FAILURE AND DROP OUT PREDICTION OF COLLEGE STUDENTS USING DATA MINING

Sagar N. Gunjal, Swapnil M. Ide, Amol P. Jadhav, Akshay A. Kekan,Hemant B. Kokate

Student, Computer engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India

ABSTRACT

In this article we have described the prediction system which is about analysing the student's academic performance..Prediction system is also able to predict the student's academic failure chances. This system has its complete technology based guidance alongwith knowledge from past experiences . It presents users a platform where they can analyse their academic performance .The staff or co-ordinators are able to get individual performance analysis of each student .This system can be used by college students and professors for analyzing the performance of student. It also provides subject wise key skills of students in which he/she is have upper hand .This system reduces the manual work & allows the college to analyse every student's academic improvements and weak zones . Having the performance analysis of each student & its prediction of failure, the assigned mentor of that students can guide them forward for better results.

Keywords: *Prediction System, Performance Analysis*

1. INTRODUCTION

The Manual approach for the performance prediction of studens is very complex and unsure. Professors and educational institute don't have any automated process for the analysis. Even students are not having any deep analysis of their academics performance. In the era of technology, by implementing this project, we are providing platform where one can get over of all this inconvenience.

The project is about developing a system where concerned professors and educational institute can analyse student's performance in academic subjects more efficiently and with less efforts.Also this system provides the prediction of students who are at the risk of failure and dropout. System can collect large data of students with particular attributes which are necessary for the analysis.

 Also it contains the subject dependencies, means the subjects with its core area are related contentwise and conceptwise. So this can help student to analyse his/her key skills like Programming, Computation, Networking, Operating System and Microprocessor. These subject dependencies are designed with the knowledge of syllabus alongwith professor knowledge of the core areas of computer engineering.

 The system will have five phases as follows:

1. Data Collection.
2. Data Preprocessing.
3. Naïve Bayes Algorithm.
4. Prediction.

1.2 Data Collection

This phase is basically for the collection of data of students which can hava several attributes such as name, branch, address, subjectwise marks, attendance. This data of students is stored using MYSQL WORKBENCH 5.2 . System can store such large number of records of students. Here only the necessary data which belong to students is collected , the parameters which are not so necessary are not taken into consideration.

1.2 Data Preprocessing

Data preprocessing phase is carried out for removal of unwanted data. The attributes which are not necessary for the performance analysis and prediction are not considered. The attributes such as mobile number, address, etc are removed. For Preprocessing of data we execute SQL query in which the impactful attributes such as subjectwise marks, live backlog, previous backlog are taken into consideration. Also the missing values are filled.

1.3 Naïve Bayes Algorithm

As we are processing such large amount of data we are using data mining. In that we approached towards naïve bayes algorithm. Naive Bayes belongs to a group of statistical techniques that are called 'supervised classification' as opposed to 'unsupervised classification. In 'supervised classification' the algorithms are told aimed at two or more classes to which texts have previously been assigned by some human(s) on whatever basis. In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. The naïve baysian formula is as:

$$\text{Posterior Probability} = \frac{\text{Likelihood} * \text{Class Prior Probability}}{\text{Prior Probability of Predictor}}$$

Three main classes has been designed : 1. Percentage($<=45\%$; $46 - 54$; > 54)
2. Live Backlog
3. Previous Year Backlog

1.4 Prediction

The Final Prediction is displayed in this phase. The student and staff will get to know failure risk of each individual student. Also the key skills of each student will be displayed according to his/her previous and current marks in the particular subject.

2. SYSTEM DESIGN

As shown in fig.-1 Super admin is the highest authority of system having all privileges. Under super admin, we have Data Collection, Data Preprocessing, Data Mining and Prediction.
Data Collection deals with collecting and storing the data, Preprocessing is for removel of unwanted data. In Data Mining , we use naïve bayes algorithm for the actual implementation and in prediction the final result is displayed.
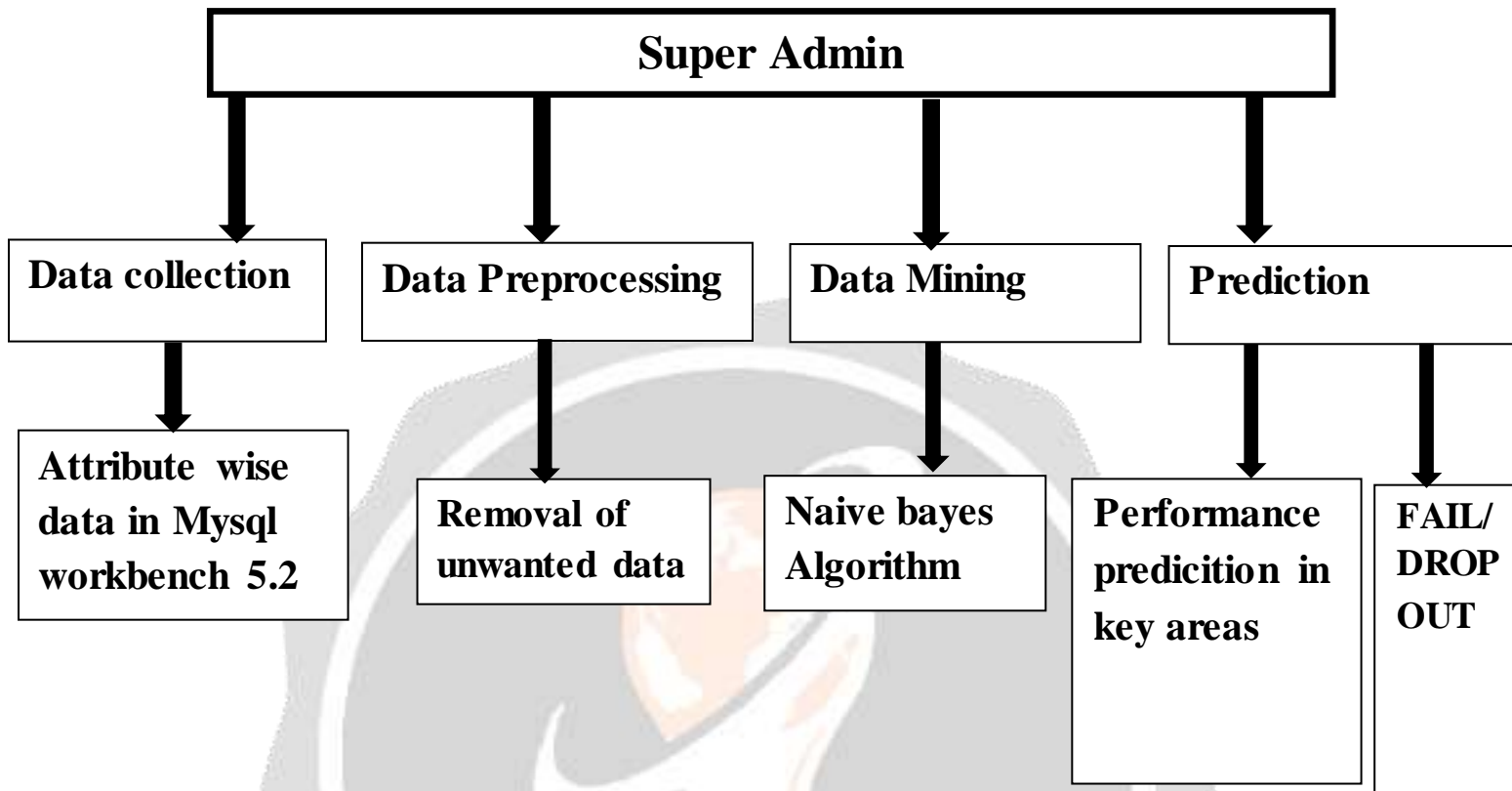
Fig. 1: System design

3. PLATFORM TO BE USED FOR IMPLEMENTATION

3.1 JAVA

Java is a set of computer software and specifications developed by Sun Microsystems, which was later acquired by the Oracle Corporation, that provides a system for developing application software and deploying it in a cross platform computing environment. Java is used in a wide variety of computing platforms from embedded devices and mobile phones to enterprise servers and supercomputers. While they are less common than standalone Java applications, Java applets run in secure, sandboxed environments to provide many features of native applications and can be embedded in HTML pages. The latest version is Java 8, the only supported version as of 2015.

3.2 MySQL

MySQL most popular and widely accepted open source relational database management system. SQL is an abbreviation for structured query language. Relational database model are based on RDBMS. The most important MySQL RDBMS features is an ability of using tables for data storage while maintaining and enforcing certain data relationships [2]. In RDBMS data in MySQL is structured in database tables. Fields and records, we used this for our project.

3.3 JSP

Java Server Pages (JSP) is a server-side programming technology that enables the creation of dynamic, platform independent method for building Web-based applications. JSP have access to the entire family of Java APIs, including the JDBC API to access enterprise databases.

Java Server Pages (JSP) is a technology for developing web pages that support dynamic content which helps developers insert java code in HTML pages by making use of special JSP tags. A JavaServer Pages component is a type of Java servlet that is designed to fulfill the role of a user interface for a Java web application. We can write JSPs as text files that combine HTML or XHTML code, XML elements, and embedded JSP actions and commands. Using JSP, we can collect input from users through web page forms, present records from a database or another source, and create web pages dynamically. JSP tags can be used for a variety of purposes, such as retrieving information from a database or registering user preferences, accessing JavaBeans components, passing control between pages and sharing information between requests, pages etc.

3.4 Apache Tomcat Server

Apache Tomcat is an open source software implementation of the Java Servlet, JavaServer Pages, Java Expression Language and JavaWeb Socket technologies. Apache Tomcat is developed in an open and participatory environment and released under the Apache License version 2. Apache Tomcat is intended to be a collaboration of the best-of breed developers from around the world.
Tomcat is an application server from the Apache Software Foundation that executes Java servlets and renders Web pages that include Java Server Page coding. Described as a "reference implementation" of the Java Servlet and the Java Server Page specifications. In order to use Tomcat for developing web applications, you must first install it (and the software it depends on). The required steps are outlined in the following subsections.Tomcat 7.0 was designed to run on Java SE 6.

4. APPLICATION

This system is aimed to have the near to perfect performance prediction of college students. Also to forecast the risk of dropout of individual student.From this the students, staff, parents of students & whole educational institute will have a deep analysis of academica performance of student.

5. CONCLUSION

This paper describes the complete idea of implementing a Failure and dropout prediction of college students.System is well structured with having perfect criterias from past experiences. The students will have an idea of his/her strong zones in core areas of computer engineering and it will guide them to the proper way of approaching for area in which they want to work.

6. REFERENCES

[1] Mr. C. MRQUEZ-VERA, Mr. C. ROMERO, Mr. S. VENTURA "PredictingCollege Failure Using Data Mining",cmarquezvera@hotmail.com.

[2] Dr. A. Padmapriya "Prediction of Higher Education Admissibility using Classification Algorithms",International Journal of Advanced Research in Computer Science and Software Engineering,2012.

[3] Mr. WILAIRAT YATHONGCHAI, Mr.CHUSAK YATHONGCHAI , Mr. KITTISAK KERDPRASOP , Mr. NITTAYA KERDPRASOP "Factor Analysis with Data Mining Technique in Higher Educational Student Drop Out",Latest Advances in Educational Technologies.

BIOGRAPHIES

Mr. Sagar Gunjal is student of Sanjivani College of Engineering and received BE degree in Pune university 2015-16

Mr.Swapnil Ide is student of Sanjivani College of Engineering and received BE degree in Pune university 2015-16

Mr. Amol Jadhav is student of Sanjivani College of Engineering and received BE degree in Pune university 2015-16

Mr. Akshay Kekan is student of Sanjivani College of Engineering and received BE degree in Pune university 2015-16

Mr. Hemant Kokate is student of Sanjivani College of Engineering and received BE degree in Pune university 2015-16