# Project Report

**Tittle:** ENVISIONING SUCCESS: Predicting University Scores using Machine Learning

**Team Members:**


- D.Surya Prakash Reddy
- K. Rishitha
- K. Aravind
- P. varsha


## 1.Introduction

### **1.1** PROJECT OVERVIEW

The University Score Prediction Web Application utilizes machine learning to forecast university scores based on parameters like education quality, alumni employment, faculty credentials, and research impact. Data cleaning, preprocessing, and feature engineering ensure model accuracy, validated through testing before deployment on hosting platforms. Continuous monitoring and maintenance guarantee ongoing relevance, offering a transparent and accountable tool for informed decision-making in higher education selection. Ultimately, the project seeks to empower users by simplifying the complex process of university selection, providing a quick, accurate, and accessible solution for making well-informed educational choices.

### **1.2** PURPOSE

The purpose of the University Score Prediction Web Application is to empower students and parents with a data-driven tool for making informed decisions about higher education. The project aims to predict university scores based on essential parameters such as education quality, alumni employment, faculty credentials, and research impact. The application simplifies the decision-making process, offering real-time score predictions. Ensuring that users receive accurate and relevant information to guide their educational journey.

## 2.LITERATURE SURVEY

### **2.1** EXISTING PROBLEM

The existing problem revolves around the lack of a comprehensive and transparent tool for individuals seeking higher education. Current university selection processes often rely on subjective factors, leading to uncertainty and uninformed decisions. Without a standardized metric, prospective students and parents face challenges in evaluating and comparing universities based on essential parameters such as education quality, alumni employment, faculty credentials, and research impact. This gap highlights the need for a data-driven solution that leverages machine learning to predict and assess university scores objectively, providing a clearer understanding of each institution's strengths and weaknesses.

### **2.2** REFERENCE

The development of the University Score Prediction Web Application is informed by relevant literature in the fields of machine learning, educational technology, and decision support systems. Key references include seminal works on predictive modeling in education, data-driven decision-making, and the integration of technology in the educational landscape. By drawing upon established research, the project aims to build a robust and effective tool that aligns with best practices and contributes to the ongoing discourse on enhancing the higher education selection process.

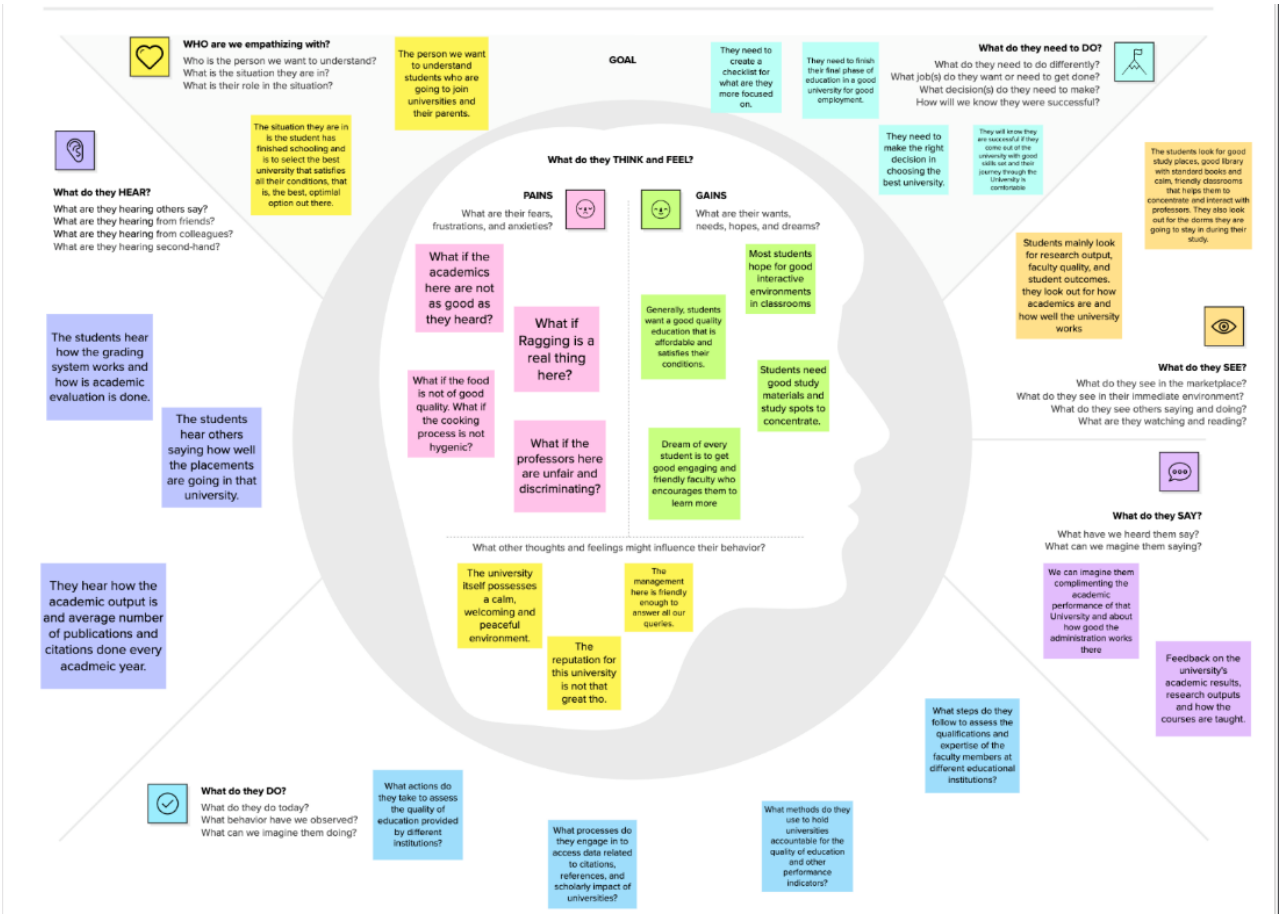
### **2.3** PROBLEM STATEMENT DEFINITION

The problem statement for this project is rooted in the absence of a standardized and transparent method for assessing and comparing universities. The challenge lies in providing prospective students and their families with a reliable tool that objectively predicts university scores based on critical parameters. This project seeks to address the deficiency in current decision-making processes by defining a clear problem statement: the need for a user-friendly web application that employs machine learning to predict and communicate university scores, fostering transparency, accountability, and informed decision-making in the higher education selection journey.

# 3.IDEATION AND PROPOSED SOLOUTION

## **3.1** EMPATHY MAP CANVAS

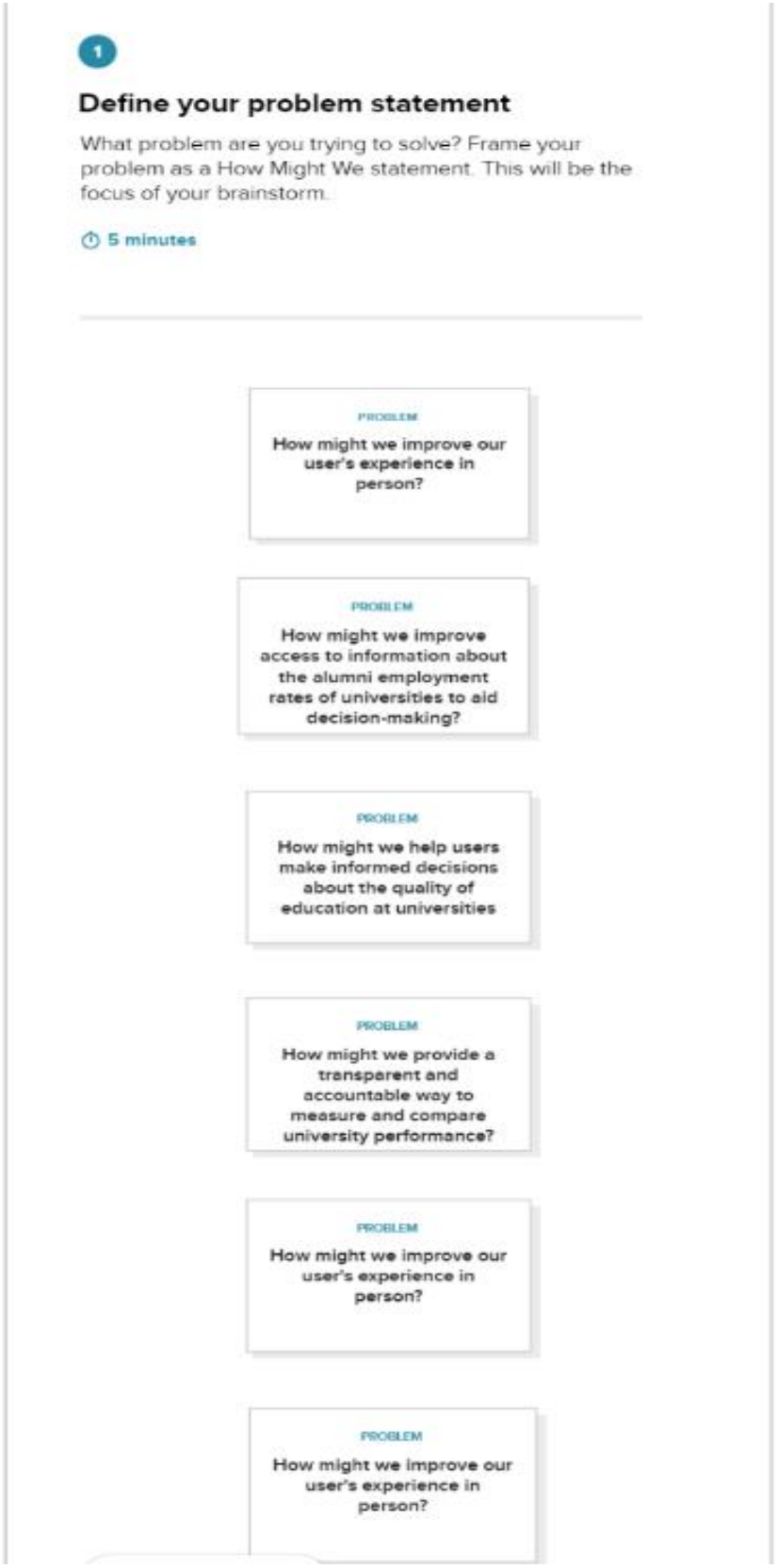
The Empathy Map Canvas was employed to gain a deep understanding of the users' perspectives and needs in the context of university selection. By empathizing with the experiences of students and parents, key insights were gathered regarding their desires for transparency, reliability, and ease of decision-making in the higher education selection process. The canvas highlighted the importance of objective information, user-friendly interfaces, and the desire for a tool that simplifies the complexity of assessing and comparing universities.




## **3.2** IDEATION AND BRAIN STROMING

During the ideation and brainstorming phase, several key features and functionalities for the University Score Prediction Web Application were conceptualized. Ideas included a clear and intuitive user interface for inputting university characteristics, real-time score predictions, and visual representations of data. Brainstorming sessions emphasized the importance of leveraging machine learning for accurate predictions, ensuring transparency in the scoring methodology,

and providing users with personalized recommendations based on their preferences. The ideation process focused on creating a solution that not only addresses the existing problem but also enhances the user experience by delivering valuable, actionable insights for informed decision-making in higher education.



2

## Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

TIP
You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

### Person 1

- Develop an AI-driven advisor to recommend universities based on individual goals and preferences.
- Establish mentorship programs
- A mobile app for easy university quality checks.
- Allow students and parents to submit reviews and feedback on their university experiences, giving prospective students valuable insights

### Person 2

- Create a dashboard that tracks universities' performance over time, holding them accountable for improvements in various areas.
- Quality Reports: Publish yearly university quality reports.
- develop a model based on measurable categories in order to predict the outcome
- Provide a tool to evaluate the effectiveness of financial aid and scholarships offered.

### Person 3

- Conduct workshops for parents and students to educate them on how to evaluate educational institutions effectively.
- Build a tool that predicts a student's likelihood of securing a job upon graduation.

### Person 4

- Advocate for government agencies to publish annual performance rankings of universities to create more transparency.
- Create an online platform for students and alumni to review and rate universities based on their experiences.
- Develop a tool that assesses a university's research impact through publications, citations, and patents, making it transparent for users.

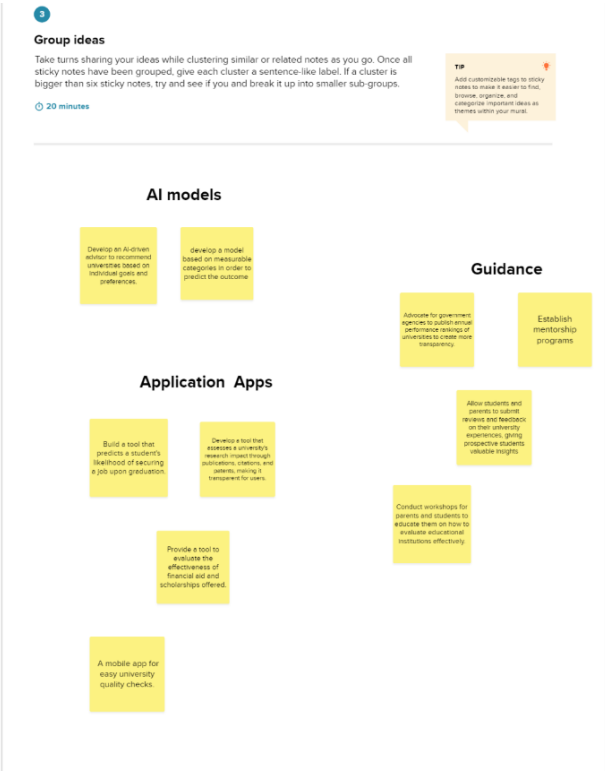
3

## Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

TIP
Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.



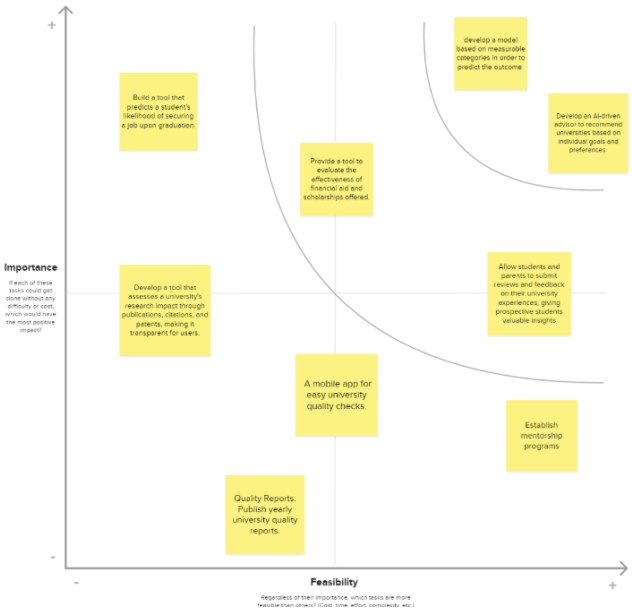

4

## Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes

TIP
Participants can use their cursors to point at where sticky notes should go on the grid. The facilitator can confirm the spot by using the laser pointer holding the **H key** on the keyboard.



# 4.REQUIRMENT ANALYSIS

## **4.1** FUNDEMENTAL REQUIREMENTS

Input Interface:
Develop an intuitive interface allowing users to input university characteristics, such as education quality, alumni employment rates, faculty credentials, publications, influence, citations, and patents.

Machine Learning Model Integration:
Integrate a machine learning model to predict university scores based on the provided characteristics, ensuring accuracy and real-time predictions.

Score Presentation:
Display the predicted university score to users in a clear and comprehensible manner, accompanied by an explanation of the scoring methodology.

## **4.2 NON-FUNCTIONAL REQUIRMENTS**

Performance:
Optimize the application for fast response times, even with concurrent user interactions, to enhance the overall user experience.

Usability:
Prioritize a user-friendly interface with clear instructions, minimalistic design, and seamless navigation for users of varying technical proficiency.

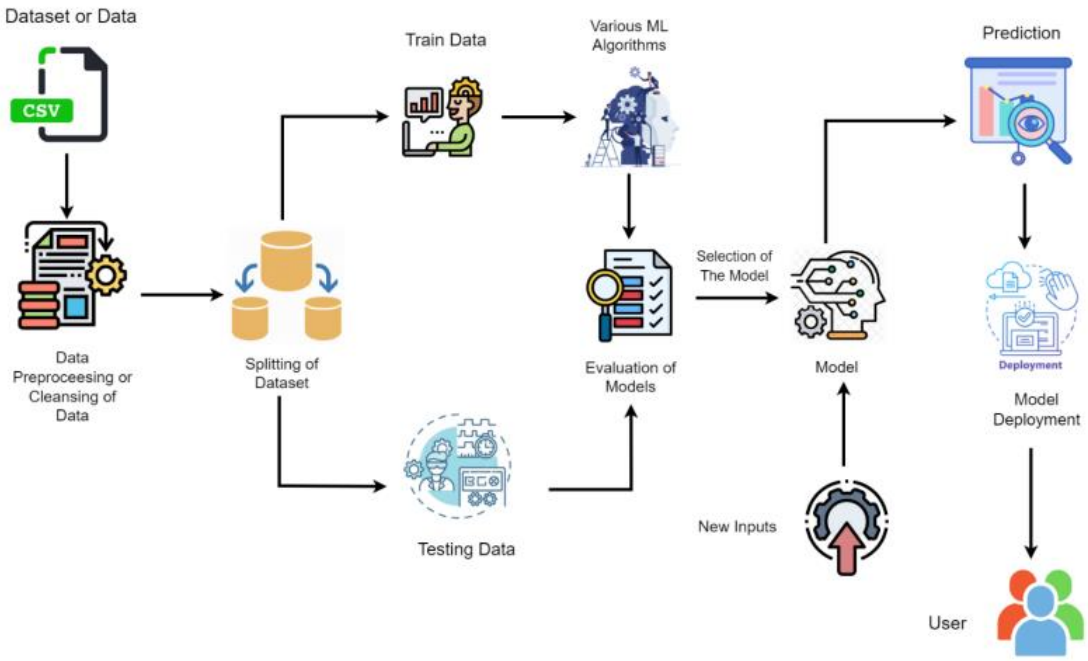
Reliability:
Ensure the reliability of the machine learning model by periodic updates and monitoring to maintain accuracy with evolving data.

# 5.PROJECT DESIGN

## **5.1** DATA FLOW DIAGRAM &USER STORIES

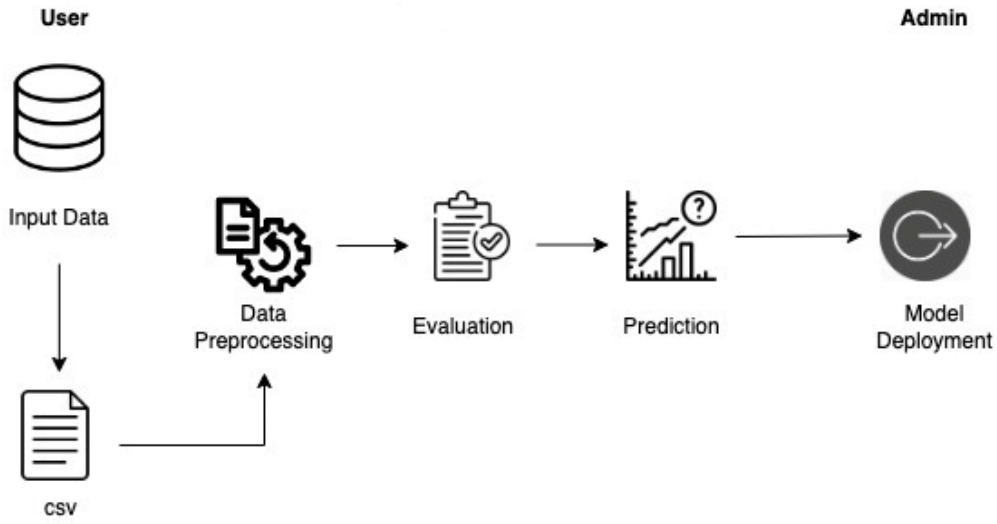
| User Type | Functional Requirement (Epic) | User Story Number | User Story / Task | Acceptance criteria | Priority | Release |
|---|---|---|---|---|---|---|
| Administrator | Model/Algorithm Building | USN-1 | Data Pre-processing and exploratory analysis | Cleaning of the data so as to achieve high accuracy. | High | Sprint-1 |
| | | USN-2 | Use various ML models on the same data. | Developing several models to find the model with best performance. | High | Sprint-1 |
| | | USN-3 | Evaluation of the Model and choosing the model with best performance. | Testing the models and selection of model with highest accuracy. | High | Sprint-2 |
| | Model Deployment. | USN-4 | Saving the best model | Yielding accurate results | Medium | Sprint-2 |
| | | USN-9 | Integrating the model with web framework | Proper Model Deployment. | High | Sprint-4 |
| | | USN-10 | Running the web application | Creating interactive user interface. | High | Sprint-4 |
| Student/Parents | Introductory Page | USN-5 | As a user, I need to know about what the website is all about in general. | Need to know about the purpose of the website. | Medium | Sprint-3 |
| | Dashboard | USN-6 | As a user, I want the dashboard to provide certain criteria to take into consideration. | I see options to give my conditions for university selection | High | Sprint-3 |
| | Data Entry and Submission | USN-7 | As a user, I want to give input of my criteria and each field is mandatory. | I can fill in my requirements of that criteria and submit them. I can submit only after filling all field. | High | Sprint-3 |
| | Result Display | USN-8 | As a user, I want to display the University's score. | The result given is almost suitable to the user conditions. | High | Sprint-3 |

**5.2** SOLUTION ARCHITECTURE




# 6.PROJECT PLANNING & SCHEDULING

**6.1** TECHNICAL ARCHITECTURE




**6.2** SPRINT PLANNING & ESTIMATION & **6.3** SPRINT DELIVERY SCHEDULE

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-1 | Model/Algorithm Building | USN-1 | Data Pre-processing and exploratory analysis | 6 | High | Varsha, Surya |
| Sprint-1 | | USN-2 | Use various ML models on the same data. | 6 | High | Rishitha, Varsha |
| Sprint-2 | | USN-3 | Evaluation of the Model and choosing the model with best performance. | 5 | High | Aravind, Rishitha |
| Sprint-2 | Model Deployment | USN-4 | Saving the best model | 2 | Medium | Surya |
| Sprint-3 | Introductory Page | USN-5 | As a user, I need to know about what the website is all about in general. | 3 | Medium | Aravind, Varsha |
| Sprint-3 | Dashboard | USN-6 | As a user, I want the dashboard to provide certain criteria to take into consideration | 3 | High | Rishitha, Aravind |
| Sprint-3 | Data Entry and Submission | USN-7 | As a user, I want to give input of my criteria and each field is mandatory. | 3 | High | Aravind, Surya |
| Sprint-3 | Result Display | USN-8 | As a user, I want to display the University's score. | 3 | High | Surya, Rishitha |
| Sprint-4 | Model Deployment | USN-9 | Integrating the model with web framework | 8 | High | Surya, Aravind, Varsha |
| Sprint-4 | Model Deployment | USN-10 | Running the web application | 4 | High | Varsha, Rishitha |

| Sprint | Total Story Points | Duration | Sprint Start Date | Sprint End Date (Planned) | Story Points Completed (as on Planned End Date) | Sprint Release Date (Actual) |
|---|---|---|---|---|---|---|
| Sprint-1 | 12 | 4 Days | 30 Oct 2023 | 02 Nov 2023 | 11 | 4 Nov 2023 |
| Sprint-2 | 7 | 4 Days | 03 Nov 2023 | 06 Nov 2022 | 9 | 7 Nov 2023 |
| Sprint-3 | 12 | 7 Days | 07 Nov 2022 | 13 Nov 2022 | 11 | 14 Nov 2023 |
| Sprint-4 | 12 | 7 Days | 14 Nov 2022 | 20 Nov 2022 | 11 | 20 Nov 2023 |

# 7.CODING & SOLUTIONING (Explain the features added in project along with the code)

## 7.1FEATURES

The features in our project are Quality of education, Alumni Employment, Quality of faculty, Publications, Influence, Citations, and Patents.

| | quality_of_education | alumni_employment | quality_of_faculty | publications | influence | citations | patents | score |
|---|---|---|---|---|---|---|---|---|
| count | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 |
| mean | 275.100455 | 357.116818 | 178.888182 | 459.908636 | 459.797727 | 413.417273 | 433.346364 | 47.798395 |
| std | 121.935100 | 186.779252 | 64.050885 | 303.760352 | 303.331822 | 264.366549 | 273.996525 | 7.760806 |
| min | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 43.360000 |
| 25% | 175.750000 | 175.750000 | 175.750000 | 175.750000 | 175.750000 | 161.000000 | 170.750000 | 44.460000 |
| 50% | 355.000000 | 450.500000 | 210.000000 | 450.500000 | 450.500000 | 406.000000 | 426.000000 | 45.100000 |
| 75% | 367.000000 | 478.000000 | 218.000000 | 725.000000 | 725.250000 | 645.000000 | 714.250000 | 47.545000 |
| max | 367.000000 | 567.000000 | 218.000000 | 1000.000000 | 991.000000 | 812.000000 | 871.000000 | 100.000000 |

**Quality of Education:**

This feature evaluates the university's overall educational experience. It considers factors such as curriculum quality, teaching methodologies, student-to-faculty ratio, and graduation rates.
**Significance:** A high-quality education feature indicates that the university provides a robust and effective learning environment.

**Alumni Employment:**

Measures the success of a university's graduates in the job market. Reflects the employability and success of alumni after completing their education.
Significance: Higher alumni employment rates suggest that the university equips students with skills and knowledge relevant to the job market.

**Quality of Faculty:**

Evaluates the credentials, expertise, and research contributions of the university's faculty members. Highlights the caliber of instructors and their impact on students' learning experiences.

**Significance:** A strong faculty contributes to a university's academic reputation and the quality of education it provides.

## Publications:

Counts and assesses the academic publications produced by the university's faculty and researchers. Demonstrates the institution's commitment to research and contributions to academic knowledge.
**Significance**: High publication rates indicate a vibrant research culture and may enhance the university's reputation.

## Influence:

Measures the impact and influence of the university's research on the broader academic community. Indicates the significance of the research conducted by the institution.
**Significance:** A university with high influence is likely to contribute groundbreaking research and innovations.

## Citations:

Counts the number of times the university's research publications are cited by other researchers. Reflects the impact and relevance of the university's research in the academic community.
**Significance:** Higher citation counts suggest that the university's research is widely recognized and respected.

## Patents:

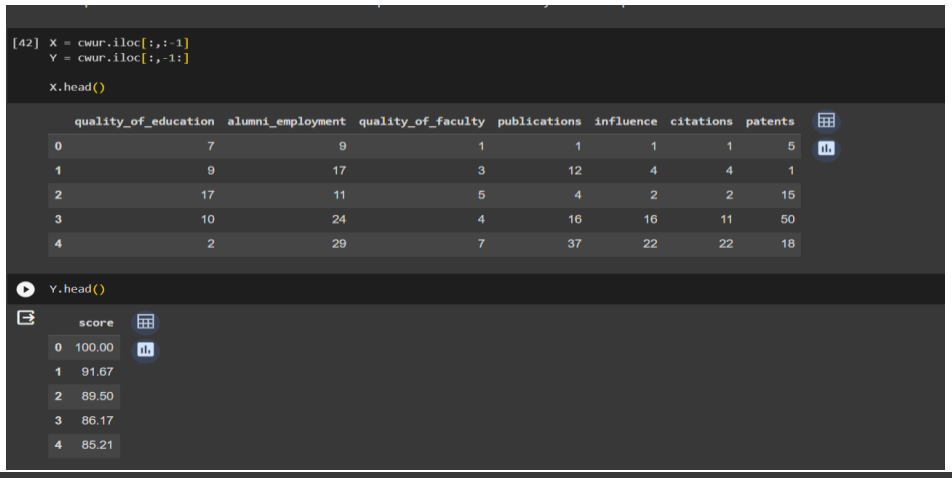
Measures the number of patents granted based on the university's research and innovations. Reflects the institution's contributions to technological advancements and innovation.
**Significance:** A high number of patents may indicate a strong focus on research with practical applications.

**Code:**

Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features

```
[31] cwur.describe(include="all")
```

| | quality_of_education | alumni_employment | quality_of_faculty | publications | influence | citations | patents | score |
|---|---|---|---|---|---|---|---|---|
| count | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 |
| mean | 275.100455 | 357.116818 | 178.888182 | 459.908636 | 459.797727 | 413.417273 | 433.346364 | 47.798395 |
| std | 121.935100 | 186.779252 | 64.050885 | 303.760352 | 303.331822 | 264.366549 | 273.996525 | 7.760806 |
| min | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 43.360000 |
| 25% | 175.750000 | 175.750000 | 175.750000 | 175.750000 | 175.750000 | 161.000000 | 170.750000 | 44.460000 |
| 50% | 355.000000 | 450.500000 | 210.000000 | 450.500000 | 450.500000 | 406.000000 | 426.000000 | 45.100000 |
| 75% | 367.000000 | 478.000000 | 218.000000 | 725.000000 | 725.250000 | 645.000000 | 714.250000 | 47.545000 |
| max | 367.000000 | 567.000000 | 218.000000 | 1000.000000 | 991.000000 | 812.000000 | 871.000000 | 100.000000 |



```
[42] X = cwur.iloc[:,:-1]
     Y = cwur.iloc[:,-1:]

     X.head()
```

| | quality_of_education | alumni_employment | quality_of_faculty | publications | influence | citations | patents |
|---|---|---|---|---|---|---|---|
| 0 | 7 | 9 | 1 | 1 | 1 | 1 | 5 |
| 1 | 9 | 17 | 3 | 12 | 4 | 4 | 1 |
| 2 | 17 | 11 | 5 | 4 | 2 | 2 | 15 |
| 3 | 10 | 24 | 4 | 16 | 16 | 11 | 50 |
| 4 | 2 | 29 | 7 | 37 | 22 | 22 | 18 |

```
Y.head()
```

| | score |
|---|---|
| 0 | 100.00 |
| 1 | 91.67 |
| 2 | 89.50 |
| 3 | 86.17 |
| 4 | 85.21 |

```
[44] Y.shape

     (2200, 1)
```

Here x and y variables are created. On x variable, features are passed. And on y target variable is passed. For splitting training and testing data we are using train_test_split() function from sklearn. As parameters, we are passing x, y, test_size, random_state.

```
[45] X_train,X_test,Y_train,Y_test=train_test_split(X,Y,train_size=0.8,random_state=42)
```

**Flask web integration:**

**APP.PY file:**

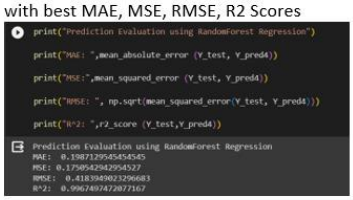
```
from flask import Flask, render_template, request
import joblib

app = Flask(__name__, static_folder='static')

model = joblib.load('model.pkl')

@app.route('/')
def front():
    return render_template('home.html')

@app.route('/predict')
def predict():
    return render_template('predict.html')

@app.route('/submit', methods=['POST'])
def submit():

    qoe = float(request.form['qoe'])
    ae = float(request.form['ae'])
    qof = float(request.form['qof'])
    pu = float(request.form['pu'])
    ni = float(request.form['ni'])
    ci = float(request.form['ci'])
    pa = float(request.form['pa'])
```

```
    input_data = [[qoe, ae, qof, pu, ni, ci, pa]]
    prediction = model.predict(input_data)

    return render_template('submit.html', prediction=prediction[0])

if __name__ == '__main__':
    app.run(debug=True)
```

# 8.PERFORMANCE TESTING

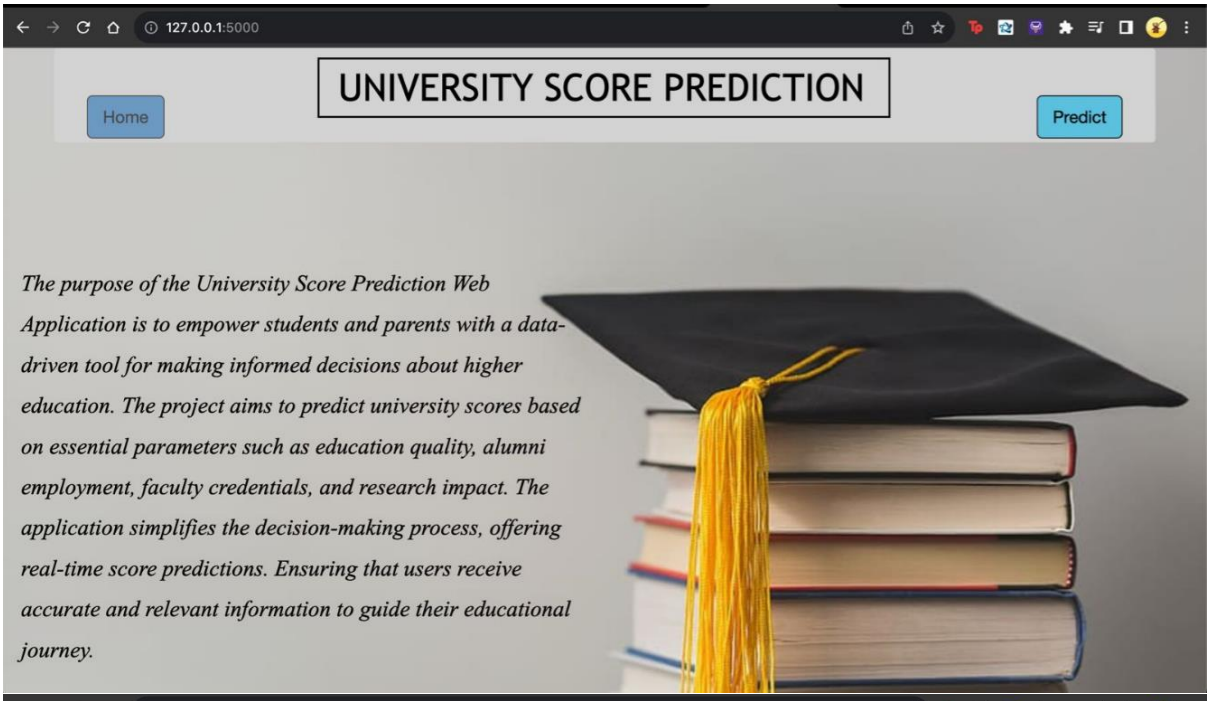
## **8.1** PERFORMANCE METRICS

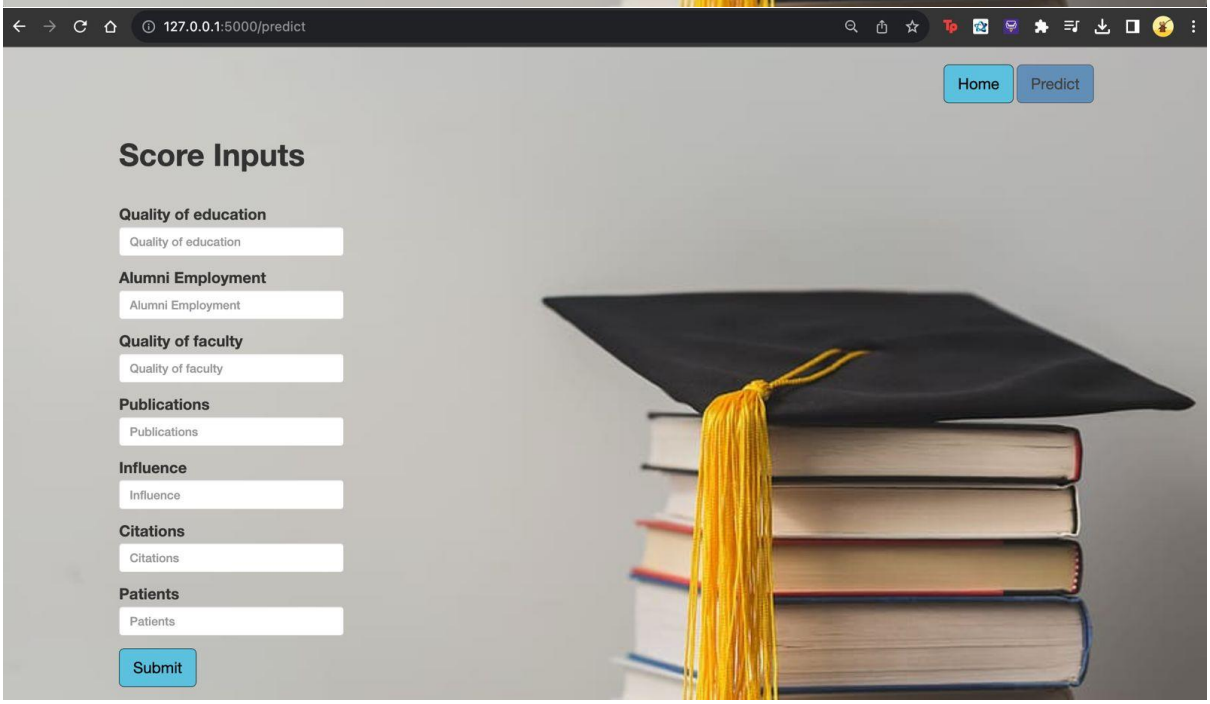
**Model Performance Testing:**.

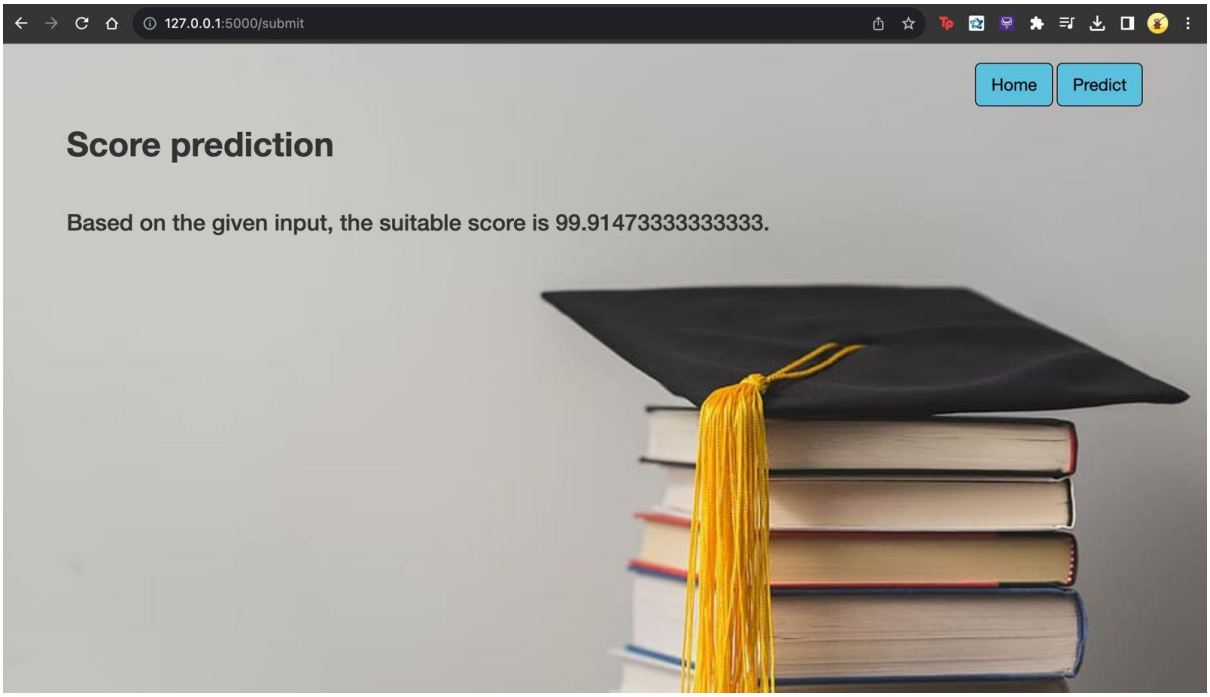
| S.No. | Parameter | Values | Screenshot |
|---|---|---|---|
| 1. | Metrics | **Regression Model:**<br>MAE - , MSE - , RMSE - , R2 score - | After performing all the Regression Models we found the best model as Random Forest Regression with best MAE, MSE, RMSE, R2 Scores<br>  |
| 2. | Tune the Model | Hyperparameter Tuning -<br>Validation Method - | GridSearch<br>  |

# 9.RESULTS

## **9.1** OUTPUT SCREENSHOTS


127.0.0.1:5000
UNIVERSITY SCORE PREDICTION
Home
Predict
The purpose of the University Score Prediction Web Application is to empower students and parents with a data-driven tool for making informed decisions about higher education. The project aims to predict university scores based on essential parameters such as education quality, alumni employment, faculty credentials, and research impact. The application simplifies the decision-making process, offering real-time score predictions. Ensuring that users receive accurate and relevant information to guide their educational journey.


127.0.0.1:5000/predict
Home
Predict
Score Inputs
Quality of education
Quality of education
Alumni Employment
Alumni Employment
Quality of faculty
Quality of faculty
Publications
Publications
Influence
Influence
Citations
Citations
Patients
Patients
Submit




# 10.ADVANTAGES & DISADVANTAGES

**ADVANTAGES:**

- **Informed Decision-Making:** The University Score Prediction Web Application empowers users by providing accurate and objective predictions, facilitating informed decisions in the complex process of university selection.
- **Transparency:** The use of machine learning models and clear visualization techniques enhances transparency in the scoring methodology, allowing users to understand how predictions are derived.
- **User-Friendly Interface:** The application offers a user-friendly interface that simplifies the input process, making it accessible to users with varying levels of technical proficiency.
- **Personalized Recommendations:** The recommendation engine provides users with personalized suggestions based on their preferences, contributing to a more tailored and relevant user experience.
- **Continuous Improvement:** The incorporation of a user profile and history feature allows individuals to track their predictions over time, fostering continuous improvement and learning.

**DISADVANTAGE:**

- **Dependency on Data Quality:** The accuracy of predictions heavily relies on the quality and relevance of the dataset used to train the machine learning model. Inaccurate or biased data may lead to unreliable predictions.
- **Overemphasis on Quantifiable Metrics:** While the application considers essential parameters like education quality and faculty credentials, it may not fully capture qualitative aspects such as campus culture, extracurricular activities, and student experiences.
- **Security Concerns:** Handling user authentication and storing personal data pose potential security risks. Robust security measures are essential to protect user information from unauthorized access or breaches.
- **Limited Scope:** The application focuses on quantifiable factors and may not address the entirety of factors influencing a user's decision. It's crucial for users to consider additional qualitative aspects when choosing a university.

- **Machine Learning Complexity:** The integration of machine learning, while providing accurate predictions, introduces complexity. Users may find it challenging to comprehend the inner workings of the model, potentially leading to mistrust or scepticism.

# 11.CONCLUSION

In summary, college grade prediction web applications address the urgent need for students and parents to make informed decisions about higher education. The project helps users ensure quality education by evaluating key factors such as teaching quality, graduate employment, faculty qualifications, and research impact. This application promotes transparency and accountability and aligns universities with performance standards. Ultimately, it serves as a valuable tool that guides users to institutions that achieve their educational goals and contributes to a more responsible and responsive higher education environment.

# 12.FUTURE SCOPE

**User Feedback Integration:**
Allow users to provide feedback on the predicted scores. This feedback can be used to improve the model and enhance its accuracy based on the experiences of students and alumni.

**Mobile Application Development:**
Extend the project by developing a mobile application, making it more accessible to a wider audience. A mobile app can provide convenience for users to access university scores on the go.

**Global Collaboration:**
Consider expanding the project's scope globally. Collaborate with educational institutions worldwide to create a more comprehensive and globally relevant scoring system.

**Personalized Recommendations:**
Implement a recommendation system that provides personalized suggestions to users based on their preferences, career goals, and academic interests.

**Ethical Considerations:**
Address ethical considerations related to data privacy, security, and fairness. Ensure that the system promotes transparency and fairness in evaluating universities.

# 13. APPENDIX

**Source Code**

https://colab.research.google.com/drive/1Gz6uHhNgqgBDk1ygMvUwjNs63MNToNkd?usp=sharing

**GitHub & Project Demo Link**

https://github.com/smartinternz02/SI-GuidedProject-609363-1698643834