Students Academic Performance Prediction Based on Square Root Data Transformation and Ensemble Technique

Wokili Abdullahi
Department of Computer Science
Federal University of Technology, Minna
Minna, Nigeria
abdullahwokil@yahoo.com

Morufu Olalere
Department of Cyber Security Science
Federal University of Technology, Minna
Minna, Nigeria
lerejide@futminna.edu.ng

***Abstract*— Data mining research is evolving rapidly in the educational sector because of the vast amount of student information used to detect and explore useful patterns applicable to student learning behaviour. Predicting students' progress is an essential task in any educational institution. To assess student performance, educational institutions may use educational data mining to improve their teaching practices and learning processes. All these modifications lead to enhancing the success of students and overall academic results. In data mining, classification is a popular technique that has been widely tested out to find student outcomes. An approach based on data transformation and the Ensemble method to predict student success is suggested in this report. The efficacy of the student's predictive model is measured using several classifiers: Error-Correcting Output Code (ECOC), K-Nearest Neighbour (KNN), Ensemble, Naïve Bayesian (NB), and Decision Tree (DT). The results obtained by training the different classifiers with square root transformed features improved the classification accuracy from 83% to 86%, thus improving the performance prediction model's overall performance. For the X-API dataset, this suggested technique also created a better prediction accuracy than related works that used the same dataset.**

Keywords—*Student Performance Prediction, Data Transformation, Educational Data mining, Ensemble, classification*

I. Introduction

In Computer Science, one of the active fields is data mining. Data mining deals with the process of extracting valuable information from raw data [1]. Data mining is crucial due to the rising amount of data and the immediate need to translate these data into practical information. With data mining, a search engine could be used to examine vast volumes of information and instantly report meaningful findings without requiring human participation [2]. The educational sector is a significant area in which data mining is gaining increasing interest. Data mining is referred to as Educational Data Mining (EDM) in the education field. EDM emphasizes that useful knowledge is obtained from educational information systems such as the course management systems, registration systems, online learning management systems, and application systems. This mined knowledge can help students at each stage of their studies, like primary to tertiary education [3]. Many user groups are interested in EDM, and these users use the data that EDM has found according to their vision and intent [4]. For example, educational data's hidden pattern can help educators develop teaching techniques, understand learners, strengthen the learning experience, and use them to boost their learning activities [5]. This secret perception will also help the administration make the necessary decisions to achieve high-quality results [6]. Educational information is obtained from multiple sources, such as educational institution databases, e-learning services and traditional surveys [3]. Predicting the academic success of students is a significant application of EDM. In the educational environment, the analysis and estimation of student performance is an integral aspect. This prediction task foresees the importance of an unknown variable that distinguishes students with outcomes such as pass or failure, grades and marks [7].

Numerous data mining techniques can be used for EDM, including Ensemble, Artificial Neural Networks (ANN), discriminate analysis, decision tree, rule induction, support vector machine, Naïve Bayes, and K-nearest neighbour [8]. A predictive classification model's quality is determined by its ability to identify unknown patterns correctly. The X-API dataset is an educational data set that several researchers have used to predict student academic performance. However, previous works by Francis and Babu [9], Amrieh et al. [5], Tuaha et al. [3] and Amrieh et al. [10] have provided accuracies of less than 83% for the prediction of the X-API dataset. Five classification algorithms were used in this research: K-Nearest Neighbour (KNN), Decision Tree (DT), Error-Correcting Output Codes (ECOC), Naïve Bayes (NB) and Ensemble.

The proposed approach's main objective is to develop the ensemble classification model based on transformed square root features that classify students' performance as low-level, middle-level, and high-level. The significant contribution of this paper consists of:

1. Presentation of a method for student performance prediction.
2. Comparative experimentation of different classifiers trained with transformed and untransformed features.

The arrangement of these studies is as follows: section two presents the relevant works, section three presents the methods used, section four presents the findings and discussion, and finally, conclusions were drawn in section five, and suggestions for future works were presented in section six.

II. Related Work

Students' viability of progress is essential to predict student performance. The significance of predicting student performance has led researchers to become more and more interested in this field. Therefore, various researches have been published to predict students' performance.

A classification model for the prediction of student performance was built by Salal, Abdullaev and Kumar [11] using a dataset of 649 examples with 33 attributes obtained from 2 Portuguese high schools: Gabriel Pereira and Mousinho da Silveira High School. The dataset includes features, such as academic, demographic and social attributes of students. The classification target class ranged from 0 to 20, rendering the classification process extremely difficult as there were only 649 examples to be trained and assessed. Based on the initial class ranges, the target class was reduced to 6 categories due to this complexity. In WEKA software, the correlation assessment, gain ratio, and information gain were used as evaluation techniques, and these new target groups were used to pick attributes. After obtaining the outcome of the attribute selection algorithms' outcome, ten different attributes were selected, which were checked to influence the prediction outcome significantly. Eight classifiers, namely the Naïve Bayes, Random Tree, REP Tree, Decision Tree, Simple Logistics, One R, and Zero R, were fed with these selected classification attributes. One R was identified to have performed better with an accuracy of 76.7334% compared to the other seven classifiers with lower accuracy value. A comparative overview of a relatively large number of classifiers was provided by the study, offering an in-depth understanding of an extensive range of techniques. In this paper, each of the methods' performance was evaluated based only on accuracy without considering other performance metrics, which could say a lot about the suitability of a technique. The classification accuracy achieved was also low, unlike similar works that used the same dataset.

Iyanda et al. [12] conducted a comparison between two Neural Networks (NN) (generalized regression NN and multilayer perceptron) to determine the best model for student academic performance prediction based on only the educational feature of the student. The dataset used was collected from the Department of Computer Science and Engineering of the Awolowo Nigeria University of Obafemi. The data collected constitutes the academic record of learners (raw scores for each course taken) as the input variable, and the associated GPA as the output parameter. Using mean square error, receiver operating features, and accuracy, the two NN models' performance was evaluated. The generalized regression NN proved to perform better with an accuracy of 95% than the multilayer perceptron. However, without considering how demographic, social, and behavioural attributes could affect a student's output, this research used only student academic attributes for prediction.

Olalekan, Egwuche, and Olatunji [13] adapted Bayes' theorem and ANN to construct a predictive model for students' graduation probability at a tertiary institution. Four variables were used for prediction: Unified Tertiary Matriculation Test, Number of Sessions at the high school level, Grade Points at the high school level and Entry Mode. The data used was collected from the Computer Science School, Federal Polytechnic, Ile-Oluji, in Ondo State, Nigeria. The data were composed of 44 examples with five attributes. The study concludes that the ANN has a 79.31% higher performance accuracy than the 77.14% obtained by the Bayes classification model. The ANN precision improved as the hidden layers increased. As compared to other previous works, the overall accuracy in this study was low because of the small size of data used. Expanding the data size would help enhance the accuracy of the classification of the model.

Magbag and Raga [14] focused on building a model to predict first-year students' academic success in tertiary education. This research aimed to allow early intervention to help students stay on course and reduce non-continuance. The data utilized in this paper were obtained from three higher education institutions in Central Luzon, primarily in the cities of Angeles, San Fernando and Olongapo. The study subjects included first-year students from 8 academic departments from 2018-2019; Arts and Sciences, Engineering and Architecture, Computer Studies, Criminology, Education, Hospitality and Tourism, Business and Accountancy, Nursing and Allied Medical Sciences. The dataset was composed of 4,762 examples. The dataset was pre-processed, and missing values were deleted, leaving 3,466 available samples. Using Correlation-based Feature Selection, Gain Ratio and Information Gain for feature rating, feature selection was carried out. Using these selected features, the NN and logistic regression models were trained and evaluated. In comparison with similar works, the scale of the dataset used rendered the scheme more robust. However, the accuracy of 76% achieved in this analysis is low.

A new prediction algorithm to determine students' progress in academia using a hybrid (classification and clustering) Francis and Babu [9] proposed a data mining technique. The analysis used information from X-API education obtained from the kaggle repository consisting of 16 attributes with 480 instances. The dataset characteristics are demographic, academic, behavioural, and additional attributes (parent school satisfaction, student absentee days and parent response survey). Using classifiers such as SVM, Naïve Bayes, Decision tree, and NN, feature selection experiments were performed.

The selection of attributes was based on the accuracy provided by each classifier after the demographic, academic, behavioural and extra attributes were trained separately. Compared to using behavioural characteristics alone, additional features alone, educational features alone or demographic features alone the academic + behavioural + extra features provided a higher classification accuracy. These selected features were used as input for K-mean clustering and the majority vote approach. When applied to the dataset's academic, behavioural, and additional features, the proposed hybrid approach achieved an accuracy of 75.47%. However, related works using the X-API education dataset achieved greater accuracy of approximately 82% compared to this study.

III. METHODOLOGY

The methods used to carry out this research work are discussed in this section. Fig. 1 demonstrates the methods and processes that have been used to achieve the purpose of this study. Each of the measures shown in Fig. 1 is discussed in the sub-sections below.

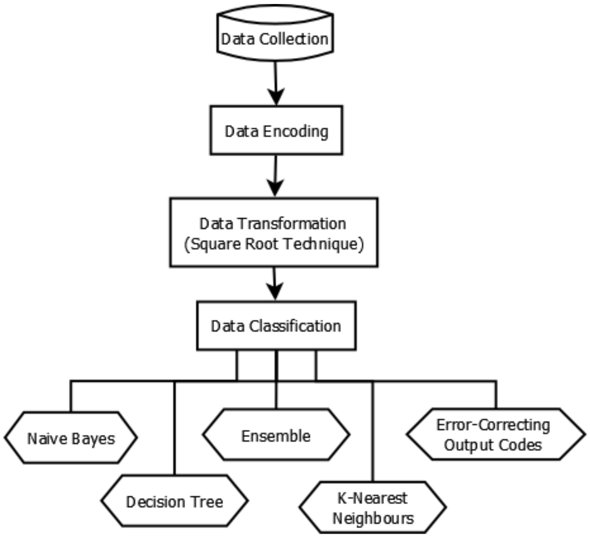

Fig. 1 Proposed System

A. Dataset

The data utilized in this research was gotten from Kaggle.com. The data is called X-API as it was collected from Kalboard 360 E-Learning system using eXperience API (X-API). The dataset is multivariate with 480 instances, 16 attributes and no missing values. The attributes are grouped into three major categories:

1. Demographic features, such as nationality and gender.
2. Academic background features such as grade level, educational level, and section.
3. Behavioural features such as viewing resources, raised a hand in class, school satisfaction, and parents' answering survey.

The dataset composes of 175 females and 305 males. The students came from various countries such as 172 students from Jordan, 179 students from Kuwait, 22 students from Iraq, 4 students from Morocco, 28 students from Palestine, 17 from Lebanon, 11 from Saudi Arabia, 12 from Tunis 9 students from Egypt, 7 from Syria, 6 from USA, Iran and Libya, and one student from Venezuela. The dataset was collected over two academic semesters: 245 student records were compiled in the first semester, and 235 student records were collected in the second semester. This dataset also contains a new category of features; this feature is parent parturition in the educational process. Parent participation feature has two sub-features: Parent Academic Satisfaction and Parent Answering Survey. Two hundred seventy (270) parents answered the survey, and 210 are not, 292 of the parents are satisfied with the school, and 188 are not. This dataset was used by [9], [3] [10] and [5]. The X-API dataset features and the description of these features are is presented in Table 1.

TABLE 1 FEATURES OF X-API DATASET AND THEIR CATEGORIES

S/ No	Attribute	Attribute Description	Data Type	Attribute Category
1	Gender	Student Gender (Male or Female)	Categorical	Demographic Attributes
2	Nationality	Nationality of the student (Lebanon, Kuwait, Egypt, USA, Saudi-Arabia, Jordan, Iran, Venezuela, Tunis, Syria, Morocco, Palestine, Lybia, Iraq)	Categorical	
3	Place of Birth	(Lebanon, Kuwait, Egypt, USA, Saudi-Arabia, Jordan, Iran, Venezuela, Tunis, Syria, Morocco, Palestine, Lybia, Iraq)	Categorical	
4	Parent Responsible	The parent who is responsible for the student (Mom or Dad)	Categorical	
5	Educational Levels	The educational level a student belongs to (lower-level, Middle-School, High-School)	Categorical	Academic Attributes
6	Section ID	The classroom a student belongs to (A, B or C)	Categorical	
7	Course	Offered courses (Spanish, English, French, IT, Arabic, Chemistry, Maths, Biology, History, Science, Quran, Geology)	Categorical	
8	Student Semester	Student school semester (First or Second)	Categorical	
9	Student Grade	The grade category student belongs (G-01, G-02, G-03, G-04, G-05, G-06, G-07, G-08, G-09, G-10, G-11, G-12)	Categorical	
10	Student punctuality to class	Amount of days of absence of a student in the class (above-7 or under-7)	Categorical	
11	Raising of Hand	Number of times a student raised their hands (0-100)	Integer	Behavioural Attributes
12	Number of visited resources	The number of times a student visited a course content (0-100)	Integer	
13	Announcements viewed	The number of time the student checks a new announcement (0-100)	Integer	
14	Discussion Group	The number of time the student participated in discussion groups (0-100)	Integer	
15	Parents Answering Survey	If the parent answered the surveys provided	Categorical	Extra Attributes

		by the school (Yes or No)		
16	Satisfaction of Parent	If parents are satisfied with the school (Yes or No)	Categorical	
17	Class	The students are categorized into three numerical intervals based on their total grade (Low-level (0-69), Middle-level (70-89), or High-level (90-100)	Categorical	Target Class/Attribute

B. Data Encoding

There are both numeric variables and categorical variables in the dataset used. In this phase, the categorical data types of attributes were converted to numeric attributes. Data encoding was done because specific machine learning algorithms such as Naïve Bayes, vector machine support and Ensemble need numeric attribute types to work. In dealing with numeric data types, machine learning models have also proven to be efficient. The label encoding technique was employed in this research. Each label was converted to an integer value. For instance, the gender, which is in a categorical data form (Male and Female), was encoded to integer 1 and 2. Table 1 and Table 2 indicate gender encoding and target class encoding, respectively.

TABLE 2 ENCODING GENDER

Gender (categorical)	Gender (integer)
Male	1
Female	2

TABLE 3 ENCODING TARGET CLASS

Target Class (categorical)	Target Class (integer)
High-Level (90-100)	1
Middle-Level (70-89)	2
Low-Level (0-69)	3

C. Data Transformation

Transformations of the data can reduce the skewness of data and the effect of outliers in the data. Transformation approaches include centring, scaling, removal of skewness, and binning. This study used the square root transformation technique to convert a skewed distribution into a normal/less-skewed distribution. The square root of all the predictor variables was derived. Square roots that were obtained were then used to train the classifiers as inputs. In equation 1, the square root transformation formula is provided. The square root of a number A is a number B such that:

$$B^2 = A \tag{1}$$

D. Data Classification

Machine learning capability lies in its ability to generalize by correctly classifying unknown information based on models developed using the training dataset. Several machine learning classification models were used for training and classification, namely, Error-Correcting Output Codes (ECOC), Naïve Bayes (NB), Decision Tree (DT), Ensemble, and K-Nearest Neighbor. In this research work students were grouped into three numerical intervals according to their overall grade:

1. Low-Level: ranges from 0 to 69,
2. Middle-Level: ranges from 70 to 89,
3. High-Level: ranges from 90-100.

Each of the five classifiers was trained to classify students into the three classes with the square root transformed data and the data that was not altered. 80% of the data was used for training, and the remaining 20% was used to test the trained models. These five classes are presented below.

1) Error-Correcting Output Codes (ECOC)

Machine learning models are built for binary classification problems, such as Support Vector Machine (SVM) and logistic regression. As such, these binary algorithms either need to be updated or not used at all for multiclass classification problems. The ECOC technique is a tool that allows the issue of multiclass classification to be interpreted as multiple problems of binary type, enabling the direct use of native binary classification models[15]. The ECOC enables the encoding of an infinite number of binary classification problems for each class [16]. ECOC designs are independent of the classifier depending on the implementation. ECOC has error-correcting properties and has shown that the learning algorithm's bias and variance can be decreased [17].

Given a classification problem with Y_c The key aim of ECOC is to create a binary or ternary "codeword" for each class. The codewords are arranged as rows of a matrix X. *Codematrix* X is defined in equation 2.

$$X \in \{-1,0,+1\}^{Y_c\ x\ L} \tag{2}$$

Where L is the code length. From the learning point of view X specifies Y_c classes to train L dichotomizes, $a, a_2, \ldots, a_L$. A classifier a_1 is trained according to the column $X\ (.\,,l)$.

2) Naive Bayes (NB)

The NB classifier is a probabilistic machine learning model based on the Bayes theorem's use with assumptions of high independence between the features. NB is used for a classification task. To predict the type of test data set, NB is fast, convenient and straightforward. In multiclass forecasting, it also suits nicely [18]. When assuming independence, a Naive Bayes classifier performs better than other models, such as logistic regression, and less data for training is needed. However, the theory of independent predictors is an important limitation of NB [19]. The Bayes theorem provides a way for P(a|b)from (a), P(b) and P(b|a) to measure the posterior likelihood. In equations 3 and 4, the posterior probability is shown in the formula. Bayes theorem provides a way of calculating posterior probability $P(a|b)$ from $P(a)$, $P(b)$ and $P(b|a)$. The posterior probability is formula is shown in equation 3 and 4.

$$P(a|b) = {p(b|a) \times P(a)}/{p(b)} \tag{3}$$

$$P(a|b) = \frac{P(b_1|a) \times P(b_2|a) \times \ldots \times P(b_n|a) \times P(a)}{P(b_1, \ldots, b_n)} \tag{4}$$

Where P(a|b) is the posterior likelihood of class (a, target) given predictor (b, attributes). P(a) is the prior probability of class. p(b|a) is the likelihood of a predictor given a category. $p(b)$ is the prior likelihood of a predictor.

3) Decision Tree (DT)

A DT is a simple and commonly used predictive modelling technique. DT is a type of supervised learning where, according to a particular parameter, the data is continually split [20]. The decision tree uses a tree-like model to go from observations on an item (represented in the branches) to conclusions on the target value of an item (defined in the leaves) [21]. Regression and classification problems can be solved using the DT algorithm. DT is easy to understand and view. It does not require normalization of data and preparation of data; it needs less effort. The decision to do strategic splits has a significant effect on a tree's precision [22]. Entropy, information gain and reduction invariance are techniques used in determining which attribute to the position at the root or the different levels of the tree.

Entropy is a measure of randomness in processed information. The larger the entropy, the more challenging it is to draw any conclusions from that data. A branch with an entropy of zero, for example, is chosen as the root node, and further division is required for a branch with an entropy greater than zero [22]. In equation 5, entropy for a single attribute is expressed.

$$E(S) = \sum_{i=1}^{n} -p_i log_2 \, p_i \qquad (5)$$

Where S is the current state, p_i is the probability of an event i of state S.

Information Gain (IG) is a statistical property that tests how well the training examples are segregated according to their target classification by a given attribute. In equation 6, information gain is expressed mathematically.

$$IG = Entropy(before) - \sum_{j=1}^{N} Entropy\,(j, after) \qquad (6)$$

Where "before" is the dataset before the split, N is the number of subsets generated by the division, and (j, after) is subset j after the division.

Reduction invariance is an algorithm that is used for problems with regression. This algorithm uses the standard formula of variance to select the best split. As the criterion to divide the population, the split with lower variance is chosen. In equation 7, the standard variance formula used in this technique is represented.

$$variance = \frac{\sum(X-\mu)^2}{n} \qquad (7)$$

Where $\mu$the mean of the values and X is the actual value and n is the number of values.

4) K-Nearest Neighbour (KNN)

The K-Nearest Neighbours (KNN) algorithm is a non-parametric supervised machine learning algorithm used to solve both classification and regression problems [23]. The KNN algorithm assumes the closeness of related objects. In KNN, an item is grouped by its neighbours' majority vote, with an object being assigned to the most common class of its k-nearest neighbours [24]. KNN does not need a training phase. KNN, however, suffers from the curse of dimensionality, and it is vulnerable to outliers. The Euclidean distance is a commonly used similarity measure in KNN [25]. The Euclidean distance is the linear distance between two points in Euclidean space. In equation 8, the Euclidean distance is expressed.

$$D(p, q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (8)$$

Where p, q are two points in Euclidean n-space, $q_i \; and \; p_i$ are the Euclidean vectors, starting from the origin of the space and n is the n-space.

5) Ensemble Classifier

An ensemble learning model combines predictions from multiple models with a two-fold goal: the first objective is to maximize prediction accuracy compared to a single classifier[5]. The second gain is more critical generalizability due to multiple advanced classifiers. As a result, solutions, where a single prediction model would have problems, can be discovered by an ensemble. A key rationale is that an ensemble can select a set of hypotheses out of a much larger hypothesis space and combine their predictions into one [26]. Via voting or weighted voting of their forecast for the final estimates, classifiers in the ensemble learning model are merged into meta-classifiers [26].

E. Performance Metrics

In this study, five performance measures were used to evaluate the proposed method. These measures are explained below.

1. **Precision:** This is a measure that computes the number of positive predictions made that are accurate. It is determined as the proportion of positive instances correctly predicted, divided by the total number of positive cases predicted. Precision is mathematically represented in equation 9.

$$Precision = \frac{Truepositives}{Truepositives + Falsepositives} \qquad (9)$$

2. **Recall:** This indicator evaluates the amount of correct positive predictions that could have been made out of all positive predictions. The formula in equation 10 represents recall.

$$Recall = \frac{Truepositives}{Truepositives + Falsenegatives} \qquad (10)$$

3. **F-Score:** this is the harmonic mean of precision and recall. The formula in equation 11 represents F-Score.

$$\mathrm{F-Score} = 2 * \frac{\mathrm{precision} * \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \qquad (11)$$

4. **Accuracy:** Accuracy can be defined as the rate of correct classifications. Accuracy is calculated in equation 12.

$$\mathrm{ACC} = \frac{\text{True Positive + True negative}}{\text{True Positive + True negative + False Positive + False negative}} \qquad (12)$$

5. **Receiver operating characteristic (ROC) curve:** is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate (recall) and False Positive Rate.

IV. RESULTS AND DISCUSSION

In this work, experiments were conducted on five algorithms: Decision Tree (DT), KNN, Ensemble, ECOC and NB classifiers for features without data transformation and transformed (square root) features. Two kinds of experiments were conducted, which are:

1. Classification of student performance using data transformed features.

2. Classification of student performance based on normal features (Features without data transformation).

The outcomes of the two experiments conducted using the five classifications techniques mentioned above are shown in Table 4 and Table 5.

TABLE 4 CLASSIFICATION WITH FEATURES WITHOUT DATA TRANSFORMATION

No Feature transform										
Algorithm	***Precision (class A)***	***Precision (class B)***	***Precision (class C)***	***Recall (class A)***	***Recall (class B)***	***Recall (class C)***	***F-Score (class A)***	***F-Score (class B)***	***F-Score (class C)***	***Accuracy***
ECOC	0.7051	0.8462	0.5750	0.6667	0.8148	0.6216	0.6857	0.8302	0.5974	0.6900
Ensemble	0.7941	0.8846	0.8250	0.8182	0.9200	0.7857	0.8060	0.9020	0.8049	0.8300
KNN	0.4412	0.7692	0.7250	0.6522	0.8696	0.5370	0.5263	0.8163	0.6170	0.6400
NB	0.7642	0.8846	0.5750	0.6667	0.8519	0.6765	0.7123	0.8679	0.6216	0.7200
DT	0.6176	0.7692	0.6750	0.7000	0.7692	0.6136	0.6553	0.7692	0.6429	0.6800

In this study, student performance is classified into three classes: low-level, Middle-level and high-level. The low-level is represented as class A, middle-level is represented as class B and high-level is defined as class C. The precision, recall and f-score of the three classes for each of the five classifiers trained with data without transformation are shown in Table 4. Table 4 shows that the ensemble method produced a higher classification accuracy of 83% than the other four classifiers. Ensemble classifier also had a better precision, recall and f-score for all the three target classes than the other four classifiers. .Fig. 2 presents a ROC curve for curve comparing the DT, NB, ECOC, Ensemble and KNN classifier trained with data that were not transformed.

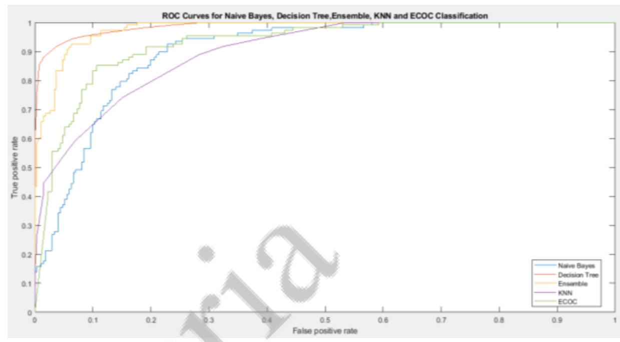

Fig. 2 ROC Curve comparing the DT, NB, ECOC, Ensemble and KNN classifier performances trained with data without transformation.

TABLE 5 CLASSIFICATION PERFORMANCE WITH TRANSFORMED FEATURES

Feature transform										
Algorithm	***Precision (class A)***	***Precision (class B)***	***Precision (class C)***	***Recall (class A)***	***Recall (class B)***	***Recall (class C)***	***F-Score (class A)***	***F-Score (class B)***	***F-Score (class C)***	***Accuracy***
ECOC	0.6970	0.9286	0.7949	0.7931	0.9286	0.7209	0.7419	0.9286	0.7561	0.8000
Ensemble	0.9091	0.9643	0.7436	0.8108	0.8710	0.9063	0.8571	0.9153	0.8169	0.8600
KNN	0.6667	0.9643	0.6923	0.7857	0.7941	0.7105	0.7213	0.8710	0.7013	0.7600
NB	0.9091	0. 9643	0.6410	0.7692	0.8438	0.8621	0.8333	0.9000	0.7353	0.8200
DT	0.8485	0.8929	0.6667	0.7368	0.8621	0.7879	0.7887	0.8772	0.7222	0.7900

From the classification result in Table 5, the ensemble method produced a higher classification accuracy of 86% compared to the other four classifiers. The Ensemble method also created a better precision, recall and f-score for all the three target classes than the other four classifiers.
Fig. 3 presents a ROC curve for curve comparing the DT, NB, ECOC, Ensemble and KNN classifier trained with a transformed feature set.

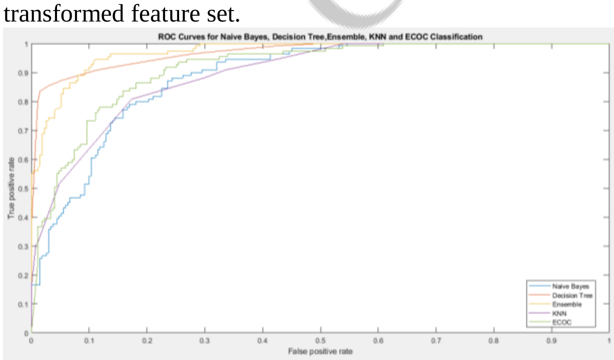

Fig. 3 ROC Curve comparing the DT, NB, ECOC, Ensemble and KNN classifier performances trained with untransformed features

Table 6 compares the accuracy of NB, DT, KNN, ECOC and Ensemble after being trained with untransformed features and transformed features. Based on the result shown in Table 6, each of the five classifiers performed better when prepared with the transformed features. Ensemble method achieved an accuracy of 86% when trained with the transformed features and achieved an accuracy of 83% when trained with untransformed features. NB achieved an accuracy of 82% when trained with the transformed features and achieved an accuracy of 72% when trained with untransformed features. ECOC, KNN and DT also achieved higher accuracy when trained with the transformed features.

TABLE 6 COMPARISON OF DT, KNN, ENSEMBLE, ECOC AND NB PERFORMANCE FOR UNTRANSFORMED FEATURES AND TRANSFORMED FEATURES

	ACCURACY	
Algorithm	***Untransformed features***	***Data Transformed Features***
ECOC	0.6900	0.8000
Ensemble	0.8300	0.8600
KNN	0.6400	0.7600
NB	0.7200	0.8200
DT	0.6800	0.7900

Fig. 4 shows a comparison of DT, KNN, Ensemble, ECOC and NB accuracy when trained with untransformed features and transformed features.

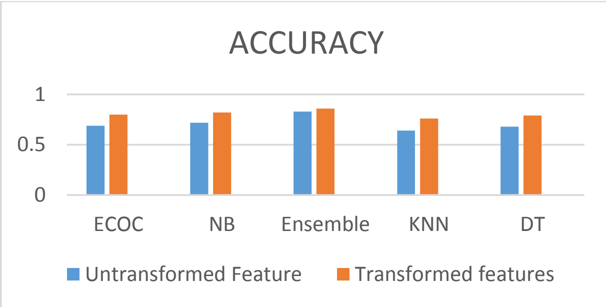

Fig. 4 Comparison of DT, KNN, Ensemble, ECOC and NB performance for untransformed features and transformed features

Table 7 and Fig. 5 compare the proposed method with related works that used the X-API dataset. From the results in Table 7, it can be seen that this study produced a better classification accuracy of 86% for the X-API dataset when compared with previous works.

TABLE 7 COMPARISON OF THE PROPOSED METHOD WITH RELATED WORKS

Algorithm	Dataset	Accuracy (%)
Ensemble (Proposed Method)	X-API	**86.0**
Artificial Neural Network [3]	X-API	78.1
Artificial Neural Network [10]	X-API	73.8
Decision Tree [5]	X-API	82.2
Clustering + Decision Tree [9]	X-API	75.5

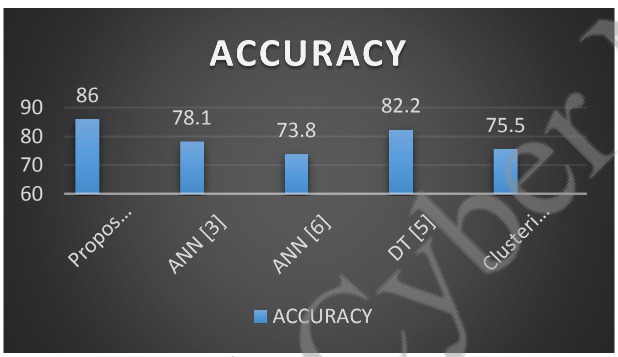

Fig. 5 Comparison of the proposed method with related works

V. CONCLUSION

This study performed a comparative result for five classifiers: KNN, DT, Ensemble, NB and ECOC in respect to student performance prediction for X-API dataset. The proposed method obtained a higher classification accuracy than previous works that used the X-API dataset. From this research, it can be established that the application of square root transformed features for training classifiers can improve the classification accuracy. Square root transformation reduces right skewness, and it also has the advantage that it can be applied to zero values. In conclusion, a system was developed which can accomplish student academic performance prediction.

VI. FUTURE WORKS

Only the square root transformation method was in this study. For future work, more transformation techniques could be applied to evaluate their effect on classification accuracy. In this study, only the X-API dataset was used. Other datasets may be considered to enhance the model robustness. Experiments may also be carried out using more data mining techniques such as genetic algorithms and discriminate analysis model.

REFERENCES

[1] S. Hussain, N. Abdulaziz Dahan, F. M. Ba-Alwi, and N. Ribata, “Educational Data Mining and Analysis of Students’ Academic Performance Using WEKA,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, p. 447, Feb. 2018, doi: 10.11591/ijeecs.v9.i2.pp447-459.
[2] S. Rajagopal, “Customer Data Clustering using Data Mining Technique,” vol. 3, no. 4, 2011, doi: 10.5121/ijdms.2011.3401.
[3] S. Tuaha, I. F. Siddiqui, and Q. Ali Arain, “Analyzing Students’ Academic Performance through Educational Data Mining,” *3C Tecnol. Innov. Apl. Pyme*, pp. 402–421, May 2019, doi: 10.17993/3ctecno.2019.specialissue2.402-421.
[4] C. Romero and S. Ventura, “Educational Data Mining: A Review of the State of the Art,” *IEEE Trans. Syst. Man Cybern. Part C, Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010, doi: 10.1109/TSMCC.2010.2053532.
[5] E. A. Amrieh, T. Hamtini, and I. Aljarah, “Mining Educational Data to Predict Student’s academic Performance using Ensemble Methods,” *Int. J. Database Theory Appl.*, vol. 9, no. 8, pp. 119–136, Aug. 2016, doi: 10.14257/ijdta.2016.9.8.13.
[6] M. B. Shah, M. Kaistha, and Y. Gupta, “Student Performance Assessment and Prediction System using Machine Learning,” in *2019 4th International Conference on Information Systems and Computer Networks*, Mathura, India, Nov. 2019, pp. 386–390, doi: 10.1109/ISCON47742.2019.9036250.
[7] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, “Student Academic Performance Prediction using Supervised Learning Techniques,” *Int. J. Emerg. Technol. Learn. IJET*, vol. 14, no. 14, p. 92, Jul. 2019, doi: 10.3991/ijet.v14i14.10310.
[8] L. M. Abu-Zohair, “Prediction of Student’s performance by modelling small dataset size,” *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, p. 27, Dec. 2019, doi: 10.1186/s41239-019-0160-3.
[9] B. K. Francis and S. S. Babu, “Predicting Academic Performance of Students Using a Hybrid Data Mining Approach,” *J. Med. Syst.*, vol. 43, no. 6, p. 162, Jun. 2019, doi: 10.1007/s10916-019-1295-4.
[10] E. A. Amrieh, T. Hamtini, and I. Aljarah, “Pre-processing and analyzing educational data set using X-API for improving student’s performance,” in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Amman, Jordan, Nov. 2015, pp. 1–5, doi: 10.1109/AEECT.2015.7360581.
[11] Y. K. Salal, S. M. Abdullaev, and M. Kumar, “Educational Data Mining: Student Performance Prediction in Academic,” vol. 8, no. 4, p. 6, 2019.
[12] A. R. Iyanda, O. D. Ninan, A. O. Ajayi, and O. G. Anyabolu, “Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 10, no. 6, pp. 1–9, Jun. 2018, doi: 10.5815/ijmecs.2018.06.01.
[13] A. M. Olalekan, O. S. Egwuche, and S. O. Olatunji, “Performance Evaluation Of Machine Learning Techniques For Prediction Of Graduating Students In Tertiary Institutions,” in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Ayobo, Ipaja, Lagos, Nigeria, Mar. 2020, pp. 1–7, doi: 10.1109/ICMCECS47690.2020.240888.
[14] A. Magbag and R. R. Jr, “Prediction Of College Academic Performance Of Senior High School Graduates Using Classification Techniques,” vol. 9, no. 04, p. 6, 2020.
[15] G. Armano, C. Chira, and N. Hatami, “Error-Correcting Output Codes for Multi-Label Text Categorization,” p. 12, 2013.
[16] T. G. Dietterich and G. Bakiri, “Solving Multiclass Learning Problems via Error-Correcting Output Codes,” *J. Artif. Intell. Res.*, vol. 2, pp. 263–286, Jan. 1995, doi: 10.1613/jair.105.
[17] S. Escalera, O. Pujol, P. Radeva, and P. Ivanova, “Error-Correcting Output Codes Library,” *J. Mach. Learn. Res.*, vol. 11, p. 4, 2010.
[18] P. Bhargavi, M. Tech, and D. S. Jyothi, “Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils,” p. 7, 2009.
[19] “Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques,” *Int. J. Sci. Res. IJSR*, vol. 5, no. 1, pp. 1842–1845, Jan. 2016, doi: 10.21275/v5i1.NOV153131.
[20] N. Patel and D. Singh, “An Algorithm to Construct Decision Tree for Machine Learning based on Similarity Factor,” *Int. J. Comput. Appl.*, vol. 111, no. 10, pp. 22–26, Feb. 2015, doi: 10.5120/19575-1376.

[21] K. David Kolo, S. A. Adepoju, and J. Kolo Alhassan, "A Decision Tree Approach for Predicting Students Academic Performance," *Int. J. Educ. Manag. Eng.*, vol. 5, no. 5, pp. 12–19, Oct. 2015, doi: 10.5815/ijeme.2015.05.02.
[22] A. S. Olaniyi, S. Y. Kayode, H. M. Abiola, S.-I. T. Tosin, and A. N. Babatunde, "STUDENT'S PERFORMANCE ANALYSIS USING DECISION TREE ALGORITHMS," p. 8, 2017.
[23] Z. Zhang, "Introduction to machine learning: k-nearest neighbours," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 218–218, Jun. 2016, doi: 10.21037/atm.2016.03.37.
[24] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 6, pp. 354–360, 2013.
[25] X. Gu, L. Akoglu, and A. Rinaldo, "Statistical Analysis of Nearest Neighbor Methods for Anomaly Detection," in *33rd Conference on Neural Information Processing Systems*, Canada, 2019, p. 11.
[26] O. W. Adejo and T. Connolly, "Predicting student academic performance using multi-model heterogeneous ensemble approach," *J. Appl. Res. High. Educ.*, vol. 10, no. 1, pp. 61–75, Feb. 2018, doi: 10.1108/JARHE-09-2017-0113.