Students' Performance Prediction based on their Academic Record

Article *in* International Journal of Computer Applications · December 2015

DOI: 10.5120/ijca2015907348

2 authors, including:

Fiseha Berhanu
Zhejiang Lab
6 PUBLICATIONS **18** CITATIONS

SEE PROFILE

Students' Performance Prediction based on their Academic Record

Fiseha Berhanu
(MSC)
College of Engineering & Technology
Lecturer at Computer Science Department
Dilla University, Dilla, Ethiopia

Addisalem Abera
(MSC)
College of Engineering & Technology
Lecturer at Computer Science Department
Dilla University, Dilla, Ethiopia

ABSTRACT
Because of rapid increasing of data in educational environment, educational data mining emerged to develop methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings in which they learn. In this paper using a concept of educational data mining students' performance is predicted based on their academic record, using a decision tree algorithm. 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 RapidMiner 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%.

Keywords
Performance, prediction, academic record, educational data mining, decision tree

1. INTRODUCTION
The increase in both instrumental educational software as well as state databases of student's information have created large repositories of data reflecting how students learn [1]. To extract knowledge from this educational data a new discipline known as Educational Data Mining (EDM) emerged; EDM is a one of the application of data mining that is conducted in educational environment [2]. Furthermore, EDM is a new growing research and emerging discipline, concerned with data from academic field to develop various methods and to identify unique patterns which will help to explore student's academic performance. EDM can also be considered as learning science, as well as a feature of data mining [2] [3]. As cited by Hany M. and Harb [2], Ramaswami M. and Bhaskaran R [4] described that prediction of student performance with high accuracy is useful in many contexts in all educational institutions for distinguishing students who are likely to have low academic achievements. On one hand, the end product of models would be beneficial to teachers, parents and educational planners.

Educational data mining uses many techniques such as decision tree, rule induction, Aprioir, neural networks, knearest neighbor, naïve Bayesian and etc.; by using these techniques, many kinds of knowledge can be discovered such as association rules, classifications and clustering [5].

This study focuses on student academic record to address attribute that determines the performance of the students. The main objective of the study is to discovering knowledge from educational data that helps to predict student academic performance using classification algorithm. Since the main objective of higher institution is providing of better quality of education, the result of this study would play a big role in a way to achieve highest level of quality of education in higher education system by generating a classification model that can be used to offer a helpful and constructive recommendations to the curriculum designers, student advisors, teachers and as well as students. On the other hand the output of the study will enhances guiding and decision making process, and helps to produce students' with excellent academic performance by minimizing student dropout and dismissal rate.

2. RELATED WORK
To discover different patterns that can improve students' performance, many studies have been conducted around education data mining. Some of related works that have been done so far discussed on this section:

In their work Mohammed M. and Abu Tair [6], carried out a case study using educational data mining (DM) techniques and algorithms in order to improve graduate students' performance, and to overcome the problem of low grades of graduate students. They applied DM techniques on fifteen years graduate students data collected from the college of Science and Technology – Khanyounis. The applied data mining techniques are: association rule analysis, classification, and clustering and outlier detection. In each of these four techniques, they extracted important knowledge and address that the possibility of applying different data mining techniques on educational data and proves its importance on educational data.

Kannammal [7], presented a systematic analysis of various features of the higher grade school public examination results data in the state of Tamil Nadu, India. To predict the performance of schools they applied through different data mining classification algorithms. Their finding helps the parents to select the right city, school and factors that contribute to the success of the results in schools of their children. Their work focused on two fold factors namely Machine Learning algorithms to predict School performance with satisfying accuracy and to evaluate the data mining technique which would give better accuracy of the learning algorithms. They found that there exist some apparent and some less noticeable attributes that demonstrate a strong correlation with student performance. They collected data through the credible source data preparation and correlation analysis. Their work addressed those public examinations results data was a very helpful predictor of performance of

school in order to improve the result with maximum level and also improved the overall accuracy.

Bhardwaj and Pal [8] conducted study on the student performance based by selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University,Faizabad, India. By means of Bayesian classification method on 17 attributes, it was found that the factors like students‟ grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance.

Suchita Borkar [9], address student's performance evaluation using association rule mining algorithm based on various attributes of the dataset of 60 students from a single department. In this study important rules are generated to measure the correlation among various attributes between attributes like students graduation percentage, attendance, assignment work, unit test performance and how these attributes affect the student's university result. They use Weka and real time data set available in the college premises to preprocess and process the dataset. They observed that the associations they get from Apriori algorithm are not identical with the correlation values of the attributes.

So far no paper is done in Ethiopia and educational data mining is not attracted researchers attention. However, from these work discussed in this paper the researchers observed that the possibility of applying different DM techniques on different educational data to obtain different knowledge. Hence, in this work the variable that determines the performance of student is examined by classification algorithm.

3. TOOLS, METHODS AND MATERIALS

3.1 The dataset and tools

This study was conducted on 5 years [2009-2014] undergraduate student datasets of department of horticulture, College of Agriculture, Dilla University; the dataset is integrated from two datasets. The first dataset is student university placement data. This dataset contains student's data like name, Identification number (ID), ESLCE1 (Ethiopian School Leaving Certification Exam) result and etc. Totally this dataset contains eight (8) attributes; it is sent to every public university annually by Ministry of Education of Ethiopia. The dataset used in this work specifically contains lists of information of students that are placed to Dilla University. The second dataset is known as student master file it contains student name, ID, lists of student course grades, course codes, and semester status; this dataset contains 43 course code2 attributes and six other attributes that describes profile and status of students, it contains a total of 49 attributes.

To preprocess and process these datasets RapidMiner (RM) software has been used. This tool is selected for several reasons: it is the world-leading open-source system for data mining, and it is available as a stand-alone application for data analysis. Moreover, thousands of applications of RM give their users a competitive edge in more than 40 countries [10].It is also the most powerful, easy to use and provide intuitive graphical user interface for the design of analytic processes [11].

3.2 Methods

As cited by Romero [12] , C. Romero [13], discribed that the EDM process converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice. He described the EDM process does not differ much from other application areas of DM, like business, genetics, medicine, etc., because it follows the same steps as the general DM process: preprocessing, DM, and post-processing. In similar way, in this paper the same DM steps and methodologies applied to process a student datasets (see on figure 1).

The steps and carried out process that is seen on figure 1 is discussed as follows:

1. The Data Integration

combines data from multiple sources into a coherent data store, as in data warehousing. These sources may include multiple databases, data cubes, or flat files [14]. In this study two datasets are integrated from different datasets. During this step attributes that has the same name is merged as one attribute and attribute that has the same value is adjusted to a common value. In addition, the same attribute that have different name representation is adjusted to common attribute name. Generally, in this step data inconsistence is resolved and one integrated dataset is created in this step; a total of 199 list of student record was found. Sample record of the dataset are shown on table 4.

2. Preprocessing

Before applying DM algorithms it is necessary to carry out some pre-processing tasks such as cleaning, integration, discretization and variable transformation [15]. This step helps to: reduce confusion during learning, and to get better input data for data mining techniques [14]. Hence, after data integration the preprocessing is done before loading the dataset to the RapidMiner. Irrelevant attributes, highly inconsistent attributes, attribute that have same value and attribute that contain high missing value were removed from the dataset. During this processes attributes are reduced from the datasets and the numbers of attribute become less than the attribute in the original dataset.

[1] ECLSE: In Ethiopian educational system after the students are graduate from high school (grade 11 and grade 12) they will take national examination, this examination is known as ECLSE. On this examination if the student scores a pass mark they will join public higher institution in Ethiopia.

[2] Course code: is a code given for every course the student takes in every semester. It stands for course titles in the curriculum. E.g. the course code "COMP203" is stands for course "basic computer skill". In this study the course code is directly taken from the existing horticulture department curriculum.

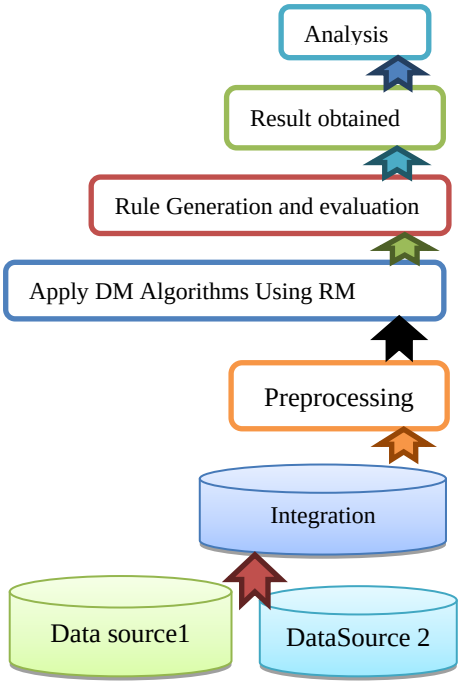

Figure 1: Used work methodology and steps.

For example, attributes such as the Student_Name, Date_of_Birth, Student_ID, etc. are not selected to be part of the mining process; because they do not provide any knowledge for the data processing and they present personal information of the students, also they have very large variances or duplicates information which make them irrelevant for data mining. The lists of selected 49 attributes are seen in Table 1.

Table 1: List of selected attributes

Attributes & Description	Value
SEX : Gender of student	{Male or Female}
Y1 Status(year 1 status) :The status of student in first year	{P(pass),D(Dismiss),DL(Dean List),DG(Dismissal for God),DO(Dropout), W(warning), WD(Withdrawn), Probation}
Y2 (Year 2 status): The status of student in second year.	{P,D,D in 1st Year, DG in 1st Year, DL, DO, DO in 1st Year, Probation, WD, WD in 1st Year, FWD}
Y3 STATUS (year 3 status): The status of student in third year.	{P,D,D in 1st Year,D in 2nd Year, DG in 1st Year, DL,DO in 1st Year, FWD(forced withdrawn), WD, WD in 1st Year, WD in 2nd Year}
School_Code: A Unique Identification number given for all high schools in Ethiopia.	A number that contains a group 8 digits: e.g. 11220101
ESLCE: The examination result that lets students to join higher institution. The exam is corrected out of 700.	{310,311,312,…700}
Courses codes	CEST201,HORT201,PISC221,HORT202,PISC342,COMP203,STAT201,ENGL201,AGEM212/252,HORT203,PISC212,PLSC231,PISC241,HORT301,HORT331,HORT311, HORT302,HORT303,HORT304,HORT321,HORT351,HORT305,HORT306,HORT307, PLSC322/HORT203, HORT322/PLSC203, AGEM473,HORT461/341,HORT432,HORT412,AREM341,HORT341,HORT452,HORT442,HORT401,HORT402,HORT403,HORT433,AREM346,HORT404,AREM342,HORT 405, HORT406: List of courses the student takes at each academic year.	{A+,A,A-,B+,B,B-,C+,C,C-,D+,D,D-,F}

From table 1 the Y1 Status, Y2 Status and Y3 Status attributes value is discussed as follows. In the original dataset the semester status of student is provided in student dataset into two forms; in continues form (Semester Grade Point Average (SGPA)3/Cumulative Grade point Average4(SGPA)) and discrete format (like pass, dismiss, dean list and etc.). From these two forms of student status the discrete format is selected for classification purposes and continues value was reduced from the dataset. The categorical value and description are directly taken from Dilla University office of registrar. The detail on how the continues value is changed to its equivalent categorical value and its description will be discussed as follows:

- **WD (Withdrawn)** student who has formally withdrawn from the program within eight weeks after the beginning of the subsequent semester.
- **DO (Dropout)** a student who has not withdrawn from a program in accordance with the withdrawal procedures set forth and within specified time limit.

[3] SGPA: $\frac{\sum(\text{Taken course credit hour} * \text{Scored grade point})}{\sum \text{course Credit hours}}$

The says the SGPA is a summation of a product of scored grade with course credit hour divided by a summation of course credit hours taken in the current semester.

E.g. assume the student took 4 courses in a semester, courses named A, B, C, D, and these courses credit hour is 4, 3, 3, and 2 accordingly. Let assume again the student scored grade A, B, A, C, for each courses accordingly. So, according to Dilla University legislation the grade point equivalence for letter grade is: A=4, B=3, C=2 and D=1.

Hence, $SGPA=(4*4+3*3+3*2+2*1)/(4+3+3+2)=2.67$

Then the SGPA of this student will be 2.67

[4] $CGPA=\frac{\sum(\text{Taken course credit hour} * \text{scored grade point})}{\sum \text{Previous semster credit hours} + \sum \text{Current semster credit hours}}$

Based on the previous example that we addressed in SGPA, assume that it's a previous semester grade point average. Let say at this semester the student took 3 courses: named F, G and H; and each courses credit hour is 3, 2, 4; and scored B, F, and C latter grade in each courses.

CGPA=(4*4+3*3+3*2+2*1+3*2+2*0+4*2)/(4+3+3+2+3+2+4)
=43/21=2.05

So, the student CGPA is 2.05.

- **Probation:** it is a status granted by the respective Academic Council/Department Council to allow student who fall below the required academic standards to continue their studies.
- **Dismissal for good (DG):** If the student can't able to score semester pass grade point fixed by registrar, the student will be dismissed for good.
- **Forced Withdrawn (FWD):** a student who has a maximum of three NG (no grade) cases and couldn't correct the NG case is forced to withdraw the college.
- **DO in 1st year and DO in 2nd Year:** is a student who dropout in first year and second year.
- **D (Dismissed) in 1st Year and D in 2nd Year:** a student dismissed in first and second year.

Other description of student status value like academic dismissal (D), pass (P), and warning (W) will be seen below on table 2. For instance on this table (table 2), if the student scores SGPA between 1.25 and 1.5, the student will be academically dismissed and the rest of the status can be interpreted accordingly. On the other hand, if student score a CGPA [1.50, 1.75) the student will pass and so on.

Table 2: Dilla University student status based on SGPA and CGPA

Class Year	Semester	Previous semester status	Academic dismissal(D)		Warning		Pass		Dean List
			SGPA	CGPA	SGPA	CGPA	SGPA	CGPA	CGPA
I	I	-	[1.25-1.50)	-	[1.50,1.75)	-	>=1.75		
	II	Pass(P)	<1.00	[1.50,1.75)	[1.50-1.75)	[1.75-2.00)			
		Warned(W)	<1.75&or	[1.75-2.00)	>=1.75 &	[1.75-2.00)			
II	I	Pass	<1.00	<1.75	<1.75 &/or	[1.75-2.00)		(2.00,3.25]	>=3.25
	I	Warned	<1.75 &/ or	<2.00	No waning	consecutive			
	II	Pass	<1.00	<1.75	<1.75 &/or	[1.75-2.00)			
	II	Warned	<1.75 &/ or	<2.00	No waning	consecutive			
III	I	Pass	<1.00	<1.75	<1.75 &/or				
	I	Warned	<1.75 &/ or	<2.00	No waning	consecutive			
	II	Pass	<1.00	<1.75	<1.75 &/or				
	II	Warned	<1.75 &/ or	<2.00	No waning	consecutive			

In this step also attributes that have a high missing value on the original dataset have been removed from integrated dataset. The students who are D, DO and FWD in any of the academic year is couldn't continue the next semester and take courses, then in this dataset the value of the course they didn't take are replaced by NT (Not Take) value. For instance if the student is D(dismissed) in first year, the courses code attribute value of the rest of semester will be replaced with NT and the final status (status in year 3) of the student will be dismissed in 1st year and so on. Finally, Attribute discretization has done on attributes like: ESLCE and 43 courses code attributes using equal-frequency binning5 (see: Table 3).

Table 3: Discretized attributes

Attribute	Range
Courses	{A-,A,A+}=A,{B-,B,B+}=B {C-,C,C+}=C,{D-,D,D+}=D,F
ESLCE	(310,320]=>310,(320,330]=>320, (330,340]=>330,(340,350]=>340, (350,360]=>350,(360,370]=>380,

[5]Equal frequency binning is a method of discretizing attribute values by applying equal-width [12]

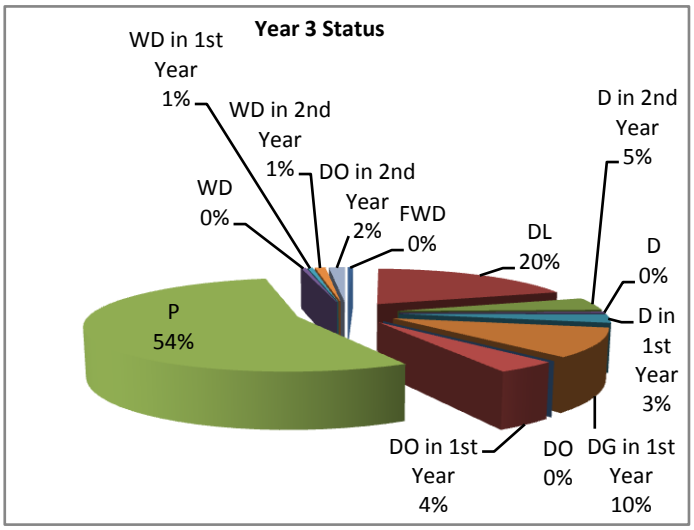

Figure 2: Student final year performance of (2009-2014)

Data was also analyzed visually to figure out the distribution of values, specifically the Y3status (final year status) of students(see figure 2). Figure 2 depicts the distribution of students in period from 2009 to 2014 according to final year (Year3) status, as it is apparent from the figure students who are graduated with P status is about 54% of the data set.

After dataset integration and preprocessing the next steps are: applying data mining algorithms, rule generation, obtain result, evaluation of result and analysis. These steps are discussed accordingly in the next sections.

4. CLASSIFICATION

Classification is a data mining task that predicts group membership for data in a given dataset [14]. As for the classification (one of the most useful educational data mining tasks in e-learning), there are different educational objectives for using classification, such as: to group students who are hint-driven or failure-driven and find common misconceptions that students possess [16].

A decision tree is a flowchart-like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an out- come of the test, and each leaf node (or terminal node) holds a class label [14]. Decision tree is so popular and have been used for classification in many application areas. It became so popular for several reasons: the construction of decision tree classifiers does not require any domain knowledge or parameter setting, decision trees can handle high dimensional data, easy to assimilate by humans, and the learning and classification steps of decision tree induction are simple and fast. On the other angle decision tree classifiers have very good accuracy [14].

5. EXPERIMENTS AND RESULTS

This section describes the rest of step illustrated on figure 1 such as: the experiments, data mining techniques used for obtaining the prediction models of students' academic status at the end of the academic year and evaluation and interpretation of the model.

Frist let see the experimentation and analysis of result. The classification algorithm experiment was conducted using RM tool in order to try to obtain the highest classification accuracy. In this experiment a decision tree classification technique is executed using all available information (49 attributes) (see: table 1). The tree classified the student dataset based on their final year status (Y3status attribute is labeled as a target class) and how other attributes affect them; the sample record of student dataset on which this algorithm were applied is seen on table 4. As it is seen in table 1 the Y3Status contains twelve (P,WD in 2nd Year,DL,D in 2nd Year,DG in 1st Year,DO in 1st Year,DO in 2nd Year,D in 1st Year,D,WD in 1st Year,WD,FWD) different values, hence the decision tree classified the dataset based on theses classes.

Table 4: Sample record from student dataset

Sex	CEST201	HORT201	…	…	HORT311	HORT302	AREM34	HORT433	School Code	AYear	ESLCE	Y1status	Y2 status	Y3 status
M	A	B	…	…	A	B	C	A	1101213	2009	>=330	P	P	P
M	B	B	…	…	C	B	A	A	2010111	2010	>=320	P	P	P
F	D	C	…	…	B	B	B	A	2012101	2010	>=310	W	P	P
F	A	A	…	…	NT	NT	NT	NT	2012101	2009	>=320	P	DO	DO in 2nd Year
M	C	C	…	…	C	D	NT	NT	2012101	2011	>=350	P	DG	DG in 2nd Year
M	C	B	…	…	C	C	B	A	1414101	2011	>=360	P	W	P
M	B	B	…	…	A	A	A	A	1001110	2010	>=370	P	DL	DL
F	D	C	…	…	A	B	C	B	1001112	2009	>=350	P	P	P
M	F	D	…	…	NT	NT	NT	NT	1001101	2009	>=340	D	D in 1st year	D in 1st Year
M	A	A	…	…	A	A	A	A	1112121	2011	>=380	DL	DL	DL

Key: **(…)** stands for list of other course code attributes not mentioned on the table.

Figure 4 illustrates the rules that resulted from applying the decision tree classification algorithm on the student data and as it is shown from the figure, the attributes that influence the category of the target class are: HORT401, HORT306, HORT461/341, HORT 203, STAT201, CEST201, AGEM212/252 HORT302, HORT452, AREM346, ENGL201, SEX, Y1Status, and Y2status.

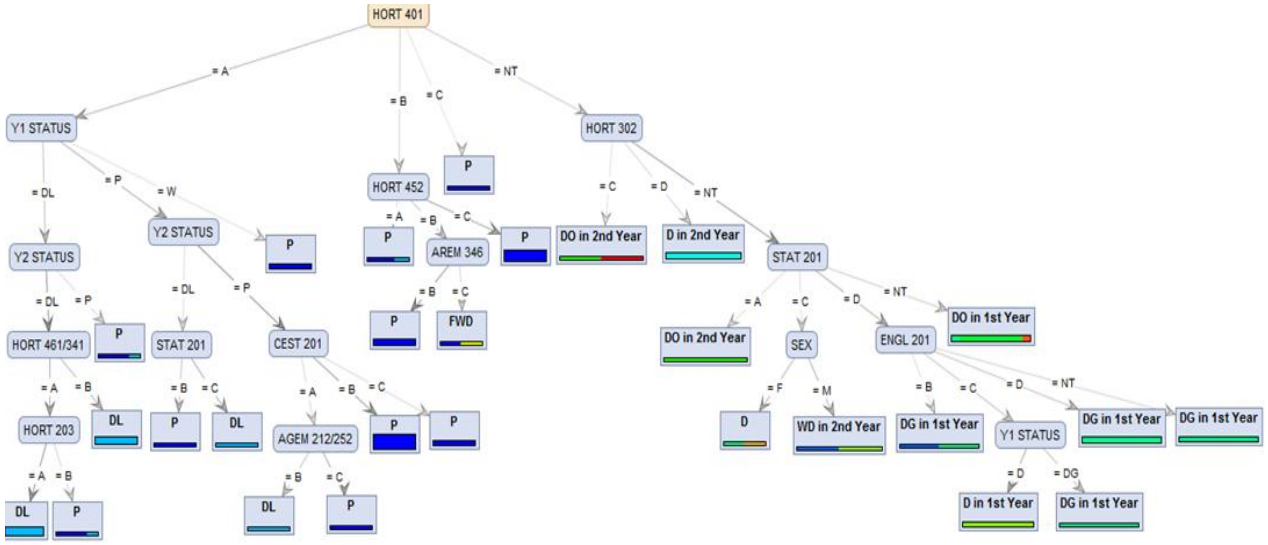

Figure 4: The decision tree result

As it is illustrated on figure 4 the tree generates 27 different rules and the rules are interpreted and discussed as follows:

1. If HORT 401 equal A, Y1 Status equal P, Y2Status equal DL, and STAT 201 equal C: then the final status of the student can be predicted as DL.
2. If HORT 401 equal A, Y1 Status equal P, Y2Status equal DL and STAT 201 equal B: then the final status of the student can be predicted as P.
3. If HORT 401 equal A, Y1 Status equal P, Y2Status equal P, CEST 201 equal A and AGEM 212/252 equal B: then the final status of the student can be predicted as DL.
4. If HORT 401 equal A, Y1 Status equal P, Y2Status equal P, CEST 201 equal A and AGEM 212/252 equal C: then the final status of the student can be predicted as P.
5. If HORT 401 equal A, Y1 Status equal P, Y2Status equal P and CEST 201 equal B: then the final status of the student can be predicted as P.
6. If HORT 401 equal A, Y1 Status equal P, Y2Status equal P and CEST 201 equal C: then the final status of the student can be predicted as P.
7. If HORT 401 equal A and Y1 Status equal W: then the final status of the student can be predicted as P.
8. If HORT 401 equal A, Y1 Status equal Dl, Y2Status equal DL, HORT461/341 equal A and HORT 203 equal B: then the final status of the student can be predicted as P.
9. If HORT 401 equal A, Y1 Status equal Dl, Y2Status equal DL, HORT461/341 equal A and HORT 203 equal A: then the final status of the student can be predicted as DL.
10. If HORT 401 equal A, Y1 Status equal DL, Y2Status equal DL and HORT461/341 equal B: then the final status of the student can be predicted as DL.
11. If HORT 401 equal A, Y1 Status equal DL and Y2 Status equal P: then the final status of the student can be predicted as P.
12. If HORT401 equal NT and HORT302 equal C: then the final status of the students can be predicted as DO in 2nd Year.
13. If HORT401 equal NT and HORT302 equal D: then the final status of the students can be predicted as D in 2nd Year.
14. If HORT401 equal NT, HORT302 equal NT, STAT201 equal A: then the final status of the students can be predicted as DO in 2nd year.
15. If HORT401 equal NT, HORT302 equal NT, STAT201 equal C, and SEX equal F: then the final status of the students can be predicted as D.
16. If HORT401 equal NT, HORT302 equal NT, STAT201 equal C, and SEX equal M: then the final status of the students can be predicted as WD in 2nd Year.
17. If HORT401 equal NT, HORT302 equal NT, STAT201 equal D, and ENGL 201 equal B: then the final status of the students can be predicted as DG in 1st year.
18. If HORT401 equal NT, HORT302 equal NT, STAT201 equal D, and ENGL 201 equal C and Y1 Status equal D: then the final status of the students can be predicted as D in 1st year.
19. If HORT401 equal NT, HORT302 equal NT, STAT201 equal D, and ENGL 201 equal C and Y1 Status equal DG: then the final status of the students can be predicted as DG in 1st year.
20. If HORT401 equal NT, HORT302 equal NT, STAT201 equal D, and ENGL 201 equal D: then the final status of the students can be predicted as DG in 1st year.
21. If HORT401 equal NT, HORT302 equal NT, STAT201 equal D, and ENGL 201 equal NT: then the final status of the students can be predicted as DG in 1st year.

22. If HORT401 equal NT, HORT302 equal NT and STAT201 equal NT: then the final status of the students can be predicted as DO in 1st year.
23. If HORT401 equal B, HORT452 equal B, and AREM346 equal B: then the final status of the students can be predicted as P.
24. If HORT401 equal B, HORT452 equal B, and AREM346 equal C: then the final status of the students can be predicted as FWD.
25. If HORT401 equal B and HORT452 equal C: then the final status of the students can be predicted as P
26. If HORT401 equal B and, HORT452 equal A: then the final status of the students can be predicted as P
27. If HORT401 equal C then the final status of the students can be predicted as P.

Based the above generated rules, the rules are discussed by categorized it in to four major classes: class excellent (DL), class good (P), class poor (D, DG) and class cases (DO, WD, FWD).

Class excellent (DL): Rule number 1, 3, 9 and 10 contains the student who has finished their study with DL status. This implies that for the student in order to graduate with excellent (CGPA>3.25) grade; they have to focus and score latter A or B grade in course mentioned in the rules. In addition they have to maintain or improve their academic record from semester to semester.

Class good (P): During data preprocessing in this work figure 2 illustrates the final year performance of the student. It shows that most (54%) the performance of the student is dominated by pass. Similarly in the decision tree model most of the rule is dominated by pass class. Like DL class in rule number 2,4,5,6,7,8,11,23,25,26,27 for students to graduate with P (CGPA 2-3.25) class; they have to focus and score letter grade C or above in particular course and must maintain their semester performance.

Class Poor (D, DG): Rule number 17, 18, 19, 20, and 21 illustrate that the student who scored C and below C with course STAT201 and ENGL201 will be dismissed in the first academic year. Rule number 15 illustrate those female students who has scored latter grade C with STAT201 course will be dismissed in first academic year. In addition Rule number 13 describes that the student who score grade D and below D in course HORT302 will be dismissed in the second academic year.

Class case (DO, WD, FWD): Rule number 12, 14, 16, 22 and 24 shows that the student interrupts from their study by several cases.

In general specific courses, semester status dominates the rules and also determines the fate of the student's weather to graduate with poor or excellent grade. On the other hand, from list of attributes none course related attributes like: ESCLCE and school code can't determine the performance of the students.

After decision tree algorithm is applied on the dataset the performance model was evaluated. The classification model was evaluated using a test dataset based on their classification accuracy. In the other word evaluation of the performance of the classifier is also made in terms of different confusion matrices (True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), False Negative Rate (FNR), Relative Operating Characteristics (ROC), the number of correctly classified instances, number of leaves, the size of the trees and execution time. In this work the confusion matrix of the classifier model was analyzed in terms of the following variables shown on Table 5 using RM.

Table 5: Sample summary of Confusion Matrix

		Actual performance		Total
		Positive	**Negative**	
Predicted performance	**Positive**	TP	FP	TN+FP
	Negative	FP	TN	FN+TP
	Total	FP+TP	TN+FN	TN+FP+ FN+TP

Key: TN (True Negative), False Positive (FP), FN (False Negative), True positive (TP).

The performance of the experiments is measured in the following manner as depicted below:

The effectiveness and efficiency of the model is computed in terms of recall (True Positive Rate) and precision (Positive Predicted Value) using RM tool as seen on table 6. As a result, recall can be computed for both positive and negative classes. So the formula that was applied by RM to calculate the recall is TP/ TP+FN. Similarly to compute the precision of the model the formula that was used by RM is TP/ FP+ TP. Based on these formula using RM tool the confusion matrix was generated and the result has seen on table 6. The overall performance the tree was calculated using accuracy formula (Accuracy = (TP + TN) / (TN + FP + FN + TP).From the confusion matrix result the RM tool calculate the accuracy based on the accuracy formula and get an accuracy of 84.95%, which is promising result.

Table 6: The confusion matrix result of the classificationmodel

	T P	T WD in 2nd Year	T DL	T D in 2nd Year	T DG in 1st Year	T DO in 1st Year	T DO in 2nd Year	T D in 1st Year	T D	T WD in 1st Year	T WD	T FWD	Class precision
pred. P	102	1	8	0	0	0	1	0	0	1	1	1	88.70%
pred. WD in 2nd Year	0	0	0	0	1	0	1	0	1	0	0	0	0.00%
pred. DL	6	0	31	0	0	0	0	0	0	0	0	0	83.78%
pred. D in 2nd Year	0	0	0	9	0	0	0	0	0	0	0	0	100.00 %
pred. DG in 1st Year	0	1	0	0	16	0	1	3	0	0	0	0	76.19%
pred. DO in 1st Year	0	0	0	0	1	8	0	0	0	0	0	0	88.89%
pred. DO in 2nd Year	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. D in 1st Year	0	0	0	0	2	0	0	3	0	0	0	0	60.00%
pred. D	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. WD in 1st Year	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. WD	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. FWD	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
class recall	94.44 %	0%	79.49%	100%	80%	100%	0%	50%	0%	0%	0%	0%	

Key: T (True) and Pred. (predicted)

6. CONCLUSION

From the study the researchers observed that Educational data mining (EDM) is an area full of exciting opportunities for researchers and practitioners. It is also the field that can assists higher educational institutions with efficient and effective ways to improve institutional effectiveness and student learning. On the other hand, this paper showed how useful data mining can be used in higher education particularly to predict students' performance using their academic record. The study was conducted with purpose that contribute to optimal managerial decision taking that can prevent students' dropout and dismissal, and to improve student learning abilities that enable students to graduate with good academic point.

The decision tree algorithm was implemented using RM tool, on preprocessed 49 attributes of 199 records; a total of 27 rules were generated. So, the result of decision tree model reveals that specific courses, student academic status in 1st and 2nd year and sex are attributes that determine the performance of student. On the other word, the result also implies that the student must sustain or improve their academic status and give focus on particular courses, in order to graduate with excellent academic result. The results presented in this paper are a part of a larger research which is used to make student performance predication and to be presented to the higher education institution managers, curriculum designers, teachers and student advisors to offer a better guideline for students' scholastic circumstances, their focuses regarding to course they take, and to predict some important aspects of their final year status.

In our future work the dataset will be further analyzed the student data using different datamining algorithms and techniques. Finally, the authors recommend all the higher institution to apply the same procedures mentioned in this study in order to help their student in different ways.

7. ACKNOWLEDGMENT

The authors wish to acknowledge Dilla University Research and Dissemination office for funding this research and Dilla University Office Registrar and Alumina for their cooperation in providing the necessary data.

8. REFERENCES

[1]. An open repository and analysis tools for fine-grained, longitudinal learner data. K. Koedinger, K. Cunningham,A. Skogsholm, and B. Leber. Montreal, QC, Canada, : s.n., 2008. 1st Int. Conf. Educ. Data Mining,. pp. 157-166.

[2]. Selecting Optimal Subset of Features for Student Performance Model. Hany M. Harb, Malaka A. Moustafa. 2012, International Journal of Computer Science Issues, pp. 253-262.

[3]. Knowledge Mining in Supervised and Unsupervised Assessment. Anwar, M. A., and Naseer Ahmed. s.l. : 2nd International Conference on Networking and Information Technology IPCSIT Vol 17, 2011. 2nd International Conference on Networking and Information. Vol. Vol. 17.

[4]. A Study on Feature Selection Techniques in Educational Data Mining. R, Ramaswami M. and Bhaskaran. 2009, Journal of Computing, pp. 253-262.

[5]. Data Mining Approaches on Detection of Students' Academic Failure and Dropout: A Brief Survey. Devikala.D M.phil and Kamalraj.N MCA, M.phil. s.l. : International Journal of Computer Trends and Technology (IJCTT), Aug 2014, Vol. volume 14 number 3. ISSN: 2231-2803.

[6]. Mining Educational Data to Improve Students' Performance:. Mohammed M. Abu Tair, Alaa M. El-Halees. 2012, International Journal of Information and Communication Technology Research, pp. 140-146.

[7]. Predicting Performance of Schools by Applying Data Mining Techniques on Public. Kannammal, J. Macklin Abraham Navamani and A. 2014, Research Journal of Applied Sciences, Engineering and Technology , pp. 262-271.

[8]. Data Mining: A prediction for performance. Pal, B.K. Bharadwaj and S. No. 4, s.l. : International Journal of Computer Science and Information Security (IJCSIS), 2011, Vol. Vol. 9.

[9]. Predicting Students Academic Performance Using Education Data Mining . Suchita Borkar, K. Rajeswari. 2013, International Journal of Computer Science and Mobile Computing, pp. 273-279.

[10]. Mawuna Remarque KOUTONIN. The Best Data Mining Tools You Can Use for Free in Your Company. silconafrica. [Online] [Cited: March 8, 2013.] http://www.siliconafrica.com/the-best-data-minning-tools-you-can-use-for-free-in-your-company/.

[11]. RapidMiner Studio-Rapid Miner. [Online] RapidMiner, 2015. [Cited: july 2015, 2015.] https://rapidminer.com/products/studio/.

[12]. Educational Data Mining: A Review of the State of the Art. Romero, Cristobal. 2010, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, pp. 601-618.

[13]. Knowledge discovery with genetic programming for providing feedback to courseware author. C. Romero, S. Ventura, and P. De Bra. 2004, User Model. User-Adapted Interaction: J. Personalization Res. , pp. 425-464.

[14]. Kamber, Jiawei Han & Micheline. Data Mining: Concepts and Techniques second edition. San Francisco : Morgan Kaufmann, 2006.

[15]. La deserción escolar en américa latina. León, E. Espíndola and A. no. 30, s.l. : Revista Iberoamer Educ., 2002, Vol. 1.

[16]. M.V. Yudelson, O. Medvedeva, E. Legowski, M. Castine, D. Jukic, C. Rebecca, Mining student learning data to develop high level pedagogic strategy in a medical ITS, in: AAAI Workshop on Educational Data Mining, 2006, pp. 1–8.