University Admission Eligibility Prediction using IBM Watson

Team members

R Naga Tarun – 20MIS0205

G V S Naga Phani – 20MIS0313

P Bhimesh – 20MIS0062

Ch. V Sasi Kumar – 20MIS0294

1. INTRODUCTION

1.1 Overview

University admission eligibility prediction refers to the process of using data and predictive models to assess the likelihood of an individual's admission into a particular university or academic program. It involves analysing various factors and criteria that universities consider during their admission process and using that information to estimate an applicant's chances of being accepted.

To predict admission eligibility, data from past admission cycles are utilized and include factors such as academic performance (such as GPA), GRE Score, TOEFL Score, letters of recommendation, and other relevant information provided by applicants. Machine learning and statistical techniques can be applied to this data to build predictive models.

University admission eligibility prediction can provide valuable insights to both applicants and universities. For applicants, it can offer an estimate of their chances of being admitted to specific universities or programs, allowing them to make informed decisions about where to apply. For universities, it can help streamline the admission process by identifying highly qualified candidates or predicting the yield rate (the percentage of admitted students who accept an offer) for a given year.

1.2 Purpose

The use of university admission eligibility prediction projects can have several practical applications and benefits. Here are some of the potential outcomes that can be achieved through this project:

Applicant Guidance: The prediction models can help prospective students assess their chances of being admitted to specific universities or programs. This information can guide them in making informed decisions about where to apply, helping them target institutions that are more likely to accept them.

Efficient Resource Allocation: Universities can utilize admission eligibility prediction to allocate their resources effectively. By estimating the likelihood of admission for each applicant, universities can prioritize their efforts and resources on candidates who are more likely to be accepted.

Enrolment Planning: By predicting the yield rate, universities can estimate the number of admitted students who are likely to accept their offer of admission. This information helps in managing class sizes, making waitlist decisions, and ensuring a balanced student population.

Targeted Outreach and Recruitment: Universities can use admission eligibility prediction to identify potential candidates who may be a good fit for their programs

Continuous Improvement: By analysing the performance of the prediction models over time, universities can gather feedback on the effectiveness of their admission criteria and processes.

2 LITERATURE SURVEY

2.1 Existing problem

While university admission eligibility prediction projects offer several benefits, there are also potential challenges and concerns associated with their use. Here are some of the existing problems that can arise:

- **Limited Predictive Accuracy:** Admission eligibility prediction models are based on historical data and statistical analysis, which means they are only as accurate as the data they are trained on.
- **Subjective Factors:** University admissions often involve subjective elements such as interviews, essays, and letters of recommendation that are difficult to quantify and incorporate into prediction models.
- **Lack of Contextual Information:** Admission eligibility prediction models typically rely on quantifiable data, such as test scores and GPA, but may not consider the context in which applicants achieved those results. Factors like socio-economic background, educational opportunities, or personal challenges faced by applicants may influence their performance
- **Overreliance on Predictive Models:** While admission eligibility prediction models can provide insights, they should not be the sole

determinant of admission decisions. The models should be used as decision support tools rather than replacing the comprehensive evaluation performed by admissions committees. Overreliance on predictive models may overlook exceptional cases or unique qualities that could contribute to an applicant's potential.

- **Ethical Considerations:** There are ethical concerns surrounding the use of predictive models in admissions. Bias and discrimination can inadvertently be encoded in the models if the training data reflects historical biases or if certain factors disproportionately affect different demographic groups. It is crucial to ensure fairness, transparency, and accountability in the development and application of these models to mitigate these concerns.

2.2 Proposed Solution

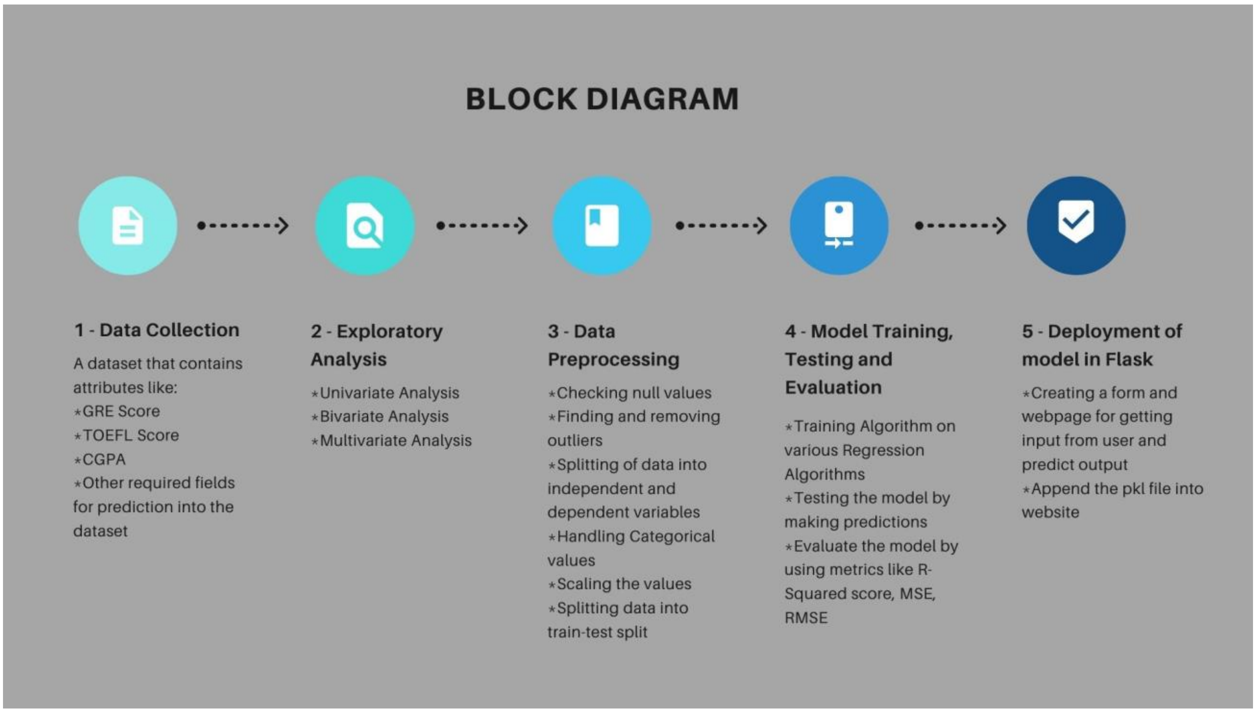
To address the existing problems associated with university admission eligibility prediction, the following solutions can be considered:

- **Enhanced Data Collection:** Collecting a comprehensive set of data that includes both quantitative and qualitative aspects of an applicant's profile can provide a more holistic view.
- **Contextual Analysis:** Incorporating contextual information about an applicant's background, such as socio-economic factors or educational opportunities, can help provide a more nuanced understanding of their achievements and potential.
- **Regular Model Updating:** Admission eligibility prediction models should be continuously updated and refined to reflect changing admission criteria, applicant profiles, and other relevant factors. This involves retraining the models with the most recent data and considering feedback from admissions committees to improve their accuracy and relevance.
- **Fairness and Bias Mitigation:** It is crucial to address potential biases in the data and models used for admission eligibility prediction. This can be achieved by conducting regular audits of the data, applying fairness-aware algorithms, and ensuring diverse representation in the development and evaluation of the models. Transparency in the model's algorithms and decision-making process should also be prioritized.

- **Educating Applicants:** Providing clear information to applicants about the limitations and uncertainties of admission eligibility prediction can help manage expectations and promote informed decision-making.
- **Continuous Evaluation and Feedback:** Regular evaluation and feedback loops involving admissions committees, applicants, and other stakeholders can help identify shortcomings and areas for improvement in the prediction models and the admission process as a whole.

3 THEORETICAL ANALYSIS

3.1 Block Diagram



3.2 Hardware and Software Requirements

<u>**Hardware Requirements:**</u>

A computer with following requirements:

- **Computing Resources:** Sufficient computational power is required to train and evaluate the predictive models.
- **Storage Capacity:** Adequate storage capacity is needed to store the datasets, preprocessed data, trained models, and any intermediate results generated during the project.
- **Memory (RAM):** Sizable memory capacity is beneficial, when working with large datasets or complex models. Sufficient RAM enables efficient processing and manipulation of data during training and evaluation.

- **Software Requirements:**
- **Anaconda Navigator:** It has the most of applications pre-installed like Jupyter Notebook, Flask
- **Web Browser:** To run and view the output based on the input values entered by the user
- **Python:** To run the application the latest version of python and the frameworks related to python and Machine Learning.

4 EXPERIMENTAL INVESTIGATIONS

When working on the proposed solutions for addressing the problems associated with university admission eligibility prediction, several key analysis and investigations can be conducted. Here are some areas to consider:

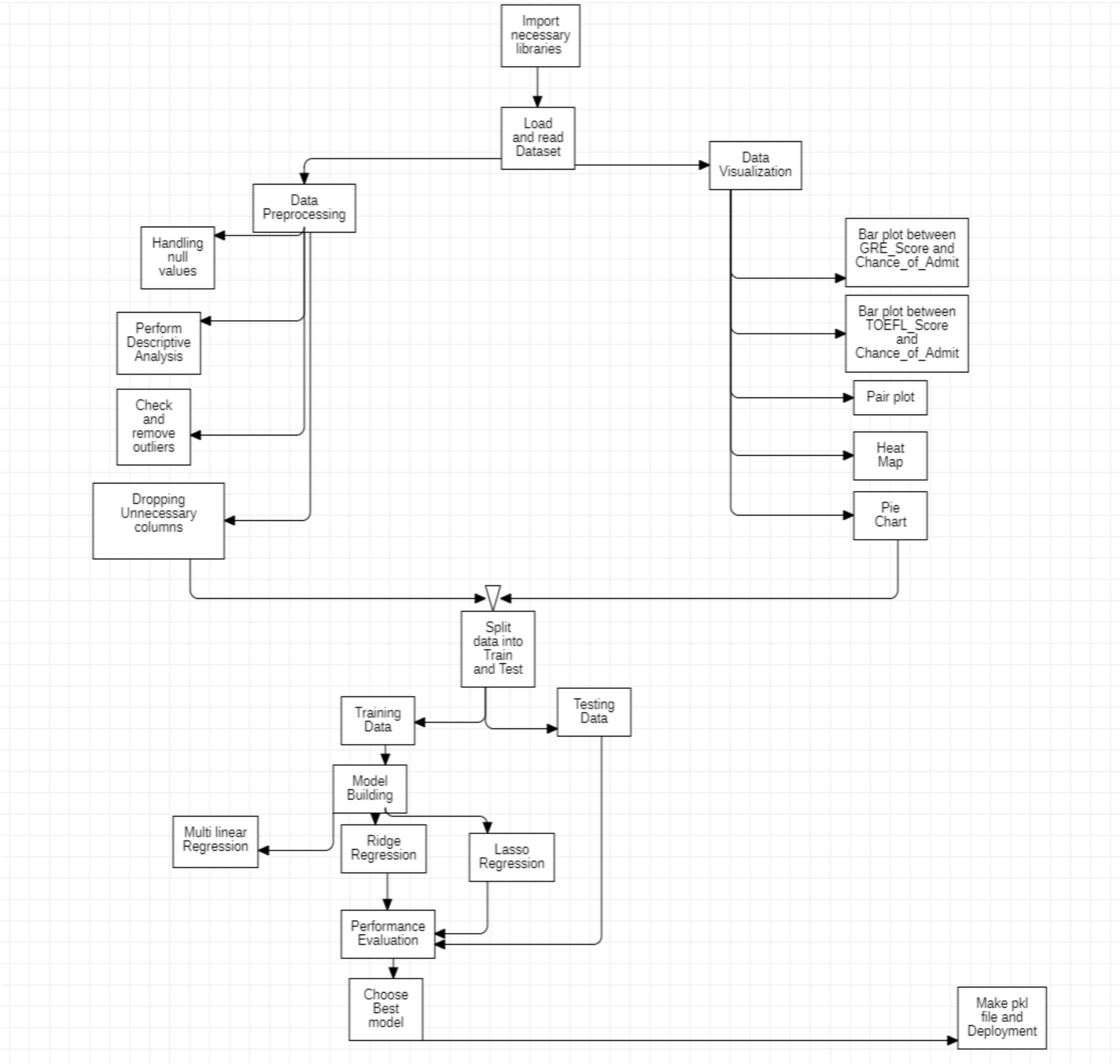
Data Analysis: Analyse the available data used for admission eligibility prediction, including historical applicant profiles and admission outcomes. Assess the quality, completeness, and representativeness of the data. Identify any biases or limitations present in the dataset that may impact the accuracy and fairness of the predictive models.

Evaluation Metrics: Define appropriate evaluation metrics to assess the performance of the predictive models. This may involve metrics such as accuracy, precision, recall, or fairness measures to evaluate how well the models predict admission outcomes and mitigate bias.

Bias Assessment: Conduct a comprehensive analysis of potential biases in the data and models. Examine whether there are any disparities in predictions or admission outcomes based on demographic or other sensitive attributes. Use fairness evaluation techniques to identify and mitigate biases, ensuring that the models do not discriminate against any particular group.

Contextual Understanding: Investigate the contextual factors that may influence an applicant's profile and potential for success. This may involve examining socioeconomic indicators, educational opportunities, or other relevant variables. Explore how to incorporate this contextual understanding into the predictive models to provide a more holistic assessment.

5 FLOW CHART


Import necessary libraries
Load and read Dataset
Data Visualization
Data Preprocessing
Handling null values
Perform Descriptive Analysis
Check and remove outliers
Dropping Unnecessary columns
Bar plot between GRE_Score and Chance_of_Admit
Bar plot between TOEFL_Score and Chance_of_Admit
Pair plot
Heat Map
Pie Chart
Split data into Train and Test
Training Data
Testing Data
Model Building
Multi linear Regression
Ridge Regression
Lasso Regression
Performance Evaluation
Choose Best model
Make pkl file and Deployment

6 RESULT

Train-Test Split

In [42]:
```
from sklearn.model_selection import train_test_split
```

In [43]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=9)
```

Build the Model

In [44]:
```
#Multi Linear Regression
from sklearn.linear_model import LinearRegression
model=LinearRegression()
```

In [45]:
```
#importing Ridge Regression
from sklearn.linear_model import Ridge
r=Ridge()
```

Train the model

In [46]:
```
#Training Linear Regression Model
model.fit(x_train.values,y_train.values)
```

Out[46]:
```
▾ LinearRegression
LinearRegression()
```

In [47]:
```
#Training Ridge Regression Model
r.fit(x_train,y_train)
```

Out[47]:
```
▾ Ridge
Ridge()
```

Testing the Model

In [48]:
```
#Testing Linear Regression Model
pred=model.predict(x_test)
```

```
C:\Users\rnt44\anaconda3\lib\site-packages\sklearn\base.py:413: UserWarning: X has feature names, but LinearRegression was fitt
ed without feature names
  warnings.warn(
```

In [49]:
```
#Testing Ridge Regression Model
pred1=r.predict(x_test)
```

Measure the performance using Metrics

In [50]:
```
from sklearn import metrics
```

In [51]:
```
# R Squared for MultiLinear Regression
print(metrics.r2_score(y_test,pred))
```

```
0.840435811146728
```

In [52]:
```
# R Squared for Ridge Regression
print(metrics.r2_score(y_test,pred1))
```

```
0.8327761441584973
```

In [53]:
```
#MSE (Mean Square Error)  for MultiLinear Regression
print(metrics.mean_squared_error(y_test,pred))
```

```
0.0032605346350277576
```

In [54]:
```
#MSE (Mean Square Error) for Ridge Regression
print(metrics.mean_squared_error(y_test,pred1))
```

```
0.0034170522702652666
```

In [55]:
```
#RMSE(Root Mean Square Error) for MultiLinear Regression
print(np.sqrt(metrics.mean_squared_error(y_test,pred)))
```

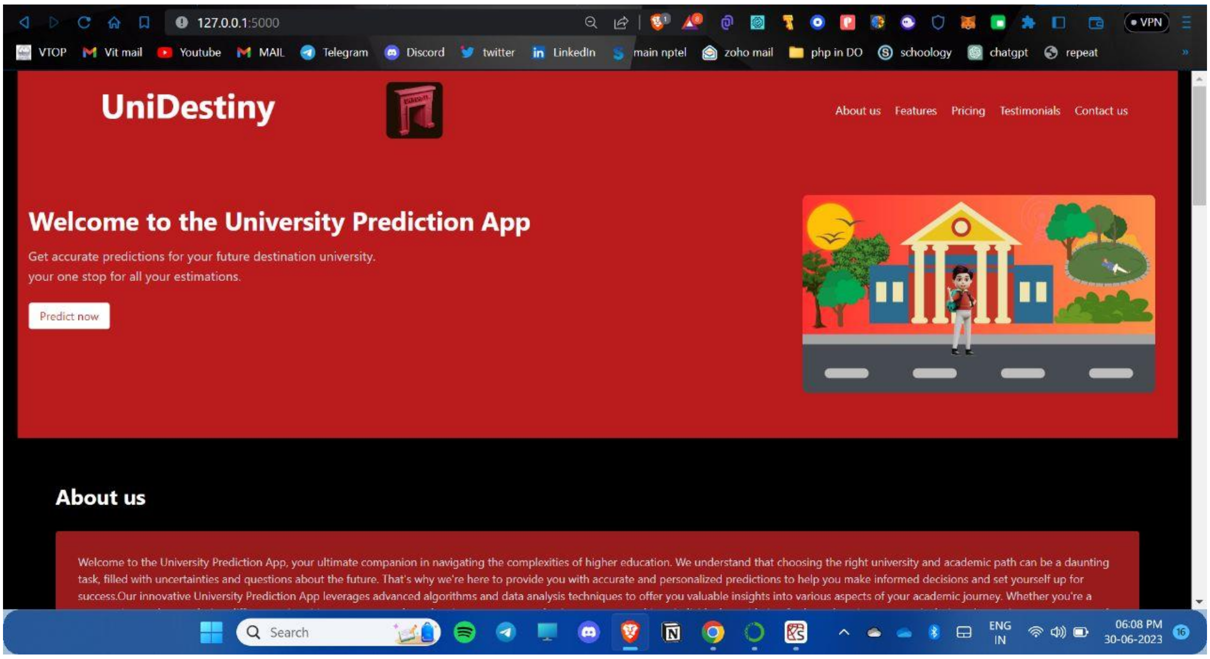
```
0.0571010913645944
```

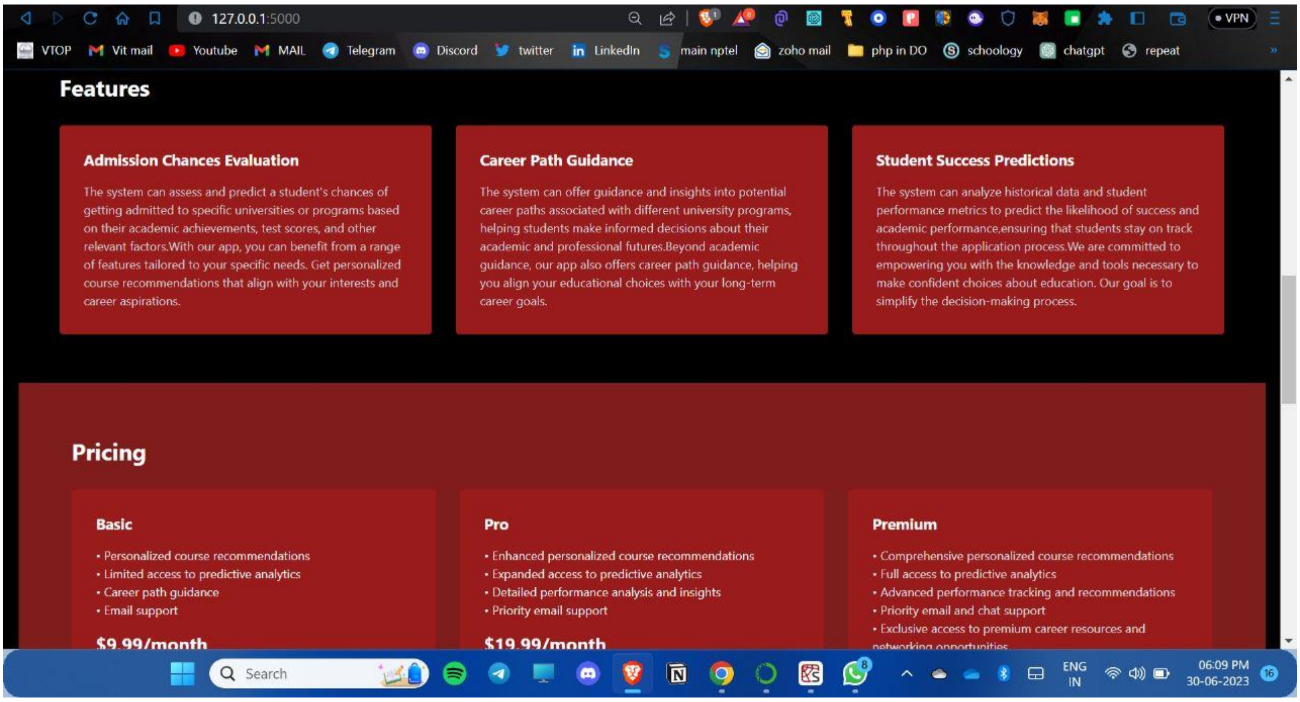
In [56]:
```
#RMSE(Root Mean Square Error) for Ridge Regression
print(np.sqrt(metrics.mean_squared_error(y_test,pred1)))
```

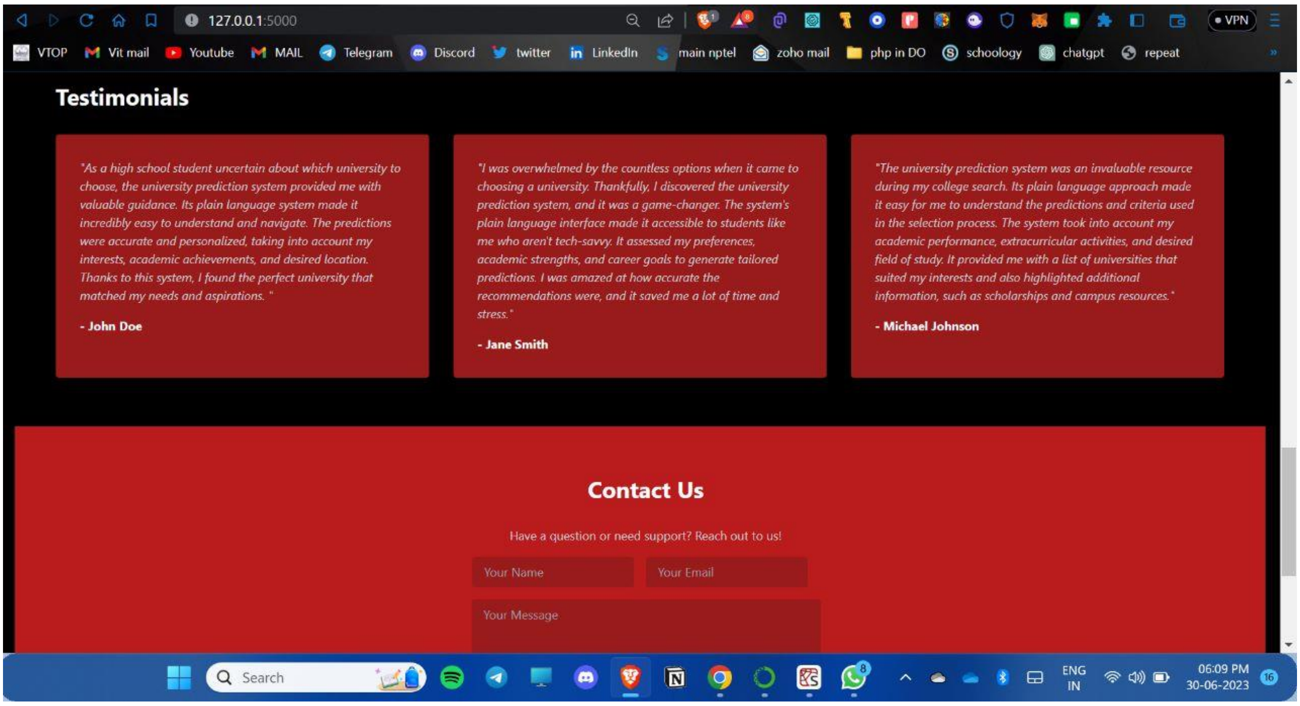
```
0.058455558078469035
```

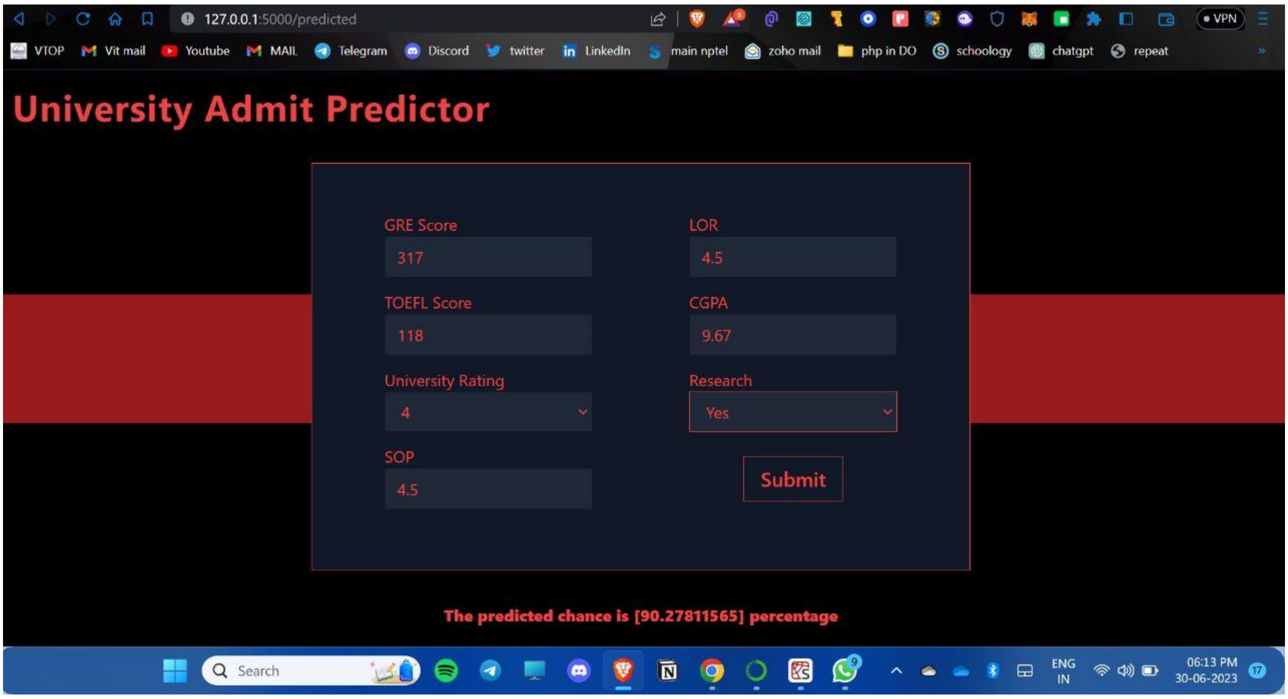
```
In [57]: def calc(a,b,c,d,e,f,g):
             myarr = np.array([a,b,c,d,e,f,g])
             mydf = pd.Series(myarr, index=["GRE_Score", "TOEFL_Score", "University_Rating", "SOP", "LOR", "CGPA", "Research"])
             removed_outliers.columns = removed_outliers.columns.astype(str)
             frames = [removed_outliers,mydf.to_frame().T]
             result = pd.concat(frames,ignore_index=True)
             mydfs = scale.fit_transform(result)
             ans = pd.DataFrame(mydfs)
             return ans[0][400],ans[1][400],ans[2][400],ans[3][400],ans[4][400],ans[5][400],ans[6][400]
         #    print(myans)
         #    return myans

In [58]: model.predict([calc(337,118,4,4.5,4.5,9.65,1)])

In [60]: calc(337,118,4,4.5,4.5,9.65,1)

Out[60]: (0.9400000000000004,
          0.9285714285714288,
          0.75,
          0.875,
          0.857142857142857,
          0.9007352941176472,
          1.0)

In [61]: model.predict([[337,118,4,4.5,4.5,9.65,1]])

Out[61]: array([[54.26945998]])

In [62]: model.predict([[337,118,4,4.5,4.5,9.65,1]])

Out[62]: array([[54.26945998]])

In [63]: model.predict([[0.94,0.928571,0.75,0.875,0.8571,0.900735,1]])

Out[63]: array([[0.94748759]])

In [64]: import pickle
         pickle.dump(model,open("model.pkl","wb"))
```

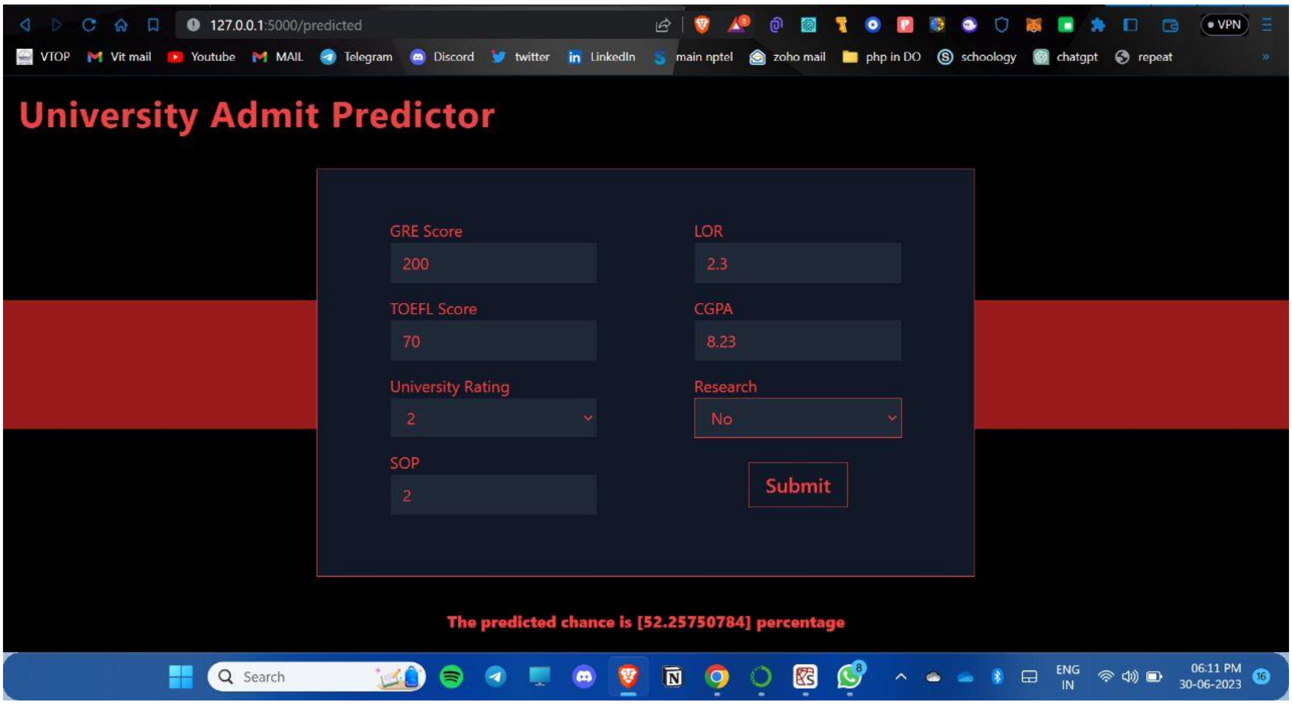
Output




127.0.0.1:5000
Features
Admission Chances Evaluation
The system can assess and predict a student's chances of getting admitted to specific universities or programs based on their academic achievements, test scores, and other relevant factors.With our app, you can benefit from a range of features tailored to your specific needs. Get personalized course recommendations that align with your interests and career aspirations.
Career Path Guidance
The system can offer guidance and insights into potential career paths associated with different university programs, helping students make informed decisions about their academic and professional futures.Beyond academic guidance, our app also offers career path guidance, helping you align your educational choices with your long-term career goals.
Student Success Predictions
The system can analyze historical data and student performance metrics to predict the likelihood of success and academic performance,ensuring that students stay on track throughout the application process.We are committed to empowering you with the knowledge and tools necessary to make confident choices about education. Our goal is to simplify the decision-making process.
Pricing
Basic
• Personalized course recommendations
• Limited access to predictive analytics
• Career path guidance
• Email support
$9.99/month
Pro
• Enhanced personalized course recommendations
• Expanded access to predictive analytics
• Detailed performance analysis and insights
• Priority email support
$19.99/month
Premium
• Comprehensive personalized course recommendations
• Full access to predictive analytics
• Advanced performance tracking and recommendations
• Priority email and chat support
• Exclusive access to premium career resources and
06:09 PM
30-06-2023


127.0.0.1:5000
Testimonials
"As a high school student uncertain about which university to choose, the university prediction system provided me with valuable guidance. Its plain language system made it incredibly easy to understand and navigate. The predictions were accurate and personalized, taking into account my interests, academic achievements, and desired location. Thanks to this system, I found the perfect university that matched my needs and aspirations."
- John Doe
"I was overwhelmed by the countless options when it came to choosing a university. Thankfully, I discovered the university prediction system, and it was a game-changer. The system's plain language interface made it accessible to students like me who aren't tech-savvy. It assessed my preferences, academic strengths, and career goals to generate tailored predictions. I was amazed at how accurate the recommendations were, and it saved me a lot of time and stress."
- Jane Smith
"The university prediction system was an invaluable resource during my college search. Its plain language approach made it easy for me to understand the predictions and criteria used in the selection process. The system took into account my academic performance, extracurricular activities, and desired field of study. It provided me with a list of universities that suited my interests and also highlighted additional information, such as scholarships and campus resources."
- Michael Johnson
Contact Us
Have a question or need support? Reach out to us!
Your Name
Your Email
Your Message
06:09 PM
30-06-2023


127.0.0.1:5000/predicted
University Admit Predictor
GRE Score
317
LOR
4.5
TOEFL Score
118
CGPA
9.67
University Rating
4
Research
Yes
SOP
4.5
Submit
The predicted chance is [90.27811565] percentage


127.0.0.1:5000/predicted
University Admit Predictor
GRE Score
200
LOR
2.3
TOEFL Score
70
CGPA
8.23
University Rating
2
Research
No
SOP
2
Submit
The predicted chance is [52.25750784] percentage

7 ADVANTAGES & DISADVANTAGES

Advantages of the proposed solution:

- **Improved Decision Support:** The solution provides decision support to admissions committees, helping them make more informed and objective decisions by considering both quantitative and qualitative factors.
- **Enhanced Transparency:** The solution promotes transparency by providing applicants with insights into their admission chances and the factors influencing those predictions. This transparency can help applicants make more informed decisions about where to apply.
- **Resource Optimization:** By predicting admission eligibility, the solution enables universities to allocate their resources more efficiently, focusing on highly qualified applicants and optimizing the admission process.

Disadvantages of the proposed solution:

- **Limited Predictive Accuracy:** The predictive models may not always accurately predict admission outcomes due to the inherent complexity of the admission process and the potential for subjective evaluation factors that are difficult to capture quantitatively.
- **Data Limitations:** The accuracy and fairness of the predictive models heavily rely on the quality and representativeness of the available data. Biases or missing data in the training dataset can impact the performance and fairness of the models.
- **Ethical Considerations:** The use of predictive models in admissions raises ethical concerns, such as privacy, transparency, and potential unintended consequences. It is important to carefully address and mitigate these ethical considerations to ensure responsible and accountable use of the solution.

8. APPLICATIONS

The proposed solution of university admission eligibility prediction can be applied in various areas within the realm of university admissions. Here are some key areas where this solution can be implemented:

Undergraduate Admissions: The solution can be applied to predict the eligibility and admission chances of prospective undergraduate students. It can help both applicants and universities in making informed decisions regarding admissions.

Graduate Admissions: Graduate programs, such as master's or doctoral programs, can benefit from this solution. It can assist in assessing the eligibility and admission prospects of applicants applying to specialized programs.

Program-Specific Admissions: The solution can be customized and applied to specific programs within a university. For example, it can be used to predict eligibility for engineering programs, business schools, or healthcare-related programs.

Scholarship and Financial Aid Allocation: The solution can aid in determining the eligibility and likelihood of receiving scholarships or financial aid. It can assist universities in allocating financial resources effectively based on predicted admission outcomes.

Waitlist Management: Universities often maintain waitlists for applicants who may be considered for admission if spots become available. The solution can help predict the likelihood of waitlisted applicants being admitted, aiding in waitlist management decisions.

International Admissions: The solution can be adapted to assess the eligibility and admission chances of international students applying to universities. It can help applicants evaluate their chances of admission and assist universities in international student recruitment efforts.

Enrolment Planning and Yield Prediction: Universities can utilize the solution to predict the yield rate—the number of admitted students who accept the offer of admission. This information can aid in enrolment planning, class size management, and ensuring a balanced student population.

Transfer Admissions: The solution can be applied to predict the eligibility and admission prospects of transfer students from community colleges or other institutions. It can assist both transfer applicants and universities in the evaluation process.

Recruitment and Outreach: The solution can aid universities in targeting and recruiting prospective students who have a higher probability of being admitted. It can help universities focus their outreach efforts on qualified candidates who are more likely to enroll.

Institutional Research and Analysis: Universities can use the solution to gather insights and perform data analysis on admission trends, applicant profiles, and the effectiveness of their admission criteria. It can contribute to institutional research efforts and inform strategic decisions.

9. Conclusion

In conclusion, the proposed solution of university admission eligibility prediction offers several benefits and considerations for improving the admission process. By incorporating both quantitative and qualitative factors, including contextual information. It enhances transparency by offering insights into admission chances and helps optimize resource allocation for universities.

However, it is important to acknowledge the limitations and challenges associated with predictive models in admissions. These include limited predictive accuracy, subjective factors challenges, potential biases in data, ethical considerations, and the complexity of incorporating contextual factors. These aspects need to be carefully addressed to ensure responsible and fair use of the solution.

Overall, the proposed solution has the potential to provide valuable decision support to admissions committees, help applicants make informed choices, and optimize the admission process. It is crucial to continuously monitor, evaluate, and adapt the solution based on feedback and evolving needs, ensuring its effectiveness, fairness, and ethical soundness.

10. Future Scope

In the future, several enhancements can be made to further improve the proposed solution of university admission eligibility prediction. Some potential enhancements include:

1. Refinement of Predictive Models: Continuously refining and updating the predictive models with new data and advanced techniques can improve their accuracy and predictive power. Incorporating state-of-the-art machine learning algorithms, ensemble methods, or deep learning architectures may lead to better performance.

2. Fine-grained Contextual Analysis: Enhancing the understanding and incorporation of contextual factors can provide a more nuanced assessment of applicants. This can involve exploring additional contextual variables, conducting in-depth analyses of their impact, and developing more sophisticated models to capture the interplay between contextual factors and admission outcomes.

3. Integration of Real-Time Data: Integrating real-time data, such as updated test scores or academic achievements, into the predictive models can ensure that the predictions are based on the most recent information. This requires establishing efficient data pipelines and implementing mechanisms to capture and process real-time data effectively.

4. Explainability and Interpretability: Enhancing the explainability and interpretability of the predictive models can help build trust and acceptance. Developing techniques to provide transparent explanations of the model's predictions, such as feature importance rankings or decision rules, can aid in understanding the factors influencing admission eligibility.

5. Dynamic Bias Mitigation: Implementing dynamic bias mitigation techniques that adapt to evolving societal changes and shifting biases can help ensure fairness in admission eligibility predictions. Regularly evaluating the models for potential biases and updating the fairness-aware algorithms accordingly can enhance their effectiveness in addressing bias.

7. Collaboration with Stakeholders: Fostering collaboration and engagement with stakeholders, including admissions committees, applicants, and experts in the field, can provide valuable insights and perspectives. Involving stakeholders in the development, evaluation, and improvement processes can ensure that the solution meets their needs and addresses their concerns effectively.

8. Expansion to Other Educational Levels: Extending the solution to other educational levels, such as high school admissions or professional certifications, can offer broader applications and benefits. Adapting the models and criteria to specific educational contexts can support fair and efficient admissions across various educational programs.

11 BIBLIOGRAPHY

1. Kruthika CS, Apeksha B, Chinmaya GR, Madhumathi JB, Veena MR (2021) University Admission Prediction using Machine Learning. Am J Glob J Res Rev Vol: 8 No:7.

2. DINO IENCO, RUGGERO G. PENSA and ROSA MEO, “From Context to Distance: Learning Dissimilarity for categorical Data Clustering,”

3. Hana Bydžovská. Course Enrollment Recommender System: Proceeding of the 9th International Conference on Educational Data Mining, P. 312 – 317.

4. Waters, A. and Miikkulainen, R. (2013). GRADE: Machine Learning Support for Graduate Admissions.

5. L. Breiman, Accuracy Predictors, Machine Learning, 24(2):123-140,1996.

APPENDIX

A. Source Code

Link:

https://drive.google.com/file/d/1LvEzBF4gDnjek4SO17A-KbLWG2mHgRq/view?usp=sharing