Automated Performance Evaluation System

Sana Akhai
Information Technology Department
Shah and Anchor Kutchhi Engineering College
akhaisanaa@gmail.com

Ruchi Karia
Information Technology Department
Shah and Anchor Kutchhi Engineering College
ruchikaria95.rk@gmail.com

Aniket Mahadik
Information Technology Department
Shah and Anchor Kutchhi Engineering College
aniketmahadik20@gmail.com

Akshat Shah
Information Technology Department
Shah and Anchor Kutchhi Engineering College
shahakshat47@gmail.com

Manya Gidwani
Information Technology Department
Shah and Anchor Kutchhi Engineering College
manya.gidwani@sakec.ac.in

Abstract: Accurately predicting student performance is useful in different contexts in universities. Educational data mining (EDM) is an emerging discipline, concerned with various approaches such as predicting student performance, Analysis and data visualization, providing feedback for supporting instructors, recommendations for students and so on that automatically extracts meaning from large data generated by or related to people's learning activities in an educational setting. For example, identifying exceptional students for scholarships is an essential part of the admissions process in undergraduate and postgraduate institutions, and identifying weak students who are likely to fail is also important for allocating limited tutoring resources. One of the biggest challenges is to improve the quality of the educational processes so as to enhance student's performance. The results of these studies give insight into techniques for accurately predicting student performance.

Keywords: EDM, Data Visualization, Universities, Limited Tutoring Resources, Recommendations.

I. INTRODUCTION

Data mining is a subfield of computer science. It is the computational process of discovering patterns in large data sets involving methods at the combination of artificial intelligence, machine learning, statistics, and database systems. The actual task of data mining is the automatic or semi-automatic analysis of large data to extract previously unknown, interesting patterns like groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies i.e. association rule mining, sequential pattern mining. This usually involves using different database techniques. These patterns can then be seen as a kind of summary of the input data and may be used in further analysis or, for example, in predictive analytics. Educational data has also become a vital resource in this modern generation, contributing much to the society. Educational institutions such as school, colleges are becoming more competitive because of the number of institutions growing rapidly. The institutions are focusing more on improving various aspects and one important factor among them is quality learning. An Educational system, in order to look behind learning of each student, it needs a complete, accurate and deep understanding of the student's characteristics and performance. The data and information gained from the learning system can be used as an indicator for monitoring of the potential student failure in the school. Personalization indicates adaptive pacing, customization of the instruction process in accordance with the learner's background, his/her knowledge, skills, and characteristics in order to successfully put him/her in proper learning processes.

A. Review of Literature

The main purpose of the data mining model is implementing in the various algorithms to explore the patterns of data. These patterns are useful for selecting the proper algorithm which also increases the accuracy of our project[3].All student performance in the examination conducted during the semester via the educational system are analyzed and decision tree techniques are utilized in order to extract knowledge from student performance and predict final grade[3]. Because of its potential to educational institutes, data mining in higher education is a research field and this area of research is gaining popularity [2]. Data Mining can be used in the educational field to enhance our understanding of learning process to notice on identifying, extracting and evaluating data related to the learning process of student [1].

II. PROPOSED SYSTEM

There is a methodology for work which governs a sequence of stages. The methodology starts from the problem definition, then data collection, Attribute selection, Nominal conversion, file conversion using WEKA tool implementation. Comparative analysis of efficient classification algorithm is done to predict student's performance by the creation of student model.

In the designing of student model J48 algorithm provide a maximum accuracy in classifying the instances in an efficient way. The student model is created in Net Beans using java coding. The J48 algorithm identified by comparative analysis of classification algorithm is taken for designing.

The student model is designed to view models such as

1. Login credential

2. Student detail record

3. Student outcome analysis

J48 is a tree based learning approach, based on iterative dichtomiser (ID3) algorithm. It uses divide-and-conquer algorithm to split a root node into a subset of two partitions till leaf node (target node) occur in the tree.
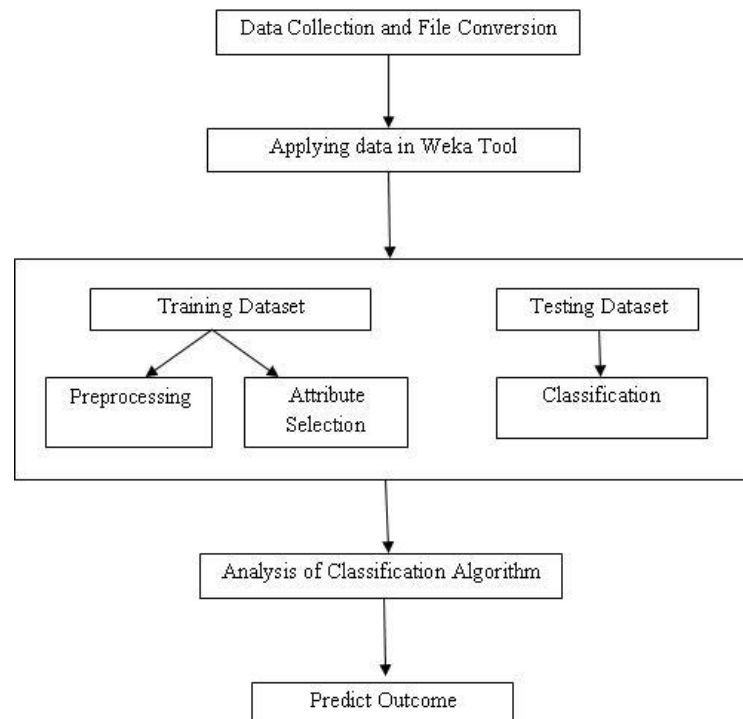

Fig. 1 System Framework

III. ANALYSIS AND EXPERIMENTS

The implementation of this study consists of two types of methods: I.

Analysis:

The process of analysis method is done through different phases, including a literature review.

II. Methods for designing:

Design models and application are carried out based on the result of an analysis of the problems.

Educational data mining (EDM) has a vast amount of data that has to be organized in a consistent manner. To organise, analyse and classify students details J48 algorithm is been used based on academic records. Therefore, to enhance the existing and previous system the proposed model is designed by collecting Students Personal, Academic data and other extra data from the teachers of the institution and Thereby Grouping the student's performance based on certain conditions.

Thereby forming two different classes based on students record

  • Pass Student

  •Fail Student

According to the analysis, we have suspected that G1 and G2 grades would have a high impact. Three input configuration were decided for testing the data mining model:

A: With all variables except G3 [The Output] i.e. Including G1 and G2

B: Including G1

C: Without G1 and G2

Table 1: Comparison of Classifiers

Input Setup	Naive Bayes	J-48	Random Forest
A	71.89%	76.70%	66.32%
B	48.10%	54.43%	45.31%
C	27.84%	31.69%	28.35%

We have done binary Classification where we have set some upper limit to declare the student as pass or fail.

Table 2: Attribute - G1, G2, G3

Class Name	Upper Limit
PASS	20
FAIL	9

After merging and making different data combinations from the dataset we have got the following results:

Algorithm Used	Correctly Classified	Incorrectly Classified
Random Forest	70.63%	29.36%
J48	75%	24%
Random Tree	65%	34%
Decision Stump	75.94%	24%
Multilayer perception	68%	31%
Naive Bayes	70%	29%
REP tree	75%	24%

IV. IMPLEMENTATION

In the implementation of the system first, we take the preprocessed data and use them in the JAVA. To train the code we add the training algorithm, in this case, we use J-48 Classification algorithm. Classification is the process of building a model of classes from a set of records that contain class labels. Decision Tree Algorithm is to find out the way the attributes-vector behaves for a number of instances. This algorithm generates the rules for the prediction of the target variable. With the help of tree classification algorithm, the critical distribution of the data is easily understandable .J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open source Java Implementation of the C4.5 algorithm. The WEKA tool provides a number of options. In this algorithm, it generates the rules from which particular identity of that data is generated. The Improvement we did in J48 Is it uses WEKA as API in java code implemented in NetBeans. WEKA is a comprehensive open source Machine Learning toolkit, written in Java. These functions provide a basic interface to WEKA to allow the transformation of data back and forth and to use most of the features available in WEKA, such as training classifiers. By doing so the accuracy rate of the J48 algorithm has increased to a large extent as compared to the accuracy of the same algorithm previously. The proposed algorithm works as: The csv data file is loaded from WEKA into java code. Then the refining of the dataset is done. Later the J48 classifier is applied. At the end, the results are obtained that is the accuracy and error rate is calculated. Fig. 1 has shown the flow chart of the proposed algorithm. For applying J-48 in the code we use default weka function. We used the types of tree classifiers like Decision tree (DT) and random forest (RF) to predict student grade and extract rules from the prediction results. We classified the attributes to each student as an input data and Student's grades as an output data. To evaluate the prediction performance, each evaluation has been done for each classifier separately, to check which classifier is giving us higher accuracy. We have used 10-fold cross validation.
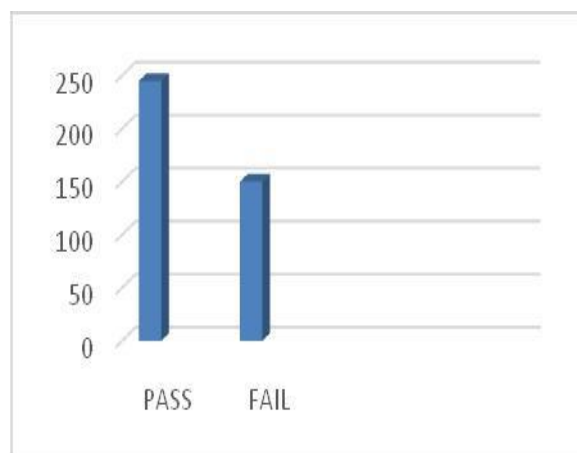
Fig 6: Predicted Outcome

CONCLUSION

The education data mining main focus is to analyze the education system. The model focuses on analyzing the prediction accuracy of the student's performance.The dataset that comprises of all academic and personal factors of the students.

This model can be useful in the educational system like Universities and Colleges. By this model, we can know the academic status of the students in advance and can concentrate on students to improve their academic results and placements. Thereby improve their standards and reputations. As a result, the quality of education can be improved. The results of the data mining algorithms for the classification of the students based on the attributes selected reveals that the prediction rates are not uniform among the algorithms.

REFERENCES

[1] N. T. Nghe, P. Janecek, & P. Haddawy, "A comparative analysis of techniques for predicting academic performance". In Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual (pp. T2G-7).

[2]Y. Kurniawan, & E. Halim, "Use data warehouse and data mining to predict student academic performance in schools: A case study. In Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on, August 2013, pp. 98-103.

[3]Foteini Grivokostopoulou,Isidoros Perikos & Ioannis Hatzilygeroudis, "Using Semantic Web Technologies and Data Mining Techniques to Analyze Students Learning and Predict Final Performance",). In Teaching, Assessment, and Learning for Engineering (TALE), 2013 IEEE International Conference on, December 2014, pp. 488-494.

[4]O.N.Pratiwi,"Predicting student placement class using data mining"In Teaching Assessment and Learning forEngineering(TALE),2013 IEEE International Conference on, August 2013,pp.618-621.

[5]H.Waguih,"Predicting Student Success: An Application of DataMining Techniques in Higher Educational Systems," in Egyptian Computer Science Journal,vol.30,No. 2,2008.

[6]https://archive.ics.uci.edu/ml/datasets/Student+Performance

[7]https://weka.wikispaces.com/Use+WEKA+in+your+Java+code