CAPSLG: College Admission Predictor and Smart List Generator

*Kiran Kumari, †Meet Kataria, ‡Viral Limbani and §Rahul Soni
*Assistant Professor, †‡§Student,
Department of Information Technology, K. J. Somaiya College of Engineering, Mumbai, India
Email: *kirankumarisinha@somaiya.edu, †meet.kk@somaiya.edu, ‡v.limbani@somaiya.edu, §soni.rv@somaiya.edu

Abstract—**Students face a lot of difficulties to secure an admission in the college of their choice. The current scenario of an engineering admission process is little complicated and not so easy in terms of selecting an appropriate college according to the scores and field of interest. Accurate choice, varying with the entrance exam result and academic scores, is very important to the candidates to fill in the application form. There are many colleges offering multiple engineering courses. So it becomes troublesome for students to organize and list-out the proper colleges of their choice for courses according to their performance score. The CAPSLG system consists of a smart list generator working together with the help of college predictor, to aid students in the admission process. The college admission predictor uses historical colleges cut-off students admission data for predicting the most probable colleges. The system analyzes student academic merits, background, and college admission criteria. Based on that, it predicts the likelihood of a university college that a student may enter. The smart list generator would enable the student to prepare the list of colleges, which could be needed to be filled in during the admission process. The system would also get feedback from the users, which would prove helpful for prediction evaluation and improving the performance factor.**

Index Terms—**College, Admission, Predictor, Seats, Feedback**

I. INTRODUCTION

In recent years, competition in the industry has increased exponentially. The increased competition has also led to issues such as unemployment and high demand for newly emerging skills. In such a competitive environment, it becomes very mandatory for a student to secure admission and training from the best suitable institute. This would help the students to improve their skills as per the requirements in the industry and secure an appropriate placement.

Thus, securing admission in the best college plays a very vital role in deciding the future of a particular student. Though the admission process for engineering courses has become easier than before, it still involves some risk factors that are hard out to figure out by the student. Admissions in engineering colleges in the state of Maharashtra or any state is based upon common entrance test (CET) and since more than 1.5 lakh seats are to be allotted in more than 200 engineering colleges and over 35 different branches of engineering, for students belonging to many categories like open, home university, outside home university, reserved category (SC, ST, OBC, etc.). Thus, the problem becomes more complex and students struggle to understand which colleges they are likely to get admitted in.

The candidates have to prepare a list of around 15-20 colleges best suitable for them, based on their academic scores and preference. But, the selection of colleges to be filled in the admission process should be well planned as a wrong choice would result in poor allotment of college. As pressure from the society and competition with the fellow candidates is more, students often end up choosing the wrong college preferences which place the student into under-qualified college.
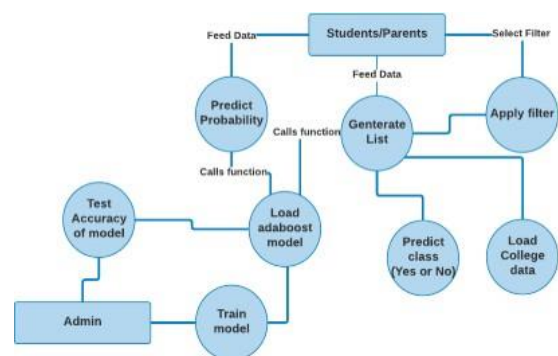

Fig. 1. Process Diagram

In order to prepare an efficient list, the candidates have to go through the online resources like websites or mobile applications of all the desired colleges individually. Then, view the cut-off for the past year and then select a list of colleges they are eligible for and shortlist them according to their preferences. To minimize the stress of students we came up with the idea of a computer-aided method which aims to automate this process and remove the risk holding factor of searching the number of eligible and best colleges within their vicinity. The users need to provide their academic score and the application predicts the best-suited college according to their preference, course, area and cut-off of colleges. The system also generates the list of colleges based on the above criteria which will help them to shortlist the desired college and fill the admission application more reliably.

Section II describes the literature survey for the project. Section III describes the methodology adopted for development of CAPSLG system. Section IV describes the result obtained from dataset analysis and snapshots of CAPSLG system.

Section V concludes this paper and Section VI elaborates the future work that can be done on the project.

II. RELATED WORK

A. Prediction of the Admission Lines of College Entrance Examination based on machine learning [1]

In this paper, Zhenru Wang and Yijie Shi compared two important machine learning algorithms to predict the college admission from the previous years' records. The paper shows different statistics based on the result of both algorithms. The first algorithm implemented by them was Random forests prediction algorithm. This model shows the accuracy of around 80% i.e. around 80 results out 100 predicted was correct. The second algorithm implemented was Adaboost. The model trained with Adaboost algorithm shows an accuracy of around 90% i.e. 90 accurate predictions from 100 results. The resources used for training of model and testing data are as follows: the system is windows 7, and python is used for the program. The conclusion of the paper suggests that the AdaBoost algorithm is much more efficient and reliable than Random forests while dealing with admission prediction process using machine learning.

B. Hybrid Recommender System for Predicting College Admission [2]

In this paper, Abdul Hamid M Ragab, Abdul Fatah S. Mashat and Ahmed M Khedra 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 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. In addition to the high prediction accuracy rate, flexibility in an advantage, as the system can predict suitable colleges that match the student's profile and the suitable track channels through which the students are advised to enter. The trust-ability is achieved since students responses positively increasing as long as they allocated to the most suitable college which satisfies their desire. The design is proposed only of Saudi Arabian Universities.

C. Shiksha.com [3]

This is a web-based application which provides the guidance and solutions for the educational queries. They developed a college predictor which takes the student data as an input and shows the possible college according to the colleges criteria. The application probably uses data mining algorithm with the previous years cutoffs as a training data for the model.

The website provides a list of colleges based on rank rather than scores. The filters are restricted to selection of branches and colleges, also the list generated by the website cannot be exported.

D. Analysis of Algorithms

TABLE I
ANALYSIS ON MULTIPLE DATASET

Dataset Name	Random Forest Score	Adaboost Score	Decision Tree Score
Breast Cancer	0.95213	0.97872	0.90957
Iris	0.92	0.96	0.64
Wine	0.98305	0.9322	0.49153

A comparative study of machine learning algorithms namely Decision Tree, Random Forest and Adaboost was performed on variety of database. The sample database included breast cancer, iris and wine, which were imported from the sklearn dataset API [4]. The comparison proved the Adaboost algorithm better than the rest of the two algorithms.

III. METHODOLOGY

The web-based application, Shiksha.com, is a propreitary software. The CAPSLG system is an open source software application hosted on GitHub. The system includes filers based on fee structure, review grades and location of the college, which aren't present in the former application. The CAPSLG system displays distance as well as travel duration to the college location and also lets the user export the list of colleges shortlisted through the application.

The project includes application of machine learning algorithm on the data-set of cut-off lists of colleges for past 3 years. The data-set was collected manually from the state-level engineering admission website [5]. The data-set was pre-processed 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, branch, category, score) and their respective class labels (admission).

The web-based application is built on a Python-based framework to render web pages called Django. The Django framework allows integration of python variables with the user interface. It follows an MVT (Model-View-Template) structure. The model includes the algorithms used for the implementation of the project. The template includes the HTML and CSS pages, which render the user-interface and its styling. The view controls the flow of the application. It maps the functionality with the interface.

With reference to the aforementioned paper and analytical study of multiple machine learning algorithms, it is concluded that the AdaBoost algorithm ensures higher accuracy as compared to the Random Forest and Decision Tree algorithm. Thus, the Ensemble AdaBoost Classifier from the scikit-learn library of Python is used for classification of the data. The model is trained for the individual colleges under the Mumbai

region. The model is prior trained and stored in pickle format. During the prediction, the stored pickle is loaded as a model and the same can be used for prediction. This improves the time efficiency of the CAPSLG system.

The AdaBoost algorithm is a binary classifier, which classifies the given input into two classes. It classifies the input data based on the classification from multiple decision trees. The majority of the classification obtained from the determined number of decision trees is the classification of the algorithm. This improves the accuracy of the AdaBoost algorithm. It also extracts the features which hold more weight in the prediction of the results.

The project also includes generation of a smart list, which can be useful for the student to fill in the application forms during the admission process. The smart list is generated based on the preferences opted by the student. The student has an access to a list of filters, which he/she can apply for a better list of colleges. The filters that the student can apply to the list of the colleges include fee structure, review grade and distance of the college from the student's location. The filters are implemented using Asynchronous JavaScript And XML (AJAX), thus the filters get applied to the list of colleges without redirection or refreshing of the web page.

IV. RESULTS AND DISCUSSION

The performance of the machine learning algorithms may vary for the database for the CAPSLG system. Hence, a comparative study of the algorithm was also performed on a sample of CAPSLG system's database, which included year, branch, category and cut-off scores as features. The following results were obtained from the study:

TABLE II
ANALYSIS ON CAPSLG SAMPLE DATASET

Learning Algorithm	Prediction Score
Decision Tree	0.9629629629629629
Random Forest	0.9703703703703703
AdaBoost Score	0.9777777777777777

Thus, the AdaBoost algorithm proved more efficient for the prediction from the database for the CAPSLG system. Following are the snapshots of the developed system:

Fig. 2. Individual college predictor

Fig. 3. User input for smart list

Fig. 4. Smart list and filters

Fig. 5. Exported data in PDF format

http://ssrn.com/link/2019-ICAST.html

V. CONCLUSION

The current scenario of the admission process in colleges for higher studies is a hectic process and requires a lot of analysis to filter out best suitable college. The student may not do the required analysis and, as a result, may secure admission in a less suitable college. This process needs to be eased where an application caters the need to perform the required analysis and output the best suitable college.

The web-application includes a user-friendly interface. It requires the user to fill in some of the mandatory details and provides the student with the probability of securing admission in the college. The student can also avail a list of colleges according to the filters applied, which can be used for application in the admission process. Thus, our application would prove helpful for the students to successfully secure admission in the best suitable and desired college in a hassle- free manner.

VI. FUTURE WORK

The project can be scaled in multiple dimensions for future work. Currently, the prediction of appropriate college selection is limited to Mumbai region, this can gradually be increased to cover entire Indian region, offering service to students of multiple streams, grades and entrance exams. The process of data feeding to the system is manual which can be automated, where the admin/web-master need not feed-in empirical data manually. Further research and advancements in data mining or machine learning can lead to replacement of AdaBoost Classifier with a better alternative. Thus, the accuracy of the system can be improved.

REFERENCES

[1] Zhenru Wang, Yijie Shi, "Prediction of the Admission Lines of College Entrance Examination based on Machine Learning", *2nd IEEE Interna- tional Conference on Computer and Communications*, 2016, pp. 332- 335.

[2] Abdul Hamid M. Ragab, Abdul Fatah S. Mashat, Ahmed M. Khedra, "Hybrid Recommender System for Predicting College Admission", *12th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2012, pp. 107-113.

[3] Higher Education in India — Shiksha

https://www.shiksha.com

[4] Sklearn Sample Datasets

https://scikit-learn.org/stable/modules/classes.html#module-sklearn.datasets

[5] Engineering Admission Cut-off

http://fe2018.mahacet.org/StaticPages/HomePage.aspx