



Evaluation of Student Classification Based On Decision Tree

Arpit Trivedi

Smt.Chandaben Mohanbhai Patel institute of Computer Applications,
CHARUSAT, India

Abstract: *This paper put forward a simple approach for categorizing student data using decision tree based approach. In this discussion one can identify several different alternatives used for categorization followed by simple architecture. In this study, a database for five subjects' marks of 100 students for each four different classes named with Excellent Student, above average student, below average student and Poor Student. For taking measures of category of specific student, a frequency measure is used as a feature extraction. The most frequent five subject marks of each of the students are used to develop a trained classifier. This trained classifier is used for predicting the class (Excellent Student, above average student, below average student and Poor Student) for indefinite student automatically.*

Key Terms: *Decision Tree, Feature Extraction, Text Categorization (TC), Frequency Measure, Student Data*

I. INTRODUCTION

Text categorization is applied for classifying a document in to predefined set or classes of categories or class tags. For student result, marks of subjects are needed to be classified to place them in proper columns. The current study focuses on the implementation of TC system for automatic predication of the class for student using decision tree approach. The performance of TC system is analyzed to further improve the quality of the system and technique for the type of database used. We use decision tree because it supports the divide and concur technique for training set. Decision Tree is a type of classifier using which we can categorize our data in terms of internal nodes and leaf. Each node consists of set of cases. In this study internal nodes are the records of students and leaf is marks of each subject.

The paper is prepared as follows. Section II provides the details of student's marks for the subjects. In Section III there is architecture of proposed text categorization. Conclusion and some future plans for further work is mention in Section IV. Finally, the minority vital references are mentioned.

II. DATABASE

A quantity of records has been collected for developing system for classifying the student according to marks of five subjects of each student. We can classify a student based on 4 categories: Excellent, Above Average student, Below Average Student and Poor Student. For each of the 4 classes, we collected five subjects' marks for 100 student summing total of 500 subjects' marks. The data of marks for five subjects has been taken from the Examination taken for each subject for testing the baseline TC.

III. DESIGN OF SYSTEM

In this paper, a collection of five subjects' marks of 100 students has been taken for developing student result classification system. The output of this system contains a number of students classified into prescribed classes for their result according to their marks of five subjects. Training set consist of marks of five subjects for 100 students which are preprocessed first then after classifier is used to predict the class for each student.

The TC system works in two phases which are as follows:

1. Training Phase
2. Testing Phase

Training phase is used to preprocess and features are collected automatically then after with the use of decision tree algorithm is used to train the features which are marks of five subjects for each student. For the development of this mechanism which is completely based on feature selection technique for the training process. The decision tree is a simple method using which the predictions are made for the test data based on the best matches with training instances. Testing phase consists of preprocessed data and a feature from which class of each student is predicted on the basis of output is generated. The proposed architecture for the result classification of student contains following set of activities.

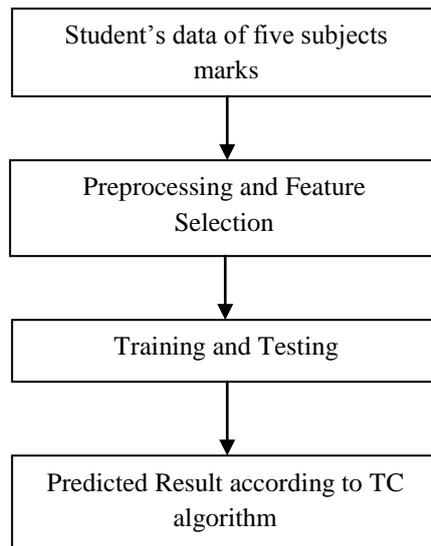


Figure.1: Activities of Text Classification Process

Using above process as shown in Figure.1 design the system consist of marks of five subjects for 100 students. The binary decision tree is used to classify the student result by the process which starts from root node until it reaches to leaf node that predicts the class for each student.

Among 100 students with five subjects marks used in study, 20 students of each class are used as a training set. The overall accuracy of this TC system using decision tree is 80% predicting accurate classes for 47 students among 80 student's five subjects' marks. Table 1 show the matrix after classification is done.

Table.1: MATRIX FOR CLASS BASED ON MARKS OF STUDENTS

	Extra Ordinary	Above Average	Below Average	Poor
Extra ordinary	14	1	2	3
Above Average	2	8	6	4
Below Average	2	5	8	5
Poor	2	5	4	9

IV. CONCLUSION

Using this method, the teachers can easily measure the performance of students. This paper provides a prediction model for student grading scheme for any institution or academic organization. The result of this paper shows the model based on five subjects marks of each student. There are many algorithms available to classify the text or document and we can also get accurate information using those algorithms. This study proposed an elementary approach to classify the category of the student based on his\her marks of five subjects. Decision tree based model of TC predicts the classes accurately as the test data is collected from the same source.

REFERENCES

- [1] Member, IACSIT Pratiksha Y. Pawar and S. H. Gawande, "A Comparative Study on Different Types of Approaches to," *International Journal of Machine Learning and Computing*, vol. 2, no. 4, pp. 423-426, August 2012.
- [2] Krishnendu Ghosh Zahid Hasan, "A Decision Tree Based Text Categorization for," *IJETET*, 2013.
- [3] Jungyun Seo Youngjoong Ko, "Automatic Text Categorization by Unsupervised Learning".
- [4] Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao Kalpesh Adhatrao, "PREDICTING STUDENTS' PERFORMANCE USING ID3 AND C4.5 CLASSIFICATION ALGORITHMS," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 3, no. 5, pp. 39-52, September 2013.
- [5] "Classification of Student's data Using Data Mining Techniques for Training & Placement Department in Technical Education," *International Journal of Computer Science and Network (IJCSN)*, vol. 1, no. 4, pp. 121-126, August 2012.
- [6] Fabrizio Sebastiani, "Text Categorization".
- [7] R. S. Bichkar R. R. Kabra, "Performance Prediction of Engineering Students using Decision Trees," *International Journal of Computer Applications*, vol. 36, no. 11, pp. 8-12, December 2011.
- [8] Thanassis Tiropanis and Hugh C Davis Farhana Sarker, "Students' Performance Prediction by Using Institutional Internal and External Open Data Sources".