PREDICTION PROBABILITY OF GETTING AN ADMISSION INTO A UNIVERSITY USING MACHINE LEARNING

Fawwaz Adib Khan[1], Syed Faraz Ahmed[2], Mohammed Faizan Khan[3], Mohammed Mateen Ahmmed[4]

[1,2,3]B.E. Student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad

[4] Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad

mateen.mohd@lords.ac.in

Abstract

In today's environment, students often struggle to find the right higher education institution that matches their profile. While there are advisory services and online applications available for university recommendations, they often charge high consultancy fees and their accuracy is questionable. Therefore, the objective of this research is to develop a model that accurately predicts the likelihood of admission into universities based on student profiles. This model will not only predict admission chances but also analyze how scores correlate with the probability of acceptance using historical data. The proposed model utilizes linear regression and random forest algorithms, with the CatBoost algorithm demonstrating the highest accuracy.

Keywords: Machine learning, Linear regression, CatBoost

I. Introduction

A person's education significantly shapes their future, underscoring the importance of quality education. Upon graduation, individuals often face uncertainties about pursuing higher studies and selecting the best university. Many Indian students particularly favor universities in the United States due to their global recognition. Despite India having reputable universities, graduates encounter challenges in securing admissions and job placements, given the limited employment opportunities. Seeking guidance, students invest considerable time and money in consultancy offices, advisors, blogs, and websites, although these resources are not always highly accurate or reliable.

Educational institutions, when processing admission inquiries, employ knowledge mining to extract pertinent details from the data they collect. This technique uncovers hidden insights that traditional queries and reports may miss, aiding in evaluating admission trends over years of student data.

This paper introduces a machine learning model that considers factors such as GRE scores, TOEFL scores, university rankings, statement of purpose, strength of recommendation letters, undergraduate GPA, and work experience to predict admission probabilities. The trained model provides robust statistical insights into admission likelihoods during testing scenarios, offering an unbiased assessment of admission probabilities.

The quality of one's education determines their destiny, hence it plays a significant part in one's life. After graduating, individuals frequently have a

number of questions in their minds about going back to school and selecting the finest university. Most students favour institutions of higher learning that are well-known worldwide. Therefore, a greater proportion of Indian students prefer to pursue higher education in the United States. Even though India is home to some well-regarded universities, graduate students sometimes struggle to gain admission to these institutions.

II. Literature Survey

1) A Comparison of Regression Models for Prediction of Graduate Admissions

Author: Acharya MS

Acharya MS conducted a study titled "A Comparison of Regression Models for Prediction of Graduate Admissions," where various models such as Linear Regression, SVM, and Random Forest were employed. The study focused on comparing their performance using error functions [1].

2) Modeling the Graduate Admission Process for American Universities.

Author: Narendhra Gupta

Narendhra Gupta explored "Will I Get in? Modeling the Graduate Admission Process for American Universities," treating admission prediction as a classification problem. Factors like GPA and other scores were considered, utilizing a large dataset for modeling purposes [2].

3) Automated Admission System for Educational Domain

Authors: Sushrutha Mitra, Soumya Sahoo

Sushrutha Mitra and Soumya Sahoo developed "A Quality Based Automated Admission System for Educational Domain," framing the admission prediction as a classification problem to forecast student admission into specific universities [4].

4) Artificial Intelligence in Education and Student Performance Evaluation

Author: Dr. N. M. Saravana Kumar

Dr. N. M. Saravana Kumar implemented Artificial Intelligence in "Imparting Education and Evaluating Student Performance," alongside other technologies. The study included score analysis versus admission chances and recommended universities based on profile similarity [5].

5) College Admission Predictor

Authors: Nagineni Dharani, Sathya Ragava

Nagineni Dharani and Sathya Ragava developed a "College Admission Predictor" that considers student details against cutoff scores to predict admission probabilities [5].

6) GRADE: Predicting Admission Likelihood Using Past Data

GRADE uses past admissions data to predict the likelihood of new applicants being accepted, assigning a score akin to human reviewers' assessments. It also identifies key applicant characteristics affecting predictions [6].

7) Predictive Value of Undergraduate Measures for Graduate Level Results

The study used bootstrapping to ascertain the predictive value of undergraduate-level variables for graduate performance, highlighting that undergraduate performance can explain 54% of the variation in graduate outcomes [7].

8) Binary Classification of Admission Acceptance

Various classifiers were evaluated based on metrics such as accuracy, precision, recall, F-measure, and area under the receiver operator curve to predict if students offered admission would accept it [8].

This compilation summarizes diverse approaches and methodologies used in predicting graduate

admissions, leveraging machine learning techniques and large datasets to enhance accuracy and decision-making in university admissions processes.

III.System Analysis

Upon graduation, students often face uncertainties when considering pursuing higher education and selecting the best university. Many students from India prefer universities in the United States due to their global recognition. However, despite reputable universities in India, graduates encounter challenges in gaining admission to highly ranked institutions and securing employment due to limited job opportunities. The current system relies on evaluating trends from admission forms collected over years to guide decision-making. Disadvantages of existing system:

1. Challenges in handling missing or inconsistent data.

2. Dependency on data mining techniques that may not yield precise predictions.

3. Algorithm: Classification and Regression Trees.

The major goal of this initiative is to assist students in reducing the amount of time and money they must spend at educational consulting companies. Additionally, it will assist students in reducing the number of applications they submit by suggesting universities where they stand the best chance of being admitted, saving them more money on application fees.

The proposed system employs machine learning techniques to predict the probability of admission into universities. The dataset is divided into 80% training and 20% testing sets, comprising 400 and 100 profiles respectively. Data preprocessing is crucial to clean and prepare the dataset for predictive algorithms. Special attention is given to handling missing data to ensure robust model performance. The developed model focuses on predicting the Chance of Admit and recommends universities where similar profiles have higher admission probabilities. Advantages of proposed system:

1. Accurate prediction of admission chances using machine learning algorithms like Linear Regression and Catboost.

2. Recommendations provided for universities based on profile similarity.

Algorithm: Machine learning, Linear regression, Catboost.

IV.System Study

The feasibility study of this project evaluates its viability and outlines a preliminary business proposal along with cost estimates. This phase ensures that the proposed system is practical and does not overburden the organization. Key considerations in this feasibility analysis include Economic Feasibility, Technical Feasibility, and Social Feasibility.

Economic Feasibility: This assessment examines the economic impact of implementing the system within the organization. It is crucial to justify the expenditure on research and development, considering the budget constraints. The proposed system is economically feasible as it leverages mostly freely available technologies, with only necessary customizations requiring additional investment.

Technical Feasibility: This study assesses the technical requirements of the system to ensure it does not strain available resources excessively. The system must have moderate technical demands, minimizing changes required for implementation. This approach avoids placing undue technical burdens on both the organization and its clients.

Social Feasibility: This aspect evaluates the acceptance of the system by its users. Effective user training is essential to ensure users perceive the system as beneficial rather than intimidating. User acceptance depends on the methods used to familiarize them with the system and build their confidence. Encouraging constructive feedback from users is crucial for refining the system to meet their needs effectively.

In summary, the feasibility study ensures that the proposed machine learning-based system for predicting admission probabilities into universities is economically viable, technically feasible, and socially acceptable, paving the way for its successful implementation and adoption.

V. System Design

The system records the information once the administrator adds the designated seats to a file. The admission allocation process moves more quickly, made simpler and takes less time overall. In which students can register with their personal and academic information to forecast college acceptance, and administrators can assign seats to students. Administrators can add batch and college information. Utilizing this Application made entrance seat allocation simpler and allowed for system implementation.

The entire admission allocation process took less time overall, and the allocation process moved more quickly. It aids students in picking the appropriate college.
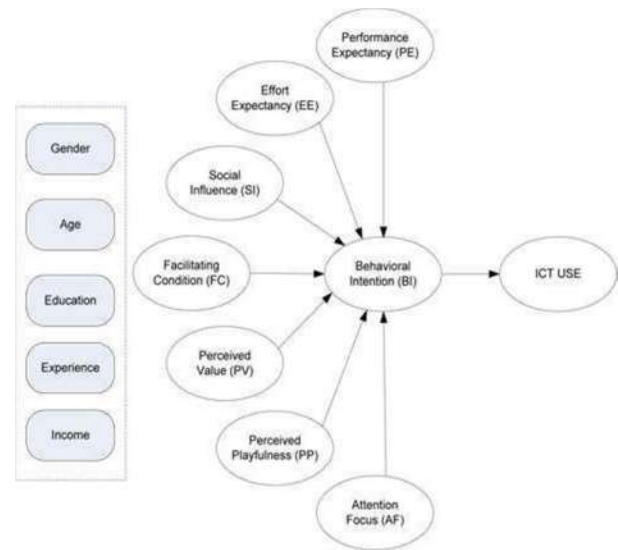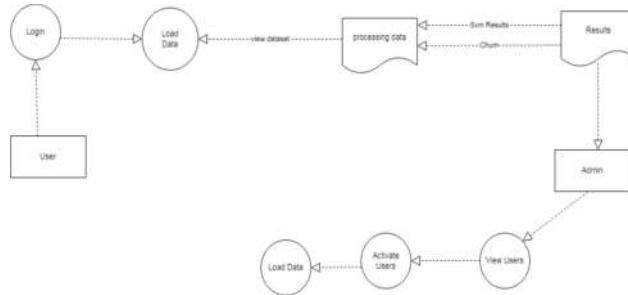

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

The data flow diagram (DFD) is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

AI allows for the simulation of human intelligence processes. Computer systems enable the learning, reasoning, and self-correction capabilities. In addition to Artificial Intelligence, other technologies are skillfully coupled to produce outstanding applications. To develop a customized teaching-

learning environment, we integrate the evolving role of Artificial Intelligence and its approaches into new educational paradigms. For a thorough learning and assessment process, features like recognition, pattern matching, decision making, reasoning, problem solving and so on are used besides knowledge-based systems and supervised machine learning for a complete learning and appraisal process.

Data flow diagram is also known as bubble chart. It may be used to represent a system at any level of abstraction. It may be partitioned into levels that represent increasing information flow and functional detail.

UML stands for Unified Modelling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artefacts of software system, as well as for business modelling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

goals:

The Primary goals in the design of the UML are as follows:

Provide users a ready-to-use, expressive visual modelling Language so that they can develop and exchange meaningful models.

Provide extendibility and specialization mechanisms to extend the core concepts.

Be independent of particular programming languages and development process.

Provide a formal basis for understanding the modelling language.

Encourage the growth of OO tools market.

Support higher level development concepts such as collaborations, frameworks, patterns and components.

Integrate best practices.

11

A use case diagram in the Unified Modelling Language (UML) is a type of behavioural diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

Class diagram:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.
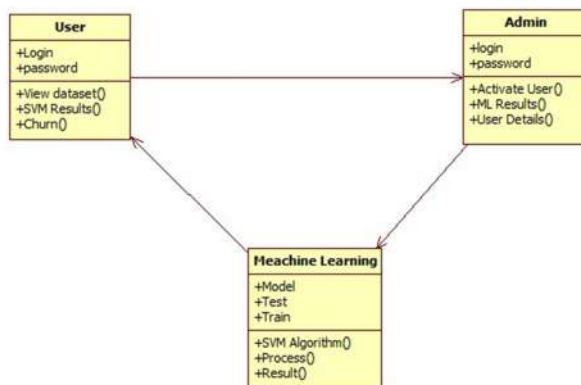

SEQUENCE DIAGRAM:

A sequence diagram in Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams
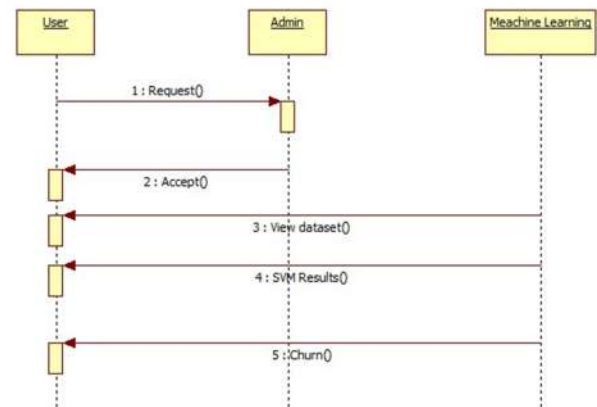
Activity diagram:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.
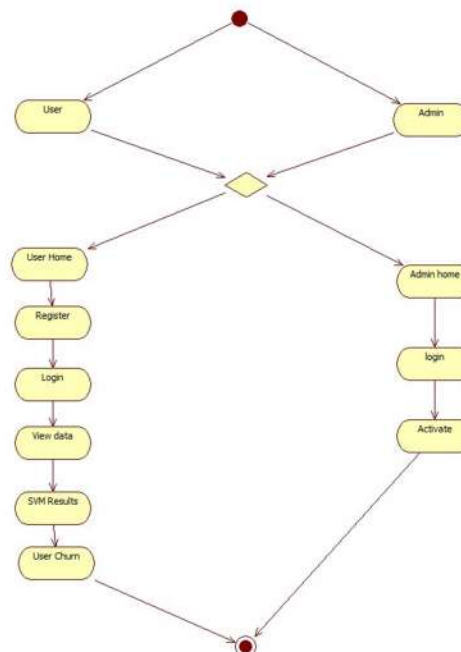
VI. Modules Description

User:

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the user. Once admin activated the user then user can login into our system. User can upload the dataset based on our dataset column matched. For algorithm execution data must be in float format. Here we took Three Customer Behaviour dataset for testing purpose. User can also add the new data for existing dataset based on our Django application. User can click the Classification in the web page so that the data calculated Accuracy and F1-Score, Recall, Precision based on the algorithms. User can click Prediction in the web page so that user can write the review after predict the review that will display results depends upon review like positive, negative or neutral.

Admin:

Admin can login with his login details. Admin can activate the registered users. Once he activate then only the user can login into our system. Admin can view the overall data in the browser. Admin can click the Results in the web page so calculated Accuracy and F1-Score, Precision, Recall based on the algorithms is displayed. All algorithms execution complete then admin can see the overall accuracy in web page.

Data Preprocessing:

A dataset can be viewed as a collection of data objects, which are often also called as a records, points, vectors, patterns, events, cases, samples, observations, or entities. Data objects are described by a number of features that capture the basic characteristics of an object, such as the mass of a physical object or the time at which an event occurred, etc. Features are often called as variables, characteristics, fields, attributes, or dimensions. The data preprocessing in this forecast uses techniques like removal of noise in the data, the expulsion of missing information, modifying default values if relevant and grouping of attributes for prediction at various levels.

Machine learning:

Based on the split criterion, the cleansed data is split into 60% training and 40% test, then the dataset is subjected to four machine learning classifiers such as Support Vector Machine (SVM). The accuracy, Precision, Recall, F1-Score of the classifiers was calculated and displayed in my results. The classifier which bags up the highest accuracy could be determined as the best classifier.

VII. Conclusion

Therefore, based on the aforementioned discussion, it can be concluded that regardless of the industry or sector, every organization must prioritize understanding and addressing customer churn. Customer retention, achieved through meeting customer demands effectively, is crucial for fostering loyalty. A robust model for predicting admission probabilities into universities will assist educational institutions in making informed decisions. 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 reccomendation) which are one of the crucial factors they have to consider. Leveraging advanced machine learning techniques, such as support vector machines, can significantly enhance the accuracy of admission prediction models. This report has emphasized the importance of customer retention strategies alongside the development of predictive models, particularly highlighting the application of support vector machines in optimizing admission prediction algorithms.

VIII. References

[1] Acharya MS, Armaan A, Antony AS (2019) A comparison of regression models for admissions. predictionof graduate

[2] Gupta N, Sawhney A, Roth D (2016) Will I get in? modeling the graduate admission process for American universities. In: 2016 IEEE 16th international conference on data mining workshops(ICDMW). IEEE

[3] Roa, Annam Mallikharjuna, et al. "College Admission Predictor." Journal of Network Commu-

nications and Emerging Technologies (JNCET), Volume 8, Issue 4, April (2018).

[4] Mishra, S. and Sahoo, S. (2016). A Quality Based AutomatedAdmission System for Educational Domain, pp. 221–223, International conference on Signal Processing, Communication, Power and Embedded System (SCOPES) 2016.

[5] Kumar, NM Saravana. "Implementation of artificial intelligence in imparting education and evaluatingstudent performance." Journal of Artificial Intelligence 1, no. 01 (2019): 1-9.

[6] Austin Waters and Risto Miikkulainen," Austin Waters and Risto Miikkulainen", Proceedings of the Twenty-Fifth Innovative Applications Conference,2013. of Artificial Intelligence Judith Zimmermann,

[7]Kay H. Brodersen, Kay H. Brodersen ,"A Model-Based Approach to Predicting Graduate-Level Performance Using Indicators of Undergraduate-Level Performance", Journal of Educational Data Mining, Volume 7, No 3, 2015.

[8] Kanadpriya Basu , Treena Basu ,Ron Buckmire and Nishu Lal ," Predictive Models of Student College Commitment Decisions Using Machine Learning", Data,2019.

[9] A.Sivasanagri,V.Shivani,Y.BindhuD.Deepa, R.Vignesh "prediction probability of getting an admission into university using machine learning" IEEE 2021.

[10] Rao, Annam M, et al. Modeling a "College Admission Predictor." Journal of Network Communications and Emerging Technologies (JNCET), Volume 8, Issue 4, April (2018).

[11] Mishra. S and Sahoo. S. (2016). A Quality Based Automated Admission Process System for Educational Domain, pp. 221–223, International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016.