Machine Learning Based – Graduate Admission Forecast

D.Rajeswararao [1], M Yellamma[2], Katady Veer Ratna Pavan[3], Goutham[4]

[1] Professor, Department of CSE , VR Siddhartha Enginnering college, Vijayawada, India.
Mail: rajeshpitam@gmail.com

[2] Assistant Professor, Department of CSE , Sreenidhi Institute of science and technology, India.
Mail:yellammam@sreenidhi.edu.in

[3]UG Scholars, Department of CSE, **SRM University , Tamil Nadu**, India.

Mail:kvrpavan10@gmail.com

[4]UG Scholars, Department of CSE, **Sreenidhi Institute of Science & Technology**, India.

Mail: bhagamsaigoutham10@gmail.com

ABSTRACT:

In response to the highly competitive job market at present times, an increased interest in graduate studies has arisen. This has not only burdened applicants but also led to an increased workload on admission faculty members of universities. Any chance of abridging the admission process impelled applicants and faculty workers to look for faster, efficient, and more accurate methods for predicting admissions All predictive methods proved to show accurate results, however; certain methods proved to be more promising than others were. Predictions were obtained within short time frames, which in turn will cut down the time in the admission process.

INTRODUCTION

The world markets are developing rapidly and continuously looking for the best knowledge and experience among people. Young workers who want to stand out in their jobs are always looking for higher degrees that can help them in improving their skills and knowledge. As a result, the number of students applying for graduate studies has increased in the last decade. This fact has motivated us to study the grades of students and the possibility of admission for master's programs that can help universities in predicting the possibility of accepting master's students submitting each year and provide the needed resources. The dataset presented in this paper is related to educational domain. Admission is a dataset with 500 rows that contains 7 different independent variables which are:

• Graduate Record Exam1 (GRE) score. The score will be out of 340 points.

• Test of English as a Foreigner Language2 (TOEFL) score, which will be out of 120 points.

• University Rating (Uni.Rating) that indicates the Bachelor University ranking among the other universities. The score will be out of 5.

• Statement of purpose (SOP) which is a document written to show the candidate's life, ambitious and the motivations for the chosen degree/ university. The score will be out of 5 points.

• Letter of Recommendation Strength (LOR) which verifies the candidate professional experience, builds credibility, boosts confidence and ensures your competency. The score is out of 5 points

• Undergraduate GPA (CGPA) out of 10

• Research Experience that can support the application, such as publishing research papers in conferences, working as research assistant with university professor

(either 0 or 1). One dependent variable can be predicted which is chance of admission, that is according to the input given will be ranging from 0 to 1. It is worth to mention that all tests will be done using R language. Models will be created using Weka, and statistical test will be performed using PHStat.

Specific preparation plays a crucial part in your life. Thus education preparation students often have multiple questions about universities which they can get admission and scholarship and accommodation. One of the main concerns is getting admitted to their dream university. It's seen that students still choose to obtain their education from universities that are known internationally. And when it comes to international graduates, the United States of America is the first preference of the majority of them. With most world-renowned colleges, Wide variety of courses available in each discipline, highly accredited education and teaching programs, student scholarships, are available for international students. According to estimates, there are more than 10 million international students enrolled in over 4200 universities and

colleges including both private and public across the United StatesMost number of students studying in America are from Asian countries like India, Pakistan, Srilanka, Japan and China. They are choosing not only America but also UK, Germany, Italy, Australia and Canada. The number of people pursuing higher studies in these countries are rapidly increasing. The background reason for the students going to abroad universities for Masters is the no. of job oppurtunities present are low and number of people for those jobs are very high in their respective countries. This inspires many students in their profession to pursue postgraduate studies. It is seen that there is quite a large number of students from universities in the USA pursuing Masters in the field of computer science, the emphasis of this research will be on these students. Many colleges in the U.S. follow similar requirements for student admission. Colleges take different factors into account, such as the ranking on aptitude assessment and academic record review. The command over the English language is calculated on the basis of their performance in the English skills test, such as TOEFL

and IELTS. The admission committee of universities takes the decision to approve or reject a specific candidate on the basis of the overall profile of the applicant application.

400 applicants have been surveyed as potential students for UCLA. The University weighs certain aspects of a student's education to determine their acceptance.

The objective is to explore what kind of data is provided, determine the most important factors that contribute to a student's chance of admission, and select the most accurate model to predict the probability of admission.

The dataset contains several parameters which are considered important during the application for Masters Programs. The parameters included are :

1. GRE Scores (out of 340)

2. IELS GPA (out of 10)

3. University Rating (out of 5)

4. Statement of Purpose and Letter of Recommendation Strength (out of 5)

5. Undergraduate GPA (out o f 10)

6. Research Experience (either 0 or 1)

7. Chance of Admit (ranging from 0 to 1)

In today's education world there are many number of students who want to pursue higher education after engineering or any graduate degree course. Higher education in the sense, some people want to do M.tech through GATE or through any educational institute entrance examination and some people want to do MBA through CAT or through any respective educational institute entrance examination and some people want to do Masters in abroad universities. we are focusing on only the students who want to pursue their higher education in abroad universities. Generally Higher education in abroad universities means we have many options like canada, USA ,UK Germany, Italy,

Australia etc. But we are focusing on only the students who want to do their Masters in America. Students who want to do masters in America have to write GRE (Graduate Records Examination) and TOEFL/IELTS (Test of English as a Foreign Language/International English Language Testing System). Once they have attended the exams they have to prepare their SOP(statement of purpose) and LOR(letter of

recommendation) which are one of the crucial factors they have to consider. These LOR and SOP plays a vital role if the student was looking for any scholarship. Then the students have to choose the universities they want to study or apply, we cannot apply to all the universities that will lead to lot of application fees. Here comes the problem that the student dont know to which university he might get admission. There are some online blogs which help in these matter but they are not that much accurate and dont consider all the

factors and there are some consultancy

offices which will take lot of our money and time and sometimes they will give some false information.so our goal is to develop a model which will tell the students their chance of admission into a respective university. This model should consider all the crucial factors which plays a vital role in student admission process and should have high accuracy. The model name is UAP. To access this model we will develop a simple user interface.

LITERATURE SURVEY

This section includes the literature review of previous research on the assessment of student enrolment opportunities in universities. Numerous programs and studies have been carried out on topics relating to university admission used many machine learning models which helps the students in the admission process to their desired universities. Previous research done in this area used Naive Bayes algorithm which will evaluate the success probability of student application

into a respective university but the main drawback is they didn't consider all the factors which will contribute in the student admission process like TOEFL/IELTS, SOP, LOR and under graduate score. Bayesian Networks Algorithm have been used to create a decision support network for evaluating the application submitted by foreign students of the university. This model was developed to forecast the progress of prospective students by comparing the score of students currently studying at university. The model thus predicted whether the aspiring student should be admitted to university on the basis of various scores of students. Since the comparisions are made only with students who got admission into the universities but not with students who got their admission rejected so this method will not be that much accurate.

is review of the literature covers 5 major areas within the graduate admissions process: (a) the decentralized nature of graduate

admissions; (b) the types of materials that are collected as part of the application process, including standardized test scores; (c) the

variety of admissions models that are used to make admissions decisions; (d) training and tools used by graduate admissions committees

to support the graduate admissions process; and (e) admitting diverse applicants. e literature review concludes with a discussion of

the ndings and the areas for future research.

Keywords Graduate admissions; admissions

is review of the literature covers 5 major areas within the graduate admissions process: (a) the decentralized nature of graduate

admissions; (b) the types of materials that are collected as part of the application process, including standardized test scores; (c) the

variety of admissions models that are used to make admissions decisions; (d) training and tools

used by graduate admissions committees

to support the graduate admissions process; and (e) admitting diverse applicants. e literature review concludes with a discussion of

the ndings and the areas for future research.

Keywords Graduate admissions; admissions

is review of the literature covers 5 major areas within the graduate admissions process: (a) the decentralized nature of graduate

admissions; (b) the types of materials that are collected as part of the application process, including standardized test scores; (c) the

variety of admissions models that are used to make admissions decisions; (d) training and tools used by graduate admissions committees

to support the graduate admissions process; and (e) admitting diverse applicants. e literature review concludes with a discussion of

the ndings and the areas for future research.

Keywords Graduate admissions; admissions p

This review of the literature covers 5 major areas within the graduate admission process.

- The Decentralized nature of graduate admisssions.

- The Types of materials that are used to make admission decisions.

- Training and Tools used by Graduate Admission Committees to support the Graduate Admission Process.

- Admitting diverse applicants.

[1] proposed a comparative approach by developing four machine learning regression models: linear regression, support vector machine, decision tree and random forest for predictive analytics of graduate admission chances. Then compute error functions for the developed models and compare their performances to select the best performing model out of these developed models the

linear regression is the best performing model with R2 score of 0.72. Janani Pet al.

[2] proposed a developed project uses machine learning technique specifically a decision tree algorithm based on the test attributes like GRE, TOEFL, CGPA, research papers etc. According to their scores the possibilities of chance of admit is calculated. The developed model has 93% accuracy. NavoneelChakrabartyet al.

[3] proposed a comparison of different regression models. The developed models are gradient boosting regress or and linear regression model. Gradient boosting regress or have to score of 0.84. That surpassing the performance of linear regression model. They computed different other performance error metrics like mean absolute error, mean square error, and root mean square error. ChithraApoorva et al.

[4] proposed different machine learning algorithms for predicting the chances of admission. The models are K-

Nearest Neighbor and Linear Regression, Ridge Regression, Random Forest. These are trained by features have a high impact on the probability of admission. Out of the generated models the linear regression model have 79% accuracy.

• One amazing work by Acharya et al. has looked at between 4 changed relapse Calculations which are: Linear Regression, Support Vector Regression, Decision Trees and Random Forest, to anticipate the opportunity of concede dependent on the best model that showed the least MSE which was multilinear relapse.

• Also, Chakraborty et al. thought about between both linear regression and gradient boosting regression in foreseeing possibility of concede; call attention to that gradient boosting regression showed better outcomes.

• Gupta et al. fostered a model that reviews the alumni affirmation measure in American colleges utilizing ML procedures.

The motivation behind this investigation was to direct understudies in tracking down the best instructive establishment to apply for. Five AI models were underlying this paper including Naïve Bayes, SVM (Linear Kernel), AdaBoost, and Logistic classifiers.

• Waters and Miikkulainen proposed an astounding article that aides in positioning graduation affirmation application as per the degree of acknowledgment and upgrades the presentation of inspecting applications utilizing measurable measurable AI.

• S. Sujay applied linear regression to anticipate the shot at conceding graduate understudies in expert's projects as a rate. Be that as it may, no more models were performed.
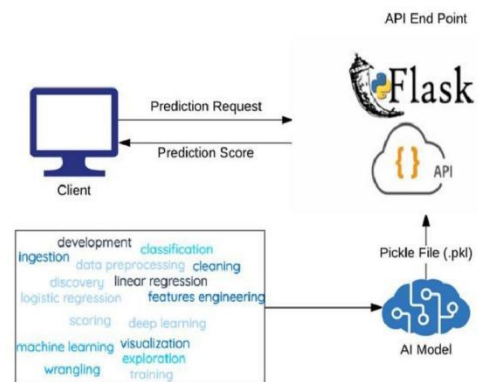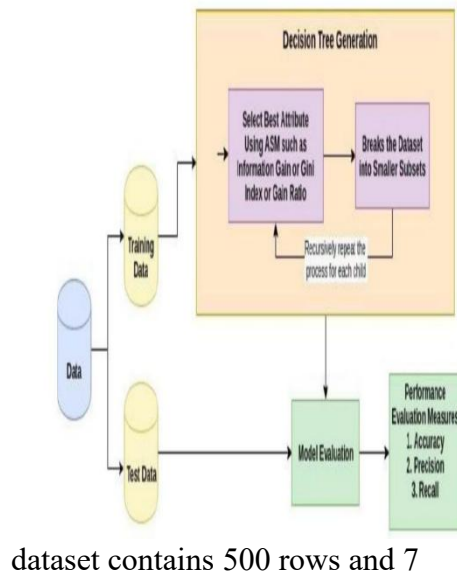
EXISTING SYSTEM

In the Existing System, many machine algorithms are used to the prediction of Graduate Admission. The existing system compares the four machine learning algorithms on the basis of accuracy. The algorithms are Linear

Regression, Support Vector Regression, Random forest Regression, Decision Tree Regression. In this system Linear Regression performs the best on the dataset with low MSE and high R2 score. Figure 2, shows sample data set, the

dataset contains 500 rows and 7

independent variables of data.

Fig 2.1: Existing System Architecture of Graduate Admission Forecast

PROPOSED SYSTEM

The proposed system consists of four regression models. Out of those we use Linear Regression using Dimensionality Reduction which is also a high accurate model. A user interface is provided

through which an actor can interact with the system. The algorithm with improved accuracy will act as a backend for the user interface. Whenever any actor (Student/Consultancy) provides the data to the user interface it

will show the result of Chance of Admission which is ranging 0 to 1

Fig : Proposed System of Graduate Admission Forecast

User Manual There are several steps for using this: 1. Initially the user has to open our website and provide all the requested values. 2. The user has to give his GRE score within the range of 270 to 340. 3. The user has to give his TOEFL score within the range of 100 to 120. 4. The user has to give his LOR number within the range of 1 to 5 5. The user has to give his CGPA score

within the range of 6.5 to 10 6. In the same way he/she has to select values within given ranges or given options. 7. All the given inputs are displays in screen and prediction is also displayed output is ranging from 0 to 1.

Benefits of the Proposed System

• High R2 score

• Low Root Mean Square Error (RMSE)

ARCHITECTURE

In this section, we will describe the architecture of the Student Admission Predictor system. The figure below explains the flow of the system:

• The student will enter his/her profile details using the user interface developed in shiny.

• The user interface code will interact with the KNN and Decision Tree models to provide the users with the required result.

• The KNN algorithm will be used to determine the chance of the student of securing

admission in a particular university based on his/her profile.

• The Decision Tree algorithm will be used to determine the rank of college to which is most suitable for the student based on his/her profile and provide the student with the list of universities which fall in that rank.

• Once the models have been executed the result will be provided to the student as the
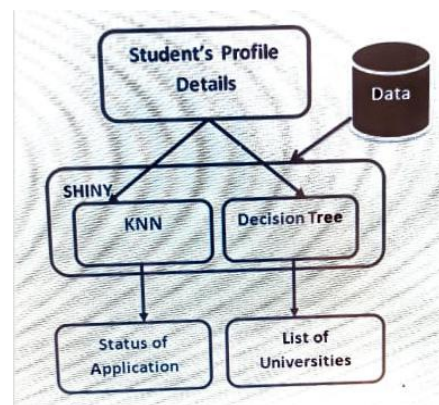
output on the user interface.

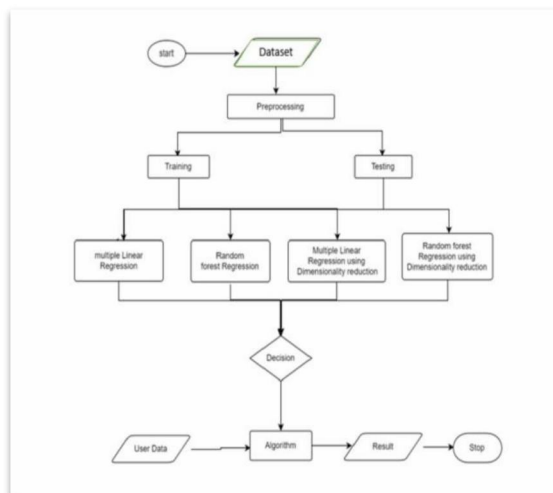Fig 4.2 : Architecture Of Graduate Admissions Forcast

FLOW CHART:

Fig . Flow Chart Graduate Admission Forecast

IMPLEMENTATION

Data-set

This section describes, in brief, the data that has been used for the research. Data from multiple sources was used in this project, the major amount of data was extracted from public website Yocket(Yocket), data regarding the rankings, fees and enrolment in colleges was obtained from a leading educational consultancy firm The Mentors Circle in India. Data from both the sources was integrated together to form a staging data-set. For predicting the chance of a student getting shortlisted in universities the final data-set was divided into multiple data-sets each representing a particular university. For predicting the list of universities suitable for students based on their profile data of all the students the staging data-set was updated only to have records of students who had successfully secured admission in the universities. Below table shows the different features of the data-sets.

Data-set Extraction and Transformation

Data related to the college ranking was collected in .csv format, the data related to students profile was extracted from (Yocket (2017)) using data extraction tool provided by (Mozenda (n.d.))in .csv files. Data being from public portal had multiple records with missing and irrelevant values; data cleaning was performed in Microsoft Excel by deleting the records having unwanted and missing values. Unwanted columns were removed from the data-set. Once the data-set was cleaned data was transformed to be suitable for the model. The

original data-set had TOEFL or IELTS score as a representation of language, to have a consistent metrics for the language score of all the records were converted to IELTS scale using the conversion table.(Prescholar (2017)). Similarly, the UGPA score of the students was represented in terms of percentage and

CGPA; all the records of percentage were converted to CGPA by multiplying percentage score by 9.5. The values of Intake, Status and Rank fields were changed as shown below to have numerical values for the data to achieve better results for KNN

Fig. Data Of Data-Set Extraction and Transformation

RESULT:

Fig.Project path of code

Rank	Value
Top 10-30	1
Top 40-60	2
Top 100	3

Status	Value
Accept	1
Reject	0

Intake	Value
Fall	0
Spring	1

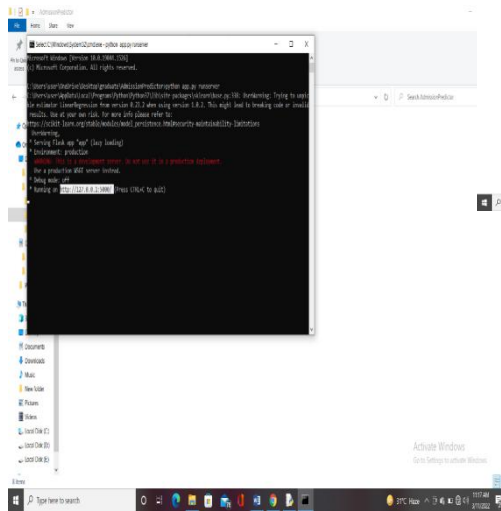This tells us in which folder does the project is present Screenshot Testing screen of code
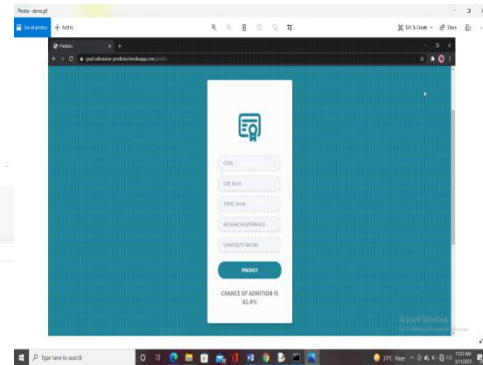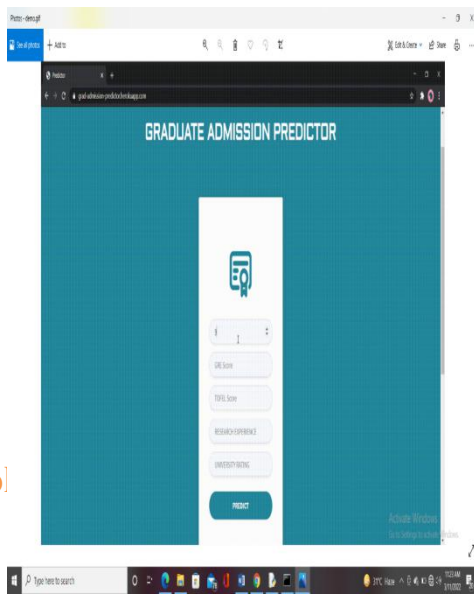

Fig. Testing screen of code

Fig.Output screen of code

This gives us the output of the project

This screenshot tell us about the execution part whether the code is running successfully or not.

CONCLUSION & FUTURE SCOPE

The main objective of this research was to develop a prototype of

the system that can be used by the students aspiring to pursue their education in the USA. Multiple machine learning algorithms were developed and used for this research. KNN proved to best-fit for development of the system when compared with the Logistic regression model. The model can be used by the students for evaluating their chances of getting shortlisted in a particular university with an average accuracy of 75%. Decision Tree algorithm was used to predict the universities which were best suitable for a student based on their profile. The decision tree algorithm proved to be 80% accurate. A simple user interface was developed to make the application interactive and easy to use for the users from the non-technical background. Shiny library from R was used to create the user interface. The overall objective of the research was achieved successfully as the system allow the students to save the extra amount of time and money that they would spend on education consultants and application fees for the universities where they have fewer chances of securing admission. Also, it will help the students to make better and faster decision regarding application to the universities. As discussed earlier in the limitation of the

research we have created the models based only on the data of Indian Students studying Masters in Computer Science in the USA, we have considered only ten universities with different rankings. In future, more data related to additional universities and courses can be added to the system. Also, the system can be enhanced to a web-based application by making changes to the Shiny code. Other classification algorithms can be evaluated to resolve the problem if they perform better than the current algorithm the system can be easily updated to support the new algorithm by changing the server code in the Shiny app.

REFERENCES

[1] Abdul Fatah S; M, A. H. (2012). Hybrid Recommender System for Predicting College Admission, pp. 107–113.

[2] Bibodi, J., Vadodaria, A., Rawat, A. and Patel, J. (n.d.). Admission Prediction System Using Machine Learning.

[3] Eberle, W., Simpson, E., Talbert, D., Roberts, L. and Pope, A. (n.d.). Using Machine Learning and Predictive Modeling to Assess Admission Policies and Standards.

[4] Jamison, J. (2017). Applying Machine Learning to Predict Davidson College ' s Admissions Yield, pp. 765–766.

[5] Mane, R. V. (2016). Predicting Student Admission decisions byAssociation Rule Mining with Pattern Growth Approach, pp. 202–207.

[6] MasterPortal (2017). MasterPortal.
URL:
http://www.mastersportal.eu/countries/82/united-states.html

[7] Mishra, S. and Sahoo, S. (2016). A Quality Based AutomatedAdmission System for Educational Domain, pp. 221–223.

[8] Mozenda (n.d.). Mozenda.
URL:
https://www.mozenda.com/

[9] Nandeshwar, A., Chaudhari, S., Sampath, V., Flagel, A., Figueroa, C., Sugrue, P., Ahlburg, D. and Mcpherson, M. (2014). Predicting Higher Education Enrollment in the United States : An Evaluation of Different Modeling Approaches, International Journal of Operational Research 19(26): 60–67. URL:

http://ezproxy.bethel.edu/login?url=http://inderscience.metapress.com/content/fr20314q1862653k/%5Cnhttp://nandeshwar.info/wp content/uploads/2008/11/DMW VU Project.pdf Prescholar (2017). Yocket.
URL:
https://www.prepscholar.com/toefl/blog/toefl-to-ielts-conversion

[10] Thi, N., Hien, N. and Haddawy, P. (2007). A Decision Support System for Evaluating International Student Applications, pp. 1–6.

[11] Times, T. E. (2015). The Economic Times.
URL:
https://economictimes.indiatimes.com/wealth/personal-financenews/around-3-lakh-indian-students-go-abroad-annually-33-take-property backedloans/articleshow/48703656.cms?intenttarget=no

[12] Waters, A. and Miikkulainen, R. (2013). G RADE: Machine Learning Support for Graduate Admissions. Yocket (2017)

[13] Yocket.
URL: https://yocket.in