Student Performance Prediction And Analysis

Ditika Bhanushali[1], Seher Khan[2], Mohommad Madhia[3], Shoumik Majumdar[4]

Bachelor of Engineering, Dept. of Computer Engineering, RGIT, Mumbai, Maharastra, India[1-4]

Abstract: The main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. This project is developed to justify the capabilities of students in various subjects. In this, the classification task is used to evaluate students' performance and as there are many approaches that are used for data classification, the decision tree method and probabilistic classification method is used here. By this task we extract knowledge that describes students' performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling. In addition to this, we will also compare the results generated by two classification algorithms, namely ID3 and Naïve Based algorithm, and thereby determine which algorithm is more accurate.

Keywords: Data Set, Data Mining, Classification Task, ID3 Algorithm, Naïve Based Algorithm.

I. INTRODUCTION

Over the past 35 years, a vast amount of knowledge has been accumulated on text mining for Information Retrieval (IR). Using automated text mining algorithms to discover knowledge from natural language texts provides numerous challenges but also offer unique possibilities. One of the most natural forms of storing information is in the form of natural language texts .This can be easily interpreted by a human but it is still a great challenge for computers to derive meaning from this data. However, computers do offer an important advantage over human capabilities: computing power. This means that computers can find patterns, which are non-trivial recurrences, within data faster and more accurate than their human counterpart, but this can only be done if the structure of the data is known. Natural language does contain implicit grammatical structure, but these structures are deeply complex and vary across different languages. The main aim of this project is to use data mining methodologies to study students' performance in the courses. Data mining provides many tasks that could be used to study the student performance. In this research, the classification task is used to evaluate students' performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information like Attendance, Class test, Seminar and Assignment marks were collected from the student's management system, to predict the performance at the end of the semester. This paper investigates the accuracy of data mining classification methods for predicting student performance.

II. NAÏVE BAYES ALGORITHM

Naive Bayes has been studied extensively since the 1950s. Also known as Nave Bayesian, it is a statistical learning algorithm based on Bayes rule to compute joint probability. It assumes conditional independence amongst the attributes. This is used as a classification tool by first dividing the data into independent classes and calculating the probability distribution for each attribute of each class. For classification, the Nave Bayesian finds the probability for the unknown in any given class and selects the class with the highest probability. The basis of Naive Bayes algorithm is Bayes' theorem or alternatively known as Bayes' rule or Bayes' law. It gives us a method to calculate the conditional probability, that is, the probability of an event based on previous knowledge available on the events. It is a popular method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines. It is one of the most used data mining classifiers among the ones used for data prediction.

lll. ID3 ALGORITHM

The ID3 algorith was the first of three Decision Tree implementations developed by Ross Quinlan. It builds a decision tree for the given data in a top-down fashion, starting from a set of objects and a specification of properties Resources and Information. Each node of the tree, one property is tested based on maximizing information gain and minimizing entropy, and the results are used to split the object set. This process is recursively done until the set in a given sub-tree

is homogeneous (i.e. it contains objects belonging to the same category). The ID3 algorithm uses a greedy search. It selects a test using the information gain criterion, and then never explores the possibility of alternate choices.

IV. AIM AND OBJECTIVES

The main objective of this project is to use data mining methodologies to study students performance in the courses. Data mining provides many tasks that could be used to study the student performance. In this research, the classification task is used to evaluate students performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information like Attendance, Class test, and Assignment marks was collected from the students management system, to predict the performance at the end of the semester.

The overall vision for the Performance Prediction System is that it will fulfill the following objectives:
• To create a user friendly web interface on which the system can be implemented.
• To be able to predict the student performance using the Naive Bayes and ID3 algorithms.
• To determine the more efficient data mining classifier among the two classifiers used i.e. Naive Bayes classifier, and the ID3 classifier.
• To be able to make the performance prediction methodology more efficient and accurate.

V. PROBLEM STATEMENT AND SCOPE

Educational organizations are one of the important parts of our society and playing a vital role for growth and development of any nation. Educational data mining is the application of data mining. It is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. Educational data mining is an emerging trend, designed for automatically exploring the unique types of data from large repositories of educationally related data. Quite often, this data is extensive, fine grained, and precise. The main objective of this paper is to use data mining methodologies to study students performance in the courses. Data mining provides many tasks that could be used to study the student performance. In this research, the classification task is used to evaluate students performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information like Attendance, Class test, Seminar and Assignment marks were collected from the students management system, to predict the performance at the end of the semester. This paper investigates the accuracy of Decision tree techniques for predicting student performance. The faculty cannot find out students abilities and their interest easily so that they can enhance them in it. Thus it may affect with poor university results, placement and career of individual. The impact is- it help us from fulfilling mission and vision of the institute. If the project get successful then it will be great help for faculty to enhance education system.

VI. PROPOSED SYSTEM

Nowadays, e-education and e-learning is highly influenced. Everything is shifting from manual to automated systems. The objective of this project is to predict performance of a student based on certain attributes of the student such as his semester attendance, unit-test marks, last semester exams and aggregate CGPA of the student in the previous semesters. In the proposed system, we will predict the performance of the students using two difference data mining classifiers, namely ID3 data mining classifier and the Naive Bayes classifier. The main aim of the proposed system, is to find the more efficient data mining classifier amongst the two. This would result to finding out a more efficient and time saving algorithm to predict the performance of a student. The new system will be cost and time efficient. This will have simple operations.

VI. METHODOLOGY

The main aim of the system is to predict the future performance of the student using certain data of the student such as pervious semester marks, attendance records, etc. After predicting the student performance, the system will also compare the results generated by two classification algorithms and there after determine which of them is more accurate and efficient.
The data to be provided as the input must have the values of the attributes classified into specific variables, for example, the student marks for the previous semester can be classified as good if marks $>= 70\%$, average if $70\% >$ marks $>=55\%$ and poor if marks $<55\%$. The attributes used in this system are: previous semester marks, attendance, project marks, seminar attendance, unit test marks, extra curricular activities, assignments and practical evaluation. This data is then normalized and fed as an input to the system. Using this normalized data, the system runs the ID3 and Naives Bayes algorithm on it and classifies the data. This classified data is then used to predict the final semester marks of the student.

VIII.ANALYSIS

The very first phase in any system developing life cycle is preliminary investigation. The feasibility study is a major part of this phase. A measure of how beneficial or practical the development of any information system would be to the organization is the feasibility study. The feasibility of the development software can be studied in terms of the following aspects:

1. Operational Feasibility: The Application will reduce the time consumed to maintain manual records and is not tiresome and cumbersome to maintain the records. Hence operational feasibility is assured.
2. Technical Feasibility: Minimum hardware requirements: 1.66 GHz Pentium Processor or Intel compatible processor. 1 GB RAM. 80 MB hard disk space.
3. Economic feasibility: Once the hardware and software requirements get fulfilled, there is no need for the user of our system to spend for any additional overhead.
For the user, the Application will be economically feasible in the following aspects: 15 The Application will find out the more efficient algorithm to predict the student performance. Hence reducing the extra cost used on the less efficient algorithm. Our Application will reduce the time that is wasted in manual processes.

IX.HARDWARE AND SOFTWARE REQUIREMENTS

Hardware Requirements
• Processor: 1.5GHz or above.
• RAM: 4GB or more.
• HDD: 100GB or above.

Software Requirements
• Operating System: Windows XP or higher.
• Languages: Java (Frontend), Python (Backend).
• Software: Visual Studio 2010 or higher
• Database: MYSQL server.

X.FUTURE SCOPE

There are quite a few things that can be polished or be added in the future work.
• We have opted to use two data mining classifies in this project namely the ID3 and Naive Bayes classifier. There are more classieres such as the Bayesian network classifier, Neural Network classifier and C4.5 classifier. Such classifiers were not included in this paper and could be counted in future to give a more data to be compared with.
• Though, we have have taken into consideration the academic data of many students, there are still many students and ample amount of input data that could futher be used. With more and more demand for not only student but also performance prediction as a whole, there is alot of data that can be taken into consideration for more accurate results. There is alot of scope for student performance prediction in the data mining world.

ACKNOWLEDGEMENT

REFERENCES

[1] P.Veeramuthu Dr.R.Periasamy Application of Higher Education System for Predicting Student Using Data mining Techniques International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 1 Issue 5 (June 2015)
[2] Umesh Kumar Pandey , S. Pal A Data Mining view on Class Room Teaching Language IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011 ISSN (Online): 1694-0814
[3] Mrs. M.S. Mythili , Dr. A.R.Mohamed Shanavas An Analysis of students performance using classification algorithms IOSR Journal of Computer Engineering (IOSR-JCE) eISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. III (Jan. 2014), PP 63-69 [
4] G.Paul Suthan and Lt.Dr. Santhosh Baboo Hybrid CHAID a key for MUSTAS Framework in Educational Data Mining IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814
[5] S. T. Hijazi, and R. S. M. M. Naqvi, Factors Affecting Students Performance: A Case of Private Colleges, Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

[6] Y. Ma, B. Liu, C.K. Wong, P.S. Yu, and S.M. Lee, Targeting the Right Students Using Data Mining,Proceedings of KDD, International Conference on Knowledge discovery and Data Mining, Boston, USA, 2000, pp. 457-464. n

[7] A. L. Kristjansson, I. G. Sigfusdottir, and J. P. Allegrante, Health Behavior and Academic Achievement Among Adolescents: The Relative Contribution of Dietary Habits, Physical Activity, Body Mass Index, and Self-Esteem, Health Education Behavior, (In Press).

[8] J. A. Moriana, F. Alos, R. Alcala, M. J. Pino, J. Herruzo, and R. Ruiz, Extra Curricular Activities and Academic Performance in Secondary Students, Electronic Journal of Research in Educational Psychology,Vol. 4, No. 1, 2006, pp. 35-46.

[9] M. Bray, The Shadow Education System: Private Tutoring And Its Implications For Planners, (2nd ed.), UNESCO, PARIS, France, 2007.