IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Prediction of Student's Performance using Selected Classification Methods: A Data Mining Approach

[1]Abba Babakura; [2]Abubakar Roko; [3]Aminu Bui; [4]Ibrahim Saidu; [5]Jobson Ewalefoh

[1]Computer Science Unit, Department of Mathematics, UDUS, Sokoto, Nigeria

[2]Computer Science Unit, Department of Mathematics, UDUS, Sokoto, Nigeria

[3]Computer Science Unit, Department of Mathematics, UDUS, Sokoto, Nigeria

[4]Computer Science Unit, Department of Mathematics, UDUS, Sokoto, Nigeria

[5]Department of Political Science, UNISA, South Africa

Abstract - Educational Data Mining (EDM) research have emerged as an interesting area of research, which are extracting useful knowledge from educational databases for purposes such as predicting student's success. The extracted knowledge helps the institutions to improve their teaching methods and learning process. In this paper, we applied Decision Tree, Naïve Bayes and Neural Network classification methods for predicting the student's performance based on the grade level. This aim to resolve the problem of difficulty in predicting the performance of student's in institutions. The objectives of this paper are to (i) implement three classification methods independently on the student's performance dataset, and (ii) determine the best method among the three classification methods. The results shows that the Decision Tree produces the highest accuracy rate of 77.778%, followed by the Neural Network with accuracy rate of 70.886% and the Naïve Bayes produces the lowest at accuracy rate 66.865%. The result recommends that Decision Tree is used in predicting student's performance rather than Naïve Bayes and Neural Network.

Keywords- *Educational Data Mining, Prediction, Student performance, Decision Tree, Neural Network and Naïve Bayes*

1. Introduction

The ability of a system model to predict a student's performance is significantly important in educational environments. Student's academic performance is based upon some diverse factors like socio-economic, psychological and other environmental variables [1]. Knowledge about these related factors and their core effect on student performance is important in determining the final grades of each students. Recently, much attention has been paid to educational mining research.

Educational mining has emerged as very important area of research to reveal presentable and applicable knowledge from large educational data repositories. Data mining models are used to obtain the hidden information and desired benefits from these large repositories [2]. There is a critical demand for academic institutions to maintain and to integrate large datasets of learners for multipurpose decision making. However, predicting student's performance becomes more challenging due to the large volume of data in educational databases [3]. Figure 1 below shows the application of data mining in educational systems. Student performance evaluation system can help in decision making for predicting the grades levels of students in an academic year.

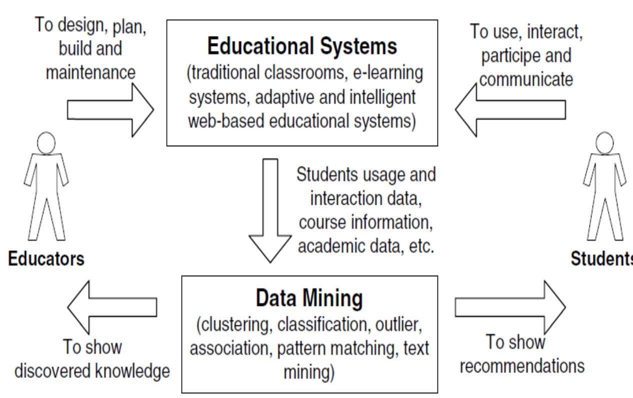


Fig. 1 Cycle of data mining application in educational systems.

In the past, mostly student performance is predicted by using different types of feature sets such academic record, family income, and family assets [4,5,6]. Family income and expenditure feature sets play an important role in student performance prediction. This paper evaluated the performance of three classification methods, which are support vector machine, decision tree and neural network for the diagnosis of the student performance dataset. Extensive experimentations are conducted to evaluate the success and failure of student at the end of the year selecting some feature sets which affect the student's performance significantly.

The main contributions of this paper are as follows:

1. In this paper, some feature sets are identified that significantly affect the performance of each student.
2. The performances of the models are tested with 3 runs of different validations to comprehensively experiment on the selected feature sets as compared to the traditional single run of test on the performance models.
3. The prediction will help the teachers to identify the weak students and help them to improve their performance.
4. The overall finding is that system models built based on some selected feature sets can improve prediction
5. of student's performance. The rest of the paper is arranged as follows: Section 2 describes the related work. This is preceded with development of methodological framework elaborated in section 3. The experimental results and discussion are outlined in section 4 and finally accompanied by conclusion in section 5.

2. Related Work

The research problem of student's performance prediction can be analyzed through diverse angles. In the current literature, a number of complimentary approaches provide a baseline for such an analysis.

In an ideal scenario, a rich dataset with student identity along with numerous characteristics could be the basis for predicting the performance of each student.

Predicting student's performance is an important task in web-based educational environments. To build a predictive model, there are several DM techniques used, which are classification, regression and clustering. The most popular technique to predict students' performance is classification.

The implementation of data mining methods and tools for analyzing data available at educational institutions, defined as Educational Data Mining (EDM) [7] is a relatively new stream in the data mining research. Extensive literature reviews of the EDM research field are provided by Romero and Ventura [7], covering the research done in the area between 1995 and 2005, and by Nithya [8], for the period after 2005. It is remarkable that most often attracting the attention of researchers and becoming the reasons for applying data mining at higher education institutions are focused mainly on retention of students, improving institutional effectiveness and enrolment management.

In the same direction, authors in [9] presented a model to predict student performance. They evaluate student success by passing grade at the exam. Parameters addressed for prediction including students' socio-demographic variables, achieved results from high school, the entrance exam, and attitudes towards studying which can have an effect on success Ramesh et al. [10] presented a valuable study to figure out factors influenced student success. They focused on parents' occupation and school type. Their obtained results from hypothesis testing reveals that type of school is not influence student performance and parents' occupation plays a major role in predicting grades [10].

An Association Rule based mining method is applied for selection of weak students in a school and is found effective [11]. Genetic Algorithm is used to assign the weights for the modeling of students' grade for three levels (binary, 3-level and 9-level) [12]. It shows that the combination of multiple classifiers leads to a significant improvement in classification. A model is proposed for predicting student performance using six machine learning techniques for distance learning education, which is quite different from the traditional educational system [13]. The experimental results show that demographic and performance features are better predictors for predicting student performance. A regression model is applied to predict the test score of subject for school students [14]. It concludes that mixedeffect models present best performance as compared to Bayesian network.

Khan [15] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a main objective to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream. The selection was based on cluster sampling technique in which the entire population of interest was divided into groups, or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socioeconomic status had relatively higher academic achievement in general.

Al-Radaideh et al [16] applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the Naïve Bayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models. However, predicting student's performance becomes more challenging due to the large volume of data in educational databases. This paper extensively experimented on the success and failure of student at the end of the year selecting some feature sets which affect the student performance significantly.

3. Methodological framework

In this section, we present the proposed framework in producing the prediction of student performance using the classification techniques. The framework outlined the steps involved in developing the models to predict the grade of student in an academic year.

The activities are detailed in the following subsections and the sequence of the activities is shown in figure 2.

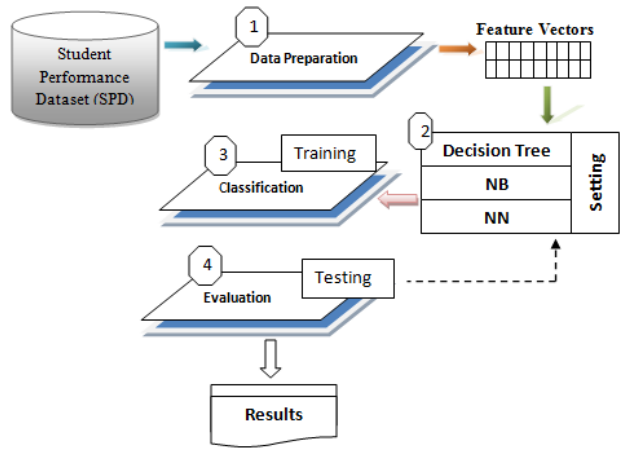


Fig. 2 The flow diagram of the research activities.

Figure 2 above describes the stages involved in the prediction of the student's performance which are namely- Data preparation, feature selections, setting up of the models, the training of the classifiers and the evaluation.

3.1 Dataset Description

This is an educational data set which is collected from learning management system (LMS) called Kalboard 360. Kalboard 360 is a multi-agent LMS, which has been designed to facilitate learning through the use of leading-edge technology (17). Such system provides users with a synchronous access to educational resources from any device with

internet connection. The dataset consists of 480 student records and 16 features. The features are classified into three major categories: (1) Demographic features such as gender and nationality. (2) Academic background features such as educational stage, grade Level and section. (3) Behavioral features such as raised hand on class, opening resources, answering survey by parents, and school satisfaction. The dataset is collected through two educational semesters: 245 student records are collected during the first semester and 235 student records are collected during the second semester. Table 1 shows the description on the student's performance dataset.

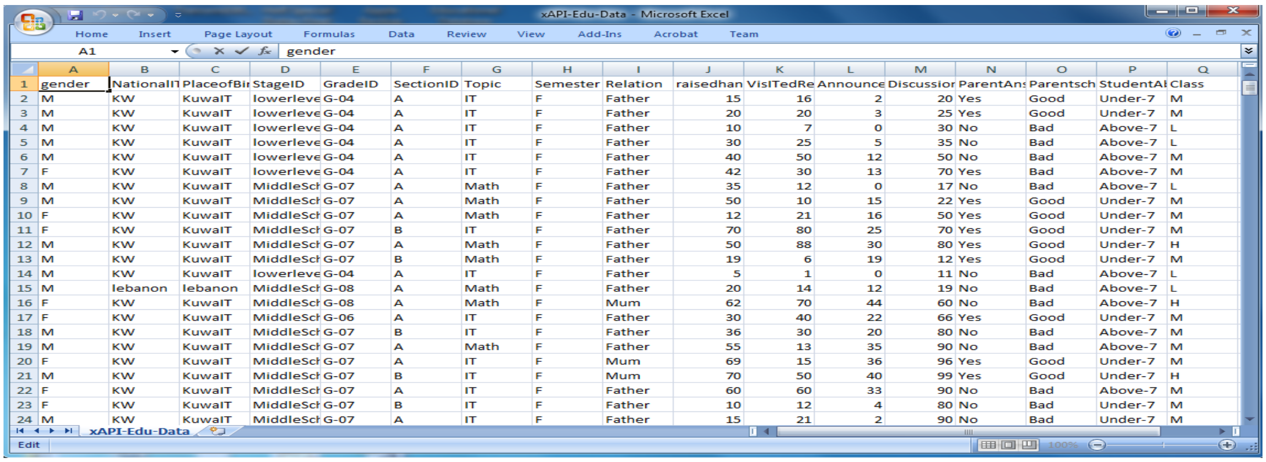
IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Table 1: Student Performance Dataset

Data characteristics	Multivariate	***Number of instances***	480
Feature characteristics	Real	***Number of features***	16
Purpose	Classification	***Number of missing values***	0

3.2 Data Preparation

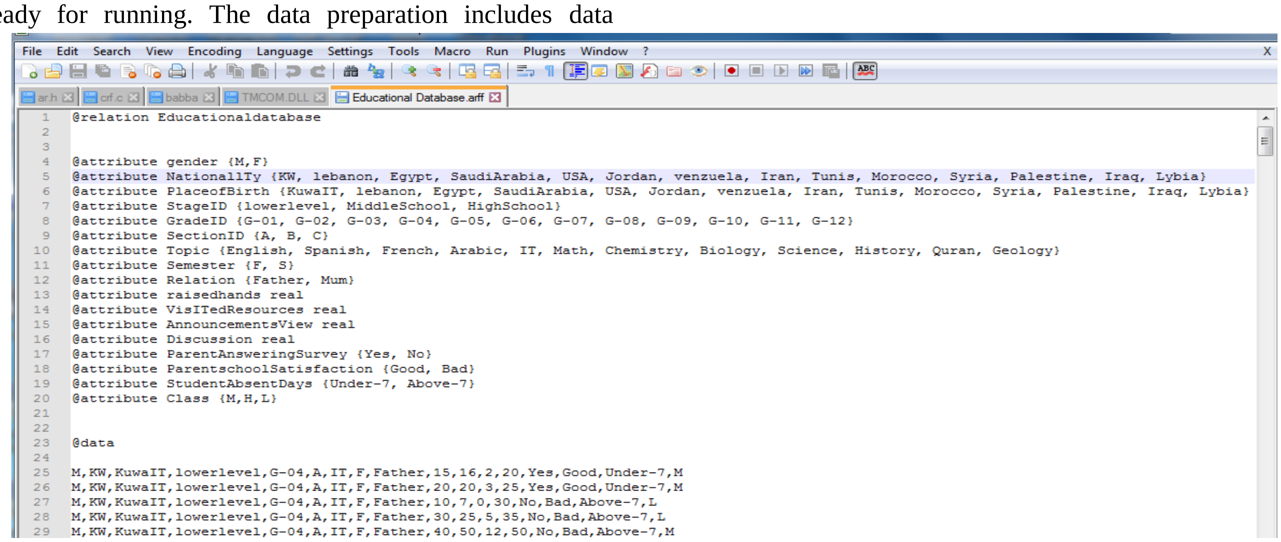
From our observation of the data and the literature review, the student performance dataset does not have observable patterns. The features are a combination of numerical and nominal values. The dataset is complete and has no missing values. Also, there is no noisy data found. The examination of the student's performance dataset is shown in figure 3 below:



	gender	NationalIT	PlaceofBir	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhan	VisITedRe	Announce	Discussior	ParentAns	Parentsch	StudentAt	Class
2	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	15	16	2	20	Yes	Good	Under-7	M
3	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	20	20	3	25	Yes	Good	Under-7	M
4	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	10	7	0	30	No	Bad	Above-7	L
5	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	30	25	5	35	No	Bad	Above-7	L
6	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	40	50	12	50	No	Bad	Above-7	M
7	F	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	42	30	13	70	Yes	Bad	Above-7	M
8	M	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	35	12	0	17	No	Bad	Above-7	L
9	M	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	50	10	15	22	Yes	Good	Under-7	M
10	F	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	12	21	16	50	Yes	Good	Under-7	M
11	F	KW	KuwaIT	MiddleSch	G-07	B	IT	F	Father	70	80	25	70	Yes	Good	Under-7	M
12	M	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	50	88	30	80	Yes	Good	Under-7	H
13	M	KW	KuwaIT	MiddleSch	G-07	B	Math	F	Father	19	6	19	12	Yes	Good	Under-7	M
14	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	5	1	0	11	No	Bad	Above-7	L
15	M	lebanon	lebanon	MiddleSch	G-08	A	Math	F	Father	20	14	12	19	No	Bad	Above-7	L
16	F	KW	KuwaIT	MiddleSch	G-08	A	Math	F	Mum	62	70	44	60	No	Bad	Above-7	H
17	F	KW	KuwaIT	MiddleSch	G-06	A	IT	F	Father	30	40	22	66	Yes	Good	Under-7	M
18	M	KW	KuwaIT	MiddleSch	G-07	B	IT	F	Father	36	30	20	80	No	Bad	Above-7	M
19	M	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	55	13	35	90	No	Bad	Above-7	M
20	F	KW	KuwaIT	MiddleSch	G-07	A	IT	F	Mum	69	15	36	96	Yes	Good	Under-7	M
21	M	KW	KuwaIT	MiddleSch	G-07	B	IT	F	Mum	70	50	40	99	Yes	Good	Under-7	H
22	F	KW	KuwaIT	MiddleSch	G-07	A	IT	F	Father	60	60	33	90	No	Bad	Above-7	M
23	F	KW	KuwaIT	MiddleSch	G-07	B	IT	F	Father	10	12	4	80	No	Bad	Under-7	M
24	M	KW	KuwaIT	MiddleSch	G-07	A	IT	F	Father	15	21	2	90	No	Bad	Under-7	M

Fig. 3 The student's performance data examination.

We examine the dataset and perform a number of preprocessing steps in order to check its quality and make it ready for running. The data preparation includes data examination, cleaning, discretization, transformation. Figure 4 shows the examination of student's performance dataset after cleaning and transformation.



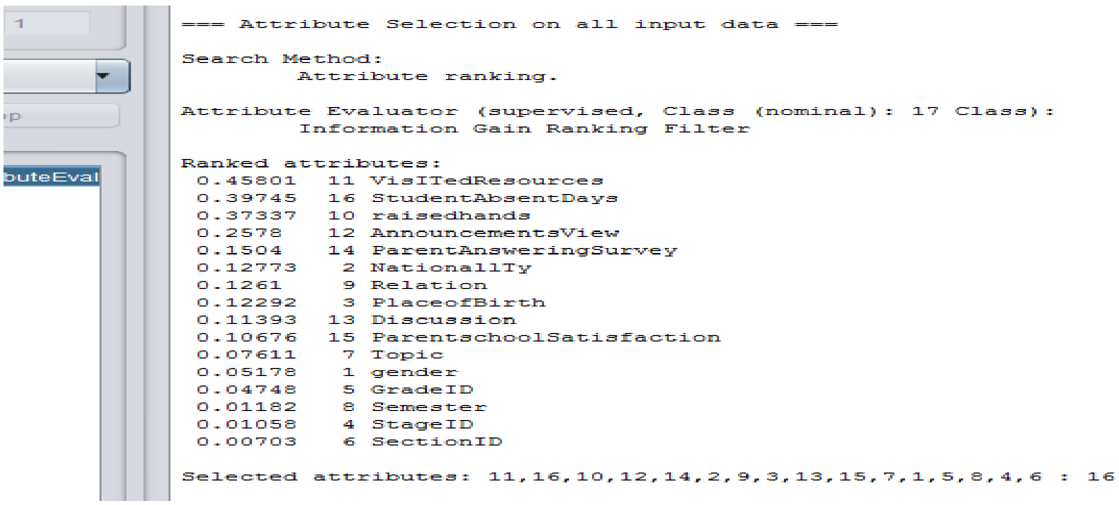
```
@relation Educationaldatabase


@attribute gender {M,F}
@attribute NationalITy {KW, lebanon, Egypt, SaudiArabia, USA, Jordan, venzuela, Iran, Tunis, Morocco, Syria, Palestine, Iraq, Lybia}
@attribute PlaceofBirth {KuwaIT, lebanon, Egypt, SaudiArabia, USA, Jordan, venzuela, Iran, Tunis, Morocco, Syria, Palestine, Iraq, Lybia}
@attribute StageID {lowerlevel, MiddleSchool, HighSchool}
@attribute GradeID {G-01, G-02, G-03, G-04, G-05, G-06, G-07, G-08, G-09, G-10, G-11, G-12}
@attribute SectionID {A, B, C}
@attribute Topic {English, Spanish, French, Arabic, IT, Math, Chemistry, Biology, Science, History, Quran, Geology}
@attribute Semester {F, S}
@attribute Relation {Father, Mum}
@attribute raisedhands real
@attribute VisITedResources real
@attribute AnnouncementsView real
@attribute Discussion real
@attribute ParentAnsweringSurvey {Yes, No}
@attribute ParentschoolSatisfaction {Good, Bad}
@attribute StudentAbsentDays {Under-7, Above-7}
@attribute Class {M,H,L}


@data

M,KW,KuwaIT,lowerlevel,G-04,A,IT,F,Father,15,16,2,20,Yes,Good,Under-7,M
M,KW,KuwaIT,lowerlevel,G-04,A,IT,F,Father,20,20,3,25,Yes,Good,Under-7,M
M,KW,KuwaIT,lowerlevel,G-04,A,IT,F,Father,10,7,0,30,No,Bad,Above-7,L
M,KW,KuwaIT,lowerlevel,G-04,A,IT,F,Father,30,25,5,35,No,Bad,Above-7,L
M,KW,KuwaIT,lowerlevel,G-04,A,IT,F,Father,40,50,12,50,No,Bad,Above-7,M
```

Fig. 4 The transformed student's data.

IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

3.3 Feature Selection

Feature selection is a fundamental task in data processing area. The objective of the feature selection process is to select an appropriate subset of features which can efficiently describe the input data, reduces the dimensionality of feature space, and removes redundant and irrelevant data. Different feature ranking techniques have been proposed for feature evaluations such as information gain and gain ratio. In this research, we applied filter-method using information gain based selection algorithm to evaluate the feature ranks, checking which features are most important to build students' performance model. Figure 5 shows the feature ranks after filter-based evaluation. During feature selection, each feature assigned a rank value according to their influence on data classification. The highly ranked features are selected while others are excluded.



```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 17 Class):
        Information Gain Ranking Filter

Ranked attributes:
 0.45801   11 VisITedResources
 0.39745   16 StudentAbsentDays
 0.37337   10 raisedhands
 0.2578    12 AnnouncementsView
 0.1504    14 ParentAnsweringSurvey
 0.12773    2 NationalITy
 0.1261     9 Relation
 0.12292    3 PlaceofBirth
 0.11393   13 Discussion
 0.10676   15 ParentschoolSatisfaction
 0.07611    7 Topic
 0.05178    1 gender
 0.04748    5 GradeID
 0.01182    8 Semester
 0.01058    4 StageID
 0.00703    6 SectionID

Selected attributes: 11,16,10,12,14,2,9,3,13,15,7,1,5,8,4,6 : 16
```

Fig. 5 Filter-based feature selection evaluation.

As shown in Figure 5 above, visited resources feature got the higher rank, and then followed by student absence days, raised the hand on classroom, parent answering survey, nationality, parent responsible for student, place of birth, discussion groups and parent school satisfaction features. As we can see the appropriate subset of features consist of ten features while other ones are excluded. Figure 6 shows the graphical view of the filter-based features selected.

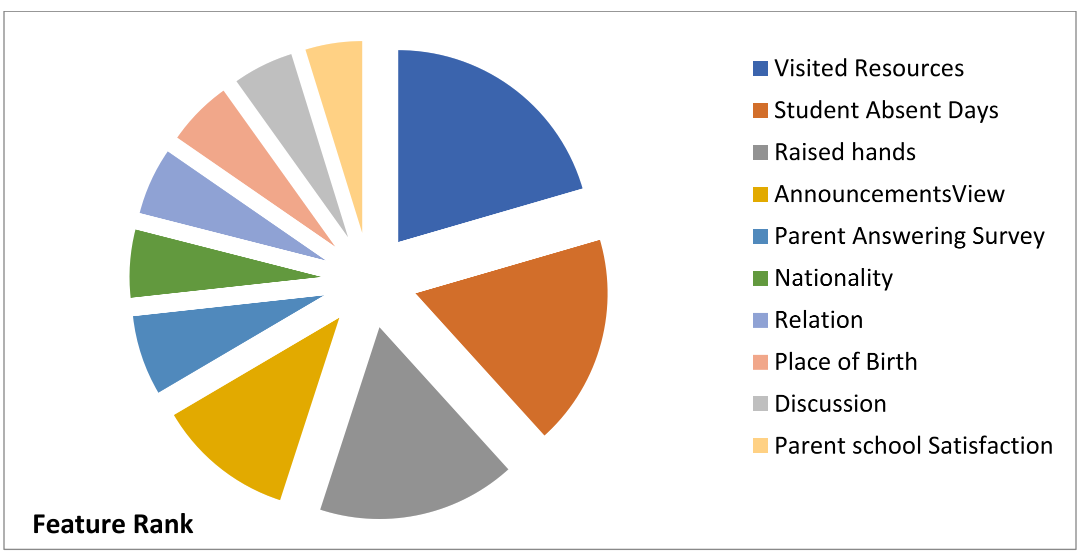


Fig. 6 Graphical view of the filter-based feature selection evaluation.

3.4 Classification Methods

We use three well-known classification methods for predicting the student's performance which respectively are Decision Tree, Naïve Bayes and Neural Network. This section presents a briefing for each of the classification methods.

3.4.1 Decision Tree:

A Decision Tree is a decision-making method that has a tree structure. It consists of four components which are a root, leaf nodes, branches and internal nodes [18]. The root connects the classes of a tree in which the leaf nodes represent the classes, the branches represent the outcomes and the internal leaves represent the processes. The classification rules are the paths from the root to the leaves [19]. The Decision Tree is represented by many algorithms and one of which is the random forest algorithm. Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object. It is used in the domains of classification, and machine learning.

3.4.2 Naïve Bayes:

Naïve Bayes is a machine learning classifier that utilizes supervised learning or statistical approach. It is based on the Bayes probabilistic theorem and uses a conditional probability to determine the outcomes [20]:

$$posterior = \frac{prior * likelihood}{evidence} \tag{1}$$

The Naïve Bayes uses the traditional classification approach setting in which a problem instance is represented by vectors of feature values known as feature vectors. The feature vectors are classified by the method to certain classes.

$$probability\left(Class_i \middle| feature_1, feature_2, \ldots\right) \tag{2}$$

The features have independent relationships (naive) in which the evaluation a feature does not affect the value of other features. This assumption reduces the accuracy of the classification; however, it also reduces the required training samples to estimate successful classification and reduces the effects of the noise in the data too. Nonetheless, the Naïve Bayes classifiers have been found to work exceptionally well when the dataset contains plenty of input features but a small number of records.

3.4.3 Neural Network:

Neural Network is one of the numerical learning methods that simulate human biological Neural Networks [21]. It consists of many nonlinear computational elements that form the network nodes or artificial neurons. The neurons are linked by weighted interconnections. One of the well-known neurons is the sigmoid that has the following form:

$$\sigma(z) = \frac{1}{1 + exp\left(-\sum_j weight_i * input_i - bias\right)} \tag{3}$$

Neural Networks are used in many research fields for classification, clustering, approximation, filtering, compression and blind source separation [21]. They are particularly useful in dealing with high complexity data that is difficult or impractical to be solved by traditional methods.

4. Experiments and Results

In this section, we detailed the experimental setup and results of the three classification methods. We ran the experiment on a PC with 4GB of RAM, 20GB of HDD using the WEKA software to evaluate the proposed classification models and comparison.

4.1 Decision Tree

In this experiment, we use a random forest function to generate the Decision Tree from the dataset. We conducted 3 tests with different data allocations of cross validation folds and split percentages [21, 22]. The test results shows that the highest accuracy is found in 1st round of test count where the data is divided into 85% training and 15% testing. Subsequently, the lowest accuracy score was observed in the 2nd round of test count where the data is divided into 75% training and 25% testing. In general, the average accuracy score is 74.418%. Table 2 shows the test results of the Decision Tree.

Table 2: Classification accuracy results of Decision Tree

Test	*Data Allocation*	*Accuracy (%)*	*RMSE*
1	85:15	77.778	0.3432
2	75:25	71.667	0.3625
3	65:35	73.809	0.3431

4.2 Naïve Bayes

In this experiment, we use an activation function to approximate the Naïve Bayes output from the dataset. We conducted 3 tests with different data allocations of cross validation folds and split percentages. The test results shows that the highest accuracy is found in the 3rd round of test count where the data is divided into 65% training and 35% testing. Subsequently, the lowest accuracy score was observed in the 1st and 2nd rounds of test counts equaling the accuracy value and Root Mean Square Error. In general, the average accuracy score is 66.865%. Table 3 shows the test results of the Naïve Bayes.

Table 3: Classification accuracy results of Naïve Bayes

Test	*Data Allocation*	*Accuracy (%)*	*RMSE*
1	85:15	66.667	0.407
2	75:25	66.667	0.407
3	65:35	67.212	0.3989

4.3 Neural Network

In the Neural Network experiment, we use a multi-layer perceptron that has two hidden layers with a sigmoid function. We conducted 3 tests with different data allocations of cross validation folds and split percentages. The Neural Network test results shows that the highest accuracy is found in 1st round of test count where the data is divided into 85% training and 15% testing. Subsequently, the lowest accuracy score was observed in the 2nd round of test count where the data is divided into 75% training and 25% testing. In general, the average accuracy score is 70.886%. Table 4 shows the test results of the Neural Network.

Table 4: Classification accuracy results of Neural Network

Test	*Data Allocation*	*Accuracy (%)*	*RMSE*
1	85:15	73.611	0.4077
2	75:25	70.000	0.4009
3	65:35	69.048	0.4188

5. Discussion

The classification accuracy results of the 10-fold cross validation are confined within 66.667% -77.778%. The Decision Tree scores the highest accuracy percentage followed by Neural Network and the Naïve Bayes. Table 5 shows the overall accuracy percentages for the classification methods based on the fold cross validation.

Table 5: The analysis of the results

Test	*1. Decision tree accuracy (%)*	*2. Naïve Bayes accuracy (%)*	*3. Neural Network accuracy (%)*	*Highest accuracy (%)*
1	77.778	66.667	73.611	1
2	71.667	66.667	70.000	1
3	73.809	67.212	69.048	1

Figure 7 shows the variation between the classification accuracy results of the decision tree, naïve bayes and neural network for the prediction of student performance based on the grade level. It can be seen that Decision Tree produced the highest accuracy rate in all the three test counts which shows that Decision Tree model performed better. It also shows the variations in their related linear measurements. However, we observed that the processing time of the neural network is dramatically longer than the other two methods used.

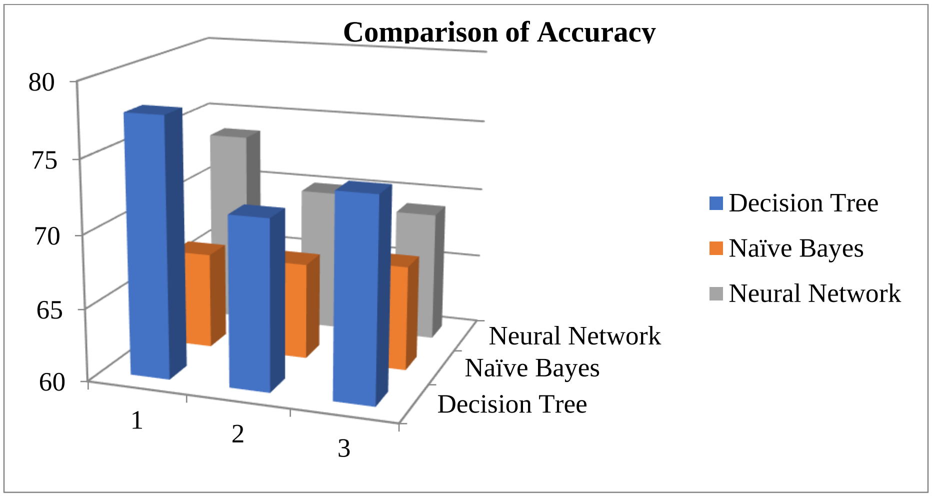


Fig.7 Variation between the classification accuracy results.

Lastly, we concluded that the decision tree boost a better prediction accuracy in predicting the student's performance with the grade level of low, medium and high.

6. Conclusion

In this research, an effort is made to find the impact of our proposed features and models on student's performance prediction. Predictions of student performance can be useful in many contexts. In this work, some feature sets are identified that significantly affect the performance of each student and grade level of student's at the end of academic year are predicted. The student's performance dataset is used to experimentally evaluate the performance of three classification methods. We implemented and tested with test counts for each of the methods and obtain the classification results. The classification results shows that the Decision Tree produces higher accuracy rate of 77.778%, followed by the Neural Network with an accuracy rate of 70.886% and the Naïve Bayes produces the lowest with an accuracy rate of 66.865%.

In future work, prediction-based researches should focus on improving student's academic performance in government institutions across developing countries.

References

[1] B. K. Baradwaj, S. Pal, "Mining educational data to analyze students' performance". IJACSA, 2012, 2: 63-69.

[2] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert systems with applications, vol. 33, no. 1, 2007, pp. 135-146.

[3] Sarker, Farhana, T. Thanassis, C. D. Hugh, "Student's performance prediction by using institutional internal and external open data sources". CSEDU: 5th International Conference on Computer Supported Education, Germany, 2013.

[4] M. Ramaswami and R. Bhaskaran., "A CHAID based performance prediction model in educational data mining," International Journal of Computer Science, vol. 7, no. 1, 2010, pp. 10-18.

[5] M. M. A. Tair and A. M. El-Halees, "Mining educational data to improve students' performance: a case study," International Journal of Information, vol. 2, no. 2, 2012, pp. 140- 146.

[6] E. Osmanbegović and M. Suljić., "Data mining approach for predicting student performance," Economic Review, vol. 10, no. 1, 2012, pp. 3-12.

[7] A. Altaher, O. BaRukab, "Prediction of student's academic performance based on adaptive neuro-fuzzy inference". IJCSNS, 2017, 17: 165-169.

[8] P. Nithya, B. Umamaheswari, A. Umadevi, "A survey on educational data mining in field of education". J Comput Sci Softw Dev, 2016, 1: 1-6.

[9] E. Osmanbegovic, S. Mirza, "Data mining approach for predicting student performance". J Econ Bus, 2012, 10: 3-12.

[10] V. Ramesh, P. Parkavi, K. Ramar. "Predicting student performance: A statistical and data mining approach". IJCA, 2013, 63: 35-39.

[11] Y. Ma, B. Liu, C. K. Wong, P. S. Yu, and S. M. Lee, "Targeting the right students using data mining," in *6th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD '00)*, New York, USA, 2000, pp. 457-464.

[12] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch, "Predicting student performance: an application of data mining methods with an educational Web-based system," in *33rd Annual Frontiers in Education (FIE 2003)*, Westminster, CO, 2003.

[13] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," *Applied Artificial Intelligence*, vol. 18, no. 5, 2004, pp. 411-426.

[14] Z. A. Pardos, N. T. Heffernan, B. Anderson, C. L. Heffernan, and W. P. Schools, "Using fine-grained skill models to fit student performance with Bayesian networks," in *Handbook of educational data mining*., 2010, pp. 417-426.

[15] Z. N. Khan, "Scholastic Achievement of Higher Secondary Students in Science Stream", Journal of Social Sciences, Vol. 1, No. 2, 2005, pp. 84-87.

[16] Q. A. AI-Radaideh, E. M. AI-Shawakfa, and M. I. AI-Najjar, "Mining Student Data using Decision Trees", International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.

[17] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah, The University of Jordan, Amman, Jordan, http://www.Ibrahimaljarah.comwww.ju.edu.jo

[18] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease". Expert Syst. Appl. 37(2), 2010, pp. 1568–1572.

[19] S. Gnanapriya, R. Suganya, G.S. Devi, M. S.Kumar, "Data mining concepts and techniques". Data Min. Knowl. Eng. 2(9), 2010, pp. 256–263.

[20] J. D. Rennie, L. Shih, J. Teevan, D. R. Karger, "Tackling the poor assumptions of naïve bayes text classifiers". In: Proceedings of the International conference on Machine Learning ICML, Vol. 3, 2003, pp. 616–623.

[21] M. Can, "Neural networks to diagnose the Parkinson's disease". SouthEast Eur. J. Soft Comput. 2(1), 2013.

[22] J. Hossain, Fazlida Mohd Sani, N., A. Mustapha, L. SurianiAffendey, "Using feature selection as accuracy benchmarking in clinical data mining". J. Comput. Sci. 9(7), 883, 2013.

Abba Babakura received the M.Sc. degree in intelligent systems from the Universiti Putra Malaysia (UPM). He is currently pursuing his Ph.D. at the Computer Science unit, Department of Mathematics, Faculty of Science, Usmanu Danfodiyo University Sokoto (UDUS). His research interests are towards decision making aspects of smart homes, Internet of Things, Data mining, Machine learning and Intelligent Building.