Higher Education Access Prediction System using Data Mining

Nikita Shirke[1], Priya Shinde[2], Shraddha Kelwalkar[3]

[1-3]Information Technology Engineering, VPPCOE & VA, Mumbai, Maharashtra, India
---***---

Abstract – *The development of educational means has become a priority for most member states and the rate regarding higher education presents a tendency to increase globally. This system can predict the stream of the student by considering the parameters like favorite score, area of interest, percentage obtained in SSC exam, average marks scored in Maths and Science subject and the score of aptitude test. The system's main objective is to offer a quick and easy way to appear the exam and it also provides results immediately after the exam. Multiple choice examination is conducted to provide a special advantage to the students that can't be found anywhere else. This software application is built to check objective answers in an online examination and allocate them to the user after verifying the answer. It can predict the streams like Science, Commerce and Arts based on the various parameters mentioned above.*

Key Words: Student Performance, exam, logistic regression, aptitude, SSC.

1. INTRODUCTION

Higher education access prediction is an area in which the stream for SSC pass-out students is predicted using some machine learning algorithms. The work done in this area includes the attributes that affect the growth of the students and data mining approaches that predict the outcomes. There is no full-proof system that can consider the various parameters like favorite score, area of interest, percentage obtained in SSC exam, average marks scored in Maths and Science subject and the score of aptitude test. We aim to build a system that can consider these parameters and can predict the results mode effectively.

Many research papers consider only the data mining approaches to predict the outcome. We are using a machine learning approach to predict the outcome. The algorithm we are using is well suited for this kind of prediction system. We are creating a dynamic system that can understand the student's opinions as well. Factors like favorite subject and area of interest are used to consider the interest of the student. We will be using an aptitude test of 20 marks to evaluate the basic knowledge of the student. The aptitude test consists of 7 Maths questions, 3 science questions, 5 logical reasoning questions and 5 verbal ability questions. The aptitude test score also helps in finalizing the stream of the student.

There are different sections present in this paper. Section 2 deals with the papers which we have studied to get an exact idea of the existing work done in the education prediction domain. In section 3 we have proposed our system. Section 4 deals with the implementation of the project. Section 5 concludes the paper.

2. LITERATURE SURVEY

We studied papers based on the area of our research i.e. Higher Education access prediction. We studied 5 research papers and drew some conclusions from the study.

In the first paper [1], the classification is employed in student information to predict the students' division on the premise of previous information. As there are several approaches that are units used for knowledge classification, the Naive theorem is employed here. Information like group action, class test, seminar, and assignment marks were collected from the students' previous information, to predict the performance at the top of the semester. The second paper [2] deals with the identification of the largest number of possible causes of attrition. Deep learning, classification algorithms, and data mining are used for the implementation.

The third paper [3] considers decision trees, classification, and data mining techniques. The aim of this paper is to enhance the performance of the student and capabilities. The data from several universities and private universities are considered. In the fourth paper [4], the focus of the author is on improving the accuracy of the model. The knowledge acquired by the usage of data mining techniques can be used to make a successful and effective decision. The fifth paper [5] focuses on course evaluation questionnaires. Data mining techniques are used in higher education more and more to give insights into educational and administrative problems in order to increase effectiveness.

This is the brief overview of the research papers studied.

3. PROPOSED SYSTEM

3.1. Model Architecture:

We are creating a model that can predict the stream of SSC students based on the historical data available.
We have created the data set by random values and the data set includes the columns like name of the student, school

---

name, pass-out year, percentage, average score, favorite subject, area of interest, aptitude score, stream.

Now, we preprocess the data. In the data preprocessing step, we have given numeric values to some of the string values like favorite subjects, area of interest and stream.
We have split our data into 70% training data and 30% testing data of the total dataset.

We are using a logistic regression model for the prediction. We will be taking the inputs from the users. The inputs include the name of the student, school name, pass-out year, percentage, average score, favorite subject, area of interest. Then we will be taking the user to the aptitude page where the user will give the aptitude consisting of sections like Maths, Logical reasoning, Science, English. The test will be of 20 marks.
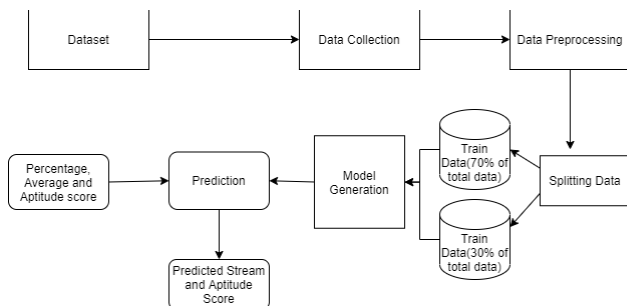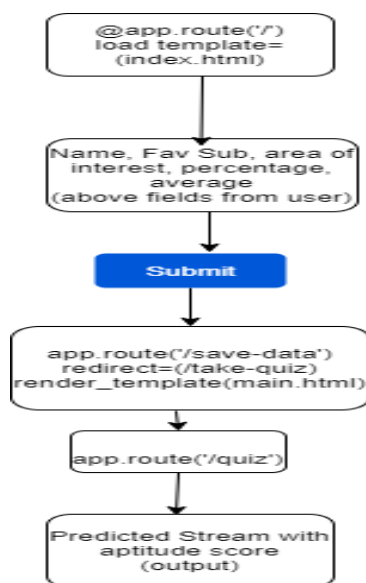
Below is the architecture for our proposed system.

Fig – 1: Proposed System Flowchart

3.2. UI Flow:

The UI flow will be as follows.

UI Flow
Fig – 2: UI Flow of the system

4. IMPLEMENTATION

4.1. Requirements:

Hardware requirements

Operating system- Windows 7,8,10
Processor- dual core 2.4 GHz (i5 or i7 series Intel processor or equivalent AMD)
RAM-4GB

Software Requirements

Python
Pycharm
PIP 2.7
Jupyter Notebook
Chrome

4.2. Model used and Result:

a) What is Data Mining?:

Data mining is the process to use the data available to predict some useful information. In data mining, we use past data to get some useful outcomes. In this project, we are following the data mining steps.

b) Logistic Regression:

Logistic Regression is one of the methods used in data mining. It is a machine learning algorithm. It is a supervised learning algorithm. Supervised learning means using the data available and the data of records. Using the records we can predict the outcome.
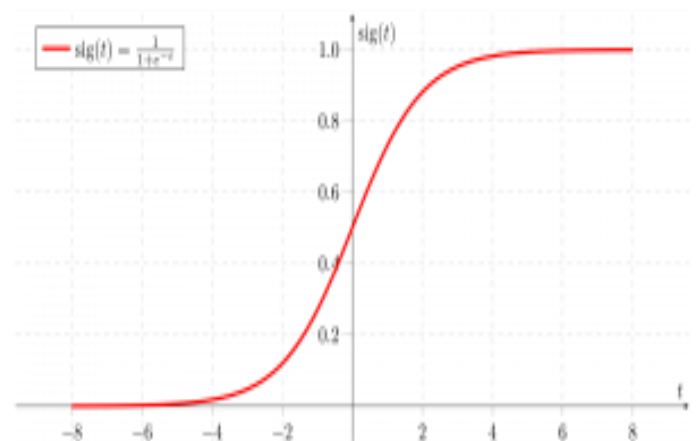

Fig – 3: Sigmoid Function for logistic regression (6).

The accuracy for the logistic regression algorithm used in our project is 73.19%.

c) Random Forest Classifier:

It is an ensemble technique. By using Random Forest algorithm we can perform classification tasks.
The accuracy for the random forest classifier used in our project is 67.04%.

5. CONCLUSION AND FUTURE SCOPE

We used machine learning algorithms for the prediction. We found that the logistic regression algorithm gives the highest accuracy with 73.19%. Since we are considering the percentages and average score of the Maths and the Science subject, we have an edge over other prediction methods used for the higher education access prediction. Our system can give the precise prediction of the stream to the user i.e. student. Even if the student is getting good marks in aptitude but not has a good percentage and the average score then the prediction will vary according to the percentages and average. So, this system can help a student in the stream prediction i.e. Science, Commerce, and Arts. This system can guide the user to take the correct decision.

In the future, we aim to make a system that can store the data of the students and admin. The system can have a login and signup included. We can look to improve the accuracies of the algorithms used. We can do feature scaling to improve the accuracies.

REFERENCES

1. IEEE Paper On Analysis Of Educational Data Mining Using classification by Chitra Jalota ,Rashmi Agarwal ,2019 .
2. IEEE Paper On Early Prediction of College Attrition using Data Mining by Luiz Carlos B. Martins , Rommel N. Carvalho, Ricardo S. Carvalho, Marcio C. Victorino , Maristela Holanda ,2017.
3. IEEE Paper On Predicting Instructor Performance using Data Mining Techniques in Higher Education, Mustafa Agaoglu ,2016.
4. IEEE Paper On Prediction for Performance Improvement of Graduate Students using Classification, Kamal Bunkar, Rajesh Bunkar, Umesh Kumar Singh, Bhupendra Pandya, 2012.
5. IEEE Paper On Prediction of Students Performance using Educational Data Mining Tismy Devasia, Vinushree T, Vinayak Hegde, 2016.
6. https://www.geeksforgeeks.org/understanding-logistic-regression/