SOMAIYA
V I D Y A V I H A R
K J Somaiya Institute of Engineering and Information Technology
An Autonomous Institute affiliated to University of Mumbai

University of Mumbai

A MINI PROJECT REPORT ON

"COLLEGE ADMISSION PREDICTION"

SUBMITTED BY

MIHIR PATEL, Roll No: 23
TIRTH PATEL, Roll No:24
HITESH PATIL, Roll No:25
PARTH PATIL, Roll No:26

GUIDED BY

PROF. Pankaj Deshmukh

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE
K. J. SOMAIYA INSTITUTE OF ENGINEERING & INFORMATION TECHNOLOGY
SION (E), MUMBAI 400022

Academic Year-2022-23

SOMAIYA
VIDYAVIHAR
K J Somaiya Institute of Engineering and Information Technology
An Autonomous Institute affiliated to University of Mumbai

CERTIFICATE

This is to certify that the project titled **COLLEGE ADMISSION PREDICTION** has been partially completed under our supervision and guidance, by the following students:

MIHIR PATEL
TIRTH PATEL
HITESH PATIL
PARTH PATIL

In the partial fulfillment of AI Mini Project of semester V in the Department of Artificial Intelligence & Data Science, during the academic year 2022-2023. The said work has been assessed and is found to be satisfactory.

(Internal Guide)		(External Examiner)
Prof. Dr. Milind Nemade (Head of Department)	College Seal	Dr. Suresh. Ukarande (Principal KJSIEIT)

SOMAIYA
V I D Y A V I H A R
K J Somaiya Institute of Engineering and Information Technology
An Autonomous Institute affiliated to University of Mumbai

ACKNOWLEDGEMENT

"I would like to acknowledge that college admission prediction is a multifaceted and intricate process that hinges on evaluating an applicant's potential for acceptance into an educational institution. This assessment takes into account a multitude of factors, encompassing academic achievements, standardized test scores, the rigor of coursework, involvement in extracurricular activities, letters of recommendation, personal statements or essays, and sometimes even interviews.

It is important to recognize that the criteria and methods used in predicting college admissions can vary significantly from one institution to another. Each college or university often has its own unique priorities and values when evaluating applicants, which can make predictions inherently complex. Moreover, the landscape of college admissions is not static; it evolves over time due to changing demographics, educational policies, and institutional goals.

In this process, historical data plays a vital role. Institutions often analyze the profiles of previously admitted students to develop predictive models. These models are designed to estimate the likelihood of admission for future applicants with similar profiles. However, it's crucial to remember that these predictions are not absolute guarantees, as various unpredictable factors can come into play during the admissions process.

Furthermore, many institutions practice a holistic admissions approach, considering not only academic metrics but also an applicant's unique qualities, experiences, and potential contributions to the campus community. This holistic review approach can make predictions challenging because it goes beyond quantitative data.

To obtain the most accurate and up-to-date information on college admission prediction for a specific institution, it is highly recommended to directly contact the admission office of that college or university. Additionally, referring to the institution's official website and admission guidelines can provide valuable insights into their specific criteria and processes. It is crucial to recognize that admission predictions are subject to change, and the final decision rests with the admission committee, which may prioritize various aspects of an applicant's profile in making their decision."

ABSTRACT

College admission prediction is a complex and dynamic process that involves assessing an applicant's probability of gaining acceptance into a higher education institution. This assessment considers a wide range of factors, including academic performance, standardized test scores, extracurricular activities, recommendation letters, personal statements, and, in some cases, interviews. The specific methods and criteria for predicting college admissions can vary significantly among institutions, making it a multifaceted endeavor.

Historical data plays a pivotal role in the prediction process, as institutions often analyze past admissions to develop predictive models. These models aim to estimate the likelihood of admission for future applicants with similar profiles. However, it is essential to understand that such predictions are not definitive guarantees, as unexpected elements can influence the admissions process.

Many colleges and universities adopt a holistic admissions approach, which evaluates applicants beyond quantitative metrics. This approach considers an applicant's unique qualities, experiences, and potential contributions to the campus community. Consequently, predicting admissions can be challenging, as it goes beyond numerical data.

To obtain the most accurate and current information on college admission prediction for a particular institution, individuals are encouraged to contact the institution's admission office and refer to official admission guidelines. Recognizing that admission predictions are subject to change and that the final decision rests with the admission committee is crucial for prospective students navigating the admissions process.

CONTENTS

Sr No.	Title	Page No.
1	Introduction	5-6
2	Literature Survey	6-7
3	Mathematical Modeling	8-10
4	Implementation	11-12
5	Results and Conclusion	12-14
6	References	14-15

INTRODUCTION

The world business sectors are growing quickly and constantly searching for generally advantageous information and experience among individuals. Youthful specialists who need to hang out in their positions are continually searching for Higher degrees that can help them in working on their abilities and information. Thus, the quantity of understudies applying for Graduate examinations has expanded in the last decade. One of the principal concerns is getting conceded to their fantasy University. It's seen that understudies actually decide to get their schooling from universities that are known Universally. What's more, with regards to international alumni, the United States of America is the primary inclination of most of them. With most incredibly famous universities, Wide assortment of courses accessible in each order, exceptionally authorized instruction and educating programs, understudy grants are accessible for international understudies.

As per gauges, there are in excess of 10 million international understudies enlisted in more than 4200. Universities and Colleges including both private and public across the United States. Generally, the number of understudies concentrating in America are from Asian nations like India, Pakistan, Sri Lanka, Japan and China. They are picking America as well as the UK, Germany, Italy, Australia and Canada. The quantity of individuals seeking after higher investigations in these nations are quickly expanding. The foundation justification for the understudies going to abroad Colleges for Masters is the quantity of open positions present are low and number of individuals for those positions are exceptionally high in their separate nations. This moves numerous understudies in their calling to seek after Postgraduate investigations. It is seen that there is a significant huge number of understudies from Universities in the USA seeking after Masters in the field of Computer Science, the emphasis of this exploration will be on these understudies. Numerous schools in the U.S. follow comparative prerequisites for understudy affirmation. Schools consider various variables, for example, the positioning on fitness appraisal and scholastic record audit. The order over the English language is determined based on their exhibition in the English abilities test, for example, TOEFL and IELTS. The entrance advisory board of universities takes the choice to endorse or reject a particular up-and-comer based on the general profile of the candidate application. The dataset taken in this undertaking is identified with the instructive area. Confirmation is a dataset with 400 lines that contains 7 distinct autonomous factors which are:

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

• Trial of English as a Foreigner Language (TOEFL) score, which will be out of 120 focuses.

• University Rating (Uni.Rating) that demonstrates the Bachelor University positioning among different colleges. The score will be out of 5.

• Statement of direction (SOP) which is a record written to show the applicant's life, drive and the inspirations for the picked degree/college. The score will be out of 5 focuses.

• Letter of Recommendation Strength (LOR) which confirms the applicant's proficient experience, fabricates validity, supports certainty and guarantees your ability. The score is out of 5 focuses.

• Undergraduate GPA (CGPA) out of 10.

• Research Experience that can uphold the application, like distributing research papers in gatherings, filling in as examination right hand with college teacher (either 0 or then again 1).

One ward variable can be anticipated which is the possibility of affirmation, that is as per the input given will be going from 0 to 1.

LITERATURE SURVEY

There have been a few ventures and studies performed on subjects connected with understudies induction into colleges. Numerous individuals utilized various AI models to make a framework that would assist the understudies with shortlisting the colleges appropriate for them likewise a subsequent model was made to assist the universities with settling on enrolment of the understudy. Naive Bayes calculation was utilized to foresee the probability of progress of an application, and numerous order calculations like Linear Regression, Random Backwoods, Naive Bayes calculations were thought about and assessed in view of their exactness to choose the best contender for the school.
Limit of this examination was that it did just depend on the GRE, TOEFL and Undergraduate Score of the understudy and missed on thinking about other significant variables like SOP and LOR archives quality, past work insight, specialized papers of the understudies and so forth. Various projects and studies have been completed on themes connecting with college affirmation utilized many AI models which makes a difference between the understudies in the confirmation cycle to their ideal colleges. Bayesian Networks Algorithms have been utilized to make a choice encouraging group of people for assessing the application put together by unfamiliar

understudies of the college. This model was created to estimate the advancement of forthcoming understudies by contrasting the score of understudies at present learning at college. The model consequently anticipated whether the hopeful understudy ought to be owned up to college based on different scores of understudies. Since the examinations are made exclusively with understudies who got confirmation into the colleges however not with understudies who got their affirmation dismissed so this strategy won't be just a lot precise. Past examination done in this space utilized Naive Bayes calculation which will assess the achievement Prediction for College Admission utilizing Machine Learning.

Past exploration done in this space utilized Naive Bayes calculation which will assess the achievement likelihood of the understudy application into an individual college however the fundamental downside is that they didn't consider every one of the elements which will contribute in the understudy affirmation process like TOEFL, GRE ,SOP, LOR and furthermore undergrad scores In the past exploration it is a very time taking cycle and complex collaboration which expects days to complete the attestation in abroad universities and it moreover asks costs for applying to the schools. So this University Admission Prediction application helps the students with vanquishing all of the issues as to process and besides can save money and time. The chance of acceptance into specific schools too shown rapidly in this application.

Abdul Fatah S (2012) fostered a model named "Cross breed Suggested System for Predicting College Admission" that can give the rundown of colleges which are best appropriate for an understudy in view of their scholastic records and school affirmation rules. The model was created by applying information mining strategies and information revelation rules to the previously existing confirmation expectation arrangement of the college. Mane (2016) directed a comparative exploration that anticipated the chance of an understudy getting confirmation in school in light of their Senior Secondary School, Higher Secondary School and Normal Entrance Examination scores utilizing the example development way to deal with affiliation rule mining[11]. The exhibition of both the models was great, the main disadvantage was the issue proclamation was single University-Centric.

Mishra and Sahoo (2016) directed research from a college perspective to foresee the probability of an understudy signing up for the college[2]. After the request about various courses in the college. They involved the K-Means calculation for bunching the understudies in light of various variables like input, family pay, family occupation, parent's capability, inspiration, and so forth to foresee on the off chance that the understudy will select at the college or not. Contingent on the closeness of the ascribes among the understudies, they were gathered into groups. What's more, choices were made. The target of the model was to increment the enlistment of the understudies in college.

MATHEMATICAL MODELLING

Mathematical modeling is a fundamental component of college admission prediction, enabling a systematic and data-driven approach to assess the likelihood of an applicant's acceptance into an educational institution. This section focuses on the development and application of mathematical or statistical models to make predictions about admission outcomes.

1. Variables and Factors:
    - Mathematical models used in college admission prediction consider a wide range of variables and factors. These typically include an applicant's academic performance (e.g., high school GPA), standardized test scores (e.g., SAT, ACT), extracurricular activities, recommendation letters, and other relevant information.

2. Model Selection:
    - Researchers and institutions select appropriate mathematical models based on the specific context and objectives of the prediction task. Common modeling approaches include logistic regression, decision trees, neural networks, and machine learning algorithms.

3. Data Collection and Preprocessing:
    - Data collection is a critical step, involving the gathering of historical admission data and applicant profiles. Preprocessing steps may include data cleaning, normalization, and feature engineering to ensure the quality and relevance of the data.

4. Equations and Algorithms:
    - The heart of the modeling process lies in the creation of equations or algorithms that leverage the collected data to make predictions. These equations represent the relationship between the predictor variables (e.g., test scores, GPA) and the outcome variable (admission status).

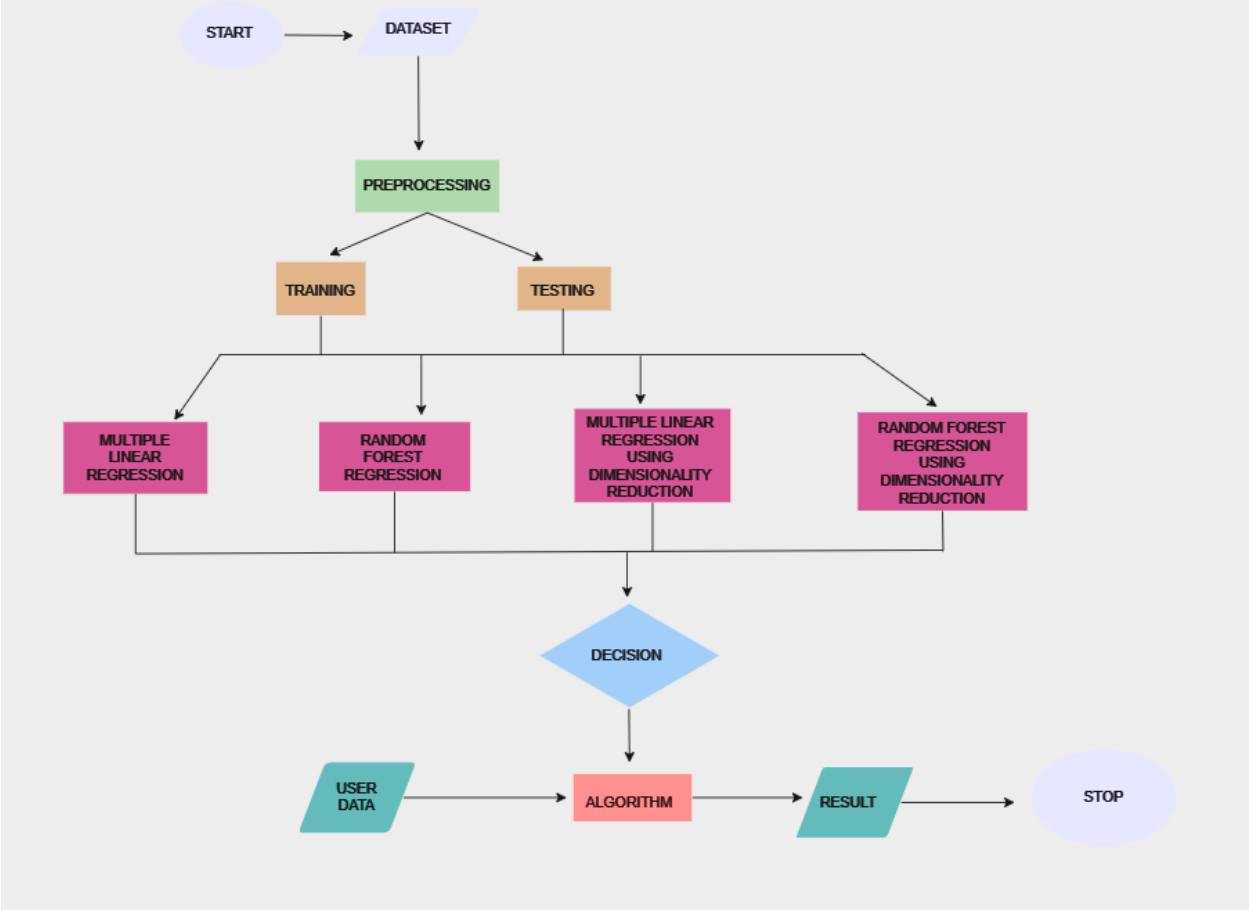
5. Training and Validation:
    - Mathematical models are typically trained and validated using historical data. This process involves splitting the data into training and testing sets

to assess the model's performance in predicting admission outcomes accurately.

6. Predictive Accuracy:
    - The effectiveness of a mathematical model in college admission prediction is often measured by its predictive accuracy, which indicates how well it can correctly predict whether an applicant will be admitted or not.

7. Assumptions and Limitations:
    - It's essential to be transparent about any assumptions made during modeling and acknowledge the limitations of the model. Factors like unpredictability in individual applications and changes in admission policies can affect prediction accuracy.

8. Continuous Improvement:
    - The field of college admission prediction is dynamic, with ongoing efforts to improve models and adapt to changing admission criteria and practices. Researchers often refine models based on real-world feedback and new data.


START
DATASET
PREPROCESSING
TRAINING
TESTING
MULTIPLE LINEAR REGRESSION
RANDOM FOREST REGRESSION
MULTIPLE LINEAR REGRESSION USING DIMENSIONALITY REDUCTION
RANDOM FOREST REGRESSION USING DIMENSIONALITY REDUCTION
DECISION
USER DATA
ALGORITHM
RESULT
STOP

IMPLEMENTATION

College Admission Prediction is planned in such a manner that it is extremely simple to apply and enlist. Effective framework ought to be conveyed and clients ought to have certainty that the framework would work proficiently and really. The more complex the framework being executed the more in question will be the framework examination and plan exertion expected for execution. Understudies are frequently stressed over their chances of affirmation in colleges which are abroad. The point of this model is to help understudies in shortlisting colleges with their profiles. The anticipated result gives them a fair thought regarding their confirmation chances in a specific college. This investigation ought to likewise help understudies who are right now planning or will get ready to get a superior thought in request to get their confirmation in the United States for aces. It is carried out so that the understudies need not stress about their time and cash, since it is an efficient and cash saving application. The significant exercises in the execution plan are cost assessment, plan and achievement assurance, project staffing, quality control designs, and controlling and observing plans.
As per the writing overview this model depiction gives the general data about its alleged parts, which is open in different manners. The remark location is performed utilizing different calculations in view of the precision of the models the best model is chosen.The models utilized here are

RANDOM FOREST CLASSIFIER

In a random forest classifier the choice trees are constructed. It is a troupe calculation which utilizes frail specialists and gathers them and becomes more grounded as entirety. Random forest, similar to its name infers, comprises countless individual choice trees and backwoods, lets out a class forecast and the class with the most votes becomes our model's prediction. The central idea driving irregular timberland is a straightforward however strong one — the insight of groups. Countless moderately uncorrelated models (trees) working as a board will beat any of the singular constituent models. The low connection between models is the key, uncorrelated models can deliver outfit forecasts that are more precise than any of the singular expectations. The explanation is that the trees shield each other from their singular blunders. While certain trees might be off-base, numerous different trees will be correct, so as a gathering the trees can move in the right bearing.

MULTIPLE LINEAR REGRESSION

Multiple linear regression (MLR), likewise referred to just as multiple regression, is a measurable strategy that utilizes a few illustrative factors to anticipate the result of a reaction

variable. The objective of numerous straight relapses is to display the direct connection between the illustrative (free) factors and reaction (subordinate) factors. Fundamentally, various relapses are the augmentation of ordinary least-squares (OLS) relapse since it includes more than one informative variable.

RESULT

The underneath figure shows that the Linear Regression has more accuracy than various models. In the figure the Linear Relapse shows 82% accuracy, Decision Tree shows 61% precision and the Random Forest Regressor shows 80% precision. We can wrap up with the Linear Regression model is useful and gives a prevalent result when diverged from various models.

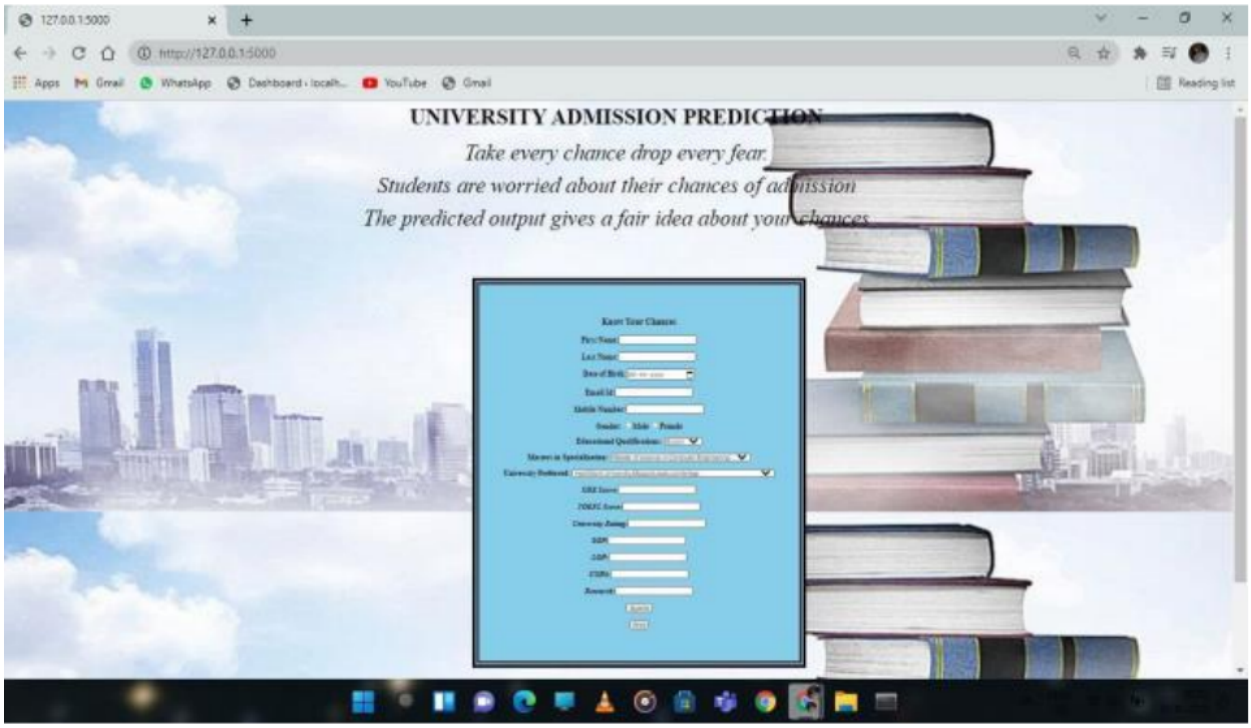


Figure 1: Web interface of University Admission Prediction

The figure shows the web point of interaction of the Project. Here the clients enter their test results to actually look at whether he/she can get down to that college or not, on the forecast rate.

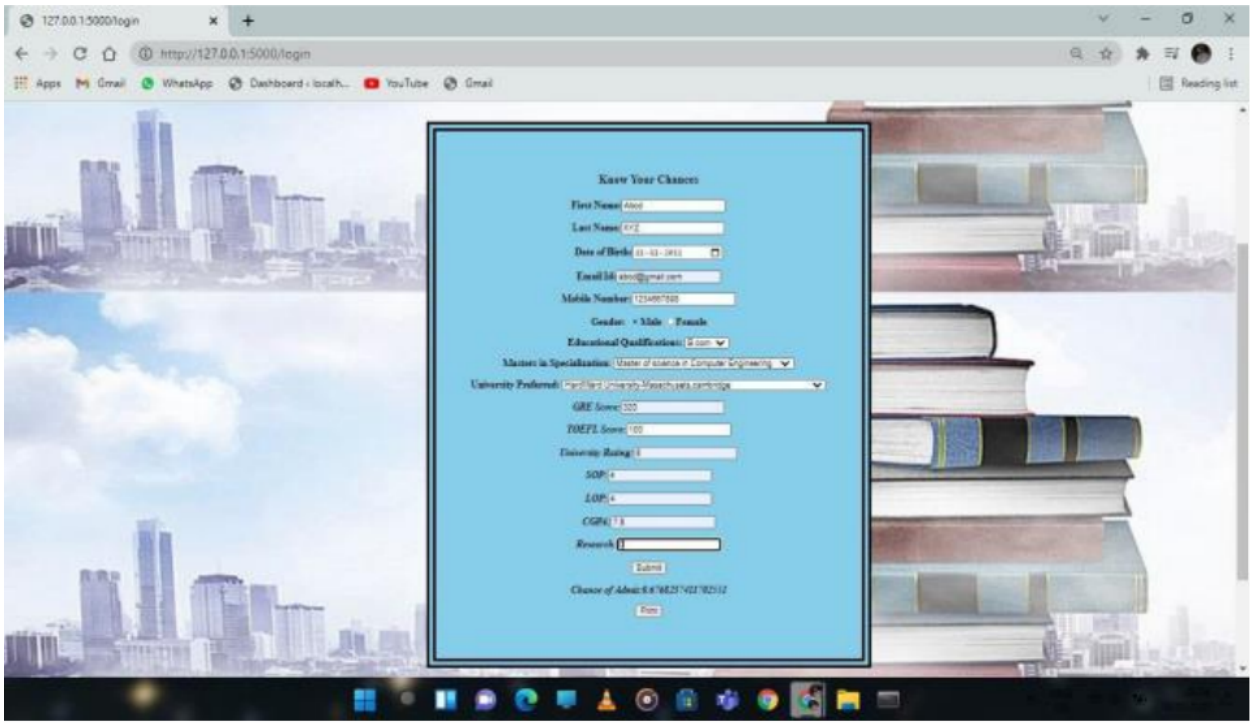


Figure 2: Output of University Admission Prediction

The above figure shows the web connection point of the venture. At the point when the client enters the subtleties followed by squeezing the anticipate button, it shows the level of Chance of Admit. By giving the legitimate data sources we got 62% of affirmation possibility to the college. The fundamental target of this model was to foster a model of the framework that can be utilized by the understudies trying to seek after their schooling in the USA. Different AI calculations were created and utilized for this exploration.

Direct Regression demonstrated to best-fit for improvement of the framework when contrasted and the Logistic relapse model. The model can be utilized by the understudies for assessing their chances of getting shortlisted in a specific college with a typical exactness of 75%. A straightforward UI was created to make the application intelligent and simple to utilize for the clients from the non-specialized foundation. Cup application was utilized to make the UI. The overall objective of the investigation was achieved successfully as the system grants the students to save an extra proportion of time, cash that they would spend on preparing experts, and application charges for the universities where they have fewer potential outcomes getting attestation. Additionally, it will help the understudies to settle on better and quicker choices in regards to application to the colleges.

CONCLUSION

Understudy confirmation issue is vital in instructive establishments. In this undertaking tends to AI models to foresee the opportunity of an understudy to be conceded. This will assist students with staying alert somewhat early expecting they get a potential chance to get recognized.AI models were performed to expect the chance of an understudy to get owned up to an expert's program. The AI models included numerous direct relapses, irregular timberland, Multiple Linear Regression with Backward Elimination and arbitrary woodland relapse within reverse end. Tests show that the Linear Regression model outperforms different models.
Our point is predicting the "Chance of Admit" considering the different limits that are given in the dataset. We will accomplish this point by utilizing the Linear Regression model. In light of the information that we have, we will part out information into preparing and testing sets. The Training set will have features and checks on which our model would be ready. The name here is the "Chance of Admit". In the event that you think from a no-specialized angle, name is essentially the yield that we need and elements are the boundaries that drive us towards the result. When our model is prepared, we will utilize the prepared model and run it on the test set and foresee the result. Then, we will differentiate the expected results and the veritable results that we want to see how our model performed. This entire course of preparing the model utilizing includes and known marks and later testing it to foresee the result is called Supervised Learning.

REFERENCES

1. S. Acharya, A. Armaan, and A. S. Antony, “A Comparison of Regression Models for Prediction of Graduate Admissions,” Kaggle, 2018

2. López-Martín, Y. Villuendas-Rey, M. Azzeh, A. Bou Nassif, and S. Banitaan, “Transformed k-nearest neighborhood output distance minimization for predicting the defect density of software projects,” J. Syst. Softw., vol. 167, p. 110592, Sep. 2020

3. S. Acharya, A. Armaan, and A. S. Antony, “A comparison of regression models for prediction of graduate admissions,” ICCIDS 2019 - 2nd Int. Conf. Comput. Intell. Data Sci. Proc., pp. 1–5, 2019.

4. Chakrabarty, S. Chowdhury, and S. Rana, “A Statistical Approach to Graduate Admissions’ Chance Prediction,” no. March, pp. 145–154, 2020.

5. Sujay, “Supervised Machine Learning Modelling & Analysis for Graduate Admission Prediction,” vol. 7, no. 4, pp. 5–7, 2020

6. N. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, “Systematic ensemble model selection approach for educational data mining,” Knowledge-Based Syst., vol. 200, p. 105992, Jul. 2020

7. Bruggink, T. H., and Gambhir, V. 1996. Statistical models for college admission and enrollment: A case study

8. for a selective liberal arts college. Research in Higher Education 37(2):221–240.

9. Deerwester, S.; Dumais, S.; Furnas, G.; Landauer, T.; and Harshman, R. 1990. Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6):391–407.

10. Koh, K.; Kim, S.; and Boyd, S. 2007. An interior-point method for large-scale l1-regularized logistic regression. Journal of Machine Learning Research 8(8):1519–1555.

11. Moore, J. S. 1998. An expert system approach to graduate school admission decisions and academic performance prediction. Omega 26(5):659–670.