

PREDICTION OF STUDENT RESULT USING DECISION TREE

¹Anand Choudhary, ²Kajal Ranpise, ³Kiran Mane,
⁴Snehal Lahare, ⁵Dr. Roshani Raut

^{1,2,3,4,5}Dr. D. Y. Patil School of Engineering & Technology, Lohegaon Pune (India)

ABSTRACT

Currently, huge amount of data is stored in educational database. These database contains useful information and can be used for prediction of students result. To construct a predictive model for Students academic result, these needs analyzing data mining methods and classification techniques. Measuring Student result using classification technique such as decision tree. The C4.5 decision tree algorithm is applied on engineering student's data to predict their result in the final exam. This prediction will help to identify the weak students and help them to score better marks. The outcome of the decision tree predicted is the number of students who are likely to pass, fail or promoted to next year. It can be used for improvement of academic results.

Keywords: *Prediction, Data mining, Classification algorithms, Decision trees.*

I.INTRODUCTION

Data mining is the analysis step of the "knowledge discovery in databases" process. Data mining concepts and methods can be applied in various fields like marketing, medicine, customer relationship management, engineering, web mining etc. [6] [10]. Educational data mining is a new technique of data mining that can be applied on the data related to the field of education. There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining [1], concerns with developing methods that discover knowledge from data originating from educational environments. Educational Data Mining uses many techniques such as Decision Trees, Neural Networks, Naive Bayes, K- Nearest neighbor, and many others. [3] [6]. One of the important facts in institution is the rapid growth of educational data. The main goal of any educational institution is to improve education quality. Using these techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for prediction student results. Student's result is the most concern for any institution. By predicting student's result and giving appropriate tutoring to the weak students, the quality of education provided by the institutions can be improved The classification task is used to evaluate student's result and as there are many approaches that are used for data classification, the decision tree method is used here. Decision tree are used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree C4.5 algorithm.



II.BACKGROUND AND RELATED WORK

Presently educational institutions compile and store huge volumes of data such as student enrollment and attendance records, as well as their examination results. Growth in educational data points to the fact that distilling amounts of data requires a more sophisticated set of algorithms. This issue led to the emergence of the field of Educational Data Mining (EDM) [1]. Data mining concept is used to predictive modeling approach. This is used for extracting this hidden information. The predictive models will help to understand how well or how poorly the students in his/her class will perform, and hence the instructor can choose proper pedagogical and instructional interventions to enhance student learning outcomes. The implementation is done in Hadoop framework with Map Reduce and Revolutionary R Enterprise RRE [9]. The predictive models obtains the university students data like attendance, class test, seminar and assignment marks from the students previous database, to predict the performance at the end of the semester [5]. Measuring student performance using classification technique such as decision tree. The task can be processed based on the several attributes to predict the performance of the student activity respectively. The improvement of Prediction/ classification techniques which are used to identify the skill expertise based on their academic performance by the scope of knowledge, The specific needs of studies to improvement for giving the details about the results, such as the accompaniment of students along their learning process, and the taking of timely decisions in order to prevent academic risk and desertion. Lastly, some recommendations and thoughts are laid out for the future development of performance. Helps to analyze the slow learner that are likely study in poor which are used to improve their skill as early to achieve the goal [8]. Four different classification techniques, – decision tree algorithms, support vector machines, artificial neural networks, and discriminant analysis are used to build classifier models [4]. Decision tree algorithms are applied on engineering student's data to predict their performance in the final exam. The outcome of the decision tree predicted the number of students performance [6]. Classification techniques are ID3, Cart, C4.5, Naive Bayes Classify [6][2][7]. For predicting the cognitive skill of students which can be evaluate by conducting the online test. The comparative performance of C4.5 algorithm, Naive Bayes classifier algorithm which one is well suited accuracy for predicting the skill [7].

III. PROPOSED METHODOLOGY

3.1 Classification Techniques for Prediction:

Generally, the classification of data has two step processor learning and a classification step which is used to predict class labels for training data. In Training Data used Flat Files. Flat files are actually the most common data source for data mining algorithms. Flat files are simple data files in text or binary format with a structure known by the data mining algorithm to be used. In fig. represents the methodology of research process can be learned for training data which are analyzing by a classification algorithm. In classification step, test data are used to estimate the accuracy of classification rules. In this study, our discussion focuses on the classification techniques such as C4.5 technique.

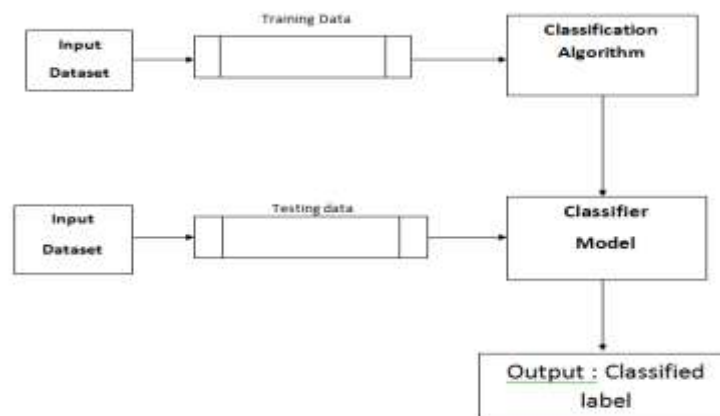


Fig. Prediction System

Dataset:Table 1

Number of Sample	Number of Attributes	Number of Classified Label
100	4	6

1. Number of Sample-Number of Students data which can be Classify
2. Number of Attribute-Input data like Roll Number of students, Previous Result, Unit Test Marks, and AttendancePercentage
- 3.Number of Classify Result- Output data like students will be in First Class with Distinction or First Class or Higher Second Class or Second Class or Pass Class or Fail.

A. Data Collection -The data mining techniques follow the possible value of required attributes used to predict the data .The data set used in this study was obtained from Dr. D Y Patil School Of Engineering and Technology, Lohegaon. Initially size of the dataset is 100.

B. Data Set- From Table II, illustrate the 100 input training dataset, using which the predictive model is built. The training dataset contains information like Unit test Marks of current Year(Test 1,Test 2),Previous Result(FE,SE,TE),Attendance count in Percentage.

Table II

Roll.No	Unit Test Marks of current Year		Previous Results			Attendance Percentage
	Test 1	Test 2	FE	SE	TE	BE
1	12	11	55%	56%	58%	60%
2	12	9	50%	55%	57%	70%
3	5	10	60%	62%	63%	75%

4	7	5	63%	60%	65%	55%
5	0	8	44%	48%	55%	40%
6	15	13	75%	73%	75%	80%
.
.
.
.

3.2 C4.5 Classification Technique

C4.5 builds decision trees from a set of training data in the same way as ID 3, using the concept of information entropy. The training data is a set of already classified samples

At each node of the tree, C4.5 chooses the attribute of the data that effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sub lists.

This algorithm has a few base cases

- * All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- * None of the features provide any information gain. In this case, C4.5 creates a decision node higher using the expected value of the class.
- * Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

```
Input: an attribute-valued dataset D

1: Tree = { }

2: if D is "pure" OR other stopping criteria met then

3: terminate

4: end if

5: for all attribute a ∈ D do

6: Compute information-theoretic criteria if we split on a

7: end for

8: a best = Best attribute according to above computed criteria

9: Tree = Create a decision node that tests a best in the root

10: D v = Induced sub-datasets from D based on a best

11: for all D v do

12: Tree v = C4.5(D v )

13: Attach Tree v to the corresponding branch of Tree

14: end for

15: return Tree
```

IV.ALGORITHM

Generate decision tree:

1. Check if algorithm satisfies termination criteria
2. Computer information-theoretic criteria for all attributes
3. Choose best attribute according to the information-theoretic criteria
4. Create a decision node based on the best attribute in step 8
5. Induce (i.e. split) the dataset based on newly created decision node in step 6

6. For all sub-dataset in step 10, call C4.5 algorithm to get a sub-tree (recursive call)
7. Attach the tree obtained in step 13 to the decision node in step 9
8. Return tree

V. CONCLUSION

The Prediction system allows predicting the academic result and generating reports to provide a base for making academics decisions for higher authorities. Decision tree method is used on student's database to predict the student's result on the basis of student's database. By implementing this system intended to take more accurate prediction on students likely grade in an exam by taking prediction. This study will help the student's to improve the student's result, to identify those students which needed special attention to reduce failing ration and taking appropriate action at right time.

REFERENCES

- [1] Ashish Dutt, MaizatulAkmar Ismail, and TututHerawan, "A Systematic Review on Educational Data Mining", 2016 IEEE
- [2] Vipin Kumar · J. Ross Quinlan, "Top 10 algorithms in data mining", Springer-Verlag London Limited 2007
- [3] Carlos Márquez Vera, Cristóbal Romero Morales and Sebastián Ventura Soto, "Predicting of school failure and dropout by using data mining techniques", The IEEE Journal of Latin-American Learning Technologies (IEEE-RITA) , Feb 2013
- [4] Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," IEEE Access , 2016.
- [5] B.K. Bharadwaj and S. Pal. "Mining Educational Data to Analyze Students' Performance", International Journal of Advance Computer Science and Applications (IJACSA), 2011.
- [6] Surjeet Kumar Yadav and SaurabhPal , "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification", World of Computer Science and Information Technology Journal (WCSIT), 2012.
- [7] M. Mayilvaganan ,D. Kalpanadevi , "Comparison of Classification Techniques for predicting the Cognitive Skill of Students in Education Environment", 2014 IEEE International Conference on Computational Intelligence and Computing Research
- [8] Ankita A Nichat, Dr. Anjali B Raut, "Predicting and Analysis of Student Performance Using Decision Tree Technique" International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), April 2017.
- [9] Midhun Mohan M G & Siju K Augustin Dr. Kumari Roshni V S, "A Big Data Approach for Classification and Prediction of Student Result Using Map Reduce", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)
- [10] Brijesh Kumar Baradwaj Saurabh Pal, "Mining Educational Data to Analyze Students Performance", (IJACSA) International Journal of Advanced Computer Science and Applications, 2011