Student Performance Forecasting

[1]E. Soumya, [2]K. Harshini, [3]S. Naveen, [4]M. Sudheer, [5]B.S.V. Kalki

[1]Associate Professor, [2,3,4,5]B.Tech,

[2,3,4,5]*Department of Computer Science and Engineering, St. Martins Engineering College, Hyderabad, Telangana, India.*

ABSTRACT

The students tend to write different kinds of exams one such is the mid-term exams. The students get marks based on how they have performed. So here we are using this performance prediction to evaluate the student's future score. This helps the students to know where they stand and helps them to improve themselves. This project is based on data mining process to evaluate the scores. C4.5 is the algorithm used for this process.. Admin and user will use the system. Here user will be the student. Admin can add faculty details, student details and the student marks with basic information. Admin must add academic details of the student, like his marks attained for each exam in every subject. The faculty can view the score of the students and predict it. The student gets to view the final predicted score that is in the form of report which consists of every subject for that semester. This system can be used in schools, colleges and other educational institutions.

Index Terms- Student performance prediction, algorithm, report.

1. INTRODUCTION

There are many colleges that provide education for many students. Large numbers of students are graduated every year. The number of students appearing for the exam tends to increase if the students fail in that subject. This leads to multiple students appearing for the re-examination.

Understanding and analyzing the factors for poor performance of the students lead to the idea, this project helps to evaluate the students based on the score they attain in the college.

Although colleges collect an enormous amount of students' data, but this data remains wasted and the data is left unused as it is not used in a proper manner.

If colleges could get to know the reason for low performance earlier and is able to predict students' marks, this knowledge can help them in taking measures beforehand, so as to improve the performance of such students. It will be a positive situation for all the stakeholders of colleges i.e. management, teachers, students and parents. Students will be able to identify their weaknesses beforehand and can improve themselves. Teachers will be able to plan their lectures as per the need of students and can provide better guidance to such students. Parents will be reassured of their ward performance in such colleges. Management can bring in better methods and rules to increase the performance of these students with additional facilities. Eventually, this will help in producing skillful workforce and hence sustainable growth for the country.

There are many students who write supply exams due to less score attained in the final exams. Sometimes the reason is unknown or the student feels that the subject is very hard. If colleges could get to know the reason for low performance earlier and is able to predict students' marks, this knowledge can help them in taking measures beforehand, so as to improve the performance of such students.

Analysis and prediction with the help of data mining techniques have shown worthy results in the area of fraud detection, predicting customer behavior, financial market, loan assessment, real-estate assessment and intrusion detection. It can be very effective in Education System as well. It is a very powerful tool to reveal hidden patterns and precious knowledge.

Substantial work is done towards the usage of data mining techniques in Education, but still there are many untouched areas and no unified approach is followed. This project presents an overview of how the project is done and how it is helpful for the students and the faculty for each and every subject.

The paper firstly consists of how this project is done and what are the improvements that have occurred. The implementation and the techniques used are also a main part of this paper.

This helps the student to improve themselves in the forth coming exams by preparing themselves beforehand. This also helps the faculty to know the number of students who are weak in their subject and helps them to find innovative and easier ways to teach the students.

2. LITERATURE SURVEY

The idea of using data mining in higher education has been put forward by many researchers and authors who have explored and discussed the performance of several students.

V.Ramesh, et al. have attempted to find suitable prediction techniques using data mining tool WEKA to enhance the quality of higher educational institutions.

Guan Li has compared the accuracy of data mining methods to classifying students in order to predict student's class grade.

J.F. Superby led an examination to research to decide the components to be considered we will utilize a model adjusted from that of Philippe Parmentier (1994). As it were the thought is to decide whether it is conceivable to foresee a choice variable utilizing the illustrative factors which we held in the model.

2.1 DATA MINING

Data Mining is the process of mining data from large data center that consists of many data bases. The data base consists of huge amount of data that is uploaded, viewed, modified and deleted. The data is mainly taken as raw data. This can be used in research centers, educational institutions, banks, hospitals, private sectors, public sectors, government sectors, and social media. One such is Aadhar card database. There is huge amount of data that is being added every day. The data is measured in bytes as per now terabytes data exists. Using of data mining is simpler as it stored in tabular formats which provide faster and easier access.

2.1.1 CLASSIFICATION

Process of categorization of the data is called as classification. It's the process in which the data is recognized and understood. As per data mining the classification is used for creating classes and adding data as per the classes the data is saved and modified. For building a classifier we need to learn regarding the data. The algorithm is used for classification. The algorithm builds classifier. The classifier is built from the training set which has tuples and associated class labels. The tuples have all the data about the training set. These tuples are nothing but data and also known as samples. If the classification is done correctly then the data can be easily accessed.

2.1.2 CLUSTERING

Clustering is the process in which the objects are formed into groups as per their common properties. Objects that belong to similar classes form a single cluster and the objects that have different properties don't form a cluster. In the data mining the clustering is done based on the data that has meaningful sub classes. The sets of data is categorized and formed in the classification. It can be used in market research, data analysis, image processing and much more. The clustering is used for classifying documents on the web for discovery of the information.

2.2. CLASSIFICATION OVER CLUSTERING

The data is to be mainly formed into different sets this can be done under classification process. In the classification we have a set of classes that are predefined. This helps to know which object belongs to which class. In this we have prior knowledge of classes. We classify into some known classes. The classification algorithms are supported.

Whereas clustering is grouping if similar objects, which means that the objects that have any relation in between them. The knowledge of the classes is null. The clustering happens based on the similar patterns. It doesn't fully support the classification algorithms as it can lead to mismatch.

2.3. ISSUES REGARDING CLASSIFICATION

2.3.1. MISSING DATA

The missing data means that the data is not completely present. This can occur due to the addition of incorrect data, irrelevant data or incomplete data. This can occur due to no response also. The missing data can refer to multiple things in the data. It can occur due to invalid data also. Sometimes the missing values are ignored or exclude the records that consist of the missing values. But avoiding these records can lead to any issue as they maybe interlinking of the data or references to that data. So we use a null value or some common value in such places. In our case, the chances of getting missing values in the training data are very less. The training data is to be retrieved from the admission records of a particular institute and the attributes considered for the input of classification process are mandatory for each student. The tuple which is found to have missing value for any attribute will be ignored from training set as the missing values cannot be predicted or set to some default value. Thinking about low odds of the event of missing information, overlooking missing information won't influence the exactness unfavorably.

3. OVERVIEW OF THE SYSTEM

The architecture can be taken as a process in which the flow of the work occurs and how it is reflecting in the project. The steps include the way in which the data is gathered and processed to get a final output. The main architecture of the system has four main components based on which it works.

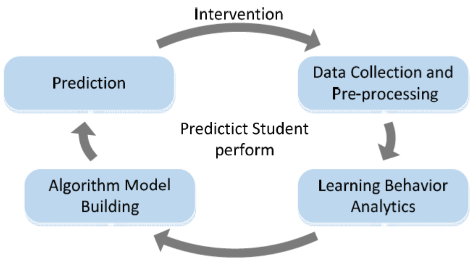
Intervention
Prediction
Data Collection and Pre-processing
Predictict Student perform
Algorithm Model Building
Learning Behavior Analytics

Figure 3.1: Architecture

The architecture mainly consists of the Collection of the data by the admin and the pre-processing of the data. In this the admin gets to add the faculty, student and marks accordingly. The learning behavior analytics helps in gathering the complete information and processing it. It includes how the subject is learnt and how it is performed such that the some marks attained and how are these marks going to be evaluated. The algorithm model building includes the use of the C4.5 algorithm. The algorithm is used to predict the students score. The prediction is the final step that takes the score and evaluates the student. A final score card/report is generated.

The modules mainly are categorized as admin, faculty and the student. In this the admin gets to login with his secure credentials and add faculty, subjects and the students. With regard of the student's roll numbers the admin gets to post each students marks attained in the mid exam per subject. This data gets saved in the database.

The admin creates login for each and every student and faculty. Based on that the faculty gets to login into his/her id and check the marks that have been uploaded by the admin. The faculty can view the marks of the whole section. The faculty gets another option named as Notifications that helps the faculty to know the students who are below the average mark. By checking their marks the faculty can upload important material to the students so that they can study from the provided material and get a better result in the final exams.

The faculty can predict the marks of the student. The student can login into the provided id and check his/her marks gained for every subject in each semester. This helps the student to know where he/she stands and how much he/she needs to improve himself or herself. The student can download the study material provided by the faculty of each subject and prepare from that. The student can check the score card which tells his/her average mark as well as the final score that is predicted based on the performance.

4. METHODOLOGY USED

The existing system mainly consists of the schools, colleges to assess students based on their regular mock tests or the mid-term exams. The students are primarily conducted exams for some specified marks i.e., consider 50marks. According to the marks they have attained the faculty enter their marks for every exam and take an average of it.

Previously the faculty used to write down the exams marks in a marks sheet and perform the sum followed by average of the marks by using a calculator. This has been manual until the use of excel sheets have been increased.

Most of the times the times the faculty tend to enter the marks in an excel sheets and attain the sum of the marks of one student and drag the cursor to perform the same for the rest of the students. According to that they calculate the average by applying a formula.

C4.5 is a well-known algorithm used to generate an output. It is an extension of the ID3 algorithm used to overcome its disadvantages. The C4.5 algorithm can be called as a statistical classifier. The C4.5 algorithm made a number of changes to improve the ID3 algorithm.

Some of these are:

- Handling training data with missing values of attributes
- Handling differing cost attributes
- Pruning after its creation
- Handling attributes with discrete and continuous values

Give the preparation information a chance to be a set S = s1, s2,….. of the characterized tests . Each example S1 = x1,x2…. Is where x1, x2… … speak to properties or highlights of the example. The planning data is a vector

C = c1, c2… . Where c1, c2,.. Address the class to which every illustration has a place with. At every hub of the tree, C4.5 picks one characteristic of the information that most viably parts informational collection of tests S into subsets that can be one class that can be one class or the other. It is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The trait factor with the most noteworthy standardized data pick up is considered to settle on the choice. The C4.5 algorithm then continues on the smaller sub-lists having next highest normalized information gain. The implementation mainly includes three modules admin, faculty and student. The admin gets to login with secure credentials and add students, faculty, student marks. With the secured credentials the faculty logs in and checks the student's marks. Apart from that the student marks are categorized as below 15 and above 15 this helps them to know where they stand. The faculty can provide study material to the students who are below the average category. The faculty predicts the future marks of the student individually. The faculty can also view the marks of the entire student's present in the class. The students get to login and check his/her results. They can view the material provided by the faculty. The student can also view the report that is generated based on the marks attained by them.

The techniques used are as follows:

For the storage of the data in the database we are using SQLyog so that huge amount of data can be stored in the database. Any change can be reflected soon and the query writing is simpler. The data changed can be reflected easily when we refresh the page. Data is clearly visible in the table in the form of rows and columns.

The coding is mainly done in Netbeans which is a flexible platform for coding. Since the project is mainly based on Webpages we use Html and java for the coding. The linking of the pages is done using Java code. The connection between the Webpages and the SQL is done by creating a connection and path. This helps in storing the data in the exact place that we need.

For supporting the java code we use JavaSE for the importing and using of the packages and making the code run successfully.

For the designing the required layout for the project we use the packages of templatemo. This is used for HTML5 CSS templates. The project is mainly a website so the errors can be reflected easily and the debugging depends on the loops used and the packages used.

The uploading of different types of materials is possible and the download is simple. It doesn't have any issue if the path is created correctly.

5. RESULT AND DISCUSSION

The output hold the information on the student side which includes the marks that are predicted by the faculty. A score card or report is generated that helps the students to know their capability. The report is calculated based on the marks that are attained in the mid-term exams. The algorithm runs in the faculty module. The mid-term marks are taken an average of the two exams and the algorithm runs. The main part of this is that the algorithm runs based on the queries and the logic given. We kept a basic logic that the students who attain below 15 marks in the average score will be getting the study materials that are uploaded by the faculty as per subject. The faculty can post the study material via a document format or a pdf format. The uploaded data is being stored in the database by using a specified path which can be downloaded and used by the students accordingly.

The predicted report is based on the scores they may attain in the annual exam if their practice and performance is just the same. If the student doesn't prepare well the marks may not be considered as the exact prediction. In this project, prediction parameters such as the reports are generated are not updated dynamically within the source code. Later on, we intend to make the whole usage dynamic to prepare the forecast parameters itself when new preparing sets are bolstered into the web application. Also, in the current implementation, we have not considered extra-curricular activities and other vocational courses completed by students, which we believe may have a significant impact on the overall performance of the students. Considering such parameters would result in better accuracy of prediction.

6 CONCLUSIONS

The admin gets to add all the details thus avoiding confusion of having multiple admins or multiple faculty who add marks each time. This can lead to concurrency or missing values. The faculty can get to login based on the credentials provided. Materials can be uploaded and all the students marks for their subject. The students get to know where they stand in a class of students. The study materials provided help the student to study more and perform well in the final exams. If the predicted score is less, then the student can get to know before-hand. The final performance is based on practice.

7. REFERENCES

[1] R.S. Bakar, K Yacef, "The state of educational data mining in 2009: A review and future visions", *JEDM-Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3-17, 2009.

[2] J.I.M.S. Barracosa, *Mining Behaviors from Educational Data*, 2011.

[3] [online] Available: http://library.queensu.ca/webedu/grad/Purpose_of_the_Literature_Review.pdf.

[4] [online] Available: http://ar.cet1.hku.hk/am_literature_reviews.htm.

[5] B. K. Baradwaj, S. Pal, "Mining educational data to analyze students' performance", *arXiv preprint arXiv:1201.3417*, 2012.

[6] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, V Honrao, "Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms", *arXiv preprint arXiv*, vol. 3, no. 5, pp. 39-52, September 2013.