iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50835

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

University Recommendation System for Abroad Studies using Machine Learning

Yash Kadam[1], Yash Kadulkar[2], Vishwas Moolya[3], Shruti Agrawal[4]
[1, 2, 3]*Students,* [4]*Asst. Prof., B.E. Information Technology, Vidyalankar Institute of Technology Mumbai, Maharashtra*

Abstract: ***Finding the right college or university to pursue postgraduate studies can be a daunting task, particularly when considering admission requirements and available academic re- sources. This can lead to students applying to institutions that are beyond their reach, wasting time and money on applications they are unlikely to be accepted to. This research study suggests creating a recommendation system to help graduate admission applicants choose the best university for them as a solution to this problem. The system leverages historical data of previously enrolled graduate students and considers several factors, including academic records, standardized test scores, academic standing, and university characteristics, to provide personalized recommendations to admission seekers.***

Keywords: Recommender System, Random Forests, Decision Trees, XGBoost

I. INTRODUCTION

Finding out which institutions or universities one has a decent chance of getting into based on one's academic record, including GPA, TOEFL, and GRE, can be challenging for those who are interested in pursuing postgraduate study. Sadly, some people make the error of applying to universities that are out of their financial grasp and that have strict scoring criteria, which could eventually hurt their future employment opportunities. Consequently, candidates should only submit applications to colleges and universities where they have a good chance of being accepted. Otherwise, they will just be wasting their time and money.

The dataset utilised for processing contains a variety of variables, including the university name, verbal and quantitative test scores, and GRE, TOEFL, or IELTS scores. Many colleges and graduate programmes use the GRE test, often known as the Graduate Record Examinations, as part of their admissions procedure. While applying to colleges, additional considerations are made in addition to standardised test scores, including letters of reference, declarations of purpose, extracurricular activities, and research papers.

It can be difficult to choose which schools to apply to base on one's academic record, scores on standardised tests like the GRE and TOEFL, and GPA. This is particularly valid if a person has completed their undergraduate degree and decides to pursue a postgraduate degree in their chosen field. Sadly, a lot of applicants could submit applications to universities that do not match their score requirements, which might be a waste of time and money. Using standardised test results to apply to several colleges might significantly raise the expense of the application procedure. Unfortunately, there aren't many effectivesolutions out there to aid this issue, which is why this system was created.

The research study's main objective is developing a system of recommendations that will help graduate admission applicants choose the best university for them. To do this, the system will examine the historical data of previously enrolled graduate students and use pertinent data to offer individualised suggestions to applicants. The academic history, academic standing, and test scores will all be considered by this recommendation mechanism. By utilising this data, the system will offer insightful information that admission hopefuls can use to decide which university will effectively fulfill their educational needs and objectives. Overall, the recommendation system will be a useful tool for people looking to improve their educational and career options by pursuing a graduate degree.

II. PROBLEM STATEMENT

The process of finding a suitable college for further studies has always been a challenge for students. Despite the avail-ability of various college prediction apps and websites, the lack of articulate information regarding colleges and the time consumed in searching for the best deserving college often make the process tedious for students. Additionally, students may not have access to all the information necessary to make informed decisions about their college choices, leading to stressful research and the need for counsellors.[2] The aim of this research paper is to design and develop a college prediction system that provides a probabilistic insight into college administration for overall rating, cut-offs of the colleges, admission intake, and preferences of students.

©IJRASET: All Rights are Reserved

The system aims to streamline the college admissions process by helping students to avoid spending time and money on counsellors and stressful research related to finding a suitable college.[2]

III. LITERATURE REVIEW

Aishwarya Nalawade and Bhavana Tiple are the authors of the study article "The University Recommendation System for Higher Education." A recommendation system created especially for higher education environments is presented in this article. The system, which seeks to give students indi- visualized suggestions for course selection, job development, and other academic activities, is discussed by the authors along with its creation and evaluation. The study concentrates on utilising machine learning and data analytics methods to improve students' educational experiences in higher education schools. The findings of this research could have a big impact on how well recommendations are granted to higher education students environments, which would eventually help those students succeed academically [1].

Alcina Judy, Kesha D'cruz, Janhavi Kathe, and Kirti Mot- wani wrote a study article titled "Recommendation System for Higher Studies using Machine Learning." The article provides a suggestion system that offers tailored advice for graduate studies by utilising machine learning techniques. The authors put forth a novel strategy that generates pertinent suggestions for appropriate higher education programs, courses, and universities while incorporating a student's scholastic experience, job objectives, hobbies, and preferences. The creation and assessment of the system, including the choice and use of machine learning algorithms, data preprocessing, and model evaluation approaches, are also covered in the article. The study's goal is to help students choose their higher education choices wisely, eventually increasing their chances of succeeding in professional success. By utilising machine learning to enhance the quality and relevance of recommendations given to people and assisting them in making the best decisions for their higher studies, the results of this article may have significant ramifications for the field of higher education [3].

Vandit Manish Jain and Rihaan Satia wrote a study titled "College Admission Prediction Using Ensemble Machine Learning Models." In order to forecast whether students will be admitted to college, the paper provides a prediction model that makes use of ensemble machine learning algorithms. The writers present a process for predicting the probability of admittance to a specific college or institution that takes into account a number of variables, including scholastic history, standardised test results, extracurricular activities, and personal writings. The creation and assessment of the model, as well as
the choice and use of ensemble machine learning algorithms, feature selection strategies, and model evaluation metrics, are also covered in the article.

By making precise predictions of students' chances of admission to particular colleges or universities, the study hopes to help students make educated choices about their college application approach. The findings of this research have significant ramifications for students, advisors, and admissions officials because they shed light on the college admissions procedure and assist students in deciding what is best for their future scholastic and professional aspirations.

Overall, "College Admission Prediction using Ensemble Machine Learning Models" makes a significant addition to the fields of machine learning and higher education by enhancing the precision and applicability of projections relating to college entry by utilising the power of ensemble algorithms [2].

The research paper titled "Recommendation System for Higher Studies at Abroad via Machine Learning Techniques" is authored by Prof. Pooja Bhatt, Miss. Manali Shah, and Miss. Priyanshee Soni. The paper proposes a recommendation system that uses methods for machine learning to provide personalized recommendations for students who want to pur- sue higher studies abroad. The system takes into consideration different factors such as the student's academic background, preferences, and goals to generate a list of recommended universities and programs. The suggested method can aid students in finding the most suitable schools and programmes while also saving them time and effort, increasing their possibilities of success. This system can also give colleges and education advisors insights into student tastes and requirements, assisting them in providing pertinent services.

Overall, the paper makes a significant addition to the fields of higher education and machine learning by offering a workable answer to the problems that students and colleges encounter when trying to choose the best course of study and school [6].

A thorough review of the issues and solutions in recommender systems is given in the research article "Recommender Systems Challenges and Solutions Survey" by Marwa Hussien Mohamed, Mohamed Helmy Khafagy, and Mohamed Hasan Ibrahim. The authors analyse several problems that these systems confront, including data quality, scalability, diversity, privacy, and security, and they provide potential solutions. They discuss several methods for data pre-processing, advanced algorithms for machine learning, and privacy-preserving methods that may be applied to enhance recommender system performance. A critical overview of the state-of-the-art in the topic is also provided by the authors, who also identify many possible areas for future research avenues.

Overall, the article is a useful resource for scholars and practitioners in the field ofrecommender systems by offering a blueprint for developing recommender systems that are efficient and dependable and can cater to the requirements of various applications and domains [7].

An improved method for the movie recommendation system is presented in the study "An Improved Approach for Movie Recommendation System" by Shreya Agrawal and Pooja Jain. To provide movie suggestions, the authors suggest an algorithm that combines the user-based collaborative filtering and item-based collaborative filtering methodologies. The cosine similarity technique and the Jaccard coefficient are also included in the recommended approach to solving the issue ofdata sparsity. On the MovieLens dataset, the authors assessed the effectiveness of the suggested algorithm and contrasted it with other recommendation methods. The experimental findings shown that the suggested strategy performed and was more precise than alternative methods [9].

The research also includes a hybridization methodology that takes use of the advantages of both user-based and item-based collaborative filtering approaches. The suggested approach successfully addresses the cold start issue and offers consumers personalised according to their preferences. The report also provides a thorough assessment of the suggested approach using several evaluation measures. The proposed approach by the authors makes a significant contribution to the field of movie recommendation systems and is pertinent to researchers, practitioners, and business experts who work in the discipline of recommendation systems due to its potential to enhance the performance and accuracy of such systems [9].

The creation of a recommendation system to help prospective graduate students choose appropriate graduate schools to apply to is presented in the research article titled "Graduate School Recommender System: Assisting Admission Seekers to Apply for Graduate Studies in Appropriate Graduate Schools". Based on the student's academic profile, such as their GPA, GRE scores, and undergraduate major, the algorithm is intended to suggest graduate institutions. To create the recommendation engine, the authors combined collaborative filtering methods with machine learning algorithms [10].

In the graduate school admissions procedure, when there are many institutions to select from and students are sometimes overwhelmed, the report emphasises the value of personalised recommendations. The algorithm performed well in terms of accuracy and relevancy of suggestions when tested using a dataset of graduate programmes in the United States. Overall, the study adds to the expanding body of knowledge on recommendation systems in the field of education and offers a workable way to help students choose graduate schools.[10].

The creation of a hybrid collaborative filtering-based recommender system is discussed in the article "A real-time recommender system based on hybrid collaborative filtering" by W. Yuan-hong and T. Xiao-qiu, which was presented at the 5th International Conference on Computer Science & Education in 2010. To provide customised suggestions for users, the system incorporates approaches for user- and item-based collaborative filtering. To enhance the accuracy of suggestions, the authors suggest a hybrid strategy that incorporates the advantages of the two approaches. The system's real-time operation incorporates aspects like user feedback and preference change to raise the accuracy of recommendations over time. The suggested system is a potential method for developing real-time recommender systems since it is easily adaptable to new domains [16].

IV. EXPERIMENTS AND METHODOLOGY

A. Data Collection

The automated technique of obtaining data from webpages is called web scraping. We need particular data items for our project, including university names, scores from the GRE, TOEFL, and IELTS, among other pertinent details. Yocket is an online resource for students looking to pursue higher education overseas that offers details about colleges, courses, entrance criteria, and other pertinent information. Data from the Yocket website will be extracted using the Quick Web Scraper tool.

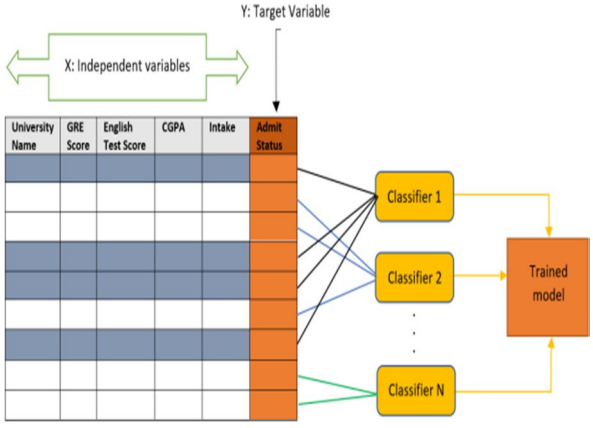
Finding the data source and the data components to be retrieved is the first stage in web scraping. The university's name, GRE, TOEFL, IELTS, and other pertinent data were among the data items we needed, and Yocket was the data source we used. We specified the data pieces we wished to retrieve using the Quick Web Scraper tool.

The next step in the data collection process is configuring the web scraper. We used the Instant Web Scraper toolconfigure our scraper by providing the website URL, selecting the data elements, and defining the scraping parameters. We also selected the export format for the data, which was a CSV file.

After configuring the web scraper, we ran the scraper to extract data from the Yocket website. The Instant Web Scraper tool navigated through the website, located the required data elements, and extracted the information. We monitored the scraping process to ensure the data collection was reliable and accurate.

B. Algorithms Used

1) *XGBoost Algorithm:* Classification models figure out a new instance's class or group based on its characteristics. To put it another way, classification models are a form of supervised learning that uses labelled data to learn in order to forecast outcomes for fresh, untainted data. Classification models are used to find patterns and connections in data that is used to distinguish between various groups or divisions. A classification algorithm could be used, for instance, to determine, based on an email's text and other characteristics, whether or not it is spam. Decision trees are a sort of supervised learning algorithm, and XGBoost (Extreme Gradient Boosting) is a machine learning algorithm created to enhance their efficiency. In order to develop a more accurate forecast, XGBoost buildsan ensemble of decision trees and iteratively improves it. A succession of decision trees that are trained on various segments of the data are the fundamental concept behind XGBoost. Every decision tree is built to fix the flaws of the one before it, improving the general prediction over time. In order to reduce mistakes, the training procedure includes determining the ideal weights for each tree. An ensemble technique called XGBoost combines an array of weak models to produce a powerful model. XGBoost can catch more intricate connections in the data and lessen overfitting by merging numerous weak models. Regularization methods are used by XGBoost to avoid overfitting, which can improve generalisation efficiency on untested data. In order to simplify the model and enhance its efficiency with fresh data, XGBoost adds fines to the loss function. A gauge of feature significance is offered by XGBoost, which can be used 1 to pinpoint which features are most pertinent for a particular job. XGBoost is able to increase accuracy and decrease noise in the data by concentrating on the most crucial characteristics. Overall, XGBoost is a potent machine learning method that combines numerous weak models, employs regularisation strategies, gradient-based optimization, feature importance, and handles missing values to produce high precision.
2) *Decision Tree Algorithm:* The decision tree is a popular machine learning technique for classification and regression issues. It divides the data into more homogeneous groups iteratively based on the values of the input characteristics. The final predictions or outcomes are represented by the tree's leaves, and each partition is a node in the tree. Because they are straightforward and easy to envisage, decision trees are a popular choice for many.so that they may manage categorical and continuous input characteristics as well as binary and multi-class classification challenges. Decision trees are a well-liked option for projects where the interpretability of the model is essential because they are simple to interpret and comprehend. Decision trees are useful because they can reveal information about the model's decision-making process, which is useful for applications like fraud detection, medical diagnosis, and credit assessment. Because they can capture complex nonlinear interactions between the input elements and the output variable, decision trees are a useful tool for predictive modelling. They can manage variable interactions and record complicated judgement limits that other linear models might struggle to do. To enhance the efficacy of the model, decision trees can be merged with other machine learning techniques like random forests or boosting. These ensemble techniques can lessen overfitting and raise forecast accuracy. In light of their capacity to manage complicated relationships, feature selection, scalability, and interpretability, decision trees are a flexible and effective machine learning, a technique that may be applied in several projects.
3) *Random Forest Algorithm:* Multiple decision trees are combined in the random forest ensemble learning method to increase the model's efficiency and robustness. Each decision tree in the random forest is trained using a random subset of the input attributes and a random piece of the train data. Random forest is often used in categorization and regression machine learning applications. It is well known for its ability to handle complex nonlinear interactions between input and output parameters and for its capacity to reduce overfitting by integrating the predictions of many weak learners.



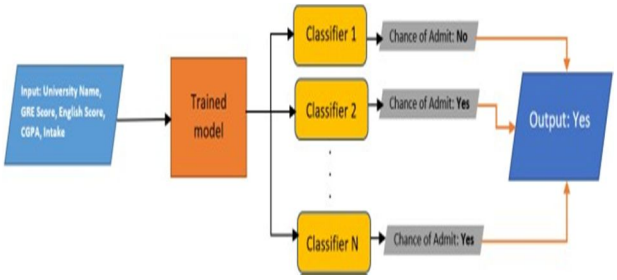


Fig.1 A visualisation of Random Forest Classifier

Particularly for complicated and nonlinear relationships between the input variables and the goal variable, random forest is renowned for its high precision in result prediction. It is a dependable option for predictive modelling jobs becauseit is resistant to data noise and anomalies. Estimates of feature importance provided by random forest can be used to pick features and obtain understanding of the underlying data.

C. Performance Metrics

Performance metrics are measurements that are employed to evaluate the effectiveness and productivity of a system, process, or individual. They are used to assess how well a work or project is doing and whether its objectives are being met. Performance metrics provide a way to make data-driven decisions to improve outcomes and precisely gauge progress. In the context of machine learning, performance measures are used to evaluate how well a model performs in terms of producing accurate predictions. In machine learning, accuracy, precision, recall, F1-score, and AUC-ROC are frequently used success metrics. The success metric to utilise is determined by the specific problem and the desired outcome.

A confusion matrix is a table that is used to assess the effectiveness of a machine learning model when solving a classification issue. It is a matrix that summarises a set of instances expected and actual categorization.

The confusion matrix is frequently set up as a chart with four columns for each of the four possibilities that could occur in abinary classification problem:

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Fig.2 Representation of Confusion Matrix in a tabular formatTrue Positive (TP) - The model correctly predicts the positive class. False Positive (FP) - The model incorrectly predicts the positive class. True Negative (TN) - The model correctly predicts the negative class. False Negative (FN) - The model incorrectly predicts the negative class.

Accuracy, precision, and memory are some examples of performance measures that can be calculated using the confusion matrix. As an illustration, recall can be computed as TP / (TP + FN), precision as TP / (TP + FP), and accuracy as (TP + TN) /(TP + TN + FP + FN).

A helpful instrument for assessing a machine learning model's performance is the confusion matrix, particularly for datasets with imbalances where one class may be more common than the other. One can find the areas where the model is off and correct it by looking at the uncertainty matrix. The proportion of accurately classified positive samples to all positivesamples that have been classified is known as precision.

$$Precision = TP/(TP + FP) \quad (1)$$

The recall is defined as the ratio of the total number of Positive samples by the number of Positive samples that were correctly identified as Positive. The model's capacity to identify positive samples is gauged by recall. More positive samples are found when recall is higher.

$$Recall = TP/(TP + FN) \quad (2)$$

An evaluation metric for a classification that is defined as the harmonic mean of recall and precision is the F1-Score or F-measure.

$$F\ 1score = 2*(precision*recall)/(precision+recall) \quad (3)$$

D. Flowchart of our Project

Figure 3 illustrates the procedures needed to forecast the likelihood of admission using machine learning.

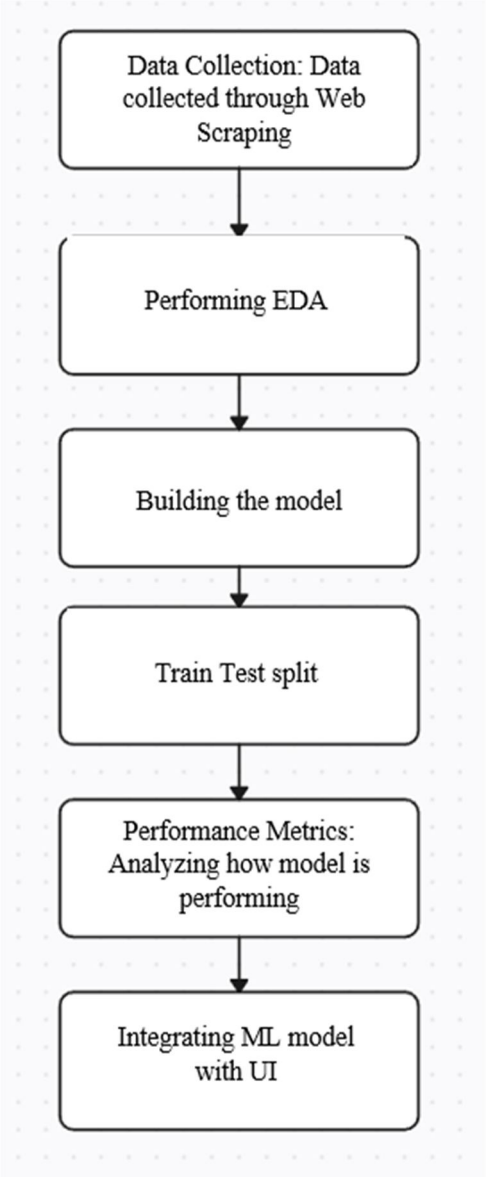


Fig.3 Methodology for solving the problem statement using Machine Learning

V. RESULTS AND CONCLUSION

Millions of students register at colleges each year to start their academic careers. Most of them lack the necessary tools, background knowledge, and caution, which leads to a host of issues like going to the incorrect university or college, which further wastes their time, money, and effort. We have made an effort to assist students who are having trouble selecting the best institution for them with the aid of our initiative. Going to colleges where one has a decent chance of admission is crucial, as opposed to going to colleges where one may never be accepted. Students will only submit to colleges where they have a high chance of admission, which will help to reduce costs. For such individuals, our prepared models function with anacceptable degree of precision and could be very helpful. For youngsters in our age group who want to continue higher education at the institution of their choice, this initiative has a bright future.

TABLE I
Model comparison with respect to their accuracies

MODEL	ACCURACY
Random Forests	0.76326
Decision Trees	0.76972
XGBoost	0.81495

A. Random Forest Classifier

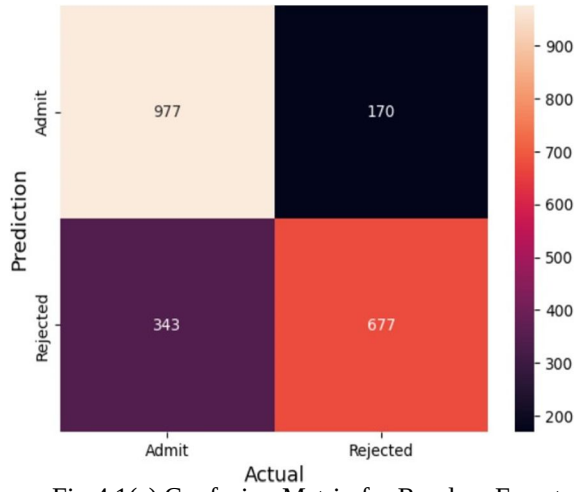


Fig.4.1(a) Confusion Matrix for Random Forest

```
Precision : 0.7992916174734357
Recall    : 0.6637254901960784
F1-score  : 0.7252276379217997
```

Fig.4.1(b) Performance metrics for Random Forest

B. Decision Tree Algorithm

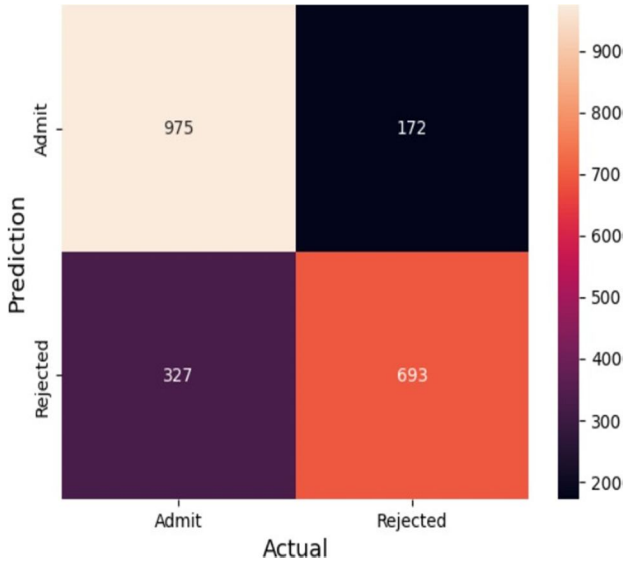


Fig.4.2(a) Confusion Matrix for Decision Tree

```
Precision : 0.8011560693641618
Recall    : 0.6794117647058824
F1-score  : 0.7352785145888594
```

Fig.4.2(b) Performance metrics for Decision Tree

C. XGBoost Algorithm

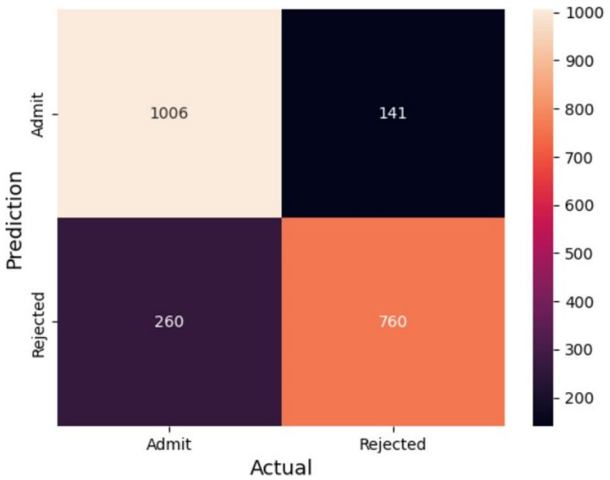


Fig.4.3(a) Confusion Matrix for XGBoost Algorithm

```
Precision : 0.8435072142064373
Recall    : 0.7450980392156863
F1-score  : 0.7912545549193128
```

Fig.4.1(b) Performance metrics for Random Forest

We have applied Random Forests, Decision Trees and XGBoost on the attributes such as CGPA, GRE, and TOEFL. By comparing the graphs of each model, it can be concluded that XGBoost has the best accuracy in this recommendation system, at 81%, compared to Random Forests and Decision Trees, which are 76% and 76%, respectively.

REFERENCES

[1] M. Hassan and M. Hamada, "Smart media-based context-aware recommender systems for learning: A conceptual framework," 2017 16th International Conference on Information Technology Based Higher.

[2] Education and Training (ITHET), Ohrid, 2017, pp. 1-4.Jain, Satia. (2021, December). College Admission Prediction using Ensemble Machine Learning Models. International Research Journal of Engineering and Technology (IRJET), 08(12), Article e-ISSN: 2395- 0056I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271350.

[3] Judy, D’cruz, Kathe, Motwani. (2020, April). Recommendation System for Higher Studies using Machine Learning. International Research Journal of Engineering and Technology (IRJET), 07(04), Article e-ISSN:2395-0056.

[4] Nalawade, Tiple. (2020, March). The University Recommendation System for Higher Education. International Journal of Recent Technology and Engineering (IJRTE), 08(06), Article ISSN: 2277-3878.

[5] E. Soldatova, U. Bach, R. Vossen and S. Jeschke, "Creating an E-Learning recommender system supporting teachers of engineering disciplines," 2013 International Conference on Interactive Collaborative Learning (ICL), Kazan, 2013, pp. 811-815.

[6] Bhatt, Shah, Soni. (2020, July). Recommendation System for Higher Studies at Abroad via Machine Learning Techniques. International Journal of Advanced Research in Science Technology (IJARST), 07(03), Article ISSN (Online) 2581-9429.

[7] M. H. Mohamed, M. H. Khafagy and M. H. Ibrahim, "Recommender Systems Challenges and Solutions Survey," 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, Egypt, 2019, pp. 149-155.

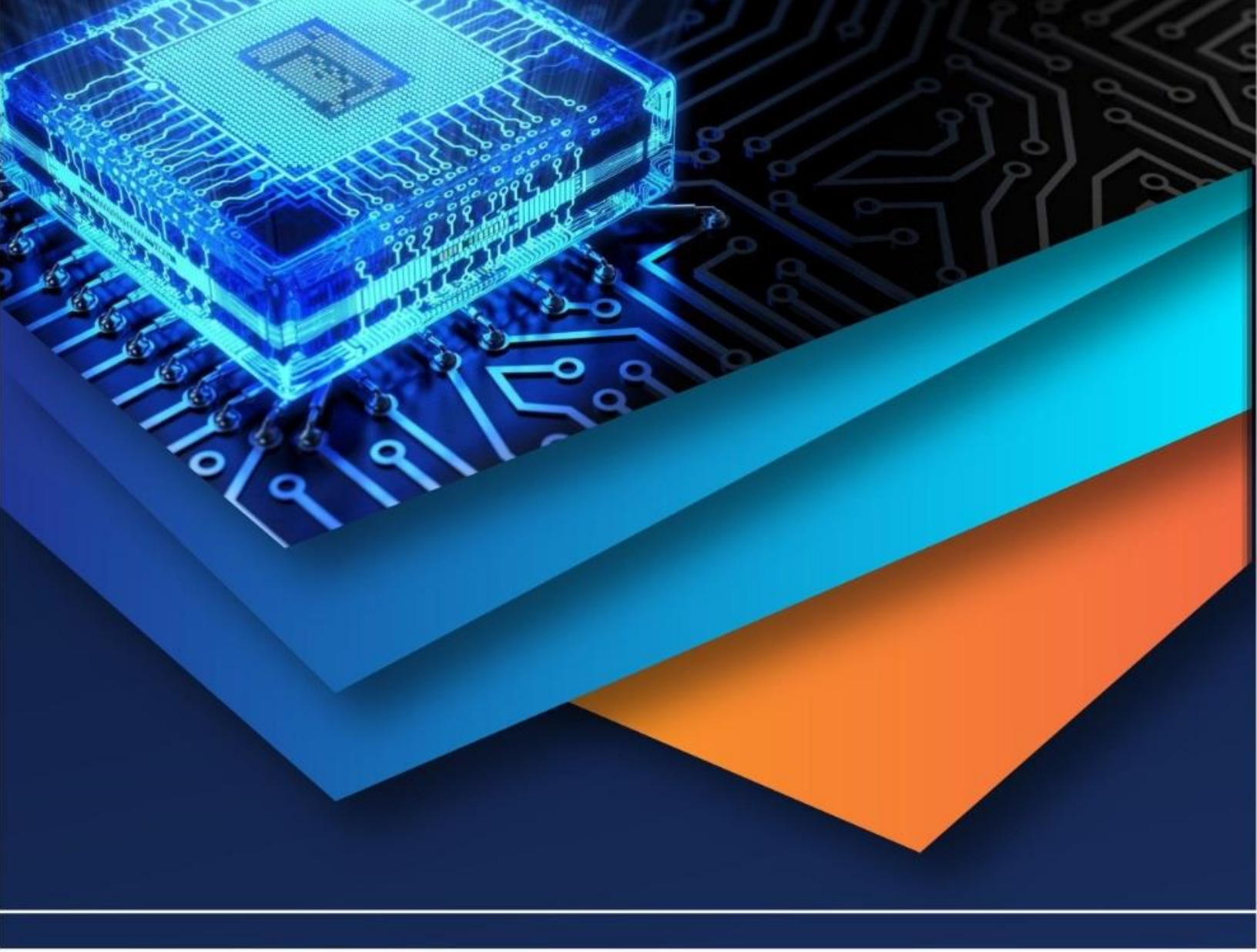
[8] M. H. Ansari, M. Moradi, O. NikRah and K. M. Kambakhsh, "CodERS: A hybrid recommender system for an E-learning system," 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), Tehran, 2016, pp. 1-5.

[9] Shreya Agrawal (ME Student) and Pooja Jain (Assistant Professor), An Improved Approach For Movie Recommendation System, 978-1-5090-3243 3/17/$31.00©2017 IEEE

[10] Mahamudul Hasan, Shibbir Ahmed,Deen Md.Abdullah, and Md.Shamimur Rahman, Graduate School Recommender System: Assisting Admission Seekers to Apply for Graduate Studies in Appropriate Graduate Schools, by 978-1-5090-1269-5/16/$31.00 ©2016 IEEE

[11] Usue Mori, Alexander Mendiburu, and Jose A.Lozano, “Similarity Measure Selection for Clustering Time Series databases,” IEEE Transactions on Knowledge and Data Engineering. Vol. 28. No. 1. January 2016.

[12] J. Bobadilla et al. “Knowledge-Based System” Elsevier B.V.
[13] D. Rodriguez-Cerezo, M. Gomez-Albarr´n and J. Sierra, "Supporting Self-Regulated Learning in Technical Domains with Repositories of Learning Objects and Recommender Systems," 2011 IEEE 11th International Conference on Advanced Learning Technologies, Athens, GA, 2011, pp. 613-614.
[14] Hector Nunez, Miquel sanchez-Marre, Ulises Cortes Joaquim Comas, Montse Martinez, Ignasi Rodriguez-Roda, Manel Poch, “A Comaparative study on the use of similarity measure in case-based reasoning to improve the classification of environmental system situations,”, ELSEVIER, EnvironmentalModeling and Software (2003)
[15] Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi, and Zhi Wang, “Friend book: A Semantic based Friend Recommendation System for Social Networks IEEE Transactions on Mobile Computing.
[16] W. Yuan-hong and T. Xiao-qiu, "A real-time recommender system based on hybrid collaborative filtering," 2010 5th International Conference on Computer Science & Education, Hefei, 2010, pp. 1909-1912.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)