International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 01 | Jan 2022 www.irjet.net p-ISSN: 2395-0072

University Recommendation Support System using ML Algorithms

Dipti Babel[1], Ashutosh Rathi[2], Sharvari Rodge[3], Saurabh Thorat[4], Prof.M.D.Salunke[5]

[1-5]Department of Computer Engineering, JSPMNTC (RSSOE), Pune, India, Savitribai Phule Pune University

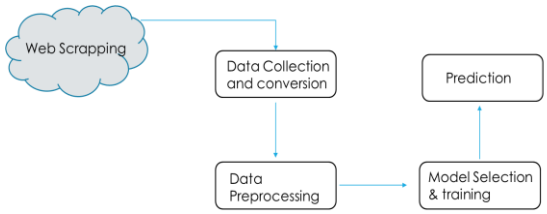
---

Abstract: For a prospective undergrad student, choosing which universities to apply to is a mystery. Students often wonder if their profile is good enough for a particular university. In this article, we have addressed this problem by modeling a recommendation system based on various classification algorithms. The required data was extracted from www.edulix.com and a dataset was created with profiles of students with admission/rejection from 45 different universities in the US.Based on this data set, various models were trained and a top 10 university list was proposed in order to maximize the chance of a student being accepted into this university list.

1.INTRODUCTION

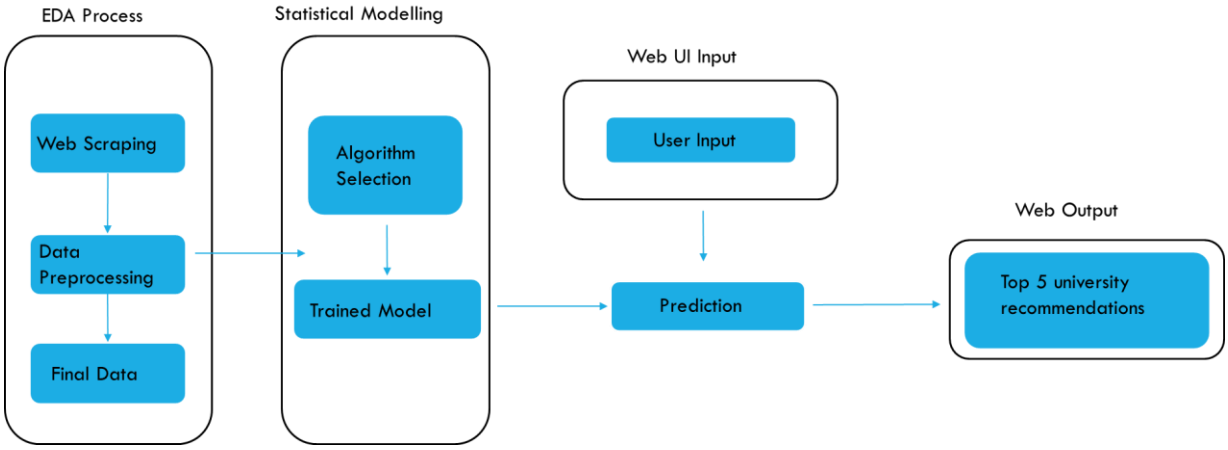
Number of students completing postgraduate studies abroad. The process for obtaining fully funded graduate study opportunities is very systematic and competitive. Many students apply to different universities in different countries. The universities offer admission to suitable candidates based on their academic profile, Test results, work experience and research.

But in this entire process, college selection is the most important step in applying for admission to college. Unique platform that can shortlist the universities/ colleges that attract applicants. The findings from the database of selected applicants are sufficient to find answers to questions such as: Which factors determine the funding opportunities for applicants at a particular graduate school? Student categories are usually assigned to the M.Sc. o Ph.D. from graduate school? What are the key factors required to receive graduate funding after selecting the right graduate school? Data mining techniques are very helpful in uncovering this kind of hidden knowledge about basic and complex data types. The main objective of this research is to design and develop a referral system for college entrance applicants that can help them select the graduate school that fits their full profile based on academic data from students who have already had the opportunity to apply abroad to study. The proposed system analyzes the data from these datasets, selects the data characteristics, executes the classification algorithm of the Support Vector Machine (SVM) and the machine learning algorithm K Nearest Neighbours (KNN) on it and suggests suitable universities to the applicants accordingly.



2. Need of Recommendation Systems

In today's fast-paced world, any technological innovation affects the importance of higher education, especially those that act as hubs for the latest research and trends. Since then, America has become one of the top travel destinations for any student around the world. If you want to undertake postgraduate studies in the United States of America, choosing a suitable university and admission are a challenge. no actual statistical relationships. From the student's point of view, the costs for the application and the commitment to the process are also high. In order to guide students efficiently, the college referral system was developed based on student contributions. Since the problem is extensive, a selected list of 45 universities has been considered for the sake of simplicity.



3. Literature Review

Much work has been done in the past on the use of data mining techniques in education. Few recommendation systems have been developed to suggest courses and colleges based on the student's academic performance. These systems used the decision tree classifier and fuzzy media clustering techniques using the WEKA tool. and it should help students choose a sequence that suits their skills. Another referral system has been developed to assist students in their academic path.in making decisions about course choice based on the student's schedule, order, and teachers. The model was trained on the

Basis of data from the last 7 years for a specific university and the classifiers for each subject were modeled on the basis of the cumulative GPA, some referral systems have been modeled to help the university get to know their students by tracking their time, extracurricular activities and accomplishments, and academic potential, and helping them identify and categorize students as needed using two algorithms and Kmeans . However, there was no access to all of the data sets that were used in the above-mentioned work. Although there are similarities with the subject covered in this document, it is not appropriate to directly compare the results with previous work as the data set used in this document is completely different.

4. Proposed Architecture

4.1. Data Set Collection

The first step in setting up a recommendation system is the identification of the data set. For this special problem, the academic details and background information provided in the application process represent the basic data. Classification model for the recommendation system, these data must be sorted with appropriate labels. These basic data for the application process are not readily available on the Internet for direct consumption. Although there were few forums that gave important information related to the scores, the most distinctive information was about student research. Interest and knowledge on a particular topic are unknown. However, this entire approach is based on making the most of the information available. Figure 2 shows the different number of admissions for each graduate college depending on the Bachelor universities. The University of Mumbai (1587), the National Institute of Technology (1467), Visvesvaraya Technological University (1426) and Anna University (1032) were found to be among the undergraduate universities with the most admissions Universities The 'Edulix' forum is one of the most popular forums for students aiming for postgraduate studies. This is the contact point for students who would like to take part in discussions and inquiries about all information about graduate studies. This forum essentially gathers the academic details of its users in order to compare their profile with previous experiences. From all of this data, some data such as the candidate's undergraduate university, CGPA, GRE and number of research publications, work experience, etc.

4.2. Data Scraping

Initially the list of forty five universities was narrowed down, and had enough information to be scrapped. Universities with skew data were born down. Then a crawler was engineered to induce the list of scholars and therefore the links to their profiles on Edulix. Once the distinctive set of students was identified, the data was scraped from every profile then the desired data was extracted from the hypertext markup language by mistreatment the python library 'Beautiful Soup'. The tabular structure of Edulix's net page, helped to spot the required data labels and points. The standard way of

accessing the desired parts by mistreatment the XPath failed to estimate for this case, as a result of the hypertext markup language was distorted in several cases.

4.3. Data Pre-processing

About 45000 samples of information were obtained as a result of scraping. Every sample corresponds to the profile of a student. The information points extracted enclosed GPA, college Boy University, GRE verbal score, GRE quantitative score, GRE analytical writing score, variety of journal publications, number of conference publications, trade expertise, analysis experience, situation experience and the following major. Cleansing the data of undergraduate universities had to be done, since this field was simply a text box and not a get field. Thus, input from totally different students created anomalies and this was corrected by trimming the string and removing areas found in them. The GRE scores (Verbal, Quantitative and AWA) were additionally clean since they contained immeasurable each previous and new versions of the examination. Equally the purpose average|GPA|standard|criterion|measure|touchstone} scores out there were supported totally different point systems, so all the GPA scores were uniformly scaled to four point scale. Also, bound categorical options just like the student's college boy university and department to which they apply were thought of as separate features. a complete of 1435 distinct undergraduate universities and fifty-three distinct majors were obtained when filtering and every of those were used as binary options.

4.4. Feature Extraction

The most necessary property of a feature is its correlation with the anticipated output. alpha analysis was done by plotting the feature values for 2 totally different universities and observing their variation. Variation of options CGPA and GRE for two different universities(Purdue and NJIT),has been shown in Figure three and Figure four respectively.

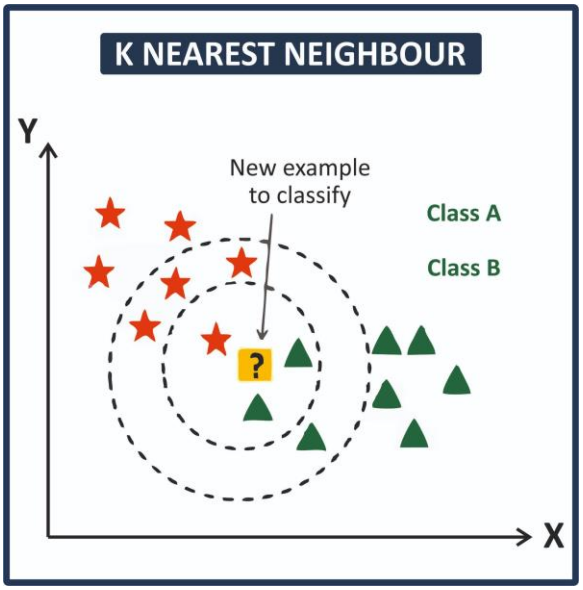
Initially, once all the features within the information set were thought-about the accuracy was relatively low(40%). The forward choice algorithm was accustomed to choose the simplest set of features for the model. Within the 1st iteration of the algorithm, the single best feature was known that best describes the variance within the data. Within the second iteration, the simplest feature was mounted and also the next best feature was found. This method was recurrent until the accuracy now not improved. Supported this method, collegian university, analysis experience, GRE, and grade point average were found to be the foremost effective options. once mistreatment the forward choice algorithm, the accuracy improved. Throughout this process, a scenario arose, when the accuracy didn't show any improvement, even if the best features were chosen. This was as a result of the numerical features like CGPA Associate in Nursing GRE scores were supported totally different scales, and then had an adverse implication on the model. However, once scaled from zero to one, there was a major improvement within the accuracy. Hence, all the numerical variables were then normalized to a scale of 0 to 1 by mistreatment the subsequent formula,

$$X = \frac{X - Xmin}{Xmax - Xmin}$$

wherever X is the price of any feature.

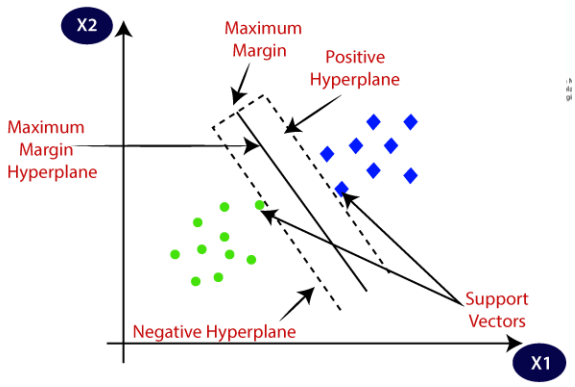
4.5. KNN

The K Nearest Neighbours algorithm is a nonparametric method of classification and regression. In the case of the KNN classification, the output is a class membership. An object is classified by the majority of its neighbors, and the object is assigned to the most common class among its k nearest neighbors (k is a positive integer, typically small) If k = 1 then the object is simply assigned to class des closest unique item Neighbor.



4.6. SVM

Support Vector Machine(SVM) could be a supervised machine learning rule used for each classification ANd regression. tho' we are saying regression issues additionally its best fitted to classification. The target of the SVM algorithm is to search out a hyperplane in an N-dimensional area that clearly classifies the information points. The dimension of the hyperplane depends upon the quantity of options. If the number of input features is two, then the hyperplane is simply a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes troublesome to imagine once the number of features exceeds three.



5. CONCLUSIONS

The best university may be suggested to the scholars as per their requirement. Will be Associate in Nursing intelligent recommendation system that helps students to see their eligibility for a University supported University admission criteria. It'll conjointly embrace several parameters like GRE, TOFEL score, University rank, budget, weather, and so on for recommending the right university.

A substantial variety of scholars are conferred with the chance to pursue instruction once the completion of their undergrad studies in countries completely different than their home countries. The records of those students that have with success gained admission can be constructive and worthy for alternative students hoping to achieve admission and facilitate them in their call making. Data processing and Machine Learning are the paradigms that may explore and supply exemplary results. Therefore the past records of fortunate graduate candidates hold utmost importance in choosing acceptable instruction institutes for graduate applicants who want to pursue higher studies. Conclusive of this research, we've planned a graduate university recommendation system that may apply SVM to classify a graduate university that's probably appealing to an applicant and also the KNN algorithm program to come up with a university with similar necessities and qualifications.

ACKNOWLEDGEMENT

We would like to thank Prof. Salunke M.D for his guidance and suggestions.

REFERENCES

[1] A Sabic, Md El-Zayat https://ieeexplore.ieee.org/document/8025933 2nd IEEE International conference in April, 2010

[2] M Hassan, Shibbir Ahmed, Deen Md.Abdullah https://ieeexplore.ieee.org/document/7434430 5th IEEE International conference, 2016

[3] Murtala Ismail, Usman Haruna, Garba Aliyu, Idris Abdulmumin https://ieeexplore.ieee.org/document/7760053 IEEE International conference, 2020

[4] Huma Samin, Tayyaba Azim https://ieeexplore.ieee.org/document/8693719 IEEE Access Volume: 7, 2019

[5] Can Ozturan, Suleyman Uslu, Mehmet Faith Uslu https://ieeexplore.ieee.org/document/7991812 10th IEEE International conference, 2016