Comprehensive Decision-Making Guide Predicting Colleges Based on User Profile Using Ensemble Machine Learning Model

Sopan Bhere[1], Sejal Dudhat[2]
AIML Department, SPPU University

Abstract: Making the right career decision is very importance for everyone. This paper aims to build a model that can help students to pick right colleges based on their profiles. For approximate prediction we plane build machine learning model in order to provide expected results. The datasets contain details about student (like Entrance Exam Score, 10th score, 12th score, cast etc.) and college (college Name, previous two-year cut-off etc.), for this various algorithm have been used i.e., Ensemble Machine learning, Linear Regression, Neural Network, Random Forest Decision Tree and Data Mining Technique for gathering the relevance data. This will assist student to know in advance if they get a chance to get recognized.
Keywords: Ensemble Machine Learning, Data Mining, Linear Regression, Neural Network, Random Forest

I. INTRODUCTION

In recent years, competition in the industry has increased exponentially. The increased competition has also led to an issue such as unemployment and high demand of newly emerging skills and technology in such competitive world, it becomes very mandatory for a student to secure admission and training from the well suitable institute. This would help the student to improve their skills as per the industry requirement. The admission process to the university after high school is a tedious process for the student and the parent. The universities mostly considered the high school percentages, and entrance percentage of that university as a criterion for the admission process This proposal deals with the two problems. Using the data available, to predict in which college a student has high chances of getting an admit, and also which stream. It also deals with providing relevant results by learning from previous system states and revising itself each time. In the system proposed, a person can enter their scores in the respective fields provided. The system then processes the data entered and produces an output of the list of colleges that a person could get into, with their scores. This is relatively quick and helps conserve time and money. In order to achieve this, we have proposed a novel method utilizing Machine Learning algorithms. To maximize the accuracy of our model, we have taken into consideration not one; but several machine learning algorithms. These algorithms include Neural Networks, Linear Regression, Decision Tree and Random Forest. More about these algorithms will be covered in the Algorithms section of this paper. These Algorithms are then compared and the algorithm which has the best key performance indicators will be used to develop the Prediction System. We also look forward to incorporate clustering of universities based on a profile and then classifying them as less likely, highly likely acceptance etc.

II. PROBLEM STATEMENT

Universities play an important role in the growth of the society, also they are the social and economic growth of the country. On the other hand, might be extremely challenging because pupils are unaware of the admission standards. They waste a lot of time and money. Admission into professional colleges for engineering degree course is based on scores of the Common Entrance Test (CET). Students are allotted colleges based on these scores. Seats are allotted on the basis of availability of seats in CAP rounds. The lowest score accepted in a college for a certain CAP round is known as the cut-off score. Universities under DTE collect data about CET scores and admissions from each college under that particular university. Analysing this extensive data provides us with an opportunity to predict the admission pattern for a particular score, branch and even a CAP round. Presently there are no such resources to sort out colleges based on the parameters of marks, branches and CAP rounds. Due to absence of such resources a student would be less informed regarding the colleges he is eligible in. Here we propose a technique to make use of Decision Support System to assist in providing a student with such decisions. The decisions taken by the system should not only focus on present decisions, but also should take past decisions into account. A great number of researches and studies have been done on graduation admission datasets using different types of machine learning algorithms. One impressive work by Sopan B. has compared between 5 different regression algorithms, which are: Linear Regression, Support Vector Regression, Multilayer

Perceptron, Decision Trees and Random Forest, to predict the chance of admit based on the best model that showed the least MSE which was multilinear regression. Sopan B. developed a model that studies the graduate admission process in Pune universities using machine learning techniques. The purpose of this study was to guide students in finding the best educational institution to apply for. Five machine learning models were built in this paper including SVM (Linear Kernel), AdaBoost, and Logistic classifiers. In this paper, Sejal Dudhat proposed the novel design for college admission hybrid recommender based on data mining techniques and knowledge discovery rules, for tackling college admissions prediction problems. This system consists of two cascade hybrid recommenders working together with the help of college predictor, for achieving high performance. The first recommender assigns students tracks for preparatory year students. While the second recommender assigns the specialized college for students who passed the preparatory year exams successfully. This predictor algorithm uses previous students' admission data of colleges GPA(grade point Average) for predicting most probable colleges. It looks over student academic merits, background, student records, and the college admission criteria. Then, predicts the possibility of university colleges that a student may enter. A great number of researches and studies have been done on graduation admission datasets using different types of machine learning algorithms. One impressive work by Sopan B. has compared between 5 different regression algorithms, which are: Linear Regression, Support Vector Regression, Multilayer Perceptron, Decision Trees and Random Forest, to predict the chance of admit based on the best model that showed the least MSE which was multilinear regression. Sopan B. developed a model that studies the graduate admission process in Pune universities using machine learning techniques. The purpose of this study was to guide students in finding the best educational institution to apply for. Five machine learning models were built in this paper including SVM (Linear Kernel), AdaBoost, and Logistic classifiers. In this paper, Sejal Dudhat proposed the novel design for college admission hybrid recommender based on data mining techniques and knowledge discovery rules, for tackling college admissions prediction problems. This system consists of two cascade hybrid recommenders working together with the help of college predictor, for achieving high performance. The first recommender assigns students tracks for preparatory year students. While the second recommender assigns the specialized college for students who passed the preparatory year exams successfully. This predictor algorithm uses previous students' admission data of colleges GPA(grade point Average) for predicting most probable colleges. It looks over student academic merits, background, student records, and the college admission criteria. Then, predicts the possibility of university colleges that a student may enter.

III. RESEARCH METHODOLOGY

A. Problem Understanding

Initially first we have to spend some time on what are the problems or concerns students having during their pre admission period and we should set the solutions to those problems as objectives of this research.

B. Data Collection

The primitive step to building a model for our use case is choosing the right dataset. For our predictions, we chose a dataset which contains all the important attributes that would affect the chances of admit. This is followed by data cleaning where we handle missing values present in various fields. Once the data is ready to be analyzed, we use various tools and libraries to visualize the data and perform analysis. This includes visualizing bar graphs and the correlation matrix.

C. Architecture
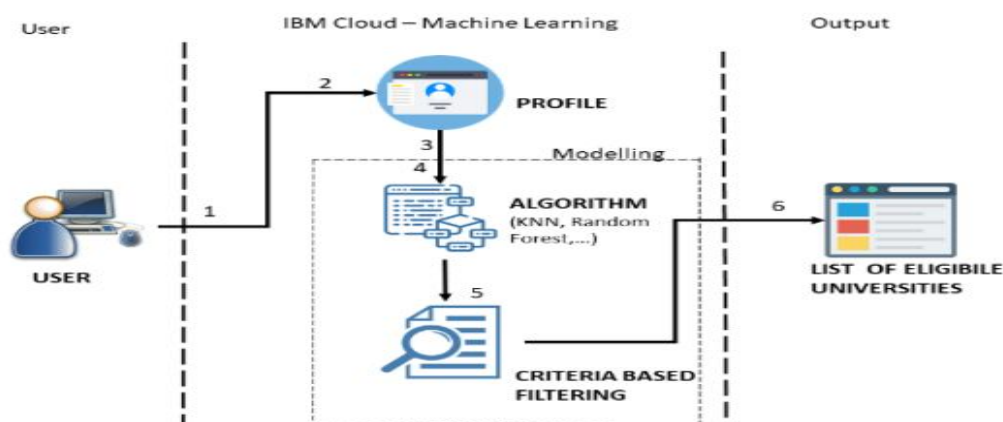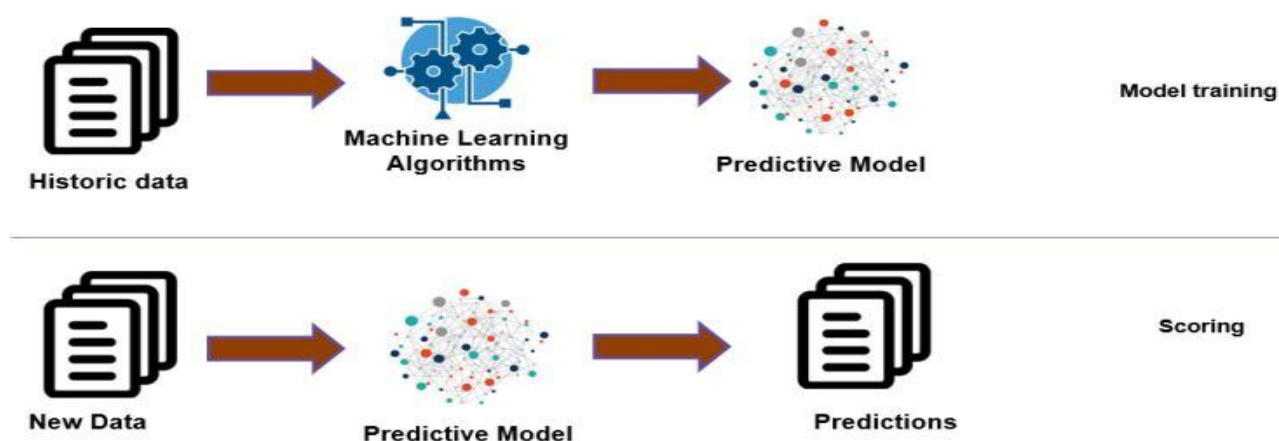
D. Data Preprocessing

Data should be cleaned that is removing the noise in the data and filling the missing values or extreme values and finalizing the attributes/factors which will have crucial importance in student admission process.

The greater part of this present reality information is:

1) Missing Data
2) Noisy Data
3) Inconsistent Data

The project includes application of machine learning algorithm on the data-set of cut-off lists of colleges for past 2/3 years. The data-set was collected manually from the state level engineering admission website. The data-set was preprocessed in the required format for application of machine learning algorithms. The pre-processing included converting data from '.pdf' to '.csv' format and segregation of data into attributes (year, category, score) and their respective class labels(admission).

Table 4.4: Descriptive csv datasets

Sr No.	College Names	Cutoff	cast	cet_mark	twelth_mark
1	Ramchandra Sanskriti Krida mandal's College of pharmacy,Jalna	70.94	open	80	78
2	Samarth College of pharmacy, Deulgaon, Raja	70.8	open	70	80
3	D. K. Patil Institute of Pharmacy ,Loha , Nanded	69.86	open	90	68
4	Dr. R.N. Lahoti Institute of Pharmaceutical college , Sultanpur	69.22	open	68	47
5	Mangaldeep Institute of pharmacy, Aurangabad	69.18	obc	47	76
6	Mauli college of pharmacy, Tondar, Udgir	68.78	sc	67	67
7	Godavari Institute of Pharmacy, Kolpa	66.98	sc	76	57
8	MDA scholl of pharmacy, Kolpa	66.79	nt	89	59
9	Institute of Pharmacy ,Badlapur	65.71	open	90	90
10	Rajashri Shahu college of pharmacy, Markhel	64.39	open	90	76

E. Data Visualization

After analysing the data, we will be able to know what the features and labels are, so from the above data, the label we have to consider is chance of getting college, and then we have to consider the parameters that influence or play a major role in predicting college.

F. Model Selection

Model selection is a cycle that can be applied both across various kinds of models and across models of a similar sort arranged with various model hyper parameters. several ML models have to be developed using various machine learning algorithms for admission to a particular university and the user interface has to be developed to access those models.

G. Train and Test Model

Training a model, we split the model into two segments which are 'training data' and 'testing data'. The classifier is training utilizing 'training data set', and afterward tests the performance of classifier on 'test data set'

H. Model Evaluation

Assessment is an integral of the model development process. It assists with tracking down the best model that addresses the information and how well the picked model will function later on. To work on the model hyper- boundaries of the model can be tuned and the exactness can be improved. Confusion matrix can be utilized to improve by expanding the quantity of genuine positives and genuine negatives.
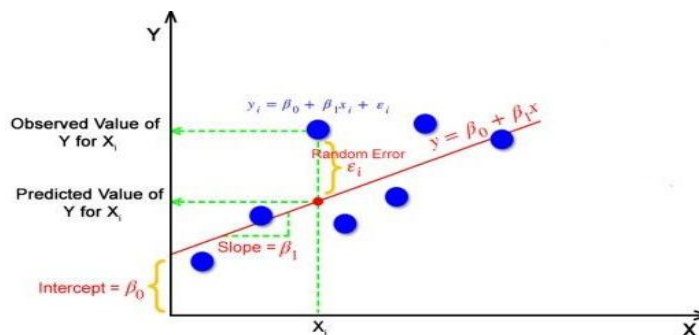
IV. ALGORITHM

A. Multilayer Perceptron Neural Network

Multilayer perceptron is a supervised deep artificial neural network machine-learning algorithm used to predict value of a dependent variable of a dataset according to weights and bias. Weights are updated continuously when finding any error in classification. The first layer is the input layer, more than one layer are presented next as hidden layers where each layer will contain a linear relationship between the previous layer, and the final layer is output layer that makes decisions and predicts. Forward pass and backward pass can be performed. Forward pass is where information flows from left to right, in other words, the flow will be input, hidden layers, and output in order. On the other hand, backward weights will adjust according to the gradient flow in that direction**.**
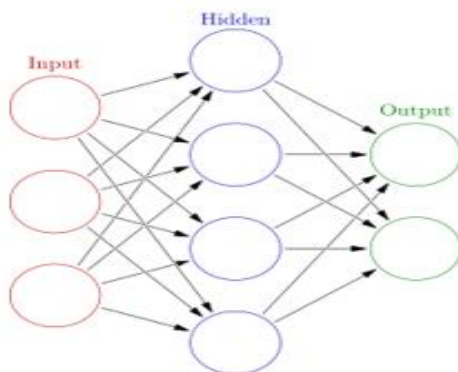

B. Linear Regression

It is an algorithm based on supervised learning of computers. It does the role of regression. Regression models a predictive goal value based on the independent variables. Mostly it is used to figure out the relation between variables and forecasting. Different regression models vary on the basis–the form of relationship between dependent and independent variables, are considered, and the number of independent variables used.

C. Decision Tree

A decision tree is a non-parametric Supervised learning algorithm which is utilized for both classification and regression task. Classification is a two-step process, learning and prediction. At first, model is developed based upon the given training data in its learning step. Then the model is used to predict response for the given data in prediction step. One of the most popular classification algorithms and easiest to learn and understand in decision tree. Decision tree algorithm is also used for solving classification and regression problems. Decision trees use a class label for predicting, for a record it.
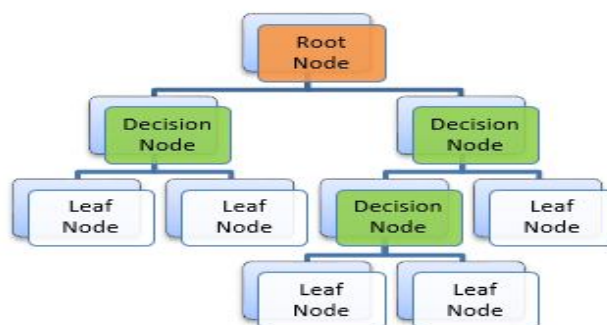
D. Random Forest

The random forest is a machine learning algorithm which is widely used in regression and classification problems. Decision trees are built upon multiple different samples and then take their majority vote for average and bifurcation in case of regression. Random forest has the ability to handle a data set which contains continuous variables in case of regression and categorical variables in case of classification. Hence, it provides good results for classification problems. In industry lingo, reason behind forest works algorithm works so well is: Any huge quantity of moderately uncorrelated trees working as a body will outperform any of the individual constituent models.
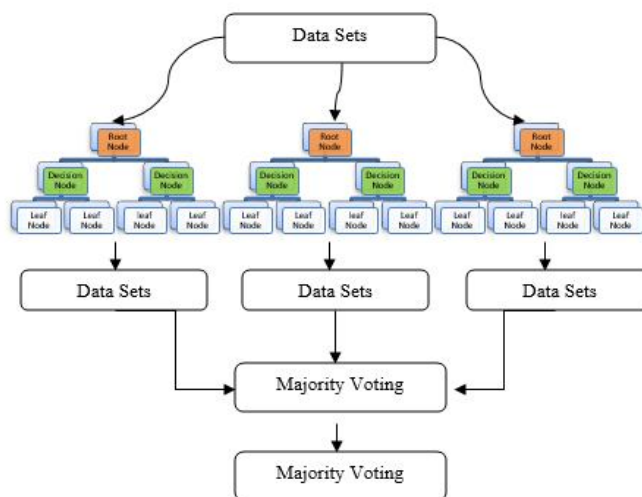
V. CONCLUSIONS

Every year millions of students apply to universities to begin their educational life. Most of them don't have proper resources, prior knowledge and are not cautious, which in turn creates a lot of problems as applying to the wrong university/college, which further wastes their time, money and energy. With the help of our project, we have tried to help out such students who are finding difficulty in finding the right university for them. It is very important that a candidate should apply to colleges that he/she has a good chance of getting into, instead of applying to colleges that they may never get into.

VI. ACKNOWLEDGMENT

REFERENCES

[1] Vandit Manish Jain, Rihaan Satia, "College Admission Prediction Using Ensemble Machine Learning Models", December 2021.

[2] Dr. Aruna Kumari B. N, Vishnu Sastry H K et.al, "An Automated Prediction Model for College Admission System", June 2021

[3] Md. Omaer Faruq Goni, Md. Abu Ismail Siddique et.al., "Graduate Admission Chance Prediction Using Deep Neural Network", September 2020.

[4] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Multi-split Optimized Bagging Ensemble Model Selection for Multi-class Educational Data Mining," Appl. Intell., vol. 50, pp. 4506–4528, 2020.

[5] M. S. Acharya, A. Armaan, and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," Kaggle, 2018.

[6] N. Gupta, A. Sawhney, and D. Roth, "Will i Get in? Modeling the Graduate Admission Process for American Universities," IEEE Int. Conf. Data Min. Work. ICDMW, vol. 0, pp. 631–638, 2016.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)