Analytical Study of Some Selected Classification Algorithms in WEKA Using Real Crime Data

Obuandike Georgina N.

Department of Mathematical
Sciences and IT
Federal University Dutsinma
Katsina state, Nigeria

Audu Isah

Department of Mathematics and
Statistics
Federal University of Technology
Minna, Niger State

John Alhasan

Department of Computer Science
Federal University of Technology,
Niger State, Nigeria

Abstract—**Data mining in the field of computer science is an answered prayer to the demand of this digital age. It is used to unravel hidden information from large volumes of data usually kept in data repositories to help improve management decision making. Classification is an essential task in data mining which is used to predict unknown class labels. It has been applied in the classification of different types of data. There are different techniques that can be applied in building a classification model. In this study the performance of these techniques such as J48 which is a type of decision tree classifier, Naïve Bayesian is a classifier that applies probability functions and ZeroR is a rule induction classifier are used. These classifiers are tested using real crime data collected from Nigeria Prisons Service. The metrics used to measure the performance of each classifier include accuracy, time, True Positive Rate (TP) Rate, False Positive (FP) Rate, Kappa Statistic, Precision and Recall. The study showed that the J48 classifier has the highest accuracy compared to other two classifiers in consideration. Choosing the right classifier for data mining task will help increase the mining accuracy.**

Keywords—Data Mining; Classification; Decision Tree; Naïve Bayesian; Tp Rate; component; formatting

I. INTRODUCTION

In this era of digital age and with the improvement in computer technology, many organizations usually gather large volumes of data from operational activities and after which are left to waste in data repositories. That is why [1] in his book said that we are drowning in data but lack relevant information for proactive management decision. Any tool that will help in the analysis of these large volumes of data that is being generated daily by many organizations is an answered prayer. It was this demand of our present digital age that gave birth to the field of data mining in computer science [2].

Data Mining is all about the analysis of large amount of data usually found in data repositories in many organizations. Its application is growing in leaps and bounds and has touched every aspect of human life ranging from science, engineering to business applications [3]. Data mining can handle different kinds of data ranging from ordinary text and numeric data to image and voice data. It is a multidisciplinary field that has applied techniques from other fields especially statistics, database management, machine learning and artificial intelligence [3].

With the aid of improved technology in recent years, large volumes of data are usually accumulated by many organizations and such data are usually left to waste in various data repositories. With the help of data mining such data can now be mined using different mining methods such as clustering, classification, association and outlier detection method in order to unravel hidden information that can help in improved decision making process [4].

Crime is a social sin that affects our society badly in recent times. Thus, to control this social sin, it is needful to put in place effective crime preventive strategies and policies by analyzing crime data for better understanding of crime pattern and individuals involved in crime using data mining techniques. Understanding the capability of various methods with regards to the analysis of crime data for better result is crucial. Classification is the data mining technique of focus in this paper. The performance of some selected classifiers such as J48, zeroR and Naïve Bayes are studied based on metrics such as accuracy, True Positive (TP) Rate, False Positive (FP) Rate, Kappa statistics, precision, recall and time taken to build the classification models.

The rest of the sections are discussions on the classifiers and their performance analysis with real crime data collected from the Nigeria Prisons Service in 2014.

II. CLASSIFICATION

Classification is the act of looking for a model that describes a class label in such a way that such a model can be used to predict an unknown class label [3]. Thus, classification is usually used to predict an unknown class labels. For instance, a classification model can be used to classify bank loans as either safe or unsafe. Classification applies some methods like decision tree, Bayesian method and rule induction in building its models. Classification process involves two steps. The first step is the learning stage which involves building the models while the second stage involves using the model to predict the class labels.

A record E with $n-$ attributes can be represented as $E = (e_1, e_2, \dots e_n)$ each of the records E belongs to a class of attributes $(A_1, A_2, \dots A_n)$. An attribute with discrete value is termed categorical or nominal attribute and this is normally referred to as class labels. The set of records that are used to

build classification models are usually referred to as training records. The model can be represented as a function $Y = F(e)$ which denotes the attribute Y of a particular record E. This function can be represented as rules, decision trees or mathematical formulae.

III. DECISION TREE

It is a well known classification method that takes the form of tree structure and it is usually made up of:

1) Testing node which holds the data for testing the condition

2) Start node is the parent and usually top most node.

3) Terminal node (leaf node): is the predicted class label

4) Branches: represents results of a test made on an attribute.

Figure 1: is a sample decision tree that predicts the purchasing interest of a customer in computer. Rectangular shapes are used for testing nodes while oval shapes are used for result nodes. It is mostly binary while others are non binary.
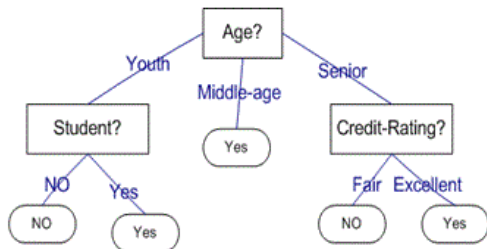

Fig. 1.   A simple Decision Tree

Source: (Jiawei et al, 2011)

B. Building Decision Tree

Decision tree can be built using different methods, the first method developed was ID3 (Interactive, Dichotomiser) which later metamorphosed into C4.5 classifier. J48 classifier is an improved version of C4.5 decision tree classifier and has become a popular decision tree classifier. Classification and Regression Trees (CART) was later developed to handle binary trees. Thus, ID3, J48 and CART are basic methods of decision tree classification [5].

C. Decision Tree Algorithm

Algorithm
Parameters Dataset T and its fields Set of Attributes A Selection Technique for the Attribute Result Tree Classifier Procedure 1) A node is Created (call it E) 2) Check if all records R is in one group G and write node

Algorithm

E as
     the last node in the that Group G
3) If $A = 0$ (no attribute)
4) then write E as the last node
5) Use Selection technique for attributes on (R, A) to get the
     Best splitting condition
6) Write the condition on node E
7) Check if attribute is discrete and allows multiway split then
     It is not strictly binary tree
8) For all output O from splitting condition, divide the records and build the tree
9) Assign $R_0 = Set\ of\ all\ records\ in\ output\ O$
10) If $R_o = 0$ then
11) Node E is attached with a leaf labelled with majority class R
12) Otherwise node E is attached with node obtained from
     Generate Decision Tree (R_o, A)
13) Next
14) Write E

Fig. 2.   Decision Tree Algorithm

Source: (jiawei, et al 2011)

IV.   NAÏVE BAYESIAN

This is a classification method that is based on Bayes' theorem which is used to predict class labels. This classifier is based on probability theorem and is named after Thomas Bayes who is the founder of the theorem [6].

Suppose R is a record set, it is considered as evidence in Bayesian theorem and R depends on n-features. Assume rule T implies that R ∈ K class, the condition that T is true if R ∈ K is given by P(T ∩ R).

For example, suppose a dataset R is described by age and educational qualification and R is a person within the age of 20 - 34 and has no educational qualification and T is a rule that someone within that particular age limit and educational qualification is likely to commit an offense then P(T ∩ R) implies that someone is likely to commit an offense if its age and educational qualification is within the limit.

P(T) is a general probability which implies that anyone is likely to commit offense not minding the age and educational qualification and other things that might be considered thus P(T) is not dependent on R. In order words, P(T ∩ R) is the probability of R when satisfied rule T. That is to say that a person is likely to commit an offense if the age and educational qualification is within the rule. P(R) is the probability that someone from the given dataset is within the age limit and a given educational qualification level. Bayes' theorem is given as in equation   1.

$$P(T\ \cap\ R\ ) = \frac{P(R \cap T)P(T)}{P(R)}, \text{provided } P(R) > 0 \qquad (1)$$

V.   ZEROR CLASSIFIER

It is a rule based method for data classification in WEKA. The rule usually considers the majority of training dataset as

real Zero R prediction. Thus, it focuses on targeted class labels and ignores others. Zero R is not easily predictable; it only serves as a baseline for other classifiers [7].

VI. ABOUT WEKA

It is machine learning software developed at university of Waikato in New Zealand. It is an open source software and can be freely downloaded from this web site address http://www.cs.waikato.ac.nz. It accepts its data in ARFF (Attribute Related File Format). It has different algorithms for data mining and can work in any platform. The Graphical User Interface (GUI) is as shown in figure 3 [8].
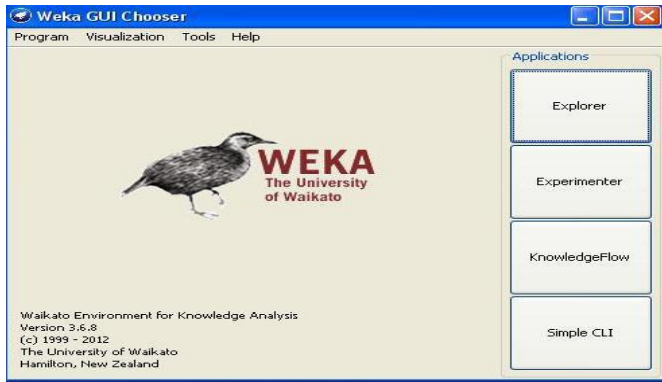

Fig. 3.   WEKA GUI Chooser

VII. EXPERIMENTS

A. Evaluation Metrics

The parameters considered while evaluating the selected classifiers are:

1) Accuracy: This shows the percentage of correctly classified instances in each classification model

2) Kappa: Measures the relationship between classified instances and true classes. It usually lies between [0, 1]. The value of 1 means perfect relationship while 0 means random guessing.

3) TP Rate: Is the statistics that shows correctly classified instances.

4) FP Rate: Is the report of instances incorrectly labelled as correct instances.

5) Recall: Measures the percentage of all relevant data that was returned by the classifier. A high recall means the model returns most of the relevant data.

6) Precision: Measures the exactness of the relevant data retrieved. High precision means the model returns more relevant data than irrelevant data.

7) Time: Time taken to perform the classification [9;10].

B. Datasets

A real crime data collected from selected prisons in Nigeria were used to perform this experiment. The dataset were converted to Attribute Related File Format (ARFF) form for easy processing by WEKA. The dataset was divided into two: training set and test set. The former was used to train the model while the other was used to test the built model. A cross validation process was applied in dividing the dataset into

training and test set. The process divides the data into equal parts usually k = 10 and the model was trained using k − 1 fold and kth fold was used as test set. The process was repeated K times to allow for both training and testing of each set.

C. Testing of J48 Classifier on crime data

J48 classifier is an enhanced version of C4.5 decision tree classifier and has become a popular decision tree classifier. It builds its model using a tree structure which usually made up of the following:

1) Testing node which holds the data for testing the condition

2) Start node is the parent and usually top most of the node.

3) Terminal node (leaf node): is the predicted class label

4) Branches: represents results of a test made on an attribute.

```
=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Combined_Prisons
Instances:   1733
Attributes:  4
             Offence
             Age
             Edu-Qualification
             Occupation

Time taken to build model: 0.76 seconds


=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1025        59.146 %
Incorrectly Classified Instances     708        40.854 %
Kappa statistic                      0.1516
Mean absolute error                  0.2769
Root mean squared error              0.3783
Relative absolute error             93.9487 %
Root relative squared error         98.5845 %
Total Number of Instances           1733


=== Detailed Accuracy By Class ===

TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
0.266    0.09     0.515      0.266   0.351      0.627     Secondary
0        0.003    0          0       0          0.663     Tertiary
0.022    0.013    0.2        0.022   0.04       0.586     Primary
0.913    0.758    0.613      0.913   0.733      0.61      NONE
0.591    0.456    0.51       0.591   0.514      0.613     weighted Avg


=== Confusion Matrix ===

  a   b   c   d   <-- classified as
122   2  10 324 |   a = Secondary
 10   0   2  51 |   b = Tertiary
 30   0   5 193 |   c = Primary
 75   3   8 898 |   d = NONE
```

Fig. 4.   Run information for J48 classifier

D. Naïve Bayes Classifier evaluation on Crime data

```
=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes
Relation:     Combined_Prisons
Instances:    1733
Attributes:   4
              Offence
              Age
              Edu-Qualification
              Occupation
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier


Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         984               56.7802 %
Incorrectly Classified Instances       749               43.2198 %
Kappa statistic                        0.0813
Mean absolute error                    0.279
Root mean squared error                0.3771
Relative absolute error                94.6346 %
Root relative squared error            98.2499 %
Total Number of Instances              1733

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area  Class
0.157     0.077     0.424       0.157    0.229       0.643     Secondary
0         0.002     0           0        0           0.709     Tertiary
0.031     0.015     0.233       0.031    0.054       0.612     Primary
0.92      0.834     0.592       0.92     0.72        0.634     NONE
0.568     0.496     0.478       0.568    0.477       0.636 Weighted Avg

=== Confusion Matrix ===

  a   b   c   d   <-- classified as
 72   0  10 376 |   a = Secondary
  7   0   2  54 |   b = Tertiary
 24   2   7 195 |   c = Primary
 67   1  11 905 |   d = NONE
```

Fig. 5.   Run Information for Naïve Bayes Classifier

E. ZeroR Classifier Evaluation

It is a simple classification method that works with mode for the prediction of nominal data and mean for the prediction of numeric data. It is usually referred to as majority class method.

```
=== Run information ===

Scheme:weka.classifiers.rules.ZeroR
Relation:     Combined_Prisons
Instances:    1733
Attributes:   4
              Offence
              Age
              Edu-Qualification
              Occupation
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: NONE

Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         984               56.7802 %
Incorrectly Classified Instances       749               43.2198 %
Kappa statistic                        0
Mean absolute error                    0.2948
Root mean squared error                0.3838
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              1733

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area  Class
0         0         0           0        0           0.497     Secondary
0         0         0           0        0           0.483     Tertiary
0         0         0           0        0           0.495     Primary
1         1         0.568       1        0.724       0.498     NONE
0.568     0.568     0.322       0.568    0.411       0.497 Weighted Avg

=== Confusion Matrix ===

  a   b   c   d   <-- classified as
  0   0   0 458 |   a = Secondary
  0   0   0  63 |   b = Tertiary
  0   0   0 228 |   c = Primary
  0   0   0 984 |   d = NONE
```

Fig. 6.   Run Information for ZeroR

VIII. RESULT DISCUSSION

Table 1 shows the tabulation of various results obtained from the three classifier used in this work while figure 7 is the graphical representation of the results.

TABLE I.   TABULATED RESULT

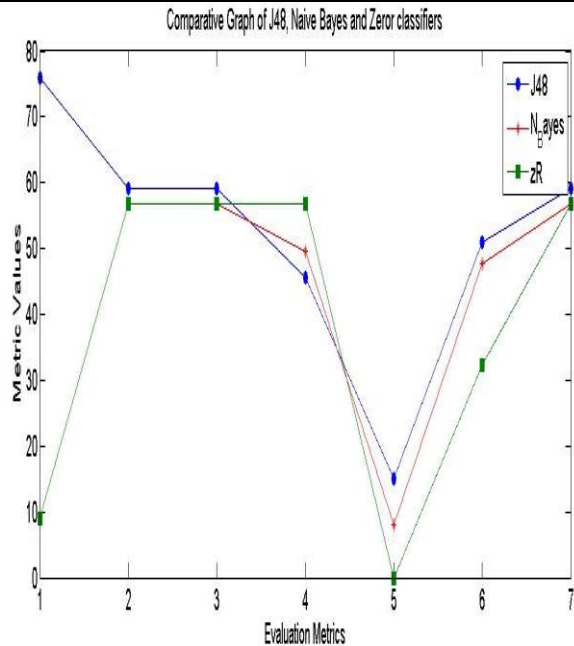| Evaluation Metrics	J48	Naïve Bayes	ZeroR
Time	0.76 Secs	0.09 Secs	0.09 Secs
Accuracy	59.15%	56.78%	56.78%
`TP Rate	0.591	0.568	0.568
FP Rate	0.456	0.496	0.568
Kappa	0.15	0.0813	0
Precision	0.51	0.478	0.322
Recall	0.591	0.568	0.568

Fig. 7.   Graph of the three Classifiers

The study shows that the J48 classifier has higher accuracy of 59.15 while both Naïve Bayesian and ZeroR classifier has accuracy of 56.78 each. The J48 though took more time of 0.76 seconds to build the model compare to 0.09 seconds each for both Naïve Bayesian and ZeorR classifier, where time is not the main metric for evaluation of the performance, the j48 classifier can be said to have performed better than Naïve Bayesian and ZeroR classifiers.

IX.   CONCLUSION

The advancement in data mining has been accompanied with development of various mining techniques and algorithms. Choosing the right technique for a particular type of data mining task is now becoming difficult. The best way is to perform a particular task using different techniques in order to choose the one that gives the best result. This work performed a comparative analysis of three classification techniques J48, Naïve Bayesian and zeroR to see which one that will give the best result using real crime data collected from some selected Nigerian prisons. There by proposing a frame work for choosing a better algorithm for data mining tasks. The J48 seems to have performed better than Naïve Bayesian and ZeroR classifiers using crime dataset and thus can be recommended for the classification of crime data. However, further work can be carried out using a different dataset and other classification techniques in WEKA mining tool or any other mining tool.

REFERENCES

[1]   J. Naisbitt " *Megatrends"*, 6th ed., Warner Books, New York. 1986.

[2]   T. ZhaoHui and M. Jamie "Data Mining with SQL Server 2005",Wiley Publishing Inc, Indianapolis, Indiana, 2005.

[3]   H. Jiawei, K. Micheline, and P. Jian "Data mining: Concept and Techniques" 3rd edition, Elsevier, 2011.

[4]   M. Goebel and L.Gruenwald "A survey of data mining and knowledge discovery   software tools",   ACM SIGKDD Explorations Newsletter, v.1 n.1, p.20-33, 1999.

[5]   Aman Kumar Sharma, Suruchi Sahni, "A Comparative Study of Classification Algorithms for Spam Email Data Analysis", IJCSE, Vol. 3, No. 5, 2011, pp. 1890-1895

[6]   Anshul Goyal, Rajni Mehta, "Performance Comparison of Naive Bayes and J48 Classification Algorithms", IJAER, Vol. 7, No. 11, 2012, pp.

[7]   S. K. Shabia and A. P. Mushtag " Evaluation of Knowledge Extraction Using Variou Classification Data Mining Techniqes", IJARCSSE, Vol. 3, Issue 6, pp. 251 – 256 , 2013.

[8]   I. Witten and E. Frank " Data mining: Practical Machine Learning Tools and Techniques with Java Implementations", San Francisco: Morgan Kaufmann   publishers, 2000.

[9]   Hong Hu, Jiuyong Li, Ashley Plank, "A Comparative Study of Classification Methods for Microarray Data Analysis", published in CRPIT, Vol. 61, 2006.

[10] Milan Kumari, Sunila Godara, "Comparative Study of Data Mining Classification Methods in cardiovascular Disease Prediction", IJCST, Vol. 2, Issue 2, pp. 304-308, 2011.