Content Registered

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

Predicting Student Academic Performance Using Data Mining Techniques

[1]Ikuomola A. J. & [2]Nwanze M.N
[1]Dept of Mathematical Sciences, Olusegun Agagu University of Science and Tech. Okitipupa, Ondo State, Nigeria
[2] Department of Computer Science, Federal College of Education (Technical), Asaba, Delta State
E-mails: [1]deronikng@yahoo.com; [2]Nwanzemaureen88@gmail.com

ABSTRACT

The main objective is to develop a model that predicts student academic performance using data mining technique. The model comprises of 3 main phases namely; database, data pre-processing and the data mining algorithm. WEKA toolkit was used as selected data mining tool which contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. The final year student data for four academic sessions from Computer Science Programme, Olusegun Agagu University of Science and Technology, Okitipupa were collected from the database management system. The collected student records serve as the dataset. Classification method that makes use of the mathematical techniques such as J.48, Naïve Bayes Bayesian Network, JRip, OneR and PART were used to classify the data items into separate groups. The results from the evaluation of the classification models generated with the selected data mining shows that Naïve Bayes, Bayesian Network and PART perform above average in accurately predicting the pattern and telling critical courses that can determine the output of each students final CGPA. The result also shows that from the total of 32 students, 8 students will be graduating with second class upper, 17 students will be graduating with second class lower while 7 students will be graduating with third class division.

Keywords: Algorithm, Dataset, Data Mining, , Education, Performance

Journal Reference Format:
Ikuomola A. J. & Nwanze M.N. (2020): Predicting Student Academic Performance Using Data Mining Techniques. Behavioural Informatics, Digital Humanities & Development Journal Vol 6. No. 2. Pp 45-56. Available online at https://www.behavioralinformaticsjournal. DOI: http://dx.doi.org/10.22624/AIMS/BHI/V6N2P4

1. INTRODUCTION

Nowadays, educational institution or university operates in a very dynamic and highly competitive environment. Kabakchieva (2012) stated that universities are confronted with a severe competition trying to attract the most appropriate students who will successfully pass through the university education process and make efforts to cope with the student retention. Many universities/institutions are not in the position to guide their students because of lack of information and assistance from their teaching-learning system (Thakar et al 2015). For universities management to take important decisions quickly and timely, high-quality information is required. Thakar et al (2015) assert that although universities/institutions collect an enormous number of student data, but this data remains unutilized and does not help in any decision or policy making to improve the performance of students. Yassien et al (2017) stated that the ability to predict student performance is very important in education environment.

If university/educational institution can predict students' academic performance earlier enough and also identify the factors responsible for their low performance then this knowledge can help them in taking proactive action, by arranging proper support for the low performance students in order for the student to improve their studies and achieve success for the university administration to handle the large volume of data and the increasing complexity involved.. An advanced information technologies are required which will transform the available data into information and knowledge needed to support decision making.

In order to administer and serve student population, better assessment, analysis and prediction tools are needed by the universities. In this study data mining techniques were used in analysing and predicting the academic performance of student.Data mining is data analysis tool/technique used to identify hidden patterns in large data set (Tiwari et al 2013). These extracted patterns can be used to predict student performance and behaviour easily (Al-Barrak and Al-Razgan, 2016). The implementation of data mining in the educational sector, recently defined as educational data mining (EDM) (Romero and Ventura, 2007) is a new stream in the data mining research field. The educational information acquired through EDM helps the administrators to allocate facilities and resources more effectively (Siraj and Abdoulha, 2007).

2. LITERATURE REVIEW

Yassein *et al.* (2017) presented a prediction model for the prediction of students performance based on 150 students records collected at the Najran university in Saudi Arabia with the use of statistical package for social sciences (SPSS) and two data mining processing techniques which include clustering and classification were applied to the dataset. Clementine is the data mining tool used in the application of the classification and clustering technique to the dataset collected. Decision tree algorithm C5.4 is the classification techniques used and a two-step clustering technique for the clustering phase. After the two techniques were applied, a predictive model was derived by splitting the dataset of 150 students into a training and test set during classification, a result revealing the critical affecting factors was students' attendance in class in addition to the final exam and mid exam grades. Although, the system employed a lot of techniques in improving data preparation but the authors only employed a single classification algorithm thereby reducing the accuracy. Also, custom software wasn't developed for the model, increasing complexity for non-expert users.

Kabakchieva (2012) employed lot of classification algorithms (OneR a rule learner, J48 a decision tree algorithm, Multilayer Perceptron neural network algorithm and IBK a nearest neighbour algorithm) to create the prediction model. 10330 student's data were used in creating the model, after manual clustering and association on the dataset, 10067 instances were left to be feed into the classification algorithms, all of which were presented in WEKA data mining tool. The prediction model was not implemented in a one-click application, increasing complexity for non- expert users

Tiwari et al. (2013) showed how various data mining techniques can be applied to an educational dataset of students which includes their assignments, attendance, sessional marks, GPA, and Final grade via questionnaires presented to the students to extract useful information from the collated dataset. All attributes in the dataset were converted to nominal data while all attributes in nominal form are retained in that form, all conversions are processed manually. Three main data mining techniques were applied to the dataset namely; association, classification and clustering. Association rule was used to identify possible grade values, Classification was applied on the collated dataset for describing future situations and K-means clustering algorithm was used in the division of data into groups of similar objects. The authors only employed a single classification algorithm and clustering algorithm during classification and clustering thereby reducing accuracy.

Kala (2017) used an association rule mining technique for analysing and evaluating student's skills. He used data mining techniques in evaluating the best skill suited for each student, determining whether a student should be a coder, designer or documenter for employment by any firm and the applied clustering technique for predicting student's skill sets (i.e. student's area of strength encouraging students with higher percentage of being a coder to continue to improve their programming skills). Although, an extensive clustering technique was used in increasing accurate rule formation in the categorization of each evaluated student's skill but no specific techniques under data pre-processing, feature extraction, clustering, rule formation and rule prediction were specified.

Content Registered
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

Al-banak and Al-razgan (2016) used data mining techniques in determining the critical factors affecting student's final GPA. Data preparation and pre-processing were done manually by eliminating irrelevant attributes which consist of general and elective courses to focus only on the program mandatory courses, thereafter classification algorithm (J48 decision tree algorithm) was used to process the dataset. Saa (2016) presented educational data mining and student performance prediction. His study explored multiple factors which was assured to affect student performance in higher education and also make use of a qualitative model that classified and predicts the students' performance based on related personal and social factors. Multiple classification algorithms were used, to create qualitative predictive models which were efficient and effective in predicting the students' grades.

3. DESIGN METHODOLOGY

3.1 Design Considerations

In this design, the following are incorporated into the system:

(i) Pre-processing filters: Data is often collected for unspecified applications. Data may have quality problems that need to be addressed before applying a data mining technique e.g. Noise and filters, Missing values, Duplicate data, etc. Pre-processing is needed to make data more suitable for data mining. During this phase, pre-processing is applied to the collected data to prepare it for the mining techniques.

(ii) Attribute Selection: Once the data is collected and separated into a unique relation, it is necessary to gain insight into their structure and informative value, so that it can be prepared well for the application of data mining algorithms and methods.

(iii) Classification: Classification is a classic data mining technique based in machine learning. Classification technique is used to classify each item in a dataset into one of a predefined set of classes or group.

(iv) Visualization: before applying any data mining method to the dataset, primary useful knowledge about the attributes can be visualized.

3.2 Architecture for Prediction of Student's Academic Performance using Data Mining Technique

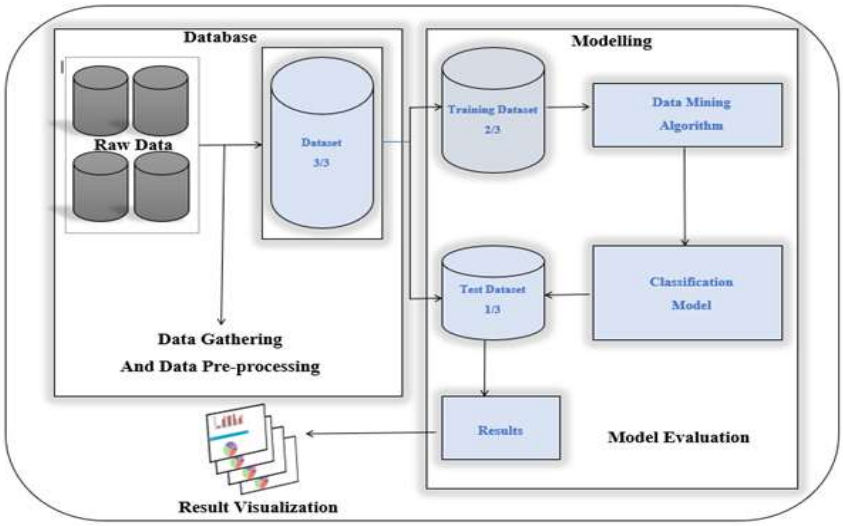

Figure 1: Architecture for prediction of student's academic performance

Figure 1 shows the architecture for the prediction of student academic performance using data mining technique. The architecture consists of three main phases namely; database, data pre-processing and data mining model. Final year student data from Computer Science Programme, Mathematical Sciences Department, Olusegun Agagu University of Science and Technology, Okitipupa, 2013 to 2018 academic sessions were collected from the school's database management system. A total of 32 student's records were used. The collected students' records serve as the historical data of the domain in consideration. Each student record had the following attributes; student matric number, all course taken by the student, course grade, CGPA, gender. The data set was divided into three parts and each time an algorithm is run, 2/3 of the data were used for training of the classification model and 1/3 of the data were used for testing and evaluation of the model. During the data pre-processing phase, some irrelevant attributes are eliminated or removed from the collected data so as to prepare the data for mining techniques. Classification method that makes use of the mathematical techniques such as J.48, Naïve Bayes Bayesian Network, JRip, OneR and PART were used to classify the data items into separate groups. The results from the evaluation of the classification models generated with the selected data mining algorithms are presented to the end-user to view.

3.3 Component of the Architecture Design

3.3.1 Database

It is a collection of data or information, which is specially organized for rapid search and retrieval by a computer. Databases are structured to facilitate the storage, retrieval, modification, and deletion of data in conjunction with various data-processing operations. The database gotten from the university after a due process of request, was received in form of a series of Microsoft excel documents, separately containing various semester performance of each students from their second year (200level) to their final year (500Level) covering a period of four years.

3.3.2 Data Pre-processing

Data gotten from the University database were noisy, contains missing values of students' scores, and inconsistency which could affect the knowledge discovery and pattern recognition process during application of classification algorithm and clustering in WEKA, the following data Pre-processing steps were carried out to ensure that the data was free of noise and inconsistency and ready for mining

a. **Data cleaning:** involves the removal of all missing values from each student's performance and assigning of zero (0) signifying that such course or activity wasn't performed by the respective student, this process was done manually by running a Visual Basic Code on the excel file to increase accuracy and reduce error occurrence on the data cleaning process.
b. **Data Integration:** it consists of combining of the various semesters results separated and scattered over numerous excel files to form a unified dataset of the student's performance from their second year to their final year. This step was performed by merging the data into a single dataset in an excel document form.
c. **Data transformation**: involves converting of the raw source data into single format, that is from the source format which is excel (xlsx) to either Comma delimited (CSV) or Attribute relation file format (arff). WEKA has the required toolkit for the data mining tool requires for collated data to be in arff or CSV form.

The following steps were performed before the actual conversion was carried out.

i. **Normalization**, where the attribute data are scaled so as to fall within a specified range such as score of 70 to 100 equates to A, only the students' scores and CGPA were normalize.
ii. **Data reduction**: this step involves the removal of unnecessary data from the gathered data, data removed involves attributes or instances that have little or no significant against our predicting class which is the CGPA. This step help in reducing large data to its bare minimum without comprising the integrity of original data itself and yet producing the quality knowledge required.

3.3.3 Data Mining Model

A model is a representation of an object, which accepts an input or set of inputs and produces a set of outputs or an output. In this work models are generated by applying a set of supervised classification algorithms (e.g. decision tree) to a set of pre-processed historical data of past occurrence and their results are used to find patterns and make possible prediction on future occurrence.

(a) Classification

Students final CGPA were predicted based on their previously grades on attempted courses offered by each student which gives an information on how specific courses affects each student's final performances (CGPA). The classification algorithms used Decision Tree (J.48), Bayesian (Naïve Bayes), Bayesian (Bayesian Network:), Rules (OneR), Rules (JRIP), and Rules (PART). Each algorithm was applied on a percentage split of 66% for testing and 33% for training (2/3 for testing and 1/3 for training).

(i) **Decision trees algorithm** is one of the most popular classification technique in data mining and an upgrade to Ross Quinlan C4.5 programs for machine learning they represent the group in a tree form, and there are many advantages over other techniques. Some of the capabilities are: (i) Ability to handle binary, missing, nominal attributes and class and, (ii) Minimum number of instances allowed is zero

(ii) **Bayesian algorithm** is also a very popular classification technique which produces a high level of accuracy compared to other techniques and works best on nominal data. The Bayesian network is a structure that gives the conditional dependants between domain variables and may also be used to illustrate graphically the probabilistic underlying relationships among domain variables. Some of the capabilities are: (i) Ability to handle binary, missing, nominal attributes and class, (ii) Minimum number of instances allowed is zero.

(iii) **OneR (One Rule)** is a simple classification algorithm that generates a one-level decision tree. OneR is able to deduce typically simple, yet precise, classification rules from a set of instances. OneR is able to handle missing values and numeric attributes showing flexibility in spite of simplicity.

(iv) **JRip (RIPPER)** is one of the basic and most popular algorithms. Classes are examined in growing size and an initial set of rules for the class is generated using incremental reduced error. JRip (RIPPER) proceeds by treating all the examples of a particular decision in the training data as a class, and finding a set of rules that cover all the members of that class.

(v) **PART** is a separate-and-conquer rule learner. The algorithm producing sets of rules called decision lists" which are planned set of rules. A new data is compared to each rule in the list in turn, and the item is assigned the class of the first matching rule. Some of the capabilities are: (i) Ability to handle binary, missing, nominal attributes and class, (ii) Minimum number of instances allowed is three

3.3.4 Result representation

The result shows the performance of each classification techniques and its accuracy level on the third split (1/3) test dataset. The accuracy level of the prediction shows the critical attributes (courses) that affect the final performance and the following estimates

(i) Correctly classified Instance

(ii) Incorrectly classified instances

(iii) Kappa Statistic is the main metric used to measure how good or bad an attribute measurement system is

(iv) Mean absolute error is a measure of difference between two continuous variables.

(v) Root mean squared error (RMSE) is a frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed. The RMSD represents the sample standard deviation of the differences between predicted values and observed values

(vi) Relative absolute error as the name suggests, the mean absolute error is an average of the absolute errors, which is the prediction and the true value.

Crossref
Content Registered

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

(vii) Root relative squared error is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values. Thus, the relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor

(viii) True positive rate is defined as the percentage ratio of correctly classified instances of a given class. True positive rate (TPR) = TP/ (TP+FN)

(ix) False positive rate is defined as the percentage ratio of falsely classified instances of a given class. False positive ratio (FPR) = FP/(FP+TN)

(x) Precision is defined as a proportion of instances that are truly of a class divided by the total instances classified as that class. Positive predictive value = TP/(TP+FP) = PS

Each result after classification can be represented either in graphical form, html, CSV, or plain text.

4. IMPLEMENTATION AND RESULTS

4.1 Dataset Description

4.1.1 Data Sources

Final year student data from Computer Science Programme, Olusegun Agagu University of Science and Technology, Okitipupa in the year 2013 to 2018 were collected from the database management system. The total of 32 student's records were used, the collected student record serves as the dataset and are presented in figure 2.

Viewer

Relation: dataset faith

No.	1: I»¿Students Matric Number String	2: Gender Nominal	3: CSC201 Nominal	4: CSC205 Nominal	5: CSC207 Nominal	6: GST221 Nominal	7: MTH201 Nominal	8: MTH211 Nominal	9: CSC202 Nominal	10: CSC204 Nominal	11: CSC206 Nominal
1	130404001	Male	B	C	C	C	D	B	E	C	E
2	130404002	Male	B	B	B	B	A	B	A	B	C
3	130404003	Male	B	B	E	E	C	D	E	C	E
4	130404004	Male	B	A	C	C	A	C	A	C	C
5	130404005	Male	A	A	C	A	A	A	A	A	A
6	130404007	Male	C	B	D	D	C	C	E	B	C
7	130404008	Male	C	B	C	E	F	C	E	F	E
8	130404010	Male	C	A	C	F	C	A	E	B	E
9	130404013	Male	A	A	C	C	A	A	B	A	A
10	130404014	Male	B	E	F	D	F	F	E	F	B
11	130404015	Male	C	B	C	D	B	C	E	B	C
12	130404016	Male	B	C	C	E	C	E	F	B	C
13	130404017	Female	C	B	F	E	E	F	E	A	C
14	130404018	Male	B	A	E	C	A	E	A	B	C
15	130404020	Male	B	C	C	E	B	F	E	B	B
16	130404021	Male	B	C	E	F	B	F	A	B	D
17	130404022	Male	A	B	C	B	A	C	B	A	E
18	130404023	Male	E	D	F	D	E	F	E	C	C
19	130404024	Male	C	B	D	C	C	C	C	C	C
20	130404025	Male	B	B	C	C	A	C	A	A	B
21	130404026	Male	B	C	D	C	A	C	E	B	E
22	130404027	Female	A	B	C	E	D	C	E	B	B
23	130404028	Male	A	B	C	D	C	E	B	C	E
24	130404030	Male	C	B	C	F	E	C	E	B	B
25	130404031	Male	B	B	B	A	B	C	B	A	B
26	130404032	Female	C	C	C	C	E	F	E	B	E
27	130404033	Male	A	A	B	C	A	A	B	B	B
28	130404035	Male	B	C	F	D	F	E	C	F	F
29	130404036	Male	A	C	C	D	E	E	E	B	E
30	14DE0404037	Male	B	A	C	C	E	F	C	B	D
31	14DE0404038	Female	C	C	D	E	E	C	F	B	C
32	14DE0404039	Male	C	B	C	B	B	C	B	A	B

Add instance | Undo | OK | Cancel

Figure 2a: Students Dataset

Content Registered

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

Viewer

Relation: dataset faith

2	28: CSC314 Nominal	29: CSC401 Nominal	30: CSC403 Nominal	31: CSC405 Nominal	32: CSC407 Nominal	33: CSC411 Nominal	34: CSC413 Nominal	35: CSC400 Nominal	36: CSC501 Nominal	37: CSC503 Nominal	38: CSC505 Nominal	39: CSC507 Nominal	40: CSC509 Nominal	41: CSC511	42: CGPA Nominal
	D	A	D	C	A	F	D	A	B	B	B	D	C	C	Second Class Lower
	A	A	A	B	A	A	A	A	B	A	B	C	A	A	Second Class Upper
	C	C	C	C	B	A	D	A	B	C	E	D	C	E	Third Class
	A	A	B	B	A	A	C	A	A	A	B	C	A	D	Second Class Upper
	B	A	A	B	A	A	C	A	A	A	A	A	B	A	Second Class Upper
	B	B	A	C	A	B	B	A	B	B	E	B	B	D	Second Class Lower
	C	A	F	C	A	F	F	A	E	B	A	F	F	B	Third Class
	B	A	B	A	A	A	B	A	A	B	A	C	F	A	Second Class Upper
	A	A	A	B	A	A	B	A	A	A	B	B	B	A	Second Class Upper
	C	B	D	C	A	F	D	A	A	C	C	D	F	D	Third Class
	D	C	D	C	A	F	C	A	C	B	F	D	B	D	Second Class Lower
	B	C	F	C	A	A	D	A	A	A	B	C	C	C	Second Class Lower
	A	C	A	C	A	F	C	A	B	A	E	C	B	B	Second Class Lower
	C	B	C	B	A	A	B	A	B	B	C	C	B	C	Second Class Lower
	B	B	B	B	A	B	B	A	C	B	E	D	B	C	Second Class Lower
	D	E	F	C	C	F	D	A	C	C	F	D	E	E	Third Class
	B	D	C	B	A	B	B	A	A	B	D	C	C	C	Second Class Lower
	C	E	C	C	B	F	F	B	C	C	E	D	F	E	Third Class
	C	B	A	C	B	B	C	A	A	B	B	C	B	A	Second Class Lower
	B	B	D	C	A	A	C	A	A	A	A	C	B	C	Second Class Lower
	A	A	C	B	A	A	C	A	A	A	A	D	B	B	Second Class Lower
	A	B	C	C	A	C	C	A	B	B	C	D	A	A	Second Class Lower
	B	A	B	B	A	A	C	A	B	B	E	B	C	A	Second Class Lower
	F	F	F	F	F	F	F	F	F	F	F	F	F	F	Second Class Lower
	A	B	C	C	A	A	C	A	A	A	C	B	A	A	Second Class Upper
	C	B	D	C	A	A	F	A	A	C	C	B	C	E	Third Class
	A	A	B	B	A	A	B	A	A	B	A	C	A	A	Second Class Upper
	D	C	E	B	B	F	C	A	F	F	F	F	F	F	Third Class
	C	E	D	C	C	F	D	B	C	C	E	D	F	E	Second Class Lower
	A	B	D	C	B	A	F	A	C	B	C	B	F	F	Second Class Lower
	A	E	C	C	A	F	C	A	C	B	C	D	C	B	Second Class Lower
	B	A	F	C	A	A	D	A	A	B	D	B	F	C	Second Class Upper

Add instance | Undo | OK | Cancel

Figure 2b: Students Dataset

4.2 Analysis

The data were analyzed using WEKA toolkit. WEKA is a collection of supervised and unsupervised machine learning algorithms and data processing tools. It contains various tools for data pre-processing, classification, regression, clustering, association rules and visualization. There are many learning algorithms implemented in WEKA including Bayesian classifier, Trees, Rules, Functions, Lazy classifiers and miscellaneous classifiers. The algorithms can be applied directly to a data set. WEKA is also a data mining software developed in JAVA, it has a GUI chooser from which any one of the four major WEKA applications can be selected. For this study, the Explorer application was used.

Crossref
Content Registered

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

The Explorer window of WEKA has six tabs. The first tab is pre-process that enables the formatted data to be loaded into WEKA environment. Once the data has been loaded, the pre-process panel shows a variety of information as shown in figure 3.

Figure 3: Weka GUI Chooser

4.3 Data Visualization

After loading the data to WEKA, we got some primary useful knowledge about the attributes before applying any data mining method by using the visualizing technique in the software. For instance, the result shows that most of the students will be graduating with second class lower in figure 4, figure 5 shows the graph of students' final performance while figure 6 shows the visualization graph of distribution of class CGPA and some course distribution

.

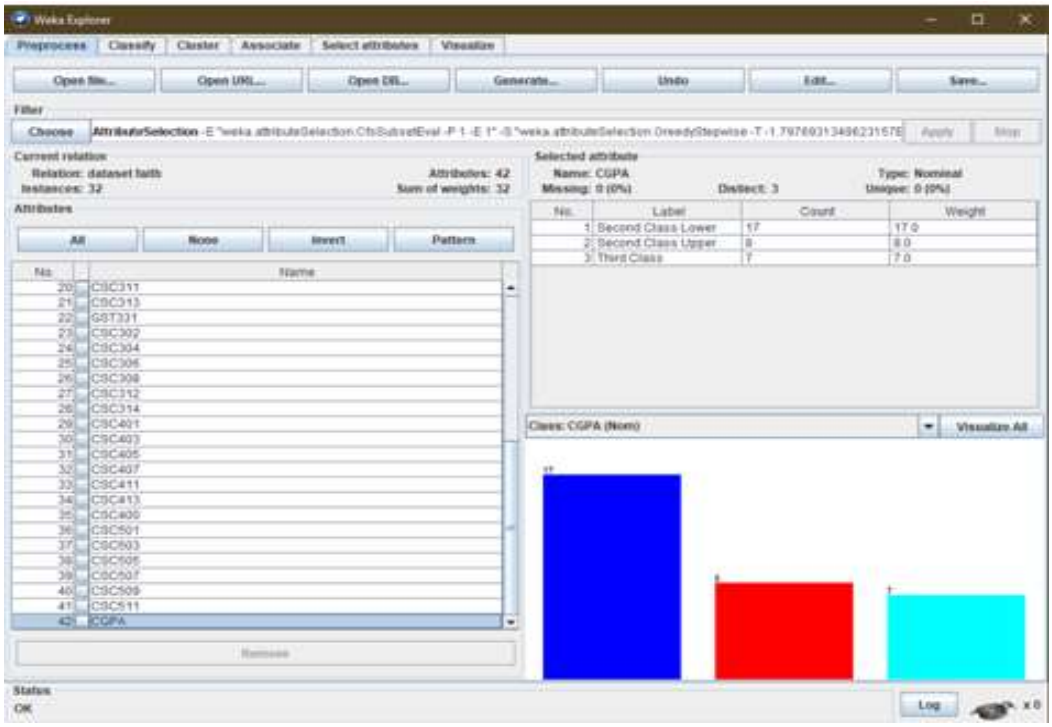

Figure 4: WEKA Explorer Interface

Crossref
Content Registered

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

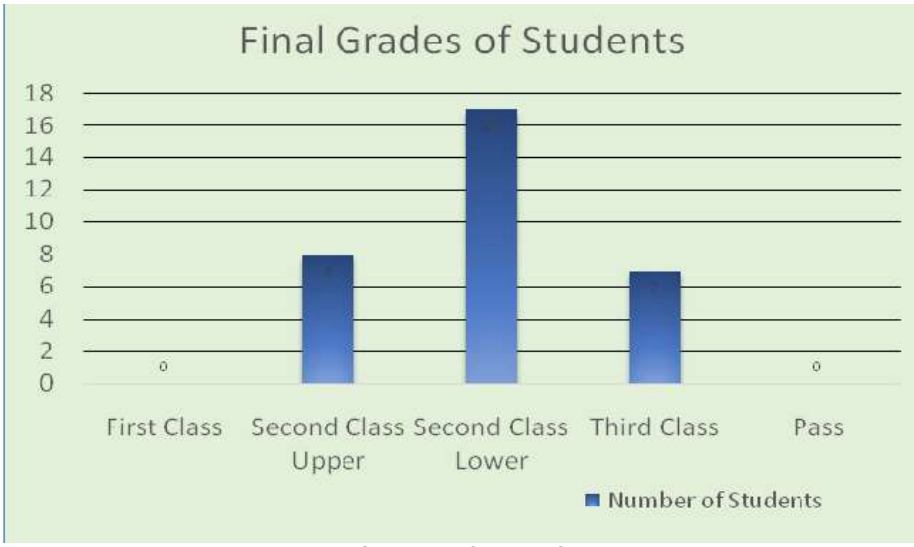

Figure 5: Students' final performance

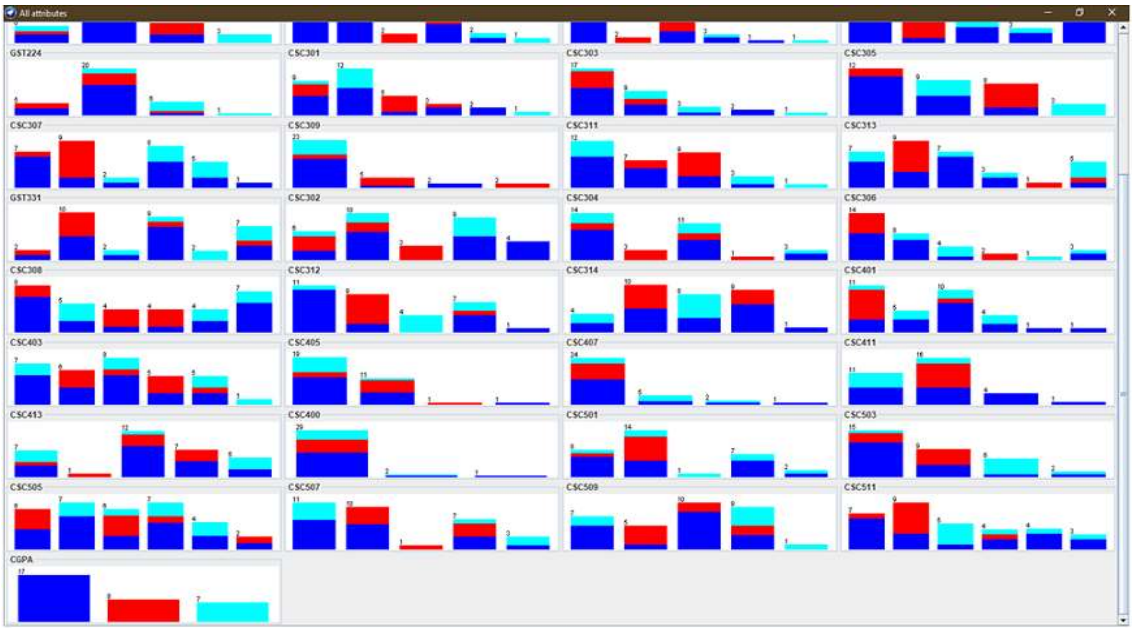

Figure 6: Visualization Graph of distribution of class CGPA and some course distribution

4.4 Classifiers

There are many classifiers available in WEKA but decision tree (J.48), Bayesian (Naïve Bayes and Bayesian Network) and Rules (JRIP, oneR and Part) were used as the classification technique.

Table 1 shows the comparison of results before feature reduction with all 41 features on various classification schemes while figure 7 shows the graphical representation of the false positive and false negative rate of each classifier respectively.

Crossref
Content Registered

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

Table 1 : Comparison of Results of Training Data before Feature Reduction with all 41 Features using Principal Component Analysis on various Classification Schemes

SCHEMES	TP (%)	FP (%)	PRECISION (%)	RECALL (%)	F-MEASURE (%)	ROC AREA (%)	TIME TAKEN TO BULD EACH MODEL
PART	72.7	14.4	77.7	72.7	72.6	80.7	0.02
OneR	36.4	40.5	28.8	36.4	32.1	47.9	0
JRip	54.4	29.5	6.636	54.5	29.1	61	0.02
NaiveBayes	81.8	11	81.8	81.8	81.8	95.8	0
J48	36.4	40.5	28.8	36.4	32.1	49.4	0
BayesNet	81.8	11	81.8	81.8	81.8	97.3	0.01

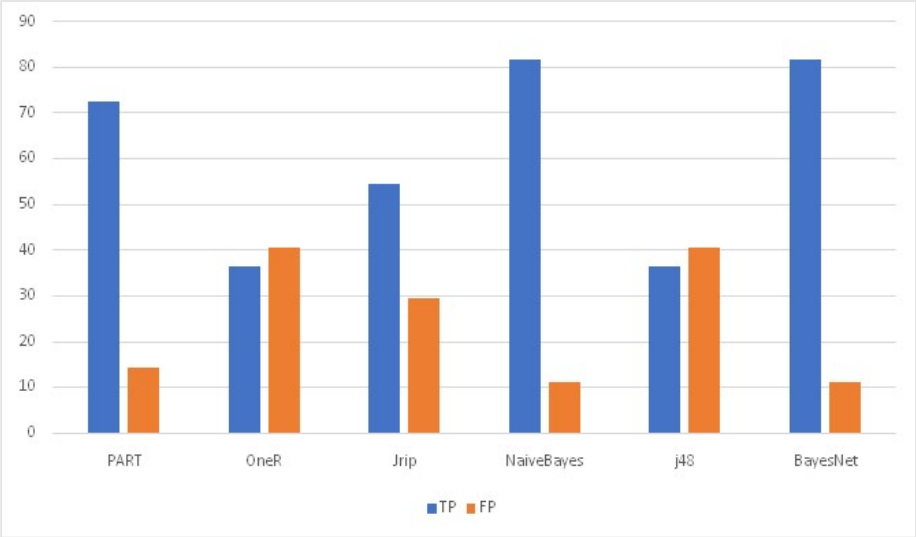

Figure 7: The True Positive Rate and Fasle Positive Rate of each Classifier before Feature Reduction

By analyzing the results obtained in table 1 above using Kuang (2007) metrics, it was found that the results of true positive rate (TPR) for PART = 72.7%, OneR = 36.4%, JRIP = 54.4%, NaiveBayes = 81.8%, J48 = 36.4% and BayesNet = 81.8% respectively, which implies that NaiveBayes and BayesNet have the highest number of truly classified instances compared with PART, OneR . JRIP and J48.

For false positive rate (FPR), PART = 14.4%, OneR = 40.5%, JRIP = 29.5%, NaiveBayes = 11%, J48 = 40.5% and BayesNet = 11% respectively, which again shows that NaiveBayes and BayesNet have the lowest rate of producing false classified instances compared with PART, OneR . JRIP and J48.

Upendra and Jain (2012) and Subbulakshmi et al. (2010) were of the opinion that any good classification scheme should have the highest TPR and the lowest FPR. It is clear from table 1 and figure 7, that the NaiveBayes and BayesNet schemes give better TPR, and FPR than the other four schemes.

Content Registered

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 6, No 2, June, 2020

Table 2 shows the comparison of results (after feature reduction) with 9 features on various classification schemes while Figure 8 shows the graphical representation of the false positive and fasle negative rate of each classifier respectively.

Table 2: Comparison of Results of Training Data after Feature Reduction with 9 Features using Principal Component Analysis on various Classification Schemes

SCHEMES	TP (%)	FP (%)	PRECISION (%)	RECALL (%)	F-MEASURE (%)	ROC Area (%)	TIME TAKEN TO BULD EACH MODEL
PART	72.7	14.4	77.7	72.7	72.6	80.7	0.02
OneR	36.4	40.5	28.8	36.4	32.1	47.9	0
JRip	45.5	37.1	37.7	45.5	40.9	54.2	0
NaiveBayes	90.9	11	92.4	90.9	90.4	94.7	0
J48	36.6	40.5	28.8	36.4	32.1	49.4	0.02
BayesNet	81.8	11	81.8	81.8	81.8	94.7	0

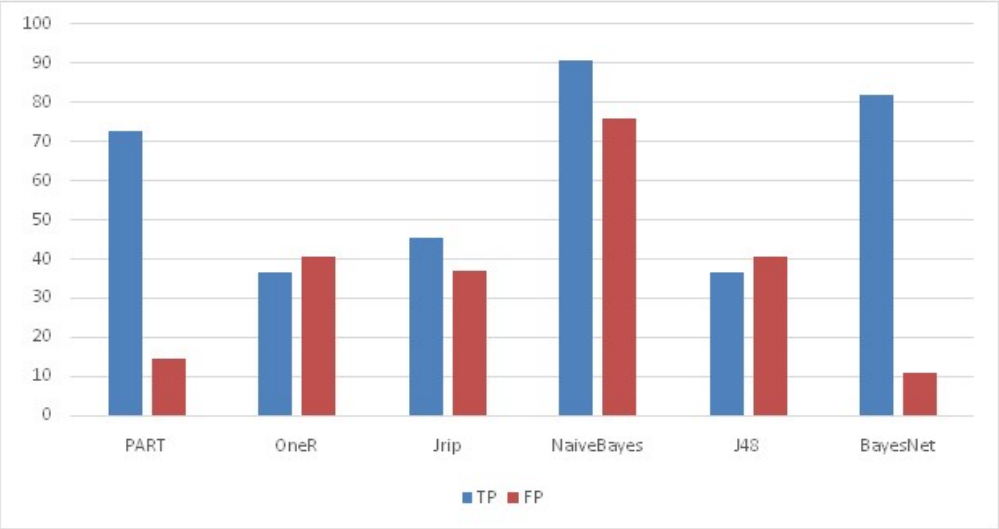

Figure 8: The True Positive Rate and False Positive Rate of each Classifier after Feature Reduction

By analyzing the results obtained in table 1 above using Kuang (2007) metrics, it was found that the results of true positive rate (TPR) for PART = 72.7%, OneR = 36.4%, JRIP = 45.5%, NaiveBayes = 90.9%, J48 = 36.6% and BayesNet = 81.8% respectively, which implies that NaiveBayes and BayesNet have the highest number of truly classified instances compared with PART, OneR . JRIP and J48.

For false positive rate (FPR), PART = 14.4%, OneR = 40.5%, JRIP = 37.1%, NaiveBayes = 11%, J48 = 40.5% and BayesNet = 11% respectively, which again shows that NaiveBayes and BayesNet have the lowest rate of producing false classified instances compared with PART, OneR . JRIP and J48.

5. CONCLUSION

The dataset was experiments with one decision tree algorithm (J.48), two Bayesian classifiers (Naïve Bayes and Bayesian Network) and three rule classifiers (JRip, OneR, PART). The dataset contains 32 instances, in which 20 instances were randomly selected and used as the test and 11 instances were used as the training dataset in which the truly classified instances were recorded as 36.36%, 90.90%, 81.81%, 45.45%, 36.36% and 72.72% respectively. After analysing student's data using WEKA toolkit, from the result it was found that Naïve Bayes, Bayesian Network and PART perform above average in accurately predicting the pattern and telling critical courses that can determine the output of each students final CGPA. The result also shows that from the total of 32 students, 8 students will be graduating with second class upper, 17 students will be graduating with second class lower while 7 students will be graduating with third class division.

REFERENCES

1. Al-Barrak M.A. and Al-Razgan M. (2016). Predicting Students Final GPA using Decision Trees: A Case Study. International Journal of Information and EducationTechnology, 6(7).
2. Kabakchieva D. (2012). Student Performance Prediction by using Data Mining Classification Algorithms. University of National and World Economy Sofia, Bulgaria. 1 (4).
3. Kala K. (2017). Student's Skills Evaluation Techniques using Data Mining. IOSR Journal of Computer Engineering. 19(1).
4. Kuang L. V. (2007). DNIDD: A dependable network IDS using the CSI-KNN, M.Sc. Thesis, School of Computing, Queen's University, Ontario, Canada.
5. Romero C. and Ventura S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135-146.
6. Saa A.A. (2016). Educational Data Mining and Students'Performance Prediction. International Journal of Advanced Computer Science and Applications, 7(5).
7. Siraj F. and Abdoulha M.A. (2007). Mining enrolment data using predictive and descriptive approaches, Knowledge-Oriented Applications in Data Mining, 53-72.
8. Subbulakshmi T., Matthew G. and Shalinie S. M., 2010. Real time classification and clustering of intrusion detection system alerts using machine learning algorithms. *International Journal of Artificial Intelligence and Application, 1(1):1-9*
9. Thakar P., Mehta A. and Manisha (2015). Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. International Journal of Computer Applications, 110(15).
10. Tiwari M., Singh R. and Vimal N. (2013). An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education. International Journal of Computer Science and Mobile Computing. 2(2),.53 – 57.
11. Upendra and Jain Y. K. (2012). An empirical comparison and feature reducation performance analysis of intrusion detection. International Journal of Control Theory and Computer Modelling, 2(1).
12. Yassein N. A., Helali R.G.M. and Mohomad S.B. (2017). Predicting Student Academic Performance in KSA using Data Mining Techniques. . Journal of Information Technology & Software Engineering.