Prediction of Students' Cumulative Grade Point Averages (CGPAs) at Graduation: A Case Study

Suleiman Khalifa Arafa Ibrahim
(MSC)
Department of Computer Science and Information Technology
Comboni College of Science and Technology, Sudan

Mahmoud Ali Ahmed, PhD
Faculty of Mathematical Sciences
Khartoum University, Sudan

ABSTRACT

Predicting student's performance achievement based on their academic grades continues to be one of the most popular applications of educational data mining and, therefore, it has become a valuable source of knowledge that has been used for different purposes in particular universities. This paper aims to implement the prediction method J48 to predict the students final CGPAs at graduation based on their academic data. For that matter, two different scenarios were investigated in this study. The students' GPAs from the first 3 years were used for prediction in the first scenario, whereas the students' GPAs from the first 2 years scores were used in the second scenario. In this study, students academic data for those who did their graduation from the department of Information Technology during the period [2007-2015] , at Comboni College of Science and Technology, SUDAN. As the results indicate, the prediction J48 method performed reasonably well in predicting the student GPA at graduation in both scenarios. According to the 10-fold cross validation test, J48 algorithm produced the accurate prediction result of 83.3333 % for the first scenario. The experiment was repeated for the second scenario. J48 algorithm again produced the accurate prediction results with 81.0345 % was obtained with 10- fold cross validation test. In conclusion, the prediction algorithm J48 in data mining able to accurately predict student CGPA at graduation well in advance, which can identify students needing extra help to improve their academic performance. Moreover, in this study, the Apriori algorithm was used to extract hidden patterns from the graduates' data. Finally, based on the achieved results, this study could offer helpful feedback and recommendations to the department planners to take corrective measures to assist weak students and in turn, increase their chances of graduating with better grade.

Keywords
Institutions of Higher Education, Educational Data Mining, prediction of students' final CGPAs, CCST, J48 Algorithm

1. INTRODUCTION

Many years of practice show that data mining (DM) is a process, and its successful application implemented in various domains such as financial industries and banking have been advanced through the techniques of data mining [1]. In the last few years, education domain become as a novel data mining target of application for knowledge discovery [2]. In this regard, Institutions of higher education (IHEs) are at the core of educational systems in which extensive research and development is performed in a competitive environment.

One of the main challenges faced today's IHEs is the lack of significant knowledge at the educational main processes such as counseling students, evaluating academic achievement and improve decisions making. Therefore they are not able to give suitable recommendation to the students regarding their performance status. In this regard, data mining techniques are analytical tools that can be used to extract meaningful knowledge from large educational data sets. DM as a new powerful technology can be defined as the automated process of extracting useful knowledge and information including, patterns, and associations from large data sets that are unknown [2].

Nowadays, IHEs are very focused on their students' academic performance to successfully preparing the next generation for their future ahead. Given the importance of academic performance, an IHE would benefit from being able to predict its students' future academic performance. This would allow the IHE to take corrective measures to assist its at-risk students and to enhance their chances of graduating. One way to predict students' academic performance is the use of data mining techniques [3]. DM is considered as the most suitable technology in giving additional insight into educational entities such as; students, teachers and managerial administrators. It acts as an active automated assistant in helping them to make better decisions on their educational activities. Students' academic performance can be measured in terms of, course grades [4], semester Grade Point Average (GPA) [5], cumulative GPA (CGPA) and final GPA [6]. In our case study, student's academic standing is calculated in the form of the cumulative grade point average (CGPA) taking into account all the courses they have taken over the whole degree or diploma program. This paper aims to investigate the capabilities of data mining application in higher learning institution help in predicting students' performance and in turn improves decisions making. In here, our main idea of this study is to predict well in advance the students' CGPAs at graduation in order to reveal whether a student tends to have a low CGPAs so that extra efforts can be made to improve the student's academic performance and, in turn, improve CGPAs. There are several DM techniques can be used in predicting the students' performance. However, classification is one of the most commonly used data mining technique in predicting the students' performance in educational institutes [7] [8]. In this study, the classification algorithm named decision tree J48 was used for mining the academic students' performance. The choice of this prediction algorithm is due to its superior capability in classification problems. Decision tree rules can be also transformed into the production rules which are easier for humans to interpret and understand. Furthermore, Apriori algorithm, as one of the most popular algorithms for association rule mining was used to discover the hidden patterns from the dataset. The dataset used in this study was collected from academic Affairs of Comboni College of Science and technology (CCST) in

Sudan. The data consisted of 522 graduated student records of students of the Department of Information Technology who completed their study in the college [2007-2015] were considered in our study. To find the high influence attributes, feature selection was also conducted to make the mining process faster, valuable and meaningful. Although there are different commercial software applications available that are adapted to DM, but in this study the commercial WEKA software has been used to apply the data mining techniques. Waikato Environment for Knowledge Analysis (WEKA) is a free software project composed of a collection of machine learning algorithms for solving real world DM problems [9]. It was developed by the University of Waikato (New Zealand). The system is written in JAVA and distributed under the terms of the GNU General Public License. WEKA provides implementations of learning algorithms, such as classification techniques, association rules, and clustering. In addition, WEKA includes a variety of tools for preprocessing data such as transforming datasets and cleaning and facilities of data visualization. WEKA tool comes with a GUI, which should make it easier to use and operates on the predication that the user data is available as a flat file or relation. This means that each data object is described by a fixed number of attributes that have of a specific type included alphanumeric or numeric values. WEKA accepts the data in specific formats, such as ARFF (Attribute-Relation File Format (ARFF), CSV (Comma Separated Values) and C4.5 format. These specific formats have been taken into account in dealing with this study. The paper is organized as follows: Section 2 presents a review of related works; Section 3 presents the study framework proposed by the authors and the data description used in this study. Section 4 showed the experiments and experimental results of the data mining algorithms on the selected dataset. The final conclusion and future work are given in Section 5.

2. RELATED WORKS

Muluken Alemu in [10] investigated the potential applicability of data mining technology to predict student success and failure cases on University students' datasets. Data set of totaling 11,873 regular undergraduate students records was used to build and test both decision tree and Naïve Bayes models using data analysis tool WEKA 3.7 software. In case of identifying those students who are not performing well and those who are at the risk of failing (CGPA below 2.00) with highest prediction accuracy is that of decision tree at 92.34% as compared to naïve Bayes classifier. The results obtained from this research have shown the potential applicability of data mining technology to classify university students' academic performance as failure/success. In concluding, the research results offer a helpful and constructive recommendations to the academic planners in universities of learning to enhance their decision making process. Furthermore, the result can be used in the curriculum structure and modification in order to improve students' academic performance.

Mohammed M. and Abu Tair in [11] gave a case study using educational data mining techniques and algorithms to improve graduate students' performance, and overcome the problem of low grades of graduate students. In doing so, they applied DM techniques on fifteen years [1993-2007] graduate students data collected from the college of Science and Technology,

Khanyounis. Then, they applied data mining techniques to discover knowledge. Particularly they discovered association rules. Then they applied two classification methods which are Rule Induction and Naïve Bayesian classifier to predict the Grade of the graduate student. Furthermore, they also clustered the students into groups using K-Means clustering algorithm. Finally, they used outlier detection to detect all outliers in the data, using two outlier methods which are Distance-based Approach and Density-Based Approach. Each one of these tasks can be used to improve the performance of graduate student.

In another similar work done by, Fiseha and Addisalem [12], a decision tree algorithm was applied to predict students' performance based on their academic record. The data was collected from the college of Agriculture, Department of Horticulture –Dilla University. The data include five years period [2009-2014]; the preprocessing, processing and experimenting was conducted using Rapid Miner tool. During processing among a total of 49 various attributes which will help to improve the student's academic performance 27 important rules were generated. From the generated model specific courses, sex, academic status in 1st and 2nd year of the students determines the performance of student. Finally, the decision tree algorithm was tested and it provides a promising result of accuracy of 84.95%.

Sadiq et. Al in [13], Students' academic performance was evaluated based on academic and personal data collected from 3 different colleges from Assam, India. The total number of records was 300 with 24 attributes. Two attributes were dropped in the phase of data cleaning. Using feature selection, 12 highly influential attributes were selected. They applied four different classification algorithms namely: J48, PART, BayesNet and Random Forest. The data mining tool used in the experiment was WEKA 3.8. Based on the accuracy and the classification errors, they conclude that the Random Forest Classification method was the most suited algorithm for the dataset. Apriori algorithm was also used to find the association rule mining among all the attributes and the best rules were also displayed

3. RESEARCH METHODOLOGY

Before starting applying the data mining techniques on the given data set, there should be a frame work that should be follow to achieve the main objective of this paper.

3.1 The proposed frame work

The authors proposed a frame work that follows the same steps as the general DM process as illustrate in Figure 1. The proposed frame work is initially carried out by the process of data collection and data preprocessing. This follows, selecting the most appropriate data mining techniques and algorithms in the second step. Next, the experimental process is carried out using Weka 3.9 software tool. Finally, the results are analyzed and evaluated for each of the data mining methods to generate rules.
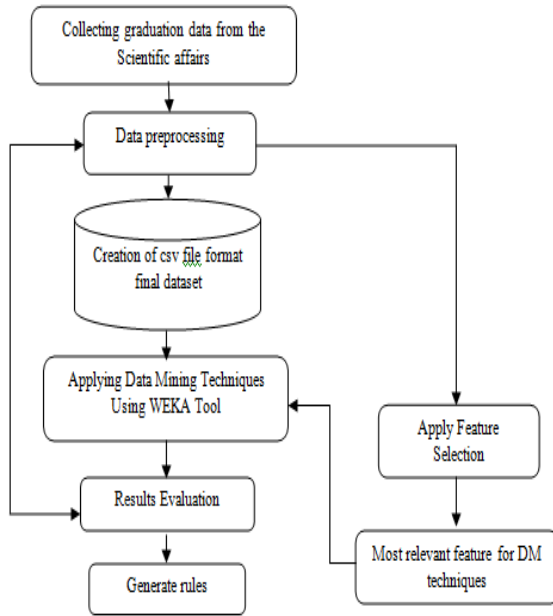
Figure 1: Proposed framework

3.2 Data collection

The data set used in this paper contains graduate students data collected from Scientific Affairs of Comboni College of Science and Technology (CCST) in Sudan. Comboni College grants their graduates a bachelor and diploma degree in four specialties, including three specialties for bachelor degree in Computer Science, Education, and English language, and one specialty for diploma in Information Technology. This department further categorized in to two specialties: Business Administration and Accounting. For the scope of the study, the previous graduates' academic data comes from the department of Information Technology was considered as a study case study. The department has three years study plan, with six semesters. The graduate students' data set consists of 522 record and 11 attribute those students who graduated within the years 2007 and 2015. The data set used in this paper contains graduate students data collected from Scientific Affairs office. In Comboni College of Science and Technology (CCST), exams are evaluated out of 100 points. The mark out of 100 is divided by 20 to get the equivalent letter grade points for each subject as follows: A = 4 or <=5; B = 3.5 or < 4; C= 3.0 or < 3.5; D = 2.5 or < 3.0; F < 2.5 which means Failure. Different subjects have different weights according to their relevance. That is translated into the credit hours of the subject. To get the result of a semester, the credit hours of each subject are calculated according to the points scored, and an average of all is made a grade point average (GPA). Thus, the attribute containing the GPAs are used as input to establish the mining model. Student's academic standing is calculated in the form of the cumulative grade point average (CGPA) taking into account all the courses they have taken over the whole degree or diploma program. After students are completed their study, the student is given a specific graduation grade as follows: 'Distinction' if student final CGPA is >= 4.0; 'Very Good' if student final CGPA is > = 3.5, 'Good' if student CGPA is >= 3.0; 'Pass' if student CGPA is > = 2.5. A student is considered graduated after the decision of the academic Council at the end of the 4 year's study for the Bachelor degree or the 3 year's study for the Diploma program. A student must have passed in all

subjects of the program and got at least a CGPA of 2.50 at the end of study. The attributes that possibly influenced their results are selected and analyzed as detailed in Table 1.

Table 1: Selected attributes and their description

Attributes	Description	Possible values
Gender	The Students' sex	Male or Female
SSCEM	The Students' Secondary School Certificate Exam Mark	Ranges between 0 to 100
SSCT	The Students' Secondary School Certificate type	Science, art, and others
Major	The students department	IT-Accounting, IT-Business
GPA_S1	The Students' Grade point Average obtained in semester One	Ranges between 0 to 5
GPA_S2	The Students' Grade point Average obtained in semester Two	Ranges between 0 to 5
GPA_S3	The Students' Grade point Average obtained in semester Three	Ranges between 0 to 5
GPA_S4	The Students' Grade point Average obtained in semester Four	Ranges between 0 to 5
GPA_S5	The Students' Grade point Average obtained in semester Five	Ranges between 0 to 5
GPA_S6	The Students' Grade point Average obtained in semester Six	Ranges between 0 to 5
Final CGPA	The Students' Final cumulative GPA obtained at graduation	Distinction, Very good, Good, Pass

The Grade Point Average (GPA) and Cumulative Grade Point Average (CGPA) are calculated according to the grading system followed by the college, which is the five points grading system. Each one of them carries 5 marks. GPA which is refers to the average of the student's scores during the semesters and rounded up to two decimal numbers as can be defined and calculate using the equation:

$$GPA = \frac{\text{sum(grade point of every subject} \times \text{credit hour)}}{\text{sum(class credit hours)}}$$

On the other side, Cumulative Grade Point Average (CGPA) which is refers to the average of all the grades accumulated over the courses taken within a specific period of education as can be calculated using the equation:

$$CGPA = \frac{\text{sum(grade of every subject} \times \text{credit hour)}}{\text{sum(credit hours of the all studied subjects)}}$$

Before the dataset is loaded into the WEKA data mining analysis software tool to carry out a series of experiments, it is of interest to observe and study the dataset in the case study terms of data redundancy, irrelevant data, missing values to improve the quality of the data set.

3.3 Data preprocessing

Data pre-processing is an important step in the knowledge discovery process. It is the step before applying data mining algorithm, it transforms the original data into a suitable form to be used by a particular mining algorithm. Therefore, before applying DM algorithms it is necessary to carry out some pre-

processing tasks such as data cleaning, data transformation and feature selection [14]. This step helps to get better input data for data mining techniques

3.3.1 Data cleaning

Data cleaning is one of the main data preprocessing tasks. The data set needs to be verified to make sure the values are correct and missing values need to be filled or removed. This can be done by replacing the missing value(s) with some constant. For numeric fields, the missing values could be replaced by the field mean; for categorical values, the mode may be used. Finally, missing values may be replaced with imputed values based on the other characteristics of the record. As part of the data preparation, the unsupervised filter technique so called ReplaceMissingValues was used to fill the missing values.

3.3.2 Feature selection

The objective of feature selection process in data preprocessing is to select an appropriate subset of features which can efficiently removes redundant and irrelevant data [15]. This process can play an important role in improving the data quality therefore the performance of the learning algorithm. Feature selection methods are categorized into wrapper-based and filter-based methods. Filter method uses variable ranking techniques to rank the features where the highly ranked features are selected and applied to the learning algorithm. In this study, 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.

3.3.3 Data transformation

Data transformation is another data preprocessing task applied to a given dataset. Data transformation can facilitate a better interpretation of information. In this context, some data mining algorithms may need the data to be transformed to categorical features and sometimes numerical values to represent class labels. Data transformation can facilitate a better interpretation of information. For some data mining algorithms, the data may need to be transformed to achieve normality and in other situations, categorical variables and sometimes numerical values may need to perform classification to represent class labels. There are several unsupervised global methods for transforming continuous attributes into discrete attributes, such as the equal-width method, equal-frequency method. In our case study, a manual discretization method is used which allows specifying the cut-off points. Finally, the required data set is transformed in a comma separated value (CSV) format that is suitable for WEKA software as shown in Figure 2.

```
@relation 'Most Update IT data'
@attribute Gender {M,F}
@attribute SSCEM numeric
@attribute CertiFicate {Art,SC}
@attribute Depart {Acct,Mang}
@attribute GPAS1 numeric
@attribute GPAS2 numeric
@attribute GPAS3 numeric
@attribute GPAS4 numeric
@attribute GPAS5 numeric
@attribute GPAS6 numeric
@attribute 'Final CGPA' {'Very good',Distinction,Good,Pass}

@data
M,79.4,Art,Acct,4.01,4.31,4.27,4.33,3.59,3.8,'Very good'
M,76.1,Art,Acct,4.25,4.27,3.96,3.71,3.41,3.73,'Very good'
F,75.9,Art,Acct,3.67,3.85,4.21,4.02,3.89,4.28,Distinction
M,75,Art,Mang,3.95,4.01,4.43,4.43,4.53,4.2,Distinction
M,57.3,SC,Mang,3.34,2.96,3.5,3.67,3.32,3.65,Good
M,78.4,Art,Acct,3.31,3.05,3.44,3.14,3.11,3.63,Good
M,66.1,Art,Acct,3.68,3.95,4.04,3.71,3.38,3.5,'Very good'
M,66.3,Art,Acct,2.83,2.72,3.23,3.46,3.29,3.56,Good
M,69.6,Art,Acct,3.57,2.9,3.52,3.21,2.87,3.06,Good
M,65.9,Art,Acct,3.73,3.4,3.61,3.75,4.04,4.04,'Very good'
M,63.9,SC,Mang,3.58,3.62,3.62,3.27,3.11,3.34,Good
M,62.3,SC,Mang,3.29,3.03,3.69,3.66,3.45,3.63,'Very good'
F,77.3,Art,Acct,2.57,2.5,2.98,2.97,3.04,3.53,Good
M,52.1,SC,Acct,2.79,2.5,3.1,2.5,3.78,3.46,Good
F,74.7,Art,Acct,3.85,3.8,3.73,4.24,3.6,3.87,'Very good'
M,54.3,Art,Acct,3,3.18,2.79,3.38,2.99,2.83,Good
M,76.7,SC,Acct,4.15,3.75,4.08,4.43,3.66,3.5,'Very good'
M,65.3,Art,Acct,3.47,3.44,3.72,3.7,3.38,3.57,Good
M,73.3,Art,Acct,3.42,3.49,3.32,3.84,2.97,3.2,Pass
M,71.3,Art,Acct,3.44,2.81,2.93,3.54,3.2,7,Good
```

Figure 2: Final dataset in CSV format

4. APPLICATION OF DATA MINING TECHNIQUES: Results and Analysis

This section describes the application of data mining techniques used for obtaining the prediction model of students' performance at the graduation. Further, the evaluation and interpretation of the model results are provided.

4.1 Classification

Classification is perhaps the most familiar and most popular data mining technique frequently performed by simply applying knowledge of the data. The purpose is to be able to use the model to predict the class of objects whose class label is unknown [16]. Predicting student's performance is an important task in educational environments. To build a predictive model, there are several DM techniques used, which are classification, regression and clustering. However, the most popular technique to predict students' performance is classification. There are several methods under classification such as Decision Tree (DT), Neural Networks (NN), Naive Bayes (NB) and K-Nearest Neighbor (K-NN). In this paper, the classification technique was executed through the application of the decision tree (J48) algorithm. The main benefit of applying the decision tree algorithm was that its results can be easily interpreted and its graphical representation that summarizes a model of implicit decision rules. The process of applying J48 algorithm in our case study includes two different scenarios. In the first scenario, students' GPA course scores from the first 3 years were used to predict the students CGPAs performance at graduation. Whereas, in the second scenario, the students' GPA of the first and second year course scores were used to predict performance of students early as possible before the graduation. If students' performance predicted to be low, so that extra efforts can be made to improve performance. For the validation objective of the proposed decision tree J48 model, 10-Fold Cross Validation test was used. In the 10-fold cross validation, the input dataset was randomly divided into 10 approximately equal subsets. At each time, one of the 10

subsets was used as the test set while the other 9 subsets (10-1) were used to form the training dataset. The purpose of the evaluation of a classification model is to get a reliable assessment of the quality of the target approximation represented by the model, which is called the model's predictive performance. Although, there are different evaluation measures can be used for the classification model performance depending on the intended application of the model. However, in this study, the accuracy performance of the J48 prediction model was evaluated in terms of correctly classified instances, incorrectly classified instances, and the kappa statistic and confusion matrix.

In this implement, the researchers build a classification tree model using the C4.5 method named in Weka 3.9 as J48 for predicting the students CGPA performance at graduation. The obtained result found that J48 has correctly classified 427 out of 522 instances. This means that about 81.8008 % of the instances were correctly classifies. Whereas, 95 of the instances were classified incorrectly which means that 18.1992 %. On the other hand, the kappa statistic became 0.7275, which indicates the existence of stronger statistical dependence as shown in Figure 3.
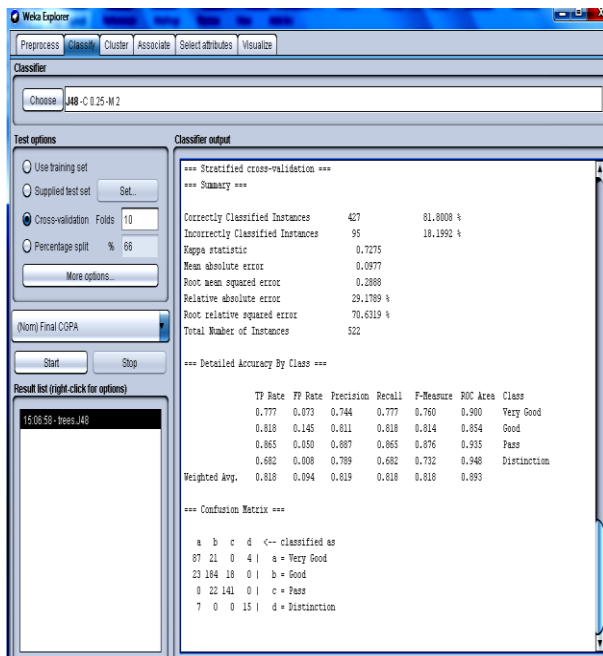

Figure 3: Classification result using all available data

The next step is to apply the feature selection method in order to select the most relevant attributes to the category of the class target. For that matter, information-gain attribute evaluation based on filter method was used to rank out the attributes as shown in Table 2.

Table 2: Features rank based on information-gain

Average Merit	Average Rank	Attribute
0.829 +- 0.026	1.6 +- 0.8	7 GPAS3
0.815 +- 0.013	1.9 +- 0.83	8 GPAS4
0.804 +- 0.016	2.5 +- 0.5	6 GPAS2
0.64 +- 0.024	4 +- 0	5 GPAS1
0.55 +- 0.019	5 +- 0	9 GPAS5
0.3 +- 0.028	6 +- 0	10 GPAS6
0.019 +- 0.003	7.6 +- 0.66	4 Depart
0.016 +- 0.003	8.4 +- 0.92	1 Gender
0.018 +- 0.023	8.8 +- 1.47	2 SSCEM
0.013 +- 0.003	9.2 +- 0.4	3 Certificate

As can observed from Table 2, the attributes 7,8,6,5,9,10 have got the best ranking attributes based on their Average Merit to the category of the class target. On the other side, the attributes 4,1,2,3 have got the least ranking average. Accordingly, the researchers began to gradually eliminating these attributes and assessing their weight in relation to the final prediction. Finally, after a series of experiments, the best accuracy rate is obtained when the least four attributes are eliminating from the given students dataset. Once again, the J48 classification algorithm was executed using the most relevant attributes. The obtained result indicates that the prediction accuracy of J48 improved to be (83.33% instead of 81.8008 %) as shown in Figure 4.
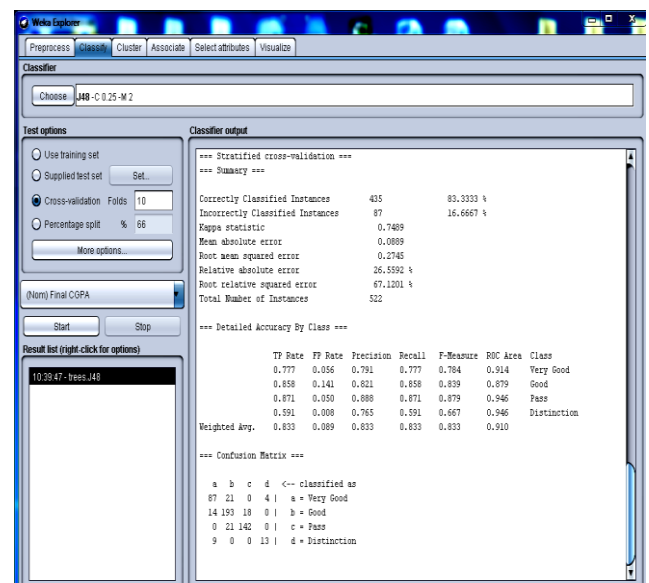

Figure 4: Classification result using most relevant attributes

Nonetheless, the main strength of individual classification trees is not from high accuracy of models, but from their interpretation ability as shown in Figure 5 and Figure 6 respectively.

```
GPAS4 <= 3.26
|    GPAS2 <= 3.03
|    |    GPAS1 <= 3.14
|    |    |    GPAS5 <= 3.22: Pass (97.0)
|    |    |    GPAS5 > 3.22
|    |    |    |    GPAS3 <= 2.98
|    |    |    |    |    GPAS4 <= 2.99: Pass (24.0/1.0)
|    |    |    |    |    GPAS4 > 2.99
|    |    |    |    |    |    GPAS5 <= 3.46: Pass (3.0/1.0)
|    |    |    |    |    |    GPAS5 > 3.46: Good (3.0)
|    |    |    |    GPAS3 > 2.98: Good (4.0)
|    |    GPAS1 > 3.14
|    |    |    GPAS4 <= 2.8
|    |    |    |    GPAS2 <= 2.85: Pass (13.0)
|    |    |    |    GPAS2 > 2.85
|    |    |    |    |    GPAS3 <= 2.86: Pass (9.0/2.0)
|    |    |    |    |    GPAS3 > 2.86: Good (4.0)
|    |    |    GPAS4 > 2.8
|    |    |    |    GPAS3 <= 2.59: Pass (4.0/1.0)
|    |    |    |    GPAS3 > 2.59: Good (27.0)
|    GPAS2 > 3.03
|    |    GPAS2 <= 3.57
|    |    |    GPAS4 <= 2.76
|    |    |    |    GPAS5 <= 3.08
|    |    |    |    |    GPAS6 <= 3.7: Pass (13.0/1.0)
|    |    |    |    |    GPAS6 > 3.7: Good (3.0)
|    |    |    |    GPAS5 > 3.08: Good (12.0)
|    |    |    GPAS4 > 2.76: Good (91.0/5.0)
```

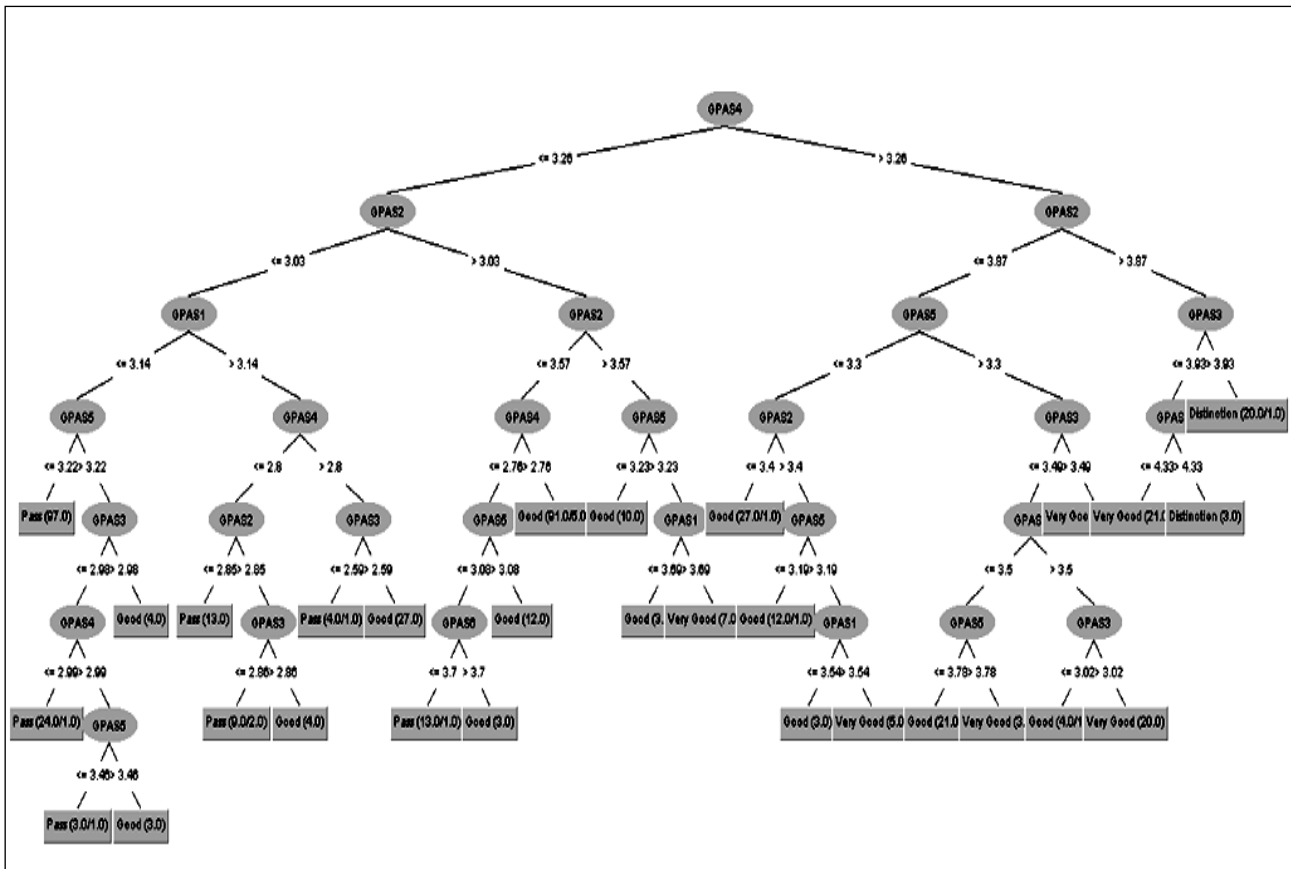**Figure 5: Partial rules generated by J48 model**

Figure 6: Prediction of students' final CGPA at graduation

As can be seen from Figure 6, GPAS4 found to be the root nod at the top of the tree. This means that, GPAS4 was the most closely related to students' final performance at graduation. Here, the tree generates 35 different rules. Decision tree rules can be easily transformed into the production rules which are easier for humans to interpret and understand. The production rules can be interpreted as follows:

1. If GPAS4 > 3.26 and GPAS2 > 3.87 and GPAS3 > 3.93, then the final CGPA of the student can be predicted as 'Distinction'.

2. If GPAS4 > 3.26 and GPAS2 > 3.87 and GPAS3 <= 3.93 and GPAS1 > 4.33, then the final CGPA of the student can be predicted as Distinction.

3. If GPAS4 > 3.26 and GPAS2 > 3.87 and GPAS3 <= 3.93

and GPAS1 <= 4.33, then the final CGPA of the student can be predicted as Very Good.

4. If GPAS4 > 3.26 and GPAS2 <= 3.87 and GPAS5 > 3.3 and GPAS3 > 3.49, then the final CGPA of the student can be predicted as 'Very Good'.

5. If GPAS4 > 3.26 and GPAS2 <= 3.87 and GPAS5 > 3.3 and GPAS3 <= 3.49 and GPAS1 > 3.5 and GPAS3 > 3.02, then the CGPA of the student can be predicted as Very Good, and so on all the other rules can be interpreted in the same way. Finally, Table 3 and Table 4 summarized the prediction performance of J48 model in terms of confusion matrix and accuracy prediction for the first scenario.

Table 3: Confusion Matrix of the first scenario

a	b	c	d	<-- classified as
87	21	0	4	a = Very Good
23	184	18	0	b = Good
0	22	141	0	c = Pass
7	0	0	15	d = Distinction

Table 4: Accuracy performance of the first scenario

J48 Model	Correctly classified instances	incorrectly classified instances	Accuracy rate	Kappa statistic
Using all attributes	427	95	81.8008 %	0.7275
Using most relevant attributes	435	87	83.33%	0.7489

After completing the experiment of the first scenario, the next step is to apply the prediction J48 model for our second scenario. As mention before, the second scenario is designed to predict the students CGPAs performance earlier to their graduation using only the GPAs grades from the first and second year of study. The obtained result found that J48 has

correctly classified 435 out of 522 instances. This led to 83.3333% of the instances were correctly classifies. Whereas, 95 of the instances were classified incorrectly which means that 18.1992 % of the instances as shown in Figure 7.
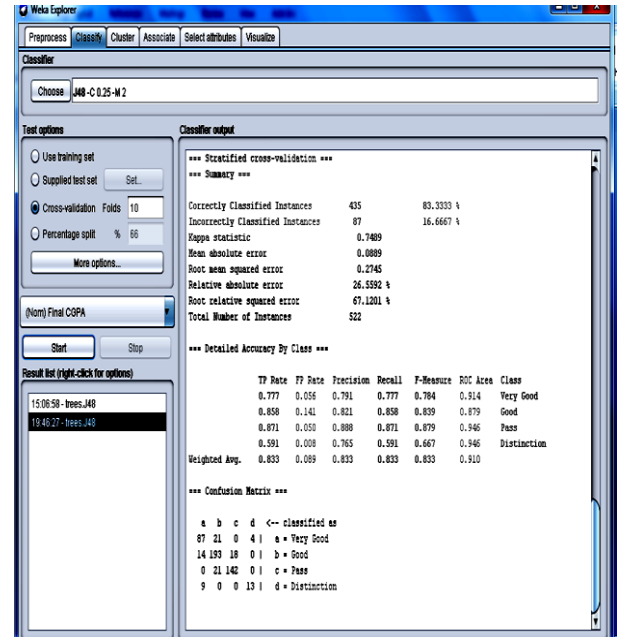

Figure 7: Classification result for the second scenario

The achieved result indicates that the accuracy of the model became slightly lower of 81.0345 % comparing to the prediction accuracy of the first scenario which was 83.33%. However, it was promising result in predicting the students CGPAs at graduation earlier as possible based on their GPAs courses score from the first and second year of study. Figure 8 shows the decision tree obtained for the second scenario experiment.
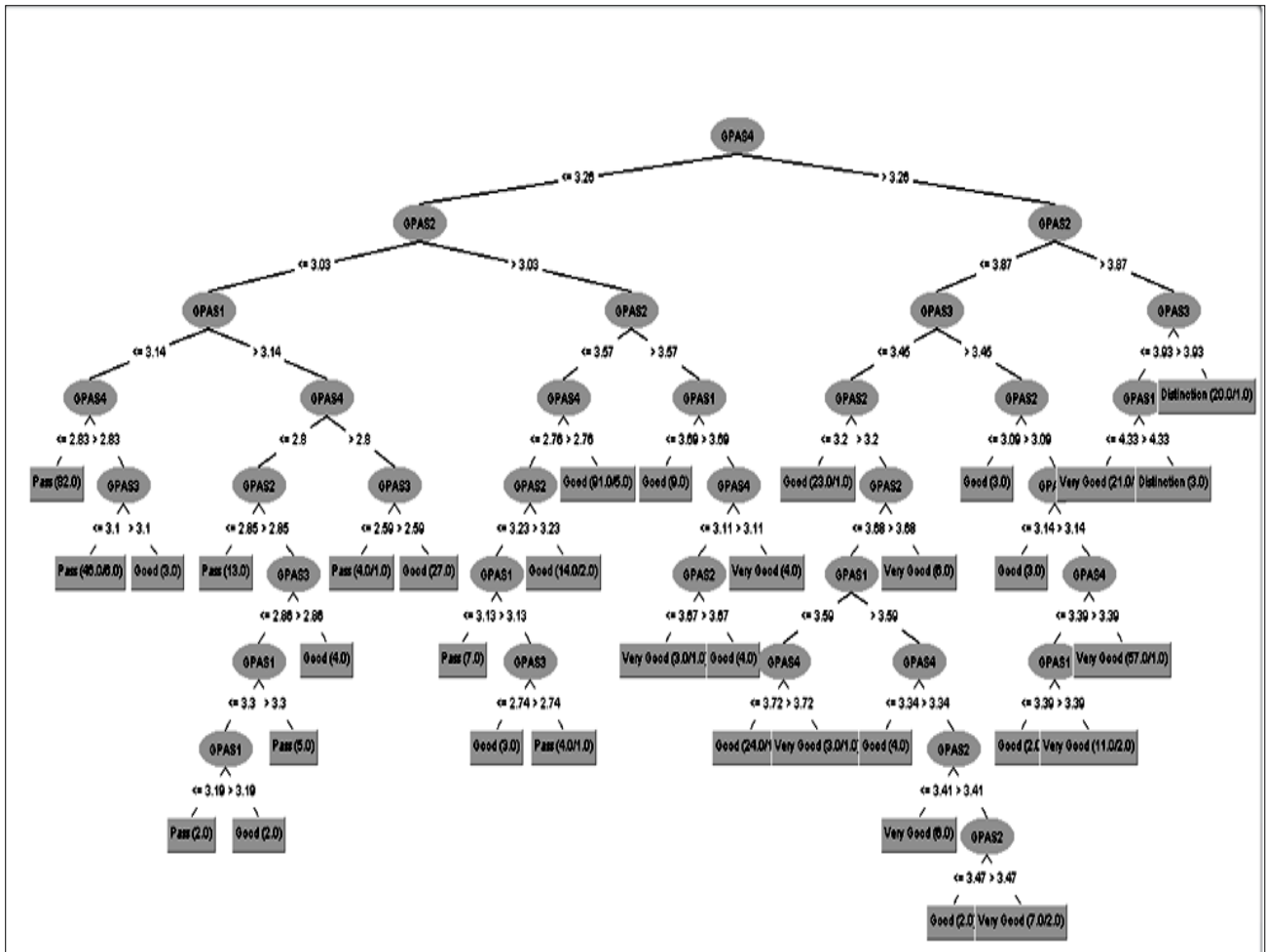
Figure 8: Prediction of students' CGPA at graduation using only GPAs from first and second year

Once again, as it is illustrated on Figure 8, the tree generates 35 different rules. Decision tree generate rules can be easily transformed into the production rules which are easier for humans to interpret and understand. The production rules can be interpreted as follows:

1. If GPAS4 > 3.26 and GPAS2 > 3.87 and GPAS3 > 3.93, then the final CGPA of the student can be predicted as Distinction.

2. If GPAS4 > 3.26 and GPAS2 > 3.87 and GPAS3 <= 3.93 and GPAS1 > 4.33, then the final CGPA of the student can be predicted as Distinction

3. If GPAS4 > 3.26 and GPAS2 > 3.87 and GPAS3 <= 3.93 and GPAS1 <= 4.33, then the final CGPA of the student can be predicted as Very Good.

4. If GPAS4 > 3.26 and GPAS2 <= 3.87 and GPAS3 > 3.45 and GPAS2 <= 3.09, then the final CGPA of the student can be predicted as Good.

5. If GPAS4 > 3.26 and GPAS2 <= 3.87 and GPAS > 3.45 and GPAS2 > 3.09, and GPAS1 <= 3.14, then the final CGPA of the student can be predicted as Good.

6. If GPAS4 <= 3.26 and GPAS2 <= 3.03 and GPAS1<= 3.14 and GPAS4 <= 2.83, then the CGPA of the student can be predicted as Pass.

7. If GPAS4 <= 3.26 and GPAS2 <= 3.03 and GPAS1<= 3.14 and GPAS4 > 2.83 and GPAS3 > 3.01, then the CGPA of the

student can be predicted as Good.

8. If GPAS4 <= 3.26 and GPAS2 <= 3.03 and GPAS1<= 3.14 and GPAS4 > 2.83 and GPAS3 <= 3.01, then the CGPA of the student can be predicted as Pass.

9. If GPAS4 <= 3.26 and GPAS2 <= 3.03 and GPAS1 < 3.14 and GPAS4 > 2.8 and GPAS3 > 2.59, then the CGPA of the student can be predicted as Good

10. If GPAS4 <= 3.26 and GPAS2 <= 3.03 and GPAS1 < 3.14 and GPAS4 > 2.8 and GPAS3 <= 2.59, then the CGPA of the student can be predicted as Pass, and so on all the other rules also can be interpreted in the same way. Therefore, the discovered hidden information can provide helpful and constructive knowledge to the department to make earlier intervention to assist at risk students to improve their performance and in turn, enhance their chance to graduate with high grade. Furthermore, the prediction result achieved by the J48 model for the second scenario was summarized in Table 5 and Table 6 consecutively.

Table 5: Confusion Matrix of the second scenario

a	b	c	d	<-- classified as
83	26	0	3	a = Very Good
23	18	18	0	b = Good
0	25	138	0	c = Pass
4	0	0	18	d = Distinction

Table 6: Accuracy performance of the second scenario

Correctly Classified Instances	423	81.0345 %
incorrectly Classified Instances	99	18.9655 %
Kappa value = 0.7151		

4.2 Association rule mining

Association rule mining is another well studied data mining tasks. It discovers relationships among attributes in databases, producing if-then statements concerning attribute-values [17]. This may take the form of attempting to find out which variables are most strongly associated with a single variable of particular interest, or may take the form of attempting to discover which relationships between any two variables are strongest. In this context, the most important relationships are identified by using support and confidence criteria. The association rule can be viewed as a two parts process. They are antecedent (if part) and a consequent (then part). In this study, the association rule technique was applied by primarily focusing on the Apriori algorithm to find the correlation based data mining works [18]. In this study, authors applied Apriori algorithm over discredited graduates' dataset in order to discover useful knowledge. In doing so, the minimum support 0.03, and the minimum confidence 0.9 and numRules = 100 which is the maximum number of rules to be performed were used. Figure 9 depicts the association rules discovered from the graduate students dataset.

```
1. GPAS2='(2.534-2.758)' GPAS4='(2.51-2.73)' 33 ==> Final CGPA=Pass 33    conf:(1)
2. GPAS3='(2.544-2.768)' GPAS5='(2.49-2.73)' 28 ==> Final CGPA=Pass 28    conf:(1)
3. Major=Mang GPAS2='(2.534-2.758)' GPAS3='(2.544-2.768)' 28 ==> Final CGPA=Pass 27    conf:(0.96)
4. GPAS2='(2.534-2.758)' GPAS3='(2.544-2.768)' 41 ==> Final CGPA=Pass 39    conf:(0.95)
5. Certificate=Art Major=Mang GPAS2='(2.534-2.758)' 38 ==> Final CGPA=Pass 36    conf:(0.95)
6. Certificate=Art GPAS3='(2.544-2.768)' GPAS4='(2.51-2.73)' 32 ==> Final CGPA=Pass 30    conf:(0.94)
7. GPAS3='(2.544-2.768)' GPAS4='(2.51-2.73)' 46 ==> Final CGPA=Pass 43    conf:(0.93)
8. Certificate=Art GPAS2='(2.534-2.758)' GPAS3='(2.544-2.768)' 29 ==> Final CGPA=Pass 27    conf:(0.93)
9. Certificate=Art GPAS2='(2.534-2.758)' 57 ==> Final CGPA=Pass 53    conf:(0.93)
10. Gender=M Certificate=Art GPAS4='(2.534-2.758)' 37 ==> Final CGPA=Pass 34    conf:(0.92)
11. Gender=F GPAS4='(2.51-2.73)' 36 ==> Final CGPA=Pass 33    conf:(0.92)
12. Major=Mang GPAS3='(2.544-2.768)' GPAS4='(2.51-2.73)' 32 ==> Final CGPA=Pass 29    conf:(0.91)
13. Major=Mang GPAS2='(2.534-2.758)' 52 ==> Final CGPA=Pass 47    conf:(0.9)
14. Certificate=Art Major=Mang GPAS4='(2.51-2.73)' 40 ==> Final CGPA=Pass 36    conf:(0.9)
15. Certificate=Art Major=Mang GPAS2='(3.206-3.43)' 30 ==> Final CGPA=Good 27    conf:(0.9)
```

Figure 9: Rules resulted by the Apriori algorithm

In interpretation of the discovered rules in the association rules model: as an example, the first rule says that, if the students GPAS2 is [2.534-2.758] and GPAS4 is [2.51-2.73], then the student graduate with CGPA grade "Pass". Rule 2 says that, if the students GPAS3 is [2.544-2.768] and GPAS5 is [2.49-2.73], then the student graduate with CGPA grade "Pass". Rule 3 says that if the graduate student if the student major is 'manage' and GPAS2 is equal [2.534-2.758] and GPAS3 is [2.544-2.768], and then the student graduate with CGPA grade "Pass". Once again, in rule 8 if student certificate is 'Art' and GPAS2 [2.534-2.758] and GPAS3 is [2.544-2.768], then the student graduate with CGPA 'PASS' and so on all the other rules can be interpreted in similar ways.

In concluding, the obtained rules provided knowledge to the department to help weak students to improve their performance. as well as, the these rules can help the administration in evaluating and enhancing the educational process requires deep and enough knowledge for assessing, planning and taking decisions especially during the college admission process.

5. CONCLUSION

Predicting students' performance from their academic performance is one of the most popular applications of educational data mining and, therefore, it has become a valuable source of information that has been used for different purposes. In this context, EDM is a field of data mining application that can assist higher educational institutions particularly to predict their students' performance using their academic record and accordingly help institutions in improving managerial decision making and student learning. The main objective of this paper is to predict performance of students which is measured in terms of cumulative grade point average (CGPAs) at graduation. The proposal was applied to the graduate students academic data for students completed their study in the department of Information technology degree at Comboni College of Science and Technology. The data include eight years period [2007-2015]. In this paper, the prediction algorithm (J48) was applied to predict students' final CGPAs performance at graduation. Two different scenarios were investigated in this study. The first scenario was designed to predict the final CGPAs at graduation of students according to their GPAs of course scores completed during their 3 years of study. As 10-fold cross validation result indicates that J48 algorithm obtained accurate prediction of 83.3333% for the first scenario. On the other hand, the second scenario used the students' GPAs from the first 2 years scores to predict the final CGPAs at graduation. As the 10- fold cross validation test, the result indicates, that it is possible to predict the student graduation performance, which is measured by CGPA using only scores of the first, and second year courses score at accuracy value equals to 81.0345 %. In conclusion, for the two scenarios, results indicated that CGPAs after the first 3 years of study demonstrated improved predictions for J48 algorithm. Furthermore, in this paper the Apriori algorithm was applied with the purpose of discovering useful knowledge from the graduate students' dataset. As the result indicates, helpful rules were discovered. In concluding, the study proved that data mining techniques can be used for predicting performance of students in institutions of higher education, especially universities. As the study indicates, the discovered results capable to provide helpful feedback and recommendation to the college management, department planners to offer a better guideline for students' regarding to their last year performance. This work is part of other large research study. Therefore, in the coming future work, the given graduate students data can be further analyzed and evaluated using several prediction techniques and algorithms in data mining like k-nearest Neighbor (K-NN), Naïve Bayes using more expanded dataset.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Romero, C. and Ventura, S. 2007.Educational data

mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 2007, 135–146.

[2] Han, J. and Kamber, M. 2001.Data Mining: Concepts and Techniques. Simon Fraser University, Morgan Kaufmann publishers.

[3] Peña-Ayala, A. 2014. Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications, 41(4 PART 1), 1432–1462.

[4] Durairaj, M., & Vijitha, C. 2014. Educational data miningfor prediction of student performance using clusteringalgorithms. International Journal of Computer Science andInformation Technologies (IJCSIT), 5(4), 5987–5991.

[5] Kolo, K. D., Adepoju, S. A., & Alhassan, J. K. 2015. Adecision tree approach for predicting students' academicperformance. International Journal of Education and ManagementEngineering, 5, 12–19.

[6] Tekin, A. 2014. Early prediction of students' grade point averages at graduation: A data mining approach. EurasianJournal of Educational Research, (54), 207–226.

[7] Dekker, G.W., M. Pechenizkiy, and J.M. Vleeshouwers. 2009. Predicting students drop out: A case study.EDM '09-Educational Data Mining: 2nd International Conference on Educational Data Mining, 2009. 2: p. 10.

[8] Sikder, M.F., M.J. Uddin, and S. Halder. 2016. Predicting Students Yearly Performance using Neural Network: A Case Study of BSMRSTU.5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016. 5: p. 6.14.Millán, E., T. Loboda, and J.L. Pérez-de-la-Cruz, Bayesian networks for student model engineerin.Computers and Education. Elsevier Ltd, 2010. 55(4): p. 20.

[9] Witten, I.H. and Frank, E. 2005. Data Mining Practical Machine Learning Tools and Techniques. San Francisco, CA: Morgan Kaufmann Publishers

[10] Muluken Alemu Yehuala, 2015.Application Of Data Mining Techniques For Student Success And Failure Prediction, a Case Of Debre_Markos University.INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 4, ISSUE 04, APRIL 2015.

[11] Mohammed M. Abu Tair, Alaa M. El-Halees, 2012.Mining Educational Data to Improve Students' Performance:A Case Study. International Journal of Information and Communication Technology Research, Volume 2 No. 2, February 2012.

[12] Fiseha Berhanu and Addisalem Abera, 2015. Students' Performance Prediction based on their Academic Record. International Journal of Computer Applications (0975 – 8887)Volume 131 –No.5, December2015

[13] Sadiq Hussain, Neama Abdulaziz Dahan, Fadl Mutaher Ba-Alwib, Najoua Ribata,Educational,2018. Data Mining and Analysis of Students' Academic Performance Using WEKA. Indonesian Journal of Electrical Engineering and Computer ScienceVol.9, No.2, February2018, pp. 447~459

[14] C. Romero, J. R. Romero and S. Ventura, 2014. "A survey on pre-processing educational data", In Educational Data Mining. Springer International Publishing, (2014), pp. 29-64

[15] A. G. Karegowda1, A. S. Manjunath2 and M.A. Jayaram3, 2010. "Comparative study of attribute selection using gain ratio and correlation based feature selection", International Journal of Information Technology and Knowledge Management, vol. 2, no. 2, (2010), pp. 271-277.

[16] Hämäläinen, W., & Vinni, M. 2011.Classifiers for educational data mining. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), Handbook of Educational Data Mining (pp. 93-106). Boca Raton, FL: CRC Press. (2011).

[17] Agrawal, R., Imielinski, T., and Swami, A.N. 1993.Mining association rules between sets of items in large databases. In Proceedings of SIGMOD, Washington, DC, (1993). pp. 207–216.

[18] Shrivastava, A.K. and R.N. Panda, 2014.Implementation of Apriori Algorithm using WEKA.KIET International Journal of Intelligent Computing and Informatics, 2014. 1(1): p. 4.