

Attributes Selection for Predicting Students' Academic Performance using Education Data Mining and Artificial Neural Network

¹Suchita Borkar

¹Asstt Prof., MCA Department, PCCOE, Pune

²K.Rajeswari

²Assoc. Prof., Department of Computer Engineering, PCCOE, Pune

ABSTRACT

Education Data mining plays an important role in predicting students' performance. It is a very promising discipline which has an imperative impact. In this paper students' performance is evaluated and some attributes are selected which generate rules by means of association rule mining. Artificial neural network checks accuracy of the results. A Multi-Layer Perceptron Neural Network is employed for selection of interesting features using 10 – fold cross validation. The artificial neural network selects 5 out of 8 attributes based on the accuracy obtained for correctly classified data. It is observed that in association rule mining important rules are generated using these selected attributes. The Experiment is conducted using Weka and real time data set available in the college premises.

Keywords

Educational Data Mining; Apriori algorithm, Association Rule Mining; Neural network, Multi-Layer Perceptron.

1. INTRODUCTION

Data Mining is the process of Knowledge Discovery in Databases. It helps to discover hidden patterns in a large dataset. Educational Data Mining (EDM) is a rising discipline, which helps to develop methods which will explore unique types of data from education database and assist to predict students' academic performance. EDM can be considered as learning science, as well as a feature of data mining [11]. Students' learning process is a very complex issue to assess. Data Mining plays a very crucial role in predicting students' academic performance which will help in proposing improvements..

2. PREVIOUS STUDIES

EDM has been applied in various studies for exploring hidden pattern to improve students' academic performance.

Ali and Kerem studied the dataset of students of Istanbul Eyup I.M.K.B.Vocational Commerce High School and found the relationship between the student performance and course. In their finding they have generated a rule that shows if a candidate is unsuccessful in numerical course in 9th class then those students are likely to be unsuccessful in 10th class. Such results were generated for different courses. This study can facilitate students to choose their appropriate profession by revealing the relation between their concern fields. [1]

Tiwari et al., conducted a study on engineering students to evaluate their performance by applying data mining techniques to assist them in decision making. They used K-Means algorithm to cluster students. The result predicted that if students are poor in attendance and assignment then there is 75% probability that their grades are poor. [2]

Sen and Ucar compared the achievements of Computer Engineering Department students in Karabük University by means of various factors such as age, gender, type of high school graduation and the students studying in distance education or regular education through data mining techniques. They have taken the dataset of 3047 records. In their study they have used NN architecture called multilayer perceptron (MLP) with back propagation type supervised-learning algorithm to produce both classification and regression type prediction models and decision tree for achieving the highest possible prediction accuracy. The results revealed that as the age of the student increases the success score decreases and students success rate is much better in distance than in formal education, students coming from vocational high school are more successful in cultural lessons than those taking vocational lesson. [3]

Baradwaj and Pal have discussed methods to achieve high quality in higher education. They have made use of various data mining algorithms like classification algorithm to estimate the accuracy of data. Clustering algorithm was used to cluster the objects which are used as preprocessing approach for attributes. Association rules were used to find the correlation between frequent item set with confidence value less than one. Neural Network was used to derive patterns from complicated or imprecise data. Through this study they tried to identify weak students needing special attention. [4]

Ramaswami and Bhaskaran developed a predictive data mining model to identify academically weak students and attributes that affect their performance using CHAID prediction model. The attributes were selected on the basis of chi-square values. If chi-square values of attributes are greater than 100 they are given due considerations and consider the highly influencing variables with high chi-square values. [7]

In our research we have studied the dataset of 60 MCA students to predict their university result. In our work we have proposed that some selected attribute are more influencing for student's academic performance and generate association rules.

3. DATA COLLECTION AND PREPROCESSING

In this study, we have considered students who are pursuing Master of Computer Application (MCA) degree from Pune University. Neural network technique is used for selecting the attributes from a set of attributes and based on the accuracy of correctly classified data important attributes are identified and rules are generated. The Attributes used are Schooling Education, Previous knowledge of computer, Father/Mother is educated, Graduation percentage, Attendance%, Assignment%, Unit Test%, University result%. On the basis

of the data collected these attributes will predict student’s performance in the university examination.

Table 1. Attributes and Its Possible Values

In a given dataset Data Pre-Processing technique is used to identify noise data, missing values, irrelevant and redundant data Data for the above mentioned attributes is depicted in percentage.

Attributes	Description	Values
Schooling Education	Medium in which schooling education is done	English/Non English
Previous programming knowledge	Previous Knowledge about programming	Yes/No
Father/Mother is educated	Whether either father or mother is educated or not	Yes/ No
Graduation%	Percentage of marks obtained in graduation.	Good, Avg, Poor
Attendance%	Attendance of the student.	Good, Avg, Poor
Assignment%	Assignment performance given during the semester.	Good, Avg, Poor
Unit Test Performance%	Percentage marks obtained by a student in Unit Test.	Good, Avg, Poor
University Result%	Percentage marks obtained by the student in university examination.	Good, Avg, Poor

Table 2. Categorization of Attributes

Attribute	Range
Graduation%	Graduation% $\geq 70\%$ = Good. 60% \leq Graduation% $< 70\%$ =Avg. Graduation% $> 60\%$ =Poor.
Attendance%	Attendance% $\geq 70\%$ = Good. 60% \leq Attendance% $< 70\%$ Avg. Attendance% 60% =Poor.
Assignment%	Assignment% $\geq 70\%$ = Good. 60% \leq Assignment% $< 70\%$ Avg. Assignment % $> 60\%$ =Poor.
Unit Test%	Unit Test% $\geq 70\%$ = Good. 60% \leq UnitTest% $< 70\%$ =Avg. UnitTest % $> 60\%$ =Poor.
University Result%	University Result% $\geq 70\%$ = Good. 60% \leq UniversityResult% $< 70\%$ =Avg. University Result % $> 60\%$ =Poor.

4. METHODOLOGY

In our research, we have used WEKA tool free software, implemented in java language and which uses ARFF file format as an external representation format. WEKA [10] is open source software that offers a collection of machine learning and data mining algorithms for data pre-processing, classification, regression, clustering, and association rules We have compared the results between the accuracy obtained by Neural Network on all the attributes and accuracy obtained by applying neural network technique on selected attributes (Fig 1).

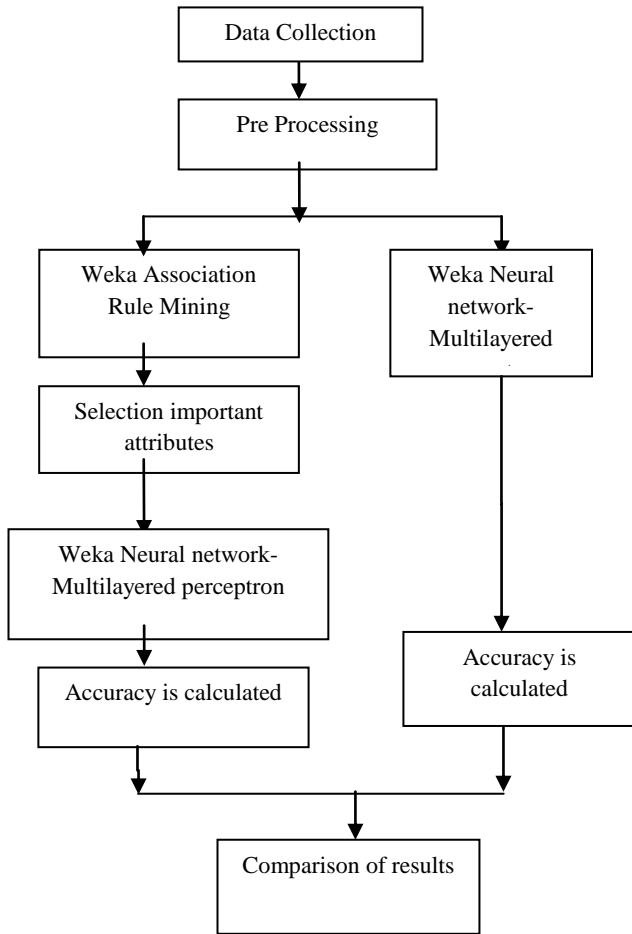


Fig 1. Work Methodology

4.1 Association Rule Mining

In EDM, **association rule learning** is a conventional and well researched method for determining interesting relations between attributes in large databases [11]. Association rule Mining is mainly intended to recognize strong rules from databases using different measures of support and confidence.

The preliminaries necessary for performing data mining on any data are discussed below.

Let $I = \{I_1, I_2, I_3, \dots, I_m\}$ be a set of items. Let D , the task relevant data, be a set of database transactions where each transaction $T \subseteq I$. Each transaction is an association with an identifier, called transaction identification (TID). Let A be a set of items. A transaction T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subset I, B \subset I$, and $A \cup B = \emptyset$. [11]

Support (s) and confidence (c) are two measures of rule interestingness. They respectively reflect the usefulness and certainty of the discovered rule. A support of 2% of the rule $A \Rightarrow B$ means that A and B exist together in 2% of all the transactions under analysis. The rule $A \Rightarrow B$ having confidence of 60% in the transaction set D means that 60% is the percentage of transactions in D containing A that also contains B .

A set of items is referred to as an item set. An item set that contains k items is a k -item set. The occurrence frequency of an item set is the number of transactions that contain the item set. If the relative support of an item set I satisfies a

prescribed minimum support threshold, then I is a frequent item set. The association rule mining can be viewed as a two-step process:

- 1) Find all frequent item sets: Each of these item sets will occur at least as frequently as a predetermined minimum support count.
- 2) Generate strong association rules from the frequent item sets: The rules must satisfy minimum support and confidence. These rules are called strong rules. [11]

4.2 Apriori Algorithm

Apriori is a seminal algorithm proposed by R. Agarwal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties. The following lines state the steps in generating frequent item set in Apriori algorithm. [3]

Let C_k be a candidate item set of size k and L_k as a frequent item set of size k . The main steps of iteration are:

- Find frequent set L_{k-1}
- Join step: C_k is generated by joining L_{k-1} with itself (Cartesian product $L_{k-1} \times L_{k-1}$)
- Prune step (apriori property): Any $(k - 1)$ size item set that is not frequent cannot be a subset of a frequent k size item set, hence should be removed
- Frequent set L_k has been achieved [11].

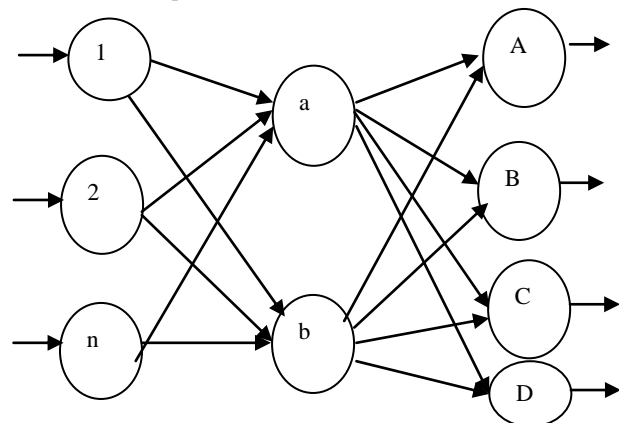


Fig 2. A Feed Forward Neural Network

There are two phases

1. Learning phase where the network learns by modification of weights.
2. Testing phase where an unknown input is tested for proper learning of neural network.

In Fig.2, the nodes 1, 2, 3... n represents the input nodes of input layer. Nodes a,b represents nodes of hidden layer. Nodes A, B, C, D represents nodes of output layer.

The Back-propagation algorithm [13] is as follows.

Input

- The samples used for training
- The rate of learning - l
- A feed – forward multi layered fully connected network.

Method

Network is initialized with random values of weights and biases.

Repeat till termination conditions are met

```
{
for each sample S in the set of training samples
{
// Propagate the inputs forward
for each hidden or output layer unit j
{
• Sum of products of weight and output of a particular node
with the bias assumed is the input value for the next layer
• For each unit j, compute the output using exponential
activation function.
    • Back propagate the errors with the following steps
    • for each unit j in the output layer
    • Error j = Out j ( 1- Outj ) ( Tru j – Out j ) ;
    • for each unit j in the hidden layers , from the last to
    the first hidden layers
    • Error j = Out j ( 1- Outj ) Σk Error k w jk;

//Compute the error wrt the next higher layer , k
    • for each weight wt ij in network
    • Increment the weight, bias values
    • wtij = wtij + Δ wt ij ;
    • for each bias θ j in network
    • Δ θb j = (lr)Err j ;
    • θb j = θ j + Δ θb j ;
}}
```

5. RESULTS AND DISCUSSION

The dataset of 60 students from MCA course was obtained from M.C.A department of Pimpri Chinchwad College of Engineering, Pune University. In this paper we have selected interesting attributes using multilayered perceptron from large set of attributes and various association rules between attributes have found to check how these selected attributes affect the student’s university result. Number of association rule can be found for different confidence values.

5.1 Attribute Selection

For the dataset of 60 students 8 attributes were considered. Multilayered Perceptron was applied on the given data and checked the accuracy of correctly classified data for the considered attributes. In the dataset 10 cross fold validation was used and every single attributes was removed subsequently and check the accuracy of the data. The attributes taken were;

1. Schooling Education.
2. Previous programming knowledge.
3. Father/Mother is educated.
4. Graduation%.
5. Attendance%.
6. Assignment%.
7. UnitTest%.
8. Universityresult%.

Multilayered Perceptron was applied on dataset and the accuracy was checked.

Table 3. Sample Pruning

	~F1	~F2	~F3	~F4	-	~F8
Schooling Education	X					
Previous programming knowledge		X				
Father/Mother is educated			X			
Graduation%				X		
Attendance%						
Assignment%						
UnitTest%						
Universityresult%.						

The Table above shows the accuracy is obtained by removing the attributes one by one. The description of features is given in the rightmost column. The results obtained by applying 10 cross fold validation on a dataset are if all attributes are considered then the accuracy of correctly classified data is 44.5% while if we remove the first three attributes i.e if we considered attributes as graduation%,Attendance%,assignment%,UnitTest%,Universit yresult% the accuracy of correctly classified data is 46%.It shows the accuracy increases after removing some of the attributes. We can check that if by applying association rule mining algorithm on these selected attributes some interested rules is generated.

5.2 Association Rule Mining

The analysis for generated association rules is as follows:

Rules generated for 90% confidence and 0.1 supports are:

IF Attendance% = Good And Assignment%=Poor THEN UnitTest%=Poor
IF Attendance%=Good And Assignment%=Poor THEN UniversityResult%=Poor

Rules for confidence 87% confidence and 0.1 supports are:

IF Assignment% = Poor THEN UniteTest%=Poor And UniversityResult%=Poor
IF Attendance%=Good And Assignment%=Poor THEN UnitTest%=Poor,UniversityResult%=Poor

Rules for confidence 70% confidence and 0.1 supports are:

IF Attendance%=good and Assignment%=Poor and UnitTest%=Poor THEN UniversityResult%=poor
IF Graduation%=Poor THEN UnitTest=Poor

The association rules for different confidence values can be interpreted in a way that the students' performance will be poor in unit test if either their attendance is poor or assignment is poor or both. Also their university performance will be affected by the poor performance in unit test. So we can interpret that to get the good university performance students have to be good in their assignment, attendance and Unit Test. Also Unit Test performance will also have an impact on the student's graduation performance.

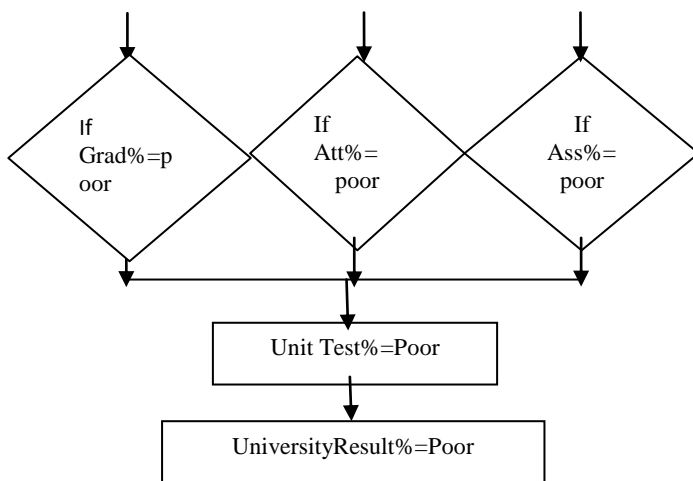


Fig 3. Interpretation of Association Rule Mining

Above figure shows the interpretation of rules generated from our data using apriori algorithm. The result shows that if a student scores poorly in graduation and performs poorly in attendance and assignment then there are chances that he/she will perform low in unit test. This will result in poor performance in University result. So to improve the student's performance in university result students should perform good in graduation, attendance, assignment and unit test. Therefore we can say that the selected attributes which give correctly classified data as 46% are giving good association rule after applying apriori algorithm.

6. CONCLUSIONS

The paper presented the potential use of EDM using association rule mining algorithm and artificial neural network in predicting students' performances in university result. The analysis revealed that some interesting attributes are observed from backpropagation algorithm on 10 cross fold validation and by applying association rule mining algorithm some dependent attributes like Unit test, Assignment, Attendance and graduation percentage are found which will help to predict the student's university performance. The results reveal that the student's performance level can be improved in university result by identifying students who perform poorly in unit Test, Attendance, Assignment and graduation and giving them additional guidance to improve the university result. Additionally, by applying multilayered perceptron with 10 cross validation on a set of attributes we get 46% correctly classified data on some selected attributes. Though the percentage of correctly classified data is low we can have some further study on the same. In our future work

this will be further analysed. Also our next paper will include fuzzy logic in categorized values as good, average and poor for better results.

7. ACKNOWLEDGEMENT

The authors wish to acknowledge MCA department of Pimpri Chinchwad College of Engineering for their support in providing the necessary data.

8. REFERENCES

- [1] Ali Buldua, Kerem Üçgün., Data mining application on students' data. *Procedia Social and Behavioral Sciences* 2 5251–5259, 2010.
- [2] Singh, Randhir. *An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education*, 2013.
- [3] Baha Sen, Emine Ucar. Evaluating the achievements of computer engineering department of distance education students with data mining methods. *Procedia Technology* 1 262 – 267, 2012.
- [4] Baradwaj, Brijesh Kumar, and Saurabh Pal. Mining Educational Data to Analyze Students' Performance. *Arxiv preprint arxiv: 1201.3417*, 2012.
- [5] Castro, Félix, et al. Applying data mining techniques to e-learning problems. *Evolution of teaching and learning paradigms in intelligent environment. Springer Berlin Heidelberg*, 183–221, 2007.
- [6] Huebner, Richard A. "A survey of educational."
- [7] Ramaswami, M., and R. Bhaskaran. A CHAID based performance prediction model in educational data mining. *Arxiv preprint arxiv: 1002.1144*, 2010.
- [8] Kumar, Varun, and Anupama Chadha. Mining Association Rules in Student's Assessment Data. *International Journal of Computer Science Issues* 9.5: 211–216, 2012.
- [9] Cristo'bal Romero, Sebastia'n Ventura, Enrique Garcí'a, 2007. Data mining in course management systems: Moodle casestudy and tutorial. *Received 5 March 2007; received in revised form 19 May 2007; accepted 25 May 2007*.
- [10] <http://en.wikipedia.org/wiki/Weka>
- [11] Anwar, M. A., and Naseer Ahmed. Knowledge Mining in Supervised and Unsupervised sssessment Data of Students' Performance." *2011 2nd International Conference on Networking and Information Technology IPCSIT vol. Vol. 17*. 2011.
- [12] <http://statistics.about.com/od/Formulas/ss/Correlation-Coefficient.htm>
- [13] Jiawei Han and Micheline Kamber, "datamining Concepts and Techniques", Elsevier Second Edition
- [14] RAJESWARI, K., and V.VAITHIYANATHAN. "ATTRIBUTE SELECTION USING ARTIFICIAL NEURAL NETWORKS–A CASE STUDY OF ISCHEMIC HEART." *Journal of Theoretical and Applied Information Technology* 46.1 (2012).
- [15] Borkar, Suchita, and K. Rajeswari. "Predicting Students Academic Performance Using Education Data Mining." (2013).