RESULT ANALYSIS USING CLASSIFICATION TECHNIQUES

Varsha Namdeo Anju Singh Divakar Singh Dr. R.C Jain

ABSTRACT

Data Mining is a powerful tool for academic intervention. The educational institutions can use classification for comprehensive analysis of students' characteristics. In our work, we collected student's data from engineering course. And then apply four different classification methods for classifying students based on their Final Grade obtained in their Courses. We compare these algorithms of classification and check which algorithm is optimal for classifying students' based on their final grade.

By this task we extract knowledge that describes students' performance in end semester examination. This work will help to the institute to improve the performance of the students.

I. INTRODUCTION

A. Data Mining

Data mining [2] [3] is the emerging field of applying statistical and artificial intelligence techniques to the problem of finding novel, useful, and non-trivial patterns from large databases.

Data mining techniques [3] [10] are the result of a long process of research and product development.

From the beginning, data mining research [4] has been driven by its applications. While the Sales & Marketing industries have long recognized the benefits of data mining, data mining techniques can be effectively applied in many areas. Recently, research in association rule mining, classification rule mining, and pattern discovery in combinatorial databases has provided many paradigms that are applicable in the wide application areas. As a result, more and more organizations have become interested in data mining.

A typical data mining architecture [2] is shown in "Fig. 1," which gives the overview of data mining process.

Newly developed web-based educational technologies offer researchers unique opportunities [9] to study how students learn and what approaches to learning lead to success.

Although using data mining in higher education is a recent research field, there are many works in this area. That is because of its potentials to educational institutes.

Data mining applications in higher education given in [5], they concluded with that the Data mining is a powerful analytical tool that enables educational institutions to better allocate resources and staff to proactively manage student outcomes and improve the effectiveness of alumni development.

One of the biggest challenges that higher education faces today is predicting the paths of students [5].Institutions would like to know, for example which student will enroll in particular course, and which students will need assistance in order to graduate.

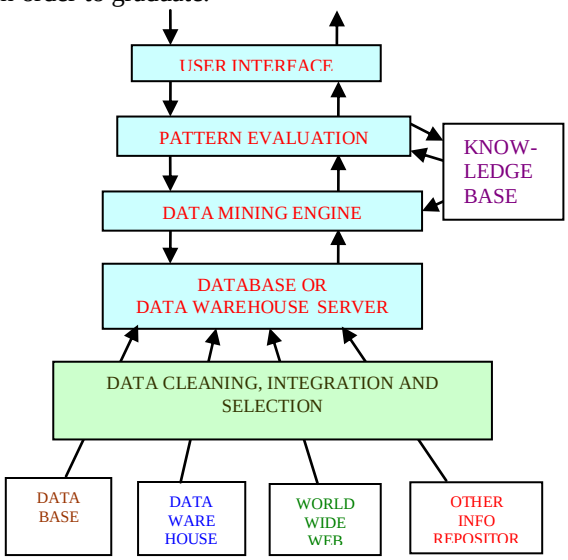


Fig. 1. Architecture of a typical data mining system**.**

One way to effectively address these student challenges is through the analysis and presentation of data or data mining. Data mining is based on four essential methods classification, categorization, estimation and visualization.

B. Educational Data Mining

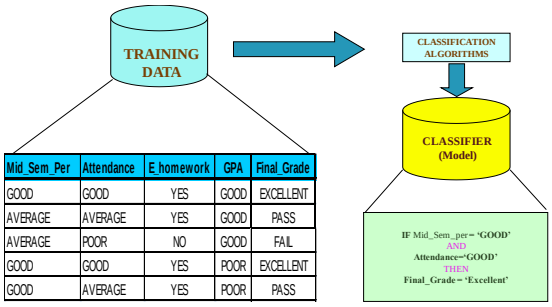
Educational Data Mining (**EDM**) is the process of transforming raw data compiled by education systems in useful information that could be used to take informed decisions and answer research questions.

n recent years, there has been increasing interest in the use of data mining to investigate scientific questions within educational research [6], an area of inquiry termed educational data mining.

Educational data mining (also referred to as "EDM") is defined as the area of scientific inquiry centered on the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in.

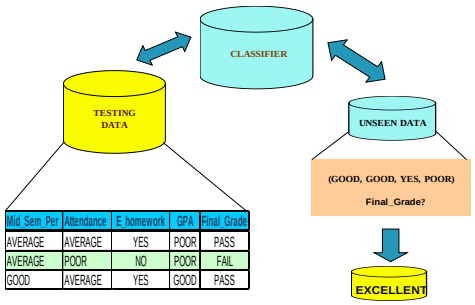
C. Classification

Classification [11] is a data mining (machine learning) technique used to predict group membership for data instances. Fig. 2 shows the classification process.



Mid_Sem_Per	Attendance	E_homework	GPA	Final_Grade
GOOD	GOOD	YES	GOOD	EXCELLENT
AVERAGE	AVERAGE	YES	GOOD	PASS
AVERAGE	POOR	NO	GOOD	FAIL
GOOD	GOOD	YES	POOR	EXCELLENT
GOOD	AVERAGE	YES	POOR	PASS

Fig. 2(a) Classification Process (1): Model Construction



Mid_Sem_Per	Attendance	E_homework	GPA	Final_Grade
AVERAGE	AVERAGE	YES	POOR	PASS
AVERAGE	POOR	NO	POOR	FAIL
GOOD	AVERAGE	YES	GOOD	PASS

Fig.2(b) Classification Process (2): Use the Model in Prediction

D. Outline of the Paper

The rest of the paper is organized as follows: Section 2 summaries previous works in educational data mining. Section 3 gives a general description of proposed work which include collection and preparation of data and applying classification methods. Section 4 gives the experimental results. Finally in section 5 we conclude this paper with a summary and 6 gives outlook for future work.

II. PREVIOUS WORKS

In [1], author used educational data mining to analyze learning behavior. They collected the data from database course. After preprocessing the data they applied data mining techniques to discover association, classification, clustering and outlier detection rules.

In educational data mining, given works of a student, one may predict his/her final grade.

In their case study they used decision tree to represent logical rules of student final grade. The benefit of this method is that it can predict low grades on time.

In their work, Alaa El-Halees suggested that classification could be done using more data mining techniques such as neural nets, naive bayes classifiers etc.
From the above work we concentrate on the classification part, and in the proposed work we have been applied decision tree (ID3), Multilayer Perceptron, Decision Table as well as naïve bayes Network for classifying students based on their final grade.

III. PROPOSED WORK

A. Data Collection and preparation

We collect the data from student database. From which we take five attributes attendance, mid_ Sem_percentage, gpa, e_homework and final_grade for analysis.

Data Preparation

We discretized the numerical attributes to categorical ones. For example, we grouped all grades into three groups Excellent, Pass & Failure as described in table below.

TABLE I
VALUES OF FINAL GRADE

Final_percentage	Final_grade
X < 35%	Fail
35% ≤ X ≤ 70%	Pass
X > 70%	Excellent

In the same way, we descretized other attributes such as attendance, e_homework, mid_ sem_ percentage and gpa.

Finally the most significant attributes presented in following table:-

TABLE II
THE SYMBOLIC ATTRIBUTE DESCRIPTION

Attribute	Description	Possible Values
Mid- Sem- per	Percentage of marks obtained in mid semester exam	Average Good, fail
Attendance	Attendance in one semester	Poor, Good, Average
E_ home work	Online exercise given by teacher done or not	Yes, No
GPA	Grade Point Avg. for General performance i.e. in lab or class or extra curricular	Good, Poor
Final – Grade	Final – Grade obtained in current semester	Pass, Fail, Excellent

B. Generating Classification Rules for Student Data by Applying Decision Tree Algorithm

A Decision Tree is a classification scheme which generates a tree and a set of rules.

Given entropy as a measure of the impurity in a collection of training examples, we can now define a measure of the effectiveness of an attribute in classifying the training data. The measure we will use, called *information gain*,[7] is simply the expected reduction in entropy caused by partitioning the examples according to this attribute.

The ID3 algorithm [14] can be summarized as follows:

1. Take all unused attributes and count their entropy concerning test samples
2. Choose attribute for which entropy is maximum
3. Make node containing that attribute

C. *Generating Classification Rules For Student Data by Applying Multilayer Perceptron classification method*

A multilayer perceptron shown in 'Fig. 3' is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate output. It is a modification of the standard linear perceptron in that it uses three or more layers of neurons (nodes) with nonlinear activation functions, and is more powerful than the perceptron in that it can distinguish data that is not linearly separable, or separable by a hyperplane.

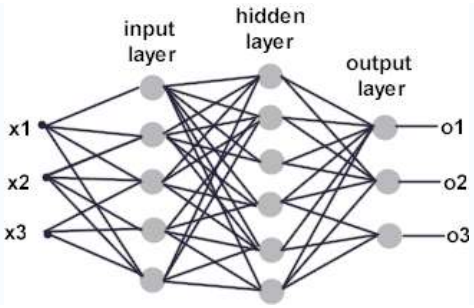


Fig 3. Multilayer Perceptron [18]

Although the backpropagation algorithm can be used very generally to train neuralnetworks, [15] it is most famous for applications to layered feedforward networks, or multilayer perceptrons.

The algorithm

Let's diagram the network as----

$$\mathbf{x^0} \xrightarrow{W^1,b^1} \mathbf{x^1} \xrightarrow{W^2,b^2} \cdots\cdots\cdots \xrightarrow{W^L,b^L} \mathbf{x^L} \qquad (1)$$

where $x^l \in R^{nl}$ for all l = 0, . . . , L and W^l is an $n_l \times n_{l-1}$ matrix for all l = 1, . . . , L.There are L+1 layers of neurons, and L layers of synaptic weights. We'd like to change the weights W and biases b so that the actual output x^L becomes closer to the desired output d.

The backpropogation algorithm consists of the following steps.

1. Forward pass. The input vector x^0 is transformed into the output vector x^L , by evaluating the equation

$$x_i^l = f(u_i^l) = f\sum_{j=1}^{nl-1} W_{ij}^l x_j^{l-1} + b_i^l$$

for l = 1 to L.

2. Error computation. The difference between the desired output d and actual out-
put x^L is computed.

$$\delta_i^L = f'(u_i^L)(d_{i\_} x_i^L) \qquad (2)$$

3. Backward pass. The error signal at the output units is propagated backwards through the entire network, by evaluating

$$\delta_j^{l-1} = f'(u_j^{l-1}) \sum_{i=1}^{n_l} \delta_i^l W_{ij}^l \qquad (3)$$

from l = L to 1.

4. Learning updates. The synaptic weights and biases are updated using the results of the forward and backward passes,

$$\Delta W_{ij}^l = \eta \delta_i^l x_j^{l-1} \qquad (4)$$

$$\Delta b_j^l = \eta \delta_i^l \qquad (5)$$

These are evaluated for l = 1 to L. The order of evaluation doesn't matter.

D. Generating Classification Rules For Student Data by Applying Decision Table classification method

Decision Table Classifier has Flat training set data with most attributes stripped off. It contains only "important" attributes.

The Functional Definition

A decision table [8] has two components:

1. A schema, which is a list of attributes.

2. A body, which is a multiset of labeled instances. Each instance consists of a value for each of the attributes in the schema and a value for the label.

The set of instances with the same values for the schema attributes is called a cell.

Given an unlabeled instance, x, the label assigned to the instance by a decision table classifier is computed as follows. Let I be the set of labeled instances in the cell that exactly matches the given instance x, where only the attributes in the schema are required to match and all other attributes are ignored. If I= Ø, return the majority class in I, breaking ties arbitrarily. Otherwise (I = Ø), the behavior depends on the type of decision table used:

1. A DTMaj returns the majority class in the decision Table.

2. A DTLoc removes attributes from the end of the list in the schema and tries to find matches based on fewer attributes until one or more matches are found and their

majority label is returned. This increases the cell coverage until training instances match x.

E. Generating Classification Rules For Student Data by Applying Naïve Bayesian Network classification method

A Bayesian network, as shown in 'Fig. 4' belief network or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independencies via a directed acyclic graph (DAG).

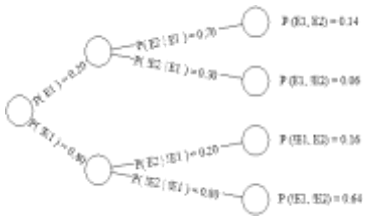


Fig 4. Example of Bayesian Network [17]

A Bayesian network is graph-based model [16] of joint multivariate probability distributions that captures properties of conditional independence between variables. Such models are attractive for their ability to describe complex stochastic processes, and since they provide clear methodologies for learning from (noisy) observations.

In order to measure the performance of a classification model on the test set [12][13], the classification accuracy or error rate are usually used for this purpose. The classification accuracy is computed from the test set where it can also be used to compare the relative performance of different classifiers on the same domain. However, in order to do so, the class labels of the test records must be known. Moreover an evaluation methodology is needed to evaluate the classification model and compute the classification accuracy. Here we use 10-fold-cross-validation method.

IV. EXPERIMENTAL RESULTS

Test Data

Mid_Sem_per	Attendance	E_homework	GPA	Final_Grade
GOOD	GOOD	YES	GOOD	EXCELLENT
AVERAGE	AVERAGE	YES	GOOD	PASS
AVERAGE	POOR	YES	GOOD	FAIL
AVERAGE	POOR	YES	GOOD	FAIL
GOOD	POOR	YES	GOOD	PASS
GOOD	GOOD	YES	GOOD	EXCELLENT
GOOD	GOOD	YES	POOR	PASS
AVERAGE	AVERAGE	NO	POOR	FAIL
GOOD	GOOD	YES	GOOD	EXCELLENT
AVERAGE	AVERAGE	YES	POOR	PASS

Id3

Attendance = GOOD
| GPA = GOOD: EXCELLENT
| GPA = POOR: PASS
Attendance = AVERAGE
| E_homework = YES: PASS
| E_homework = NO: FAIL
Attendance = POOR
| Mid_sem_per = GOOD: PASS
| Mid_sem_per = AVERAGE: FAIL

Results from Decision Trees using Id3

Time taken to build model: 0 seconds

Correctly Classified Instances	7	70	%
Incorrectly Classified Instances	3	30	%

Results from Multilayer perceptron classifier

Time taken to build model: 0.06 seconds

Correctly Classified Instances	6	60	%
Incorrectly Classified Instances	4	40	%

Results from Decision Table classifier

Time taken to build model: 0 seconds

Correctly Classified Instances	3	30	%
Incorrectly Classified Instances	7	70	%

Results from Naïve Bayesian Network classifier

Time taken to build model: 0.02 seconds

Correctly Classified Instances	3	30	%
Incorrectly Classified Instances	7	70	%

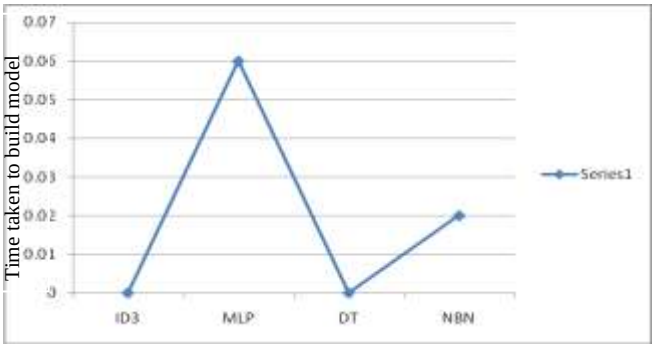


Fig. 5. Time Graph

"Fig. 5," gives the analysis chart which shows time taken by Different classifier.

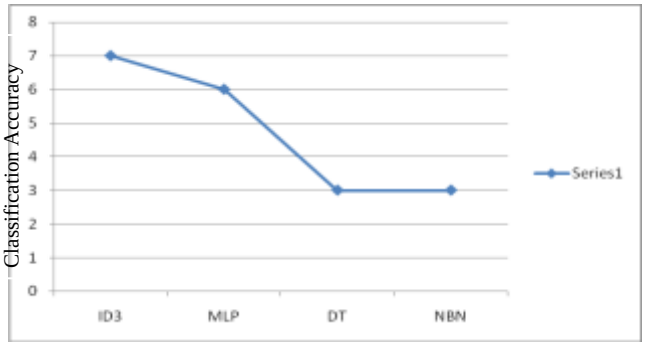


Fig. 6. Accuracy Graph

"Fig. 6," gives the Analysis chart which shows relative classification accuracy of algorithms.

V. CONCLUSION

This work is an attempt to use Data Mining techniques to analyze students' academic data and to enhance the quality of higher educational system.

In this work we applied four classification methods on student data i.e. Decision tree (ID3),Multilayer Perceptron, Decision Table and Naïve Bayes Network Classification method.

We notice that ID3 Classifier is most suitable method for this type of student dataset.

The Higher managements can use such classification model to improve students' performance according to the extracted knowledge.

VI. FUTURE WORK

One of the most attractive future works is to collect large data set from the university database and apply these classification methods on such data.

These classification methods can also be applied to predict other student outcomes such as dropouts or alumni pledges.

REFERENCES

[1] Alaa El-Halees, "Mining Students Data to Analyze Learning Behavior: A Case Study." *Department of Computer Science, Islamic University of Gaza P.O.Box 108 Gaza, Palestine alhalees@iugaza.edu.ps.,2008*

[2] Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques", *Morgan Kaufmann Publishers,2006*.

[3] Arun K. Pujari." Data Mining Techniques", *Universities Press (India) Private Limited,2005*.

[4] M. Stonebraker, R. Agrawal, U. Dayal, E.J. Neuhold andReuter, "DBMS Research at a Crossroads: The Vienna Update", *proc. 19th Very large data bases Conf., 1993*.

[5] Jing Luan "Data mining application in higher education". *Chief planning and Research Officer, Cabrillo College founder knowledge Discovery,2006.*

[6] Ryan S.J.d. Baker, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA "Data Mining for Education",*2008.*

[7] dms.irb.hr/tutorial/tut_d**trees**.php

[8] Ron Kohavi and Daniel Sommerfield "Targeting Business Users with Decision Table Classifiers" *Data Mining and Visualization Silicon Graphics, Inc.,1998.*

[9] Behrouz Minaei-Bidgoli 1, Deborah A. Kashy 2, Gerd Kortemeyer3, William F. Punch 4 "Predicting Student Performance: an Application of Data Mining Methods with an Educational Web-Based System" *IEEE 2003*.

[10] M. Hegland, "Algorithms for Association Rules", *Lecture Notes in Computer Science, Vol. 2600, Jan 2003.*

[11] Mike Chapple, About.com

[12] Tan P. N., Steinbach M & Kumar V. "Introduction to Data Mining" *Pearson Education, 2006*.

[13]Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar "Mining Student Data Using Decision Trees", *The 2006 International Arab Conference on Information Technology.*

[14]en.wikipedia.org/wiki/ID3_algorithm

[15] Sebastian Seung "Multilayer perceptrons and backpropagation learning 9.641" Lecture 4: September 17, 2002

[16] NirFriedman , Michal Linial,Iftach Nachman, Dana Pe'er "Using Bayesian Networks to Analyze Expression Data"

[17] Daryle Niedermayer, "An Introduction to Bayesian Networks and their Contemporary Applications", I.S.P., PMP, B.Sc., B.A., M.Div.December 1, 1998

[18] http://dms.irb.hr/tutorial/images/ann.jpg