

Review on Application of Data Mining Techniques for Student Performance Analysis

¹C. Sudhakar Reddy, ²Dr K Venugopala Rao

¹Associate Professor, ²Professor

¹Dept of Information Technology, ²Department of Computer Science
G. Narayanamma Institute of Technology & science, Hyderabad, India.

Abstract: For an educational organization the primary objective is continuous improvement in the quality of education. In the current scenario of Digitization lot of information is stored about the programs offered, information about various stakeholders of the system which includes employees of the organization, students enrolled for various programs along with their progress in the respective Areas/Activities. Improvement of quality of education can be done by identifying the factors contributing to student's performance, identifying strength and weakness of students, suggestions of suitable courses for the students, improvement in the Design of the course structure etc. Hence it is essential to have a systematic approach to analyze the huge voluminous data and to provide an appropriate data structure for better decision making by the corresponding Teams responsible for the same.

Index Terms – Classification, Data Mining,

I. Introduction

Most of the educational institutes' stores data about various stake holders of the system. Primary stake holders of the system are students. Data about students includes academic performance, participation in co curricular activities, demographic information etc. Many tools are available for analyzing student's data which includes data mining, machine learning Techniques etc. In this paper we present an overview of various data mining functionalities used by different analysts for analyzing data. We have organized this paper as follows.

Section II: Introduction of the concept of data mining and various data mining functionalities. Section III: overview of various data mining functions used by analyst for student performance analysis applications in the context of educational data mining.

Section IV: Explanation about the application of data mining Techniques for classification of student based on their grade.

Section V: Conclusion and Future scope.

II. Data Mining

Data Mining is the process of discovering knowledge from the data. Data Mining is a step by step procedure used for discovering and extracting knowledge from the data available as data present in databases may exhibit the following characteristics

1. Noisy Data: Values stored are not valid or may not be filled
2. Continuous Values: Each Sample may be representing different values
3. Inconsistency: Different users would have stored data in different formats (eg: Representation of Date in different formats)
4. For the process of knowledge discovery a subset of the huge data is sufficient, so it is essential to identify the data necessary for the analysis.

Steps in Data Mining

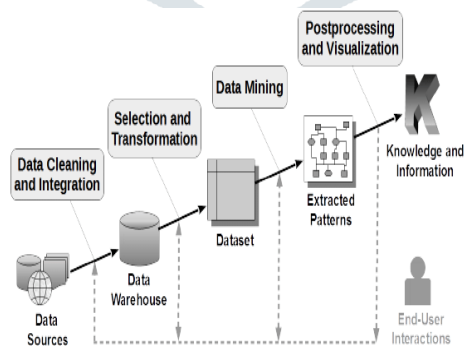


Figure 1

Data Cleaning – This process fills missing values and invalid values are replaced with appropriate values

Data Integration – This process involves collection of data from multiple data sources

Data Selection - This process is responsible for identifying the data necessary for performing data mining task

Data Transformation –This process involves on version of the data to common format and if the range of values are very high they are converted to suitable range using normalization

Data Mining – In this process an appropriate data mining functionality is applied to extract useful patterns

Pattern Evaluation – In this process extracted patterns are evaluated against given set of chosen thresholds and patterns which satisfies the given thresholds are considered as interesting patterns. But specifying thresholds may be tricky for better training we may provide some domain knowledge and patterns extracted earlier can be given as input for this module for better results.

Knowledge Presentation – An appropriate representation of the patterns must be chosen to present the patterns.

Data mining functionalities

Classification:

It is the process of constructing a model that distinguishes classes. Membership of the data in a given class can be obtained by using model. Decision Tree, Naive bays, Neural networks are widely used for classification

In the context of educational data mining we may find the grade of the student, using classification, generating a model for finding whether student is going to get placement or not or to find what factors are contributing to student's success or failure.

Regression Analysis

Regression analysis is a statistical methodology used for numeric prediction. It is used for representing relationship between multiple variables.

Linear regression is used for representing relationship as a straight line equation that best fits two variables so that one variable can be computed using the other variable.

Multiple linear regression is used for representing relationship between more than two variables. In case of student performance analysis, we may use regression for predicting future grade using grades scored in previous years

Clustering

It is a process of organizing objects into clusters such that objects in the same cluster exhibit high degree of similarity similar and objects from different clusters have high degree of dissimilarity. For clustering we must specify no of clusters and threshold for similarity metrics. Characteristics are derived for each cluster.

Clustering is used for building successful and unsuccessful student profiles.

Association Analysis

It is useful for discovering interesting relationships hidden in large data sets, Relationships are expressed using association rules.

For example

{bread} -> {butter}

Indicates sales of butter is strongly related to sales of bread. The strength of the rule is indicated by using confidence and support. After extracting all the rules whose support and confidence are above the specified threshold are considered as strong association rules. Apriori algorithm, FP growth Algorithm are widely used for mining association rules.

Association analysis can be used for finding most adaptive students for a given course

III. Background

In the paper “**Student Performance Evaluation Using Data Mining Techniques for Engineering Education**” [1] authors have used WEKA as tool for Students performance in the course “Network Theory “is analyzed using the hybrid of fuzzy inference system and neural network. When compared with classical approach hybrid approach has provided the better results

In the paper “**predicting student performance using data mining techniques**” [3] authors have used SVM, Naive Bays, and Decision tree algorithms by considering various attributes about the student. In this paper authors have identified student's performance as function of not only academic performance but also behavioral and demographic attributes.

In the paper “**An Efficient use of Ensemble Methods to Predict Students Academic Performance**” [5] authors have used ensemble methods (bagging, boosting, and voting) for improving the accuracy.

In the paper “**Student's Performance in the Context of Data Mining**” [16] authors have used classification and clustering algorithms for analyzing student's performance.

In the paper “**Using Data Mining Techniques to Predict Students at Risk of Poor Performance**” [14] authors have used CHAID, C & R Tree for the assessment of risk level for achieving good honors degree.

For the improving the quality of education various authors has used data about student's academic performance and behavior information and demographic information to perform the following

Prediction of student's grade.

Prediction of student's placement.

Impact of various factors (academic performance, demographic information) on student's performance.

Identifying characteristics of slow learners and fast learners.

Identification of relationship between different courses so that appropriate courses can be suggested for students.

Identifying factors contributing to student's dropout.

Finding the impact of assignments on final student's performance.

IV. Related work

We have collected the data about the student’s academic performance along with demographic attributes and applied bagging, boosting techniques for generation of models to predict student’s grade.

Bagging

Bagging is a technique that is carried out in multiple iterations. In each iteration a training set D_i is sampled with replacement from the given of data D . A classifier model M_i is learned for each D_i to classify an unknown tuple X each classifier returns a class prediction, which is counted as one vote. The bagged classifier M^* counts the votes and assigns the class with the most votes to X .

Boosting

In boosting weights are assigned to each training tuple. A series of k -classifiers is iteratively learned. After a classifier M_i is learned, weights are updated to allow the subsequent classifier M_{i+1} . The finally classifier M^* , combines the votes of each individual classifier.

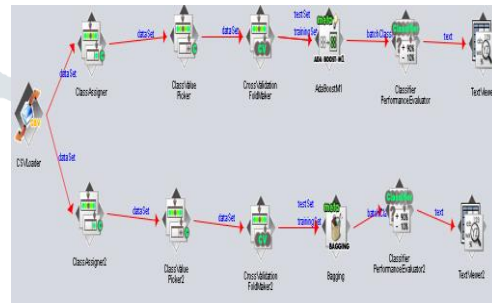


Figure2

Using WEKA knowledge flow environment as shown in the figure 2 models are generated for classification using bagging and ada boosting and the performance of these methods are indicated in the table 1

Table 1

Algorithm	TP Rate	FP Rate	Accuracy
Bagging	0.832	0.585	83.20
Ada Boosting	8.17	8.17	81.67

V. Conclusion and Future scope

Data Mining can be used as a tool for systematic analysis of data stored by educational institutions across various dimensions which includes assessment of student’s performance, identifying factors contributing to student’s success or failure, finding students adaptability for a given course etc. Most widely used data mining functionalities is classification, prediction, clustering. To improve the accuracy authors have proposed ensemble methods like bagging, boosting. With the help of agent based mining performance can be improved with the support of distributed data mining. For feature extension various feature selection algorithms can be used to identify the most relevant futures for analyzing student performance. By applying feature selection methods identify the attributes most relevant for data analysis then apply data mining algorithms.

References

[1] Veena Deshmukh, Srinivas Mangalwede, Dandina Hulikunta Rao, “Student Performance Evaluation Using Data Mining Techniques for Engineering Education”, *Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 6, 259-264 (2018)*.

[2] A.Dinesh Kumar , R.Pandi Selvam, K.Sathesh Kumar , “Review on Prediction Algorithms in Educational Data Mining” *International Journal of Pure and Applied Mathematics ,Volume 118 No. 8 2018, 531-537*.

[3] Astha Soni, Vivek Kumar, Rajwant Kaur, D. Hemavathi, “predicting student performance using data mining techniques “*International Journal of Pure and Applied Mathematics Volume 119 No. 12 2018, 221-227*.

[4] Mashael Al luhaybi, Allan Tucker and Leila Yousefi, “The Prediction of student failure using classification methods: a casestudy” pp. 79–90, 2018. CS & IT-CSCP 2018.

[5] Pooja Kumari , Praphula Kumar Jain , Rajendra Pamula “An Efficient use of Ensemble Methods to Predict Students Academic Performance”, 978-1-5386-3039-6/18 2018 IEEE.

- [6] Khokhoni Innocentia Mpho Ramaphosa, Tranos Zuva, Raoul Kwuimi, Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools, 978-1-5386-3060-0/18 2018, IEEE.
- [7] CH.M.H.Sai Baba, AkhilaGovindu , Mani Krishna Sai Raavi , VenkataPraneethSomisetty “student performance analysis using classification techniques”, International Journal of Pure and Applied Mathematics ,Volume 115 No. 6 2017, 1-7.
- [8] Bhavesh Patel, Chetan Gondaliya “Student Performance Analysis Using Data Mining Technique” IJCSMC, Vol. 6, Issue. 5, May 2017, pg.64 – 71.
- [9] N.V. Krishna Rao, Dr. N Mangathayaru , Dr. M. Sreenivasa Rao “Evolution and Prediction of Radical Multi-Dimensional E-Learning System with Cluster based Data Mining Techniques” 978-1-5090-4257-9/17 2017 IEEE 701-707.
- [10] MukeshKumar, Prof.A.J. Singh “Evaluation of Data Mining Techniques for Predicting Student’s Performance”, I.J. Modern Education and Computer Science, 2017, 8, 25-31.
- [11] Md. Hasibur Rahman, Md. Rabiul Islam “Predict Student’s Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques”. 2nd International Conference on Electrical & Electronic Engineering (ICEEE), 27-28 December 2017.
- [12] Anoushka,Jain,Tanupriya,Choudary, Praveen Mor,A.Sai Babitha . “Intellectual Student Performance Analysis by comparing various data mining techniques”, 978-1-5386-1144-9 2017 IEEE.
- [13] Maryam Zaffar , Manzoor Ahmed Hashmani , K.S.SAVITA ,”Performance Analysis of Feature Selection Algorithm for Educational Data Mining”, 978-1-5386-0790-9/17 2017 IEEE.
- [14] Zahyah Alharbi, James Cornford, Liam Dolder and Beatriz De La Iglesia” Using Data Mining Techniques to Predict Students at Risk of Poor Performance” ,978-1-4673-8460-5/16 2016, IEEE.
- [15] MAO Keji,YU Mingyuan,CHEN Qingzhang, “Discovering the most adaptive students of one course by data mining”, 978-0-7685-3817-4/09 ,2009, IEEE, 242-245
- [16] Kamran Shaukat, Iqra Nawaz, Samreena Aslam,Sana Zaheer,Usman Shaukat , “Student’s Performance in the Context of Data Mining” 978-1-5090-4300-2/16 2016 IEEE

