International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Early Prediction of Student Success Using a Data Mining Classification Technique

[1]Mohamed Hegazy Mohamed, [2]Hoda Mohamed Waguih

[1]Demonstrator at Shorouk Academy, Management Information, System, Cairo, Egypt

[2]Associate Professor, Computer and Information Systems Department, Sadat Academy for Management Sciences, Cairo, Egypt

Abstract: ***Nowadays the amount of data stored in educational databases is increasing rapidly. These databases contain hidden information which can be used for the improvement of student academic performance in higher education systems. Predicting student's academic performance beforehand can help management, faculty as well students to make timely decisions. A data mining technique is used to study the data available in the educational field, and bring out hidden knowledge for decision making. Data mining classification technique was used through three decision tree methods, namely: J48, Rep tree, and Random Tree. This paper is an attempt to apply the data mining techniques, particularly classification, to help students in enhancing their results and the quality of the higher educational system by early prediction of student success using decision tree methods. In this study, the main attributes that affect the student performance were determined to predict students' final grade early. For this purpose, we have used real data obtained from the managerial higher institute 'Tammoh' in Giza-Egypt for first year students. Weka data mining tool was used to generate and compare the classifier models of the selected algorithms and results were reported. A ranker search method was then applied to rank the best five attributes in data by using Info Gain ranker to filter the most important rules of the selected classifier model. Final results show that applying J48 algorithm with ranker search method can enhance the generated rules and help management predict early weak students and take appropriate decisions to prevent them from failure and thereby enhance students' academic performance.***

Keywords: Data Mining, Classification, Decision Trees, Higher Education.

1. Introduction

In this age, we are overwhelmed with data. Analysis of this data needs powerful tools such as data mining. Data mining is used to extract the meaningful information from large data using some algorithms. [5] Data mining concepts and methods can be applied in various fields like marketing, medicine, real estate, customer relationship management, engineering, web mining, educational data mining, etc. Educational data mining is a new emerging application of data mining that can be applied on the data related to the field of education. There are increasing research interests in using data mining in education. One of the problems in education which are solved by applying data mining is prediction of student performances, whose aim is predicting values of an unknown variable (result or grade) which describes a student. Educational Data Mining uses many techniques: Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others. By using these techniques, many kinds of knowledge can be discovered such as association rules, classifications, and clustering. Classification techniques are supervised learning techniques that classify data items into predefined class label. It is one of the most useful techniques in data mining to build classification models from an input data set. The used classification techniques commonly build models that are used to predict future data trends. There are several algorithms for data classification such as decision tree and Rule. With classification, the generated model will be able to predict a class for given data depending on previously learned information from historical data, the discovered knowledge can be used for predicting the final grade for student in a particular course in the first semester. Examinations play a vital role in any student's life. The marks obtained by the student in examinations decide his future. Therefore, it becomes essential to predict whether the student will pass or fail the examination. If the prediction says that a student tends to fail in the examination prior to the examination, then extra efforts can be taken to improve his studies and help him to pass the examination. In this paper, we make prediction about fail and pass class for students based on final grade by using J48, Rep tree and Random tree algorithms.

This study is more useful for identifying the weak students, and the identified students can be individually assisted by the educators so that their performance is better in future. This study investigates the accuracy of some classification techniques for predicting performance of a student. The main objectives of this study are: to identify highly influencing predictive variables on the academic performance of higher secondary students, find the best classification algorithm on student data, and predict the grade at higher secondary examination. The next sections contain the related works that covered the area of educational data mining and how to use Education Data Mining to enhance the educational process through predicting student final grades. In Section 3, the decision Tree and the three algorithms used in this paper (J48, Rep tree and Random tree) were discussed. Then, the experiment of this research was discussed including data set description, and how we collected data set from real world in higher education system, classification model and important factors for predicting student's academic Performance in Section 4. In section 5 results and discussion on the existing prediction methods were discussed. Finally, conclusions and future work are outlined in section 6.

2. Related Work

Data mining is a powerful analytical tool that enables educational institutions to better allocate resources, staff, and proactively manage student outcomes. The management system can improve their policy and enhance their strategies so as to improve the quality of the education system. Many of the earlier researchers have used data mining techniques, and capabilities for improving student academic performance in higher education systems.

In [3], the authors use Naive Bayes classification algorithm to build a model for predicting academic performance of students enrolled in the bachelors of computer applications programs in five colleges at Awadh university in India. The constructed model was applied on a data set of 300 records for students including: 226 males and 74 females. The data set was described by features representing academic, social, demographic, and psychological attributes. The results show that the most affecting variable in predicting student academic performance are student's grade in the senior secondary school, place of residence and their language.

In [1], the authors conducted a comparative study for predicting student level performance using three classification algorithms: Naive Bayes, C4.5 and ID3. The authors applied the algorithms on real data sets that were collected from department of computer science at Bishop Heber college. The study included 1000 students. The performance of these students was analyzed based on their assignment mark, seminar mark, internal marks, and their external marks. Final results show that Naïve Bayes algorithm out performs the two other algorithms and provides 81% of accuracy and can be better used in predicting student level performance.

In [6], the authors applied a number of classification algorithms using J48, Naïve Bayes, Multilayer perceptron, and SVM and compared them. They developed a J48 classifier with 10-fold cross validation on a data set of 162 instances representing student records at Machakos University College in India. They used the decision tree (J48) classifier model and they were able to build a model that has the ability to load it, and fetch data for prediction from the database of the Higher learning institution. The raw dataset that was sourced from the University database systems and files for this study the data set was described using 10 attributes. The results obtained from experiments with the classifier showed that it is capable of performing classification with an accuracy of 94.4% for dataset obtained from the database systems.

In [7], the authors presented a comparative analysis using different classifiers on students' data to identify weak students and take the necessary action for improving their performance to improve academic results. The data was collected from the database of final year students for Information technology undergraduate course. The authors applied data mining classification techniques on the available data set to discover knowledge. The classifiers used in the study were naïve bayes, NB tree, bayes net, IBK, and J48. They applied the selected classifiers on data sets and conducted a comparative study on the results. The results show that the J48 classifier which acts as java implementation of c4.5 algorithm was chosen as the best one on the used dataset.

In [10], the authors compare the accuracy of decision tree and Bayesian network algorithm for predicting academic performance of undergraduate and post graduate students at two different academic institutes. In this context, the authors conducted two case studies and made predictions for 4 classes (Fail, Fair, Good, and Very Good), 3 classes (Fail, Good, and Very Good) and 2 classes (Fail and Pass). They use 10-fold cross-validation for building and testing the model. The results show higher accuracy for the decision tree classifier.

In [4], the authors predict student academic performance by using student personal and pre-university characteristics. The data set belongs to 10330 students of a Bulgarian educational sector, each instance being described by 20 attributes (gender, birth year and place, place of residence, and country, place and total score from previous education, current semester, total university score, etc.). They applied C4.5 decision tree, Naive Bayes, Bayesian networks, K-nearest neighbors (KNN) and rule learner algorithms to classify the students into 5 classes, Excellent, Very Good, Good, Average or Bad. The results show that the best accuracy obtained by all these classifiers is 66.3%. The predictive accuracy for the Good and Very Good classes (which contain most students) for all classifiers was around 60% –75%.

In [2], the authors used data mining classification algorithms to predict the results of students currently in the second year, based on the results obtained by students in the first year of engineering major. The data set used was in the form of a Microsoft Excel 2003 spreadsheet, and had details of each student such as full name, application ID, gender, caste, percentage of marks obtained in board examinations of classes X and XII, percentage of marks obtained in physics, chemistry and mathematics in class XII, marks obtained in the entrance examination, and admission type. They applied an implementation of the decision tree ID3 and C4.5 algorithms on the collected data set to identify promising students and provide them some attention and improve those who would probably get lower grades. The results show that the accuracy of the ID3 algorithm is 75.145% and that of C4.5 is 75.145%.

Most of the previous studies focused on the use of data mining classification techniques for predicting student academic results based on enrollment data and performance of students for final examination. The main point in our research, unlike the previously mentioned researches, is making use of the most relevant factors affecting the final grade through a decision tree technique. In this context, we applied an implementation of three decision tree algorithm on the collected data set; conduct a comparative study on the obtained results to choose the most accurate algorithm in predicting the student final grade based on the selected

features. We, then, make use of a Ranker Search method to filter the most important rules of the selected algorithm.

3. Decision Tree and Used Algorithms.

A decision tree is a powerful and popular method for both classification and prediction techniques. It is a flow-chart-like tree structure. The attribute that has the maximum value of information gain is selected as a root node where each internal node is denoted by rectangles, and leaf nodes are denoted by ovals. All internal nodes have two or more child nodes. All internal nodes contain splits, which test the value of an expression of the attributes. Arcs from an internal node to its children are labeled with distinct outcomes of the test. Each leaf node has a class label associated with it. It is commonly used for gaining information for the purpose of decision -making. The process in decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

The attractiveness of tree-based methods is due largely to the fact that decision trees can easily be converted to classification rules. Rules are a good way of representing information or bits of knowledge. A rule-based classifier uses a set of IF-THEN rules for classification. An IF-THEN rule is an expression of the form IF condition THEN conclusion. Rules can readily be expressed in English so that humans can understand them [2].

Decision trees are produced by algorithms that identify various ways of splitting a dataset into branch-like segments. These segments form an inverted decision tree that originates with a root node at the top of the tree. The object of analysis is reflected in this root node as a simple, one-dimensional display in the decision tree interface. The name of the field of data that is the object of analysis is usually displayed, along with the spread or distribution of the values that are contained in that field. It uses real data-mining algorithms to help for decision making with classification technique.

Most commonly Decision trees algorithms used in the educational field because the construction of decision tree classifiers does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision trees can handle multidimensional data. Their representation of acquired knowledge in tree form is intuitive and generally easy to assimilate by humans. The learning and classification steps of decision tree induction are simple and fast. In general, decision tree classifiers have good accuracy. In this research, we used this decision tree learning algorithms through J48, Rep tree and Random Tree algorithms. J48 algorithm is a successor to ID3 developed by Quinlan Ross (Quinlan, 1986). It is also based on Hunt's algorithm and handles both categorical and continuous attributes to build a decision tree. In order to handle continuous attributes, it splits the attribute values into two partitions based on the selected threshold such that all the values above the threshold as one child and the remaining as another child. It also handles missing attribute values. J48 uses Gain Ratio as an attribute selection measure to build a decision tree. It removes the biasness of information gain when there are many outcome values of an attribute. Also, J48 calculate the gain ratio of each attribute. The root node will be the attribute whose gain ratio is maximum. It uses pessimistic pruning to remove unnecessary branches in the decision tree to improve the accuracy of classification. Entropy and Information Gain measures are used by J48 to construct a decision tree. Entropy is a formula that calculates the homogeneity of a sample where the completely homogeneous sample has entropy of 0 and equally divided sample has entropy of 1. Entropy uses the formula described below to calculate the homogeneity of a training set (T).

$$\text{Entropy (T)} = -\sum_{i=1}^{n} Pi \; log2 \; Pi$$

Rep Tree algorithm is the Fast decision tree learner. It Builds a decision or regression tree using information gain or variance and prunes it using reduced-error pruning (with back fitting), and only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces (i.e. as in C4.5). Random tree uses Class for constructing a tree that considers K randomly. We have a random subset of attributes to deals with limitation of decision tree. The value of random subset is based on operator, in this way we can solve the classification as well as regression and prediction problem.

4. Experiment

The data sets used in our experiment are real instance examples representing student's records of the first academic year from the managerial higher institute 'Tammoh' in Giza, Egypt. A total of 8080 records and 9 attributes throughout the years 2007 to 2015 are taken for the analysis of this research. Students' data in first academic year include grade details and personal information as well as family background details. The attributes are student's gender, student's place of birth, student's age, high school major, high school grade, student's attendance, father Education, student's department and student's final grade representing the decision attribute. The detailed description of the attributes is presented in table (1).

Data preparation and pre-processing are very important steps in any data mining process that usually consumes the bulk of the effort invested in the entire data mining process. Some preprocessing was conducted on the data set in order to clean the data and prepare it prior to the mining process. In this context, we ignored some tuples to handle missing values in some of the conditional attributes and eliminated some irrelevant attributes, like student name and id because they did not affect the data analysis process. Data tuples from multiple sources are then merged into a coherent data source, and a discretization process is applied to transform some attributes such as the final grade, age and attendance from numeric to nominal attributes. Finally, prior to applying the selected classification algorithms, the target

data set was transformed to the specific input data formats used by Weka.

Table 1: Student Related Attributes in The First Year

ID	*Variable Name*	*Description*	*Domain*
1	Gender	Student's Gender Type (binary: 'F' - female or 'M' - male)	{M, F}
2	Place of Birth	Student's Place of Birth Type (Nominal: 'U' - urban or 'R' - rural)	{U, R}
3	Age	Student's Age Nominal [A:18-20 B:21-23 C:23-25]	{A, B, C}
4	HS_Major	High School Major: Type Nominal Literary_General_Secondary 'L' Scientific_General_Secondary = 'S' Management and Services = 'MS' Commercial = 'Com' Industria_3_Years = 'I3' Industria_5_Years = 'I5'	{L S MS Com I3 I5}
5	HS_Grade	High School Grade: Type Nominal A = Excellent from 85%: 100% B = Good from 61%: 84% C = Fair from 50%: 60%	A B C
6	Attendance	Student's Attendance in the End Year: Type Nominal H = High from 81%: 100% M = Medium from 61%: 80% L = Low from 51%: 60% P = Poor from 0%: 50%	H M L P
7	Final Grade	High Excellent >=93 & <100 A+	A+
		Excellent >=85 & <93 >=80 & <85 A	A
		High Very Good >=80 & <85 B+	B+
		Very Good >=75 & <80 B	B
		High Good >=70 & <75 C+	C+
		Good >=65 & <70 C	C
		High Acceptable >=60 & <65 D+	D+
		Acceptable >=50 & <60 D	D
		failure >=0 & <50 F	F
8	Father_Edu	Father Education. High level = H / Medium = M / Low=L	H / M /L
9	Department	Student's Department (Accounting = A /Management = M /MIS)	A / M/ MIS

Three different decision tree classification algorithms were compared in this experiment. The first; J48 is an implementation of the C4.5 decision tree classifier model by Quinlan. The second is Rep tree, a fast decision tree learner algorithm, and the third is the Random tree algorithm. The three algorithms have been compared with the help of implementations provided by the Waikato Environment for Knowledge Analysis (Weka) [5]. Weka is a comprehensive suite of Java class libraries that implement many state-of-the-art machine learning and data mining algorithms.

The performance of a classification algorithm is determined by how accurately it classifies a given set of examples. In the current experiment, we tested the generated classifier models using the k-fold Cross Validation (CV) mode because the data set was limited. The k-fold CV refers to a widely used experimental testing procedure where the dataset is randomly divided into k disjoint blocks of objects, then the data mining algorithm is trained using k-1 blocks, and the remaining block is used to test the performance of the algorithm; this process is repeated k times. At the end, the recorded measures are averaged. In our experiments, we choose k = 10 that is we used a 10-fold cross-validation for measuring the error rate of each of the three classifiers. Table (3) presents a summary of the results of testing the generated classifier models. As shown from the results, using the first classifier, J48, 6917 instances were correctly classified, and 1163 were not correctly classified giving an accuracy of 85.60 %. The classification accuracy for the Rep Tree was 85.54 % and that of the random tree was 85.28 %. Thus, comparing the three selected classification algorithms for the current problem, the J48 Pruned decision tree classifier is performing better.

Finally, we discovered in this experiment that the generated rules were 38 rules by using J48 classifier, and that in this generated rule many rules were not important for predicting student final grade so we used a Ranker Search method to filter the most important rules of all obtained rules in order to prioritize the if – then rules that were generated based on the best 5 attributes related for student as showed in Table (2). Finlay, we obtained the best 19 rules in the decision tree as showed in table (7).

Table 2: High potential Attributes

ID	*Name of the Variable*	*Rank Values*
1	High School Major	1.5092
2	Attendance	1.2982
3	Father Education.	1.0346
4	Department	0.9828
5	High School Grade	0.896

5. Results and Discussion

This section will discuss the results analysis of the recent works for predicting student academic performance. In this study, decision tree classifiers were used through J48, Rep Tree and Random Tree. We have addressed the prediction of class final grade of students based on the attributes taken. This study will help in identifying those students with poor performance. The main goal of the current research is to predict final grade for students who need special attention. All results are provided in figure (1) that shows the accuracy for three algorithms and the predicting numbers of each class for final grade attribute as showed in tables (4) for J48 algorithm, Rep Tree in tables (5) and Random Tree tables (6). Finally, it has been shown that J48 algorithm gives the best performance with accuracy 85.60% at time 0.9 seconds, number of leaves was 38 and size of the tree was 51. In order to evaluate the obtained results, five performance measures are used: the accuracy, TP rate, FP rate, precision and recall. The accuracy refers to the percentage of correctly classified records in the testing dataset. The accuracy measures by TP rate and TN rate divided by TP rate, TN, FP and FN as showed in equation 1.TP rate represents the number of examples predicted positive that are actually positive. FP rate is the number of examples predicted positive that are actually negative. Precision is TP rate divided by predicted TP rate and FP rate. Precision measures the percentage of records that the model classified as good that are actually Good as showed in equation 2. Recall is the TP rate divided by predicted TP rate and FN rate. Recall is the TP rate (also referred to as sensitivity)

recall measures the true positives recognition rate as showed in equation3. Finally, the best 19 rules were discovered of a total of 38 rule by eliminating the lower priority features using Ranker Search method technique by WEKA tool as sowed in table (7). All measures are calculated as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Equation 1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Equation 2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Equation 3}$$

Table 3: Statistical Analysis of Classifiers

Algorithm	Time /Sec	Model Evaluation			
		Correctly Classified		Incorrectly Classified	
		#	%	#	%
J48	0.09	6917	85.60	1163	14.39
REP Tree	0.08	6912	85.54	1168	14.45
Random tree	0.05	6891	85.28	1189	14.7153

Table 4: Precision and Recall for J48 Classifier

Algorithm	*Class*	*TP Rate*	*FP Rate*	*Precision*	*Recall*
J48	A+	0.881	0.021	0.858	0.881
	A	1.000	0.019	0.845	1.000
	B+	1.000	0.022	0.822	1.000
	B	1.000	0.024	0.775	1.000
	C+	1.000	0.011	0.889	0.914
	C	0.653	0.034	0.789	0.804
	D+	0.565	0.009	0.940	0.588
	D	1.000	0.017	0.878	1.000
	F	1.000	0.005	0.944	0.971

Table 5: Precision and Recall for REP Tree Classifier

Algorithm	Class	TP Rate	FP Rate	Precision	Recall
REP Tree	A+	0.791	0.020	0.850	0.791
	A	1.000	0.023	0.820	1.000
	B+	1.000	0.023	0.818	1.000
	B	0.981	0.034	0.741	0.981
	C+	0.905	0.019	0.826	0.905
	C	0.675	0.016	0.884	0.675
	D+	0.569	0.009	0.930	0.569
	D	0.996	0.020	0.835	0.996
	F	0.996	0.018	0.843	0.996

Table 6: Precision and Recall for Random Tree Classifier

Algorithm	Class	TP Rate	FP Rate	Precision	Recall
Random Tree	A+	0.832	0.020	0.859	0.832
	A	0.988	0.018	0.851	0.988
	B+	1.000	0.026	0.790	1.000
	B	0.987	0.031	0.747	0.987
	C+	0.916	0.018	0.842	0.916
	C	0.704	0.018	0.863	0.704
	D+	0.570	0.008	0.944	0.570
	D	0.993	0.015	0.879	0.993
	F	0.992	0.017	0.851	0.992

The performance comparison on the basis of accuracy among algorithms is shown in Figure 1

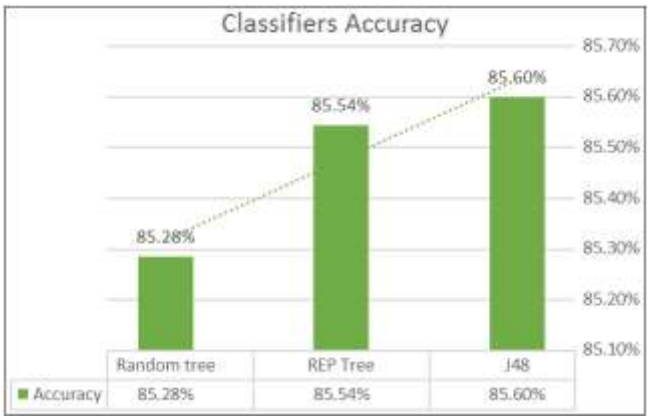

Figure 1: Accuracy Comparison of Classifiers

Table 7: J48 Classifier Rules Output

ID	*J48 Rules*
1	If the Department is Management and High School Major is Industria_3_Years Then Final Grade is B
2	If the Department is Management and High School Major is Commercial Then Final Grade is A+
3	If the Department is Management and High School Major is Industria_5_Years Then Final Grade is C+
4	If the Department is Management and High School Major is Management and Services Then Final Grade is D
5	If the Department is Management and High School Major is Scientific_General_Secondary Then Final Grade is A
6	If the Department is Management Information System and Father Education is High Level Then Final Grade is A+
7	If the Department is Management Information System and Father Education is Medium Level Then Final Grade is B
8	If the Department is Management Information System and Father Education is Low level and Attendance is poor Then Final Grade is C
9	If the Department is Management Information System and Father Education is Low level and Attendance is Average and High School Major is Industria_3_Years Then Final Grade is B+
10	If the Department is Management Information System and Father Education is Low level and Attendance is Average and High School Major is Commercial Then Final Grade is B+
11	If the Department is Management Information System and Father Education is Low level and Attendance is Average and High School Major is Industria_5_Years Then Final Grade is A+
12	If the Department is Management Information System and Father Education is Low level and Attendance is Average and High School Major is Management and Services Then Final Grade is B+
13	If the Department is Management Information System and Father Education is Low level and Attendance is Average and High School Major is Scientific_General_Secondary Then Final Grade is B+
14	If the Department is Management Information System and Father Education is Low level and Attendance is Good and High School Major is Industria_3_Years Then Final Grade is D
15	If the Department is Management Information System and Father Education is Low level and Attendance is Good and High School Major is Commercial Then Final Grade is D+
16	If the Department is Management Information System and Father Education is Low level and Attendance is Good and High School Major is Industria_5_Years Then Final Grade is D+
17	If the Department is Management Information System and Father Education is Low level and Attendance is Good and High School Major is Management and Services Then Final Grade is D+

18	If the Department is Management Information System and Father Education is Low level and Attendance is Good and High School Major is Scientific_General_Secondary Then Final Grade is A+
19	If the Department is Management Information System and Father Education is Low level and Attendance is Excellent Then Final Grade is D

6. Conclusions and Future Work

The ability to predict a student's performance is very important in educational environments. In this Research, a case study was presented that shows how to use data mining classification techniques to predict student's academic performance through the early prediction of their final grade in a managerial higher institute in Giza Egypt. We aimed at using the knowledge extracted from the student's final grade in the first academic year to enhance the final grade in the next year and focus on students that have low final grade, and take the appropriate actions. The implementation of three data mining classification algorithms, J48, Rep-tree and Random-tree, was used for conducting our experiment and results are reported. In the process of evaluation of the classifier models, J48 Pruned decision tree showed better results compares with the two other classifiers, Rep Tree and Random Tree. Present research shows that academic performance of students is not always depending on their own effort, but other factors may have significant influence over student performance. Therefore, we need a method to evaluate the most important factors that have a big impact on the student final grade to enhance the results of the selected classifier. In this context, we applied a Ranker Search method in order to rank the rules generated from the selected algorithm and select the most significant ones. The decision tree method through using J48 algorithm helps the management to predict early weak students and can take appropriate decisions to prevent them from failure and there by enhance students' performance. For future work, we will generalize the study and add the elective and general courses to get more accurate results, and we will compare many universities and institutes in the private education sector. We will extend the experiment using other data mining techniques, such as neural network and clustering.

References

[1] A. Rajalakshmi, K. Sivaranjani and M. Phil, (2015)," A Comparative Study on Student Performance Prediction", International Conference on Engineering Trends and Science & Humanities (ICETSH), ISSN: 2348 – 8387, PP 53-56.

[2] Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., and Honrao, V, (2013), "Predicting Students Performance Using ID3 and C4.5 Classification Algorithms.", Vol.3, No.5, pp.39-52.

[3] Bhardwaj, B.K., Pal, S., (2011)," Data mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security, 9(4):136–140.

[4] D. Kabakchieva, K. Stefanova, V. Kisimov, (2011), "Analyzing University Data for Determining Student Profiles and Predicting Performance", Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, the Netherlands.

[5] Ian H Witten, Eibe Frank and Mark A Hall, (2011), "Data Mining - Practical Machine Learning Tools and Techniques (3rd Ed)", Elsevier, ISBN 978, PP 3-5.

[6] Joseph Mutuku Ngemu, Elisha Opiyo Omulo, William Okelo Odongo and Bernard Manderick, (2015)," Student Retention Prediction in Higher Learning Institutions: The Machakos University College Case", International Journal of Computer and Information Technology, Volume 04, Issue 02, PP 489 – 497.

[7] Kritika Mendiratta and Simple Sharma, (2015), "A Decision Support System for Predicting Student Performance", International Journal for Scientific Research & Development, Vol. 3, Issue 03, ISSN (online): 2321-0613, PP 1028-1030.

[8] Max Bramer, (2007), "Principles of Data Mining", Springer.

[9] Mrinal Pandey and Vivek Kumar Sharma, (2013), "A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction", International Journal of Computer Applications, 61(13)1-5.

[10] T. N. Nghe, P. Janecek, P. Haddawy, (2007)," A Comparative Analysis of Techniques for Predicting Academic Performance", Proceedings of 37thASEE /IEEE Frontiers in Education Conference.