Predictive Analytics in Engineering Institute using Python

Priyanka Mhetre[1], Sneha Patil[2], Prof S.M.Patil[3]
BharatiVidyapeeth College of Engineering, Belpada, Navi Mumbai, India

Abstract:
The current education system has a fixed syllabus. Since syllabus is fixed and not based on student's interest, it sometimes fails to deliver the planned outcome. This paper proposes a predictive model which can be used as a tool for education management. This tool will help in managing the outcomes of courses and developing appropriate curriculum. Through this predictive model we can predict the future outcome of courses and students by analyzing the past history of students related to that courses and courses related to students. This model will help to increase the success of planned outcomes.

Keywords: Predictive model, Python, Naïve Baye's algorithm, data mining, data analytics, curriculum management, data preprocessing.

I.   INTRODUCTION

In current education system, we fail to know about the student's preferable subjects, reasons for failure of students, teachers, course outcomes and organizations. This drawback of education system results an increase in failure rates. To overcome this problem we are proposing a solution based on predictive analytics. This predictive model will help us to predict about the student based on the past history and will help us to improve the progress of student, teacher as well as the organization. It will also help us to develop the curriculum based on student's interest.

II.   LITERATURE REVIEW

1. Predictive Analytics in a Higher Education Context.
Authors: J. Rajni and D. B. Malaya.
This paper proposes that, Predictive analytics is the process of forecasting future courses of action by analyzing historical and current facts. It's now a priority in many organizations because it can suggest the most favorable future planning by letting decision makers combine data about the four W's--that is, what, who, where, and when--to analyze why and how. It plays a vital role in higher education planning. Higher education plays an important role in a nation's socioeconomic development. As an educational management tool, predictive analytics can help to improve education quality by letting us analyze critical issues in education such as enrollment management and curriculum development. This article presents an analytical study of the prospects of predictive analytics in education planning, focusing on technical education. As a case study, the authors discuss an All India Council for Technical Education-sponsored project at Delhi Technical University, India.[1]

2. Predictive analytics using data mining technique:
Author: H. Gulati.
This paper presents analysis of data set using data mining algorithms. The outcome after analysis is considered to be the major factors that affect students dropping out of the open courses the most (dropout rate). Before applying classification algorithms some feature selection algorithms are also used so as

to get refined prediction results. Such analysis and prediction information will help college management and teachers to make necessary changes for imparting better education. Tool used for feature selection and mining is weka.[2]

3. Analytics within UAE higher education context.
Authors: A.Marks and M.Al-Ali
This paper presents the major findings from a study conducted with four different universities in the U.A.E regarding their use of the learning analytics (LA) capabilities available within their learning management systems (LMS). Data was collected from an online survey instrument, in-depth interviews with IT directors and academic administrators, and a case study in Zayed University. The study reveals that while the examined universities seem to recognize and appreciate the new analytics functions available within learning management systems, they seem to struggle with forming a consistent and effective approach to link those functions to improve the decision-making process.[3]

III.   PROPOSED MODEL

To build the predictive model following steps can be followed.
Step1: Data collection from user in a structured format.
Step2: The main working of the system i.e. analytics is performed on the data.
Step3: Generation of patterns.
Step 4: Decision making.
This system is to be used by an organization to improve the organizational success by analyzing student's marks, attendance and their performance records to find out the reasons for failure rate and the areas where teachers or students are lagging.
This system, only process the data and gives us conclusions but doesn't give any suggestion on it. So user need to find their own suggestion on the problem.

Proposed system will work in the following steps:
Step 1: Data Collection and its Preprocessing.
Student Dataset: This dataset contains attributes of student based on which we are going to analyse the data and generate patterns that can be used to improve the student's progress and

curriculum. The attributes present in the dataset may be unique_id, marks, subject_id, practical marks, internal marks, term work marks, etc. The data collected need to be preprocessed by finding the missing values and processing them.

Step 2: Data Analytics.

The proposed system uses python libraries such as pandas, numpy, plotly, scipy, sklearn, tkinter, etc. With the help of these libraries "Naïve Bayesian" algorithm is implemented for performing analysis and analytics on the data of engineering institute. Naïve Bayes Algorithm: This algorithm is used here for performing prediction using decision trees. Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In a machine learning classification problem, there are multiple features and classes, say, C_1, C_2, C_3,…C_k. The main aim of this algorithm is to calculate the conditional probability of an object with a feature vector X_1, X_2,…X_K belongs to a particular class C_i.

$$P\left(C_i \middle| x_1, x_2, \ldots, x_n\right) = \frac{P\left(x_1, x_2, \ldots, x_n \middle| C_i\right).P\left(C_i\right)}{P\left(x_1, x_2, \ldots, x_n\right)}$$
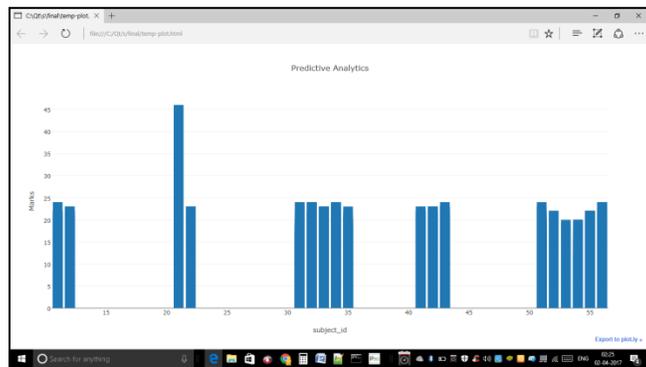
For $1 \leq i \leq k$

Figure.1. Output of Analysing Student data.

In the above figure, the Bar graph is showing the student's performance in each subject.

Figure.2. Output of Naïve Bayes algorithm.

In the above output, the probability of the student's result is calculated using Naïve bayes algorithm. For the above example attributes considered are Unique_id, subject_id, marks, year and the result is predicted.

IV. CONCLUSION

By using the proposed predictive model we can analyse student's performance based on student's marks and attendance. Here we will use the dataset in engineering institute, process it, analyse it and this will help Teachers as well as Administrators to find out the reasons behind low success rate of education system.. In this way the proposed predictive model will help for betterment of education system.

V. REFERENCES

[1]. Rajni and D. B. Malaya, "Predictive Analytics in a Higher Education Context," in *IT Professional*, vol. 17, no. 4, pp. 24-33, July-Aug. 2015. doi: 10.1109/ MITP.2015.68

[2]. H. Gulati, "Predictive analytics using data mining technique," *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2015, pp. 713-716.

[3]. Marks and M. Al-Ali, "Analytics within UAE higher education context," *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, 2016, pp. 1-6. doi: 10.1109/ICBDSC.2016.7460396