Research Paper

Comprehensive Guide and Assistance System for Future Studies and Academic Pursuits

Udayan Iyer[1]* , Bhavesh Sonje[2] , Gaurav Suvarna[3] , Bushra Shaikh[4]

[1,2,3,4]Department Of Information Technology, SIES GST, Navi Mumbai, India

*Corresponding Author: iyerudayanganeshan19@siesgst.ac.in

Abstract: Education is a deliberate pursuit aimed at achieving specific goals such as imparting knowledge, developing skills, and fostering moral values. These objectives encompass the advancement of comprehension, rationality, empathy, and ethics. Critical thinking plays a vital role in distinguishing education from indoctrination, as emphasized by numerous studies. Higher education is provided by academic institutions such as universities and colleges, leading to the attainment of degree certificates. Those who pursue tertiary education have better prospects of securing well-paying jobs and forging unique career paths. Furthermore, they are more likely to cultivate profound critical thinking and reasoning abilities that contribute to personal development. The quality of education varies among universities, influenced by the curriculum and teaching methods employed. Therefore, it is essential to carefully consider one's options and aspirations before making a decision. Given the circumstances, individuals should be prepared for the demanding college admission process, which can be time-consuming.

Keywords: Higher Education, Linear Regression, Random Forest, Education, Decision Tree, College Prediction, Artificial Neural Network.

1. Introduction

Education holds great significance in our lifelong journey, as it allows us to pursue higher learning opportunities tailored to our preferences and aspirations. This opens doors to diverse fields of study and educational institutions worldwide. When planning our educational path, crucial questions arise, such as selecting the right university and choosing the appropriate course of study. These decisions involve considerations of job prospects, expenses, housing, and more. Ultimately, every student's ultimate goal is to secure admission to their dream university. However, the admissions process is highly competitive, placing significant pressure on students to excel in academics, extracurricular activities, and other relevant areas. Student demographics have evolved over the years, with over 8 million international students currently pursuing education in the United States alone across public and private universities. This increasing number indicates a growing trend of higher education and intensifying competition.

Given this context, it is essential for students to adequately prepare for the demanding college admissions process. Motivated by this challenge, our project aims to assist students in their pursuit of higher education. Our objective is to provide comprehensive support through various initiatives. To develop a solution, we must first understand the core of the problem: the college admissions process. This process can be summarized in several key steps:

1. Researching desired subjects, colleges, and related exams.
2. Considering admission timelines, tuition fees, and scholarships.
3. Taking required exams.
4. Receiving exam results.
5. Identifying suitable colleges based on individual profiles.
6. Offering guidance on selecting the right college according to individual profiles.
7. Submitting applications to universities.

Our project focuses on the evaluation phase of the admissions process. At this stage, students have received their exam results, completed their resumes, and gathered the necessary certifications for their applications to be considered by universities. Our project takes into account the factors universities consider during this evaluation. It provides additional insights into students' profiles, helping them make informed decisions throughout the application process. Our software calculates acceptance probabilities for each university, provides students with an understanding of their global ranking, and enables them to analyze their situation realistically. Additionally, the project assists students in researching colleges based on predicted outcomes. Recognizing the value of second opinions based on personal or professional experiences, we offer such suggestions through our website, utilizing blog portals and forums as mediums of communication.

The prediction software recommends universities that are likely to accept users. Overall, this project aims to offer a comprehensive analysis of user profiles, along with research guidance and personal advice, to support students throughout the college admissions process.

2. Literature Review

Goutam Kumar Kundu et al. have provided necessary information about higher studies & about the exams to get admission into colleges. [1]

Moses Acquah & Na Chen et al. describes the process of acquiring a visa for various countries. The authors of the paper "Machine Learning & Higher Education" Heath Chamber & Craig Chamberlain, told us that higher education has the potential to utilize machine learning not only for improving its decision-making processes but also to investigate its applications in promoting positive outcomes. [2]

Thiago Nazareth & Flavia Bernardini el at. told us that Recommendation Systems have the potential to assist students and managers in selecting academic disciplines, leading to improved academic performance and enhancing student learning, engagement, and retention. These Recommendation Systems can be built using Educational Data Mining techniques, which incorporate machine learning algorithms to identify and anticipate situations where students may drop out or discontinue their studies, thereby helping to minimize such instances. [3]

Kiran Fahd & Khandakar Ahmed et al. told us that the field of machine learning (ML) has undergone significant advancements and is now being utilized in higher education (HE) for diverse data analysis purposes. Research has demonstrated that this emerging area in educational technology offers valuable insights into various aspects of educational excellence. Conducting a comprehensive examination of ML applications could have a beneficial influence on the higher education sector. However, there is currently a lack of a systematic review of the existing literature in higher education, which hinders the ability to leverage the overall trends and patterns identified through ML. [4]

Ghadeer Abdul et al. tell us that "this paper, we apply different data mining approaches for the purpose of examining and predicting students' dropouts through their university programs. For the subject of the study we select a total of 1290 records of computer science students Graduated from ALAQSA University between 2005 and 2011. The collected data included student study history and transcript for courses taught in the first two years of computer science major in addition to student GPA , high school average , and class label of (yes ,No) to indicate whether the student graduated from the chosen major or not. In order to classify and predict dropout students, different classifiers have been trained on our data sets including Decision Tree (DT), Naive Bayes (NB). These methods were tested using 10-fold cross

validation. The accuracy of DT, and NlB classifiers were 98.14% and 96.86% respectively. [5]

Elaf Abu Amrieh & Ibrahim Alijarah et al. tell us that "Educational data mining has received considerable attention in the last few years. Many data mining techniques are proposed to extract the hidden knowledge from educational data. The extracted knowledge helps the institutions to improve their teaching methods and learning process. All these improvements lead to enhance the performance of the students and the overall educational outputs". [6]
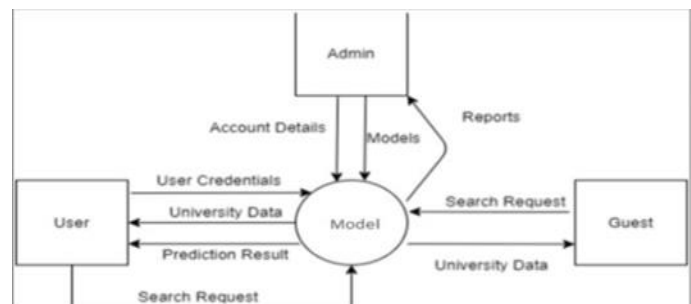
3. Proposed System

Figure 3.1 Diagrammatic flow diagram of future studies guide & Assistance System

The data flow diagram of our project is shown in Figure 3.1. The administrator logs in and resumes his duties of supporting users with authorization, security, and feedback. Additionally, it oversees and maintains data sets for prediction models. On the website, users can browse the university database and view information about their profiles. The prediction tool can be used by logged-in users. The user can also get guiding features through various media. Unless they register or log in, a guest user is not able to utilize any of those functions. Visitors are only permitted to browse the forum or website with restricted access. [7]
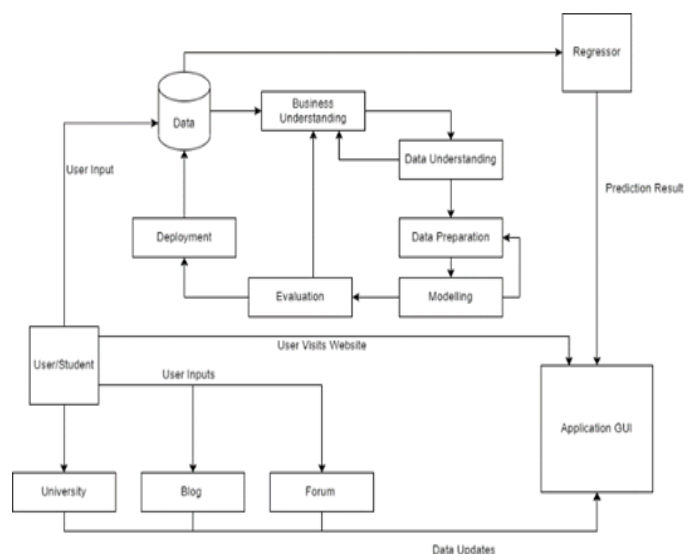
4. Implementation Details

Figure 4.1 Block diagram

Figure 4.1 states the block diagram is shown which describes the flow of the website. It starts from user visiting as guest, after authorization process the user logs into the website, now it can start receiving and sending data using controller which manages the flow between database and views. It returns created and updated view to the user. The pre-trained models take the input from the user and gives the predicted odds for the user with respect to a particular university. The data uploaded for models are analyzed and goes through a cyclic process of development from applying the model to result evaluation. In the end, it deploys the result and sends feedback which improves the model producing better results in future.

The User is depicted as being the main actor in our model, whereas the admin and cloud are support actors. User actor techniques include logging in, signing up, browsing university databases, accessing blogging portals, and participating in forums. The admin actor also handles user query resolution, authorization, security, and mechanisms for cloud-based data set deployment. The models for our prediction system are trained, tested, compared, studied, and deployed through the cloud service. [8]

S. No	GRE Score	TOEFL Score	Uni	SOP	LOR	CGPA	R.P	Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65

Figure 4.2 Data set snippet

This project utilizes data from a variety of sources, with the bulk of the data coming from public websites like Kaggle. The data set was cleaned of unnecessary parameters in order to predict a student's likelihood of admission to any university. The resulting data set is depicted in Figure 4.2. One dependent parameter and seven independent parameters make up the data set. Along with the student's SOP and LOR points, competitive exam results like the GRE and TOEFL are taken into account. Under graduation, it is determined that a student's CGPA and published research paper are significant factors in their profile. All of these factors have an impact on the likelihood of acceptance for each rated university. In order to get the highest degree of variance in the admission results, we decided on 8 factors.

There are several ways to perform prediction analysis in the field of machine learning, including regression, classification, etc. Regression lets us anticipate the outcomes with a precision value, in contrast to other approaches like classification, which only gives us acceptance or rejection values, thus we chose it because this project seeks to deliver the percentage likelihood of admission. The regression models used in our project are listed below.

1.Linear Regression (LR): One of the simplest and most well-known Machine Learning algorithms is linear regression.

This statistical technique is employed in predictive analysis. For variables that are real or numerical, linear regression makes predictions. A dependent (y) variable and one or more independent (y) variables are shown to have a linear relationship via linear regression, so known as linear regression. Since linear regression shows a linear relationship, it determines how the dependent variable's value changes in relation to the independent variable's value. The line of regression's goodness of fit describes how well it matches the data set of observations. Optimization is the process of selecting the model that best fits the situation from a pool of available model. [9]

2.Artificial Neural Network (ANN): An ANN is composed of a group of nodes known as artificial neurons that resemble the neurons in a biological brain in some ways. Other neurons can receive signals from each link. An artificial neuron processes whatever signal it receives carefully and can transmit the signal to other neurons that are connected to it. The "signal" at a connection is a real number, and each neuron's output is determined by some non-linear function of the sum of its inputs. The relationships are referred to as edges. The weight of the neurons and edges naturally changes as learning progresses. The signal intensity at a link is altered by the weight. A threshold in neurons can determine whether a signal is only broadcast if the signal's totality exceeds that value. [10]

3.Decision Tree (DT): This algorithm for supervised machine learning. The most extensively used categorization method is characterized by its straightforward logic, efficiency, and interpretability. The model operates by separating the data set into numerous smaller subsets depending on various conditional logic, resulting in a tree-like structure. The decision nodes, leaf nodes, and branches make up the decision tree's essential parts. Decision nodes are nodes with several branches, leaf nodes are nodes without branches, and the top node is known as the decision node. The nodes are linked to one another by branches based on various conditions. Calculating the entropy and information gain for the data set results in the creation of the root and decision nodes. [9]

4.RF(Random Forest)-A well-known machine learning algorithm that uses the supervised learning method is called Random Forest (RF). Machine learning problems including classification and regression can be solved with it. It is based on the idea of ensemble learning, which is the practice of integrating various classifiers to solve a challenging problem and enhance the performance of the model. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, predicts the final output. As the name suggests, "RF is a classifier that contains several decision trees on many subsets of the dataset and takes the mean to improve the predictive accuracy of that dataset. [6]

5. Results and Discussion

Figure 5.1 Landing Page

This is the page that comes first when we load our website.This page explains what the website is about,what is its purpose & how it serves the students etc.

Figure 5.2 Contact Us Page

This is the Contact page of our project where students can contact us if they have any doubts regarding exams,visa processing etc.
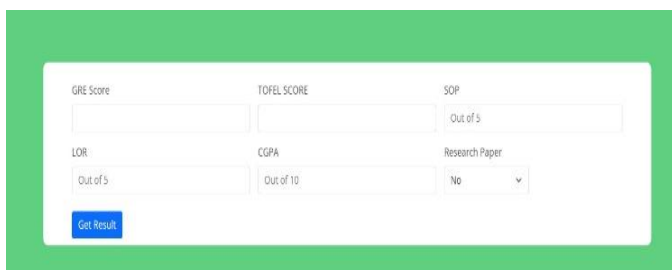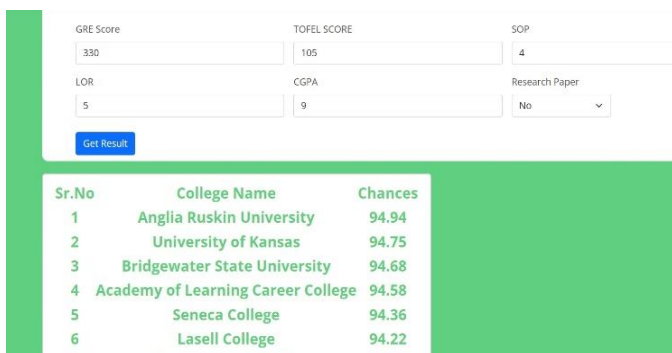

Figure 5.3 College List Predictor

Figure 5.4 College List Predictor results

In college list predictor,students enter their gre,toefl scores & other information & a list of colleges are predicted in which they can get admission into based on their scores.

Figure 5.5 College Odds Predictor

Figure 5.6 College Odds Predictor Result

In college odds predictor,students enter their gre,toefl scores &other information & the college they in which they want to get admitted , then this predictor predicts the odds of the student getting into that college.
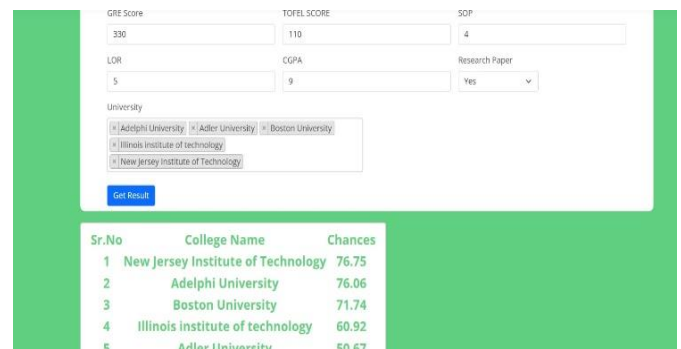

Figure 5.7 Preference Order Predictor

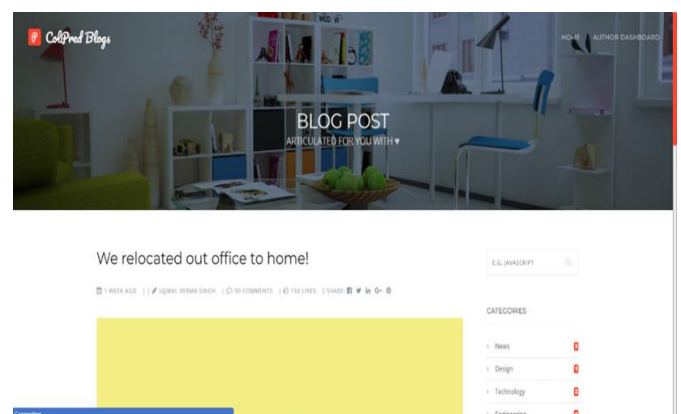In preference order predictor,colleges are predicted based on the students prerference order.

Figure 5.8 Blogs page

This is the blog page of our website.Where we can post blogs about colleges,exams,universities, entrance exams & many more.
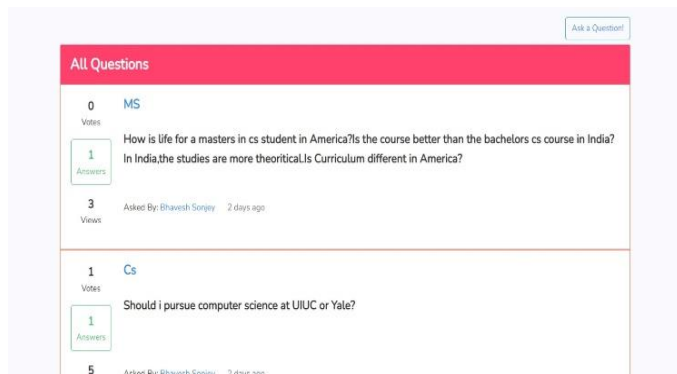
Figure 5.9 Forums Page

This is the forum page of our website.Where students can ask questions regarding their doubts about colleges,exams & many more.People can answer to their questions & people can upvote or downvote their answers if they agree or disagree to it.
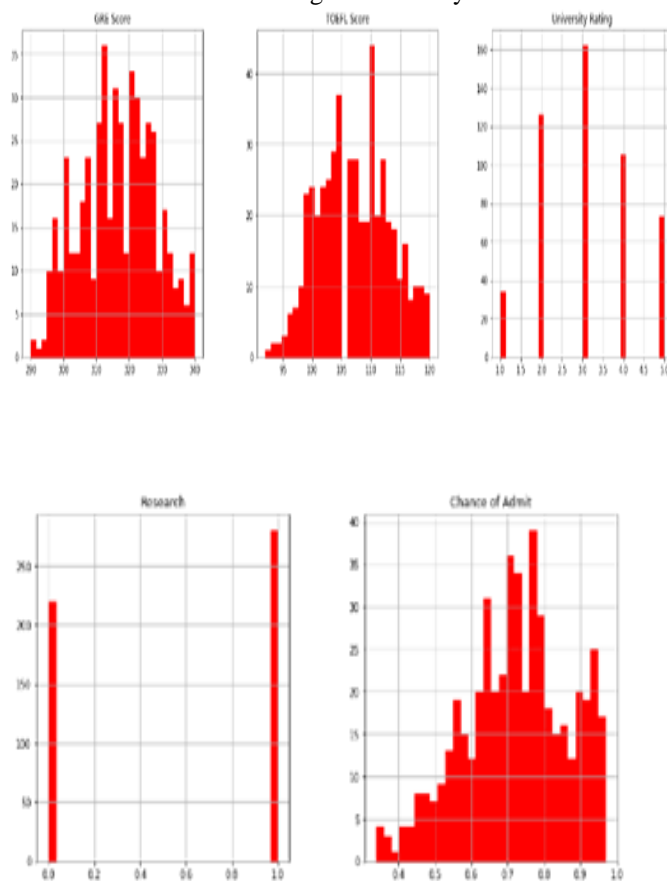
ML Algorithm Analysis

Figure 5.10 Exploratory Data Analysis

Exploratory Data Analysis is seen in Figure 5.10. Metrics from the data set are displayed as a bar graph or histogram. According to research, undergrad CGPA and competitive examinations like the GRE and TOEFL produce a Bell Curve where the majority of students score in the average range. SOP and LOR have a 5-point metric scale, while the project only has a 2-point metric scale that can be either 0 or 1. The illustration demonstrates that the chance of admission is vague and not tied to any one parameter.
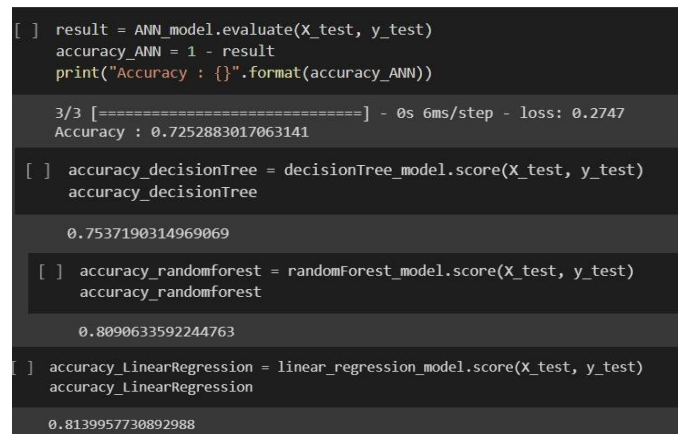

Figure 5.11 ML Algorithms accuracy

From the above figure 5.11,we can see that Linear regression gives a accuracy of 81.39% for our project which is better than the accuracy given by Random forest,decision tree,ANN.So we can say that Linear Regression is the best algorithm for our project.
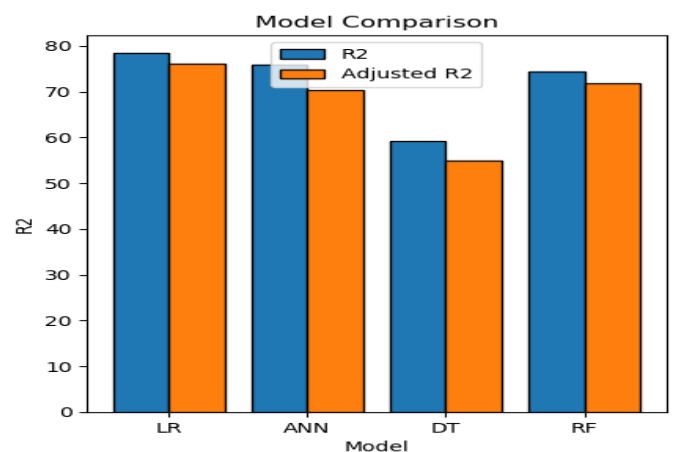

Figure 5.12 R2 & Adjusted R2 Comparision

Figure 5.12 compares four regression models with respect to their R2 and adjusted R2 values. It is clear that linear regression yields the highest R2 and Adjusted R2 values, which are correlated with the model's accuracy. As a result, when compared to alternative models, we have a stronger preference for linear regression.
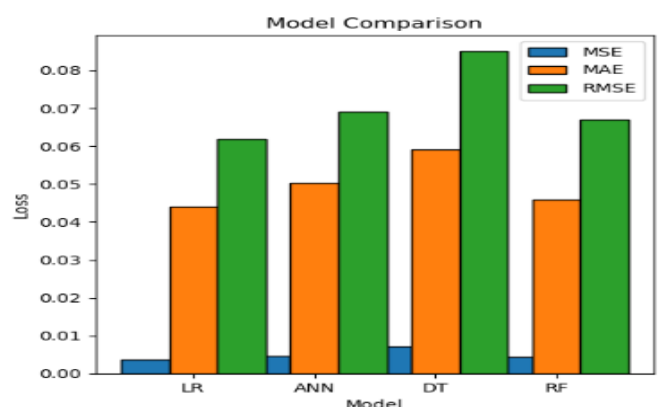

Figure 5.13 MSE,MAE,RMSE Comparison

The comparison of our models with the MSE, RMSE, and MAE values is shown in Figure 5.13. We can clearly see the differences between each model from the predicted graph. Lower the value, better prediction results should be anticipated because the metrics obtained here provide us values for the error rate. It is obvious that for each metric, Linear Regression gives us the lowest values of all. Therefore, after analyzing these 3 indicators, we once more choose linear regression above other models.

6. Conclusion and Future Scope

In this project, we conducted an extensive examination of different approaches to implementing machine learning techniques. We utilized various evaluation measures to assess the performance of regression models. By analyzing the results obtained for each metric, we determined that linear regression was the most suitable model for our specific project. We designed the architecture of the selected model. Our website offers comprehensive solutions, ranging from predicting university admission outcomes and providing research assistance, to offering reliable advice for university choices through various channels. The predictor takes user credentials and generates accurate results based on their profile. Additionally, the website includes a database of universities, along with blogs and forums to assist users in addressing their queries at all levels of complexity, helping them with research prior to submitting their applications and providing detailed information to alleviate any doubts or uncertainties.

There are several ways to enhance this project, such as incorporating newer technologies, gathering additional data, and utilizing more advanced models. The web application can be enriched by introducing more features and enhancing user connectivity across all three phases. To enhance the user experience, it is beneficial to leverage modern website development tools and the latest frameworks. The end-to-end framework can be improved by utilizing more extensive datasets and training the models for longer durations on powerful computers. By conducting multiple training and testing iterations, the framework can become more robust and capable of handling various realistic scenarios. Conducting surveys among stakeholders can provide valuable insights and improve understanding. Furthermore, integrating an AI assistant feature can offer quick solutions and eliminate the need for human interaction. While the project currently focuses on engineering and specific countries, expanding its scope to other fields like MBA, MSc, MBM, etc., can enhance its applicability. Employing better software development methodologies and fostering increased collaboration will contribute to improving the project's overall design.

Conflict Of Interest

Conflict of interest refers to a situation where an individual or entity has competing interests that could potentially compromise their ability to provide unbiased guidance or assistance. In the context of a future studies guide and assistance system, it is crucial to recognize and mitigate any conflicts of interest that may arise. This is because future studies aim to provide objective and reliable information to individuals seeking guidance for their academic pursuits, and any bias or partiality can undermine the integrity of the system.

To ensure the absence of conflicts of interest, the future studies guide and assistance system should adopt transparent policies and procedures. These should include strict guidelines for disclosing any affiliations, financial interests, or potential biases that could impact the information or recommendations provided. The system should also prioritize independent research and analysis, seeking input from a diverse range of experts and sources to ensure a comprehensive and well-rounded perspective.

Additionally, the future studies guide and assistance system should maintain a clear separation between the provision of information and any commercial or promotional activities. It should avoid endorsing or promoting specific products, services, or institutions that may create conflicts of interest. By prioritizing transparency, independence, and integrity, the system can establish trust among users and maintain its credibility as a reliable source of unbiased guidance for future studies and academic pursuits.

Author's Contribution

Corresponding Author (Mr. Udayan Iyer) - Content Creation: With a strong aptitude for research and writing, this college student will take charge of content creation for the website. They will conduct extensive research on future studies, academic pursuits, and related topics, and generate informative and engaging content. Their role also includes organizing and structuring the information, optimizing it for search engines, and collaborating with the mentor and other team members to ensure accuracy and relevance.

2nd Author (Mr. Bhavesh Sonje) - Web Development: As a college student specializing in web development, their contribution lies in designing and building the technical infrastructure of the website. They will leverage their skills in HTML, CSS, and JavaScript to create the necessary web pages, implement functionality, and ensure responsive design across various devices. Their role involves collaborating with the mentor and other team members to understand the requirements and deliver a user-friendly and visually appealing website.

3rd Author (Mr. Gaurav Suvarna) - User Experience (UX) Design: This student specializes in user experience (UX) design and will contribute to the website's overall usability and user satisfaction. They will conduct user research, create wireframes and prototypes, and design intuitive navigation and interaction elements. By considering the needs and preferences of the target audience, they will help create a seamless and enjoyable user experience. Their collaboration

with the mentor and other team members will ensure the website's design aligns with the goals of the future studies guide and assistance system.

4th Author (Ms. Bushra Shaikh) - Guidance and Oversight: The mentor, an experienced professional in the field of future studies or web development, will provide guidance and oversight throughout the website's creation. They will mentor the college students, offering valuable insights, best practices, and industry knowledge. The mentor will help refine the project's objectives, review progress, and provide feedback to ensure the website meets high standards. They will also assist in coordinating the efforts of the team, fostering collaboration, and offering support to address any challenges that arise during the development process.

The manuscript underwent a thorough review and editing process by all authors, and they have collectively endorsed and approved the final version.

Acknowledgement

We acknowledge that every individual involved in this work has provided their consent for their contribution to be acknowledged.

References

[1] Goutam.Kumar.Kundu ,"Higher Education Quality." ICTACT Journal On Management Studies, May 2016, Vol.**2**, Issue.**2**, **2016**.

[2] Moses Acquah & Na Chen, "Visa Processing ETS." ICTACT Journal On Management Studies. **2016.**

[3] Osmanbegovic, E. and Suljic, M, "Data mining approach for predicting student performance." Economic Review: Journal of Economics and Business, Vol.**10**, Issue.**1**, pp.**3-12, 2012.**

[4] Thakar, P, "Performance analysis and prediction in educational data mining: A research travelogue." arXiv preprint arXiv:1509.05176.2015.

[5] Alyahyan, E. and Düştegör, D, "Predicting academic success in higher education: literature review and best practices." International Journal of Educational Technology in Higher Education, **17**(**1**), p.**3, 2020.**

[6] H S Divyashree , "Student Strategizing in Education system using a Machine Learning Model" in the proceedings of the 2018 International Journal of Computer Sciences and Engineering, India, pp.**584-585, 2018.**

[7] Lubna Mahmoud Abu Zohair , "Prediction Of Student's Performance By Modelling Small Dataset Size."**2019**.

[8] Martin Liz-Dominguez, Manuel Caeiro- Rodriguez, Martin Llamas-Nistal, And Fernando Mikic-Fonte , University Of Vigo, Spain, "Predictors And Early Warning Systems In Higher Education."**.2019**.

[9] Erman Yukselturk, Serhat Ozekes, Yalın Kılıç Türel,, "Predicting Dropout Student: An Application Of Data Mining Methods In An Online Education Program."**2014**.

[10] B. Mounika, Voola Persis , " A Comparative Study of Machine Learning Algorithms for Student Academic Performance.". In the proceedings of the 2019 International Journal Of Computer Sciences & Engineering, India, pp.**723-724, 2019**

AUTHORS PROFILE

Mr. Udayan Iyer: is currently pursuing his Bachelor of Engineering in Information Technology from SIES GST, Navi Mumbai, India. His fields of interest are Machine Learning & Data Science

Mr. Bhavesh Sonje: is currently pursuing his Bachelor of Engineering in Information Technology from SIES GST, Navi Mumbai, India. His fields of interest are Machine Learning & Data Science.

Mr. Gaurav Suvarna: is currently pursuing his Bachelor of Engineering in Information Technology from SIES GST, Navi Mumbai, India. His fields of interest are Machine Learning & Data Science.

Ms. Bushra Shaikh: is currently working as assistant professor in SIES GST, Navi Mumbai, India. She is a graduate in BE-Information Technology & ME- Information Technology. She has published 8 papers so far & has 8 publications.