PREDICTIVE ANALYTICS OF EDUCATION
DATA BASED LEARNING PATTERNS–A LITERATURE REVIEW

Mamta Saxena, Sachin Gupta
E-Mail Id: 17ca9001w@mvn.edu.in, sachin.gupta@mvn.edu.in
MVN University, Haryana, India

Abstract- In the recent years, the scenario of education has changed a lot. It has moved from one platform to another by using various new techniques and technologies. In this sector, the new technology that is growing very fast is Educational Data Mining. Though the term “Data Mining” is very popular but when it get added with the education sector it is named as mining the educational data on various prospects to find the useful patterns of data that can serve Educators, Stakeholders, Academicians and other personnel to find out the learning patterns of students and to retain their intensity of failure. In this paper we are discussing about classification algorithm and feature selection method that can be used to find the most relevant attribute for predicting the performance of students.

Keywords: Educational Data Mining, Classification, Feature selection.

1. INTRODUCTION

Data Mining is a technique that can be used to extract the appropriate data from a large volume of data. Data is available in a voluminous amount but we need to extract the data to find the useful patterns that can be used in decision making. When the techniques and methods of data mining get associated with education sector to help an academician, Stakeholders and Educators to retain the intensity of students from failure then it is termed as Educational Data Mining.

There are various techniques that are available in Data Mining like Association, Clustering, Classification, Prediction, Neural Networks, Pattern Recognition, and Decision Trees. From the above techniques we choose classification to find the useful patterns from students’ data. The data of students can be collected either from educational institutes as the institutes maintained their ERP system to manage student’s data or data can be collected from learning management systems that have vast amount of data stored with them.

Data mining technique is used to mine various type of data like Spatial Data, Transactional data, Data Warehouse data, Sequential Data and Multimedia data to find out useful patterns by finding out frequent patterns from data. Various number of steps are included while mining the data. It includes cleansing the data, integration of data, selection of data, by applying operations like aggregation, association, and mining of data, evaluation of pattern and in last representation of knowledge.

EDM helps academicians to predict the performance of students on the basis of various factors that include socio-economic factors, their academic performance in previous semester and end semester, psychological factors, personal factors and other environment factors that are responsible for it. These predictions can be helpful for learners to retain the failure intensity of students.

Dealing with large volume of data that is unstructured and in different file format then extracting useful patterns from it, and fitted it with small amount of database then finding relevant patterns to retain failure intensity of students are the challenges faced by data miners.

2. RELATED WORK

Though Educational data mining is a new field to research but still various works have done so far in this area.[1] used classification to predict the division of students and ID3 decision tree algorithm to predict the performance of students for end semester exams on the basis of various academic factors like attendance, tests seminars, projects. They also discussed how these factors can help educators to identify students who need special guidance and to reduce the failure intensity of students.

[2] Described various socio-psychological factors that affect the performance of students. They used multidisciplinary approach to predict student’s behavior that inhibits the performance of students in higher education and used a qualitative model to describe it.[3] have used Bayesian Classification method to predict the grade and retention intensity of students on the basis of data that is stored previously in student’s database.[4] have used Bayes Classification Prediction model to distinguish between slow learners and high learners for this they have collected data of 300 students to categorize them and filling and transforming the missing values on the basis of prediction model.

[5] have used Decision tree algorithm to predict a new model by using previous data model of students to retain intensity of students and help teachers to supervise those weak students who need special attention and guide them for better future performance.[6] have used C4.5, CART decision tree algorithm to do comparative study of students and help them to identify weak areas and to improve it for better results in the final exams.

[7] have taken the statistics of private colleges and discussed that other than educational parameters there are various other factors that affect the performance of students like mother’s education, mother’s age, parent’s level of income.[8] have used fuzzy association rule mining to predict the end semester performance of students by taking

DOI Number: https://doi.org/10.30780/specialissue-ICAASET021/008
Paper Id: IJTRS-ICAASET2021-008

pg. 42
attributes like attendance, mid semester marks, their academic records were collected from previous database to identify special attention needed students and taking necessary action to improve their performance in next semester exams.

[9] have used Apriori Algorithm and Random Forest Classifier on the data collected from 3 different colleges of Assam to predict the final semester results on the basis of highly influential factors that affect the academic performance of students.[10] have compared various classification algorithm like NaiveBayes, Bayesian Network, ID3, J48 and Neural Network to predict the performance of students and compared result to find which algorithm performs better.

3. DATA MINING AND CLASSIFICATION ALGORITHM

Data mining is a technique used to find the useful patterns from a large volume of data. It is used in the education sector to find out the relevant patterns by using attributes that are selected by using Feature selection method. Classification algorithm is used in educational data mining to classify the data and to categorize them on the basis of certain items. Classification algorithm uses a classifier to predict them. It uses training data sets to work. The test data is used to estimate the accuracy of classification rules. Classification algorithm includes Bayesian Network, Naïve Bayes, J48, Random Forest, Multilayer perceptron and logical Regression.

Classification algorithm tells about the academic performance of students and help educators and learners to set the educational settings on the basis of results they get from classification algorithm. [9] In today’s world the academic performance of students is determined by two assessments i.e. by internal assessment and end semester exams. The internal assessment helps the teacher to predict the students on the basis of various attributes that includes class test, attendance, projects and seminars attended by the students, their lab works while end semester only includes minimum marks needed by the students to get pass.

**Table-3.1 Some Attributes of Dataset**

| Attributes | Description |
|------------|-------------|
| G          | Gender of student whether Male or Female; |
| CAT        | Category of student whether General, SC, ST, OBC; |
| TNP        | Class X\textsuperscript{th} Percentage; |
| TWP        | Class XII\textsuperscript{th} Percentage |
| IP         | Internal Assessment Percentage; |
| EP         | End Semester Percentage; |
| ATLEN      | Class attendance |
| BACKP      | Back in any Paper; |
| MS         | Marital Status; |
| FAS        | Family Size; |
| OF         | Occupation of Father; |
| OM         | Occupation Of Mother; |

Student’s Academic Data get collected on the basis of these attributes and then preprocessing of data get started. It includes cleaning of data, selecting relevant data then applying next step of framework.

3.1 Applying Feature Selection Method

Feature selection is defined as the process of reducing the number of inputs for performing processing operation and finding out the meaningful inputs from the data provided. Feature selection method is used to find out most appropriate attribute that best describe the data.

3.2 Classification Analysis

Classification Analysis is analysis of data that groups the data provided in the form of inputs into various categories for accurate and correct analysis of data and to make decision or to predict about the behavior of inputs supplied as training data sets and by using useful patterns it converted it into test data to make decision and to predict about the behavior with the help of an algorithm.

3.3 Evaluation and Performance Analysis

After applying Classification algorithm next step of student’s academic framework is to evaluate the student’s data and analyze the performance of students to check for the retention intensity of students. After that certain measures are taken by the stakeholders and educators to enhance the learning skills of students and to keep a track of weak
students to improve their retention intensity on the basis of it these people find out the next learning needs of students.

CONCLUSION AND FUTURE WORK

While applying Classification algorithm for predicting the academic performance of students it results in Confusion Matrix and for this various parameters like sensitivity, accuracy, precision and mean absolute error are used. Decision Tree Models are used to classify students in various groups of Good, Average and Bad but it cannot categorize them on the grounds of their learning skills of programming so it is better to use Clustering algorithm for that. Association Rule Mining which is another method also used to find out the interesting patterns from student’s data and it doesn’t used predictive models to predict about students grades from training datasets collected.[7] ID3 decision algorithm can be used to build the decision tree by following top-down approach to minimize the frequency of questions asked while learning to avoid learner’s distraction. If the algorithm apply greedy approach to optimize the performance of algorithm the performance can be improved.[10] The following research problems will be solved by applying Classification techniques to analyze the learning pattern of students and then applying Feature selection method to find the most appropriate attributes. Various Algorithms can be used for analyzing Programming skills of students, learning patterns, analyzing academic performance. We Implement the existing algorithms and proposed algorithms and then get the results. After that analyze the results of proposed algorithms and check their performance with existing algorithms.

REFERENCES

[1] Bharadwaj, B.K. and Pal, S., 2011. Mining Educational Data to Analyze Students’ Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.

[2] Ahmed, A.B.E.D. and Elaraby, I.S., 2014. Data Mining: A prediction for Student’s Performance Using Classification Method. World Journal of Computer Application and Technology, 2(2), pp.43-47.

[3] Pandey, U.K. and Pal, S., 2011. Data Mining: A prediction of performer or underperformer using classification. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2), 2011, 686690.

[4] Bharadwaj, B.K. and Pal, S., 2012. Data Mining: A prediction for performance improvement using classification. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011.

[5] Yadav, S.K., Bharadwaj, B. and Pal, S., 2012. Data Mining Applications: A Comparative Study for Predicting Student’s Performance. International Journal of Innovative Technology & Creative Engineering (ISSN: 2045-711), Vol. 1, No.12, December.

[6] Yadav, S.K. and Pal, S., 2012. Data mining: A prediction for performance improvement of engineering students using classification. World of Computer Science and Information Technology Journal (WCSIT). (ISSN: 2221-0741), Vol. 2, No. 2, 51-56, 2012.

[7] Naqvi, Sayyed. (2006). Factors affecting students' performance A Case Of Private Colleges. Bangladesh e-Journal of Sociology. 3.

[8] Kumar Verma, Sushil & Thakur, R.S. (2017). Fuzzy Association Rule Mining based Model to Predict Students’ Performance. International Journal of Electrical and Computer Engineering (IJECE). 7. 2223. 10.11591/ijece.v7i4.pg2223-2231.

[9] Hussain, Sadiq & Abdulaziz Dahan, Neama & Mutaher Ba-Alwib, Fadl & Najoua, Ribata. (2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. Indonesian Journal of Electrical Engineering and Computer Science. 9. 447-459. 10.11591/ijeecs.v9.i2.pg447-459.

[10] Almarabeh, H., Analysis of Students’ Performance by Using Different Data Mining Classifiers. IJ. Modern Education and Computer Science, 2017. 9(8); p. 9-15.

[11] Agrawal, S., S.K. Vishwakarma, and A.K. Sharma, Using Data Mining Classifier for Predicting Student’s Performance in UG level. International Journal of Computer Applications, 2017. 172(8): p. 6.

[12] Tan, P.-N., M. Steinbach, and V. Kumar, CHAPER 5: Rule-based Classifiers, in Introduction to Data Mining, A. Nordman, Editor. 2005, PEARSON EDUCATION: US.

[13] Fayyad, U. “From Data Mining to Knowledge Discovery in Databases.” AI Magazine 17.3(1996):37-54.

[14] Anjewierden , Anjo , B.Koloffel, and C.Hulshof. “Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes.” International Workshop on Applying Data Mining in e-Learning (ADML 2007) (2007).

[15] 15.R. Eberhart, and J. Kennedy. (1995) A New Optimizer Using Particles Swarm Theory, Proc. Sixth International Symposium on Micro Machine and Human Science (Nagoya, Japan), IEEE Service Center, Piscataway, NJ, pp. 39-43.

[16] J. Kennedy, and R Eberhart, (1995), Particle Swarm Optimization, IEEE Conference on Neural Networks, pp. 1942-1948, (Perth, Australia), Piscataway, NJ, IV, 1995.

[17] J. Kennedy and R. Eberhart. Swarm Intelligence. Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2001.