).The webserver software is used to implement functionalities such as security and identification as well as retrieval and sending of webpages. It creates a weblog file that identifies URL of the visitor, length of visit and search engines and the keywords used to find the site. Webserver software includes web development tools such as HTML editor and web page upload support.

4.2.2. E-commerce application software

This is customized application software or website for carrying out specific e-commerce transactions. Application software is computer software that enables a person to perform specific jobs with a computing device (Ike, 2021). They are also called "user programs" because they help the user to solve day-to-day problems. Typical e-commerce software must support processes such as; catalog management, product configuration, shopping cart and transaction processing.

Catalog management: it is required to deliver the customized content to the screen or the GUI used by the customer. The software used for catalog management combines the data formats into a standard format for viewing, aggregating and interacting catalog data into a central store.

Product configuration: the web-based product configuration software allows the user to build the product to their specifications without the intervention of the salesperson. For example, Dell computers and Cisco systems use configuration software to sell build-to-order and network processes to their customers over the internet.

Shopping cart: a model known as shopping cart is used by the e-commerce sites to track the items that are selected for purchase. The shopping cart allows customers to view all the items selected by them. The customers can add new items and remove the previously selected items from the shopping cart.

5. HOW TO WORK WITH E-COMMERCE PLATFORMS

E-commerce is about setting business on the internet, allowing visitors to access the e-commerce website and go through a virtual catalogue of products/services online. When a visitor wants to buy something he/she likes, they merely 'add' to their shopping basket. Items in the virtual basket can be added or deleted and when the person wants to checkout, he/she heads to the virtual checkout counter, which has his/her complete total and that will ask for name, address and other vital information as well as the mode of payment (usually via credit card). Once the person has entered his/her information, he/she then waits for delivery.

6. SOFTWARE-INSPIRED INCENTIVES FOR ENGAGING IN E-COMMERCE IN THE POST COVID-19 ERA

A basic fact of internet retailing is that all e-commerce websites are created equal as far as the location imperative of success in retailing is concerned. No site is any closer to its web customers and competitors offering similar goods and services may be only a mouse click away. This makes it vital that businesses find ways to build customer satisfaction, loyalty and relationships so customers keep coming back to their sites, especially as the post COVID-19 era still requires less interactive meetings while engaging in commercial activities to prevent a recurrence or surge in the spread of the dreaded disease. Thus, the incentives for engaging in e-commerce in the post-COVID-19 era are listed below;

1. No overcrowding is witnessed as is the case with traditional physical shopping.
2. There is a virtual community of customers, suppliers, company representative and others via newsgroup, chat-rooms and links to related sites.

3. There are personal webpages, personalized product recommendations, advertising and e-mail notices as well as e-interactive support for customers.
4. Attractive product selection, competitive prices, satisfaction guarantee and customer support after the sale.
5. Fast and easy navigation, shopping and purchasing and prompt shopping and delivery.

7. MEASURES TO SUPPORT COVID-19 RECOVERY PLANS THROUGH E-COMMERCE DEVELOPMENT

The COVID-19 pandemic and other events surrounding it have made the already exploding e-commerce trend even more important. E-commerce giants such as Amazon, Walmart, Alibaba and others have been rapidly expanding in the last few years. While traditional shopping is highly discouraged in the post-COVID-19 era, making it difficult and sometimes scary, people are increasingly inclined to shop online. Some of the reasons that customers are likely to continue shopping more and more online in the post COVID-19 era: new shopping behavior caused by quarantine and lockdowns, the fear of contacting the covid-19 from individuals overcrowded in traditional (physical) markets and scarce commodities in physical stores are available for order online.

To make e-commerce an engine of sustainable growth in developing countries and LDCs in the context of Covid-19 economic recovery plan, measures that should be adopted according to respondents in a business survey are cited in Vitale et al. (2020). These recommendations are based on the top ten measures resulting from the business survey which were especially emphasized by the survey respondents. They are:

1. Development of a national e-commerce strategy.
2. Reduction of cost for internet access
3. Reduced e-payment cost.
4. Financial incentives and liquidity support.
5. Increased internet connectivity in underserved areas.

6. More ads on available e-commerce.
7. New online consumer protection measures.
8. New e-commerce market places for essentials.
9. Skill training programs.
10. Maintenance logistics operations.
11. New-e-payments applications.
12. Market integration of informal e-commerce.
13. Initiatives to get businesses online.
14. New digital health and education solutions.
15. New logistics services offered by private operators.

8. CONCLUSION

The paper has shed light on the suitability of e-commerce as a reliable alternative for shopping in the post-covid-19 area.In terms of ensuring less physical meetings between individuals especially while engaging in normal commercial activities, the role of e-commerce cannot be easily exaggerated. Nevertheless, there are issues associated with the use of e-commerce platforms which include cost, security and the ability to use electronic mobile devices to conduct business online. These issues notwithstanding are not strong enough to dissuade individuals from embracing the e-commerce trend and switching to it.

REFRENCES

Anam, Bhahi, Akram, Hamza; Hafiz Muhammad; Khan, Ahmed Usman; Nagvi Syeda Mahwish; Bilal, Muhmmed (2020); E-commerce trends during COVID-19 pandemic. International journal of future generation communication and networking

Bokos, Yannis (2001); The emerging landscape of retail e-commerce. Journal of Economic perspective.

Bunnel, David (2001); The ebay business mode; business secrets behind the world's hottest internet company. John Wiley and sons. Pg. 71-81

Burges, Stephen; Sellito, Carmine; Karanasios, Stan (2009); effective web presence solutions for small businesses; strategy for successful implementation. IGI global publishers

Chan, E (1999); "What is e-commerce"? [www.document] URL: http://www.businesst.bf.rmit.edu.au/elsie/whatis/sld001.htm.accessed

Chan, E. and Swatman, P.M.C (1999); Electronic commerce; a component model. 3rd annual CollECTeR conference on Electronic commerce. Wellington,New Zealand.

Chaudhury, Abijit; Kiulboer, Jean-pierre (2002); E-business and E-commerce Infrastructure; technologies supporting the E-business infrastructure. McGraw Hill Education.

Deleone, William; Mclean, Ephraim (2014); Measuring e-commerce success; applying the Deleone and Mclean information system model. International Journal of Electronic Commerce.

Eger, L., Komarkova, L., Egerova, D. and Micik, M. (2021); The effect of Covid-19 on consumer shopping behaviour: Generational cohort perspective. Journal of Retailing and Consumer Services, 61, 102542

Eisingerich, Andreas, Kretschmer Tobias (2008); In E-commerce, more is more. Harvard business review. 86. Pg. 20-21

Ike, U.K (2021); E-commerce as strategy for sustaining commercial activities and averting overcrowding in the post covid-19 eraSchool of Arts and Science conference Alvan Ikoku Federal college of Education Owerri in affiliation with University of Nigeria Nsukka 2021.

Khan, Sultan David (2019); Congestion detection in pedestrian crowds using oscillating in motion trajectories. Engineering applications of AI. Vol. 85

Kohavi, Ronny; Provost, Foster (2001); Applications of data mining to electronic commerce. Springer US. ISBN 9780792373032

Laudon, Kenneth; Travel, Carol (2004); E-commerce, Business, Technology, Society. Pearson PLC.

Mckingsley & Company (2020); how covid-19 is changing consumer behavior now and forever.

[Online] available.

Meiryani W.Y. (2019); The role of IT in e-commerce. International Journal of Scientific and Technology Research volume 8 issue 01.

Power, Michael (2013); Online highs are old as the net; the first e-commerce was a drug deal. The guardian.

Sheth, J. (2020); Impact of Covid-19 on consumer behavior. Will the old habitreturn or die? Journal of Business Research, 117.

Sridhar, S. (2017); E-commerce technology made easy. (IJITR) InternationalJournal of Innovative Technology Research Volume No.5 Issue No.3 April.

Susan, Meyer (2003); Understanding the COVID-19 effect on online shopping behavior

Tkacz, Ewaryst; Kapczynsk, Adrian (2009); Internet technology development and applications. Springer science business media. Pg. 255.

Terzi, Nuray (2011); The impact of e-commerce on international trade and employment. Procedia social and behavioral sciences. 24. 745-753

Vinerean, Simona (2020); Understanding consumers' online shopping behaviour during the Covid-19 pandemic- Empirical research. Expert journal of Marketing, Volume 8, Issue2.

Vitale, A., Cyron, L., Michaud, L., Riegel, V., Barayre, C., Fredriksson, T. (2020); Covid-19 and E-commerce; impact on business and policy responses. 2020 United Nations conference on Trade and Development.

Zwass, V. (1998); electronic commerce; structures and issues. International journal of electronic commerce volume 1 No.1.

Zwass, V. (1998); Structure and Macro-Level of Electronic Commerce; From technological infrastructure to electronic marketplace [www document] URL; http://www.mhhe.com/business/mis/zwass/ecpaper.html.accessed

SMART HOME SECURITY SYSTEM USING INTERNET OF THINGS

Omenka Ugochukwu Enyinna[1], Jidere Ann[2], Nwogu Uchechukwu[3],
Department of Computer Science,
Federal Polytechnic Nekede, Owerri.
uomenka@gmail.com[1]; annchinee@gmail.com[2]; uchenwogu@gmail.com[3]

Abstract
The Internet of things (IoT) describes physical objects (or groups of such objects) that are embedded with sensors, processing ability, software, and other technologies that connect and exchange data with other devices and systems over the Internet or other communications networks without any interfering human interventions. Home security is a very valuable application of IoT and we are using it to create a low-cost security system for homes as well as for industrial use. The system will alert the owner once there is any unauthorized entry by sending a notification to the owner, and also sound an alarm to scare any intruder. The security system will use a microcontroller known as Arduino Uno to interface between the components, a Passive InfraRed motion sensor to detect whether a human has moved in or out of the sensors range, a buzzer for sounding the alarm, a WiFi module, ESP8266 to connect and communicate using the Internet, and battery for powering the components. The core advantages of such a system includes the ease of setting up, lower costs and low maintenance.
Keywords (IoT, Home Automation, Smart Home Security)

1.0 Introduction

The Internet of things (IoT) describes physical objects (or groups of such objects) that are embedded with sensors, processing ability, software, and other technologies that connect and exchange data with other devices and systems over the Internet or other communications networks. The field has evolved due to the convergence of multiple technologies, including ubiquitous computing, commodity sensors, increasingly powerful embedded systems, and machine learning (Gills,2021).

Internet Things refers to the network of connected physical objects that can communicate and exchange data among themselves without the need of any human intervention. It has been formally defined as an "Infrastructure of Information Society", because IoT allows us to collect information from all kind of mediums such as humans, animals, vehicles, kitchen appliances. Thus, any object in the physical world which can be provided with an IP address to enable data transmission over a network can be made part of IoT system by embedding them with electronic hardware such as sensors, microcontrollers, software and networking gear. IoT is different than Internet as in a way it transcends Internet connectivity by enabling everyday objects that uses

embedded circuits to interact and communicate with each other utilizing the current Internet infrastructure (Anitha,2017).

When something is connected to the internet, that means that it can send information or receive information, or do they both. This ability to send and/or receive information makes things "smart." In the Internet of Things, all the things that are being connected to the internet can collect information and then send it, receive information and then act on it, or do both procedures. Meaning that we can classify core IoT devices and activities as follows:

Table 1.0 IoT Essential Devices and Activities

DEVICE	ACTIVITIES	EXAMPLES
Sensors	Receive data from the real world	PIR Motion Sensors, Reed Sensors etc.
Actuators	Converts data to physical quantities	Relays, Servo Motors etc.
Embedded Microcontrollers and Microprocessors	Interprets and processes data it receives from its I/O peripherals using its central processor.	Arduino, Raspberry Pi, etc.
Connectivity Interfaces, Adapters and Gears	Connects to the Internet or any local communication device to exchange data and to push the data to the IoT application for any analysis.	Wifi Modules, GSM Controller Modules etc.

Traditional fields of embedded systems, wireless sensor networks, control systems, automation (including home and building automation), independently and collectively enable the Internet of things. In the consumer market, IoT technology is most synonymous with products pertaining to the concept of the "smart home", including devices and appliances (such as lighting fixtures, thermostats, home security systems and cameras, and other home appliances) that support one or more common ecosystems, and can be controlled via devices associated with that ecosystem, such as smartphones.

Home automation or domotics is building automation for a home, called a smart home. A home automation system will monitor and/or control home attributes such as lighting, climate, entertainment systems, and appliances. It may also include home security such as access control and alarm systems. When connected with the Internet, home devices are an important constituent of the Internet of Things ("IoT"). A home automation system typically connects controlled devices to a central smart home hub (sometimes called a "gateway"). The user interface for

control of the system uses either wall-mounted terminals, tablet or desktop computers, a mobile phone application, or a Web interface that may also be accessible off-site through the Internet. While there are many competing vendors, there are increasing efforts towards open-source systems. However, there are issues with the current state of home automation including a lack of standardized security measures and deprecation of older devices without backwards compatibility. Home automation has high potential for sharing data between family members or trusted individuals for personal security and could lead to energy saving measures with a positive environmental impact in the future (Home automation, 2022).
An important issue to consider when we dialogue on home automation is Security. Home security is a vital feature of home automation and sometimes the most crucial one. Home security has made a drastic change in the past few years and continues to advance. Formerly home security systems meant having an alarm system that would trigger when someone breaks into a home, but a smart secure home can do much more than that. Thus, the main objective of our work is to design and implement a system which alerts the owner of an intruder break-in by sending a notification to their smart phones. The owner will also have the ability to stop or start the alarm remotely using his smart phone. This system will help the users to safeguard homes. There has been an extraordinary growth in the number of devices being connected to the Internet since past few years. All these devices connected to the internet are part of the IoT infrastructure, which allows these devices to send and receive data among each other. This is why it is beneficial to use such an existing infrastructure for designing the proposed security system. An alarm system that sounds the buzzer is useless when a user is not present in the home to take action.

The IoT network consists of embedded electronics, sensors and software. User want to be assured that their home is protected from intruders and thieves while they are gone. This is why the proposed system keeps the owner informed in the real time about the security status of their home. The designed system informs the user as there is a break-in so that the user can take necessary actions.
This paper is organized using the IMRaD report format and it has the following sections: Section 1 discusses about the Introduction of IoT, Home Automation and Smart Home Security. Section 2 gives the details of the materials and methods to implement the proposed system, the proposed working model configuration of applications. Section 3 explains the experimental results which is followed by Section 4, Discussions and future enhancements.

2.0 Methods

Various components are required to build home automation systems. Some of the key components are listed below to give us an idea about the proposed system.

2.1 Arduino Uno

Arduino is an open-source hardware and software company, project, and user community that designs and manufactures single-board microcontrollers and microcontroller kits for building digital devices. Its hardware products are licensed under a Creative Commons license, while software is licensed under the GNU General Public License (GPL), permitting the manufacture of Arduino boards and software distribution by anyone. Arduino boards are available commercially from the official website or through authorized distributors.

Arduino board designs use a variety of microprocessors and controllers. The boards are equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards ('shields') or breadboards (for prototyping) and other circuits. The boards feature serial communications interfaces, including Universal Serial Bus (USB) on some models, which are also used for loading programs.
The microcontrollers can be programmed using the C and C++ programming languages, using a standard API which is also known as the Arduino language, inspired by the Processing language and used with a modified version of the Processing IDE. In addition to using traditional compiler toolchains, the Arduino project provides an integrated development environment (IDE).

2.1.1 Software

The Arduino integrated development environment (IDE) is a cross-platform application (for Microsoft Windows, macOS, and Linux) that is written in the Java programming language. It originated from the IDE for the languages Processing and Wiring. It includes a code editor with features such as text cutting and pasting, searching and replacing text, automatic indenting, brace matching, and syntax highlighting, and provides simple one-click mechanisms to compile and upload programs to an Arduino board. It also contains a message area, a text console, a toolbar with buttons for common functions and a hierarchy of operation menus.
The Arduino IDE supports the languages C and C++ using special rules of code structuring. The Arduino IDE supplies a software library from the Wiring project, which provides many common input and output procedures. User-written code only requires two basic functions, for starting the sketch and the main program loop, that are compiled and linked with a program stub main() into an executable cyclic executive program.

Figure 1. Arduino IDE on Windows 10

A cyclic executive is an alternative to a real-time operating system (Laplante et al, 2012). It is a form of cooperative multitasking, in which there is only one task. The sole task is typically realized as an infinite loop in main(), e.g. in C. According to Macfadden (2018), Arduino boards are actually micro-controllers rather than 'full' computers like Raspberry Pi. Arduino lacks a full operating system but can run a written code that is interpreted by its firmware. Arduino has no API and cannot provide user interactivity as there is no operating system.

Monk (2011) defined a *sketch* as a program written with the Arduino IDE. Sketches are saved on the development computer as text files with the file extension **.ino**. Arduino Software (IDE) pre-1.0 saved sketches with the extension **.pde**.

A minimal Arduino C/C++ program consists of only two functions:

- setup(): This function is called once when a sketch starts after power-up or reset. It is used to initialize variables, input and output pin modes, and other libraries needed in the sketch. It is analogous to the function main().

- loop(): After setup() function exits (ends), the loop() function is executed repeatedly in the main program. It controls the board until the board is powered off or is reset. It is analogous to the function while(1).

2.1.2 Blink example

Most Arduino boards contain a light-emitting diode (LED) and a current limiting resistor connected between pin 13 and ground, which is a convenient feature for many tests and program functions. A typical program used by beginners, akin to Hello, World!, is "blink", which

repeatedly blinks the on-board LED integrated into the Arduino board. This program uses the functions pinMode(), digitalWrite(), and delay(), which are provided by the internal libraries included in the IDE environment. This program is usually loaded into a new Arduino board by the manufacturer.

```
# define LED_PIN 13                    // Pin number attached to LED.

void setup() {
   pinMode(LED_PIN, OUTPUT);       // Configure pin 13 to be a digital output.
}
void loop() {
   digitalWrite(LED_PIN, HIGH);      // Turn on the LED.
   delay(1000);                          // Wait 1 second (1000 milliseconds).
   digitalWrite(LED_PIN, LOW);     // Turn off the LED.
   delay(1000);                         // Wait 1 second.
}
```

2.1.3 Hardware

The Arduino UNO is the best board to get started with electronics and coding. If this is your first experience tinkering with the platform, the UNO is the most robust board you can start playing with. The UNO is the most used and documented board of the whole Arduino family.

Arduino UNO is a microcontroller board based on the ATmega328P. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started.

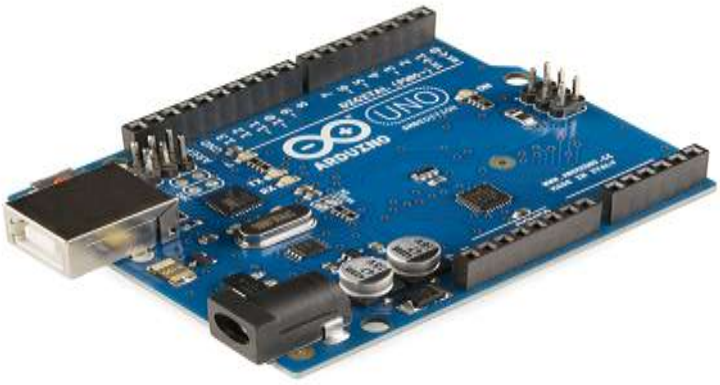


Figure 1. Arduino Uno R3 Board

2.2 ESP8266 (Wi-Fi Module)

The ESP8266 is a low-cost Wi-Fi microchip, with built-in TCP/IP networking software, and microcontroller capability. This small module allows microcontrollers to connect to a Wi-Fi network and make simple TCP/IP connections using some specific command language.

Figure 2. ESP8266 Wi-Fi Module

ESP8266 comes with capabilities of 2.4 GHz Wi-Fi (802.11 b/g/n, supporting WPA/WPA2), general-purpose input/output (16 GPIO), UART (on dedicated pins). It employs a 32-bit RISC CPU running at 80 MHz (or overclocked to 160 MHz). It has a 64 KB boot ROM, 64 KB instruction RAM and 96 KB data RAM.

It is mostly used for development of IoT (Internet of Things) embedded applications as it is used for controlling devices over the Internet. It can work with a micro-controller like the Arduino or it can be programmed to work on its own. It can connect to your router and work as a client or it can be an access point itself or both. It is IP addressable and can be a Web Server.

2.3 PIR Motion Sensor

A passive infrared sensor (PIR sensor) is an electronic sensor that measures infrared (IR) light radiating from objects in its field of view. They are most often used in PIR-based motion detectors. PIR sensors are commonly used in security alarms and automatic lighting applications. PIR sensors detect general movement, but do not give information on who or what moved. Therefore, using these sensors to detect human movement or occupancy in security systems is very common. The output of PIR motion detection sensor can be connected directly to one of the Arduino (or any microcontroller) digital pins. If any motion is detected by the sensor, this pin value will be set to "1". The two potentiometers on the board allow you to adjust the sensitivity and delay time after detecting a movement.

Figure 3. PIR Motion Sensor

2.4 Breadboard and Jump Wire

A breadboard is used to build and test circuits quickly before finalizing any circuit design. The breadboard has many openings into which route components like ICs and resistors can be connected. A typical breadboard that includes top and bottom power distribution rails is shown below figure 4. A jump wire (also known as jumper) is an electrical wire, or group of them in a cable, with a connector or pin at each end, which is normally used to interconnect the components of a breadboard or other prototype or test circuit, internally or with other equipment or components, without soldering as shown in figure 5.

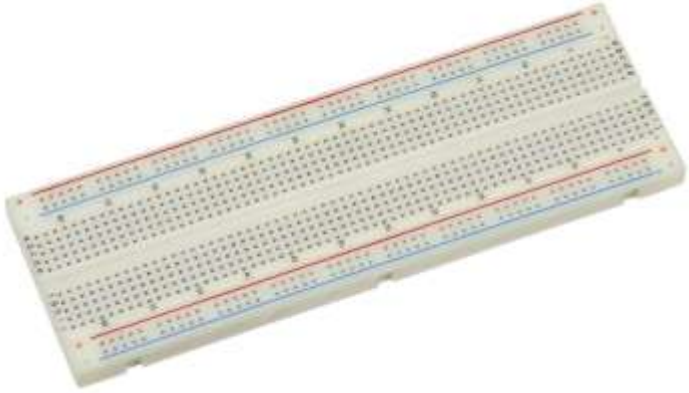


Figure 4. Breadboard **Figure 5. Jump Wire**

2.5 Proposed Model and Architecture of the System

The following architecture diagram in figure 6, represents the proposed systems structure.

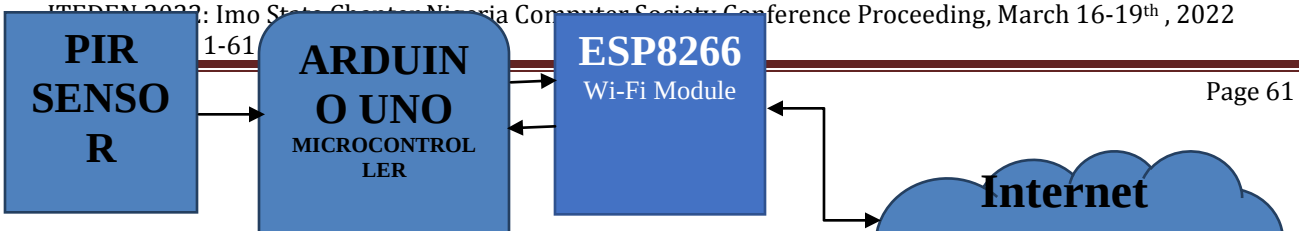


Figure 6. Architecture Diagram of Proposed Model

For us to achieve the above configuration, we will use a breadboard to connect the Microcontroller, PIR Sensor, Buzzer and the ESP8266 using jumper wires, the breadboard makes it possible to connect multiple inputs to a single pin on the Arduino board. The following sketch in figure 7, constructed using the Fritzing software, shows how the components are supposed to be connected together using the breadboard and jumper wire.

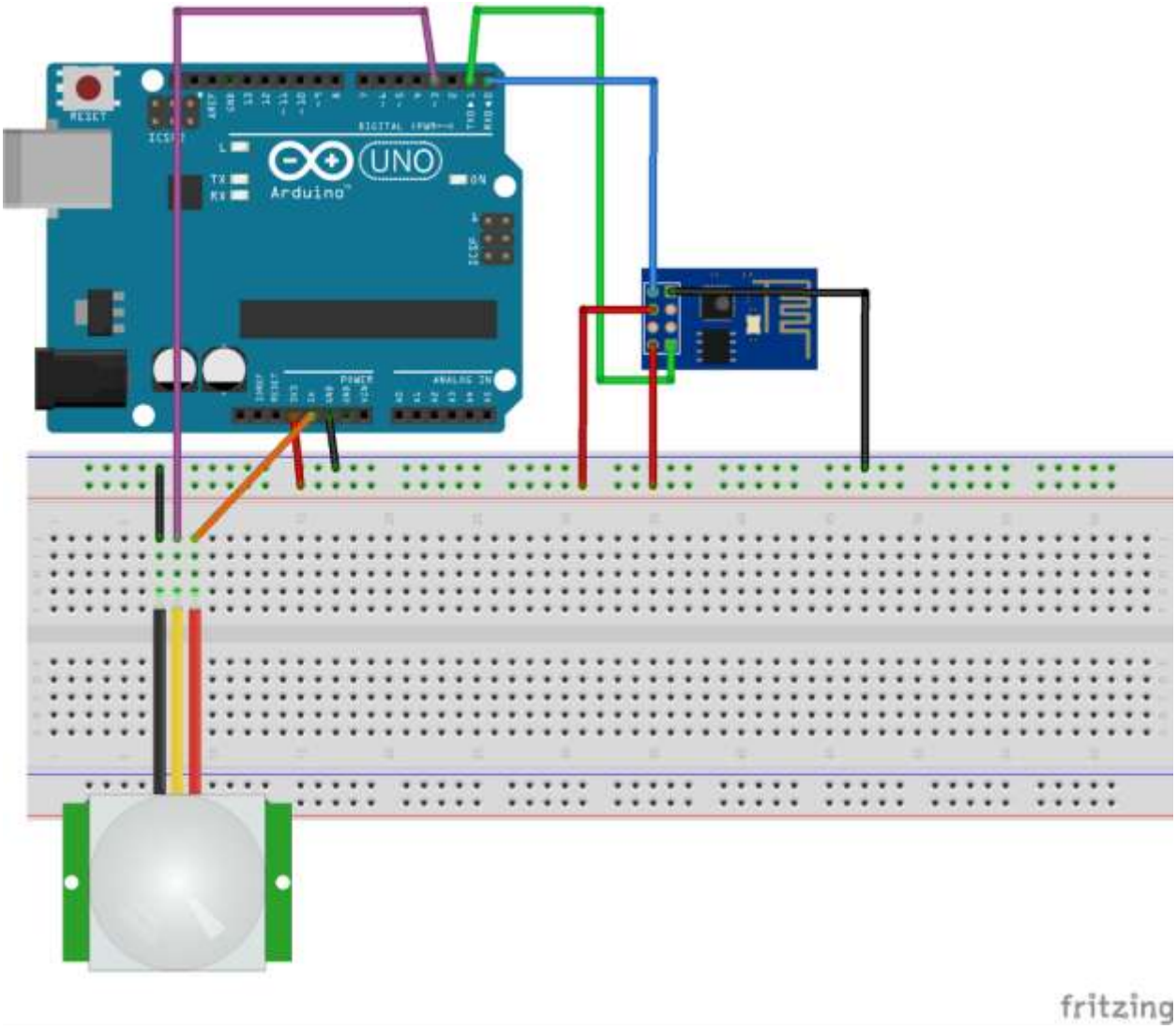


Figure 7. Fritzing Sketch Diagram.

Wiring:

Arduino	| ESP8266	TX (D1)	| RX
RX (D0)	| TX	3v3	| VCC and CH_PD

GND | GND
Arduino | PIR
D3 | OUT
5v | VCC
GND |
GND

2.5 Programming and Configuration of the System

2.5.1 Arduino Sketch for programming the microcontrollers

```
/the time we give the sensor to calibrate (10-60 secs according to the datasheet)
int calibrationTime = 30;

//the time when the sensor outputs a low impulse
long unsigned int lowIn;

//the amount of milliseconds the sensor has to be low
//before we assume all motion has stopped
long unsigned int pause = 5000;

boolean lockLow = true;
boolean takeLowTime;

int pirPin = 3;     //the digital pin connected to the PIR sensor's output
int ledPin = 13;

/////////////////////////////
//SETUP
void setup(){
  Serial.begin(9600);
  pinMode(pirPin, INPUT);
  pinMode(ledPin, OUTPUT);
  digitalWrite(pirPin, LOW);
  //give the sensor some time to calibrate
  Serial.print("calibrating sensor ");
    for(int i = 0; i < calibrationTime; i++){
      Serial.print(".");
      delay(1000);
      }
    Serial.println(" done");
    Serial.println("SENSOR ACTIVE");
    delay(50);
  }
 ////////////////////////////
//LOOP
void loop(){

     if(digitalRead(pirPin) == HIGH){
       digitalWrite(ledPin, HIGH);    //the led visualizes the sensors output pin state
       if(lockLow){
         //makes sure we wait for a transition to LOW before any further output is made:
         lockLow = false;
         Serial.println("---");
         Serial.print("motion detected at ");
         Serial.print(millis()/1000);
         Serial.println(" sec");
         delay(50);
         }
         takeLowTime = true;
       }

     if(digitalRead(pirPin) == LOW){
       digitalWrite(ledPin, LOW);    //the led visualizes the sensors output pin state

       if(takeLowTime){
        lowIn = millis();          //save the time of the transition from high to LOW
        takeLowTime = false;       //make sure this is only done at the start of a LOW phase
        }
```

```
    //if the sensor is low for more than the given pause,
    //we assume that no more motion is going to happen
    if(!lockLow && millis() - lowIn > pause){
      //makes sure this block of code is only executed again after
      //a new motion sequence has been detected
      lockLow = true;
      Serial.print("motion ended at "); //output
      Serial.print((millis() - pause)/1000);
      Serial.println(" sec");
      delay(50);
      }
    }
  }
```

```
#define BLYNK_PRINT Serial
#include <ESP8266WiFi.h>
#include <BlynkSimpleEsp8266.h>

char auth[] = "Enter your Authentication code from Blynk";
char ssid[] = "wifi";
char pass[] = "wifiwifi";
int Main = D4;       // Virtual Switch
int Buzz = D6;       // Buzzer / Hotter Pin ( RELAY DATA)
int Sensor = D1;     // Sensor ( RCWL-0516 ) Pin

int sensorval = 1;  // RCWL-0516 Temp Storage Int

void setup()
{
 Serial.begin(9600);
 Blynk.begin(auth, ssid, pass);
 pinMode(Sensor,INPUT);
 pinMode(Main,INPUT);
 pinMode(Buzz,OUTPUT);
digitalWrite(Buzz,HIGH);

}

void runer()
{
sensorval = digitalRead(Sensor);
Serial.println(sensorval);
if (digitalRead(Main) == 1 && digitalRead(sensorval) == 0)
{
 Serial.println("Intrution Detected");
  Blynk.notify("hey man someone entered your room");
  Blynk.email("INTRUTION ALERT!", "Hey man someone tried to enter your room, better be careful");
  Blynk.tweet("hey man someone entered my room not cool #PrivacyMatters #MisionCritical ");
  digitalWrite(Buzz,LOW);
  delay(7000);
  digitalWrite(Buzz,HIGH);
 }
}

void loop()
{
 runer();
 Blynk.run();
}
```

2.5.2 Configuring the Blynk App

After the user installs the Blynk app on the smartphone, an account has to be created in the app to access its services. The first time the app is opened, it will ask to either sign in or create an account. Create an account and add a new project to get started. Each project has its own authentication code which is used by the code to communicate with that particular model. To interface with our components, we need to add widgets to our model. To add widgets press „+" to add to the model. The app provides a neat interface to add all the required widgets and setting them up according to the code. The Blynk needs to be running in the background for the user to get real time notifications.

3.0 Results

The experiment was carried out in Intel Core i5 3.20GHz processor, with 8 GB RAM, 19" LCD monitor with hard disk as 500 GB. The software required are Blynk App, Arduino IDE, in windows operating system using C++ programming language. The resultant system was checked thoroughly by repeating the motion of passing in front of the PIR Sensor multiple times to see if each time a notification is sent or not and by triggering the buzzer on or off from which showed that the system works in the intended way and flawlessly. To test the endurance of the hardware, the setup was left turned on for a couple of hours and tested afterwards. The components got heated which is acceptable but still worked and the notification.

4.0 Discussion

The sensors strategically placed in the house informs the home owner as soon as an intruder enters by sending a Push notification. The user will get this notification irrespective of whether the phone is locked or unlocked or even if any other app is opened at the moment. This was the main objective of the project, which is the user feels safe and not worry about any intrusion or break-ins when he is away from home. This setup can also be used in commercial offices, making use of Reed Sensors on the entrance doors that are restricted for certain personnel, such a system will immediately inform the administrator of any unauthorized personnel trying to access such an area. Therefore, the extensibility and applicability of such a system is only limited only by the thinking. Another important component of the project is the connectivity between the ESP8266 (WiFi module) and the Blynk server. The system successfully connected to the Blynk server using the authentication token and the Blynk libraries. As a result, we were able to get the notification on our smart phones as soon as there was any change in the status of the PIR motion sensor. Also, the additional ability to control the alarm remotely is very beneficial and can be very useful in some unforeseen circumstances. It was also observed that the Blynk app worked smoothly and carried out all communication between the hardware and the app very accurately. The developed system can also be used to in industrial and commercial applications such as offices, warehouses and other areas where some areas are reserved for authorized personnel only or other places where safety and precautions are of primary

concerns such as internet server room of a big multinational companies from where corporate data can be stolen. The system can also be easily upgraded to add extra safety features such as cameras, reed sensors, etc. for increased safety. The system can also further be developed by adding an RFID scanner so that the authorized users need only carry a RFID or NFC tag with them on their person. The RFID scanner will work by scanning the tag wirelessly and if the user is authorized to enter, the alarm system will be disabled for some time so that the user can enter. As mentioned earlier, the system microcontrollers Arduino and ESP2866 can be replaced with a microcomputer known as Raspberry Pi. This offers more a robust approach to IoT application as it allows programmers and hobbyist to be more creative.

References

Anitha A. (2017) "Home security system using internet of things" Published under license by IOP Publishing Ltd IOP Conference Series: Materials Science and Engineering, Volume 263.

Gillis, A. (2021). "What is internet of things (IoT)?". *IOT Agenda*. Retrieved 17 August 2021.

Home automation (2022, February 27). In Wikipedia. https://en.wikipedia.org/wiki/Home_automation

LaPlante, Phillip A.; Ovaska, Seppo J. (2012). Real-Time System Design and Analysis (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc. pp. 84–85, 100–102. ISBN 978-0-470-76864-8.

Macfadden C. (2018) Raspberry Pi and Arduino: What's the Difference and Which Is Best for Your Project? Source: Sven.petersen/Wikimedia commons

Monk S. (2011) "Programming Arduino: Getting Started with Sketches" McGraw-Hill Education.

APP CULTURE AND THE BENEFITS OF MOBILE APPS IN CONTEMPORARY TECHNOLOGY-DRIVEN ECONOMY

Ike, U.K[1]. Ohanuma, C.K[2]. Ajaero, G.N[3]. Ovwonuri, A.O[4], Dimoji, T.E[5]

[1,3]Department of Computer Science, Imo State Polytechnic Umuagwo, Imo State.
[2]Department of Information Technology, Federal University of Technology Owerri, Imo State
[4]Department of Software Engineering, Federal University of Tech. Owerri, Imo State
[5]Department of Computer Science, Federal College of Agriculture, Ishiagu, Eboyi State
uchebenzene@yahoo.com; godsgraceng@gmail.com; Abraham.ovwonuri@futo.edu.ng; dimogit@fcaishiagu.edu.ng

ABSTRACT

The paper discusses the concept of app culture, a consequence of the pervasive applications of mobile apps in contemporary tech-driven economy. The 21st century has been defined by applications of and advancements in the field of information and communication technology. ICT has become an integral part of our daily life and has served as a big change agent in different aspects of business and society. These advancements in ICT are evident in the emerging trends, one of which is mobile applications, a dominant theme in the area of mobile computing due to its huge impacts in business and society. The paper also covers the scope of mobile apps, examines mobile apps as new application model in information systems and highlights the benefits of mobile apps in contemporary tech-driven economy.

KEYWORDS: Mobile Apps, Mobile Devices, Technology, ICT, Information Systems

INTRODUCTION

The 21st century has been defined by the applications of and advancements in information and communication technology. ICT has become an integral part of our daily life and has served as a big change agent in different aspects of business and society. It has proven game changer in resolving economic and social issues. The field of ICT has seen great advancements and changes in the last decade. These advancements are evident in the emerging trends in ICT, one of which is mobile applications which has become a dominant theme in the area of mobile computing due to its huge influence in business and society.

According to Wikipedia, mobile computing is a human-computer interaction in which a computer is expected to be transported during normal usage, which allows for the transmission of data, voice and video. Mobile computing involves mobile communication,

mobile hardware and mobile software. Ike (2022) defines mobile computing as the set of technologies, products, services and operational strategies and procedures that enables end-users to gain access to computation, information and related resources and capabilities while mobile. Mobile means access in motion and is therefore unrestricted in a fixed geographical location. Mobile can be seen as also meaning access in a fixed location via equipment that users can relocate as required, but is stationary while in operation (Ike, 2022).

Mobile devices include notebook PCs which are functionally equivalent to desktop PCs, mobile phones or smartphones and a variety of products aimed at vertical and specialized applications such as those used in medical applications, surveillance and security and so on.

Mobile applications are software designed to run on these mobile devices. They are available as download from various mobile app stores such as apple appstore, Google playstore and so on. Some of the mobile apps are available for free whereas some involve download cost. Today, many smart phones come with several apps packed in them as pre-installed software. These pre-installed apps include web browsers, email, calendar, mapping program and so on. Some pre-installed apps can be removed by an ordinary uninstall process thus leaving more storage space for desired ones. Where the software does not allow this, some devices can be rooted to eliminate the undesired apps (Ike 2022). Facebook, Whatsapp and YouTube are amongst apps categorized as social media apps which often find a place as pre-installed apps in many trending mobile device brands.

MOBILE APPLICATIONS AND TYPES

A mobile application is a computer program running on a mobile device. Inukollu et al. (2013) defines mobile apps as application software designed to run on perspicacious phones, tablet computers and other mobile devices. An app makes sense or is desired if the goal is to have an interactive engagement with users, or to provide an application that requires working more akin to a computer program than a website. Mobile applications often stand in contrast to desktop computers and web applications which run in mobile web browsers rather than directly on the mobile device (Inukollu et al. 2013). The size of a mobile device is the most significant characteristic and limitation of mobile devices because it determines other limitations of mobile devices. The most important of them are battery power, the size of the display and uncomfortable input methods (Panis et al. 2002). The limitations of mobile

devices are essential because they determine the limitations of mobile apps. The limitations of mobile apps and mobility itself make the mobile app model distinct from other application models.

The three types of mobile apps are: native, web and hybrid mobile apps (Ike, 2022).

Native apps

These are mobile applications that are designed specifically for a mobile operating system. What this means is that an app developed for any of the platforms such as iOS, android, blackberry, Symbian or windows will be functional only on the respective platform for which it was designed. Native apps are targeted towards a particular mobile platform; therefore an app that is intended for apple device cannot run on android and vice versa. Meanwhile to overcome this challenge, many businesses today, develop apps to run on multiple platforms. The main purpose for creating native apps is to ensure best performance for a specific mobile operating system. A typical example of native app is the calculator app on the iPhone.

Web-apps

These are apps that are written in HTML or CSS and typically run through a browser. Internet access is typically required for web apps to perform at optimum. With the internet access, web-based apps are able to use all features compared to offline usage. Most, if not all user data is stored in the cloud and as such, only small amount of memory space in a user device is required. However, web apps run slower than the equivalent native app. Examples of web apps are Netflix, Google docs and Drop box etc.

Hybrid apps

The concept is a mix of native and web-based apps. Hybrid mobile apps are made to support web and native technologies across multiple platforms. They are easier and faster to develop as it involves the use of single codebase which works on multiple operating system. Despite these advantages, hybrid apps are slower in performance and may not bear exactly the same look and feel on different mobile operating systems. Apps developed using Apache Cordova, Flutter, Xamarin, React Native, Sencha Touch and other frameworks could be classified as hybrid apps.

THE SCOPE OF MOBILE APPLICATIONS

Mobile apps offer information services in various areas. The scope of mobile applications is to enable the use of applications and access to important information and basic level of services in the state of mobility (Rupnik &Krisper, 2013). The scope of mobile applications does not include the several functionalities provided by classical applications. Thus, the services and functionalities offered by mobile apps are appropriate for mobility and the needs of a mobile user. In other words, mobile apps are designed to offer mobility adapted and mobility suitable mobile services. From a business perspective, (Rupnik &Krisper, 2013) posited that mobile applications must provide a basic level of functionality, access to important information and the possibility to be informed about exceptional, unexpected and other unusual situations happening within an organization that a person belongs to. Mobile applications signify the connection of a mobile user with his organization and its information system.

MOBILE APPS AS A NEW APPLICATION MODEL IN INFORMATION SYSTEMS

Mobile apps represent a new application model introduced to information systems and thus, a relatively new area of research. Mobile applications are enabled by the convergence between information technologies and telecommunication technologies necessitated by the information needs of contemporary society (Rupnik & Krisper 2003). The information society is enabled by technology, but is far more than just technology driven society. It is a complex and multidisciplinary society driven by knowledge, innovations and development (Rupnik & Krisper 2003). One of the most representative characteristic of the information society is the convergence between information and telecommunication technologies. The convergence between several technology sectors offers the opportunity for the emergence of new services and mobile applications are the consequence and the result of the convergence mentioned above (Smith, 2002).

The basic task of information systems is to provide information support for business processes. The introduction of mobile applications to information systems enables the information support of business processes in the state of mobility. (Rupnik & Krisper, 2013)

outlined the following types of mobile applications as those required by information systems in business organizations:

1. Messaging systems: enables information broadcasting for the mobile users.
2. Mobile transaction systems: provides users with transaction oriented information support.
3. Mobile decision support and controlling system: enables managers and decision makers to access important information about business processes (Sharaf, 2002).

THE CONCEPT OF APPS CULTURE

Taylor (1871) defined "culture' as an umbrella term which encompasses the social behaviour and norms found in human society, as well as knowledge, beliefs , arts, laws, customs, capabilities and habits of the individuals in these groups. People acquire culture through the learning process of enculturation and socialization, which is shown by the diversity of cultures across societies. At its most basic level, "culture" is seen as the way of life of a particular people especially as shown by their ordinary behaviour and habits. On the other hand, mobile apps or apps as have been defined earlier are computer application software basically designed to run on mobile devices.

Thus, "app culture" is a combination of two words; "apps" from the software domain and "culture" from the sociology domain. In this paper, the term is used to describe the widespread download and use of mobile applications in smartphones which cuts across all genders and spheres of human endeavor.

The mobile devices being utilitarian, user-friendly and accessible has made it the most popular and indispensible expedient for human essentials from the past few years (Malavolta et al. 2015). To say that the use of mobile phones has dramatically increased over the past decades is not an exaggeration. Most adults today are cellphone users and live in a household in which there might be a cellphone without a landline phone. According to the world mobile applications market-advanced technologies, global forecast (2010-2015), there were about 6.4 billion applications that were downloaded in 2009 alone which generated revenue of $4.5 billion in the same year (Inukollu et al. 2014). A market research conducted by IDC- a global market intelligence firm predicted that the market for mobile applications will continue to

accelerate as the number of downloaded apps is expected to increase from 10.9 billion worldwide in 2010 to 76.9 billion in 2014. Similar growth will be observed in the revenue of mobile apps worldwide surpassing $35 billion in 2014 (Inukollu et al. 2014).

As the mobile phone has transformed from a voice device to a multi-channel device and then to an internet accessing mini-computer, a large market of mobile software application has arisen. The emergence of the pervasive connectivity is changing the way people interact, share creations, and exploit vast libraries of materials that are generated for the internet. The widespread embrace of mobile technology has spawned the development of an apps culture. Smart phones now permeate cultures across the globe and as they become more powerful as connected multi-media handheld devices, a new ecosystem of computing applications is being created around them.

At the individual level, mobile apps provide the following:

1. An avenue for people to connect and communicate with others from any part of the globe. This service is provided by social media apps such as Facebook, Whatsapp and so on.
2. An avenue to share "selfies" (pictures of one's self) on social media.
3. An avenue for online shopping and sales. Ecommerce platforms provide these services.
4. An avenue to perform banking transactions through bank apps.
5. An avenue to play games and watch movies as well as listen to music.

THE BENEFITS OF MOBILE APPS IN BUSSINESS

Mobile applications have become a significant part of the technology-driven world we live in and can enhance a person's life, enjoyment and productivity. Apps can also be used by companies both large and small to streamline production and increase ease at work. Any company hence seeking to improve business processes and boost profitability significantly would have to embrace mobile presence (Ike, 2022).

Today, most of the traffic online is powered by mobile devices. Mobile apps put businesses in the right position to leverage this trend. From branding perspective to customer service, sales and marketing, there is hardly any department of a business that would not benefit from

incorporating a mobile app. Below are some of the reasons why mobile app development has become crucial in business:

Provide more value to customers: Today's business owners now rely heavily on technology to initiate and complete transactions with customers. Mobile apps are one of the major elements of modern tech that are consistently changing how consumers shop and satisfy their needs. With the increasing pace of technological advancement, consumer expectation also grows and mobile apps offer a great channel for meeting their expectations. Modern consumer behaviour prioritizes convenience and apps make that available at their fingertips. Through mobile apps, retail stores can make goods and services more accessible and easy to use. Also businesses can operate complaint management systems, online registration, intelligent database and offer much more valuable services using mobile apps.

Build a stronger brand: Since mobile devices are now a highly integral part of our everyday lives, seeing a high amount of usage time, businesses can leverage this channel to their brand benefits. The more branding elements (logos, color combination etc) a business appropriately applies within its app, the better the impact they one expect on the brand marketing results. More importantly, the more value a business offers customers via mobile app, the more interest customers will develop in the brand.

Connect with customers fast and easy: One thing that greatly influences customer satisfaction is easy access to information about the business. A dedicated mobile app is a good way to deliver to customers a reliable communication link. Unlike websites and other channels, a mobile app is much more reliable, personalized and convenient. Similarly, employees can easily access this information on the app and will help them serve customers faster and better. Beyond support, mobile apps are great for requesting and collecting customer feedback. A live chat support feature added to a business app drives customer engagement and invariably boost customer satisfaction.

Improve customer loyalty: Loyal customers are not only useful for direct revenue generation but they also serve as marketing channels. Businesses can boost customer loyalty by featuring loyalty programs on mobile apps. Loyalty programs can be integrated into mobile app experience by several ways. One is to offer customers instant rewards for using the app to

make a purchase or other kinds of transactions or providing timely personalized product recommendations and so on.

Have a competitive edge: In the digital world of marketing we live in today, keeping up with the competition may be really arduous task. Though there are many factors to keep track of, a mobile app gives a business the significant competitive edge. Ever-changing consumer behavior seem to be the primary factor behind this; the demand for instant communication, increased reliance on mobile devices and communication speed is vital for customer satisfaction compared to conventional websites and other channels. Mobile apps do a better job catering to the needs for speed. Considering the number of ways an efficient mobile strategy improves business workflow, its overall effect on business revenue generation will eventually become evident.

Reach higher customer engagement level: Mobile apps help business to engage customers at a much higher rate. Customer engagement helps businesses to increase brand loyalty and invariably revenue generation. One way to use mobile app to ensure optimum engagement is to apply segmented targeting. This involves creating multiple user categories and controlling the type of content delivered to each user segment. User segmentation allows business owners to send permeated in-apps messages, provide accurate user recommendations and understand average customer's journey with the business. Loyalty programs, discounts, freebies and continuous feature upgrades can also promote customer engagement.

Build a direct and personalized marketing channel: Another area of any business that can benefit a lot from a customized mobile app is the marketing department. The first staring advantage that digital marketers enjoy when a business goes mobile is the direct access to user information. The data collected from user sessions and entry points into a business's mobile app can be very useful for improving marketing campaigns. Once the marketers have all the data they need, an app also allows the business to deliver content to users more efficiently than other traditional marketing channels.

Utilize social media channels: Engagement is the driving force of the internet. The more time people spend engaging on the business app, the better it is for the business. When it comes to driving engagement, a god option is to integrate business app effectively with social media channels. Businesses too can begin by first advertising on social media platforms such

as Facebook and YouTube. This will get the app the needed attention immediately after launch. Also social media buttons could be added to the app and to get users to use them may require some incentives.

Offer instant, better customer service: As mentioned earlier, instantly communicating with customers is essential for good customer experience. Studies in customer service industry prove that most customers now seek support via mobile. Mobile customer service benefits both the business owner and the customers in significant ways. Customers get the chance to serve themselves, customer service personnel receive fewer questions since apps reduce the pressure on employees and customers overall experience is greatly improved.

Find valuable customer insight: To be successful in business in this 21st century requires customer insight. Any winning marketing strategy in today's competitive business environment must ensure it revolves around customer behavior and the mobile app can serve as a reliable and valuable source of customer insights. Usually customers are often willing to share some important information if business offers value in return. There are several analysis tools that provide utility to gather insights from users. A useful strategy is to track metrics such as daily, weekly and monthly users, user demographics such as age, gender, device types as well as downloads and uninstalls retention rates and so on.

Have more control: An app affords a business owner the opportunity to control how the business is run. There are several areas where businesses can make the best of the extensive control they get. This includes areas such as branding, security, scalability and engagement as well as customer interaction.

Provides unique services and features: In a bid to stay ahead of the competition, businesses would have to offer customers some unique services. To achieve this requires that new features are added to the mobile app. A good mobile app packed with unique, interesting features may provide all the advantages that a business needs to stay ahead.

Have an avenue for customer feedback: Mobile apps are never lacking in interfaces or mechanism for capturing user feedbacks. Depending on the purpose the feedback is intended, each interface has their advantages and disadvantages. Some of the customer feedback mechanism includes widgets, surveys, rate my app prompts etc.

CONCLUSION

The impacts of mobile apps to the individual and business organizations cannot be easily exaggerated. The paper has shown the usefulness of mobile apps in a business context and highlights the need for individuals and businesses to take advantage of its far reaching impacts. The scope of mobile apps however does not include functionalities provided by classical applications. Mobile apps are designed for users who are required to satisfy their information needs and accomplish tasks even while in motion.

REFERENCES

Ike, U.K., Durunna, I.L., and Orji, E,C. (2021) "Introduction to computers and information technology" Ambix publishers Owerri, Nigeria.

Ike, U.K., Udegbe, I.V., Orji, E.C., and Ukachukwu, N,T. (2022). "Computer application packages and mobile apps; theory and practice. Ambix publishers, Owerri Nigeria.

Inukollu, V.N., Keshamon, D.D., Kang, T and Inukollu M. (2014) "Factors influencing quality of mobile apps; role of mobile app development lifecycle". International journal of software engineering and applications (IJSEA) vol. 5 No.5

Malvolta, I., Rubberto, S., Soru, T., Terragni, V. (2015) "End users perception of hybrid mobile apps in the Google playstore. In proceedings of the IEEE 3rd international conference on mobile services Newyork NY USA.

Mobile application history, http;//en.wikipedia.org/wiki/mobile_app

Panis, S., Morphis, N., Felt, B., Reufenheuser, A., Bohmj, N., & Nits, S. (2002). "Mobile commerce scenarios and related business models. In proceedings of the 1st international conference on mobile business pdf. Greece Athens.

Rupnik, R., and Krisper M. (2001) "Mobile applications; the consequence and demand of information society. In proceedings of the international conference information society Slovenia Ijubljana.

Rupnik, R., and Krisper, M. (2003). "The role of Mobile applications; in information systems" In proceedings of the 2nd international conference on mobile business Vienna Austria.

Rupnik, Rok and Krisper, Marjan (2013). "Mobile applications: a new application model in information systems. Retrieved from http;//researchgate.com

Sacher, H (2001). "Uncovering the wireless interaction paradigm" (available from www.baychi.org).

Smith, H.A, Kulatilaka, N., & Venkatramen, (2002). "Development MIS practice III; Riding the wave extracting value from mobile technology. Communication of the association for information systems. http://cais.ais.net.org

Uskov, V.L (2013). "Mobile software engineering in mobile computing curriculum". Interdisciplinary Engineering Design Education Conference (IEDEC).

DEVELOPMENT OF PHISHING SITE DETECTION PLUGIN TO SAFEGUARD ONLINE TRANSCATION SERVICES

Christiana Ugochinyere Oko[1], Anthony Ifeanyi Otuonye[2]

[1,2]Department of Information Technology,
Federal University of Technology Owerri, Nigeria.

Corresponding Author's email: anthony.otuonye@futo.edu.ng

Abstract

In this research project, an efficient phishing website detection plugin service was developed using machine learning technique based on the prevalent phishing threat while using existing web browsers in critical online transactions. The study gathered useful information from a dataset consisting of 11,000 data points with 30 features downloaded from phishtank. A unique architectural framework for detecting phishing websites was designed using random forest machine learning classifier based on the aim and objectives of the study. The model was trained with 90% (9,900) of the dataset and tested with 10% (1,100). System model was carried out using key component diagrams of the Unified Modeling Language (UML). Implementation was done using the Python programming language. Front end system was developed using Microsoft Visual Studio Code, Jupiter Notebook, Anaconda Integrated Development Environment, HTML/CSS and JavaScript, and for easy integration into existing web browsers. Result obtained revealed that the proposed model had an accuracy value of 0.96, with an error rate of 0.04, and 0.97 precision, recall value of 0.99 and f1-score of 0.98 which is a better performance in comparison to exiting models. For future research efforts, we recommend more focus on improved security features for such plugins, and more phishing adaptive learning properties, for a more reasonable application to other web browsers to accurately detect real-world phishing situations.

Keywords: Plugin, Phishing website, Machine Learning, Random Forest, Online transaction.

1.0 INTRODUCTION

1.1. Background Information

The convenience and speed of an online purchase and transaction cannot be overemphasized. Payment for goods and services can be made at any time and from any location via the internet. Nevertheless, Internet users are subject to wide range of online threats that can result in monetary losses, identity theft, data breaches, damage to a company's brand and a loss of trust in e-commerce and online banking The existence of growth of Internet frauds, one of which is the phishing attack, is critical to these advancements in e-commerce [1]. This raises security concerns as some websites for example, banking and shopping websites are forged. These forged websites are called phishing websites. Phishing is a method of social engineering or cybercrime used by malicious persons to imitate real online pages in order to steal personal information from users or clients. As a result of this, individuals and institutions have lost very huge amount of money because unknown to them they give out confidential information to fraudsters.

The word "phishing" was coined from the concept that Internet criminals were 'fishing' personal and financial information from the sea of unsuspecting Internet users. In 1996, hackers who were hijacking American Online (AOL) accounts by phishing passwords from unsuspecting AOL consumers coined the term "phishing." Phishing was first used on the Internet in January 1996 alt.2600 hacker newsgroup; however, it's possible that it was first used in the printed issue of the hacker newsletter "2600." "Ph" is a hacker's substitute for "f," and it relates to the first sort of hacking, "phreaking."

John Draper (aka. "Captain Crunch"), the first hacker, created the term "phreaking." He pioneered "hacking" when he created the infamous Blue Box, a device he used in the early 1970s to hack telephone systems. "Phone Phreaking" was the name given to this first type of hacking. The blue box emitted tones that allowed the user to manipulate the phone switches, thereby allowing them to make free long-distance calls or charge calls to a different number, and so on. Many hacker pseudonyms and hacker organizations get their "ph" spelling from this source [2].

According to section 58 of the Cybercrime (Prohibition, Prevention, etc.) Act 2015 (CPPA), 'Phishing' is the illegal and deceptive method of trying to obtain sensitive information such as usernames, passwords, and credit card details by impersonating a reputable organization or enterprise in digital communication such as an email from what appears to be your bank asking a user to change their password or disclose their identity so that such information can later be used to commit fraud [3].

The Economic and Financial Crime Commission (EFCC) in 2006 reported on the amount of online crime in Nigeria as well as the country's standing among other countries with high levels of cybercrime. A retired civil worker and two other accomplices were accused of defrauding a German citizen, Klaus Wagner, of USD 1, 714,080 via the internet, according to the publication. Cloned websites, fraudulent claims, internet purchases, and other e-commerce fraud are among the most common forms of cybercrime in Nigeria, according to Ribadu, the founding Chairman of EFFC [4].

According to KPMG Forensic Services report in 2016 on "Top Five Fraud Trends in Nigeria's Commercial Banks", Nigeria experienced real losses of N485,194,350 and N6,215,987,323 due to phishing frauds in 2013 and 2014, respectively [4]. Also, according to a research published by the FBI, phishing scams cost a minimum of 2.3 billion dollars between October 2013 and February 2016 [5].

High number of COVID-19 phishing emails was received by internet users in Nigeria during the pandemic in the year 2020. This included texts with promises of various incentives, including government offers as palliatives to persuade vulnerable people in need of aid as a result of the pandemic. Attackers also emailed a counterfeit URL (https://covid-19-fg-grant.blogspot.com/?=1) promising a free Internet bundle as part of COVID-19's stay-at-home package in exchange for victims' personal information. Section 32 of the Cybercrime Act of 2015 makes phishing attacks illegal in Nigeria [6].

Planning, setup, attack, collection, and identity theft and fraud are all steps in the phishing process [7]. The process of a phishing attack begins first with a phisher creating a website imitating a real website. Then they will send the URL of the phishing website to their victims through email or SMS. When the victims click on the URL, it takes them to the fake website created by the phisher. The victims enter their credentials such as email and passwords on the fake website. The phisher then uses these credentials to access their victims' account on the legitimate website.

The methods used by attackers to steal sensitive information from their victims through phishing can be divided into two categories:

(a) Social Engineering Attacks: Social Engineering refer to the plans and strategies hackers employ to manipulate individuals to visit fake websites and submit personal details like bank information, usernames, and passwords [8]

(b) Technical Subterfuge: Technical subterfuge methods infect computers with malware in order to steal credentials directly, usually by using systems to intercept users' online account user names and passwords and distorting local navigational infrastructures to divert users to bogus websites [9].

In this study, we will focus on social engineering attacks. Some examples of social engineering attacks include:

(i) **Email spoofing:** Here, the recipients of the email are routed to a phishing website, where they are requested to enter sensitive data such as account numbers and passwords.

(ii) **Phone phishing:** A phone user receives a phone call or message on social media or Short Messaging Services (SMS), most of the time purporting to be from a bank, requesting users to call a phone number to either upgrade their account or address difficulties with their bank account.

(iii)**Web spoofing:** A phisher could create a website that appears to be legitimate, misleading victims to believe it is the genuine site and provide personal information to the phisher.

In recent times, 65 percent of phishing efforts begin with a link received in an e-mail, e-mails has been the major method for circulating phishing links. Phishing URLs are distributed through susceptible websites including blogs, forums, instant messaging (IM) on social networks, SMS and multimedia (MMS) communications, chat rooms, and fake browser programs [10].

The majority of phishing attacks begin with an unsuspecting victim receiving an electronic mail(e-mail)or a Short Message (SMS). Phishing messages may appear to be legitimate messages concerning an organization's activities and practices, online transactions, or social exchanges, and they ask recipients to take action [11]. Phishing messages are usually divided into two categories:

(i) Phishing emails containing threats (for instance, "your account has been breached," "your account will be deactivated in 24 hours," or "Urgent: account security update") or

(ii) Phishing emails containing promises of advantages or prizes (for instance, "you have won" or "congratulations").

The first sort of phishing email has been most widely employed in phishing study [12]. These messages usually contain a Uniform Resource Locator (URL) link. When the receiver opens the link, it directs them to the illegal website that has been created by the hacker where they divulge personal and sensitive information for example: login details, PIN, ATM card number. The hacker then uses this information for financial gains.

A major motivating factor for phishing attacks is financial gain. Other motivating factors have also been identified: identity theft, industrial espionage, identity trafficking, malware dissemination, password harvesting, fame and popularity, and exploiting security flaws [13]. Hackers use information stolen from victims to carry out fraudulent transactions thereby putting financial institutions at risk. In a world where technology is rapidly taking over primitive trends, there is need for improved security especially in online financial transactions to reduce phishing attacks on unsuspecting internet users.

In this study, a model that is less complex, has a high performance, available, accessible to all, and has several inbuilt functionalities is designed to detect phishing websites and reduce online fraud. This model is a plugin that can be integrated into web browsers and has the capability to notify users when they visit a phishing website thereby forestalling a phishing attack.

1.2 Objectives of Study

In this research paper, our objective is to:

(i) Develop a model for the extraction of website characteristics for use in classifying websites as either phishing or legitimate.

(ii) Develop a machine learning-based model using Random Forest classifier to classify websites based on their characteristics as either a 'phishing website' or a 'legitimate website'.
(iii) Further classify the level of severity of the phishing website detected as either 'high', 'moderate' or 'low'.

Specifically, we shall use random forest technique to detect phishing websites and the severity of the phishiness of the phishing assault. The datasets used in training the model are limited to 11,000 datasets with 30 characteristics. The internet-based phishing website detection model can be integrated into web browsers as a plugin.

2.0. Literature Review

In the past, a number of Researchers have proposed the use of various phishing prevention models using various techniques and approaches in trying to stop phishing attacks on unsuspecting internet users. Their efforts has led to the creation of many technical counter-measures such as email filtering, phishing detection software and anti-phishing toolbars. Some of these literature were carefully studied during the course of this research. Some highlights of these reviews are presented in this section.

2.1. Empirical Literature

Zhang et al., (2011), created the Carnegie Mellon Anti-phishing and Network Analysis Tool (CANTINA) which is an anti-phishing tool. Surface level characteristics, textual content, and visual content make up the content-based anti-phishing, also known as leveraging the attributes of websites. The content of a web page refers to all of the information on the page, including the domain name, URL, hyperlinks, terms, images, and forms embedded in the page. Using a text classifier, an image classifier, and a data fusion algorithm, this framework synthesizes numerous clues, i.e., textual content and visual content, from a given web page and automatically flags a phishing web page. The Naïve Bayes rules were used to model this text classifier. This program examines the content of a website and classifies it as 'phishy' or 'legitimate'.

In their e-banking phishing website detection model, Martin et al., (2011) used artificial neural network techniques to identify the major phishing characteristics and significant traits of phishing or indicators in the e-banking phishing website archive data, using URL and domain identity, security, and encryption, source code and JavaScript, page style and contents, web address bar, and social human factor as indicators. Genuine, doubtful and legitimate were the input data for each. Using these data, rules was formed and the network trained to provide output that ranged from very legitimate, legitimate, suspicious, phishy and very phishy.

A. Jain and Richarya, (2011). This paper described a method for detecting phishing emails based on link-based attributes. This work's key contribution is the use of features such as

visible links, invisible links and unmatched URLs. The algorithm, when combined with the web browser prototype, alerts users to potential phishing attacks and prevents them from visiting questionable websites. This prototype included a C#.Net implementation of a web browser. To open mails, the user will need to utilize this web browser. Because the browser's core is wrapped around Internet Explorer's engine, users will not notice a change while sending standard emails. The user is alerted of the forged email suspicion and advised to delete the email.

Chandan *et al.*, (2014) designed a phishing detection approach using Artificial Neural. A network that classifies the security of a webpage by examining the source code. They extracted several phishing characteristics to assess the websites' security and checked the webpage source code; if a phishing character is identified, the weight will be reduced. Then, based on the final weight, they determined the security percentage; a high percentage indicated a secure website, whereas a low percentage indicated a phishing website. The legitimate and phished website's final values were compared, and thus detection was performed. The authors concluded by saying that a neural network with a sufficient number of hidden units can attain satisfactory accuracy.

3.0. Methodology

3.1. Agile Methodology

For this study, the agile software methodology was adopted. The agile strategy promotes relatively shorter iterations rather than the formerly popular waterfall methodology's lengthy release cycles. The iterative approach breaks the development process into smaller parts. Each part contains the planning, design, development, and testing steps. Breaking down the objectives into subtasks is an added advantage, allowing the software developer to easily determine how much of the software has been done at any given time. This allows project managers to determine if the project is progressing as planned. When a project falls behind schedule, the agile approach quickly identifies the risk, allowing managers to interact with stakeholders to develop a risk management plan [14]. Figure 2 illustrates the agile software development process.

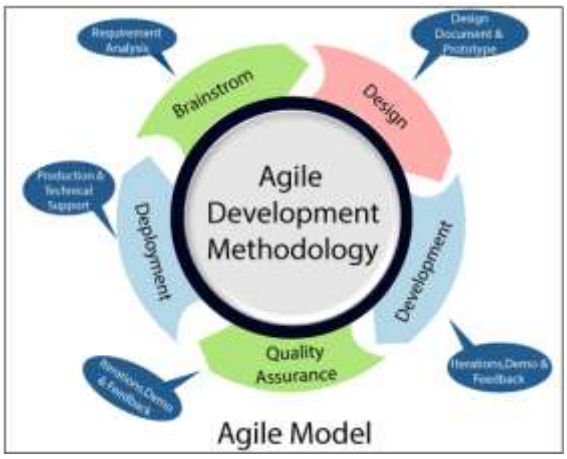


Figure 2: Agile Software Development Process

3.2. Method of Data Collection

(a) Online Repository

The dataset consist of 11,000 data points with 30 features downloaded from phishtank was used. The features are: Having IP Address, URL Length, Shortening Service, Having @ Symbol, Double Slash Redirecting, Prefix Suffix, Having Sub Domain, Secure Socket Layer (SSL) State, Domain Registration Length, Favicon, Using Non-Standard Port, HTTPS Token, Request URL, URL of Anchor, Links in Tags, Server Form Handler (SFH), Submitting Information to Email, Abnormal URL, Website Redirect Count, Status Bar Customization, Disabling Right Click, Using Pop Up Window, Iframe, Age of Domain, DNS Record, Web Traffic, Page Rank, Google Index, Links Pointing to Page, and Statistical Report.

(b) Published Articles

Under this section, the researcher downloaded and reviewed 27 articles based on phishing detection techniques, development and deployment in order to critically assess the work done on the study area and to justify the gap in knowledge established.

3.3 System Architecture

System architecture is a mental picture that explains the organization, performance, and other qualities of a system. It depicts how the users interact with the system's components.

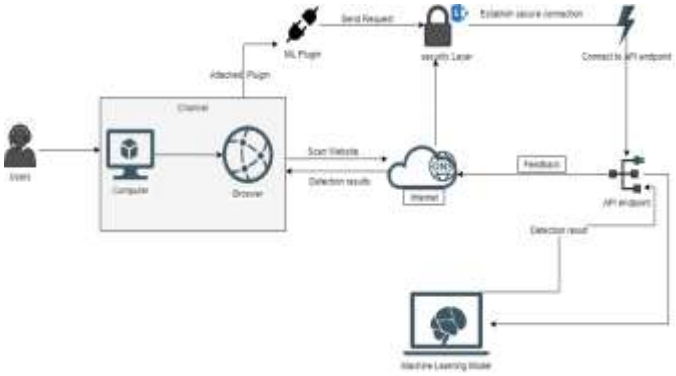


Figure 3: System Architecture of the Proposed System

3.4. Proposed Algorithm

For the Proposed Algorithm, the Random forest classifier will be adopted. It is a supervised machine learning method that uses statistics and ensemble learning. Because of its ensemble learning methodology, it is widely employed. Because trees make the forest, this method combines multiple Decision Trees for a more exact prediction, giving it the name random forest algorithm. The process of combining results from multiple trees is called aggregation.

It is random because two random processes are applied: bootstrapping and random feature selection. Bootstrapping is the process used to create new data. It ensures that there is no repetition of data for every tree. Random feature selection helps to reduce the correlation between the trees. Regression and Classification are the most common algorithms for this technique [15].

3.5. Experimental Setup

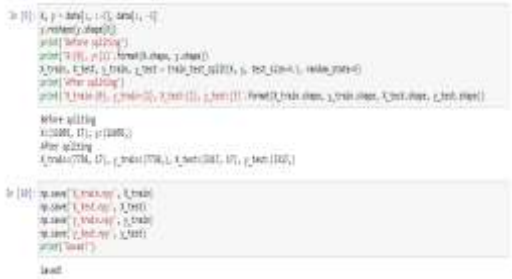
This section summarizes the experiment that was conducted for this research. We begin with downloading a collection of 30 content-based, URL-based and domain information characteristics datasets from phishtank repository. After preprocessing of dataset, it was divided into two for training and testing purpose. At first it was divided into 70% training, 30% testing dataset. To better improve the model, the test data was reduced to 10% and the training data increased to 90%. The proposed methodology was implemented in python programming language.

Dataset Preparation

Data preparation is a set of procedures for organizing datasets for machine learning. Data preparation, in a deeper context, comprises identifying the optimal data gathering method. And these methods consume the vast bulk of machine learning time.

(i) Importing the library and Dataset: This is the process of loading and reading data into python from various resources.

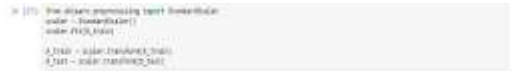
(ii) Preprocessing: Preprocessing is critical while preparing the data for the ML model after it has been collected. Outliers, inconsistency, missing values, erroneous, skewed, and trends are examples of problems that can be solved. Preprocessing the data is important since the model learns on that data alone. If we give the model inconsistent, appropriate data, the model will only return garbage, so it is necessary to ensure that the data is free of any unknown issues. In data preprocessing, we typically remove null values columns, fill missing values, make the data set consistent, and remove outliers and skewed data.



(iii) Train Test Split: To prevent overfitting, we divided our dataset into training and test groups, to enable us have a clearer perspective of how our algorithm performed throughout the testing phase. Our method is put to the test on unseen data, just like it would be in a real-world scenario. 10% is set aside for testing to ensure that the model is correctly trained and fully understands the data. Run the following script to generate training and test splits:

(iv) Feature Selection: To improve the performance of the model, we have to find relevant characteristics from a batch of sample datasets and delete the unnecessary and far less essential data without changing the target variables. When developing a machine learning model, feature selection has a significant impact on the model's performance and efficiency. This stage assists us in identifying the most important input data for the task. If we have unnecessary features, we will have overfitting and underfitting in the model. The following are some of the advantages of feature selection:

a) Reduces overfitting and underfitting
b) Improves precision
c) Saves time
d) Enhances efficiency



(v) Training and Prediction: Training with Random Forest

System Evaluation

We can calculate four distinct metrics to measure the correctness of our model using confusion matrix, Accuracy, precision, recall and f1 scores are the four measures. True

Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) scores are used to generate each measure.

Note: 11,000 dataset; Training 9,900 (90%), while testing 1,100 (10%).

a) TP (True Positive): The number of phishing websites detected.
b) FN (False Negative): The number of legitimate websites detected.
c) TN (True Negative): The number of correct Legitimate websites being classified.
d) FP (False Positive): The number of incorrect Phishing websites which are classified

In order to evaluate the performance of the model, the parameters above are used to compute the confusion matrix and this is shown in table 1.

Table 1: Confusion Matrix Evaluation

TP	TN	FP	FN
1050	10	35	05

a) **Computation of Accuracy:** Accuracy is the number of accurate classifications made from all examples in the testing dataset. Eq. 3.10 is the formula for accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.10)$$

$$\text{Accuracy} = \frac{1050+10}{1020+10+35+05}$$

$$\text{Accuracy} = \frac{1060}{1100}$$

$$\text{Accuracy} = 0.96 = 96\%$$

b) **Computation of Precision:** Precision is a metric that quantifies the accuracy of a classifier by counting the number of examples that were accurately categorised. It's the overall number of positive instances anticipated divided by the total number of positive predictions. Precision, for us, addresses the question, "How many of the URLs identified as phishing are indeed phishing?" Eq. 3.11 is the formula for calculating precision.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3.11)$$

$$\text{Precision} = \frac{1050}{1050+35}$$

$$\text{Precision} = \frac{1050}{1085}$$

$$\text{Precision} = 0.967 = 97\%$$

c) **Computation of Recall:** The number of true positives correctly predicted by the classifier from a set of all positive occurrences is measured by recall. To put it in

another way, recall is the number of instances that were overlooked. Recall is a measure of the classifier's completeness. "Of all the URLs that are genuinely phishing, how many did we identify as phishing?" is a question that recall answers for us.Eq. 3.12 is the formula for calculating recall.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3.12)$$

$$\text{Recall} = \frac{1050}{1050+05}$$

$$\text{Recall} = \frac{1050}{1055}$$

$$\text{Recall} = 0.99 = 99\%$$

d) **Computation of F1-score:** The F1-score is the weighted average of precision and recall. Eq. 3.13 is the formula for calculating F1-score.

$$\text{F1-Scores} = \frac{2\ x\ Precision\ x\ Recall}{Recall+Precision} \quad (3.13)$$

$$\text{F1-Scores} = \frac{2\ x\ 97\ x\ 99}{99+97}$$

$$\text{F1-Scores} = \frac{19{,}206}{196}$$

$$\text{F1-Scores} = 97.98\%$$

e) **Computation of Error Rate:** The error rate (ERR) is calculated by dividing the total number of wrong predictions by the total number of predictions in the dataset, equivalent to 1 minus accuracy. Eq. 3.14 is the formula for calculating error rate.

$$\text{Error rate} = 1-\frac{TP+TN}{TP+TN+FP+FN} \quad (3.14)$$

$$\text{Error rate} = 1-\frac{1050+10}{1050+10+35+05}$$

$$\text{Error rate} = 1-\frac{1060}{1100} = 1-0.96$$

$$\text{Error rate} = 0.04 = 4\%$$

According to the confusion matrix, our Random Forest algorithm classified all 1,100 records in the test set with 96% accuracy and a 4% error rate, which is excellent.

4.0. RESULTS AND DISCUSSION

Table 3 and figure 4 shows the level of accuracy of some of the existing systems alongside our proposed model. Our model attained the highest overall prediction accuracy of 96% which makes it most suitable for phishing website detection and prevention of internet fraud.

Table 3: Accuracy rate of existing models and the proposed model

Researchers	Accuracy(%)
Zhang et al. (2011)	94
Kulkarni et al. (2019)	90.39
Sahingoz et al. (2018)	94
Marjan et al. (2016)	91
Nguyen et al. (2015)	94
Ratnaparkhi et al. (2020)	94
Sonowal et al. (2020)	92.72
Proposed Phishing Detection Model	96

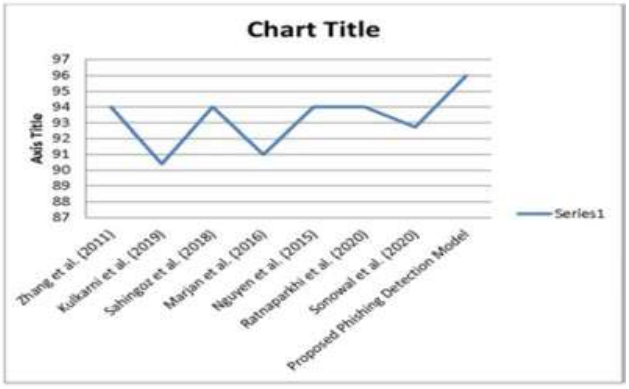


Figure 4: Line graph showing accuracy rate for detection models

Figure 5: Developed plugin interface for showing results

This interface is displayed whenever users are on a webpage. The plugin must be integrated in a web browser in order to receive notifications. Whenever it detects a phishing website it goes ahead to tell the user the level of severity of the phishing website.

5.0 CONCLUSION

In this research paper, we have developed a phishing website detection plugin for web browsers. Our proposed model when compared to existing literatures attained a high detection rate of 96% and an error rate of 4%. From our results we can say that our objectives were achieved and this model is capable of detecting phishing websites and forestalling online theft. Therefore; based on the attained prediction performance this research concludes random forest approach is a suitable algorithm for building phishing website detection models.

References

[1] A. A. Aminu, A. Abdulkarim, A. Y. Aliyu, M. Aliyu, and A. M. Turaki, "Detection of Phishing WebsitesUsing Random Forest and XGBOOST Algorithms," *Front. Knowl. J. Ser. | Int. J. Pure Appl. Sci.*, vol. 2, no. 3, pp. 2635–3393, 2019, [Online]. Available: www.smrpi.com.

[2] A. M. John-otumu, M. Rahman, and C. U. Oko, "An Efficient Phishing Website Detection Plugin Service for Existing Web Browsers Using Random Forest Classifier," *Am. J. Artif. Intell.*, vol. 5, no. 2, pp. 66–75, 2021, doi: 10.11648/j.ajai.20210502.13.

[3] A. V. Vidyapeetham, M. City, P. Poornachandran, and A. V. Vidyapeetham, "DeepAnti-PhishNet : Applying Deep Neural Networks for Phishing Email Detection," 2018.

[4] A. Abiodun, "ASHIRU: Identifying Phishing As A form of Cybercrime in Nigeria

IDENTIFYING PHISHING AS A FORM OF CYBERCRIME IN NIGERIA*," *Nnamdi Azikiwe Univ. J. Int. Law Jurisprud.*, vol. 12, no. 2, pp. 176–186, 2021.

[5] M. F. Oluwalami, "Fraud detection in banking institutions," *Int. J. Eng. Technol.*, vol. 8, no. 2, pp. 1127–1130, 2016.

[6] A. A. Ubing, S. J. B. Kamilia, A. Abdullah, N. Z. Jhanjhi, and M. Supramaniam, "Phishing Website Detection : An Improved Accuracy through Feature Selection and Ensemble Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 1, pp. 252–257, 2019, doi: 10.14569/IJACSA.2019.0100133.

[7] P. N. Mangut and K. A. Datukun, "The Current Phishing Techniques – Perspective of the Nigerian Environment," *World J. Innov. Res.*, vol. 10, no. 1, pp. 34–44, 2021.

[8] B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, "Defending against phishing attacks: taxonomy of methods, current issues and future directions," *Telecommun. Syst.*, vol. 67, no. 2, pp. 247–267, 2018, doi: 10.1007/s11235-017-0334-z.

[9] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," *Expert Syst. Appl.*, vol. 117, no. October, pp. 345–357, 2019, doi: 10.1016/j.eswa.2018.09.029.

[10] R. Alabdan, "Phishing Attacks Survey : Types , Vectors , and Technical Approaches," *Futur. Internet*, vol. 12, no. 168, pp. 1–29, 2020.

[11] J. Kolouch, "Evolution of Phishing and Business Email Compromise Campaigns in the Czech Republic," *AARMS*, vol. 17, no. 3, pp. 83–100, 2018.

[12] M. L. Jensen, M. Dinger, and R. T. Wright, "Training to Mitigate Phishing Attacks Using Mindfulness Techniques," *J. ofManagement Inf. Syst.*, vol. 34, no. 2, pp. 597–626, 2017.

[13] T. C. A. Ebot, *Explaining Two Forms of Internet Crime from Two Perspectives Toward Stage Theories for Phishing and Internet Scamming Alain Claude Tambe Ebot Explaining Two Forms of Internet Crime from Two Perspectives Toward Stage Theories for*. 2017.

[14] A. Jain and V. Richarya, "Implementing a Web Browser with Phishing Detection Techniques," *World Comput. Sci. Inf. Technol. J.*, vol. 1, no. 7, pp. 289–291, 201

Stock Market Predictions Using Artificial Neural Network Based Forecasting Model Analysis

Egbe T.P[1], Njoku D.O.[2], Oparah C. C.[1], Akandu L. N[1], Omenka U.E.[1]
[1]*Department of Computer Science, Federal Polytechnic, Nekede, Owerri, Imo State-Nigeria*
[2]*Department of Computer Science, Federal University of Technology, Owerri, Imo State-Nigeria*

Abstract –This paper has presented a Neural Networks Multi-Layer Perception (MLP) architecture based on Back Propagation Algorithm to predict stock market performance. Stock market fluctuations are the result of complex phenomena, whose effect translates into a blend of gains and losses that appear in a stock market time series that is usually predicted by extrapolation. Twenty three years representative stock prices of three different stock sales were analyzed using a feed forward neural network with back-propagation algorithm and k-means clustering algorithm. The designed system was implemented using Object Oriented Methodology and Java programming language which is full-featured and capable of developing robust applications. The developed system is found to predict the monthly stock exchange sales more accurately when compared with other methods, so that dealers and stock brokers can make proactive and successful preparation.

Keyword: Back Propagation Algorithm, Neural Networks, Predict, stock market, Multi-Layer Perception, k-Means

1. INTRODUCTION

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. Stock Market prediction is one of the hottest fields of research lately due to its commercial applications owing to high stakes and the kinds of attractive benefits that it has to offer. Forecasting the price movements in stock markets has been a major challenge for common investors, businesses, brokers and speculators. As more and more money is being invested the investors get anxious of the future trends of the stock prices in the market. The primary area of concern is to determine the appropriate time to buy, hold or sell. In their quest to forecast, the investors assume that the future trends in the stock market are based at least in part on present and past events and data. However financial time-series is one of the most 'noisiest' and 'non-stationary' signals present and hence very difficult to forecast. Stock market fluctuations are the result of complex phenomena, whose effect translates into a blend of gains and losses that appear in a stock market time series. The gain and losses can be predicted by extrapolation. The periodic variations follow either seasonal patterns of the business cycle in the economy or short-term and day-to-day variations appear at random and are difficult to predict, but they are often the source for stock trading gains and losses, especially in the case of day traders. Numerous investigations gave rise to different decision support systems for the sake of providing the investors with an optional prediction. Many experts in the stock markets have employed the technical analysis for better prediction for a long time.

Generally speaking, the technical analysis derives the stock movement from the stock´s own historical value. The historical data can be used directly to form the support level and the resistance or they can be plugged into many technical indicators for further investigation. Conventional researches addressing this research problem have generally employed the time series analysis techniques (i.e. mixed auto regression moving average (ARMA)) as well as multiple regression models [1]. The ARIMA models have one severe drawback: they assume that the volatility of the variable being modeled (e. g. stock price) is constant over time. In many cases this is not true. Large differences (of either sign) tend to be followed by large differences.

Prediction is made by exploiting implications hidden in past trading activities and by analyzing patterns and trends shown in price and volume charts. The Neural Network based forecasting seems to be better predictor than the other approaches to predict stock market [2]. In this work, we showed a method to forecast the daily stock price using neural networks

`

and the result proved to be better when compared with results of other forecasting methods

2. OVERVIEW OF FORECASTING AND PREDICTION

Forecasting is the process of making statements about what will happen in the future based on information that you have at hand. A commonplace example is a meteorologist using maps and scientific data to tell us about the possibility of rain fall, snow fall or sunshine. While Prediction is making a statement about what you think will happen in future, often but not always based on experience or knowledge .Prediction is a guess which is based on instinct eg. A fortune teller makes a prediction using a crystal ball. Both prediction and forecasting might refer to formal statistical methods employing time series, cross-sectional or longitudinal data, or alternatively to less formal judgmental methods. Usage can differ between areas of application. For example, in hydrology, the terms "forecast" and "forecasting" are sometimes reserved for estimates of values at certain specific future times, while the term "prediction" is used for more general estimates, such as the number of times floods will occur over a long period. Risk and uncertainty are central to forecasting and prediction; it is generally considered good practice to indicate the degree of uncertainty attaching to forecasts. In any case, the data must be up to date in order for the forecast to be as accurate as possible.
Cumulative Prospect Theory (CPT): This is concerned with behavior of decision makers who face a choice between two alternative actions which are associated with particular probabilities or gambles. It is a further development and variant of prospect theory. The difference from the original version of prospect theory is that weighting is applied to the cumulative probability distribution function, as in rank-dependent expected utility theory. *(Which states that the belief that an item or service's utility is a measure of the satisfaction that the consumer will derive from the consumption of that particular good or service)*? Rather than to the probabilities of individual outcomes.

In 2002, "Daniel Kahneman received the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel for his contributions to behavioral economics, in particular the development of Cumulative Prospect Theory (CPT)". A typical value functions in Prospect Theory and Cumulative Prospect Theory. It assigns values to possible outcomes of a lottery.
A typical weighting function in Cumulative Prospect Theory. It transforms objective cumulative probabilities into subjective cumulative probabilities.
The main observation of CPT (and its predecessor Prospect Theory) is that people tend to think of possible outcomes usually relative to a certain reference point (often the status quo) rather than to the final status, a phenomenon which is called framing effect. Moreover, they have different risk attitudes towards gains (i.e. outcomes above the reference point) and losses (i.e. outcomes below the reference point) and care generally more about potential losses than potential gains (loss aversion). Finally, people tend to overweight extreme, but unlikely events, underweight "average" events. The last points in contrast to Prospect Theory which assumes that people overweight unlikely events, independent of their relative outcomes.
CPT incorporates these observations in a modification of Expected Utility Theory by replacing final wealth with payoffs relative to the reference point, replacing the utility function with a value function that depends on relative payoff, and replacing cumulative probabilities with weighted cumulative probabilities. In the general case, this

$$U(p) := \int_{-\infty}^{0} v(x)\frac{d}{dx}(w(F(x)))\,dx + \int_{0}^{+\infty} v(x)\frac{d}{dx}(-w(1-F(x)))\,dx,$$

leads to the following formula for subjective utility of a risky outcome described by probability measure

$$F(x) := \int_{-\infty}^{x} dp \qquad (1)$$

where; v is the value function, w is the weighting function as shown in Eq. (1).

The claim that prediction markets can efficiently aggregate information is based on the Efficient Market Hypothesis. In certain cases, existing theoretical results regarding efficient capital markets can be applied directly.

2.1 Forecasting Methods

Though there are various methods of forecasting, this work is going to look at two major types of forecasting models namely: Conventional and Intelligent methods.

a) Conventional Methods of Stock Market Forecasting Model Construction

These types of forecasting methods are based on mathematical (quantitative) models, and are objective in nature. They rely heavily on mathematical computations. Conventional forecasting models are used to forecast future data as a function of past data; they are appropriate when past data are available. These methods are usually applied to short- or intermediate-range decisions.

`

Examples include Trend Analysis, Trend Projection method, Regression Base Approach, Time Series Analysis and others

b) Intelligent Methods

Modern forecasting techniques, such as expert systems, Artificial Neural Networks ANN), fuzzy logic, Markov Model and Delphi method etc, have been developed recently, showing encouraging results. Among them, ANN methods are particularly attractive, as they have the ability to handle the nonlinear stock exchange market and the factors affecting it directly from historical data. Banks and Financial Institutions are investing heavily in development of neural network models and have started to deploy it in the financial trading arena.

2.2 Review of Related Works

Fazel et al [3] used a type-2 fuzzy rule based expert system to develop stock price analysis. The proposed type-2 fuzzy model applies the technical and fundamental indexes as the input variables. "The model used for stock price prediction of an automotive manufactory in Asia. The output membership values were projected onto the input spaces to generate the next membership values of input variables and tuned by genetic algorithm". The type-1 method was used for inference and to increase the robustness of the system. This method was used to robustness, flexibility and error. Preethi, and Santhi [4] surveyed recent literature in the area of Neural Network, Data Mining, Hidden Markov Model and Neuro-Fuzzy system used to predict the stock market fluctuation. Neural Networks and Neuro-Fuzzy systems are identified to be the leading machine learning techniques in stock market index prediction area. Zhang et al. "[5] used Back-propagation Neural Networks (BPN for Sales Forecasting Based on ERP System. They found out that BPN can be used as an ac-curate sales forecasting system. Efendigil and Kahraman [6] utilized a forecasting system based on artificial neural networks ANNs and adaptive network based fuzzy inference systems (ANFIS) to predict the fuzzy demand with incomplete information". Sheta [7] developed fuzzy models for stock markets. He used the model for two non-linear processes, one pertaining to NASA and the other to prediction of next week S&P 500 index levels. The two steps involved in the process are (1) the determination of the membership functions in the rule antecedents using the model input data; (2) the estimation of the consequence parameters. Parameters are estimated using least square estimation. Tan, Quek & Ng [8] introduced a novel technique known as Genetic Complementary Learning (GCL) to stock market prediction and give comparisons to demonstrate the superior performance of the method. GCL algorithm is a confluence of GA and hippocampal complementary learning. Kyoungjae-Kim [9] introduced Genetic algorithm approach to instance selection (GAIS) for ANN in financial data mining has been reported. Kim introduces this technique to select effective training instances out a large training data set to ensure efficient and fast training for stock market prediction networks. The GA also evolves the weights that mitigate the well known limitations of the gradient descent algorithm. The study demonstrates enhances prediction performance at reduced training time.

A hybrid model proposed by Kuo, Chen & Hwang [11] integrates Genetic Algorithm (GA) based fuzzy logic and ANN. The model involves both quantitative factors (technical parameters) and qualitative factors such as political and psychological factors. Evaluation results indicate that the neural network considering both the quantitative and qualitative factors excels the neural network considering only the quantitative factors both in the clarity of buying-selling points and buying and selling performance. Another hybrid model involving Genetic Algorithm (GA) proposed [5] utilizes the strengths of Hidden Markov Models (HMM), ANN and Genetic Algorithm (GA) to forecast financial market behavior. Using ANN, the daily stock prices are transformed to independent sets of values that become input to HMM. The job of the GA is to optimize the initial parameters of HMM. The trained HMM is then used to identify and locate similar patterns in the historical data. A similar study investigates the effectiveness of a hybrid approach based on Time Delay Neural Networks (TDNN) and GA. The Genetic Algorithm (GA) is used to optimize the number of time delays in the neural network to obtain the optimum prediction performance.

Many researchers conclude that the application of Back-propagation Neural Networks (BPN) is an effective method as a forecasting system, and can also be used to find the key factors for enterprisers to improve their logistics management level. In [4] utilized Back Propagation neural networks (BPN) in order to forecast safety stock.

3. SYSTEM ANALYSIS AND DESIGN

3.1 Analysis of the Existing System

Neenwi et al [12] have proved by contradiction that the Nigerian stock market is not efficient but chaotic. Two years representative stock prices of some banks stocks were analyzed using a feed forward neural network with back-propagation in Matlab 7.0. The simulation results and price forecasts show that it is possible to consistently earn good returns on investment on the Nigerian stock

`

market using private information from an artificial neural network indicator.
The disadvantage of the existing system is as follows

1. The convergence obtained from back propagation is very slow and not guaranteed.
2. Time to train NN is probably identified as biggest disadvantage.
3. They used very small sample sets to train, thus model is not very efficient.
4. Back propagation learning requires input scaling and normalization.
5. Back propagation does not guarantee to find the global network minimum. While it does minimize error, there is a chance the weights will be changed to fit a local minimum in the error landscape, but the network will not be optimized.

The architecture of the present system is shown in Fig. 1. The training process requires a set of examples of proper network behavior - network inputs (close prices) and target outputs. During training the weights and biases of the network are iteratively adjusted to minimize the network performance function. The most common performance function for feed forward networks is Mean Square Error MSE - the average squared error between the networks outputs and the target outputs.

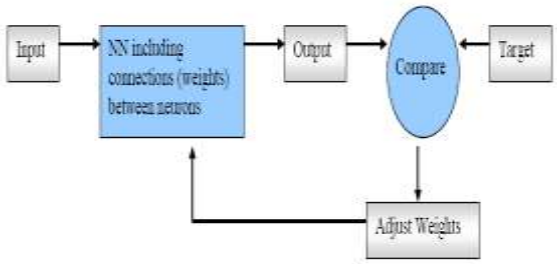


Fig. 1 Architecture of the present system [12]

3.2 Analysis of the Proposed System

We proposed hybrid model to predict stock market performance for three companies namely *(Guinness Plc, Nestle Plc and Total Plc)* in Nigeria using neural network and k-means clustering algorithm. The developed system will accurately predict the monthly stock exchange sales, so that the companies can make proactive and successful preparation. Twenty two years representative stock prices for each of the company's stock were analyzed using a feed forward neural network with back-propagation algorithm and k-means clustering algorithm. The outcome of this thesis will led to policy formulations aimed at an improved and developed market for potential gain to the benefit of rational investors even across national borders. It will allow the investor to make decisions based on probabilities of success.
The advantages of the proposed system are as follows

1. It improves the training rules of backward propagation algorithm, thus arrive at a better prediction with greater accuracy.
2. The proposed system will guarantee and obtain global network minimum using clustering algorithm.
3. It uses very large sample sets to train model efficiently.
4. It uses an Artificial Neural Network, which has the ability to use an arbitrary functional approximation mechanism in learning from observed data.
5. It will reduce time complexity of the train of the hidden layer of backward propagation.

3.3 Methodology of the Proposed System

A system development methodology (SMD) refers to the framework that is used to structure, plan, and control the process of developing an information system. A wide variety of such framework has evolved over the years, each with its own recognized strengths and weaknesses. We adopted the object Oriented Analysis and Design Methodology in the analysis of Neural Networks Multi Layer Perception (MLP) architecture based on Back Propagation Algorithm to predict the stock market performance.

1. Object oriented analysis (OOA): This is the process of defining the problem in terms of object: real world with which the system must interact, and candidate software objects used to explore various solution alternative. The nature fit of programming objects to real world objects has a big impact here in the all real world objects can defined in terms of their classes, attribute and operations.
2. Object oriented design (OOD): this is the process of defining the component, interfaces, objects, classes, attributes, and operations that will satisfy the requirement. You typically start with the candidate object defined during analysis, but add much more rigor to their definitions, then you add or change objects as needed to refine a solution. The basic step of system design is shown in Fig. 2. This methodology features the unified modeling language (UMLs): Case diagram, active diagram, the architectural design of the system and proposed algorithm.

`

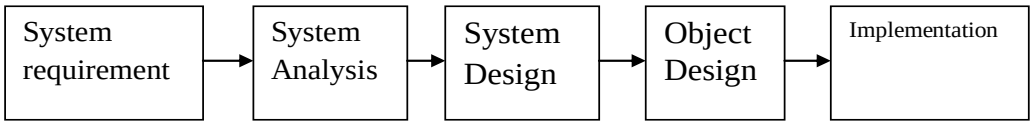


Fig. 2 Methodology of the Proposed System

a) Use Case Diagram

A Use case represents the functionality of the system as shown in Fig. 3. In this system, the user can directly communicate with the system and get the appropriate advices suggested for decision making by the intelligent system. The proposed system allows the user to provide the training data and the threshold constant. Training the Neural network requires initialization of input weights and the first input weight will be adjusted until the optimal prediction is got at last.

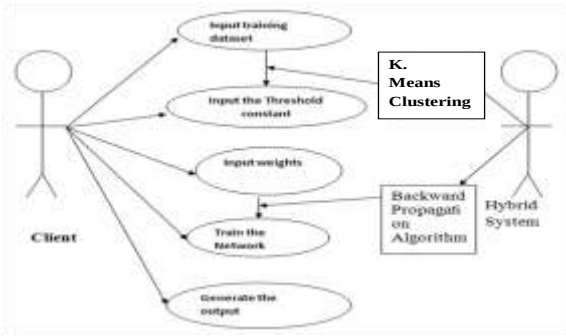
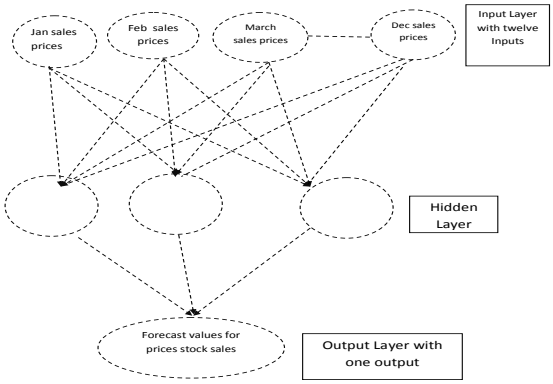


Fig. 3: Use Case Diagram of the Proposed System

c) The Neural Network Architectural Design of the Proposed System

The architecture of this model in Fig. 5 consists of twelve input variable, four intermediate variables and one output variable. It is a schematic diagram of a 9-3-1 topology. The twelve input variables are: $X1_1, X1_2, X1_3$, represents twelve months sales for each of the three companies. The hidden layer of $X2_1$, $X2_2$, and $X2_3$ are intermediate variables which interact by means of weight matrices with adjustable weights to produce the output. It uses the Activation function to manipulate the inputs from the input layer in other to make the optimal prediction. The output layer contains one output variable which is the optimal prediction.



b) High level Design of the Proposed System

The user is expected to enter the input dataset with the required output (target) value in Fig. 4. The Neural Network initializes the input weight using a random number and also the hidden layer weight with the same random number and threshold constant. It does the neural network calculation using the activation function for the input layer and the output layer, and then it calculates its error by finding the difference between the output and the target. If the error is large enough then the neural network adjust this weight and back-propagates, but if the error is small, it outputs the prediction.

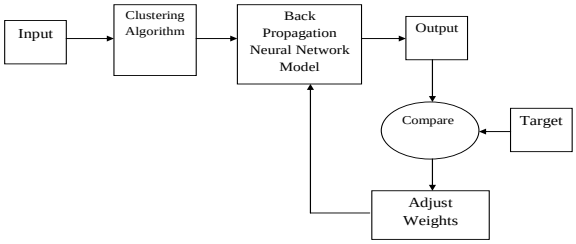
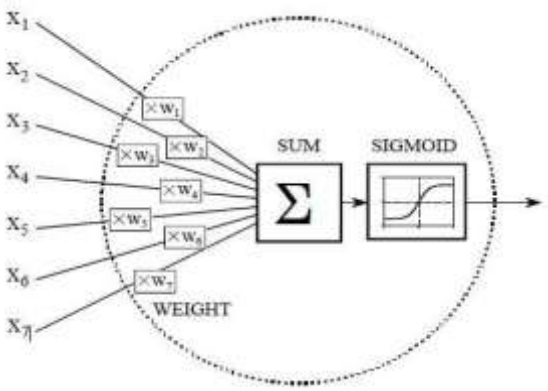


Fig. 4 High level Design of the Proposed System

Fig. 5 Neural network Architecture of the proposed system

d) Activations Function

Activations function is needed for hidden layer of the NN to introduce nonlinearity. Without them NN would be same as plain perceptions. If linear function were used, NN would not be as powerful as they are. Activation function can be linear, threshold or sigmoid function. Sigmoid activation function is usually used for hidden layer because it combines nearly linear behavior, curvilinear behavior and nearly constant behavior depending on the input value. To explain activation function Fig. 6 will be used.



`

Fig. 6 Activations Function

SUM is collection of the output nodes from hidden layer that have been multiplied by connection weights, added to get single number and put through sigmoid function (activation function). Input to sigmoid is any value between negative infinity and positive infinity number while the output can only be a number between 0 and 1.

e) Activity Diagram of the Proposed System

From the activity flow diagram in Fig.7, the client is expected to provide the input dataset and the required output. The required output is used in back propagation. The system uses it to compare its predicted value from time to time in other to get the optimal prediction. The client selects a threshold constant and looking at the input value, the neural network initializes the weight for the input layer and the hidden layer (the hidden layer uses the activation function to manipulate the inputs from the input layer in other to make optimal prediction). Then the calculation for the activation function of both the input and the output layer is done and the system calculates the error. If the error is large then the system adjust the weight and do a back propagation to the input layer for further adjustment and calculation to get an optimal prediction. The system outputs its prediction whenever the error is small. An activity diagram is essentially a flowchart, showing flow of controls from one activity to another. Unlike a traditional flowchart, it can model the dynamic functional view of a system. A flow diagram represents an operation on some classes in the system that results to changes in the state of the system.

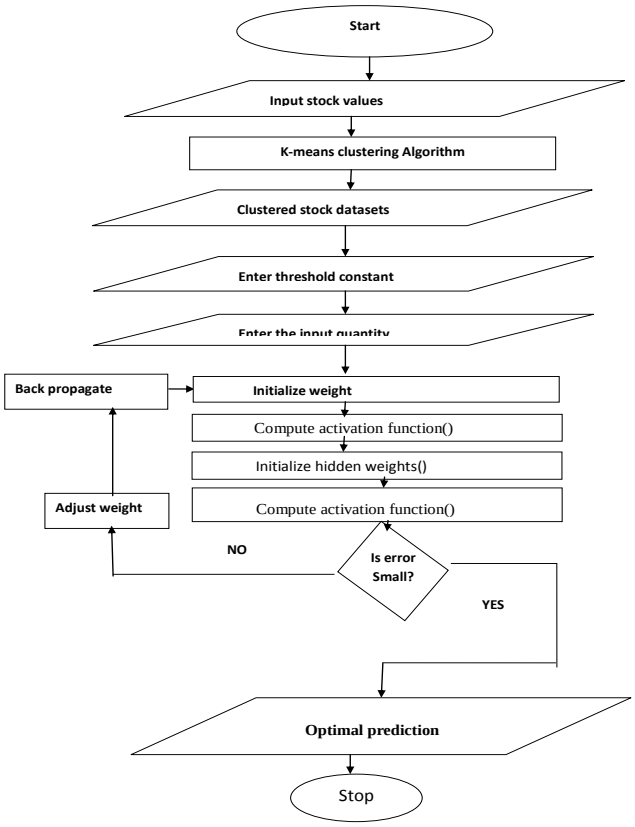


Fig. 7 Activity Diagram of the Proposed System

Proposed System Algorithm

The connection we are interested in is between neuron A (a hidden layer neuron) and neuron B (an output neuron) and has the weight WAB.

1. First apply the inputs to the network and work out the output – remember this initial output could be anything, as the initial weights were random numbers.
2. Next work out the error for neuron B. The error is *What you want – What you actually get*, in other words: $Error_B = Output_B (1-Output_B)(Target_B – Output_B)$

The "*Output (1-Output)*" term is necessary in the equation because of the Sigmoid

Function – if we were only using a threshold neuron it would just be *(Target – Output)*.

3. Change the weight. Let W^+_{AB} be the new (trained) weight and W_{AB} be the initial weight.

$W^+_{AB} = W_{AB} + (Error_B \times Output_A)$

Notice that it is the output of the connecting neuron (neuron A) we use (not B). We update all the weights in the output layer in this way.

`

4. Calculate the Errors for the hidden layer neurons. Unlike the output layer we can't calculate these directly (because we don't have a Target), so we *Back Propagate* them from the output layer (hence the name of the algorithm). This is done by taking the Errors from the output neurons and running them back through the weights to get the hidden layer errors. For example if neuron A is connected as shown to B and C then we take the errors from B and C to generate an error for A.
$Error_A = Output_A (1 - Output_A)(Error_B W_{AB} + Error_C W_{AC})$ Again, the factor "*Output (1 - Output)*" is present because of the sigmoid squashing function.
5. Having obtained the Error for the hidden layer neurons now proceed as in stage 3 to change the hidden layer weights. By repeating this method we can train a network of any number of layers. This may well have left some doubt in your mind about the operation, so let's clear that up by explicitly showing *all* the calculations for a full sized network with 2 inputs, 3 hidden layer neurons and 2 output neurons as shown in Fig. 8. W+ represents the new, recalculated, weight, whereas W (without the superscript) represents the old weight.

$W^+_{stock\_1A} = W_{stock\_1A} + \eta*\delta_A*in_{stock\_1A}$ $W^+_{stock\_2A} = W_{stock\_2A} + \eta*\delta_A*in_{stock\_2}$
$W^+_{stock\_1B} = W_{stock\_1B} + \eta*\delta_B*in_{stock\_1B}$ $W^+_{stock\_2B} = W_{stock\_2B} + \eta*\delta_B*in_{stock\_2}$
$W^+_{stock\_1C} = W_{stock\_1C} + \eta*\delta_B*in_{stock\_1C}$ $W^+_{stock\_2C} = W_{stock\_2C} + \eta*\delta_C*in_{stock\_2}$
The constant η (called the learning rate, and nominally equal to one) is put in to speed up or slow down the learning if required.

Table 1: Illustrations of Back Propagation

		Training Weights		
inputs		W1	W2	W3
A	0.35	0.1	0.8	0.3
B	0.9	0.4	0.6	0.9

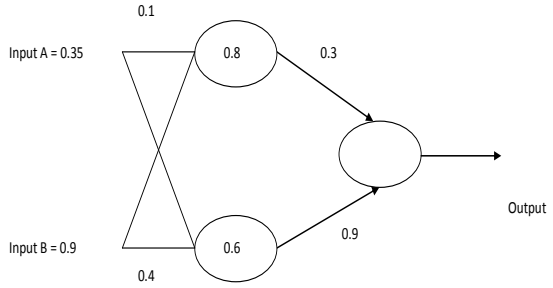


Fig. 9: Sample diagram showing back propagation
Assume that the neurons have a Sigmoid activation function and
(i) Perform a forward pass on the network.
(ii) Perform a reverse pass (training) once (target = 0.5).
(iii) Perform a further forward pass and comment on the result.

Neuron Inputs

(i) Input to top neuron = (0.35x0.1)+(0.9x0.8)=0.755. Out = 0.68.

Input to bottom neuron = (0.9x0.6)+(0.35x0.4) = 0.68. Out = 0.6637.
Input to final neuron = (0.3x0.68)+(0.9x0.6637) = 0.80133. Out = 0.69.

Input to top neuron = $(stock\_1* W_{stock\_2\ A})+(stock\_2* W_{stock\ 1\ B})$=neuron_1.
Input to bottom neuron = $(stock\_2* W_{stock\ 1\ b})+(stock\_2* W_{stock\ 1\ B})$=neuron_2.
Out = sigmoid function formula f(x) =1/ (1.0 + exp (-$Neuron_s$).

1. Calculate errors of output neurons

$\delta_{Stock\ 1\ Predict\_1} = out_{Stock\ 1\ Predict\_1} (1 - out_{Stock\ 1\ Predict\_1}) (Target_{Stock\ 1\ Predict\_1} - out_{Stock\ 1\ Predict\_1})$
$\delta_{Stock\ 2\ Predict\_2} = out_{Stock\ 2\ Predict\_2} (1 - out_{Stock\ 2\ Predict\_2}) (Target_{Stock\ 2\ Predict\_2} - out_{Stock\ 2\ Predict\_2})$

2. **Change output layer weights**

$W^+_{A\ Stock\ 1\ Predict\_1} = W_{A\ Stock\ 1\ Predict\_1} + \eta*\delta_{Stock\ 1\ Predict\_}\mathbf{1}*outA$
$W^+_{A\ Stock\ 2\ Predict\_2} = W_{A\ Stock\ 2\ Predict\_2} + \eta*\delta_{Stock\ 2\ Predict\_2}*out_A$
$W^+_{B\ Stock\ 1\ Predict\_1} = W_{B\ Stock\ 1\ Predict\_1} + \eta*\delta_{Stock\ 1\ Predict\_}\mathbf{1}*out_B$
$W^+_{B\ Stock\ 2\ Predict\_2} = W_{B\ Stock\ 2\ Predict\_2} + \eta*\delta_{Stock\ 2\ Predict\_2}*out_B$
$W^+_{C\ Stock\ 1\ Predict\_1} = W_{C\ Stock\ 1\ Predict\_1} + \eta*\delta_{Stock\ 1\ Predict\_}\mathbf{1}*out_C$
$W^+_{C\ Stock\ 2\ Predict\_2} = W_{C\ Stock\ 2\ Predict\_2} + \eta*\delta_{Stock\ 2\ Predict\_2}*out_C$

3. Calculate (back-propagate) hidden layer errors

$\delta_A = out_A (1 - out_A) (\delta_{Stock\ 1\ Predict\_1}*W_{A\ Stock\ 1\ Predict\_1} + \delta_{Stock\ 2\ Predict\_2}*W_{A\ Stock\ 2\ Predict\_2})$
$\delta_B = out_B (1 - out_B) (\delta_{Stock\ 1\ Predict\_1}*W_{A\ Stock\ 1\ Predict\_1} + \delta_{Stock\ 2\ Predict\_2}*W_{B\ Stock\ 2\ Predict\_2})$
$\delta_C = out_C (1 - out_C) (\delta_{Stock\ 1\ Predict\_1}*W_{C\ Stock\ 1\ Predict\_1} + \delta_{Stock\ 2\ Predict\_2}*W_{C\ Stock\ 2\ Predict\_2})$

4. Change hidden layer weights

`

(ii) Output error δ=(t-output)(1- output) output = (0.5-0.69)(1-0.69)0.69 = -0.0406.
New weights for output layer:
w_1^+ = w1+(δ x input) = 0.3 + (-0.0406x0.68) = 0.272392.
w2+ = w2+(δ x input) = 0.9 + (-0.0406x0.6637) = 0.87305.
Errors for hidden layers:
Once error for hidden layer nodes is known, weights between input and hidden layer can be updated. Rate of change first needs to be calculated for every weight:
δ1 = δ x w1 = -0.0406 x 0.272392 x (1- output) output = -$2.406x10^{-3}$
δ2= δ x w2 = -0.0406 x 0.87305 x (1- output) output = -$7.916x10^{-3}$
New hidden layer weights:
$w3^+$=0.1 + (-2.406 x 10^{-3} x 0.35) = 0.09916.
$w4^+$ = 0.8 + (-2.406 x 10-3 x 0.9) = 0.7978.
$w5^+$ = 0.4 + (-7.916 x 10-3 x 0.35) = 0.3972.
$w6^+$ = 0.6 + (-7.916 x 10-3 x 0.9) = 0.5928.
(iii)Old error was -0.19. New error is -0.18205. Therefore error has reduced.

f) K-means Algorithm

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroid, one for each cluster. These centroids should be placed in a cunning way because different location causes different results. So, the better choice is to place them as much as possible far away from each other.
The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point, we need to re-calculate k new centroid as barely centers of the clusters resulting from the previous step. After we have these k new centroid, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated.
As a result of this loop, we may notice that the k centroids change their locations step by step until no more changes are done. In other words, centroids do not move any more.
K-means group n objects based on features into k number of groups, each feature belongs to the cluster with the nearest mean, where k is a positive integer number. The grouping is done by minimizing the sum of squares of distance between data and corresponding cluster centroid.
The basic k-means clustering technique is described as follows
Step 1: Select k points as the initial centroid.
Step 2: Assign all points to the closest centroid.
Step 3: Re-compute the centroid of each cluster.
Step 4: Repeat steps 2 and 3 until the centroid can no longer be changed.

i. Distance-Based Clustering

A large number of algorithms, including k-means, are described as distance-based clustering because they view data as points in a metric space. The similarity or dissimilarity between two points ˆx and ˆy, both in a metric space M, is evaluated by measuring the distance between them, d (ˆx, ˆy). Given three data points ˆx, ˆy and ˆz all in M, a distance metric should satisfy the following:

1.d(ˆx, ˆy) ≥ 0
2. d(ˆx, ˆy) = 0 if and only if ˆx = ˆy
3. d(ˆx, ˆy) = d(ˆy, ˆx)
4. d(ˆx, ˆz) - d(ˆx, ˆy) + d(ˆy, ˆz)

$$d(\hat{x}, \hat{y}) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \quad (2)$$

Equation (2) is the Euclidean distance.

Given a database, D, with m attributes A1, A2, . . . ,Am of the type real value and categorical (nominal) data, i.e. mixed data, the following extended Euclidean metric can be used instead (Nguyen & Rayward-Smith, 2008):

$$d(\hat{x}, \hat{y}) = \sqrt{d_1^2(\hat{x}_1, \hat{y}_1) + d_2^2(\hat{x}_2, \hat{y}_2) + \ldots + d_n^2(\hat{x}_2, \hat{y}_{2n}) + \ldots + d_n^2(\hat{x}_m, \hat{y}_m)}, \quad (3)$$

Where ˆx = (ˆx_1, ˆx_2, ˆx_m), ˆy = (ˆy_1, ˆy_2, . . . , ˆy_m) and d_i(ˆx_i, ˆy_i) is |ˆx_i − ˆy_i|
if ˆx_i, ˆy_i Ɛ R, otherwise is the 0/1 metric. Generally, with mixed data, it is important to ensure that no attribute m dominates another with respect to the metric being used.

The Algorithm

i. Based on the necessary conditions, the k-means algorithm alternates the two steps:
ii. For a fixed set of centroids (prototypes), optimize A (·) by assigning each sample to its closest centroid using Euclidean distance.
iii. Update the centroids by computing the average of all the samples assigned to it.
iv. The algorithm converges after each iteration, as the objective function decreases (non-increasing), and converges fast.
v. Stopping criterion: the ratio between the decrease and the objective function is below a threshold.
vi. This is further illustrated in table 3.

(i) Experimental data set: {1.4, 0.9, 0.2, 3.1, 6.2, 3.6}.
(ii) Apply k-means algorithm with 2 centroids, {z_1, z_2}.

(iii)Initialization: randomly pick z_1=0.8, z_2=3.8

Table 2: First Phase-How K-means work

`

The Euclidean distance, L(Z,A) = d(ˆx, ˆy) =

$$\sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}$$

$$= \sqrt{(1.4-6.2)^2 + (0.9-3.6)^2} + (0.2-0)^2 + (3.1-0)^2$$

$$= \sqrt{(23.04) + (7.29)} + (0.04) + (9.61)$$

$$= \sqrt{39.98}$$

$$= 6.32$$

The objective function is L(Z,A) = 6.32/1.4= 4.52.
L(Z,A) = 6.32/2.5= 2.52.
The two prototypes are: $z_1 = 2.52$, $z_2 = 4.52$.
Initialization: randomly pick z1 = 0.8, z2 = 3.8.

Table 3 Second Phase- How K-means work as illustrated

Fixed	Update
0.8	{1.4, 0.9, 0.2, 0.8}
3.8	{6.2, 3.6, 3.1, 4.3}
{1.4, 0.9, 0.2, 0.8}	0.8
{6.2, 3.6, 3.1, 4.3}	4.3
0.8	{1.4, 0.9, 0.2, 0.8}
4.3	{6.2, 3.6, 3.1, 4.3}

$$\sqrt{d_1^2(\hat{x}_1, \hat{y}_1) + d_2^2(\hat{x}_2, \hat{y}_2) + \ldots}$$

L(Z,A)= $+ d_n^2(\hat{x}_2, \hat{y}_{2n}) + \ldots + d_n^2(\hat{x}_m, \hat{y}_m),$

$$= \sqrt{(1.4-6.2)^2 + (0.9-3.6)^2} + (0.2-2.6)^2$$

$$= \sqrt{(23.04 + 7.29} + 8.41 = 6.22$$ which is the Euclidean distance.

Hence, the two prototypes are: $z_1 = 1.45$ and $z_2 = 7.47$
The objective function (optimal solution) = L (Z, A) = 6.22/0.8 =7.47

(i) Starting from different initial values, the k-means algorithm converges to different local optimum.

It can be shown that $\{z_1 = 0.8, z_2 = 4.3\}$ are the global optimal solution.

ii. Dataset for Training and Testing

The daily closing price of total, Nestel and Guiness data from 1992 to 2013 is taken as training samples(approximately 1500 samples).All the inputs are normalized within a range of [0, 1] by training data = actual data/ 1000. And the result computed at 4 decimal places.

Fixed	Update
0.8	{1.4,0.9,0.2,3.1}
3.2	{6.2,3.6}
{1.4,0.9,0.2,3.1}	1.4
{6.2,3.6}	2.5
1.4	{1.4,0.9,0.2,3.1}
2.5	{6.2,3.6}

g) Design Specification

The design specifications consist of the input specification and the output specification.

i. Input Specification

The input variables includes: stock prices of Guinness Plc, Nestle Plc and Total Plc, for monthly sales for twenty two years. The input design consists of input value button. The input values buttons enable training data to be extracted from the database. The extracted data are been trained and used for prediction as shown in Fig. 10.

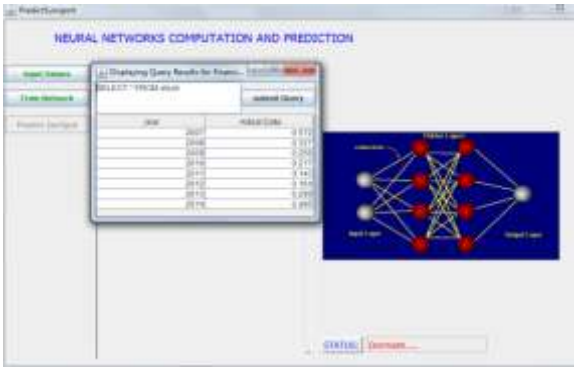


Fig. 10 Input design

ii. Output Specification

The output variables includes: Year, Actual values of stock, Predict values of stock and Closed Loop Predicted of stock. The input design consists of input value button. The train network and the predicted button used extracted date for trained and prediction as shown in Fig. 11.

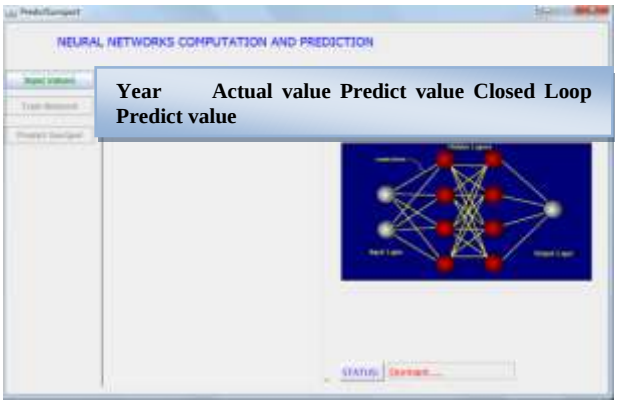


Fig.11.Output Design

`

h) Data Collection

In this work, we used data from the Nigeria stock exchange for three stock indices namely: Guinness Plc, Nestle Plc and Total Plc. The study was focused on the rate at which Customers have demanded to buy shares. The data collected were data reflecting the stock price sales from the period of twenty-two years (16/02/1997-17/04/2020). Data collection is one of the most important stages in conducting a research. The Data were collected by using the secondary method of data collection.

Normalization of Datasets

Normalization is a transformation performed on a single data input to distribute the data evenly and scale it into an acceptable range for the network. Table 3.4 shows a sample of normalized data set. We have found that input data normalization with certain criteria, prior to training process, is crucial to obtain good results, as well as to fasten significantly the calculations. The first step is to normalize the data. All the inputs were normalized within a range of [0, 1] using the following formula,

$$(X_{norm} = \frac{\text{Xorig - Xmin}}{\text{Xmax - Xmin}}) \quad (4)$$

The stock prices were preprocessed in Microsoft Excel worksheet. Where X_{norm} is the normalised value, X_{orig} is the actual currency value, X_{max} is the maximum value and X_{min} is the minimum value.

Table 4 Sample of the Normalized dataset for Nestle Plc (Source NSE 1997-2020)

S/N	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec
1	1.1210,	1.1210,	1.1210,	1.1210,	1.1210,	1.1210,	0.1640,	0.1721,	0.1825,	0.2000,	0.1790,	0.1830,
2	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,
3	0.1810,	0.1820,	0.1890,	0.1900,	0.1650,	0.1659,	0.1577,	0.1590,	0.1550,	0.1595,	0.1700,	0.1800,
4	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,
5	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1470,	0.1450,	0.1360,	0.1400,	0.1351,	0.1380,
6	0.1478,	0.1450,	0.1810,	0.1990,	0.1900,	0.1870,	0.1900,	0.1900,	0.1900,	0.1900,	0.1900,	0.1900,
7	0.1900,	0.1900,	0.1900,	0.1900,	0.1900,	0.1900,	0.2331,	0.3471,	0.3000,	0.2285,	0.2202,	0.2200,
8	0.2140,	0.1960,	0.2100,	0.2190,	0.2600,	0.2767,	0.2780,	0.780,	0.2780,	0.2780,	0.2780,	0.2780,
9	0.2780,	0.2780,	0.2780,	0.2780,	0.2780,	0.2780,	0.1482,	0.1482,	0.1482,	0.1482,	0.1482,	0.1482,
10	0.1482,	0.1482,	0.1482,	0.1482,	0.1482,	0.1482,	0.2520,	0.2520,	0.2520,	0.2520,	0.2520,	0.2520,
11	0.2520,	0.2520,	0.2520,	0.2520,	0.2520,	0.2520,	0.4102,	0.4102,	0.4102,	0.4102,	0.4102,	0.4102,
12	0.4102,	0.4102,	0.4102,	0.4102,	0.4102,	0.4102,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,
13	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.8400,	0.7756,	0.8400,	0.7756,	0.7756,	0.8400,
14	0.8400,	0.8400,	0.8400,	0.8400,	0.7756,	0.7756,	1.1210,	1.1210,	1.1210,	1.1210,	1.1210,	1.1210,
15	1.1210,	1.1210,	1.1210,	1.1210,	1.1210,	1.1210,	0.1640,	0.1721,	0.1825,	0.2000,	0.1790,	0.1815,
16	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,	0.1815,
17	0.1810,	0.1820,	0.1890,	0.1900,	0.1650,	0.1659,	0.1577,	0.1590,	0.1550,	0.1595,	0.1700,	0.1800,
18	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,	0.1911,
19	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,	0.1747,
20	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,	0.1500,
23	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,	0.2430,
22	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,	0.2300,
23	0.1810,	0.1820,	0.1890,	0.1900,	0.1650,	0.1659,	0.1470,	0.1450,	0.1360,	0.1400,	0.1351,	0.1380.

4.2 Discussion of Results

We proposed a hybrid model to predict the stock market performance using neural networks multi layer perceptron (MLP) architecture based on back propagation algorithm and k-means clustering algorithm. Back propagation is the process of back propagating errors through the network from the output layer towards the input layer during training. Back propagation is necessary because hidden units have no training target value that can be used, so they must be trained based on errors from previous layers. The output layer is the only layer which has a target value from which to compare. As the errors are back propagated through the nodes, the connection weights are changed. During the process of back propagation of errors through the network when different input patterns are presented to the network, the error is gradually reduced to a minimum. Training continues until the errors in the weights are sufficiently small to be accepted as shown in Fig. 12 to 20.

Our proposed hybrid algorithm enable the system finding the minimum error that can be attained through training, this was done by clustering 274 dataset for three stock sales of Guinness, Nestle and Total which gave a total dataset of 822. And the clustered results were used by backward propagation for training. Because, there is always a danger of getting trapped in a local minima and then being unable to find the global minimum error. The

problems of local minima were mitigated by using optimization method such as k-means clustering algorithms. The first approach to training is to run the iterations until there is no improvement in the error. The point at which the network's error has no further improvement is called convergence. The second approach is the train-test interruptions. Training is stopped after a predetermined number of iterations and the networks ability to generalize on the testing set is evaluated and training is resumed. The advantage with the convergence approach is that one can be confident that the global minimum has been reached. The drawback of the convergence method is over fitting which results from large numbers of weights. The advantage with the train-test approach is the limited degrees of freedom (weights) that the network has which is important in avoiding over fitting and gives better generalization than convergence.

i. Snap Shots of experimental results for Total, Guinness and Nestle stock prices forecasting

Different architectures were used for this experiment. The results of the experiment are presented in a graphical form and also in a table that compares the prediction accuracies of the different three stock market closed price. Overall the results show that MLP has the ability to learn the non-linear relationships that exist between the input and the output, Graphs in Fig. 12 to 14 show that MLP is able to follow the trend of the target prices.

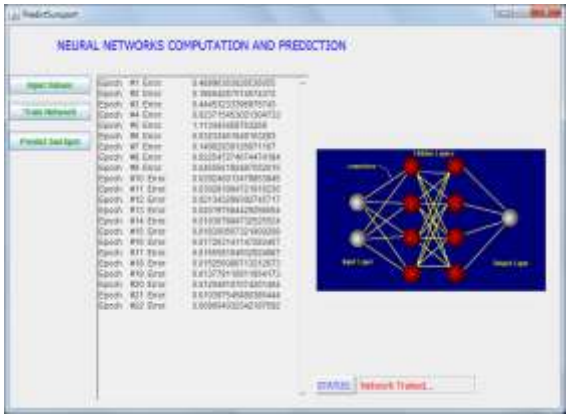


Fig.12 Shows Neural Network Training for Total Stock sales prices

For a neural network model to have a good generalization error, it has to have reached a global minimal during training and also have optimal number of weights. Form the results of the experiment the global minimal error was 0.00960.

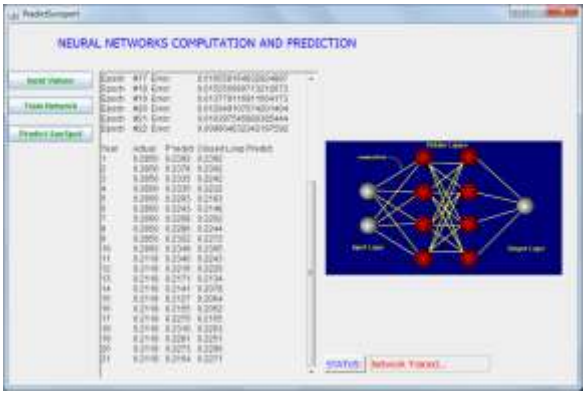


Fig. 13 Results of predicted values for Total Stock sales prices

Table 5 Actual, predicted and closed loop values for Total PLC

Month	Actual Values	Forecast Values	Closed Loop Forecast Values
1	0.285	0.2392	0.2392
2	0.285	0.2302	0.2302
3	0.285	0.2335	0.2242
4	0.285	0.2335	0.2232
5	0.285	0.2293	0.2183
6	0.285	0.2243	0.2148
7	0.285	0.2268	0.2202
8	0.285	0.2286	0.2244
9	0.285	0.2322	0.2227
10	0.285	0.2349	0.2305
11	0.2118	0.234	0.2243
12	0.2118	0.2216	0.222

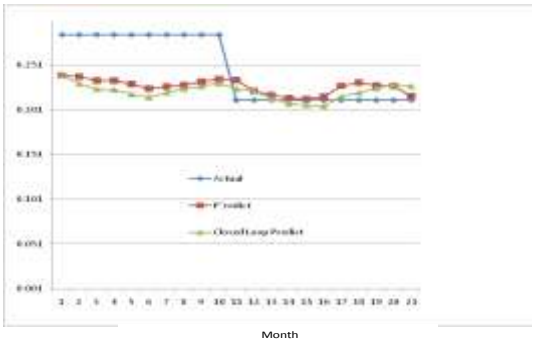


Fig. 14 Graphical representation of Actual, Predicted and Closed loop predicted values for Total stock prices sales

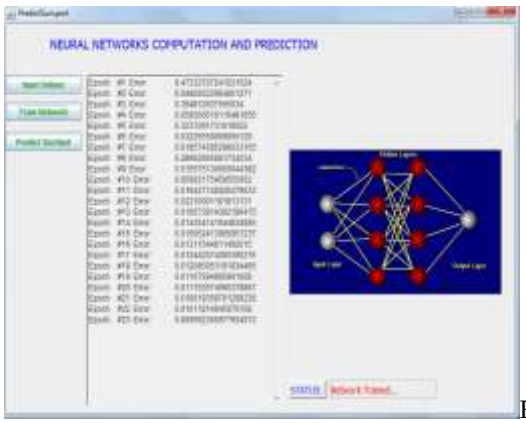


Fig. 15 Nestle analysis of network training

From the results of the experiment the global minimal error is 0.0095.

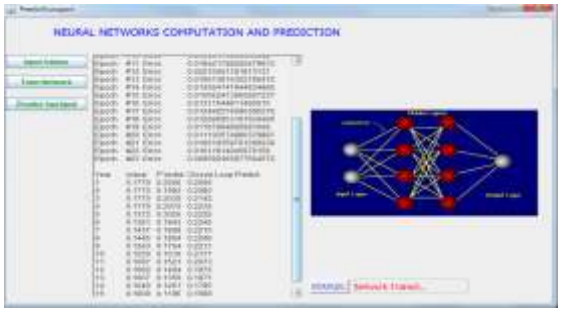


Fig. 16 Results of predicted values for Nestle Stock sales prices

Table 6 Actual, Predicted and Closed Loop Values for Nestle PLC

Month	Actual	Predict	Closed Loop Predict
1	0.177	0.2006	0.2006
2	0.177	0.1982	0.208
3	0.177	0.203	0.2142
4	0.177	0.207	0.2235
5	0.1372	0.2026	0.2225
6	0.1381	0.1843	0.2245
7	0.1437	0.1808	0.2215
8	0.1445	0.1804	0.224
9	0.1243	0.1764	0.2211
10	0.125	0.161	0.2177
11	0.1097	0.1521	0.2073
12	0.108	0.1404	0.1972

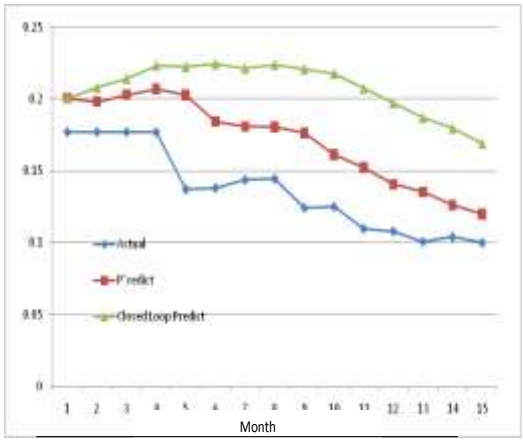


Fig.17 Graphical representation of actual, Predicted and closed loop predicted values for Nestle stock prices sales

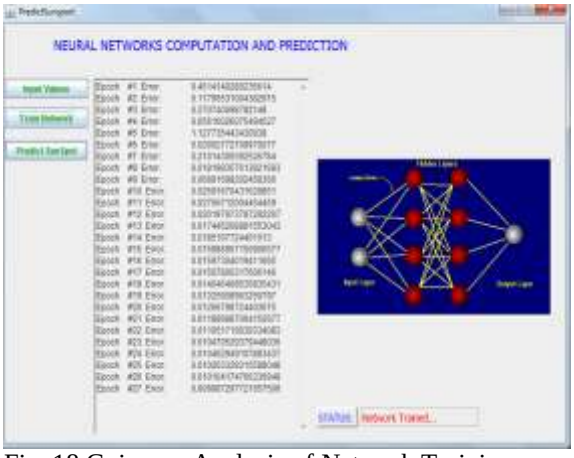


Fig. 18 Guinness Analysis of Network Training

From the results of the experiment, the global minimal error is 0.0099.

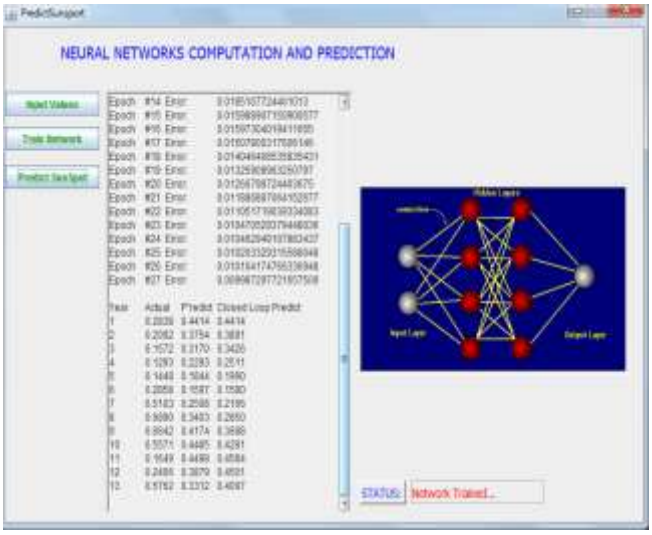


Fig. 19 Results of predicted values for Guinness Stock sales price

Table 7 Predicted values for Guinness sales prices

Month	Actual Values	Forecast Values	Closed Loop Forecast Values
1	0.2838	0.4414	0.4414

2	0.2092	0.3754	0.3891
3	0.1572	0.317	0.3426
4	0.1293	0.2293	0.2511
5	0.1448	0.1844	0.199
6	0.2858	0.1597	0.158
7	0.5103	0.2508	0.2166
8	0.689	0.3403	0.285
9	0.6642	0.4174	0.3688
10	0.5571	0.4485	0.4281
11	0.1649	0.4498	0.4584
12	0.2406	0.3879	0.4501

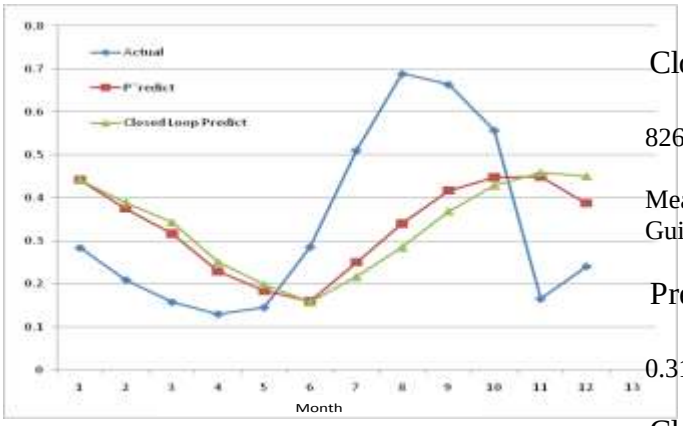


Fig. 20 Graphical representation of actual, predicted and closed loop predicted values for Guinness stock prices sales

4.3 Validation of Results for Predicted and closed loop predicted values for Nestle; Total and Guinness stock prices sales

The proposed model was validated using mean absolute percentage error (MAPE), the mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It usually expresses accuracy as a percentage, and is defined by the formula:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|,$$

where A_t is the actual value and F_t is the forecast value.

The difference between A_t and F_t is divided by the Actual value A_t again. The absolute value in this calculation is summed for every fitted or forecasted point in time and divided again by the number of fitted points n. Multiplying by 100 makes it a percentage error.

Mean absolute percentage error (MAPE) for Nestle stock prices sales:

$$\text{Predict} = \frac{1}{13}\sum_{t=1}^{13}\left|\frac{A_t - F_t}{A_t}\right| = 0.252383357$$

=25.23%

$$\text{Closed Loop Predict} = \frac{1}{13}\sum_{t=1}^{13}\left|\frac{A_t - F_t}{A_t}\right| = 0.008$$

82629 = 0.88%

Mean absolute percentage error (MAPE) for Total stock prices sales:

$$\text{Predict} = \frac{1}{21}\sum_{t=1}^{21}\left|\frac{A_t - F_t}{A_t}\right| = 0.252383357$$

=25.24%

$$\text{Closed Loop Predict} = \frac{1}{21}\sum_{t=1}^{21}\left|\frac{A_t - F_t}{A_t}\right| = 0.028$$

82629 = 2.88%

Mean absolute percentage error (MAPE) for Guinness stock prices sales. :

$$\text{Predict} = \frac{1}{12}\sum_{t=1}^{12}\left|\frac{A_t - F_t}{A_t}\right| =$$

0.310897793=31.09%

$$\text{Closed Loop Predict} = \frac{1}{12}\sum_{t=1}^{12}\left|\frac{A_t - F_t}{A_t}\right| = 0.080$$

56121= 8.06%. The results are represented in Table 8

Table 8: Mean Absolute Percentage Error (MAPE) for Total, Guinness and Nestle

	Mean absolute percentage error (MAPE)		
	Total	**Guinness**	**Nestle**
Forecast	0.1147	0.9667	2.6940
Closed Loop Forecast	0.0088	0.0806	0.0225

Table 8 contains Mean absolute percentage error (MAPE) for Total, Guinness and Nestle forecast and closed loop forecast. The validation result shows that the Mean absolute percentage error (MAPE) for Total, Guinness and Nestle are close to zero this implies that the proposed hybrid model is effective and efficient for the prediction of all the three datasets.

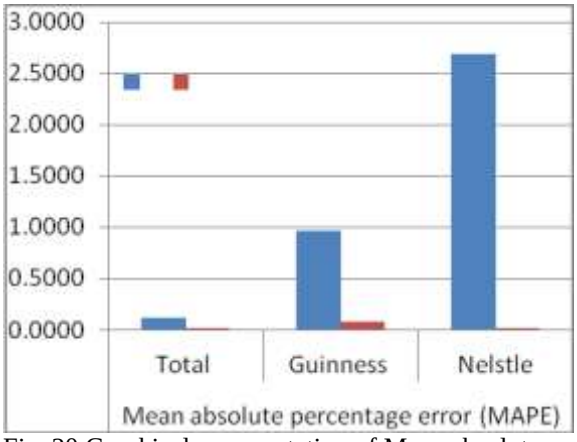


Fig. 20 Graphical representation of Mean absolute percentage error (MAPE) for Total, Guinness and Nestle Plcs for forecasting and closed loop forecast.

3. CONCLUSIONS

In Stock indices it was observed that they are highly non-linear, stochastic and volatile in nature. Predicting the stock prices accurately have always been a challenge to mankind. In order to overcome the limitations of neural networks, k-means clustering algorithm was integrated to form an enhanced model. Finally, the proposed enhanced network model is found to predict the stock prices more accurately when compared with other models. Potential investors and active day stock traders need a proactive strategy to secure their investment portfolios. The Nigerian stock market is non linear; hence it is predictable using neural networks. The use of artificial neural networks (ANNs) in a non-linear market does not require an understanding of the market dynamics. This is why it is practically feasible and profitable to use machine learning systems like neural networks to predict the behaviour of financial instruments such as stocks. From our projection results we can infer that:

i. Reasonable profit can be obtained in stock markets (especially Nigerian stock market) trading with ANNs (appendix).

ii. When forecast to buy is realized, you have to later sell the stocks at a favorable price margin above your purchase price.

iii.

iv. Forecasts to sell therefore at the reasonably higher price may not be the next day.

REFERENCES

[1] Huang, D, and Balvers, R. J (2005). "Productivity-based asset pricing: Theory and evidence," *Journal of Financial Economics*, *Elsevier*, vol. 86(2), pages 405-445, November.

[2] Sachin K,Shailesh J and Thakur R.S(2014);Stock Market Behavior Prediction Using NN Based Model. British Journal of Mathematics and Computer Science 4(17):2502-2515.

[3] Fazel Zarandi M.H, Rezaee B, Turksen I.B and Neshat E. (2017).Two Type Fuzzy Rule Based Expert in Developing Stock Price Analysis. Proceedings of the International MultiConference of Engineers and Computer Scientist 2009 Vol,IMEC 2009,March 17-19,Hong Kong

[4] Preethi ,G and Santhi. B(2012).Using K-means For Clustering data in different independent FNN models. *Journal of Theoretical and Applied Information Technology*, Vol.46 No.1

[5] Zhang, H and Huang (2019). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based Systems*. Proceedings of IEEE annual conference on Neural Network..II,pp451-458.

[6] Efendigil, O, and Kahraman (2019) (ANFIS) Forecasting System in Predicting Fuzzy Demand with incomplete Information. European *Journal of Computer Science and Information (June 2009),*. Published by European Centre for Research Training and Development UK (www.ea-journals.org) 37 Vol. 1 No. 1, pp. 30-39

[7] Sheta, A.; (2006) "Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models" , *IEEE International Conference on Fuzzy Systems*, Page(s):171 – 178

[8] Tan, T.Z.; Quek, C.; Ng, G.S.;(2005) "Brain-inspired genetic complementary learning for stock market prediction", IEEE Congress on Evolutionary Computation, 2005.Volume 3, Sept. Page(s):2653 – 2660

[9] Kyoung-jae Kim; (2006) "Artificial neural networks with evolutionary instance selection for financial forecasting", Expert Systems with Applications 30, pages 519-526.

[10] Kuo; C.H. Chen, Hwang; Y.C. (2011) "An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network", Fuzzy Sets and Systems 118 (2001) pages 21-45

[11] Hassan, Nath & Kirley, (2018) Utilizing the strengths of Hidden Markov Models (HMM), ANN and Genetic Algorithm (GA) to forecast financial market behavior. *Global Economy & Finance Journal* .Vol 45, pp 114-118.

[12] Neenwi.S,Asagba P.O and Kabari L.G(2013) Predicting The Nigerian Stock Market Using Artificial Neural Network. Euopean *Journal of Computer Science and Information* Vol.1 No.1, pp.30-39,June 2013.

ONTOLOGY- BASED TECHNIQUE FOR MEDICAL INTELLIGENCE PROCESS

Agbakwuru O.A., Amanze B.C and Agbasonu V.C

Department of Computer Science, Faculty of Physical Sciences,
Imo State University, Owerri, Imo State, Nigeria.

Abstract

The business model developed focused on expert system for health sector that uses intelligent agent to guide doctors in accurately carrying out disease control procedures. The objective of the research is to create ontology-based data integration (OBDI) system process model that can uses intelligent agent to guide doctors in accurately carrying out disease control procedures. The system developed used to manage a disease registry that consists of the concepts of the domain, the attributes characterizing each disease, the different symptoms, and treatments. A model for enhanced medical intelligence process using ontology based technique developed. The design provided for a database system for storing medical records, software for enhanced Medical Intelligence Process that would be more user-friendly, flexible, adaptive, intelligent, agile and automatic in integrating and analyzing medical data thereby helping medical practitioners at various levels to make realistic intelligent and real-time decision on critical health issues. Object Oriented Analysis and Design Methodology (OOADM) adopted in the design of the system. The system achieved integration of various patients medical records from different hospitals using ontology based and virtual data integration technique that will allow clinic data of one patient collected together to form a combinational resource, and could be accessed by physician if authority is assigned to the physician. Ontology-based data integration technique for disease control procedure achieved 95% accuracy in predicting the disease control procedure.

Keywords: Patients, production rule, OOADM, OBDI technique and Expert System

1. Introduction

Medical intelligence is the ability to detect and cure an ailment on time with minimal effort. It requires vast knowledge on disease symptoms, cure, and this can only be achieved by having

a data warehouse build from the knowledge of medical experts. To apply medical intelligence effectively, the healthcare condition of the patients ascertained. Ontology for data integration is as a data model that consists of parts such as classes, properties and relationships between them. It says the term ontology refers to a machine-readable representation of knowledge, particularly for automated inference. .The health-care condition of a patient defined as all the past and current medical and social information about the patient that may affect the professional immediate and short-term management of that patient. In this research, this information corresponds to all the diseases, syndromes and social issues that are diagnosed for the patient, the signs and symptoms (including family medical history), the problem assessments performed (i.e., medical, social, cognitive, and mobility tests), and the current interventions, either pharmacological, rehabilitative, nurse care, social care, counseling, and special medical services. In healthcare system all over the world, the amount of patient oriented data is constantly growing. More hospitals are opening up with various departments / units. For example, the Intensive Care Unit (ICU) is an extremely data intensive environment where large volumes of data from patient monitoring and observations are recorded continuously. Physicians and nurses could generate such patient oriented data from medical devices, laboratory results, electronic prescriptions, therapeutic decisions, and clinical observed values. These data disintegrated and access to the data done by requesting for it through the various hospital units. This is not only time consuming but obsolete. The following are the existing challenges of the medical system;

1. Lack of fast, accurate, reliable and intelligent software solutions that can help healthcare practitioners make decisions that would solve urgent, and in some cases, complex medical problems in real-time.
2. Cost of processing and analyzing large volumes of data in a medical environment is high most especially in terms of time consumption.

Ontology-Based Data Integration (OBDI) Technique

In computer science, ontology is a controlled vocabulary that describes objects and the relations between them in a formal way. Ontologies provide a sound basis for sharing domain knowledge between human and computer programs, or between computer programs. An ontology normally defines concepts (or classes), individuals (or instances), properties, relationships and their constraints. Logical formalization of ontology language ensures

semantic interpretation, i.e. inference, by computer programs. Ontology is a major instrument toward realization of the Semantic Web vision (Rosse, 2013).

Helena and Lidia (2009) defined ontology as a formal, explicit specification of a shared conceptualization and further defined conceptualization to be an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use explicitly defined. Ontologies allow more complete and precise domain models. Ontologies intended shared and reused, and the approach perceived to be beneficial. Ontology-based design has an advantage of being syntactically correct and semantically consistent as a model.

Vipul et al. (2008) Ontologies provide a common language to express the shared semantics and consensus knowledge developed in a domain. This research will explore this in its ontology-based integration technique phase.

Ontology based Data Integration involves the use of ontology(s) to effectively combine data or information from multiple heterogeneous sources. It is one of the multiple data integration approaches and classified as Global-As-View (GAV). The effectiveness of ontology based data integration closely tied to the consistency and expressivity of the ontology used in the integration process. (www.wikipedia.com, 2016)

Inter-application interoperability seen as schema mapping and data integration problem. In this manner, integration requires mapping systems and integration systems that uses those mappings to answer queries or translate data across data sources. There are three different categories of ontology-based integration approaches; single ontology approaches (SOIA), multiply ontology approaches (MOIA), hybrid ontology approaches (HOIA) (Bostjan and Vili, 2010)

Single ontology approach: A single ontology used as a global reference model in the system. This is the simplest approach simulated by other approaches. The SIMS (Search in Multiple Sources) system is a prominent example of this approach. The Structured Knowledge Source Integration component of Research Cyc is another prominent example of this approach. (Title = Harnessing Cyc to Answer Clinical Researchers' Ad Hoc Queries)

Multiple ontologies: Multiple ontologies, each modelling an individual data source used in combination for integration. However, this approach is more flexible than the single ontology approach; it requires creation of mappings between the multiple ontologies. Ontology mapping is a challenging issue and is focus of large number of research efforts in computer

science. The OBSERVER (Ontology Based System Enhanced with Relationship for Vocabulary heterogeneity Resolution) system is an example of this approach.

Hybrid approaches: The hybrid approach involves the use of multiple ontologies that subscribe to a common, top-level vocabulary. The top-level vocabulary defines the basic terms of the domain. Thus, the hybrid approach makes it easier to use multiple ontologies for integration in presence of the common vocabulary.

Ontologies enable the unambiguous identification of entities in heterogeneous information systems and assertion of applicable named relationships that connect these entities together. Specifically, ontologies play the following roles:

a. Content Explication: The ontology enables accurate interpretation of data from multiple sources through the explicit definition of terms and relationships in the ontology.
b. Query Model: In some systems like SIMS, the query formulated using the ontology as a global query schema.
c. Verification: The ontology verifies the mappings used to integrate data from multiple sources. These mappings may either be user specified or generated by a system. (www.wikipedia.com, 2016)

Ontology allows more complete and precise domain models. They intended to be share and reused and one of the main advantages of its design is that it has syntactically correct and semantically consistent model and reasoning over them provides retrieval of additional rules possibly not recognized during the design phase. In any domain such as that of business intelligence systems, ontology play the role of providing a common language to express the shared semantics and consensus knowledge developed in such domain. The shared semantics typically captured in the form of various domain specific ontologies and classifications. The concepts provide the shared semantics to which various data objects and data interpretations mapped to enabling integration across multiple business intelligence, data sources and domains. (Vipul et al., 2008)

After the 80s, the massive adoption of database systems inside organizations leads to the need to integrate different data repositories with possibly incompatible data schemata. The process of integrating different data residing at different sources to provide a unified view of this information known as data integration problem. As early stated in this research, data warehousing is one of the approaches to data integration problem solution. Here data originates from different sources and are submitted to a process called ETL (Extraction,

Transformation and Loading) and the stored into a new database with a single and usually de-normalized schema. The resultant database often structured to store various aggregations of the sources' data in order to speedup query processing. Architecturally, data warehousing seen as a tightly coupled approach because the integrated data reside in a single place at query time. Recent approaches to data integration as would be found in our proposed research are sometimes "loosely coupled". (Letizia, 2016)

A data integration system provides a uniform interface to distributed and heterogeneous sources. These sources can be databases as well as unstructured information such as files, HTML pages, etc. One of the most important problems within data integration is the semantic heterogeneity, which analyzes the meaning of terms included in the different information sources. As earlier stated in this research, Data integration is concerned with unifying data that share some common semantics but originate from unrelated sources.

Heterogeneity classified into four categories: (1) structural heterogeneity, involving different data models; (2) syntactical heterogeneity, involving different languages and data representations; (3) systemic heterogeneity, involving hardware and operating systems; and (4) semantics heterogeneity, involving different concepts and their interpretations. The semantic heterogeneity deals with three types of concepts: the semantically equivalent concepts, the semantically unrelated concepts, and the semantically related concepts. In the first case – semantically equivalent concepts – a model uses different terms to refer the same concept, e.g. synonymous, or some properties modelled differently by different systems, for example, the concept length may be "meter" in one system and "mile" in one another. In the second case – semantically unrelated concepts – the same term may be used by different systems to denote completely different concepts; and in the last case – semantically related concepts – different classifications may be performed, for example one system classifies "person" as "male" and "female" and other system as "student" and "professor".

The main difference between ontology and a database schema is that the latter essentially constrains the possible states of the database, while the former has typically a model theoretic semantics and thus allows inferring new knowledge (in a deductive fashion).

With the use of ontology the following advantages is feasible when used for data integration, which are;

(1) The vocabulary provided by the ontology serves as a stable conceptual interface to the databases and is independent of the database schemas,

(2) The language used by the ontology is expressive enough to address the complexity of queries typical of decision-support applications,

(3) Knowledge represented by the ontology is sufficiently comprehensive to support translation of all the relevant information sources into its common frame of reference, and

(4) The ontology supports consistent management and recognition of inconsistent data.

Ontology gives the name and the description of the domain specific entities by using predicates that represent relationships between these entities. The ontology provides a vocabulary to represent and communicate domain knowledge along with a set of relationships containing the vocabulary's terms at a conceptual level. It is therefore possible to use ontology for data integration tasks.

The features for the building of the ontologies defined in the system divided into three sub-features reusability, changeability and scalability. Reusability refers to the ability of reuse the ontologies, that is, ontologies defined to solve other problems used in the system because of either the systems support different ontological languages and/or defines local ontologies. Changeability refers to the ability of changing some structures within an information source, without producing substantial changes in the system components. Finally, scalability refers to how easy the integrated system extended with new information sources. In general, the systems use languages based on Description Logics, although some Web–based languages (OWL) have recently emerged and is quite better (Augustina et al., 2014).

The interoperability of a system seen as consequence of technical, semantic, organizational, legal and political tools. It empowers transfer and usage of data in other information resources such as;

(1) Organizational. It specifies the regulation of resource interaction.

(2) Technical. It describes the compatibility of Information Technology (IT) tools, establishment and usage of open interfaces, standards and protocols in order to ensure effective data exchange.

(3) Semantic. This characteristic ensures that data from one information system understood and interpreted in the same way in other systems.

Systems must be able to exchange data. Data exchange between information systems determined by reciprocal agreements, which are different in each case: web-based services, open standards, specifications. Direct data integration is impossible if data processed by

applied information system logic. This process performed in real time in source system changes occur, fixed time intervals automatically or manually, using popular methods: Extract, Transform and Load (ETL), data replication, federation, event-based integration, web-based technologies and open standards. The aforementioned methods have essential disadvantages in the context of heterogeneous DBMS (Database Management System): the problems of automatic update neither considered nor solved; the same data is stored in several sources. Besides, there is no possibility to get data or information messages on databases using direct access interaction. The researchers of distributed heterogeneous databases have applied ontologies to support semantic interoperability: to integrate data sources developed using different vocabularies and to see data from a different perspective. The proposed ontology-based data integration phase of our business intelligence system hope automatically perform data extraction and integration from structured, semi-structured and unstructured data sources. (Virginija and Rimantas, 2011)

Normally, the organizational data resides in multiple data sources. For typical business intelligence (BI) data integration projects, the design and development of data integration processes involve collecting facts for the integration, analyzing data structures and their descriptions. However, it is inappropriate to focus on the management of data requirements only: it is very important to discern that integration is more than data. It also covers:

Data sources: what data from where has to be integrated?

Business rules (BR): which BRs have to evaluate for data processing and keeping in data sources?

Transformations: which transformations have in order to avoid structural and semantic conflicts?

The integration of data ensures data management in the way that they unambiguously identified in information system (IS) and it is possible to transfer, transform, load and use them in other information system (IS) or source without changing program code. Ontology-guided data integration makes the process more efficient – reducing the cost, maintenance and risk of the project.

Again, the term “ontology” refers to a machine-readable representation of knowledge, particularly for automated inference. Ontology is a data model consists of these parts: classes, properties and relationships between them. The power of ontologies lies in the ability to represent relationships between the classes. The main benefit of using the ontology-based model is its runtime interpretation. One of the major advantages of the ontology model is an

assumption of open-world. The reason for the popularity clearly interpreted dissemination of knowledge between people and applications. Moreover, ontology supports the integration task as it describes the exact content and semantics of these data sources more explicitly. He ensures that if a highly descriptive semantic representation of the available knowledge built, it reused variety of business applications without the need for repeated integration exercises. Furthermore, the new knowledge gathered from different sources can build upon the current knowledge because all of it exists in a semantically consistent system. Thus, we conclude that knowledge is the foundation of all successful decisions. (Virginija and Rimantas, 2011)

An overview of the requirements, which automatically satisfied by an ontology-based process, is given in Figure 1.

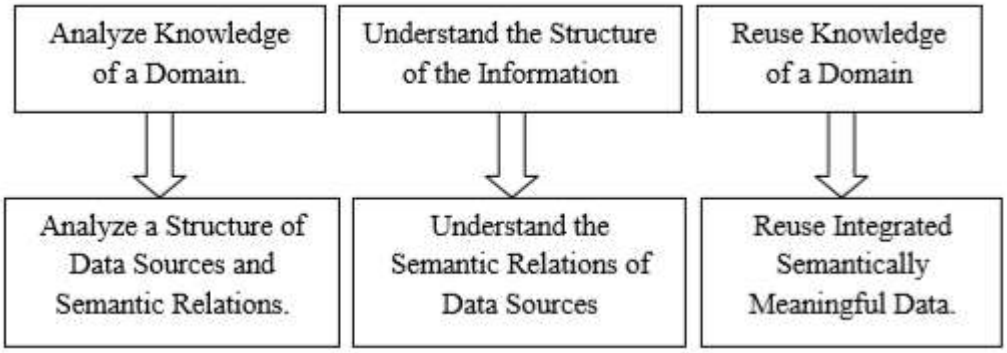


Figure 1: Transformation of Ontologies Features to Data Integration Systems Requirements (Virginija and Rimantas, 2011).

2. Review of Related Works

Table 1: Summary of Related works

Author	Techniques	Work done	Limitations
Marut, 2016	Ontology-based	clinical reminder system that link clinical guideline knowledge with patient registries	The paper didn't integrate electronic health record (EHR) standards
Madhura, 2020	open data integration platform	facilitates centralization of data assets	Lacks analytics, data visualization, monitoring and reporting functionalities for clinical decision support
Chih-Lin, 2019	machine learning	facilitating personal health care, reducing costs of health care, and improving outcomes	The work was carried out using a single hospital and may not represent the facts

			when you broaden the scope.
Sozan, 2019	Adaptative Neuro-Fuzzy System (ANFIS)	The obtained simulation results demonstrate the efficiency of using ANFIS model in the identification of heart attacks	The intelligent system was limited to heart attacks only
Serdar, 2014	Survey	A major finding of the survey is that although significant advances have been made in introducing AI technology in critical care, successful examples of fielded systems are still few and far between	Theoretical review. No practical implementation
Vishesh, 2010	cloud computing	Addressed the challenge of sharing medical data	Didn't incorporate medical intelligence
Dipti, 2010	Expert System	help a great deal in identifying those diseases and describing methods of treatment to be carried	Lacks data integration
Soltan, 2013	Expert System	Able to give appropriate diagnosis and treatment for two heart diseases namely; angina pectoris and infarction	Is limited to two heart diseases
Ighoyota, 2017	Fuzzy	The analysis clearly shows the effectiveness and accuracy in the system performance through false result elimination	Lacks data integration
Alexander, 2011	Multi Agent Enhanced Business Intelligence	Results indicate that the pMAEBI managed stores performed better (in terms of profit) than the comparison stores	Narrowed to product pricing
David, 2020	ontology-based personalization	Helps health-care professionals to detect anomalous circumstances such as wrong diagnoses, missing	Needs specialized knowledge to operate

		information, unobserved related diseases, or preventive actions	
Agustina, 2018	Framework for Comparison	This survey describes seven systems and three proposals for ontology-based data integration.	It is a survey
Bostjan, 2014	ontology based	bridges the gap between ontology based integration and service oriented architecture by enabling dynamic and transparent integration of information which is provided by services	The problem of splitting the query into static and dynamic query was not addressed fully.
Ali et al., 2018	Virtual Data Integration	They developed a Virtual – Data Integration Framework (V-DIF) that meets most of the users' expectations	concentrate mainly on data integration process and avoid or ignore the other two processes (inconsistency detection and resolution)
Ali, 2018	Mapping Approach	Provides a linkage between the fundamental components required to provide accurate and unambiguous answers to the users' queries from the integration system	Cannot use the sources of the data to resolve the duplicate through source preferences.
Vinoth, 2019	Ontology based	By using Internet of Things will help us to cure the patient in a short period of time	The paper didn't integrate electronic health record (EHR) standards
Taqdir, 2017	Intelligent-Knowledge Authoring Tool (I-KAT)	Developed technologically integrated healthcare system	Increased complexity
Richter and Weber 2016	case based reasoning	Medical dataset	Existence of many problems without solutions
Jagannathan, 2009	KNN	The performance of CBR applications was enhanced	The missing data values of some attributes have been handled while others are not

			treated
Asma, 2011	Decision tree	Prediction of presence and absence of diabetes	It is a review of techniques and no model was developed
Kapil, 2010	Support Vector Machine (SVM) and Artificial Neural Network (ANN)	Diagnosis heart disease	The accuracy is low when compared to other research
Nassim, 2014	Fuzzy logic	Modeled clinical practice guidelines	Is not a dynamic systems

3. Model of the Diagnosis Ontology and their relationships between sub-ontologies

In the healthcare sharing platform, ontologies used for describing semantic meaning of information source explicitly in order to solve semantic heterogeneity as shown in Figure 2. We adopt hybrid ontologies in our research. In effect, the proposed ontology-based and virtual data integration architectural process based on the use of ontology, which explicitly captures knowledge about different types of data sources, and virtual aspect uses mediators to bring about the real-time and agility aspect of the system. Generally, database schemas regarded as static, but ontology schemas typically assumed highly dynamic and are an evolving object(s).

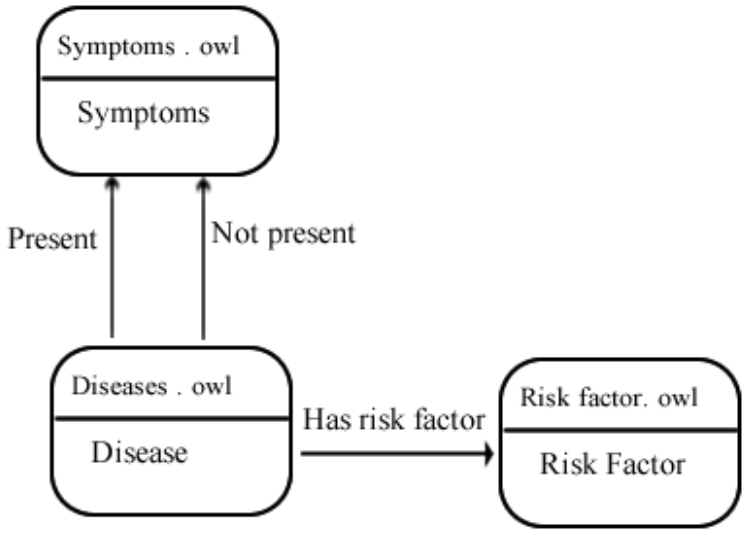


Figure 2: Diagnosis ontology model using Web-Ontology Language **(**OWL)

The proposed ontological model presented in a modular fashion, which divided into the following sections: diseases (age group and gender); symptoms with anatomy, intensity and evolution, and risk factors. A diagnosis must be an instance of the disease concept and it governed by the Catalogue of Disease. Therefore, three sub-ontologies created that shape the complete model Diseases, Symptoms and Risk Factors. Figure 2 shows the complete model of the Diagnosis Ontology and their relationships between sub-ontologies.

Algorithm

The algorithm for the proposed model is as follows

Start

Algorithm to set up the disease control dataset

Enter the disease symptoms

Enter the procedure for treatment

Create a dataset

Store in database

Stop

Algorithm for disease control

Start

Enter the symptoms

Query the disease db

Integrate all the dataset found

Match the symptoms entered with the one in the database

Use intelligent agent to filter the database

Apply disease ontology

Search for best matching case

Call Virtual data integration algorithm

Search the knowledge base

Is the disease a new case?

If yes then store in the disease dictionary else

Find the matching case

Is similar case found?

If yes search for the disease control procedure

Otherwise search other dataset from global view

Extract the suggested disease control procedure that matched the disease found from disease ontology

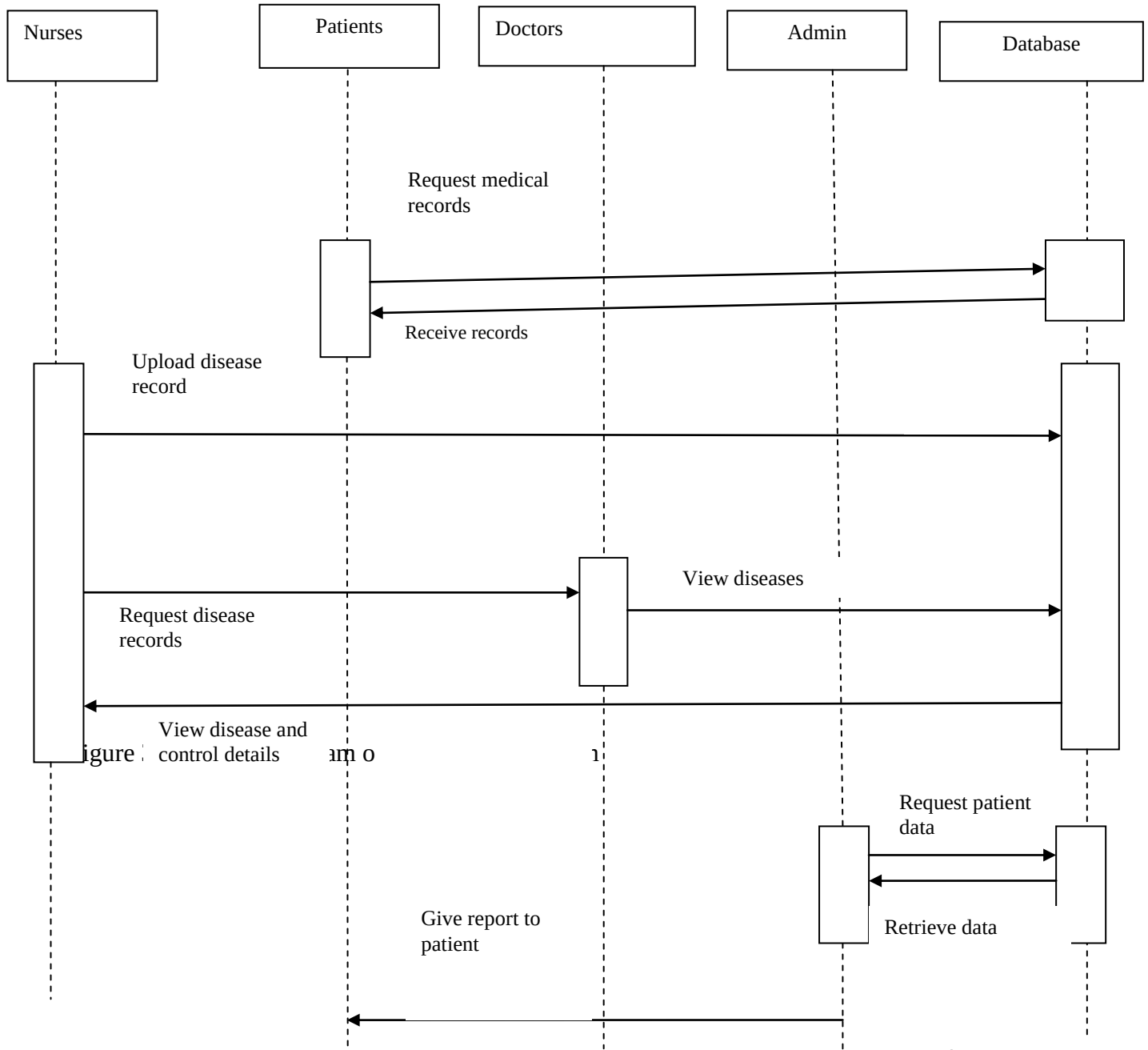
Display the suggested disease control procedure

stop

OOADM Diagrams

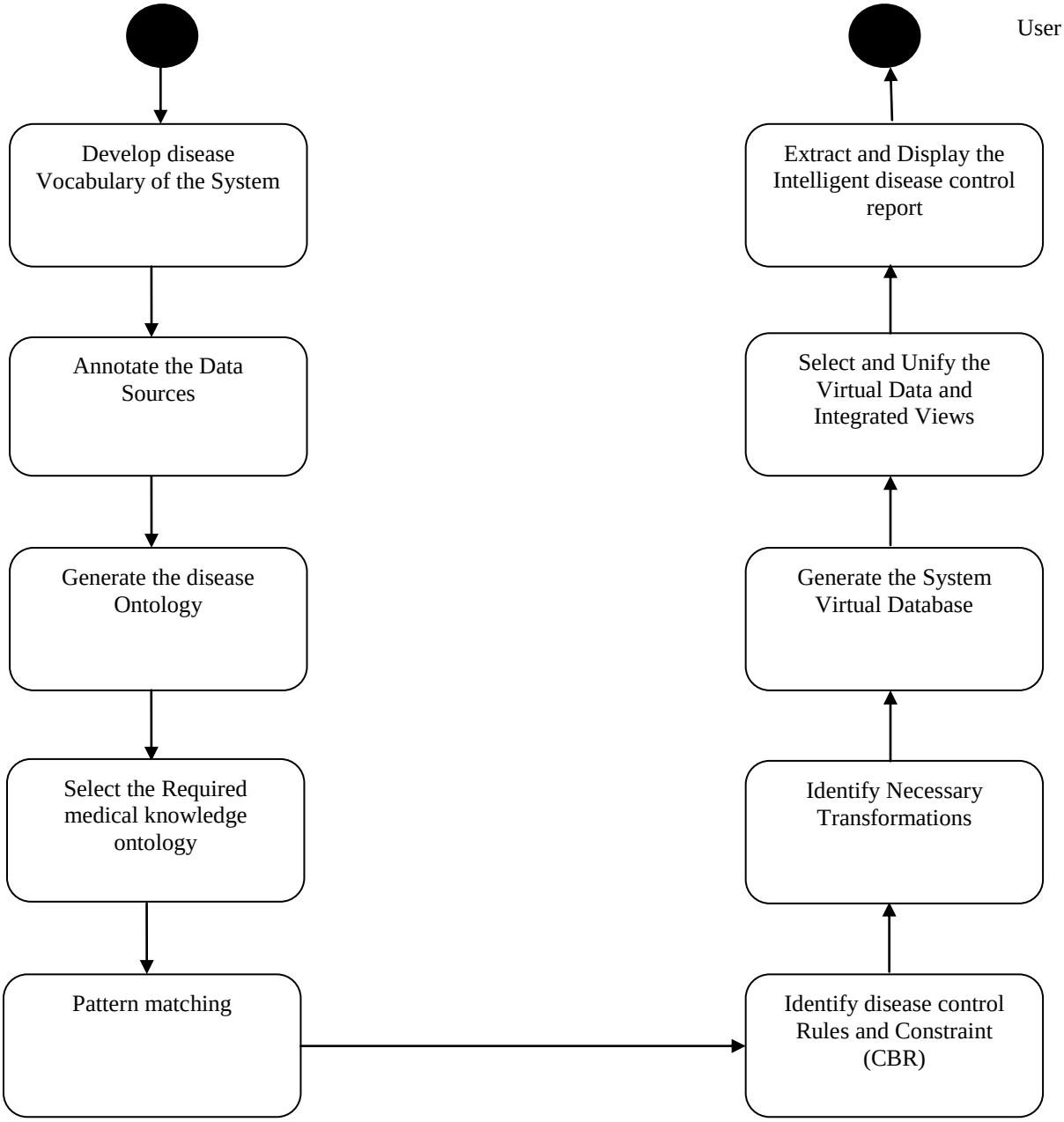
Sequence Diagram

The sequence diagram in Figure 3 shows how objects interact with one another and in what order. It depicts the objects and classes involved in the scenario.



igure 3: am o ı

Activity Diagram



Intelligence process that the research is enhancing.

Collaboration Diagram

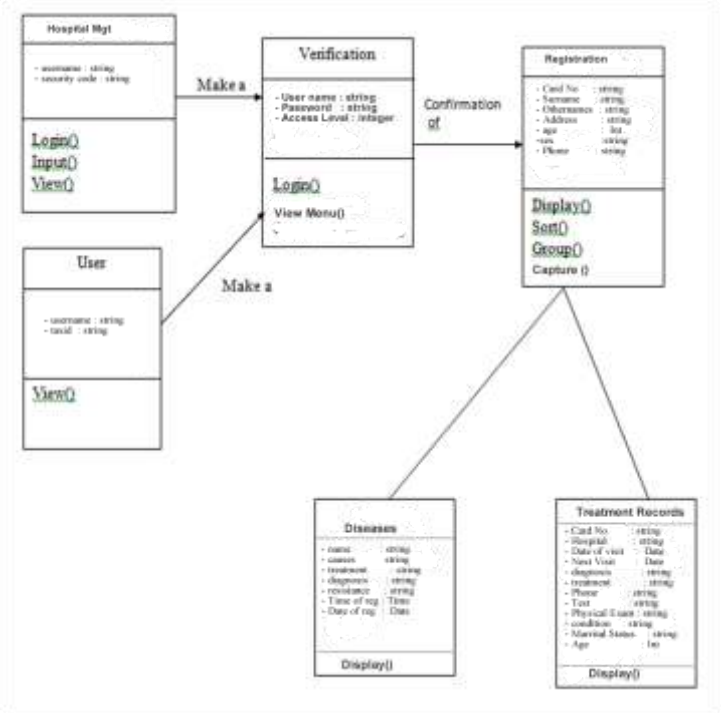


Figure 5: Collaboration Diagram of the proposed system

Figure 5 shows various information sources that is needed at various stages of the data integration processes.

Result

Table 2: **Confusion matrix applied to test dataset of Disease Control**

		Observed	
		True	**False**
Predicted	**True**	18	0
	False	1	1

Table 2 shows that out of 20 transactions, 18 diagnoses are True Positive and was predicted correctly. One diagnosis detected to be False Negative while it is not. Finally, one False Positive detected. A model of performance metrics derived from the confusion matrix as show in equation 3, which show the accuracy of the system.

Substituting the values, we have

AC = (18+1) / (18+0+1+1)

AC = 0.95 i.e. 95% accuracy in predicting the disease control procedure

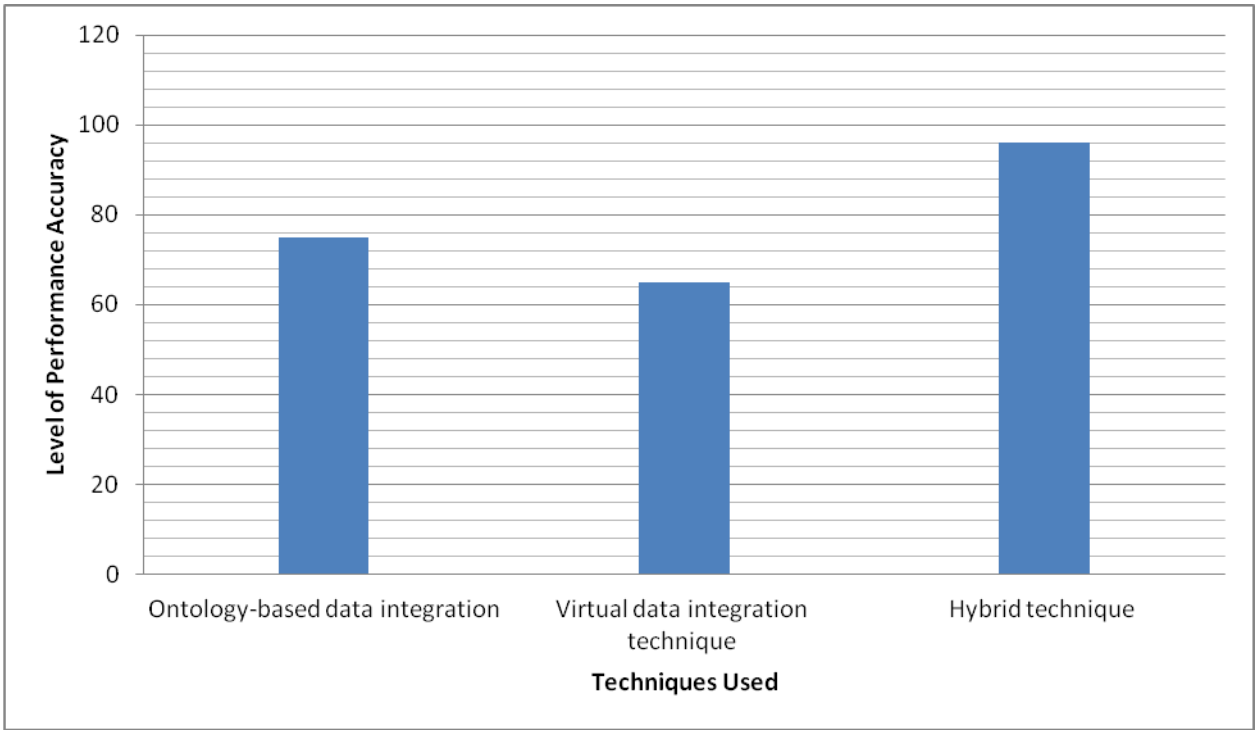


Figure 6: Comparison of level of prediction accuracy using various techniques

From Figure 4.17, one can see that the Ontology-based data integration technique for disease control procedure has 75% accuracy in predicting the disease control procedure; Virtual data integration technique for disease control procedure has 65% accuracy in predicting the disease control procedure; while Hybrid technique using both Ontology-based data integration and virtual bydata integration technique for disease control procedure has 95% accuracy in predicting the disease control procedure. This shows that the Hybrid technique outperforms the existing techniques with (95 – 75) = 20%, i.e. there is 20% improvement from the existing technique.

Conclusion

Utilizing ontology-based data integration is an attractive avenue as it is also a key factor for enabling interoperability. However, integrating vast amount of information from different sources is a difficult, complex and demanding task. The use of ontology-based data integration systems tool is automate full in the data integration. A relational database for

storing and tracking disease outbreak and control using ontology-based data integration (OBDI) achieved.

References.

Rosse, C. and Mejino, J.L.V. (2013). A reference ontology for biomedical informatics: the Foundational Model of Anatomy**.** Journal Biomedical Information, 3(6), 478-500

Helena, V. and Lidia, R. (2021) "Ontology-based Data Warehouse Development Process" Proceedings of the ITI 2009 31st International Conference. on Information Technology Interfaces, June 2225, 2009, Cavtat, Croatia, retrieved on June, 2021

Vipul, K., Kei-Hoi, C., Donald, D., Matthias, S., Scott, M., Joanne, L., Susie, S., Ivan, H. and Raymond, H. (2008). Chapter 5 An Application Of Ontology-Based Data Integration For Biomedical Research" retrieved on May, 2016

Bostjan, G. & Vili, P., (2021). Automating ontology based information integration using service orientation" WSEAS Transactions on Computers Bostjan Grasic, Vili Podgorelec 9(6), 12-26

Letizia, T. C. (2016). An Ontology-Based Data Integration System: Solving Semantic Inconsistencies" Relatore: Letizia Tanca Correlatore: Ing. Carlo Curino Tesi di Laurea di: Giorgio ORSI matr. 674222 Anno Accademico 2016.

Augustina, B., Alejandra, C. and Nieves, R. B. (2016). Ontology-Based Data Integration Methods: A Framework for Comparison", retrieved on July, 2016

Virginija, U. and Rimantas, B. (2011). Ontology-based Foundations for Data Integration. The First International Conference on Business Intelligence and Technology Copyright (c) IARIA, 2011. ISBN: 978-1-61208-160-1 retrieved on October 2014

Marut, B. (2016). Ontology-based Clinical Reminder System to Support Chronic Disease Healthcare. Article in IEICE Transactions on Information and Systems · March 2016 DOI: 10.1587/transinf.E94.D.432 · Source: DBLP

Madhura, J., Dinithi, N., Daswin, S., Damminda, A., Brian, D., Kate, E .W. (2020). A data integration platform for patient-centered e-healthcare and clinical decision support. Research Center for Data Analytics and Cognition, La Trobe University, Victoria, Australia b School of Allied Health, La Trobe University. Victoria, Australia

Chih-Lin, C. (2019). Medical decision support systems based on machine learning. PhD (Doctor of Philosophy) thesis, University of Iowa, https://doi.org/10.17077/etd.o5gmwvxk

Sozan, S. M. (2019). Intelligent System for Identification Heart Diseases. A thesis submitted to the graduate school of applied sciences of near East University

Serdar, U. (2014). Intelligent Systems in Patient Monitoring and Therapy Management. Knowledge Systems Laboratory, Stanford University, 701 Welch Road Bldg. C Palo Alto, CA 94304, USA

Vishesh, V. (2010). Personal health record system and integration techniques with various electronic medical record systems. A Thesis Submitted to the Faculty of The College of Computer Science and Engineering in Partial Fulfillment of the requirements for the Degree of Master of Science Florida Atlantic University, Boca Raton, Florida

Dipti, P. S., Santosh, K. P. (2010). An Expert System for Diagnosis of Human Diseases. International Journal of Computer Applications, 1(13), 23-34

Soltan, R.A., Rashad, M. Z. , El-Desouky, B. (2013). Diagnosis of Some Diseases in Medicine via computerized Experts System. International Journal of Computer Science & Information Technology, 5(5), 26-45.

Ighoyota, B. A. and Sujatha, P. (2017). Fuzzy Based Multi-Fever Symptom Classifier Diagnosis Model. International journal Information Technology and Computer Science, 6(10), 13-28 Published Online October 2017 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijitcs.2017.10.02

Alexander, P., Joseph, L. (2017). Multi Agent Enhanced Business Intelligence for Localized Automatic Pricing in Grocery Chains" A dissertation submitted in fulfillment of the requirements of the degree of Doctor of Philosophy for the School of Information Technology, Bond University. December 2017. www.fulltext.pdf.

David, R., FrancisReal, J. (2020). An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. Journal of Biomedical Informatics, 45 (3), 429-446

Ali, Z. (2018). A Mapping Approach for Virtual Data Integration System Processes. International Journal of Advanced Computer Science and Applications, 9(12), 24-37

Vinoth, K. (2019). Ontology based Public Healthcare System in Internet of Things. 2nd International Symposium on Big Data and Cloud Computing (ISBCC'19). Available online at www.sciencedirect.com

Richter, M.M., Weber, R.O. (2016). Case-based reasoning. Springer-Verlag Berlin

Nassim, D., Elpiniki, I. P., Jos, D., Hans, C. and Marie-Christine, J., (2014). Clinical Decision Support System based on Fuzzy Cognitive Maps, INSERM UMR_**S** 872, Eq 20, Medicine Faculty, Pierre and Marie Curie University , Paris 6, France

Development of An IoT-Based Gas Detection System

D.O. Njoku[1] , F.O. Nwokoma[1], S.A. Okolie[1], J.N. Odii[1], V.C. Iwuchukwu[1], J.E. Jibiri[2], E.O. Okechukwu[1], F.U. Madu[3]
[1]*Department of Computer Science, Federal University of Technology, Owerri, Imo State-Nigeria*
[2]*Department of Information Technology, Federal University of Technology, Owerri, Imo State-Nigeria*
[3]*Department of Computer Science, Federal Polytechnic, Nekede, Owerri, Imo State*

Abstract –Liquefied Petroleum Gas (LPG), has been more useful in these recent times. This is reasonable considering the fact that LPG is environmental friendly that produced lesser residues compared to other fuels while burning. Notwithstanding, LPG can be very hazardous when it gets leaked from its storage causing harm either by suffocation because it is twice heavier than air or explosion as it is highly flammable. This paper presents a detailed research on development of an IoT-based Gas Detection System''. This system has the ability to detect the presence of an LPG, methane, and smoke with the help of an MQ5 sensor when gas leakages is experienced. The system measures the concentration of the leaked gas in the atmosphere and displays the status which could either be "Low", "Mid", "High" concentration or "Emergency" at a worst case scenario using a 16 x 2 LCD. The system produced a sounds alarm through a buzzer upon the detection of any of the mentioned gases and sends an SMS to a predefined number; a GSM Module is being used to achieve this feature. Other materials used for this paper include an Arduino UNO microcontroller, jumper wires, SIM card, power bank (for power supply) etc. The Arduino UNO microcontroller was programmed using C++ on an Arduino IDE..

Keyword: Arduino, Detection, Emergency, Hazardous, IoT, Microcontroller, MQ5 Sensor

4. INTRODUCTION

The use of energy is important to human society for solving problems in the environment. Energy resources are used to power computers, for agriculture, transportation, garbage collection, information technology and human communication in developed countries. It is integral that we tend our energy resources; making them more convenient and safe to use. Stock Energy is the lifeblood of the modern economy is essential to life and activities of this digital age. It is estimated that 320Billion KW/hr is being exhausted per day. The sources of this consumed energy includes fossil fuels (coal and natural gas), bio fuels and biomass, fuel cells, solar, nuclear, wind, geothermal , and oceanic energy, but predominantly is the fossil fuels [1].

In the last decades, new applications for gas energy emerged. The Liquefied Petroleum Gas is one of the best choices of energy in industries and homes. It offers thermal, light and electric power, in the homes, it serves for cooking, warming the house, lightening, heating the water etc. In industries likewise it is used for food processing, metal processing, textile, refrigerants, printing, chemical production, etc. Liquefied Petroleum Gas which is also known as auto gas consists of mostly propane (C_3H_8) or butane (C_4H_{10}) it is twice heavier than air and is a green house gas, however, a clean burning one. Several lives and properties have been lost over the years to Liquefied Petroleum Gas explosions, most of these accidents could have been avoided with the help of a gas detector. Nonetheless the gas sensor (MQ5) to be used in this paper is not only able to detect the presence of LP gases but also CO and smoke. (It is important to note that the chemical composition of smoke is dependent on the nature of the burning fuel and the conditions of combustion.)

The motivation of this research is to address the necessary challenges considering the number of deaths caused by gas explosion due to leakage; this paper study hopes to save humanity by solving the following problems: MQ5 which is an excellent gas sensor will serve as substitute for the human smell organ in this work. Inability to immediately and quickly notify everyone in the environment where there is a leakage.

More often, where there are no gas alarm systems only few persons are aware of a gas leakage at its early stages. An audible alarm system is incorporated while designing the gas detector in this work to ensure everyone in the affected

Several gas explosions could have been avoided or managed adequately if the authorities are notified as soon as there are a gas leakage the objectives of this research paper is develop an IoT-based gas detection system to develop a system that

detects the presence of gas and/or smoke with the help of the gas sensor component. The paper focuses in addressing in achieving these: to produce an audible fabricated sound alarm when the gas sensor senses a gas and/or smoke using a buzzer it automatically gets silenced when the gas outflow gets regulated; to measure the concentration of the gas in the atmosphere; to display the gas leakage status in an LCD using a 16×2 LCD module; to ensure that an SMS is sent to a predefined phone number when there is a leakage using the GSM module.

5. REVIEW OF RELATED WORK

The early coal miners were able to improvised gas detectors to save them from methane explosions or displacement of oxygen in the air leading to death. Methane explosions occur in mines when a buildup of methane gas, a byproduct of coal, comes into contact with a heat source, and there is not enough air to dilute the gas to levels below its explosion point. According to [2] in the this research presented that, gas detectors would be discussed beginning from the earliest methods down to the most present technologies. Gas detectors have been in need since the discovery and use of gas as a source of energy. Alongside industrial environments and places where gases are used as a source of energy, gas detectors have been found to also be of importance in every surrounding considering the general atmospheric state of the 21st century. Gas detection has evolved over time and every new gas detector system comes with an improvement and/or diversity of its sensor. Gas detectors have so far been able to detect hydrogen sulfide, carbon monoxide, oxygen, sulfur dioxide, phosphine, ammonia, nitrogen dioxide, hydrogen cyanide, chlorine, chlorine dioxide, ozone and combustible gases[5].

Photo Ionization Gas Detector

The Photo Ionization Detector (PID) is a portable vapor and gas detector that detects a variety of organic compounds. Photo ionization occurs when an atom or molecule absorbs light of sufficient energy to cause an electron to leave and create a positive ion. The PID is comprised of an ultraviolet lamp that emits photons that are absorbed by the compound in an ionization chamber. Ions (atoms or molecules that have gained or lost electrons and thus have a net positive or negative charge) produced during this process are collected by electrodes. The current generated provides a measure of the analyte concentration. Because only a small fraction of the analyte molecules are actually ionized, this method is considered nondestructive, allowing it to be used in conjunction with another detector to confirm analytical results. In addition, PIDs are available in portable hand-held models and in a number of lamp configurations. The PID may give false positive readings for water vapor. Rain may also affect performance. High humidity can cause lamp fogging and decreased sensitivity [3]. This can be significant when soil moisture levels are high or when a soil gas well is actually in ground water. High concentrations of methane can hinder performance. Rapid variations in temperature at the detector, strong electrical fields, and naturally occurring compounds, such as terpenes in wooded areas, may affect instrument response. The PID must be re-calibrated frequently Detection limits for most PIDs are in the parts per million range. Thus they are unsuitable for most vapor intrusion indoor air investigations, where screening or action levels are normally in the parts per billion range.

The PID is used mostly to detect volatile organic compounds in soil, sediment, air and water. It is often used to detect contaminants in ambient air and soil during drilling activities and during spills to identify potential problems.

Common substances that a PID can detect and monitor include: Benzene, toluene, vinyl chloride, Hexane, Isobutylene, Jet fuel, styrene, Allyl alcohol, Mercaptans, trichloroethylene, Perchloroethylene, Propylene oxide, Phosphine

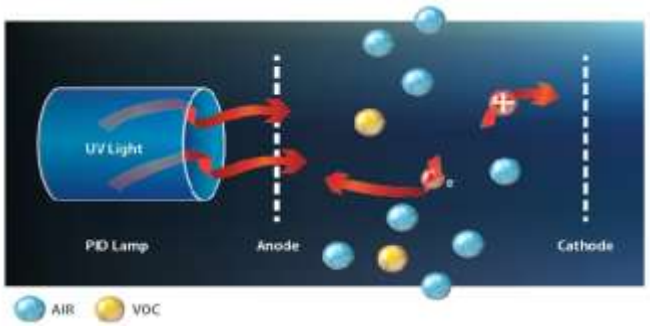


Figure 1: Process of Photoionization *[3]*

Infrared Sensors

Some gases are not very reactive and require other detection methods. Carbon dioxide (CO_2) is an example of an important gas that cannot be detected using typical electrochemical cells. Infrared (IR) sensor, is commonly used to detect CO_2 or hydrocarbons, it approaches gas detection in a completely different way. With IR sensors, the amount of gas is determined by how much light the gas absorbs, rather than by a chemical reaction. Gases absorb certain wavelengths of light, and certain gases absorb

certain frequencies. For instance, astronomers determine the composition of distant stars and galaxies by noting which wavelengths of light are missing from the spectrum coming from the object. The wavelength of light that is allowed to pass through must match the wavelength that is easily absorbed by the gas of interest. The amount of light energy received at the detector decreases as more of the "target" gas passes into the sensing cavity. For CO_2, which is commonly found in fermentation processes such as brewing, this is the best means of detection available in portable instruments. Some manufacturers also offer infrared sensors for combustible gas detection. The main drawback to IR sensors is cost. IR sensors are more complex and expensive even though they have the potential to last longer. As the technology matures, it is possible that IR sensors may overtake catalytic sensors as the preferred choice for combustible gas detection [5]

Catalytic bead sensors have been replaced by gas monitors based infrared gas detection. This is because the infrared method is free of frequent calibration requirements, silicon poisoning and the need for oxygen to be present in order to detect hydrocarbon gases. As a result, infrared detectors have become more reliable for monitoring hydrocarbon within fixed gas detection systems. The microprocessor used in typical infrared gas detection systems continuously monitors the status of the source and receiver and communicates any errors or fault conditions to the controller, transmitter or other PLC and DCS system. Since the infrared gas detection wavelength implemented in hydrocarbon detection is similar to that needed for carbon dioxide detection, these systems are also great for toxic and percent by volume monitoring of CO_2.

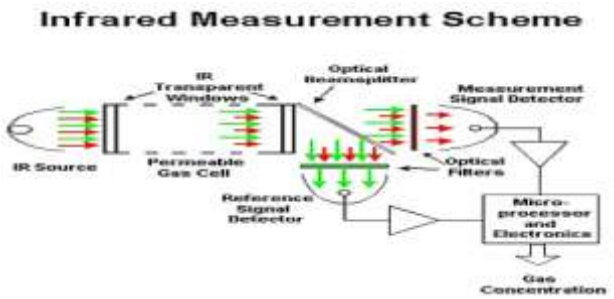


Figure 2: An Infrared Measurement Scheme [6]

Forms of Gas Detectors

Gas detectors comes in two major forms; portable and fixed.

i. Portable Gas Detectors

Portable gas detectors can be carried around or worn by individuals. They are most times powered by batteries and are for personal use. It works by sending signals to the user when in an area with toxic or combustible gas detection as well as the case of oxygen deficiency in confined spaces. In situations where the workers travel from country to country, they are useful in sampling gases, measuring the quality of the gas in the atmosphere.

ii. Fixed Gas Detectors

Fixed gas detector is designed to work at a stationary position, protecting many employees at one time unlike the latter. They are also able to detect several gases majorly used in the control room of an industry, process area of a plant or a sensitive area that might need meticulousness like a residential bedroom. A fixed gas detector has the advantage of being able to monitor an area constantly, even with the absence of an operator and make an appropriate emergency response.

Figure 3: An engineer with a portable detector [7]

Figure 4: An engineer working on a gas a fixed gas detector [7]

Gas Detectors in Different Fields

Gas detectors at the initial stages of development could only detect a single gas, but with improvement in

technology, the sensors are currently built with multiple components that allows for multiple gas detection. Gas Detectors must be part of the safety tools in any environment where gas is used as a supply of energy or is being given out as a by-product.

Over time, gas detectors have been incorporated into a wider range of systems; automobiles for engine emission control, Carbon dioxide sensors for proper ventilation check, gas monitors and alarms for CO are increasingly being needed in every field. They can also be used to detect the concentration of toxic gases in the petroleum, chemical, and pharmaceutical fields in real time.

CO_2 Hazard Monitoring

In poorly ventilated storage rooms and especially in very narrow containers or even aircraft, CO_2 concentrations can rise quickly. The danger for employees lies in particular in the fact that the gas is odourless and displaces oxygen. There is a risk of breathing difficulties, dizziness or disorientation. Concentrations of more than 10% can lead to coma. The occupational exposure limit (OEL) for CO_2 is 5000 ppm or 0,5% by volume. This limit describes the concentration in the air we breathe up to which carbon dioxide has no noticeable effect on the body. For this reason, it is crucial that CO_2 concentrations in the various areas of order picking and decanting are continuously monitored on a mobile or stationary basis [8]

Total Organic Carbon

Total Organic Carbon (TOC) Analysis is a measure of water quality. The TOC is the total amount of carbon that is found in an organic compound. It has become an important indication of approximate levels of organic contamination and can therefore be used as a suggestion of water quality. TOC is measured by oxidising the organic carbon to produce CO2, which can be quantified by measurement using a gas sensor. This process is essential because water purity is critical for a number of industries including pharmaceutical, manufacturing, power generation and water supplying. The presence of bacteria or other inorganic compounds can indicate filtration, storage or system failure.

TOC analysers help to measure the CO2 concentration produced by oxidising the organic carbon in a water sample, and allowing the TOC to be calculated [9]

Liquefied Petroleum Gas

Figure 5: Liquefied Petroleum Gas Tank [13]

Liquefied Petroleum Gas(LPG) also known as "cylinder gas" is a source of energy used for cooking, heating and lightning. This byproduct of natural gas, oil extraction and crude oil refining is a mixture of butane and propane in different proportions. LPG is falsely a toxic gas causing harm only when inhaled in large amount displacing oxygen. It is being delivered to the consumers in steel tanks. LPG in these steel cylinders can be in gaseous and liquid forms. When the cylinder gas is in use, the gaseous LPG at the top burns out and the liquid LPG at the bottom evaporates to replace the burnout gas. The liquid LPG is able to do this by drawing thermal energy from the environment.

LPG being a liquefied gas under pressure has a boiling point of approximately $0^{o}C$, and is gaseous in room temperature. This makes LPG superior to other fuels.

LPG cylinders and LPG-powered devices are safe when manufactured in compliance with strict safety standards and are equipped with safety components.*[*10] LPG can be commonly used as fuel for gas barbecue grills, gas cook tops, ovens and even vehicles as seen below.


Figure 6: LPG cooking demonstration in South [11]

Figure 7: LPG as a fuel for vehicle[12]

LPG Accidents

High Pressure Gas Safety Institute of Japan (KHK) in 2017classifies LPG accidents into four types:

i. **Leakage:** This is a situation where LPG got leaked but caused no damage like poisoning and fire. However, leakage of a very little amount of LPG could be ignored.
ii. **Leakage and explosion:** Gas leakage that leads to explosion, or explosion and fire.
iii. **Leakage and fire:** Leakage that led to fire excluding the ones mentioned in [ii].
iv. **Poisoning and asphyxiation:** This is caused by incomplete combustion, leakage of LPG, leakage of exhaust gas from exhaust pipes, etc at LPG consumption facilities.

The Internet of Things

Internet of Things is a recent technology that creates a global network of machines and devices that are capable of communicating and exchanging data with each other through the internet. Internet of Things is more intelligent than the Internet because it has the ability of creating information about the connected objects, analyze it and make decisions.

Key Features of IoT

1. Connectivity: This is the various levels of connection between devices, hardware, sensors, electronics and control systems. IoT networking mean networks are no longer exclusively tied to major providers. Networks can exist on a much smaller and cheaper scale while still being practical. IoT creates these small networks between its system devices.
2. AI: IoT makes everything smart enhancing every aspect of life with the power of data collection, AI algorithms and networks.
3. Small Devices: Devices have become smaller, cheaper, and more powerful over time. IoT exploits purpose-built small devices to deliver its precision, scalability and versatility.
4. Active Engagement: IoT uses a new paradigm for active content, product, or service engagement rather than the existing passive engagements [17]

Applications of IoT

1. Smart homes:

Though this is applied at different levels, the best is the one that blends intelligent utility systems and entertainment together.

2. Smart City:

Connected technology is being incorporated into infrastructural requirements and some vital concerns like waste management, water distribution, electricity management and more. This is made possible by giving internet access to devices making life easier for humans.

3. Wearables:

These devices serve a wide range of purposes from medical, wellness to fitness. Examples includes a smart wristwatch, smart bangle, smart shoes, smart jewelleries, smart clothings etc

4. Farming:

Tools are already being developed for Drip irrigation, understanding crop patterns, water distribution, drones for farm surveillance and more. These would assist the farmers in producing high yields.

Analysis of the Existing System

The existing system of gas leakage detection is the natural way which is by perception with the help of the sense organ for smell (the nose). The sense organ for smell is part of the chemosensory system. This comes from specialized sensory neurons, which are found in a small patch of tissue high inside the nose. These cells connect directly to the brain. Each olfactory neuron has one odor receptor. Microscopic molecules released by substances around us – whether cooking or grasses around us – stimulates these receptors. The brain identifies the smell once the neurons detect the molecules. The smells in an environment is much more than receptors, and any given molecule may stimulate a combination of receptors, creating a unique representation in the brain. These representations are stored in the brain as a particular smell, for example, a LP gas [14]

Problems of the Existing System

The key issues with the existing system are as follows:

1. Detects only gases with odours:

The existing system has no ability to detect the presence of odorless gas. This can be dangerous because some gases are odourless yet toxic. They are harmful to the body and eventually lead to death. Poisonous gases like CO could suffocate humans slowly and silently in an environment with no one realizing.

It is important to also note again that LPG and LNG are only detected by the existing system because of the harmless chemicals added to it during production to make them smell. Otherwise, LPG and LNG are naturally odourless [10].

2. Cannot be detected by people with smell disorders

There are high chances for people who have smell disorders like hyposomia (a reduced ability in detecting gas), anosmia (inability to detect gas at all), catarrh, COVID-19 or other health challenge to not be able to sense a gas leakage.

3. Inability to measure concentration of gas in the atmosphere

The existing system does not have the ability to tell how much gas is in the atmosphere.

4. No proper notification when there is a leakage

The existing system does not have any effective means of making others aware of a gas leakage. Its method of notification is verbal which is slow and inefficient indirectly leading to more damages.

6. METHODOLOGY

System Design

System Design is the phase that bridges the gap between problem domain and the existing system in a manageable way. This phase focuses on the solution domain, that is, how to implement the IoT- based Gas Leakage Detection System. The complex activities of system development are divided into several smaller sub-activities, which coordinate with each other to achieve the main objective of system development.

The proposed system design has the ability to detect the presence of all gases that are constituents of LPG, carbon monoxide, smoke and some other gases. The system also has the ability to sound an alarm when there is a leakage, send an SMS to a predefined no, and also measure the amount of gas that was leaked. The system makes varieties of sound, the sound at each point is determined by the gas concentration of the atmosphere.

A thorough listing of all requirements needed to build the proposed system is made below; these include the hardware components, software and related requirements

A list of the major hardware components needed for this research work:

1. Arduino Uno
2. MQ5 sensor
3. GSM Module
4. Buzzer
5. 16 x 2 LCD Display
6. Jumper wires

Arduino Uno

An Arduino Uno is an open-source electronic prototyping platform enabling users to create interactive electronic objects. This microcontroller board was founded by Massimo Banzi and David Cuartielles in 2005. It uses easy to learn libraries and its IDE is available for Windows, MAC and Linux. The major components of the Arduino Uno board are as follows: USB connector, power port, microcontroller, analog input pins, digital pins, reset switch, crystal oscillator and the USB interface chip.

The microcontroller which is the major component of the Arduino UNO works by interpreting data it receives from the I/O peripherals using its central processor. A microcontroller generally is used for controlling a singular function in any device it is embedded on. The elements of every microcontroller includes a:

1. **Processor(CPU):** the brain of the devices. It processes and responds to various instructions that direct the microcontroller's function.
2. **Memory:** The program memory is non-volatile, storing long-term information about the instructions that the CPU carries out. The data memory is very volatile(temporal), the information it carries is only maintained for as long as the microcontroller is connected to a power source.
3. **I/O peripherals:** This is the interface for the processor to the outside world. The input ports receive information and send it to the processor in the form of binary data. The processor receives that data and sends the necessary instructions to the output devices that execute tasks externally to the microcontroller.

Other elements of a microcontroller includes: Digital to Analogue Converter, Analogue to Digital Converter, System Bus, Serial port. The memory types are flash memory, Erasable Programmable Read-Only Memory (EPROM) and Electrically Erasable Programmable Read-Only Memory(EEPROM)

Figure 8: An Arduino Uno Board

Features: The Arduino board above operates on 5V, and it is recommended that input voltage be from 7v-12v. The DC current for 3.3v pin is 50Ma. The digital input/output pins are 14 and the analog I/O pins are 6. The SRAM memory is 2kb and the flash memory is 32kb. The EEPROM storage is 1KB and the clock speed is 16MHz

MQ5 Sensor

Sensors are mainly manufactured by Chinese and have its origin in China. MQ sensors are specially designed to have sensitivity to detect some gas like LPG, CO, propane etc. In Chinese, 'sensitive' means 'Mĭngăn' and 'Gas to' means 'Qĭ lai'. So, more or less, MQ stands for the sensors having sensitivity towards (or to) gas *[15]* The MQ-5 sensor has a sensitive filament made of SnO_2, when a combustible gas such as LPG is introduced, the filament's conductivity rises, and the amount of change in its conductance/resistance can be used to indicate the equivalent gas concentration. MQ5 sensor module is very useful for gas leak detectors, other industrial combustible gas detectors and gas monitoring devices to be used at home.

The MQ5 has two output possibilities – an analog out (A0) and a Digital out (D0). The analog out is used to detect Gas leakage and measure volume of gas leakage(by doing proper calculation of the sensor output inside program) in specific units(ppm). The digital out can be used to detect Gas leakage and hence trigger an alert system(a sound alarm, sms activation etc). The digital out gives only two possible outputs – high and low.

Figure **9: An MQ5 sensor**

The MQ5 sensor has a high quality dual panel design and a dual signal output(analog, TTL level). It has low sensitivity to alcohol and smoke and gets a higher voltage when concentration is high. It is highly sensitive to liquefied petroleum gas, natural gas and town gas. It is fire resistant and has also proven to be long lasting.

Table 1: Specifications of an MQ5 sensor

Items	Parameter name	Min	Type	Max	Unit
VCC	Working Voltage	4.9	5	5.1	V
PH	Heating consumption	0.5	-	800	mW
RL	Load resistance	Can adjust			
RH	Heater resistance	-	31	-	Ω
Rs	Sensing Resistance	3	-	30	K Ω

GSM Module

A GSM modem or GSM module is a hardware device that uses GSM mobile telephone technology to enable communication between a microcontroller and the GSM/GPSR Network. GSM Modulles are used in the following applications: Cellular Communication, Robotics, Mobile Phone Accessories, Servers, Computer Peripherals, Automobile, USB Dongles.

Figure 10: A SIM800L GSM Module *(360hub, n.d).*

Description/Features: Sim800L connects easily with arduino and other MCUs because of its inbuilt regulator. It is approximately 100% of high quality, working between 3.7 to 5v. It supports Quad-Band 850 / 900/ 1800 / 1900 MHz GSM networks in all countries across the world. The TTL serial interface is compatible with 3.3V and 5V MCU Microcontrollers, and also compatible with arduino.

16 x 2 LCD Module

LCD stands for Liquid Crystal Display. An LCD is an electronic display module that uses liquid crystal to produce a visible image. A 16x2 LCD display is a very basic module and is commonly used in various devices and circuits. A 16 x 2 LCD means it can display 16 characters per line and there are 2 such lines. 16 x 2 LCD are commonly used in DIYs and circuits. In this project, it would be used to display the status of the gas detector.

Figure 11: 16 x 2 LCD *(360hub, n.d)*

A 16 x 2 LCD has a display colour of either blue or green, a blue one was used in this project as seen above. A 16 x 2 LCD uses a standard 16 pins interface, the symbols and functions are specified as seen Table 2:

Table 2: 16 x 2 LCD has a display

Pin	Symbol	Function
1	V_{SS}	Ground (0 V)
2	V_{dd}	5V Logic supply voltage
3	V_0	Contrast adjustment
4	RS	H/L register select signal
5	R/W	H/L read/write signal
6	E	H/L enable signal
7-14	DB0-DB7	H/L data bus for 4- or 8-bit mode
15	A(LED+)	Backlight anode
16	K(LED-)	Backlight cathode

Buzzer

The buzzer is a sounding device that is able to convert audio signals into music, siren buzzer, alarm, electric bell and other sound signals. It is usually powered by DC voltage and is widely used in alarms, computers, printers, and other electronic products as sound devices.

Figure 12: A buzzer *(360hub, n.d).*

Buzzer Features and Specifications

The buzzer has an Operating Voltage of 4-8V DC and its current is rated below 30mA. The sound type is a continuous beep. It is Breadboard and Perfboard friendly. It is small in size as seen above and has a resonant frequency approximating 2300Hz.

Table 3: Buzzer Pin Configuration

Pin Number	Pin Name	Description
1.	Positive	Identified by (+) symbol or longer terminal lead.
2.	Negative	Identified by short terminal lead. Typically connected to the ground of the circuit

Jumper Wires

A jumper connector is a tiny metal connector or group of them that is used to close or open part of an electrical circuit. A jumper is made of materials that conduct electricity, and is sheathed in a nonconductive plastic block to prevent accidental circuit shorts. Jumpers are like on/off switches. They may be removed or added to enable component performance options. The jumper's main advantage is its one-time configuration, which makes it less vulnerable to corruption or power failure than firmware. Jumper altering requires that settings be physically changed.

Jumpers are of two types; male-male(m/m) and male-female(f/m). The difference between each is in the end point of the wire. (M/M) ends have a pin protruding and can plug into things, while (F/M) ends do not and are used to plug things into. Male-to-male jumper wires are the most commonly used.

Figure 13: Male-Female JumperWires

Figure 14: Male-Male Jumper Wires

Other components: USB cable, 9v connector, I2C LCD interface, LED bulb, 9V battery (or power bank) for power supply, SIM card, and USB cable.

2 Arduino IDE

The Arduino Integrated Development Environment (IDE) is a cross-platform application (for Windows, macOS, Linux) that is written in functions from C and C++. It is used to write and upload programs to Arduino compatible boards, but also, with the help of third-party cores, other

vendor development boards. The Arduino IDE contains a text editor for writing code, a message area, a text console, a toolbar with buttons for common functions and a series of menus. It connects to the Arduino and Genuino hardware to upload programs and communicate with them.

The Proposed System is able to solve the following problems when there is a leakage:

1. Detection of a gas leakage whether poisonous, non-poisonous, odourless or colourless with the help of MQ5 sensor.
2. The buzzer in the proposed system notifies everyone in the environment of a leakage synchronously when there is one.
3. The proposed system is able to measure and display the concentration of the gas in the atmosphere.
4. In several cases, the head of a household or business is not aware of a leakage. With the help of GSM Module, an SMS is sent to a predefined number (the proprietor's) immediately there is a leakage.

Circuit Design

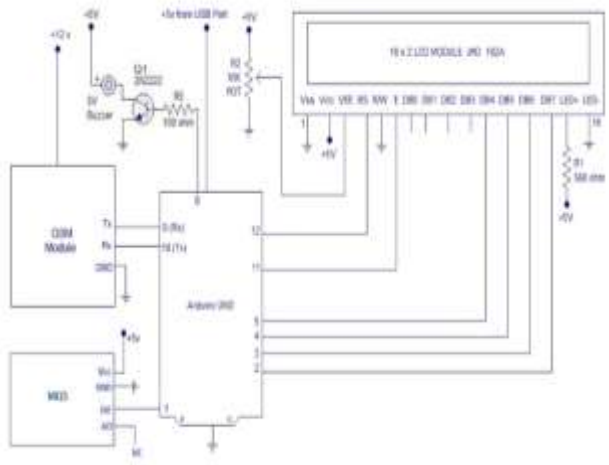
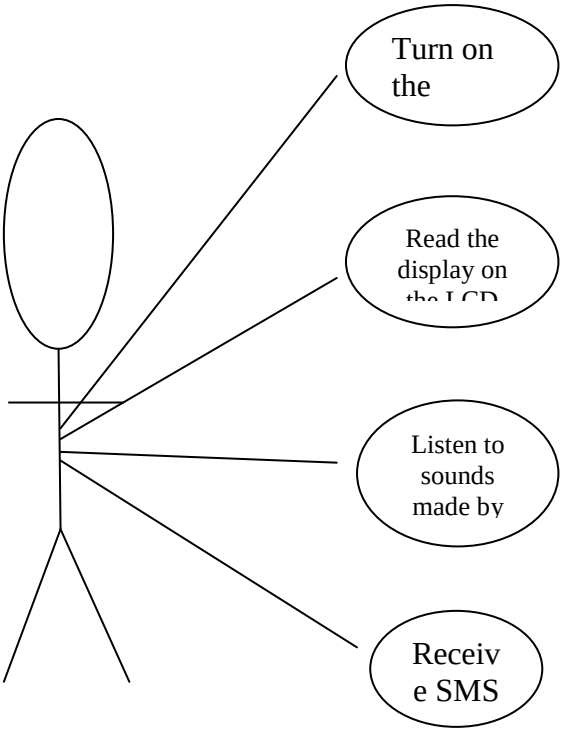


Figure: 15: Circuit Diagram of the IoT Based Gas Detection System

USE Case Diagram

The USE Case diagram below represents a user's interaction with the working system of the IoT based Gas Leakage Detection System



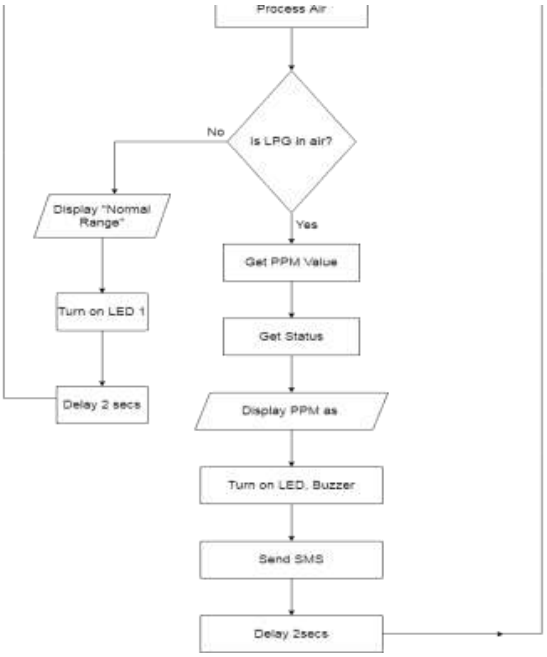


Figure 17: System Flowchart Diagram

Interconnection of the Hardware Components

As already mentioned, the hardware components used for this work include: Arduino board, MQ5 sensor, buzzer, power supply, 16 x 2 LCD and jumper wires.

The interconnection of the components is first done using a breadboard to test the circuits before soldering. A breadboard is a solderless construction base used for developing an electronic circuits and wiring. As would be seen in the next diagram, the MQ5 sensor, the GSM module, the buzzer, LCD display are being connected to breadboard and the Arduino. Power is being supplied using a 9v battery

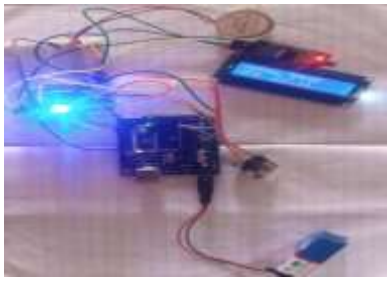


Figure 18: Interconnections Hardware Components

Precautions during Construction

1. Read the datasheet for every component noting the features, specification, technical guidelines and applications of each of them.
2. While using 5v and GND pins to power up the Arduino, ensure the 5V input is stable and steady.
3. Sequel to [2], double-check the polarity because if mixed up can potentially damage the Arduino board.
4. Soldering is done in a well ventilated place because the fumes released are very uncomfortable.
5. Solder contains lead which is poisonous so wash the hands should be washed thoroughly.
6. Always return the soldering iron to its stand after soldering.

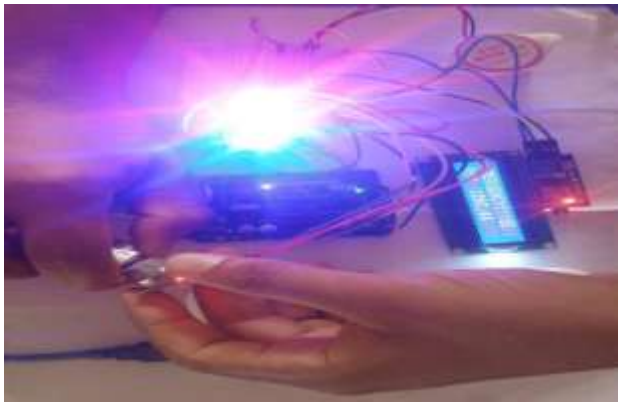


Figure 19: When the gas leakage is at Emergency

Figure 20: Display of SMS for Leakages

Packaging:

The components of the work were transferred from the breadboard to a veroboard where it was soldered. A pacterex box was used for the packaging, openings were made for the for the placement of the LCD, buzzer, 5 LED bulbs and GSM antena

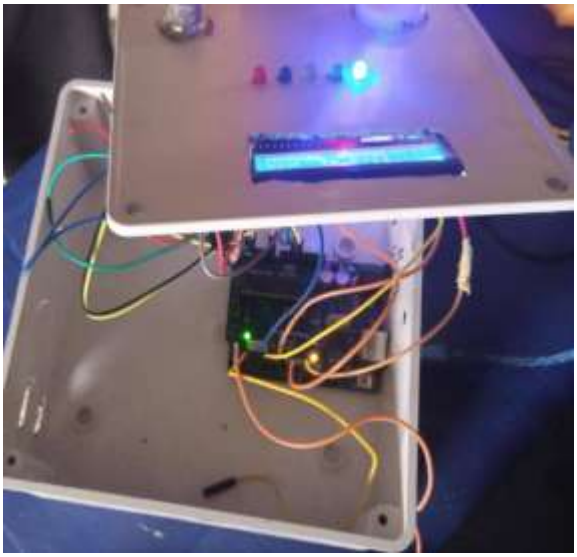


Figure 21: The packaged work I

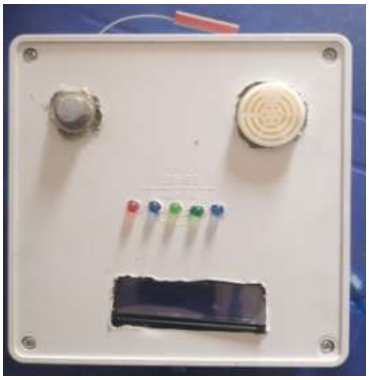


Figure 22: The packaged work II

4 Discussion

LPG releases 81% of carbon dioxide when compared to oil and 70% when compared to coal making it the more friendly to the environment than other sources of energy. LPG burns cleaner than most energy sources because it does not contain oil. LPG is also easy to transport and has a higher heating value *(Advantages of using LPG over other fuels, 2014).*

In the course of this work, an LP gas leakage and smoke detector was designed and built. This system is able to detect the presence of an LP gas and smoke with the help of the MQ5 sensor, measure and display the amount of gas in the atmosphere, sound an alarm and send a notification to a predefined phone number. We hope that over time, with improvement in technology, that gas leakage and smoke detectors would be integrated into low cost devices and into everyday consumer electronics such as mobile phones and wearable's. Gas sensors, and also every other kind of sensors will play a key role in IoT development and be used extensively in smart homes and smart city programmes. Some of the Challenges in Using LPG are:

i. When inhaled in large amounts LPG can cause suffocation since it is heavier than air.
ii. LPG is highly flammable, making it very risky to use.
iii. LPG as a vehicle fuel, does not provide power to the vehicle in mountains or rough terrains.
iv. LPG has to be supplied in a heavy steel cylinder for safety purposes.

The IoT-based gas detection system is able to solve in parts the problems stated in [i] and [ii] above about the challenges of LPG. The system having a gas sensor and buzzer ensures that there is always a sound alarm when LPG is in the atmosphere. It also prevents or reduces the damages that could have occurred in case of an explosion, this is by exclusively notifying necessary authorities once there is a leakage so immediate arrangements would be made to extinguish fire

5. Conclusion

According to [16], the President, Nigeria LPG Association,Nuhu Yakubu, in an interview revealed that the LPG industry in Nigeria had grown over 1000 percent between 2007 and 2017. The adoption rate is still considered to be low considering that 90% of the homes in Nigeria still lack LPG cylinder, that is, just three million out of the 30 million homes use gas cylinders, exposing the higher percentage to dirty fuel options such as fire woods and kerosenes, causing a high mortality rate in the country. Neighbouring African countries such as Ghana has about 70% of the cars running on LPG but sadly Nigeria is still adopting LPG for cooking. Nigeria has enormous reserve of this most diversified gas in the world and should be driving this innovation but that is not the reality. LPG infrastructure is expensive, characterized by the high cost of entry for consumers, and also the cost of LPG due to multiple devaluation of naira, he said ,was the chief reason why it is not being used in remote regions of the country. "LPG Diversification: Expanding LPG Frontiers in Nigeria.", the theme for the 8th LPG annual conference and exhibition was poised to make awareness thereby closing up the huge gap in LPG usage especially in the northern part of the country.

Reference

[1] Generator source (2019) *The Future of Power.* https://www.generatorsource.com/The_Future_of_Power.aspx

[2] Cory(2019, March 13) *A Short History of Gas LeakDetection.*DODTechnologies,Inc.https://dodtec.com/a-short-history-of-gas-leak-detection/

[3] Phol Dhanya(2021) *PID sensor for detecting fast volatile substances(VOCs)*Thai Safetywiki. https://thai-safetywiki.com/electrochemical-sensor/?hcb=1

[4] Dave Wagner(2020, January 29) *Safety in wastewater plants: Why you need portable gas detectors.*IndustrialScientific.https://www.indsci.com/en/the-monitor-blog/safety-in-wastewater-plants-why-you-need-portable-gas- detectors/?hcb=1

[5] Kinsley Jeremiah(2019) *FG projects rise in LPG consumption as price soars.* The Guardian. Energy. https://guardian.ng/energy/fg-projects- rise-in-lpg-consumption-as-price-soars/

[6] EnggCyclopedia(2021) *How infrared gas detectors work.* Safety/Loss Prevention https://www.enggcyclopedia.com/2011/11/infrared-gas-detectors/

[7] Dave Wagner(2020, January 29) *Safety in wastewater plants: Why you need portable gas detectors.* Industrial Scientific. https://www.indsci.com/en/the- monitor-blog/safety-in-wastewater-plants-why-you-need-portable-gas- detectors/?hcb=1

[8] Dräger(2018) *CO_2 Monitoring – Vaccine Transportation & Storage.* Draeger. https://www.draeger.com/en_me/Safety/CO2-Monitoring-Vaccine- Transportation-Storage

[9]Edinburg Sensors(2020). *An Overview of the Industries Requiring Gas Sensing.* News and Events. Retrieved April 05, 2021 from https://edinburghsensors.com/news-and-events/an-overview-of-the- industries-requiring-gas-sensing/

[10] Aygas(2019). *Technical Features of LPG.* AYGAS https://www.aygaz.com.tr/en/cylindergas/technical-features-of-lpg

[11] The High Pressure Gas Safety Institute of Japan (KHK)(2017) *Annual Report on Liquefied Petroleum Gas (LPG) Related Accidents* https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.khk.or.jp/Portals/0/khk/info/2018/LPG%2520related%2520accidents%2520201 7.pdf&ved=2ahUKEwi3h8KUjJLyAhWWgf0HHUlfAYcQFjAAegQIBBA C&usg=AOvVaw2YJo6Ep8QJA5lSG8SVPPtY

[12] Otomobil gazetesi(2020, May 27) *Misunderstood urban legends about LPG.* NEWS http://www.otomobilgazetesi.com/mobile/?page=haber&ID=8798

[13] San Pedrosun(2019, October 5) *LPG important issue prompts meeting with PM andButane Gas Company representative* https://www.sanpedrosun.com/government/2019/10/05/lpg-importation- issue-prompts-meeting-with-pm-and-butane-gas-company- representatives/?hcb=1

[14] National Institute on Deafness and Other Communication Disorders. (2017, May 12) *Smell Disorders. How does your sense of smell work?* https://www.nidcd.nih.gov/health/smell-disorders

[15] Choudhary Abhishek(2018) *What does MQ stand for in MQ gas sensors?* Quora https://www.quora.com/What-does-MQ-stand-for-in-MQ-gas-sensors

[16] Stanley Opara(2018, November 21) *90 per cent of Nigerian homes lack LPG cylinders, reveals NLPGA.* Energy. https://m.guardian.ng/energy/90-per cent-of-nigerian-homes-lack-lpg-cylinders-reveals-nlpga/

[17] Tutorials Point(2016) *Internet of Things.*https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.tutorialspoint.com/internet_of_things/internet_of_things_tutorial.pdf&ved=2ahUKEwjK26CG6JjyAhWoBWMBHauoAV4QFnoECAUQAg&usg=AOvVaw2HgxSpWYd5NFv7rXZrqH3C

AN INNOVATIVE FRAMWORK FOR INTEGRATED LOAN MANAGEMENT WITH QR CODE ENHANCEMENT

Anthony I. Otuonye[1], Perpetual N. Ibe[2], Irene F. Eze[3]

[1]Department of Information Technology,
Federal University of Technology Owerri, Nigeria.

[2]Department of Computer Science Technology,
Imo State Polytechnic Umuagwo Owerri.

[3]Department of Computer Science Education,
Federal Collage of Education (Technical), Umunze

Corresponding Author's email: anthony.otuonye@futo.edu.ng

Abstract

This research paper aims at design of an innovative framework scalable and integrated loan management with Quick Response Code enhancement, which will guarantee easy and better-secured loan validation and processing in microfinance banks and other lending institutions. As the number of microfinance bank customers in need of personal loans rises on a daily basis, especially in the post COVID-19 era, management is faced with the complex job of loan application verification in order to correctly determine eligibility for a loan. There is the challenge of coping with fraudulent customers who make false claims with their loan application documentations, sometimes seeking to access multiple loans from more than one microfinance banks, and using single collateral security. There is need for a central regulatory agency that links up major lending institutions in a collaborative effort to forestall incidences of multiple loan access using single loan security. Our new model will also forestall activities of some fraudulent bank officials who go as far as granting credit facilities to their friends and family members without followimg due process, and using their privileged positions to obtain unsecured loans for themselves and their relations, sometimes in excess of bank's statutory lending limits, and in total violation of the provisions of policy of microfinance banking sector. Our new model was designed using the Rapid Application Development (RAD) methodology following its high involvement and concentration on user's viewpoint as well as its ability to change system design on the go based on user demands. Timely implementation of this framework will forstall incidnces of "insider abuse" and ensures that loan approvals are done according to internal and external regulations.

Keywords: Loan Scheme, Loan security, Microficance bank, Management, QR Code,

1.0 INTRODUCTION

1.1. General Overview

The number of people seeking for bank loans these days, especially in Nigeria has risen drastically. Salary Earners and Business Executives alike, have all been caught in the mad rush for loan, either to enable them solve some pressing personal/domestic needs, or to serve as initial capital for new business initiatives, or even to support an existing business venture. Sometime these Loan Seekers are willing even to cut corners or deceive the bank management in one way or the other in order just to have their way. The desperation of some of these Loan Seekers can easily be traceable to the poor living condition in Nigeria, and in most third world countries of the

world. The poor economic situation in these countries seems to have affected both salary earners and business entrepreneurs who now see bank loan procurement as the only way to meet up with their individual and corporate financial obligations. The COVID-19 pandemic has also fuelled the worsening economy and the living condition of most citizens. Unemployment rate has risen to its highest ebb. The desperation to survive in the land has given rise to many more individuals seeking loan opportunities from lenders, both banking and non-banking institutions. Some have even argued that without such loans, it will be quite difficult to achieve most of their economic set goals and aspirations.

For Salary Earners, this can be attributed to poor remuneration, inconsistency in salary payments, inflations, as well as consistent and indiscriminate deductions in salaries, especially, those in the categories of government employees. Government workers now resort to formation of Cooperative Societies as a strong and dependable alternative to Workers' Welfare. Again, apart from assisting to meeting mutual needs of members, there is a strong belief that people can actually achieve more as a group or as members of a cooperative society. Cooporative societies have helped members in collective Estate Acquisition, accessing of soft loans from commercial banks as well as from Microfinance banks.

With an organized system, accessing of loans should be a simple and straightforward process, unlike what we see today, especially in the microfinance banking sector. The process of loan application processing should be as easy as bank management checking her client's eligibility for a loan, then approving or denying the loan request. Once approved, the customer should easily receive the funds without much delay.

This is, however, not the case, especially when it comes to individual loan application processing by management of microfinance banks. These bankers often complain that some of the individual loan seekers are fraudulent deceivers and that there is need to take time to thoroughly verify their claims.

In some cases, for instances loan applicants can go as far as seeking to access multiple loans from more than one microfinance banks at the same time, using a single collateral security. Though this is against microfinance banking policy, it takes much effort and scrutiny to discover these abnormalities and sharp practices on the part of bank customers. This is obviously one the reasons for delay in loan application processing and procurement. Despite their resolve to forestall or identify these sharp practices no matter how long it takes, microfinance banks, most times still fall short of their responsibilities. There is basically no system in place to assist them in checkmating the excesses of fraudulent customers in this regard.

Apart from fraudulent bank customers, we also have fraudulent bank staff and corrupt managers to contend with. Activities of these bankers have hampered the performance of microfinance banks. This is called "insider abuse" [18].

According to [18], "insider abuse" in microfinance bank is made manifest in the granting of credit facilities. Some researchers has pointed out that a good number of Microfinance bank directors misuse their privileged positions to obtain unsecured loans for themselves and their relations, which, in some cases are in excess of their banks' statutory lending limits, and in violation of the provisions of policy of microfinance banks. Furthermore they have been found to approve loans for their friends and relatives without proper documentation, and in most cases these loans turn out to increase cases of non-performing credit. These fraudulent activities continue to go unabated due to poor system of loan management, and lack of proper use of the digitized system.

In traditional lending systems, particularly in larger organizations, the process of loan application verification, processing, and monitoring, is often chaotic, and without an efficient verification system in place, the process of loan application processing and monitoring will continue to drag. If this is not properly addressed, instead of becoming a smooth experience, loan application processing will continue to be tedious and stressful on the part of both the lender and the borrower.

Globally, several other sectors, especially the banking sectors are testing different implementations of the Quick Response (QR) code technology which seek to provide simple and unified payment solution as well as other solutions. QR Code is the trademark for a type of matrix-based barcode (or two-dimensional bar code) first designed for the automotive industry in Japan [13]. Its fast readability and greater storage capacity has made the QR code a popular technology alternative, which is now being deployed outside the automotive industry.
Research has shown that QR codes can easily be read by any smartphone camera, point of sale (POS) terminals, or other devices. Its use ranges from advertising and promotions to merchandise tracking and coupons, to accessing media on the Internet, downloading offers, locating product information, and many more. Several implementations of QR codes also exist in other fields. [12], for example, made use of the QR technology to secure and transmit patient-sensitive information from one level of the system of health care delivery to another.

Getting a digital platform that incorporates the QR code technology is quite crucial for the survival of the lending business sector in this competitive market, especially with the increasing number of loan seekers across the country. The obvious challenges faced by both management and customers of microfinance banks in Nigeria leave very little to be desired, and as such, an effective Loan Management System is needed to streamline the process.

Despite the clear advantages of Information Technology over the manual process and its unlimited possibilities, very few microfinance banks have embraced the digital technology as a viable tool for accurate record keeping and verification. Similarly, it is no longer in doubt that the world is constantly undergoing an Information and Communication Technology (ICT) revolution. Information Technology is being deployed for more efficient goal attainment, flexibility, accurate record management, and transparency in transaction processing.

1.2. Aim and Objectives of Study

In this research paper, we aim at a design of new architectural framework of modular, scalable and integrated Loan Management process with Quick Response Code Enhancement for easey and better-secured loan processing in the microfinance banking sector and other lending institutions. Specific objectives include, to:

1. Ascertain all User Requirements for efficient loan management based on customer expectations and management regulations.
2. Design a QR code-enabled digital platform that inhibits multiple loan access on single collateral security to reduce credit risk.
3. Propose a unified QR-code generation function for faster and better-secured loan validation and monitoring process, and
4. Make recommendations for effective implementation of the new framework for more efficient goal attainment, flexibility, accurate record management, and transparency in transaction processing.

2.0. LITERATURE REVIEW

2.1. The Loan Concept

A loan is a type of debt. Like all debt instruments, a loan entails the redistribution of financial assets over time, between the lender and the borrower. In a loan, the borrower initially receives or borrows an amount of money, called the principal, from the lender, and is obligated to pay back or repay an equal amount of money to the lender at a later time. Typically, the money is paid back in regular installments, or partial repayments; in an annuity, each installment is the same amount. The loan is generally provided at a cost, referred to as interest on the debt, which provides an incentive for the lender to engage in the loan. In a legal loan, each of these obligations and restrictions is enforced by contract, which can also place the borrower under additional restrictions known as loan covenants. Even though we are focusing on monetary loans in this research project, any material object, however, might be lent. Acting as a provider of loans is one of the principal tasks for financial institutions such as banks. For other institutions, issuing of debt contracts such as bonds might be a typical source of funding.

In finance, a loan is the lending of money from one individual, organization or entity to another individual, organization or entity. It could be provided by an individual, organization or corporate body at an interest. Many borrowers have found the loans to be quite helpful in the pursuit of their individual goals either in business or other personal engagements. On the other hand, lenders also benefit from loans because of the interest attached to each loan, and as borrowers increase the more their business grows.

2.2. Understanding Loan Scheme Management Systems

Lenders make use of Loan Management Systems to streamline their process of checking for client's eligibility, loan approval or denial, fund disbursement, repayment monitoring, etc. In traditional lending systems, especially in large organization, this process is often chaotic. Sometimes it becomes quite cumbersome to keep track of important records owing to the fact that different customers have different terms and payment dates. The chaotic process can become even more complex to handle as the customer base increases.

A Loan Management System therefore is a software system that helps to automate the entire loan process and life cycle. Depending on User Requirements, such a system can help either in part or in whole, such as processing customer information, creating new loans, manage interest rates, repayment monitoring, debt collection and recovery processes, etc. A good Loan Management System can equally provide management with accurate statements or reports at intervals.

Normally, an automated loan management/lending system will do better than legacy systems in many ways. Being a digitized system, it can cater for newer generation of customers, and reduce manual errors and risks.

2.3. Quick Response (QR) Code Technology

According to [13], Quick Response (QR) Code (is a matrix-type (or two dimensional) barcode trade mark, usually optical machine readable labels attached to items that record information related to the item. They have the ability to store information both vertically and horizontally as opposed to conventional one-dimensional bar codes that can store information only in the horizontal manner [19]. QR codes often contain data for a locator, identifier, or tracker that points to a website, application, URL, text, or other types of data, which can be easily read by the cameras of mobile devices [17]. The QR code system is one of the most used types of two dimensional codes. The main reasons for its popularity include its fast readability and its greater storage capacity when compared to standard UPC barcodes. QR codes are now used for commercial tracking applications and convenience-oriented applications aimed at mobile phone users. Notable areas of QR deployment applications include product tracking, item identification, time tracking, document management, and general marketing.

A QR code is composed of black modules (Square dots) arranged in a square grid on a white background. It is capable of encoding four standardized types ("modes") of data (numeric, alphanumeric, byte / binary, Kanji) or, through supported extensions, virtually any type of data. This can be read by any imaging device such as a camera, and processed using Reed–Solomon error correction until the image can be interpreted appropriately as required. The data required is extracted from the vertical and horizontal image pattern of the QR code. The QR code can be used to display text to the user, images, videos, web-links to the user's device [10].

2.3.1. Functional Elements of a QR Code

According to Denso [9], the following are the functional elements of a QR code technology: version information, format information, data and error correction keys, finder pattern, separator, required pattern, timing pattern, position pattern, alignment pattern, and the quiet zone. Figure 1 illustrates the functional elements of the QR code technology.

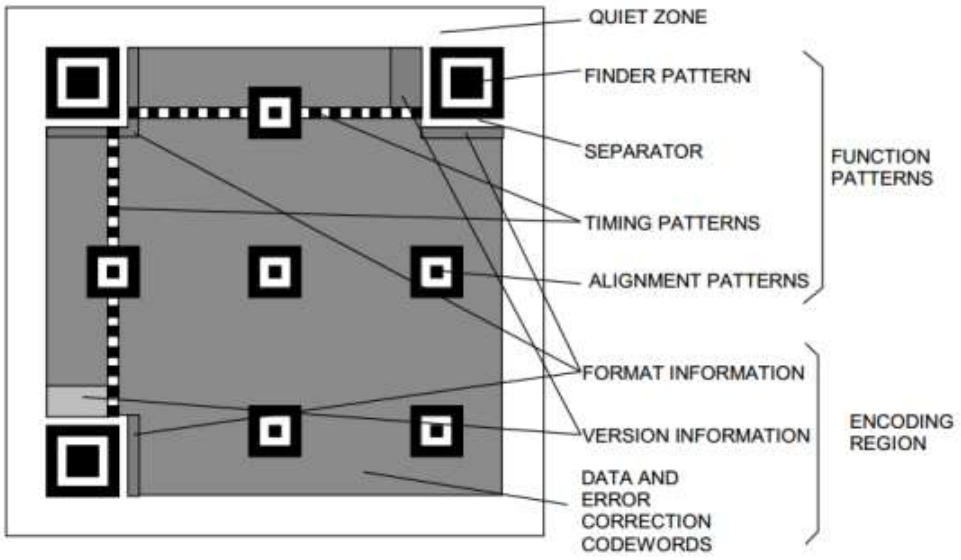


Figure 1: Functional Elements of a QR Code (Source: ISO/IEC, 2005)

1. **Version Information:** This element specifies the QR code version that is being used. There are currently 40 different versions of the QR code technology. However, version 1to **7** are the most widely used, especially for marketing purposes [9].
2. **Format Information:** This component contains information about the error tolerance and the data mask pattern, and it is this information that usually makes it easy for the code to be scanned.
3. **Data and Error Correction Keys:** This is where the actual data is stored.
4. **Finder Pattern:** This pattern identifies the symbol and decides the correct orientation it will take. It usually contains three collective structures which are positioned in the QR codes three angles.
5. **Separator:** This envelops the finder pattern and promoteseasy identification of the patterns.
6. **Required Pattern:** They are of three categories.
7. **Position Pattern –** These patterns or markings indicatethe QR code print direction or orientation.
8. **Alignment Pattern –** Additional element that aids orientation for especially large codes.
9. **Timing Pattern –** These lines assist the QR code scanner to determine how large the data matrix is.
10. **Quite Zone:** Helps distinguish the scanning zone from its surroundings.

2.4. Review Summary

Table 1 presents a summary of some of the major literature that was reviewed in the course of this research, which include names of authors and the year of their publication, work tittle, and major findings.

Table 1 Summary of literature review

S/N	Authors/Year	Tittle of Research Work	Findings	Comment
1.	Melisa Santos (2019)	Development of Automated Loan Lending and Mangament System.	He developed a Loan Lending Record Management System using Akwa Savings and Loans Ikot Ekpene as a case study. The system mostly aided registration of loan records by implementing an effective database system that facilitates easy recording and retrieval of loan management information.	The scope of this work is limited only to record retrieval. There is much more to the lending business than just record keeping. The work can therefore be improved upon.
2.	Kanu Clementina (2015)	Microfinance Bank Operations in Nigeria, Constraints and Suggested Solutions: An Evaluation.	Descriptive Analysis from SPSS was used to determine the most challenging factors in the operation of microfinance banks in Nigeria. Their suggestion was that the identified constraints to microfinance banks should be eliminated to ensure sustainability of FMBs in the country.	Their suggestion underscores the need for such systems as the one currently under study which seeks to harness the power of QR code technology to improve the system.
3.	Sonawane Shamal et al (2014)	Secure Authentication for Online Banking Using QR Code	The researchers were able to develop a fairly secure authentication for an online banking system using a two factor authentication that combined password and a camera equipped with mobile phone acting as an authentication token.	The system is a better secure method of online banking transaction than earlier applications. However, this system could not handle some of the excesses of fraudulent bank staff.
4.	Mbam and Kinglsey (2013)	Enhancing Cooperative Loan Scheme through Automated Loan management System	The Researchers were able to develop an Auto-LMS using the rich potentials of VB.NET and the SQL Server 2005. It is an innovative system that was able to manage the short-term loan scheme of a named cooperative society and keeps track of cash inflow and outflow. The system was able to handle the difficult task of maintaining all entries of users account, search records of activities, and handle loan deduction errors.	The system was neither able to enthrone the desired flexibility, nor ensure that credibility and integrity of data is fully maintained.
5.	Olorunlomerue et al (2017)	Web based Centralized Cooperative Information Management System	The Researchers were able to develop a web based centralized cooperative information management system aimed at enhancing the operations of cooperative societies, where they can log on using unique passwords to register their cooperative society as well as provide	The system lack scalability, and the ability to locate the cooperative societies using

			details of their financial statements. The system was meant to assist government gather information on cooperative societies for planning and development of commerce.	Google-enabled maps.

3.0. METHODOLOGY AND REQUIREMENT ANALYSIS

3.1. The Rapid Application Development (RAD).

The Rapid Application Development (RAD) is the methodology of choice for this project following its high involvement and concentration on user's viewpoint as well as its ability to change system design on the go based on user demands The figure 2 shows the different stages involved in RAD.

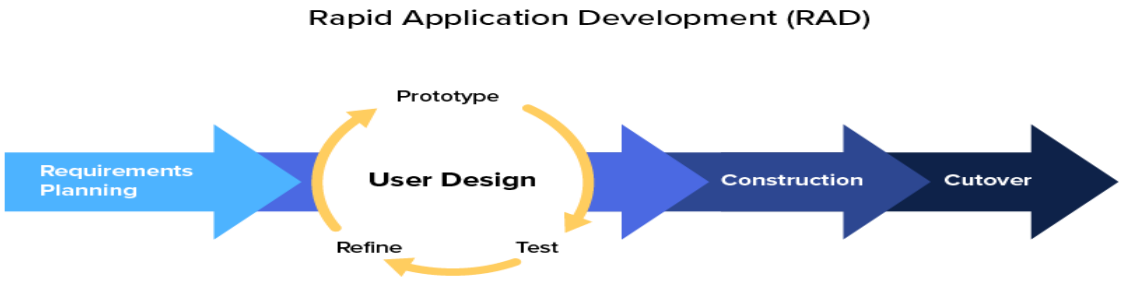


Figure 2: Stages involved in RAD (source: kissflow.com).

3.2. The RSA Algorithm

The QR code technology being proposed for use in this study will be based on the RSA cryptosystem. The RSA Algorithm is a cryptosystem that follows Asymmetric key cryptography used for bulk encryption and decryption. This algorithm basically involves four steps, which include:

a. Key generation, (b). Key distribution, (c.). Encryption and (d). Decryption

Following the four steps listed above, the RSA algorithm makes use of the following procedure to generate public and private keys:

i. Select two large prime numbers, p and q.
ii. Multiply these numbers to find n = p x q, where n is called the modulus for encryption and decryption.
iii. If n = p x q, then the public key is <e, n>, where e is a number less than n such than n is relatively prime to (p – 1) x (q – 1).

3.3 Requirement Planning

3.3.1. Data Collection Method.

There are several options on how data may be collected for research efforts of this nature. These options can however be grouped into the following methods:

- Interviews (which can be unstructured, semi-structured or structured)
- Focus groups and group interviews
- Surveys/questionnaire (online or physical surveys),
- Observations,
- Documents and records,
- Case studies, etc.

The choice of data collection method will always depend on the researcher's overall research aim and objectives, as well as practicalities and resource constraints. In our study which aims at developing a modular, scalable and integrated Loan Management System, we have a specific objective of ascertaining all user requirements for efficient loan management based on customer expectations and regulations of lending businesses. Since our research study is explanatory in nature, qualitative methods such as interviews and focus groups will be the best fit. Large-scale surveys and other quantitative data gathering techniques which produce large volumes of numerical data will not be a better fit in this study.
In order to gather relevant data for this study, we made use of three major data collection instruments, including Interviews, Document Evaluation, and Personal Observation.

3.4. Requirement Analysis

During our interview sessions with some selected staff of Microfinance Bank and other key stakeholders in the banking sector and the cooperative societies, some set of structured questions were posed to them with the goal of ensuring that accurate and honest information were gathered for user requirements analysis.
Specifically, we tried to elicit information in the following areas:

i. **Need for possible integration and collaboration among Lending Businesses**: Our microfinance banks, cooperative societies, and other lending business has suffered the incidences of multiple loan access by fraudulent customers, who try to access multiple loan facilities from different banks using the same collateral security. The following questions were therefore posed to the bankers:
    - Do you need an integrated system that serves as regulatory agent to forestall incidences of multiple loan access using a single collateral security?
    - Do you desire an easy way to collaborate with other lending businesses to checkmate activities of fraudulent microfinance bank customers?

- Would you like to efficiently manage all lending processes to endure drastic reduction in rising cases of non-performing loans?

ii. **Need for Improved Security Measures to Checkmate "Insider abuse"**: Due to the perceived challenge of "insider abuse" and the need to checkmate activities of unscrupulous bank staff, the following questions were posed to stakeholders:

- Do you require a highly secured system in place such that activities of unscrupulous bank staff will be reduced to its barest minimum?
- Do you need a technology-assisted system that will not allow loan approvals beyond the required credit limit?
- Do you require improved loam management system?
- Do you think your business requires standardized lending processes?

iii. **QR Code-enabled Loan management System**: Based on the successes of the Quick Response Code technology in other sectors and application areas, top managers of microfinance banks and cooperative societies were asked the following question:

- What is your view with regards to the QR Code technology and its ability to secure your lending business from fraudsters?
- Do you desire a better user authentication for your lending business?

iv. QR Code Enhanced Loan Management System and Basic Application Features: To gather this requirement, the following questions were posed:

- What further features do you expect your loan management system to possess?
- Do you require improved security features?
- Give examples of the kind of features a standard loan management system should have.

v. Instant Customer Update via an Alert system: There have been reported cases from unsatisfied microfinance bank customers and borrowers who complain of lack of access to their debt balances, and inadequate communication system with bank management. Therefore, in order to gather user requirements in the area of customer update and alert systems, the following questions were asked:

- Do you need a functional alert system that will quickly update your customers on loan performance?
- Do you think it will help your customers in planning, when they are aware of their loan status on a regular basis?
- What type of alert system would you like to have in place?

i. **Report Generation to improve planning and management Decision making**: The following questions were posed in the area of Report Generation and management decision making:

- Do you have any need for better report generation platforms for your management decision?

- At what intervals do you think such reports should be generated?
- What is your projected performance enhancement based on your expectations of this improved system?

At the end of our interaction session with both management and staff, including selected customers and borrowers, it was discovered that more than ninety percent (90%) of all questions were answered in the affirmative, showing the need for an improved system of loan management using the QR code technology, and from the information gathered at this stage, the following facts were established:

a. An enhanced system of QR code-enabled platform for effective loan management in our microfinance bank institutions and other lending businesses is required for a sustainable growth and development.

b. Three is need to track incidences of multiple loan access using on single collateral security
c. Existing Loan management systems in microfinance banks and cooperative societies have major weaknesses which also has been the reason for frequent and incessant complaint from unsatisfied customers and borrowers.
d. There is need to maintain higher level of user authentication using the Quick Response Code technology.
e. There is need to forestall ever increasing cases of "insider abuse" and ensure strict compliance with approved lending regulations.
f. There is need to checkmate the excesses of fraudulent bank customers.

g. There is need to speed up loan application validation and monitoring by reduced involvement of manual operations.

h. There is need to ensure timely customer update on loan performance via a good notification system.

i. There is need for efficient report generation module to facilitate effective decision making by management of microfinance banks, and

j. Need to exploit the power and potency of the Quick Response code technology as a veritable technique capable of improving security in all credit institutions.

4.0. MODEL FORMULATION, RESULT AND DISCUSSIONS

4.1. Description of Major Activities of the Proposed QR Code-Enabled System

a. Quality Control Subsystem using Integrated Approach

The proposed system will develop a quality control subsystem that will check the quality of recommended loan application by member banks before a final loan approval can be issued. There is need to carry out thorough

scrutiny and quality control of recommended loans against all internal rules and external regulations before final approval. Since the lending business is highly regulated, the quality control aspect of loan processing is a critical factor to all lending institutions. The application should be sent to a credible body that analyses all critical variables against internal rules and regulations, and this should serve as the last look at the application before final approval and loan disbursement.

If properly handled by an independent body of credible personnel, the quality control check can end all cases of "insider abuse" and reposition microfinance banks and other lending institution to render qualitative service, while reducing incidences of non-performing credit in the banking sector.

b. QR code technology-based System and Improved Security Measure

The proposed system will include a QR code technology-based module to guarantee improved security of the entire loan process. The central admin manager will have the privilege of generating a QR code for each approved loan which only the individual customer can access, and independent of his/her credit institution. During and after loan disbursement by member banks, individual customer accounts can be accessed using the generated security code at the point of final loan approval.

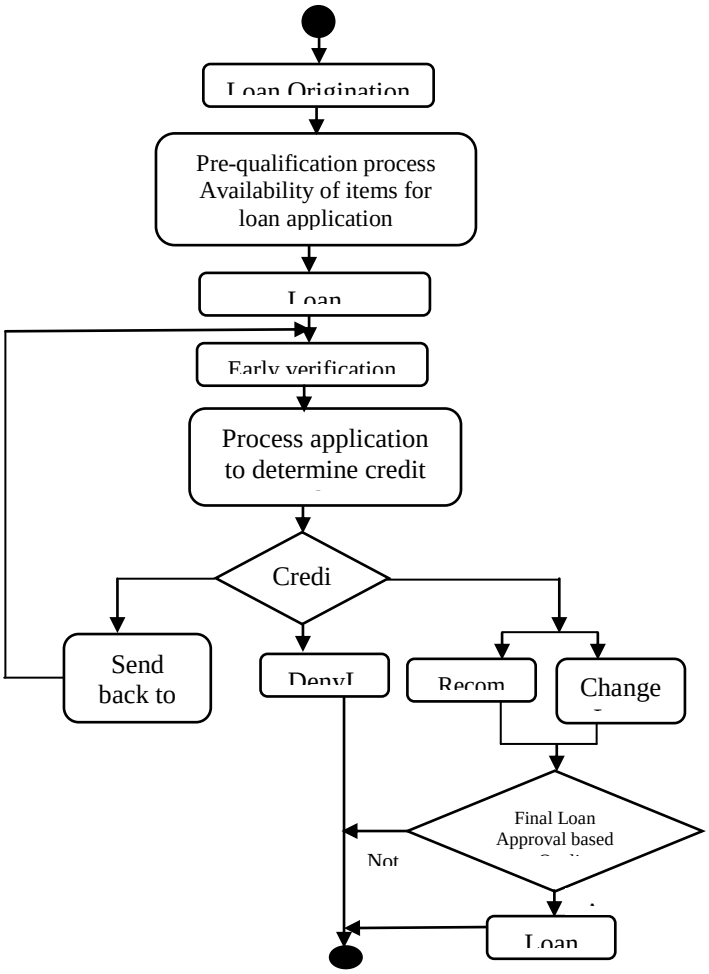


Figure 3 shows the Activity Diagram of the proposed QR code-enabled Loan Management System.

Figure 3. Activity Diagram of proposed System

4.2. New Architectural Framework

Figure 4 is the architectural framework of the new integrated loan management system. It models a QR code technology-based system that guarantees improved security of the entire loan process.

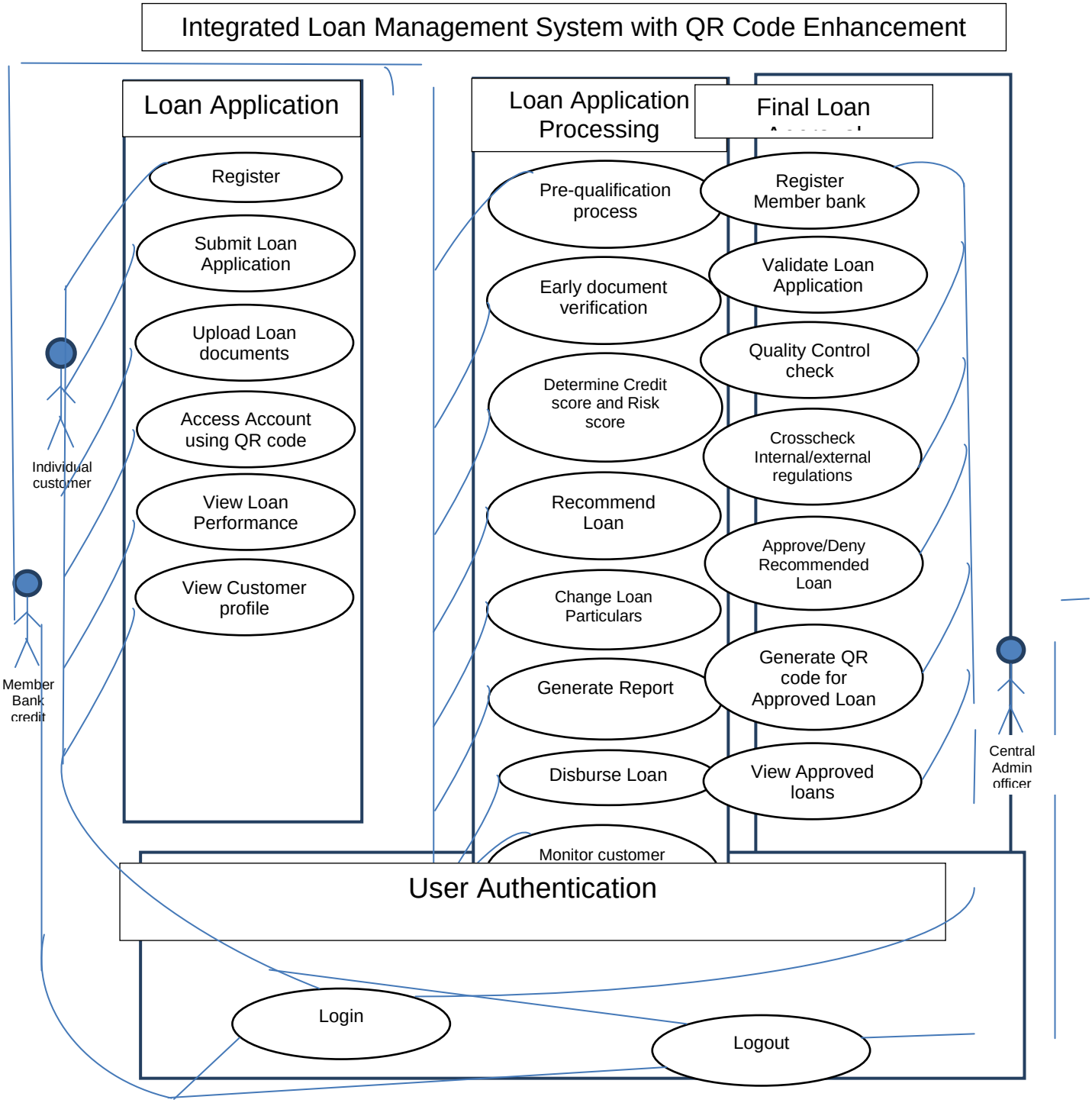


Figure 4. Architectural framework of the new system

With the new system architecture, the Central Admin Manager will have the privilege of generating a QR code for each loan application at the point of final approval. After loan disbursement by member banks, individual account holders can access their accounts using the generated QR code.

5.0. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

In this research paper, we have designed an integrated Loan Management with QR code technology enhancement for easy loan application validation and processing in all Lending Businesses.

Due to the rising number of microfinance bank customers in need of personal loans, especially in the post COVID-19 era, management is faced with the complex job of loan application verification in order to correctly determine eligibility for a loan. There is the challenge of coping with fraudulent customers who make false claims with their loan application documentations, sometimes seeking to access multiple loans from more than one microfinance banks at the same time, using a single collateral security. There arises a need therefore for an integrated loan management system to link other lending institutions in a collaborative effort to forestall incidences of multiple loan access using single security.

5.2. Recommendations

We recommend an early implementation of the findings of this research paper tittles "An Innovative Framework for Interated Loan management with QR code Enhancement". The system should be implemenmted and deployed in the Nigerian Banking sector, especially Microfinace banks, and lending institutions, including Cooperaive Societies for improved operational efficiency. There is need for a concerted effort by all stake holders, Government Agencies, Bank operators, Businessmen, and so on, to take advantage of the opportunities offered by the emergence the QR code system, Computing and Information Technology, to improve the sector.

We also encourage government should encourage the establishment of a central coordinating body in the microfinance banking sector and other lending institutions.

There should be an establishment of more stringent regulations to curb the menace of "insider abuse" in the microfinance banking sector.

REFERENCES

[1] Acha, I.A., 2012, Microfinance Banks in Nigeria: Problems and Prospects. Journal of Finance and Accounting, 1(5), 106-111.

[2] Al-Kilani, M. & Kobziez, V. (2016). An Overview of Research Methodology in Information System (IS). Open Access Library Journal, Vol. 3: 33126. doi:DOI: 10.4235/oalib.1103126

[3] Al-Husseini, O. A.& Obaid, A. H. (2018). Usage of Prototyping in Software Testing. Multi-Knowledge Electronic Comprehensive Journal for Education and Science Publications(Issue 14).

[4] Ana, I. (2008), "Microfinance Banking? What is that" Lagos the Guardian Newspaper, Saturday March 22, P.47.

[5] Beynon-Davies, P., Came, C., Mackay, H. & Tudhope, D. (2014). Rapid application development (RAD): An empirical review. European Journal of Information Systems, 211-223. doi: 10.1057/palgrave.ejis.3000325

[6] Canadi, M., Höpken, W., & Fuchs, M. (2010). Application of QR codes in online travel distribution. Inform. Commun. Technol. Tourism, 26, 137 - 148.

[7] Central Bank of Nigeria Research paper, 1-3 CBN, 2005, Microfinance Policy Regulatory and Supervisory Framework for Nigeria.

[8] Central bank of Nigeria, 2008. Guidelines and Procedures for the establishment of Microfinance banks in Nigeria, Published by the CBN.

[9] Deninzon, D., Malik, N., & Kapoor, A. (2019, June 20). Banking operations for a customer-centric world. Retrieved from McKinsey & Company:https://www.mckinsey.com/industries/financial-services/our-insights/banking-matters/banking-operations-for-a-customer-centric-world

[10] Dhanapal, C. & Ganesan, G. (2012). Measuring Operational Efficiency of Public Sector Banks in India. 2012 International Conference on Business and Management, (pp. 6-7). Phuket-Thailand.

[11] Derek Jansen and Kerryn Waren (2020): "What is Research Methodology? A Plan-language Explanation and Definition", http://gradcoach.com/what-is-research-methodology/

[12] Dube, S., Ndlovu S., Nyathi, T., & Sibanda, K. (2015). QR Code Based Patient Medical Health Records Transmission: Zimbabwean Case. Proceedings of Informing Science & IT Education Conference (InSITE) 2015, (pp. 521-520).

[13] Hirzallah, N., & Masalha, F. (2014). A Students Attendance System using QR Code. International Journal of Advanced Computer Science and Applications, 5(3), 75 - 79.

[14] Kanu Clementina (2015): "Microfinance Banks Operations in Nigeria, Constraints and Suggested Solutions: [15] An Evaluation", Global Journal of Contemporary Research in Accounting, Auditing and Business Ethics

(GJCRA), 2015 Vol: 1 Issue: 2

[16] Katharina, K., Peter, F., Peter, K., Ioannis, K., Markus, H. & Edgar, W. (2014). QR Code Security: A Survey of Attacks and Challenges for Usable Security. SBA Research, pp 1-8.

[17] Kharat, S. A., Panage, B. M., & Nagarkar, S. (2017). Use of QR code and layar app for academic library services. Librar Hi Tech News, 34, 21–28.

[18] Okpara, G. C., 2009, A Synthesis of the Critical Factors Affecting Performance of the Nigerian banking System. European Journal of Economics, Finance and Administrative Sciences, Issue 17, 34-44

[19] Shamim, H., Xiaoyan, Z., & Farjana, R. (2018). Examining the impact of QR codes on purchase intention and customer satisfaction on the basis of perceived flow. International Journal of Engineering Business Management.

A Pragmatic Approach to the Development of Cryptocurrency using Block Chain Data Structure

[1]Onyemauche U.C, [1]Okpala L.C. [2]Osundu, B, U., [3]Mbanusi, C.E., [3]Okwor, J.U.

[1]Department of Computer Science, Federal University of Technology Owerri, Imo State, Nigeria.
[2]Department of Health Information Management, Imo State College of Health Technology and Management Sciences, Amaigbo. Imo State Nigeria.
[3]Department of Computer Education, Federal College of Education (Technical)Umunze, Anambra State, Nigeria.
:Corresponding Author: osigwe.uchenna@yahoo.com

Abstract
This paper is aimed at implementing cryptocurrency using the block chain data structure. Block chain is considered by many to be a troublesome but core technology. Many researchers have realized the dire need of block chain but the research of block chain is still in its elementary level. Consequently, this study reviews the current academic research on block chain, especially in the subject area of business and economics (cryptocurrency). Block chain data structure promises doles in trust-ability, association, organization, identification, trustworthiness, and transparency. The block chain data structure offers great prospective to promote various sectors with its distinctive combination of characteristics, for example, decentralization, immutability, and transparency. Block chain Hi –Tech is revolutionary. The methodology used for this study is the Rapid Application Development (RAD). The RAD favors iterative development and the rapid construction of prototypes instead of large amounts of up-front planning. It will make life simpler and safer, changing the way personal information is stored and how transactions for good and services are made. During the process of development, the following tools were used: CSS, HTML5, JavaScript, Servlets. Java Server Pages (JSP) were used to elucidate adding fresh transactions and blocks. Block chain technology generates an enduring and immutable digital record of every transaction. This impassable digital ledger makes fraud, hacking, data theft and information loss dreadful. A proof-of-work unanimity algorithm was instigated using Java Programming language. We see auspicious possibilities in the use of this technology for science and academia. Keyword: Block Chain, Cryptography, Hacking, Bit Coin, Transactions

Introduction

A cryptocurrency (or crypto currency) is a digital strength built to work as a podium of altercation using cryptography to protect the transactions and to regulate the conception of adding units of the currency (Andy, 2021). Within cryptocurrency systems the well-being, veracity and balance of ledger are sustained by a community of conjointly skeptical parties known as miners: members of the general public using their computers to help authenticate and timestamp transactions adding them to the ledger in accordance with a particular time stamping method. Cryptocurrencies use various time stamping schemes which includes but not limited to proof-of-work, proof-of-stake, Byzantine Fault Tolerance. These time stamping schemes are sometimes referred to as consensus algorithms. The most widely used consensus algorithm in use today is the proof-of-work, which is used in the popular cryptocurrency bitcoin. The proof-of-work along with other consensus algorithms require solving complex cryptographic hash.

The rise and success of Bitcoin during the last six years proved that blockchain technology has real-world value. However, this technology also has a number of drawbacks that prevent it from being used as a generic platform for cryptocurrencies across the globe. Some notable drawback includes; risk of centralization, high transaction fee, scalability and energy consumption. How can we develop
a cryptocurrency that is more environmentally friendly, without the risk of centralization and unnecessary exorbitant fees (McDonnell, 2015).

Literature Review

Blockchain Data Structure

A blockchain, originally blockchain, is a growing list of records, called blocks, which are linked using cryptography (Raval, 2016). Blockchains which are readable by the public are widely used by crypto currencies. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented as a merkle tree root hash). By design, a blockchain is resistant to modification of the data. It is "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way". For use as a distributed ledger, a blockchain is typically managed by a peer-to-peer network collectively adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, which requires consensus of the network majority. Transaction data is permanently recorded in files called blocks. They can be thought of as the individual pages of a city recorder's record book (where changes to title to real estate are recorded) or a stock transaction ledger (Narayanan et al, 2016). Blocks are organized into a linear sequence over time (also known as the block chain). New transactions are constantly being processed by miners into new blocks which are added to the end of the chain. As blocks are buried deeper and deeper into the blockchain they become harder and harder to change or remove. Each block contains, among other things, a record of some or all recent transactions, and a reference to the block that came immediately before it. It also contains an answer to a difficult-to-solve mathematical puzzle - the answer to which is unique to each block. New blocks cannot be submitted to the network without the correct answer - the process of "mining" is essentially the process of competing to be the next to find the answer that "solves" the current block. The mathematical problem in each block is extremely difficult to solve, but once a valid solution is found, it is very easy for the rest of the network to confirm that the solution is correct. There are multiple valid solutions for any given block - only one of the solutions needs to be found for the block to be solved.

While each block contains a reference to the prior block, the collection of all blocks in existence can be said to form a chain. However, it's possible for the chain to have temporary splits - for example, if two miners arrive at two different valid solutions for the same block at the same time, unbeknownst to one another. The peer-to-peer network is designed to resolve these splits within a short period of time, so that only one branch of the chain survives.

The client accepts the 'longest' chain of blocks as valid. The 'length' of the entire block chain refers to the chain with the most combined difficulty, not the one with the most blocks. This prevents someone from forking the chain and creating a large number of low-difficulty blocks, and having it accepted by the network as 'longest'.

In every crypto currency, there is always a genesis transaction credit with a certain value. This transaction has no input. Users create transactions and submit them to the network, where they sit in a pool waiting to be included in a block.

Methodology

The software development methodology employed in this paper is the Rapid application development (RAD). The RAD favors iterative development and the rapid construction of prototypes instead of large amounts of up-front planning. The "planning" of software developed using RAD is interleaved with writing the software itself. The lack of extensive pre-planning generally allows software to be written much faster, and makes it easier to change requirements.

Analysis of the Block Data Structure

To successfully implement a crypto currency on the block chain distributed ledger, we need to understand the technical structure of a block. A block is a container data structure below are the properties of a block:

1. A block contains transactions
2. A block has size measured (e.g bytes, kilo bytes, mega bytes)
3. A block must have a header
4. An identifier (unique hash)
5. Height (the position in the block chain)

Anatomy of a Transaction

Typically, a block holds data which is usually why the ledger was developed. In the case of crypto currencies, the data held by blocks are called transactions, the transactions usually contain information about the sender of the crypto currency, the receiver and the value of the crypto currency exchanged.

Figure 3.1 Structure of a Transaction

Figure 1: Structure of a Transaction

Anatomy of a Block

The block is made up of the block header and a bundle of transactions. In this section we are going to implement the structure as a class in the Java programming language. Below is a diagrammatic representation of a block.

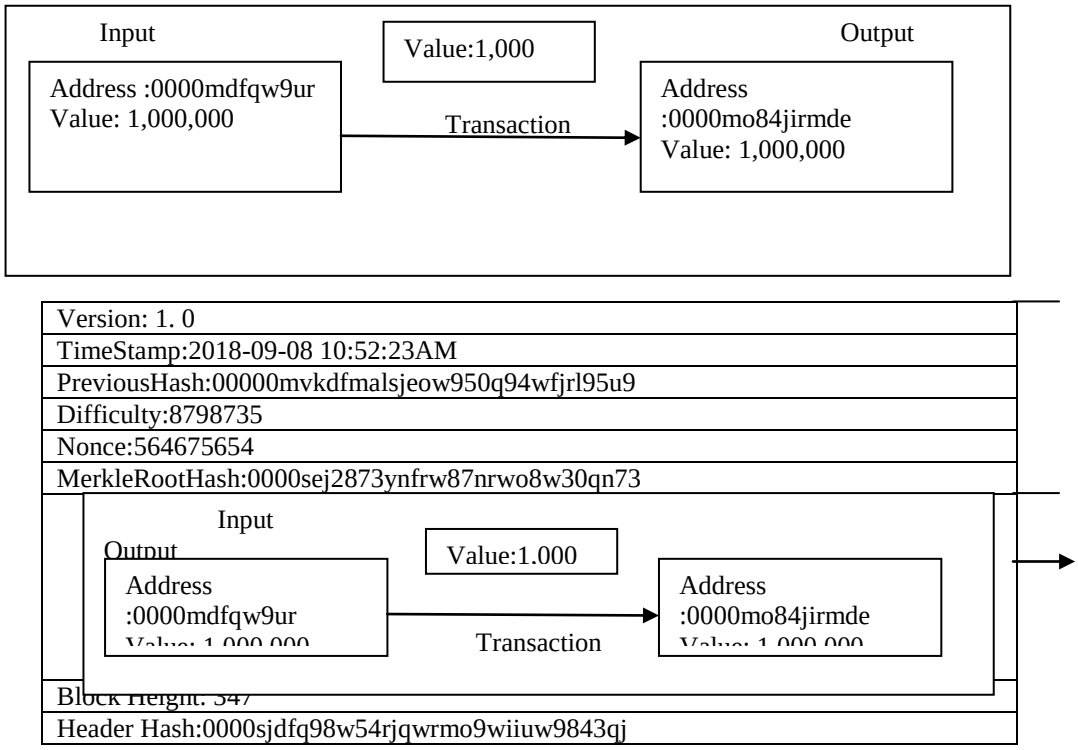


Figure 2: Structure of a Block

Results and Discussion

Class Diagrams

OOD process started by identifying the classes required to build the conference management system. These classes were discussed using class diagrams and implemented in Java Programming Language. The class diagram enable us to model via class diagrams, each class is modeled as a rectangle with three compartments. The top one contains the name of the class centered horizontally in bold face. The middle compartment contains the class attributes, while the bottom compartment contains the class behavior or operation. Figure 3 is the class diagram for the system.

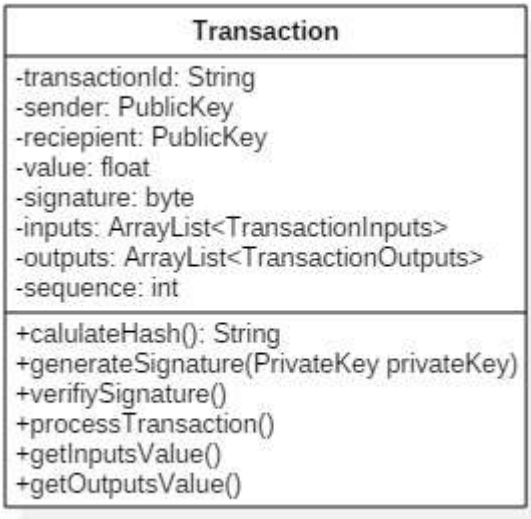


Figure 3: Transaction Class Diagram

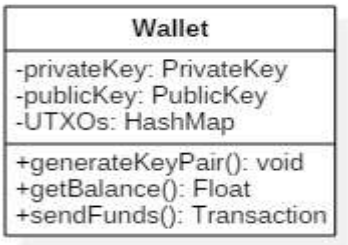


Figure 4: Wallet Class Diagram

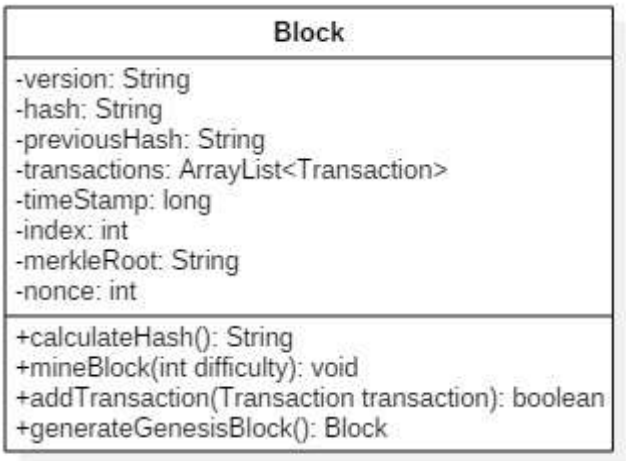


Figure 5: Block Class Diagram

Creating and Mining Genesis Block

In every cryptocurrency there is always a genesis block, that is the first block in the block chain.

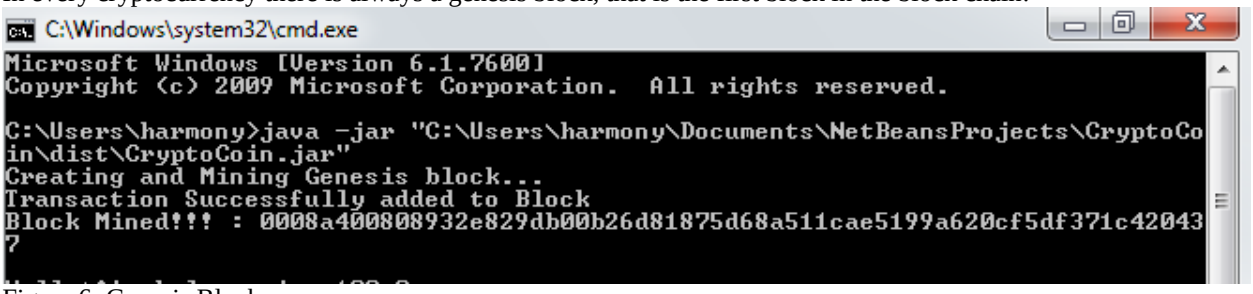


Figure 6: Genesis Block

TRANSACTIONS

In the figure 7, we perform transaction between two wallets and add them to the genesis block.

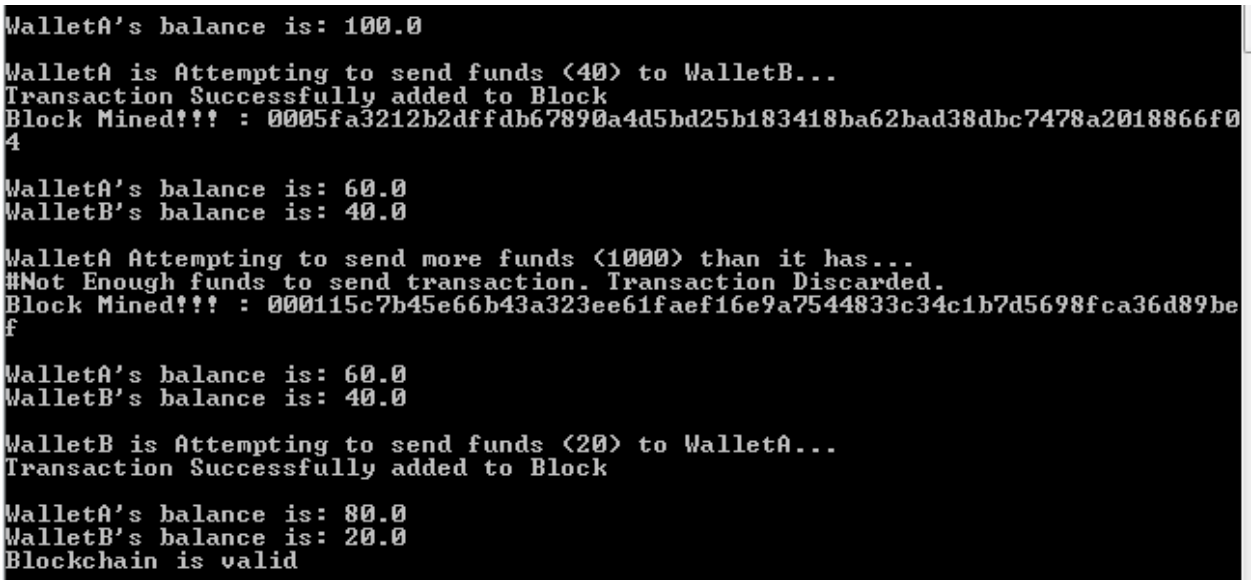


Figure 7: Result for Transaction between two wallets plus its addition to the Genesis Block.

Conclusion Recommendations and Future Work

Conclusion

Block Chain has shown its potential for transforming traditional industry with its key characteristics: decentralization, persistency, anonymity and auditability. In this paper, a comprehensive overview on blockchain using consensus algorithm was presented Firstly, an overview of blockchain technologies including blockchain architecture and key characteristics of blockchain were explored. Finally the typical consensus algorithms used in blockchain was discussed.

Recommendation and Future Work

Some vital features could not be incorporated during the process of developing this system, it is recommended that in future work, the following features be integrated.

1. Block mining should be reduced or eliminated in the next implementation because of the amount of resources it consumes.
2. The block should be implemented on a peer-to-peer network.
3. A mobile version of the application should be implemented

References

Andy Greenberg (2021). "Crypto Currency". Forbes.com. Archived from the original on 31 August 2021. Retrieved 8 January 2020

McDonnell, Patrick "PK" (2015). "What Is The Difference Between Bitcoin, Forex, and Gold". NewsBTC. Archived from the original on 16 September 2015. Retrieved 13 March 2020.

Narayanan, Arvind; Bonneau, Joseph; Felten, Edward; Miller, Andrew; Goldfeder, Steven (2016). Bitcoin and cryptocurrency technologies: a comprehensive introduction. Princeton: Princeton University Press. ISBN 978-0-691-17169-2.

Raval, S. (2016). "What Is a Decentralized Application?". Decentralized Applications: Harnessing Bitcoin's Blockchain Technology. O'Reilly Media, Inc. pp. 1–2. ISBN 978-1-4919-2452-5

EXPLORATORY APPROACH TO BIG DATA ANALYSIS USING BUSINESS INTELLIGENCE

[1]Theodora U. Onwuama, [2]Ikenna Caesar Nwandu, [3]Juliet N. Odii, [4]Francisca O. Nwokoma

[1,3,4]Department of Computer Science, Federal University of Technology, Owerri, Nigeria
[2]Department of Software Engineering, Federal University of Technology, Owerri, Nigeria
udora09@gmail.com, ikenna.nwandu@futo.edu.ng, juliet.odii@futo.edu.ng, francisca.nwokoma@futo.edu.ng

Abstract

The complexity of data in recent times gave rise to the necessity of the concept of big data. Big data comprises the large datasets generated daily from several web-based business firms which increasingly grow exponentially in complexity. The analysis of big data gives a threshold to the development of critical ideas required for business intelligence purposes. Business intelligence itself is a phenomenon that encourages the analysis of data, knowledge gathering and their application to a diversity of methods. The importance of big data analytics continues to pose some sort of concerns in decision making processes. Therefore, this paper makes an exploratory research on the use of business intelligence in big data analysis. It also discovers an easier process to decision making using business intelligence systems. A review of the components of big data and the components of business intelligence systems exposed in-depth ideas about the necessary knowledge required for business decision making which in turn leads to greater organizational productivity. The study concluded by establishing the fact that business intelligence plays a vital role in processing big data in its raw form and transforming them into useful formats that can be utilized in business policy formulations for the purpose of organization's performance assessment.

Keywords: Big Data, Business Intelligence, Analytics, Data Mining, Data Warehouse.

1. INTRODUCTION

The world today is faced with a staggering rate of generating data at an amazing speed. It has almost become a norm to entertain tremendous amount of data from social media, regular media houses, mobile applications, e-commerce, and cloud computing done over virtual platform (Setty et al., 2013). This is an indication that the structured and unstructured data that are on the increase by the day are generated from different forms and types of sources. Interestingly, the various sources of data are borne out of different motives. This phenomenon of realizing huge size of structured and unstructured data describes the term big data (Bucur, 2015). Due to the size of big data, it poses a challenge to process them into usable information using the traditional data processing approaches. As the name implies, big data is characterized by high volume, high velocity and high variety of information that gives room for creation

of new strategies that can reduce the difficulty of processing the huge sizes with the intent of having a business insight (Bucur, 2015). For the fact that big data is a collection of exponentially increasing complex data, Lcia (2011) agrees that unending cumulative lump is by no means viable to the conventional processing methods and techniques. While the fast data stream in the digital environment enables people to reach information quickly, it becomes hard to differentiate useful information among the giant data that is constantly accumulating.

With the current technologies, data is consolidated in different forms and the structural, non-structural or semi-structural components have been stored as in content form. However, with big data architecture, companies can analyze a large amount of data. Hence, they can develop their critical business ideas and manage their strategic planning for business intelligence purposes. By business intelligence (BI) we refer to a variety of software used to analyze an organization's raw data. These raw data stem from the simplest of data to big data. Business intelligence is made up of several related activities. These activities range from data mining, online analytical processing up to querying and reporting. Effective business intelligence systems give decision makers access to quality information, enabling them to accurately identify where the company has been, where it is now, and where it needs to be in future (Pablo et. al, 2014). Companies also use business intelligence to improve their decision making, cut costs and also identify new business opportunities. However, the usage and adoption of business intelligence systems remain low, despite its immense benefits, particularly among smaller institutions and companies with resource constraints (Pablo et. al, 2014).

2. DEFINING BIG DATA

With the enormous growth of data, there is now a paradigm shift in the perception of data. The boundaries in the approach built on a structured data model (relational/object-oriented model) and object application are emerging. This is the driving force that led to the era of unstructured databases. The unstructured databases are hitherto called NoSQL databases. The concept of NoSQL was first used in 1998 by Carlo Strozzi to refer to an open source database that does not use SQL interface (Abramova et. al, 2014). Strozzi prefers to refer to NoSQL as "noseequel" or "Norel" (no relational) since it is the main difference between this technology and relational model. Also, the emergence of unstructured databases is associated with Google. (Abramova et. al, 2014).

A 2008 study by International Data Corporation (IDC) predicted that over a thousand hexabytes of digital data will be generated in 2010 (Gantz, 2008). Storing and analyzing such volumes of information represents an insurmountable challenge for the current generation of database technology. For instance, companies like Google, Facebook, Amazon, and eBay generate petabytes of data on a daily basis. Facebook alone handles several petabytes of data of diverse forms ranging from pictures, sounds, texts, e.t.c. On a similar note, Google manages vast amounts of semi-structured data: billions of URLs with associated internet content, crawl metadata, geographic objects (roads, satellite images, etc.), and hundreds of terabytes of satellite image data, with hundreds of millions of users and thousands of queries per second (Dean, 2009). The scale and level of functionality required for such "big data" applications have not been anticipated by commercially available Database Management Systems, and almost invariably internet companies were forced to develop their own database solutions. But, even more traditional database applications manage increasingly large volumes of data. For example, the retail chain WalMart handles more than one million transactions per hour, and manages databases with more than 2.5 petabytes of data. These structured and unstructured data make up what is known today as Big Data.

3. ANALYZING BIG DATA

Big Data is often described as a multi-V model which helps to unleash the characteristics that qualify it as Big Data. There are up to 56 v's of big data as shown in figure 1, but this paper will discuss the 5 major v's. The 5 v's of big data are depicted in figure 2.

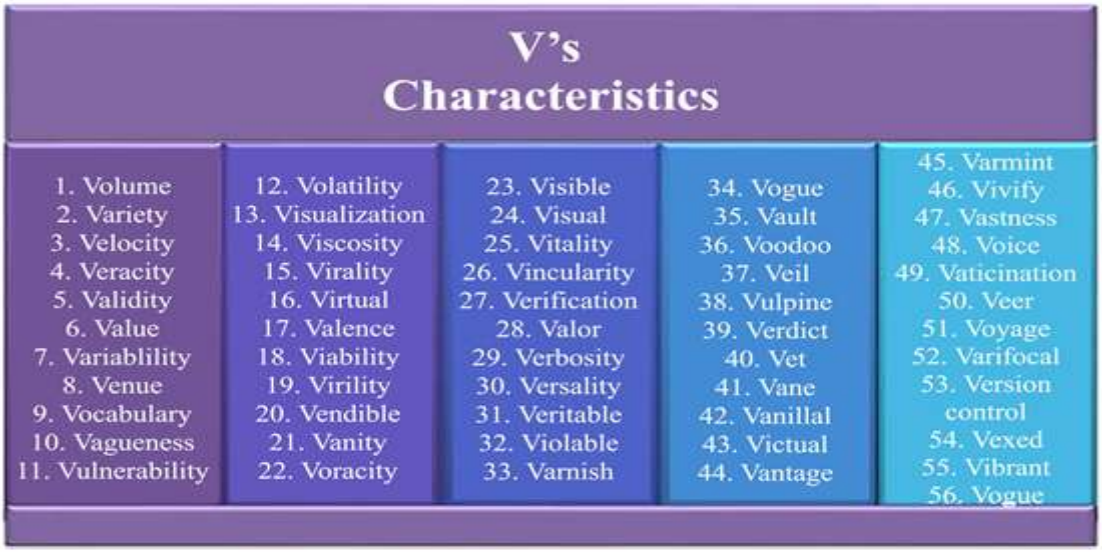


Figure 1: 56 v's of big data (Source: Hussein 2020)

In figure 2, Variety represents the data types, Velocity refers to the rate at which the data is produced and processed while Volume defines the size of data. On the other hand, Veracity refers to how much the data can be trusted given the reliability of its source (Yu et al., 2013), whereas Value corresponds the monetary worth that a company can derive from employing big data computing. Although the choice of V's used to explain big data is often arbitrary and varies across reports and articles on the web, variety, velocity, and volume are the items most commonly mentioned (Assunção et al., 2013).

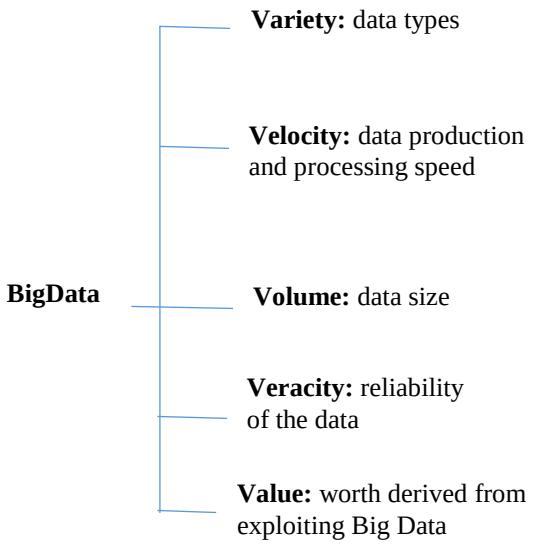


Figure 2: Characteristics nção et al., 2015)

3.1 Variety of data

Over the years, it has been observed that substantial amount of data has been made publicly available for scientific and business uses such as repositories with government statistics, historical weather information and forecasts, DNA sequencing, information on traffic conditions in large metropolitan areas, product reviews and comments, demographics (Assunção et al 2013), comments, pictures, and videos posted on social network websites, information gathered using citizen-science platforms (Bonney et. al, 2014), and data collected by a multitude of sensors measuring various environmental conditions such as temperature, air humidity, air quality, and precipitation. Handling and analyzing this data poses several challenges as it can be of different types. This is an indication that these public data could be structured, unstructured or a hybrid of both (see figure 3). An example illustrating the need for such a variety within a single analytics application is the Eco-Intelligence platform (Zhang et al., 2010).

Variety of data is shown in figure 3 while figure 4 shows velocity of data.

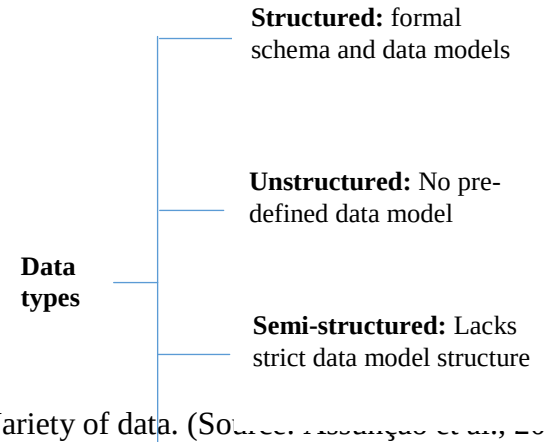


Figure 3: Variety of data. (Source: Assunção et al., 2015)

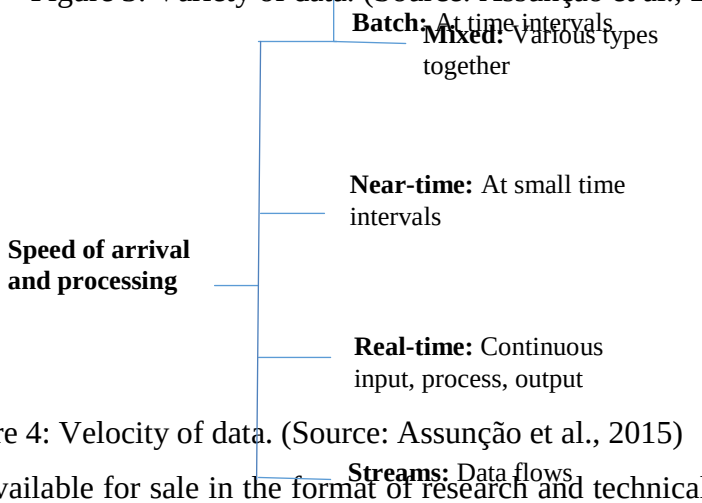


Figure 4: Velocity of data. (Source: Assunção et al., 2015)

Data is also often available for sale in the format of research and technical reports, market segment and financial analyses, among other means. This data can be used by various applications, for instance, to improve the living conditions in large cities, to provide better quality services, to optimize the use of natural resources, and to prevent or manage response to unplanned events. It is argued that a large part of data produced in the present digital age is either unstructured or semi-structured.

3.2 Velocity of Data

Data can arrive and require processing at different speeds (see figure 4). Usually, in some applications, the arrival and processing of data can be performed in batch, while other analytics applications require continuous and real-time analyses, sometimes requiring immediate action upon processing of incoming data streams. For instance, to provide active management for data centers. Wang et al. (2011) presents an architecture that integrates monitoring and analytics. The proposed architecture relies on Distributed

Computation Graphs (DCG) that are created to implement the desired analytics functions. The motivating use cases consist in scenarios where information can be collected from monitored equipment and services, and once a potential problem is identified, the system can instantiate DCGs to collect further information for analytics.

Increasingly often, data arriving via streams need to be analyzed and compared against historical information. This is because a lot of products that currently produce digital data, especially the internet, often generate data that are uncertain and noisy resulting to what is usually known as information garbage. The ever increasing need to evaluate, activate and store the arriving big data accruing from many channels necessitated the creation of data warehouse.

3.3 Volume of Data

The amount of data contained in the repository or used by organizations, networks or infrastructure refer to volume of data. Simply put, volume of data refers to the amount of data in a file or database. The volume of big data is of importance because it gives the size of datasets needed for analysis and processing. The volume of data can be measured over a period of time as the product of the quantity of data obtained and the quantified period of measurement. The volume of big data has massively grown larger than terabytes and petabytes such that they now require sophisticated technologies for processing. This means that the volume of big data is relatively too large.

3.4 Veracity of Data

By veracity we mean the quality and reliability of the data undergoing analysis. Veracity of data tries to probe into the amount of trust that can be laid on data source, data type and the processing technique. In other words, veracity of data ensures that big data is devoid of statistical bias, data lineage issues, bugs, noise, abnormalities, ethical threats, falsehood, uncertainty, obsolete, human error and inaccuracy. Veracity filters through what is important and what is not important. This process enables a better understanding of data and how to contextualize it in order to act on it. Veracity of data is an important characteristic of big data which ensures that the processing technique employed for data processing satisfies business needs and the output is pertinent to objectives (Wengel and Renee 2019).

3.5 Value of Data

Value of data refers to the ability to transform huge amount of data into business. In other words, value of data gives the monetary worth that a company can derive from analyzing its available data using the big data computing. In business organizations, big data has made it possible to know faithful partners

and created an enormous avenue for the monetization of that information. The value of data is usually perceived as an asset that is very important for making appropriate investment decisions. Value of data is an off-shoot of data assessment which places recognition on the data in order to make it valuable. The organization needs to strategize on how to float on better operations (such as more efficient delivery of products and services) for the purpose of creating value for the organization's data.

4 BUSINESS INTELLIGENCE

The analysis of Big Data is achieved through various techniques and tools such as Business Intelligence, the NoSQL databases (such as MongoDB, CouchDB, Cassandra, etc.) and also some Big Data tools (such as Hadoop, Pig, Hive, among many others). This work focuses on Business Intelligence as a basic mechanism for analyzing big data. Nithya and Kiruthika (2021) defined Business intelligence as a set of methodologies to convert a raw data set to meaningful and useful information for making decisions would help in quick computations, enhanced communication and collaboration, increased productivity of teams, efficient use of volumes of data and offers support anytime and anywhere.. Basically, the term business intelligence can be viewed from two different perspectives. Firstly, a school of thought is of the opinion that human intelligence or the ability of a common brain to be applied to business affairs can be termed business intelligence. Here, business intelligence is seen as a novelty, whereby the applications of human intellect and new technologies like artificial intelligence is used in management and decision making in different business related problems. Secondly, business intelligence can be viewed as the information which grows currency in business. In this case, the intelligent knowledge gained by experts are used alongside efficient technology in managing organizational and individual businesses. Therefore, business intelligence basically involves analyzing data, collection of knowledge and applying them to various different methods.

Business intelligence became a popular term in the business and IT communities only in the 1990s. In the late 2000s, business analytics was introduced to represent the key analytical component in BI. As a data-centric approach, Business Intelligence has its roots in the longstanding database management field. It depends so much on various data collection, extraction, and analysis technologies.

COMPONENTS OF BUSINESS INTELLIGENCE SYSTEM

(a) A business intelligence system refers to a set of integrated tools, technologies and programmed products used to collect, integrate, analyze, and make data (Koronios et al., 2010). More broadly, Arnott et al. (2004) described business intelligence systems as a collection of software platforms,

applications, and technologies that aim to help decision makers perform more effectively and efficiently. These systems are primarily meant for the creation of necessary knowledge required for business decision making (Olszak et al., 2006)). Business intelligence systems can be used to guide and improve decision making at all levels, strategic, tactical and operational (Coman, et al., 2010). In other words, the systems play vital roles at every level of business management. At managerial level, business intelligence systems provide adequate help to individuals in performing their day-to-day tasks effectively as well as providing the input to strategic and tactical decisions at the top management. Business intelligence systems improve strategic decisions by creating an enabling platform for the use of available data to make forecasts into the future based on historical results. Also, business intelligence systems provide a basis for decision making to optimize actions for overall company performance on the tactical level. As a guide at operational level decision making, business intelligence systems are used to perform just-in-time analysis of departmental performance (Olszak et al., 2007). Figure 5 shows the role played by business intelligence systems in business decision making. There are at least four specific components required by all business intelligence systems to produce business intelligence. They include: Data warehouses, ETL tools, OLAP techniques, and Data mining.

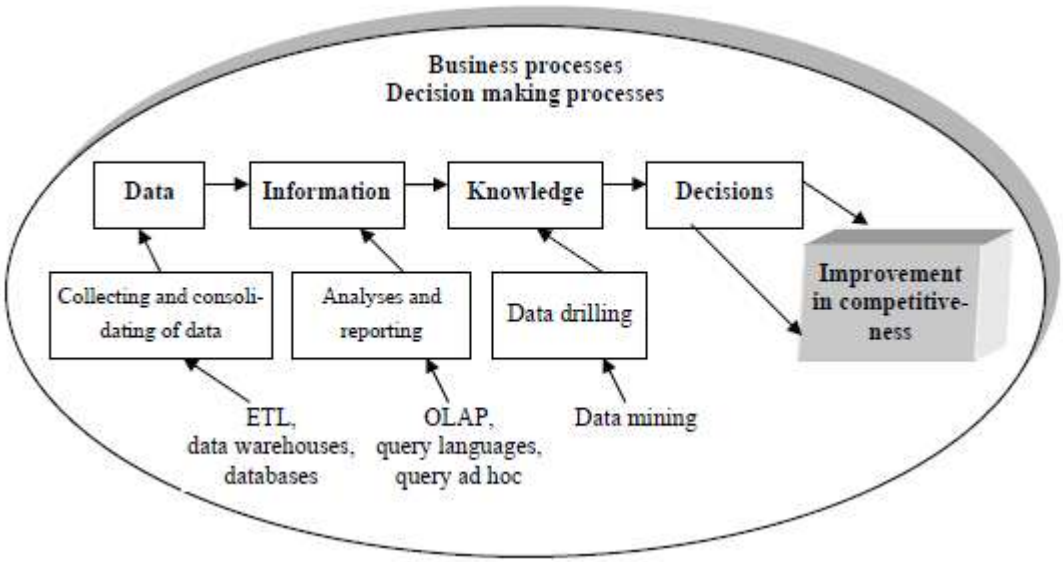


Figure 5: The Role of Business Intelligence in decision making. (Source: Olszak et al., 2007)

Data warehouse is a subject oriented, collection of data used to support decision making in organizations (Anderson et al., 2008). Data warehouse is usually seen as a data repository that is maintained separately

from an organization's operational databases. This means that data warehouses are maintained offline. Data warehousing provides architectures and tools for business executives to systematically organize, understand, and use their data to make strategic decisions (Han et al., 2012). It can be simply maintained that data warehouse is a collection of relevant historic business data which can be collectively analyzed to support and improve business decisions.

Extract-Transform-Load (ETL) tools play the role of extracting different formats of data from various and numerous sources and transforming them into a common format compatible to the data repository (data warehouse). Usually, these extracted data are pre-processed by the ETL tools into a useful format before they are loaded into data warehouse structures (Hevner et al., 2005). ETL tools and processes are viewed as a subset of data warehousing because they play supporting roles to the data warehouse. Hence, ETL processes perform mild transformations on external data before loading them into the enterprise data warehouse.

Online analytical processing (OLAP) techniques are systems used in organizing and presenting data in various formats in order to accommodate the diverse needs of different users (Han et al., 2012). OLAP can also be described as tools that allow the analysis of multidimensional data known as cubes. By cubes we mean data that are extracted from the data warehouse and used by managers in decision-making situations (Hevner et al., 2005). OLAP systems manage large amounts of historic data and provide facilities for data clustering. They also store and manage information at different levels of granularity thus, making data simpler and more useful for informed decision making. OLAP systems create room for analysis of data from multiple perspectives and further exploring into the data to discover hidden or unknown information.

Data mining is a process that turns a large collection of data into knowledge which may describe a reality or used for predicting an outcome of a decision. Data mining process involves discovering various patterns, generalizations, regularities and rules in data resources (Lloyd, 2011). Data mining should have been more appropriately named "knowledge mining from data," to avoid the anomaly of treating it as merely an essential step in the process of knowledge discovery. However, for convenience sake, the term data mining was universally adopted. The predictions generated by data mining use known variables to predict the outcome of a situation, while reality is measured by graphing, tabling, and creating formulas based on the existing data (Olszak et al., 2007).

5 ANALYZING BIG DATA USING BUSINESS INTELLIGENCE

The increasing use of social media such as YouTube, Twitter, and Weibo has contributed a huge percentage of the total data available in our present world. For instance, Walmart can handle millions of transactions per hour. Same goes for Twitter, Facebook, Webo and other social media platforms. These unprecedented large and complex data have brought the need to develop software and systems that import data streams of any size and use them to generate informational displays that point towards specific decisions. This is the essence of business intelligence because there is great need to finding lasting solution to the huge data problem. Business intelligence is capable of utilizing data streams of any size to analyze it and display crucial information. This process is known as "analytics" because it analyzes and then digests data streams in a way that is both easier to understand and points more obviously towards needed actions based on said data. The implication of this is that business intelligence systems support and help in decision making processes. This can be seen in Walmart's nappy and beer model.

Business intelligence also plays great role in strategic planning within an organization by addressing the achievement of management effectiveness. Effective business intelligence systems give decision makers access to quality information, enabling them to accurately identify where the company has been, where it is now, and where it needs to be in future. Business intelligence therefore is of paramount importance as per its sustenance and continuous development because digital systems are generating more data than ever, and new approaches are needed to handle and store this data.

6 CONCLUSION

With the enormous growth of data and speed of data processing, it becomes difficult to handle data which are generated in different varieties and formats. These data of various forms gave birth to the concept of big data. The term "big data" has become a household name as the world becomes ubiquitously digitized. This is because the digitization of all facets of human endeavor is solely responsible for the generation of unending and exponentially complex data. The need for analyzing these data is of great importance as data generated at such rate is meaningless unless properly harnessed. That is why data warehousing became popular. Data warehousing extended its functionality to employing business intelligence in data processing in order to make data more meaningful. This work paid attention to the relevance of business intelligence in big data analysis. This study carried out an exploratory study on business intelligence to ascertain how its application to companies' and/or

organizations' data are used in strategic planning and consequent effective business decisions. This study conclusively opines that application of business intelligence in the analysis of big data results in the generation of more meaningful information, essential for the formulation of business policies needed for organizational growth.

REFERENCES

Abramova V., Bernardino J. and Furtado P. (2014) *Experimental Evaluation of NoSQL Databases. International Journal of Database Management Systems (IJDMS) Vol.6, No.3DOI :10.5121/ijdms.2014.6301, 3.*

Anderson, D., Fries, H., and Johansson, P. (2008). Business intelligence: The Impact on Decision Support and Decision Making Processes. Retrieved from http://hj.diva-portal.org/smash/record.jsf?pid=diva2:3599

Arnott, D., Gibson, M., and Jagielska I. (2004). Evaluating the intangible benefits of business intelligence: review & research agenda. *The IFIP TC8/WG8.3 International Conference*, 1-11. doi:10.1.1.94.8550

Assunção M.D., Calheiros R.N., Bianchi S., Netto M.A.S., Buyya R. (2015). Big Data computing and clouds:Trends and future directions. *Journal of Parallel Distributed Computing* 79–80

Bonney R., Shirk J.L., Phillips T.B, Wiggins A., Ballard H.L., Miller-Rushing A.J and Parrish J.K. (2014) *Next Steps for Citizen Science* 1436–1437

Gantz, J. F., (2008). The Diverse and Exploding Digital Universe. IDC. IDC. http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf

Lcia, (2011). Big Data: Big Opportunities to Create Business Value. http://www.emc.com/microsites/cio/articles/big-data-big-opportunities/LCIA-BigDataOpportunities-Value.

Lloyd, J. (2011). Identifying Key Components of Business Intelligence Systems and Their Role in Managerial Decision making. Applied Information Management Program, University of Oregon, 1-76.

Muriithi G. M. and Kotzé J. E. (2013) A Conceptual Framework for Delivering Cost Effective Business Intelligence Solutions as a Service. *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference, East London, South Africa: ACM; 2013,* 96–100.

Nithya N. and Kiruthika R. (2021) Impact of Business Intelligence Adoption on performance of banks: a conceptual framework. *Journal of Ambient Intelligence and Humanized Computing* 12,3139–3150. https://doi.org/10.1007/s12652-020-02473-2

Olszak, C.M., and Ziemba, E. (2006). Business intelligence systems in the holistic infrastructure development supporting decision-making in organizations. Interdisciplinary Journal of Information, Knowledge and Management, 1, 47- 58. doi:10.1.1.99.8329

Olszak, C.M., and Ziemba, E. (2007). Approach to building and implementing business intelligence systems. Interdisciplinary Journal of Information, Knowledge and Management, Retrieved from http://www.ijikm.org/ (2)135-148

Pablo Michel Marín-Ortega, Viktor Dmitriyev, Marat Abilov and Jorge Marx Gómez (2014) ELTA: New Approach in Designing Business Intelligence Solutions in Era of Big Data Conference on ENTERprise Information Systems / ProjMAN 2014 - International *Conference on Project MANagement / HCIST 2014 - International Conference on Health and Social Care Information Systems and Technologies.* 668

Setty, K., and Bakhshi, R. (2013). What is Big Data and What Does It Have To Do With Audit? *ISACA Journal,* 3, 23-25. View at Google scholar

Wang Y and Liu Z (2009). Study on Port Business Intelligence System Combined with Business Performance Management. Proceedings of the 2009 Second International Conference on Future Information Technology and Management Engineering, Washington, DC, USA: IEEE Computer Society; 258–260.

Wang C., Schwan K., Talwar V., Eisenhauer G., Hu L., and Wolf M. (2011) A Flexible Architecture Integrating Monitoring and Analytics for Managing Large-Scale Data Centers, in: Proceedings of the 8th ACM International Conference on Autonomic Computing, 141–150

Wengel R.and Renee S., (2019). Veracity: The Most Important "V" of Big Data. Retrieved from http://www.gutcheckit.com/blog/veracity on 24/02/2021.

Yu P.S (2013) On mining big data in: Lecture Notes in Computer Science, *Springer*, 7923, 14.

Zhang X., Zhang E., Song S., Wei F. (2010) Towards Building an Integrated Information Platform for Eco-city, in: Proceedings of the 7th International Conference on e- Business Engineering (ICEBE2010), 393– 398.