Tribhuvan University

Faculty of Humanities and Social Sciences

Q&A website with Student Grade Prediction

A Project Report

Submitted to:

Department of Computer Application

Padmashree International College

In partial fulfilment of the requirements for the Bachelor in Computer Application

Submitted by:

Shekhar Ghimire

Registration No:6-2-622-38-2018

November, 2023

Under the Supervision of

Basanta Chapagain

Tribhuvan University

Faculty of Humanities and Social Sciences

Padmashree International College

SUPERVISOR'S RECOMMENDATION

I hereby recommend that this project prepared under my supervision by **Shekhar Ghimire** entitled "**Q&A Website with Student Grade Prediction**" in partial fulfilment of the requirements for the degree of Bachelor of Computer Application is recommended for the final evaluation.

....

Mr. Basanta Chapagain
Project Supervisor
Department of Computer Application
Padmashree International College

Tribhuvan University

Faculty of Humanities and Social Sciences

Padmashree International College

LETTER OF APPROVAL

This is to certify that this project prepared by **Shekhar Ghimire** entitled "**Q&A website with Student Grade Prediction**" in partial fulfillment of the requirements for the degree of Bachelor in Computer Application has been evaluated. In our opinion it is satisfactory in the scope and quality as a project for the required degree.

…………………………. Mr. Basanta Chapagain Project Supervisor Department of Computer Application Padmashree International College	……………………………… Mr. Ramesh Kumar Pudasaini BCA Coordinator Department of Computer Application Padmashree International College
…………………………. Mr. Internal Examiner	………………………………. Name: External Examiner

ABSTRACT

In the field of education and academic advancement, technology continues to shape and enhance the learning experience. This abstract presents two innovative systems that provides modern technology to serve the educational community effectively. The Student Grade Prediction System, developed on the Django framework, employs advanced machine learning techniques, with a focus on Linear Regression. This system is designed to accurately predict student academic performance, a crucial task for both educational institutions and learners. It streamlines the process of handling substantial amounts of student data, offering a seamless user experience to anticipate and comprehend potential academic outcomes. Alongside the Student Grade Prediction System, the Question-and-Answer Website creates a collaborative learning environment. This platform encourages the exchange of knowledge and insights by enabling users to ask questions and receive answers from a community of peers and experts. It fosters academic growth, encourages curiosity, and provides a space where learners can explore and expand their knowledge horizons. These systems, operating in harmony, exemplify the power of technology in revolutionizing education, both in terms of academic performance prediction and facilitating collaborative learning. They represent the future of education by merging data-driven insights and collaborative knowledge sharing.

Keywords: Student grade prediction, Django, Algorithm, Question and Answer Website.

ACKNOWLEDGEMENT

We would like to take this moment to express our heartfelt gratitude to the Department of Computer Application at Tribhuvan University for granting us the opportunity to delve into our passion and ideas within the realm of computer software through our "**Project-III**." It is with immense pleasure that we extend our deepest appreciation to our esteemed guide, Mr. **Basanta Chapagain** sir, who serves as the **Supervisor at Padmashree International College**. His unwavering guidance, encouragement, and invaluable assistance were pivotal in the successful completion of this project naming "**Q&A website with student grade prediction**". We sincerely acknowledge his useful insights and cooperative demeanour.

We also wish to extend our gratitude to Mr. **Ramesh Kumar Pudasaini** sir, the **Coordinator of the Department of Computer Application at Padmashree International College**, for his wholehearted support throughout this journey. Moreover, we are indebted to the IT officer of Padmashree International College, who generously shared his time and expertise, contributing significantly to the success of this study.

Equally, we extend our heartfelt thanks to our dear friends whose unwavering support has been a constant source of motivation throughout the preparation of this report. Additionally, we extend our sincere appreciation to all individuals, both visible and behind the scenes, who played a role in the development of this project. We eagerly anticipate hearing any feedback and welcome suggestions for further improvements. Your insights are highly valued and will be instrumental in our ongoing pursuit of excellence.

TABLE OF CONTENTS

LIST OF ABBREVIATION

CSS	Cascading style sheet
Html	Hypertext transfer protocol
JS	JavaScript
Q&A	Question and Answer

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Introduction

The Question-and-Answer Website with Student Grade Prediction is a forward-thinking online platform tailored to the dynamic educational landscape. It leverages the power of advanced machine learning techniques to offer two distinct but interconnected functionalities. First, it provides a collaborative space where users can seek and share knowledge through questions and answers, fostering an interactive learning environment. Second, it incorporates a robust Student Grade Prediction System, developed on the Django framework and powered by Linear Regression, to forecast student academic performance.

In a world where digital platforms play a pivotal role in education, this website stands out by addressing the vital needs of students and educators alike. The Question-and-Answer section encourages knowledge exchange and community building, promoting effective learning. Simultaneously, the Student Grade Prediction System offers a data-driven approach to anticipate and understand academic outcomes, significantly benefiting educational institutions. This innovative system operates at the intersection of data science, education technology, and collaborative learning, enhancing the educational experience for students and institutions alike. By combining the efficiency of a question-and-answer platform with the insights provided by student grade prediction, this website represents the future of holistic education support.

1.2 Problem Statement

Nepal's education system faces significant challenges, including limited resources, inadequate infrastructure, and a shortage of qualified teachers. These issues result in a lack of data-driven tools for understanding and supporting individual student progress, leading to high dropout rates and limited opportunities. To address these challenges, there is a critical need for a student grade prediction and analysis system that uses advanced technology and data analytics to provide personalized feedback and insights, ultimately improving the quality of education and supporting the success of all students.

1.3 Objectives

- To provide precise forecasts of student academic performance by leveraging machine learning algorithms, such as Linear Regression, Decision Trees, and Random Forest, based on comprehensive data analysis.

- To offer a user-friendly interface through a web-based platform developed with Django, facilitating seamless data input and efficient grade predictions for educational institutions and students.

1.4 Scope and Limitation

1.4.1 Scope

The Question-and-Answer website with student grade prediction aims to provide a comprehensive platform for educational questions and academic performance forecasting. It allows users to ask and answer questions on various educational topics while integrating a student grade prediction feature. The system's objectives include offering insights into student performance trends, assisting educators in developing intervention strategies, and enabling informed decision-making by educational institutions. It operates at both individual and group levels, providing personalized predictions and analysing performance patterns across student cohorts.

1.4.2 Limitation

Despite its capabilities, the system has limitations. The accuracy of student grade predictions relies on the quality and quantity of available data, making incomplete or insufficient data a potential limitation. Additionally, the model may not capture all factors affecting student performance, such as personal circumstances or dynamic changes in education. The diverse nature of students can also challenge accurate predictions, as some may not fit predefined models. Privacy and security concerns must be addressed to protect sensitive student data adequately. These limitations should be considered when utilizing the system for educational insights and predictions.

1.5 Development Methodology

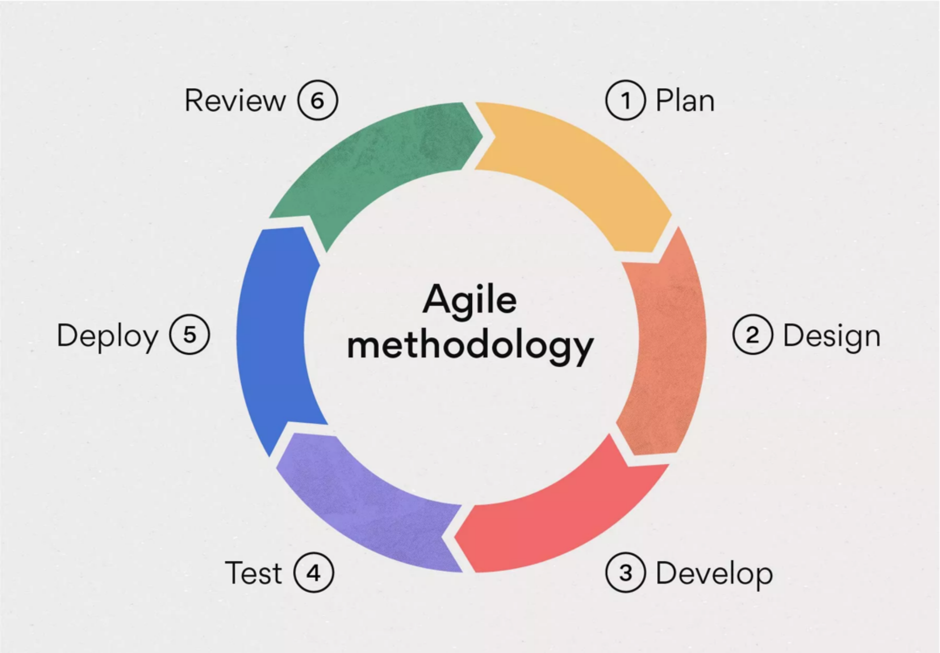


Figure1.1:Agile Methodology [1]

The Agile development methodology is found to be suitable for the Question-and-Answer website with student grade prediction. In Agile, the project is divided into short development cycles called sprints, each lasting a few weeks. Features are developed incrementally, with continuous user feedback and testing. This approach allows for flexibility and early delivery of a minimum viable product. The website evolves over time, integrating the machine learning model step by step. Continuous integration and deployment to AWS ensure regular updates. The project adapts based on feedback, and scaling and optimization are ongoing processes. Maintenance and further development continue to meet evolving needs.

1.6 Report Organization

Chapter 1: Introduction

This chapter includes introduction, problem statement, objectives, scope and report organization which identifies and focuses on core ideology of the working system of Student Grade and Performance Prediction.

Chapter 2: Background Study and Literature Review

This chapter includes the information relating project similar to project that is being developed and mentioning their work as a reference.

Chapter 3: System analysis and Design

This chapter includes the explanation of various diagrams like sequence diagram, class and object diagram, activity diagram, component diagram, deployment diagram providing the information on how the system works after the development.

Chapter 4: Implementation and Testing

This chapter includes the explanation on how the algorithms are implemented in core and provides the test cases used for validating and verifying the codes alongside the tools and technology used for the development of the system.

Chapter 5: Conclusion and Future Recommendation

This chapter includes the explanation of the things learned after the completion of project and describes future enhancement for the project in short.

Chapter 2: Background Study and Literature Review

2.1 Background Study

In a study conducted by Kumar and Pai in 2019, the primary objective was to evaluate the performance of different machine learning algorithms in the context of predicting student grades. The study utilized a dataset comprising 300 undergraduate students from an Indian university. This dataset included a wide range of information, such as demographic data, prior academic records, and examination scores. To fulfill their objective, the authors implemented five distinct machine learning algorithms: Naive Bayes, k-Nearest Neighbours (k-NN), Decision Tree, Random Forest, and Artificial Neural Network (ANN). These algorithms were employed to forecast students' grades across five academic subjects. The study's assessment was based on various metrics, including precision, recall, and the F1 score, to gauge the accuracy of the algorithms.

The outcomes of the research unveiled that the Artificial Neural Network (ANN) algorithm exhibited the most favourable performance, achieving an overall accuracy rate of 85.33%. Additionally, the Decision Tree and Random Forest algorithms demonstrated commendable performance, with accuracy rates of 80.67% and 80.33%, respectively. In contrast, the Naive Bayes and k-NN algorithms displayed lower accuracy rates, standing at 74.00% and 73.33%, respectively. The key takeaway from this study was that machine learning algorithms can be effectively leveraged to forecast student grades, and among the algorithms assessed, the ANN algorithm emerged as the most precise option for this purpose. [2]

2.2 Literature Review

While there has been an overall improvement in the educational attainment of the Portuguese population over recent decades, Portugal still faces significant educational challenges, especially concerning high student failure rates. The problem is particularly acute in core subjects like Mathematics and the Portuguese language. However, the fields of Business Intelligence (BI) and Data Mining (DM), which focus on extracting valuable insights from raw data, offer promising automated tools that can assist the education sector. This study aims to address student achievement in secondary education by employing BI/DM techniques.

To achieve this, the researchers collected real-world data, including student grades, demographic information, social factors, and school-related features, using school reports and questionnaires. The study focused on two critical subjects: Mathematics and Portuguese, and employed binary and five-level classification as well as regression tasks. They tested four DM models: Decision Trees, Random Forest, Neural Networks, and Support Vector Machines, along with three different input configurations, including the presence or absence of previous grades.

The findings of this research suggest that predictive accuracy can be high, especially when considering the availability of first and second school period grades. While past performance is a significant influence on student achievement, the study's explanatory analysis revealed other relevant factors, such as the number of absences, parents' occupations and education levels, and even student alcohol consumption. As a direct outcome of this research, more effective tools for predicting student outcomes can be developed. This can ultimately lead to improvements in the quality of education and better management of educational resources. [3]

Chapter 3: System Analysis and Design

3.1 System Analysis (Structured Approach/Object Oriented Approach)

3.1.1 Requirement Analysis

Requirement analysis for the Question-and-Answer Website with Student Grade Prediction system entails determining what the system should achieve and understanding user expectations. This process involves identifying the specific needs of end users and creating a comprehensive document known as the Software Requirement Specification (SRS). Before initiating system development, the following requirements are considered:

Comprehensive Understanding: It is essential to gain a deep understanding of the system's intended functions and purposes.

User Expectations: User needs and expectations must be clearly identified and addressed to ensure that the system aligns with their requirements.

SRS Documentation: The creation of a detailed Software Requirement Specification (SRS) document is a fundamental step. This document outlines both the functional and non-functional requirements of the system, serving as a roadmap for development.

This requirement analysis phase is vital to ensure that the Question-and-Answer Website with Student Grade Prediction system effectively meets user needs and functions as intended.

i. Functional requirements

Table 3.1:Functional Requirement for Q&A website with Student Grade Prediction

S.No.	Requirements	Description
1.	Login/Register	The system should be able to make the user/admin login and register with authentication and authorization.
2.	Ask Question/Provide Feedback	The system should be able to provide the users with feature of asking question and providing feedback for the given question in the website.
3.	Predict Student Grade	The system should be able to provide the users with grade

		prediction feature for authorized users.
4.	Logout	The system should provide feature of logging out if the user wants to end their session in the website.

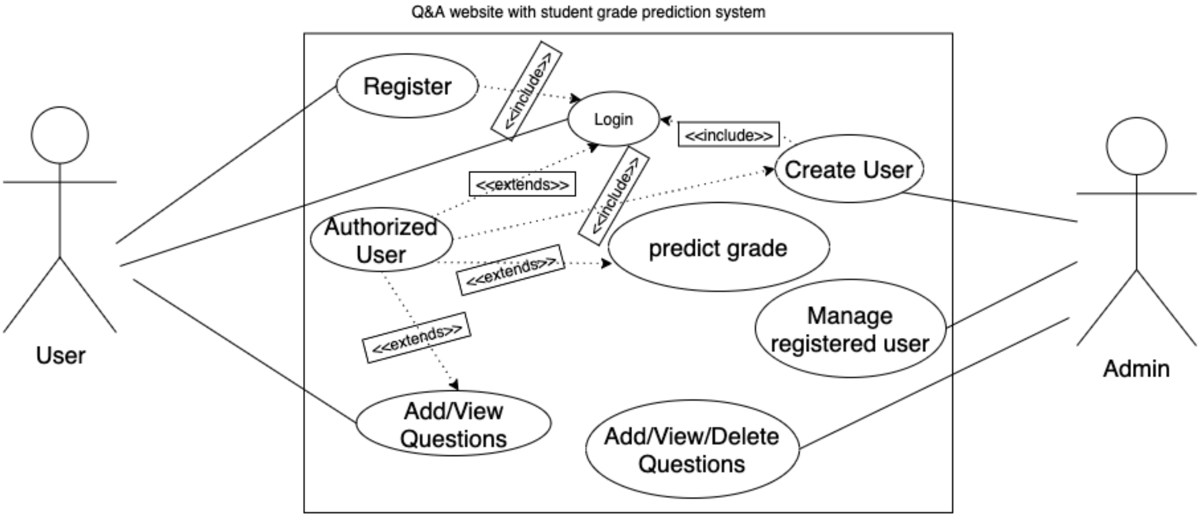


Figure 3.1:Use Case of Q&A website with Student Grade Prediction

ii. Non-functional requirements

Some of the non-functional requirement for Q&A website with Student Grade Prediction are mentioned below:

- Performance: The system should be able to process and predict grades efficiently and in a timely manner, even with a large number of users and data inputs.
- Accuracy: The grade prediction model should have a high level of accuracy in predicting student grades and performance, minimizing errors or incorrect predictions.
- Scalability: The system should be designed to handle an increasing number of users and data inputs without significant degradation in performance or accuracy.
- Reliability: The system should be reliable and available for use during required periods, minimizing downtime and ensuring continuous operation.
- Security: The system should implement appropriate security measures to protect user data and ensure confidentiality, integrity, and privacy of student information.

3.1.2 Feasibility Analysis (Technical, operational, economic)

i. Technical

The Question-and-Answer Website with Student Grade Prediction is technically feasible, benefiting from readily accessible technical resources. The system development utilized current technology, with software development resources met through open-source and free tools. Hardware requirements for coding and deployment were also readily available, without any infrastructure-related issues.

ii. Operational

The Question-and-Answer Website with Student Grade Prediction is operationally feasible, featuring a user-friendly interface that requires only basic web knowledge for access. Furthermore, the system fully aligns with the functional requirements and incorporates all necessary components.

iii. Economic

From the developer's perspective, we utilized open-source development tools, and there were no economic challenges encountered during the project. Likewise, the development costs remained well within the predefined budget, affirming that the system is economically feasible and cost-effective.

iv. Schedule Feasibility

Table 3.2:Gantt Chart for System Development

Tasks	Week 1-2	Week 3-4	Week 5-6	Week 7-8	Week 9-10	Week 10-12	Week 12-14	Week 15-16
Documentation	■	■	■	■	■	■	■	■
Planning	■	■						
Design		■	■					
Development			■	■	■	■		
Testing						■	■	
Deployment								■

Based on previous experience, the project was successfully completed within a 4.5-month timeframe (Starting March,2023 until Early July, 2023). This timeline demonstrates that the software development project proceeded without any complications and adhered to the established schedule, affirming its timeliness and feasibility.

3.1.3 Object Modelling: Object & Class Diagram

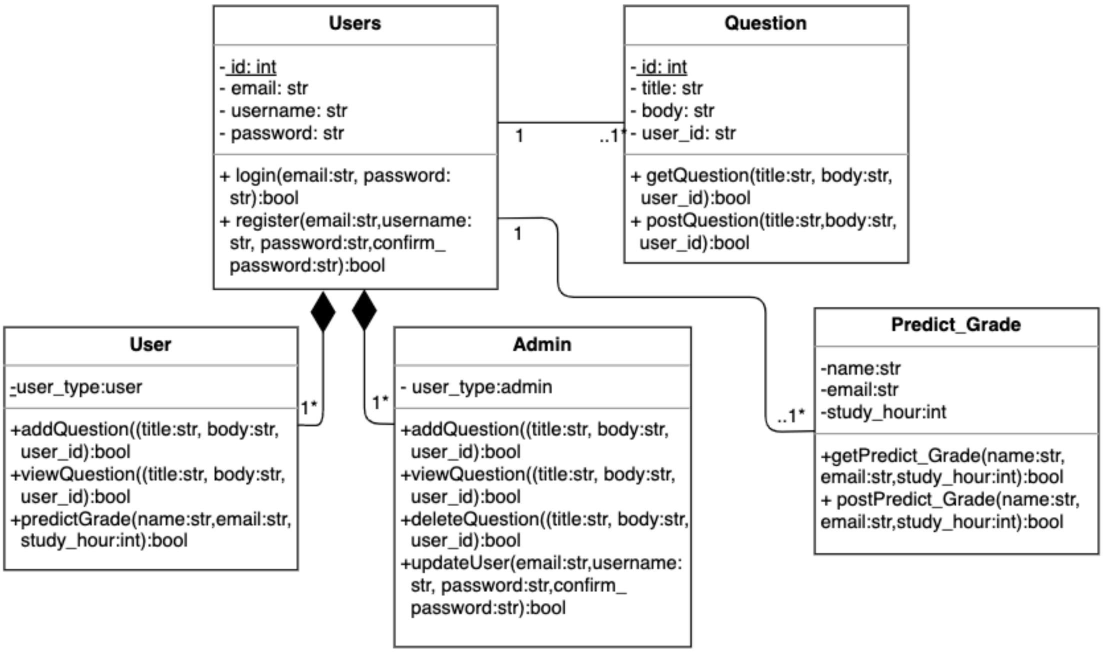

Users
- id: int
- email: str
- username: str
- password: str
+ login(email:str, password: str):bool
+ register(email:str,username: str, password:str,confirm_ password:str):bool
1
..1*
1
Question
- id: int
- title: str
- body: str
- user_id: str
+ getQuestion(title:str, body:str, user_id):bool
+ postQuestion(title:str,body:str, user_id):bool
Predict_Grade
-name:str
-email:str
-study_hour:int
+getPredict_Grade(name:str, email:str,study_hour:int):bool
+ postPredict_Grade(name:str, email:str,study_hour:int):bool
..1*
User
-user_type:user
+addQuestion((title:str, body:str, user_id):bool
+viewQuestion((title:str, body:str, user_id):bool
+predictGrade(name:str,email:str, study_hour:int):bool
1*
1*
Admin
- user_type:admin
+addQuestion((title:str, body:str, user_id):bool
+viewQuestion((title:str, body:str, user_id):bool
+deleteQuestion((title:str, body:str, user_id):bool
+updateUser(email:str,username: str, password:str,confirm_ password:str):bool

Figure 3.2:Class diagram for Q&A website with student grade prediction

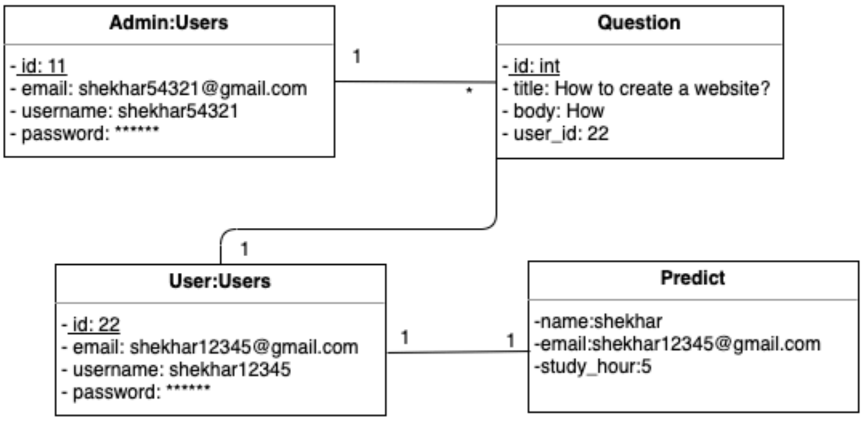


Figure 3.3:Object diagram showing single instance of class for Q&A website with student grade prediction

3.1.4 Dynamic Modelling: State & Sequence diagram

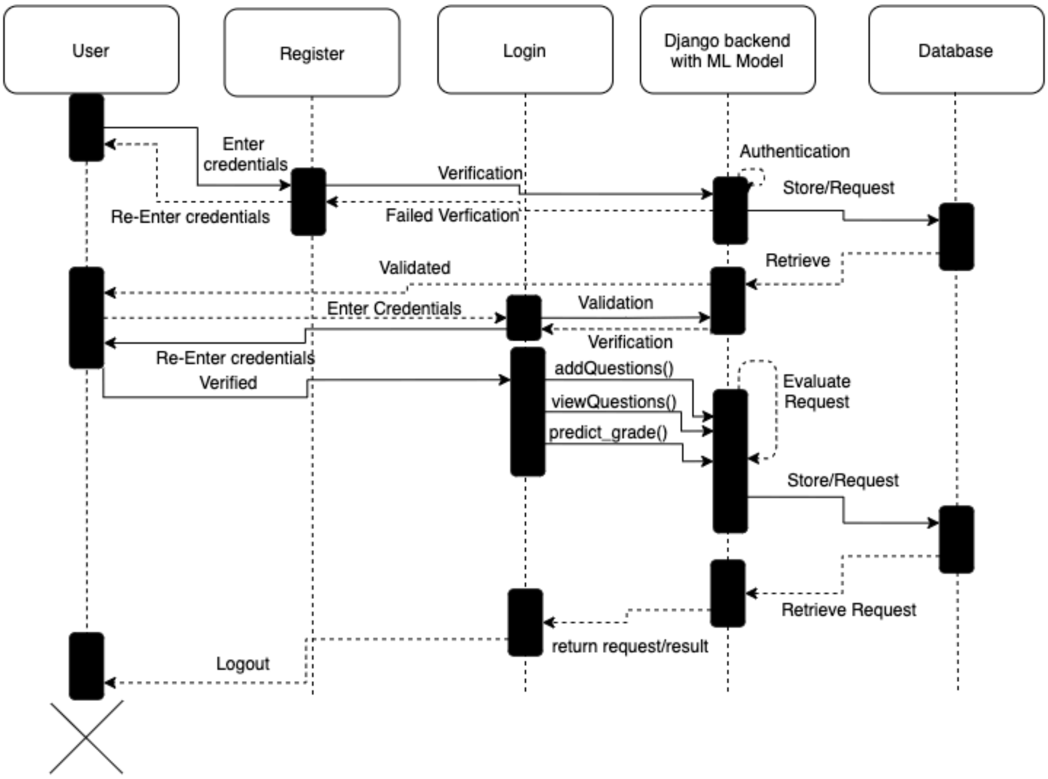


Figure 3.4:Sequence Diagram of User for Q&A website with student grade prediction

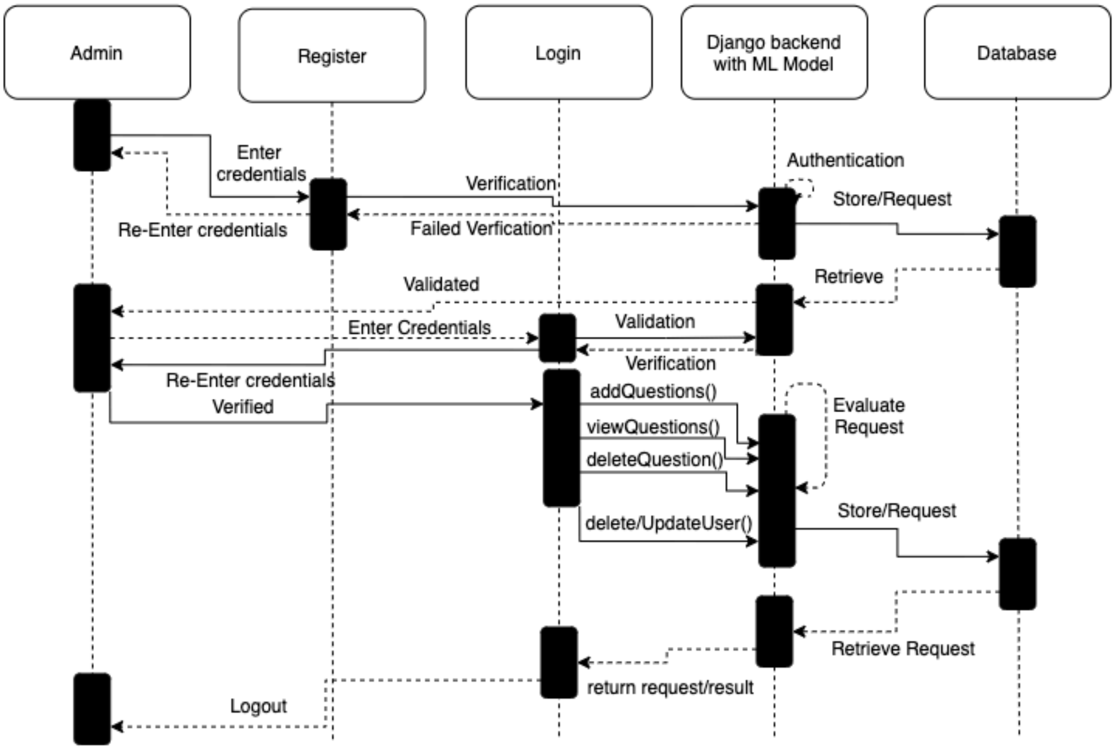


Figure 3.5:Sequence Diagram of Admin for Q&A website with student grade prediction

3.1.5 Process modelling: Activity Diagram

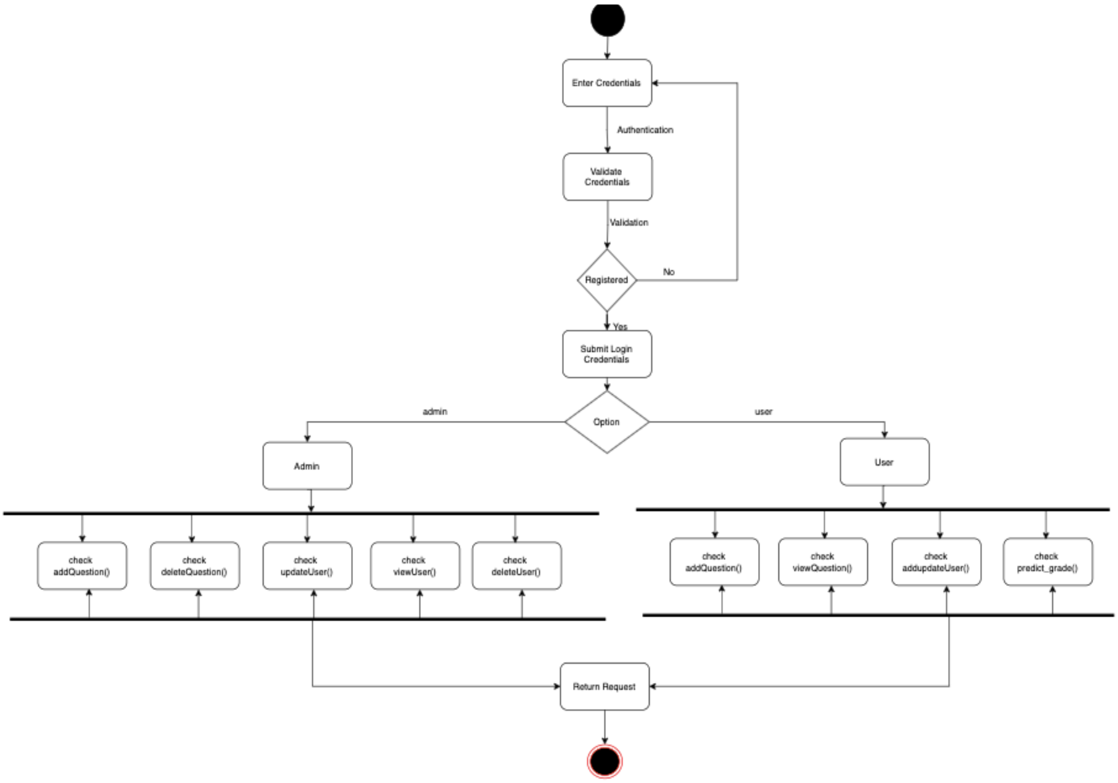


Figure 3.6:Activity Diagram for Q&A website with student grade prediction

3.2 System Design (Structured Approach/Object Oriented Approach)

3.2.1 Component diagram

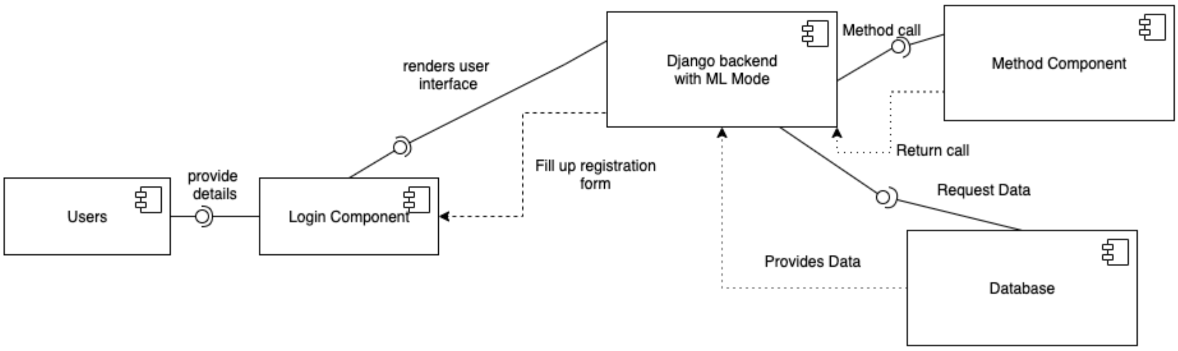


Figure 3.7:Component Diagram for Q&A website with student grade prediction

3.2.2 Deployment diagram

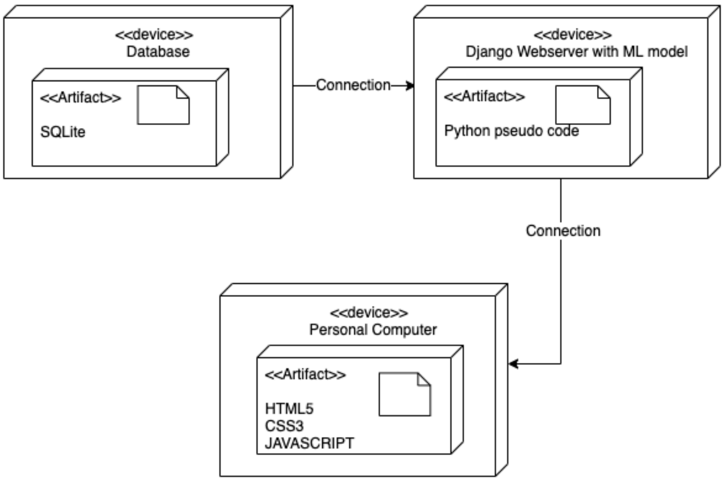


Figure 3.8:Deployment Diagram for Q&A website with student grade prediction

3.3 Algorithm details

Linear Regression

- Linear regression is a statistical regression method which is used for predictive analysis.
- It is one of the very simple and easy algorithms which works on regression and shows the relationship between the continuous variables.
- It is used for solving the regression problem in machine learning.
- Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.
- If there is only one input variable (x), then such linear regression is called simple linear regression. And if there is more than one input variable, then such linear regression is called multiple linear regression.

Below is the mathematical equation for Linear regression:

Y= aX+b

Here, Y = dependent variables (target variables),

X= Independent variables (predictor variables),

a and b are the linear coefficients

Linear Regression is one of the most important algorithms in machine learning. It is the statistical way of measuring the relationship between one or more independent variables vs one dependent variable.

Assuming we have a linear regression model, it is then calculated with the equation:

y^i=mxi+by^i=mxi+b

Where:

- y^i is the predicted value for the ith observation.
- xi is the ith feature value.
- m is the slope (coefficient) of the linear regression model.
- b is the intercept of the linear regression model.

And our data consists of n observations.
Here's how you can calculate the R-squared value step by step:

Calculating the Mean of the Actual Target Values:
y¯=1n∑i=1nyiy¯=n1∑i=1nyi

Calculating the Total Sum of Squares (SStot):
SStot=∑i=1n(yi−y¯)2SStot=∑i=1n(yi−y¯)2

Make Predictions and Calculating Residuals: Make predictions using your linear regression model and calculate the residuals (differences between actual and predicted values): Residuali=yi−y^iResiduali=yi−y^i

Calculate the Sum of Squared Residuals (SSresSSres):
SSres=∑i=1nResiduali2SSres=∑i=1nResiduali2

Calculate R-squared:
R2=1−SSresSStotR2=1−SStotSSres
In summary, the R-squared value is calculated by comparing the total variability in the target variable (SStot) with the variability that the model is unable to explain (SSres). The formula quantifies how well the linear regression model fits the data, with values closer to 1 indicating a better fit.

Chapter 4: Implementation and Testing

4.1 Implementation

4.1.1 Tools used (CASE tools, programming languages, database platforms)

CASE Tools (Computer-Aided Software Engineering): CASE tools encompass a variety of software tools used to support and automate different stages of the software development life cycle, including analysis, design, coding, testing, and documentation. Although specific CASE tools are not mentioned, the project likely utilized these tools for various development phases.

Programming Languages:

Python: Python, a versatile and widely-used programming language known for its simplicity and readability, serves as the backend programming language in this project. Python is commonly used for web development, data analysis, machine learning, and automation.

Web Framework:

Django Framework: The project leverages the Django framework, which is a high-level Python web framework. Django simplifies and accelerates web application development by offering tools and libraries for handling various aspects, including URL routing, database management, forms, and authentication.

Database Platforms:

SQLite: SQLite offers the advantage of being easy to set up and maintain, and it's also well-supported within Python and the Django web framework. It can be a good choice if your project doesn't require the scalability and advanced features provided by larger database systems.

Database Connectivity:

MySQL Connector Libraries (e.g., mysql-connector-python, pymysql): Given that Python is the backend language, MySQL connector libraries designed for Python are employed. These libraries enable Python code to connect to and interact with the MySQL database efficiently.

Web Deployment on AWS:

Amazon Web Services (AWS): In contrast to Heroku, this project deploys on AWS. AWS offers a robust cloud computing environment for hosting web applications. The deployment process involves configuring servers, databases, and necessary services on AWS to make the application accessible over the internet.

Development Tools:

Integrated Development Environment (IDE): Visual Studio Code is used for writing, testing, and debugging Python and Django code.

Version Control:

Git: Git, a distributed version control system, facilitates tracking code changes and effective collaboration among team members.

Data Analysis and Visualization:

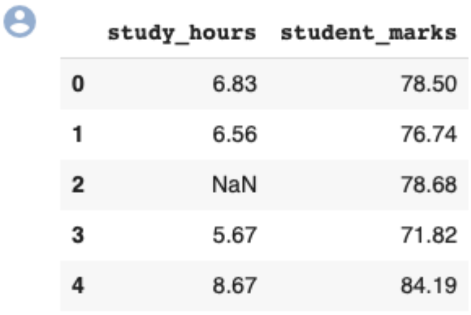
Python Libraries: Several Python libraries, including NumPy, pandas, Matplotlib, and Seaborn, are utilized for data analysis, data manipulation, and the creation of visualizations.

Each of these tools plays a critical role in various stages of the project, from development and testing to deployment and hosting. The selection of these tools aligns with the project's goals and requirements while considering the familiarity with each technology. In this case, the deployment is performed on AWS, a prominent cloud platform.

4.1.2 Implementation details of modules (description of procedures/ functions/ classes/ methods)

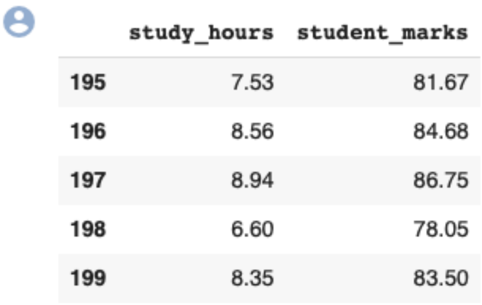
Linear Regression Model

```
# Importing Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
# Load Dataset
path = rhttps://drive.google.com/uc?export=download&id=13ZTYmL3E8S0nz-UKl4aaTZJaI3DVBGHM
df  = pd.read_csv(path)
df.head()
```



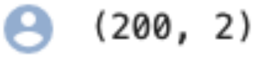
	study_hours	student_marks
0	6.83	78.50
1	6.56	76.74
2	NaN	78.68
3	5.67	71.82
4	8.67	84.19

```
df.tail()
```



	study_hours	student_marks
195	7.53	81.67
196	8.56	84.68
197	8.94	86.75
198	6.60	78.05
199	8.35	83.50

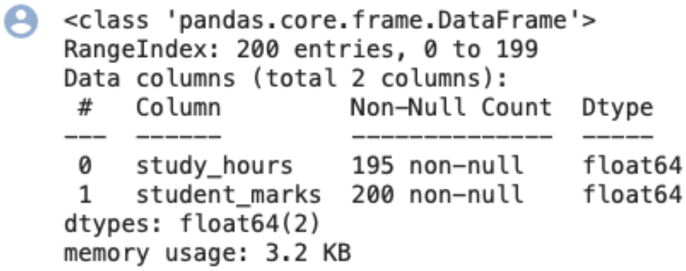
df.shape



```
(200, 2)
```

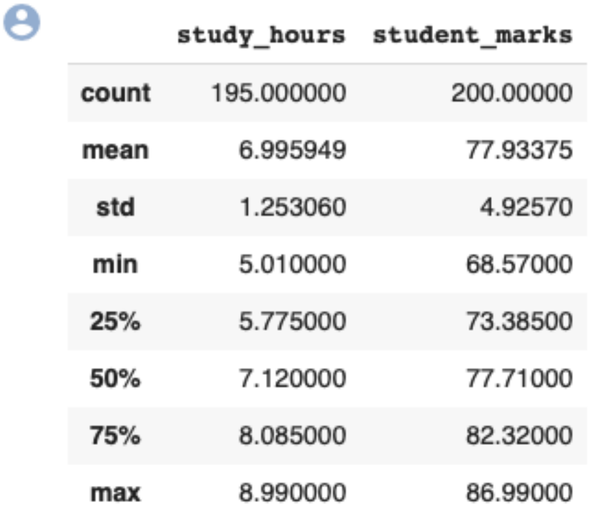
## Discover and visualize the data to gain insights

df.info()



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 2 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   study_hours    195 non-null    float64
 1   student_marks  200 non-null    float64
dtypes: float64(2)
memory usage: 3.2 KB
```

df.describe()



	study_hours	student_marks
count	195.000000	200.00000
mean	6.995949	77.93375
std	1.253060	4.92570
min	5.010000	68.57000
25%	5.775000	73.38500
50%	7.120000	77.71000
75%	8.085000	82.32000
max	8.990000	86.99000

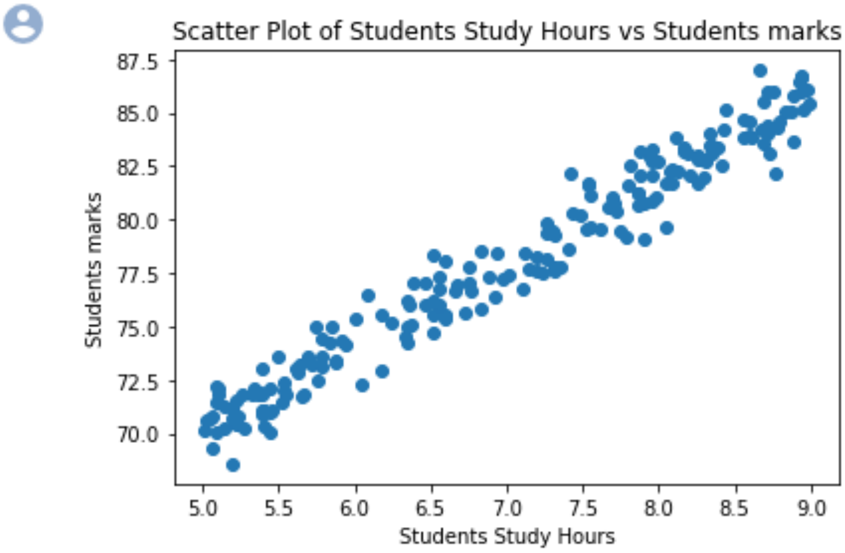
plt.scatter(x =df.study_hours, y = df.student_marks)

plt.xlabel("Students Study Hours")

plt.ylabel("Students marks")

plt.title("Scatter Plot of Students Study Hours vs Students marks")

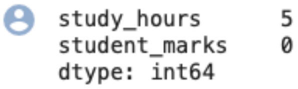
plt.show()



## Preparing the data for Machine Learning algorithms

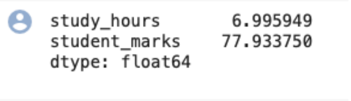
# Data Cleaning

df.isnull().sum()



```
study_hours      5
student_marks    0
dtype: int64
```

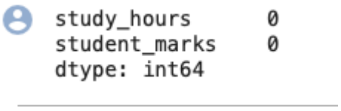
df.mean()



```
study_hours       6.995949
student_marks    77.933750
dtype: float64
```

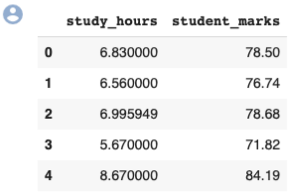
df2 = df.fillna(df.mean())

df2.isnull().sum()



```
study_hours      0
student_marks    0
dtype: int64
```

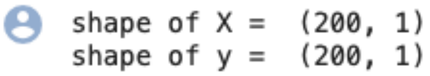
df2.head()



	study_hours	student_marks
0	6.830000	78.50
1	6.560000	76.74
2	6.995949	78.68
3	5.670000	71.82
4	8.670000	84.19

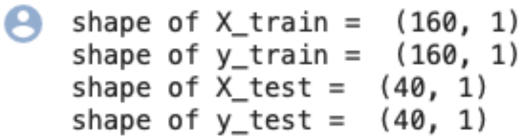
# splitting the dataset

```
X = df2.drop("student_marks", axis = "columns")
y = df2.drop("study_hours", axis = "columns")
print("shape of X = ", X.shape)
print("shape of y = ", y.shape)
```



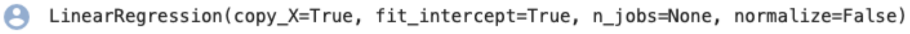
```
shape of X =  (200, 1)
shape of y =  (200, 1)
```

```
from sklearn.model_selection import train_test_split
X_train, X_test,y_train,y_test = train_test_split(X,y, test_size = 0.2, random_state=51)
print("shape of X_train = ", X_train.shape)
print("shape of y_train = ", y_train.shape)
print("shape of X_test = ", X_test.shape)
print("shape of y_test = ", y_test.shape)
```



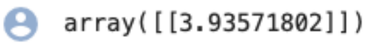
```
shape of X_train =  (160, 1)
shape of y_train =  (160, 1)
shape of X_test =  (40, 1)
shape of y_test =  (40, 1)
```

```
# Selecting a model and train it
 # y = m * x + c
 from sklearn.linear_model import LinearRegression
 lr = LinearRegression()
lr.fit(X_train,y_train)
```



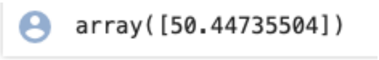
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
lr.coef_
```



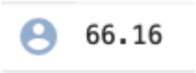
```
array([[3.93571802]])
```

```
lr.intercept_
```



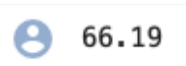
```
array([50.44735504])
```

```
m = 3.93
c = 50.44
y = m * 4 + c
y
```



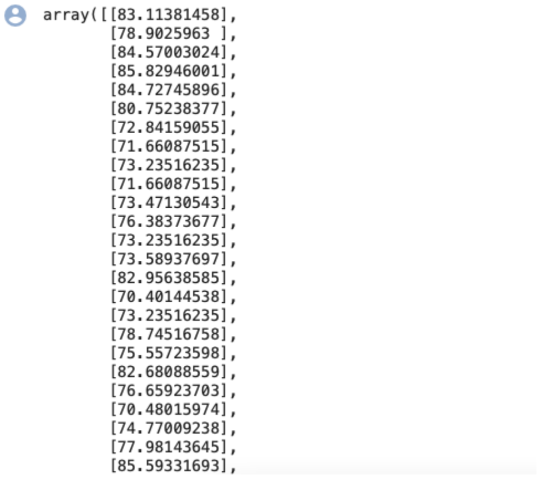
```
66.16
```

lr.predict([[4]])[0][0].round(2)

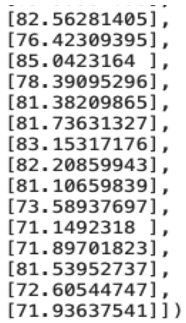


```
66.19
```

y_pred = lr.predict(X_test)

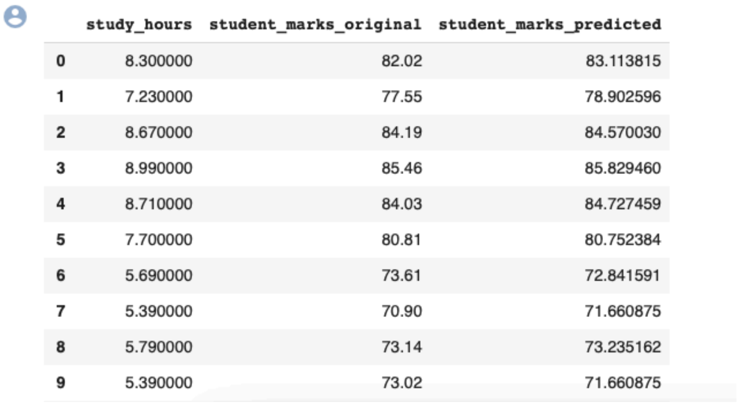


```
array([[83.11381458],
       [78.9025963 ],
       [84.57003024],
       [85.82946001],
       [84.72745896],
       [80.75238377],
       [72.84159055],
       [71.66087515],
       [73.23516235],
       [71.66087515],
       [73.47130543],
       [76.38373677],
       [73.23516235],
       [73.58937697],
       [82.95638585],
       [70.40144538],
       [73.23516235],
       [78.74516758],
       [75.55723598],
       [82.68088559],
       [76.65923703],
       [70.48015974],
       [74.77009238],
       [77.98143645],
       [85.59331693],
```



```
       [82.56281405],
       [76.42309395],
       [85.0423164 ],
       [78.39095296],
       [81.38209865],
       [81.73631327],
       [83.15317176],
       [82.20859943],
       [81.10659839],
       [73.58937697],
       [71.1492318 ],
       [71.89701823],
       [81.53952737],
       [72.60544747],
       [71.93637541]])
```

pd.DataFrame(np.c_[X_test, y_test, y_pred], columns = ["study_hours", "student_marks_original","student_marks_predicted"])



	study_hours	student_marks_original	student_marks_predicted
0	8.300000	82.02	83.113815
1	7.230000	77.55	78.902596
2	8.670000	84.19	84.570030
3	8.990000	85.46	85.829460
4	8.710000	84.03	84.727459
5	7.700000	80.81	80.752384
6	5.690000	73.61	72.841591
7	5.390000	70.90	71.660875
8	5.790000	73.14	73.235162
9	5.390000	73.02	71.660875

```
# Mathematical implementation of linear regression
def simple_linear_regression(X_train, y_train, X_test):
  mean_X = sum(X_train) / len(X_train)
  mean_y = sum(y_train) / len(y_train)
  m = sum((X_train[i] - mean_X) * (y_train[i] - mean_y) for i in range(len(X_train))) / sum((X_train[i] - mean_X) ** 2 for i in range(len(X_train)))
  c = mean_y - m * mean_X
  y_pred = [m * x + c for x in X_test]
  return m, c, y_pred
 m, c, y_pred = simple_linear_regression(X_train, y_train, X_test)
mean_y = sum(y_test) / len(y_test)
sst = sum((y_test[i] - mean_y) ** 2 for i in range(len(y_test)))
sse = sum((y_test[i] - y_pred[i]) ** 2 for i in range(len(y_test)))
score = 1 - (sse / sst)
```

Here's how linear regression works in the model:

- **Calculating Mean Values:**

Calculate the mean of the input feature (X_train) and the target variable (y_train). These are represented as mean_X and mean_y, respectively.

- **Calculating Slope and Intercept:**

Compute the slope (m) and the y-intercept (c) of the regression line using the formulas for simple linear regression: m=∑i=1n(Xi−X‾)(yi−y‾)∑i=1n(Xi−X‾)2m=∑i=1n(Xi−X‾)2∑i=1n (Xi−X‾)(yi−y‾) c=y‾−m⋅X‾c=y‾−m⋅X‾

Here, nn is the number of data points.

- **Making Predictions:**

Use the computed slope and intercept to predict the target variable for the test set (X_test). The predictions are stored in the y_pred list.

- **Evaluating Model Performance:**

Calculate the mean of the actual target variable values in the test set (y_test) to get mean_y.
Compute the total sum of squares (SSTSST) and the sum of squared errors (SSESSE):
SST=∑i=1n(yi−y‾)2SST=∑i=1n(yi−y‾)2
SSE=∑i=1n(yi−y^i)2SSE=∑i=1n(yi−y^i)2, where y^iy^i is the predicted value.
Calculate the coefficient of determination (R2R2) as the model evaluation metric:
R2=1−SSESSTR2=1−SSTSSE
The closer R2R2 is to 1, the better the model fits the data.

4.2 Testing

For the verification, validation and debugging of the system, the following test case has been designed.

Table 4.1: Test Case Design

S.NO.	Description	Requirements
1.	Admin Rights	-View/Edit/Delete users' details -View/Delete Question/Answers Details
2.	Users Rights	-Creating users (Registration) -View Questions/Answers -Predict Grade
. 3.	Duplicate Admin Credentials	-Email with Password must be different.
4.	Duplicate User credentials	- Email and Password must be different
5.	Change Password	-User and Admin should be able to change password.

4.2.1 Test cases for Unit Testing

During the testing process, individual units of source codes, sets of one or more program modules are tested to determine whether they can perform as per expectations. Each separate activities like login, dataset loader, prediction and visualization etc. part have been tested individually and all the tested activities pass with the successful result.

Table 3.2: Test Case for Admin

S.No.	Test Case ID	Test Description	Input Test Data	Expected Result	Actual Result	Test Result
1.	TC-1.1	Open Browser and enter URL	http://127.0.0.1:8000 /login	Must Display Login Page Having input field of email And password	As Expected	Passed

2.	TC-1.2	Enter Valid Admin Credentials /Admin Email and Password	Username: shekhar12345 Password: ********	Must Redirect Access to Admin Panel	As Expected	Passed
3.	TC-1.3	Enter Data in Email and Password as Empty	Username: Password:	Must Show Error Message	As Expected	Passed
4.	TC-1.4	Enter Data in Email as empty with Password	Username: Password: ******	Must Show Error Message	As Expected	Passed
5.	TC-1.5	Enter Data in Email with Password as empty	Username: admin@gmail.com Password:	Must Show Error Message	As Expected	Passed
6.	TC-1.6	Enter Data in Email and Password as invalid	Username: shekhar12345 Password: ***	Must Show Error Message	As Expected	Passed
7.	TC-1.7	Updating User Credentials	Email: shekhar1@gmail.com Username: shekhar12345 Password: ****** Confirm Password: ******	User’s Credentials Successfully Updated	As Expected	Passed
8.	TC-1.8	Deleting User Credentials	Email: shekhar1@gmail.com Username: shekhar12345 Password: ******	User’s Credentials Successfully Deleted	As Expected	Passed

9.	TC-1.9	Creating Admin Account	Username: shekhar12345 Password: ******** Confirm Password: ******	Must Create Admin Account	As Expected	Passed
10.	TC -1.10	Creating Admin Account with any one details empty in database	Email: Username: shekhar12345 Password: ******** Confirm Password: ******	Must Show Error Message (Enter your Phone number)	As Expected	Passed

Table 4.3:Test Case for User

S.No.	Test Case ID	Test Description	Input Test Data	Expected Resul	Actual Result	Test Result
1.	TC-2.1	Open Browser and enter URL	http://127.0.0.1:8000 /login	Must Display Login Page having input field of email and password with Creating Account Field	As Expected	Passed
2.	TC-2.2	Creating Account for user Login with all credentials	Email: shekhar54321@gmail.com Username: shekhar54321 Password: ******** Confirm Password: ******	Must Create User Account Successfully	As Expected	Passed
3.	TC-2.3	Creating Account	Email: Username:	Must show Validation	As Expected	Passed

		for Customer Login with all credentials Empty	Password: Confirm Password:	module to fill the form		
4.	TC-2.4	Enter Valid User Credentials/ User Email and Password	Email: shekhar54321@gmail.com Username: shekhar54321 Password: ******** Confirm Password: ******	Must redirect access to user panel	As Expected	Passed
5.	TC-2.5	Enter Data in Username and Password as Empty of User	Username: Password:	Must show error message	As Expected	Passed
6.	TC-2.6	Enter Data in Username as empty with Password of User	Username: Password: ******	Must show error message	As Expected	Passed
7.	TC-2.7	Enter Data in Username with wrong password	Username: Shekhar54321 Password:*	Must show error message	As Expected	Passed
8.	TC-2.8	Enter Data in Username and Password as invalid	Username: shekharme Password:****	Must show error message	As Expected	Passed
9.	TC-2.9	Updating User Credential in User Panel	Email: shekhar654321@gmail.cor Username: shekhar654321	Must Update User Credentials	As Expected	Passed

			Password: ********* Confirm Password: *******			
10.	TC-2.10	Updating User Credential in User Panel One field empty	Email: shekhar54321@gmail.com Username: Password: ******** Confirm Password: ******	Must Update User Credentials	As Expected	Passed

4.2.2 Test cases for System Testing

The test performed by the system for compatibility with the platform is known as system testing. Some of the Test Cases for System Testing while making Pet Management System are mentioned below in the table.

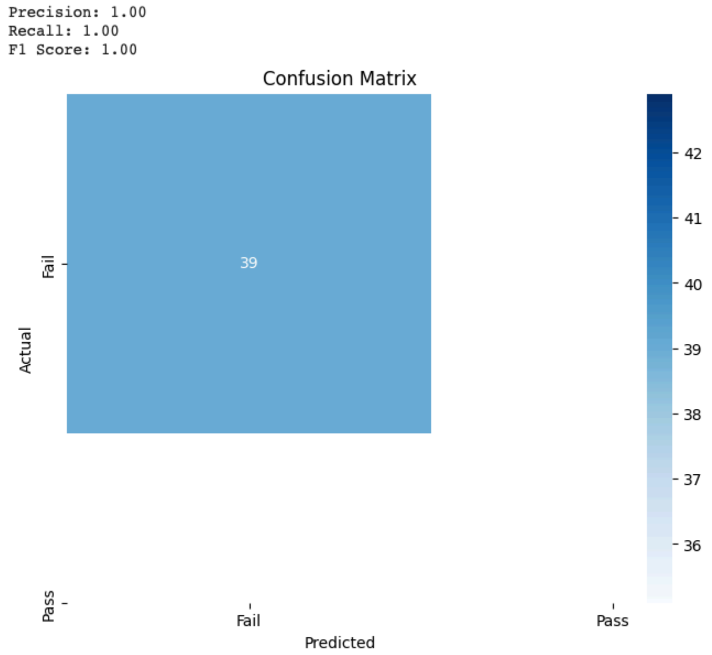
Table 4.4: Black box system testing as per test case

S.No.	Description	Requirements	Result	Test Result
1.	Admin Rights	-View/Edit/Delete users' details	Admin can view, update and delete user details	Passed
		-Add/View/Update /Delete Question/ Answer Details	Admin can view, update and delete Question/Answer Details	Passed
2.	User Rights	Creating User	User can register with validation	Passed
		View Question/Answer	User can view question and Provide feedback for answer	Passed
		Predict Grade	User can Predict Grade	Passed
		View Predicted Result	User can view the predicted result	Passed
3.	Duplication	Duplicate Admin Credentials	Admin's login with username and password is different	Passed
		Duplicate User Credentials	User's login with name and password can be same but email and phone number can be	Passed

			different	
4.	Password	Change Password	User and Admin can change password	Passed
		Confirm Password	User and Admin can confirm Password through validation	Passed

4.2.3 Result Analysis

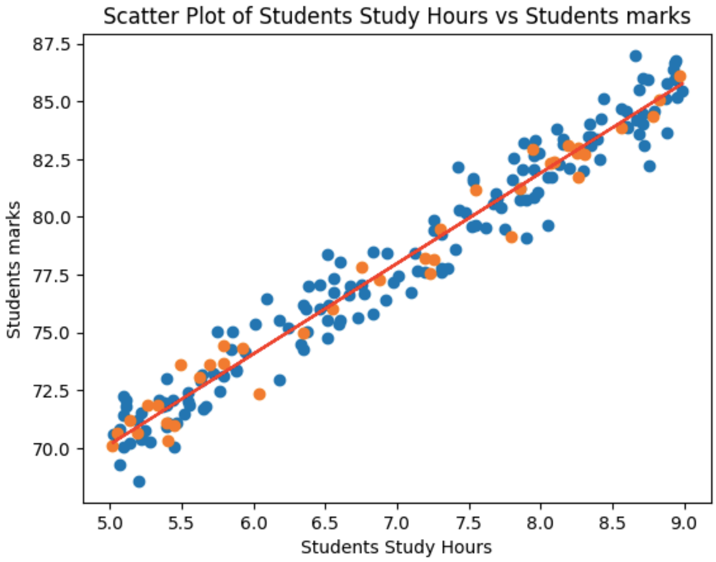
```
# Converting the problem to binary classification
pass_threshold = 40
y_test_pass_fail = (np.array(y_test) >= pass_threshold).astype(int)
y_pred_pass_fail = (np.array(y_pred) >= pass_threshold).astype(int)
# Evaluation metrics
from sklearn.metrics import confusion_matrix, precision_score, recall_score, f1_score
import seaborn as sns
# Calculate metrics
precision = precision_score(y_test_pass_fail, y_pred_pass_fail)
recall = recall_score(y_test_pass_fail, y_pred_pass_fail)
f1 = f1_score(y_test_pass_fail, y_pred_pass_fail)
conf_matrix = confusion_matrix(y_test_pass_fail, y_pred_pass_fail)
# Print metrics
print("Precision: {:.2f}".format(precision))
print("Recall: {:.2f}".format(recall))
print("F1 Score: {:.2f}".format(f1))
# Plot confusion matrix
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=['Fail', 'Pass'], yticklabels=['Fail', 'Pass'])
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```



```
Precision: 1.00
Recall: 1.00
F1 Score: 1.00
```

In above above figure, the precision, recall, and F1 score values of 1.00 each suggest that the model has performed exceptionally well in binary classification. A precision of 1.00 indicates that all predicted "Pass" instances are correct, while a recall of 1.00 signifies that all actual "Pass" instances are accurately identified. The F1 score, a balanced measure of precision and recall, is also perfect at 1.00. In essence, the model exhibits flawless accuracy in distinguishing between "Pass" and "Fail" classes based on the specified threshold of 40 marks. This outstanding performance is noteworthy, but it's prudent to assess the dataset's balance and consider the specific context to ensure the results are robust.

```
plt.scatter(X_train, y_train)
plt.scatter(X_test, y_test)
plt.plot(X_train, [m * x + c for x in X_train], color="r")
plt.xlabel("Students Study Hours")
plt.ylabel("Students marks")
plt.title("Scatter Plot of Students Study Hours vs Students marks")
plt.show()
print("R-squared score:", score)
```



```
R-squared score: 0.9736130341790977
```

After the fine tuning of the model, score state that the accuracy for the model was found to be around 97% which state that the model has high accuracy for predicting the student grade which was trained using linear regression algorithm. Despite, the high accuracy, the data was trained using different algorithm like decision tree and random forest but had accuracy lower than 70% so was not used for prediction. So, linear regression was only preferred for the analysis.

Chapter 5: Conclusion and Future Recommendations

5.1 Conclusion

In this project, we embarked on the exciting journey of developing and evaluating a student grade prediction system using various regression algorithms. The primary objective was to leverage machine learning techniques to accurately predict student grades based on relevant input features. We started by thoroughly understanding the dataset, which contained valuable information such as study time, past performance, and attendance. The dataset pre-processing phase was critical, involving data cleaning, feature selection, and splitting the data into training and testing sets. Our experimentation involved implementing three regression algorithms: Linear Regression, Decision Tree Regressor, and Random Forest Regressor. Each of these techniques offered unique insights into the student grade prediction task.

5.2 Lesson learnt/Outcome

Some of the key lessons learned after the development of Question-and-Answer Website with Student Grade Prediction are:

- **Data Understanding and Pre-processing**: Understanding and cleaning of data for accuracy.
- **Algorithm Experimentation**: Experimentation of different machine learning algorithms for specific tasks.
- **Data Splitting**: Splitting data for model training and testing to assess effectiveness.
- **Feature Selection and Engineering**: Identifying and creating relevant features for predictions.
- **Continuous Improvement**: Regularly refining and enhancing machine learning models for better result.
- **Interpretability**: Understanding of why predictions or recommendations are made.
- **Domain Knowledge**: Having a strong grasp of the domain to build effective models.

Applying these lessons lead to the development of a more accurate and valuable Q&A website with student grade prediction features.

5.3 Future Recommendations

For a Question-and-Answer Website with Student Grade Prediction, here are some future recommendations:

- **Enhanced User Experience**: Improvement of the user interface and experience to make it even more user-friendly and accessible for a diverse user base.
- **Advanced Machine Learning Models**: Explore and implementation of advanced machine learning models and techniques to enhance the accuracy of grade predictions and answers.
- **Feedback Mechanism**: Incorporation of a feedback system that allows students and educators to provide input on the quality and relevance of answers and predictions.
- **Real-time Updates**: Providing of a real-time grade predictions and continuous updates for students to track their progress throughout the semester.
- **Data Security and Privacy**: Strengthening of data security measures to ensure the protection of sensitive student data.
- **Integration with Learning Management Systems**: Integration with existing learning management systems to streamline the sharing of relevant data and improve predictions.
- **Mobile Application**: Development of a mobile app version for convenient access on smartphones and tablets.
- **Community Building**: Fostering a community of users who can share experiences and best practices related to the Q&A website.
- **Multilingual Support**: Offering multilingual support to cater to a broader international audience.
- **Analytics Dashboard**: Creating a dashboard for educators and administrators to gain insights into student performance and trends.

Implementing these recommendations can lead to a more robust and effective Q&A website for student grade prediction in the near future.

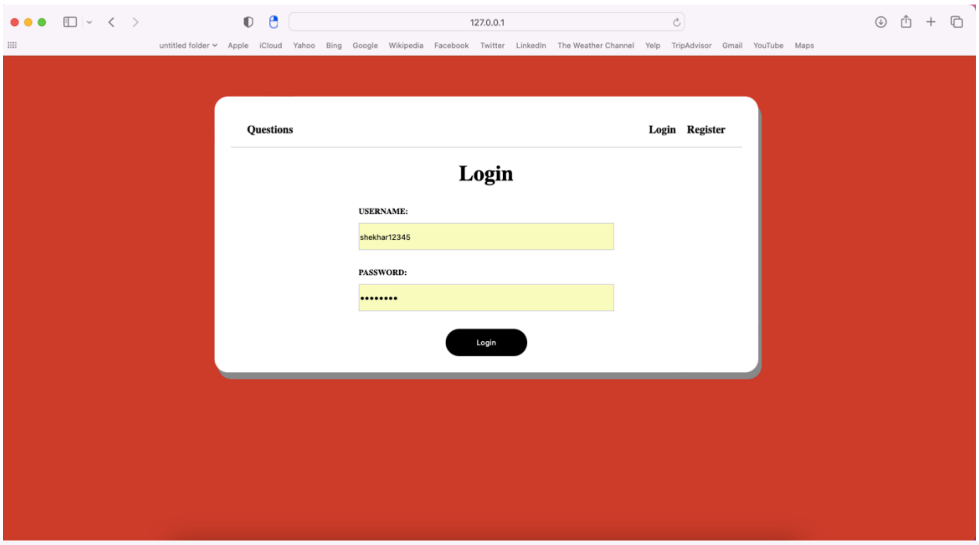
References

[1] “asana.com,” 15 October 2022. [Online]. Available: https://asana.com/resources/agile-methodology. [Accessed 10 October 2023].

[2] K. and P. , A Comparative Study of Machine Learning Algorithms for Student Grade Prediction, 2019.

[3] C.Paulo and S.Alice, Himalayan Times, [Online]. Available: http://www3.dsi.uminho.pt/pcortez/student.pdf.. [Accessed 10 October 2023].

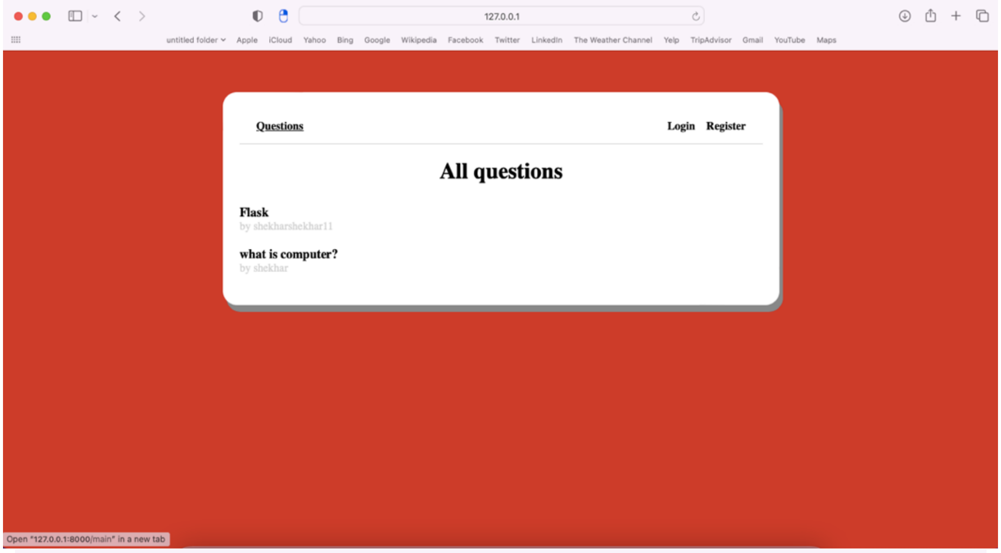
Appendices



Login

USERNAME: shekhar12345

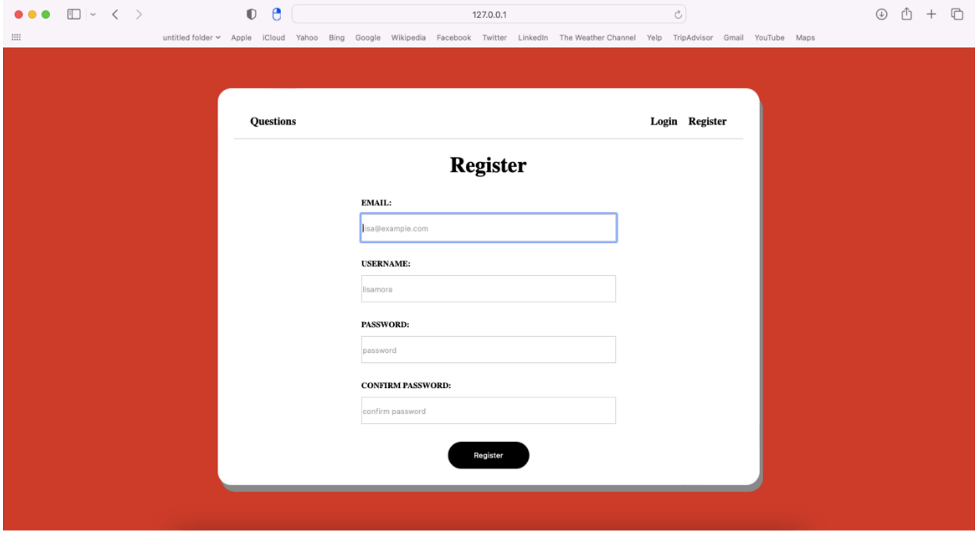
PASSWORD: ••••••••



All questions

Flask
by shekharshekhar11

what is computer?
by shekhar



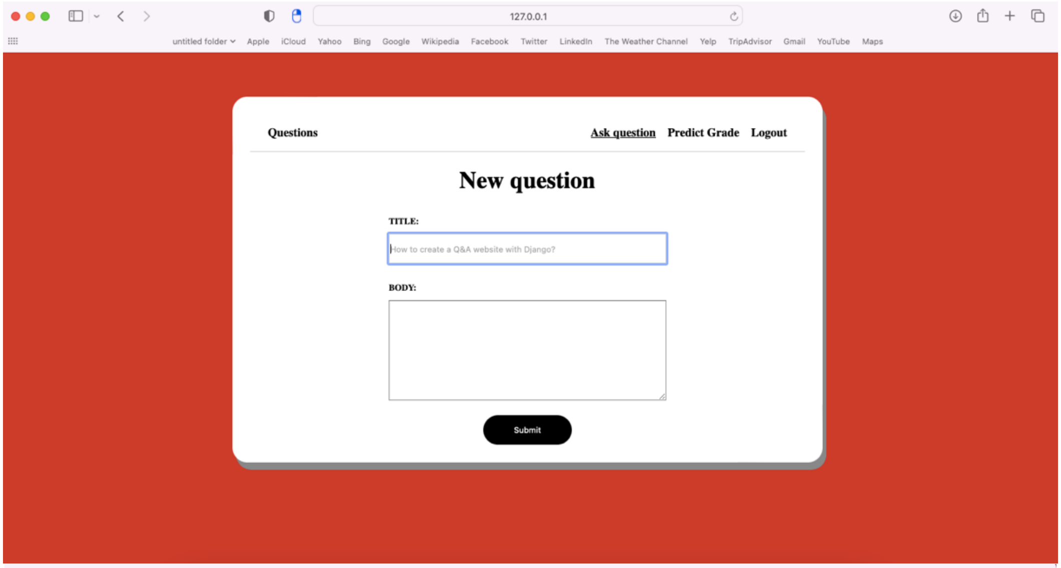
Register

EMAIL: lisa@example.com

USERNAME: lisamora

PASSWORD: password

CONFIRM PASSWORD: confirm password

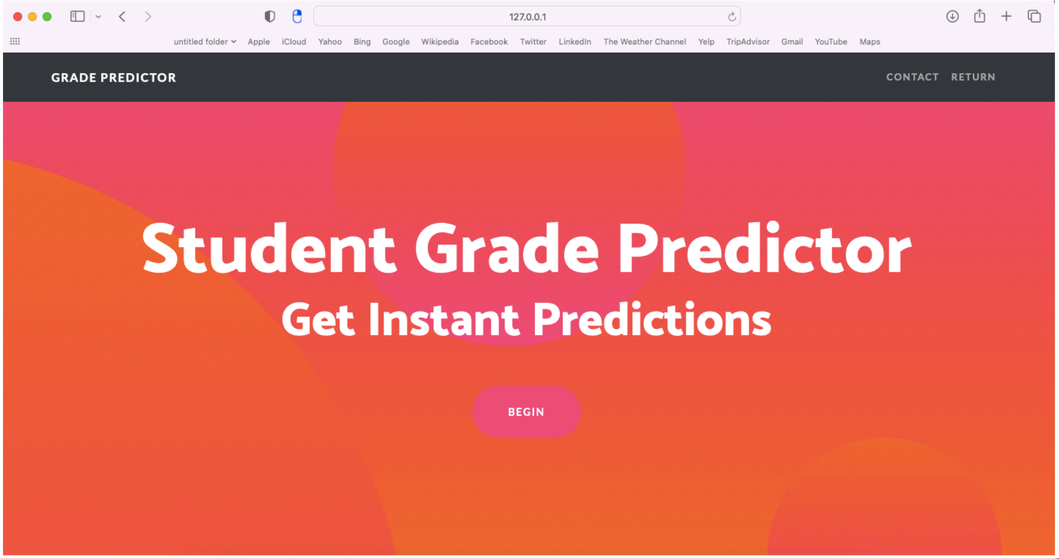


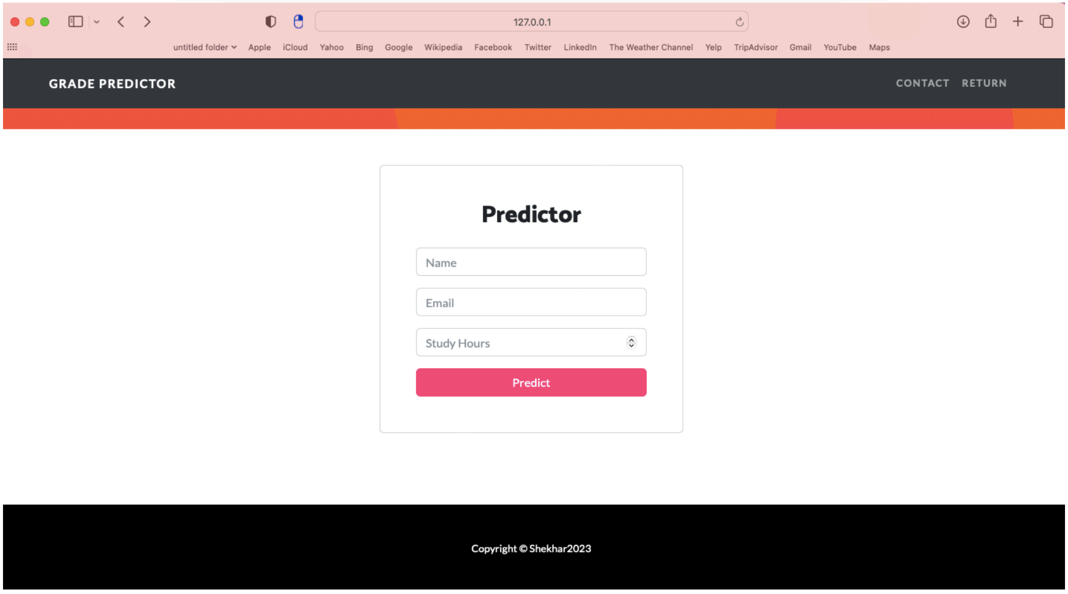
New question

TITLE:

How to create a Q&A website with Django?

BODY:





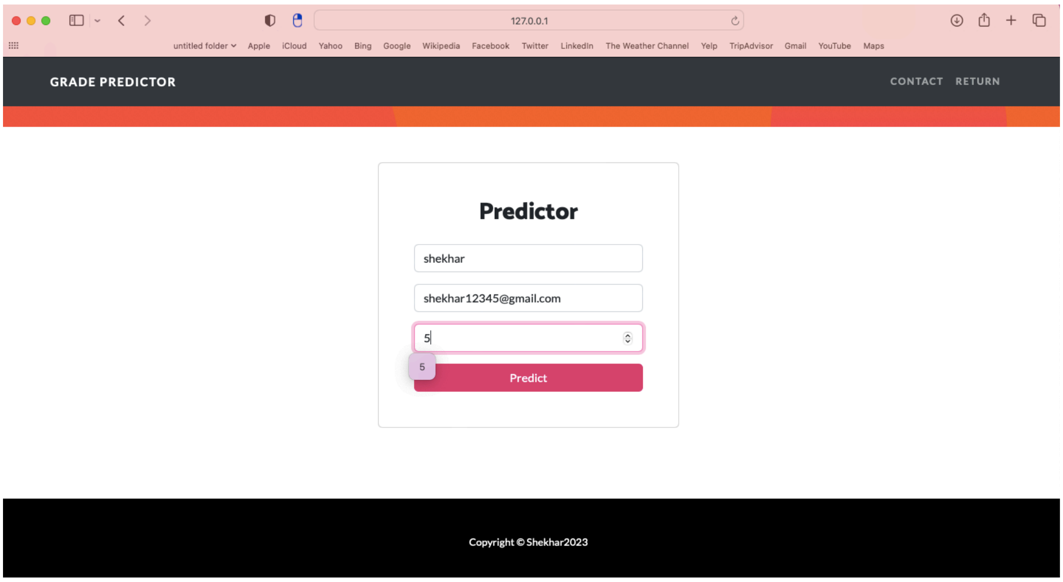
Predictor

Name

Email

Study Hours

Copyright © Shekhar2023


127.0.0.1
untitled folder Apple iCloud Yahoo Bing Google Wikipedia Facebook Twitter LinkedIn The Weather Channel Yelp TripAdvisor Gmail YouTube Maps
GRADE PREDICTOR
CONTACT RETURN
Predict

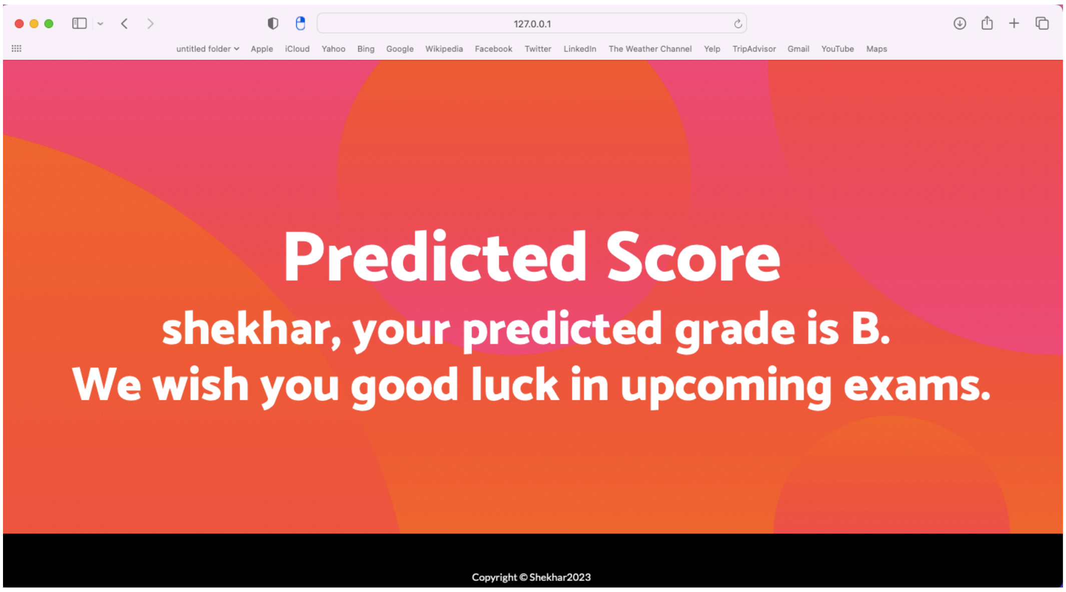
Predictor

shekhar

shekhar12345@gmail.com

5

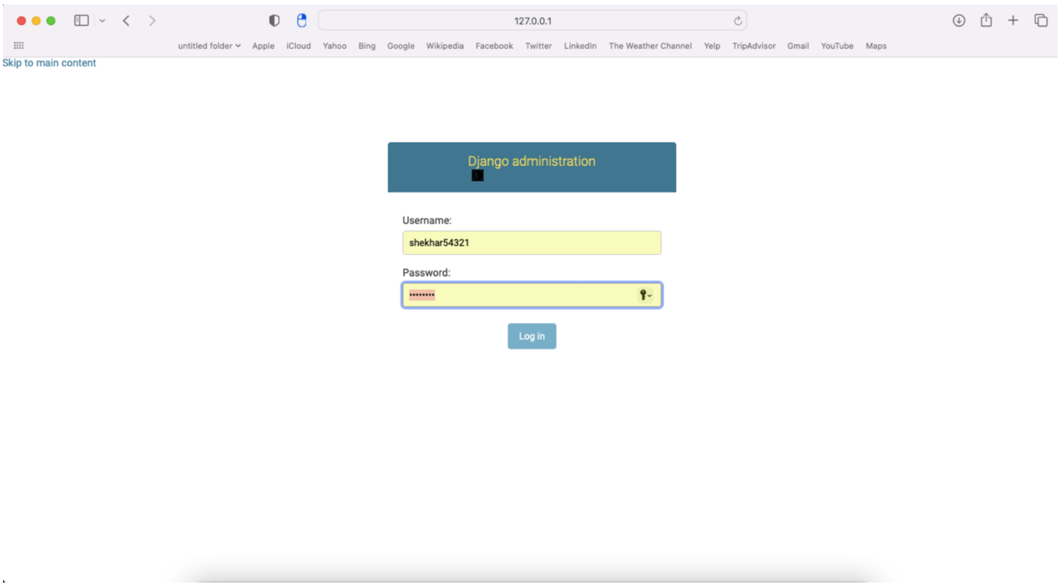
Copyright © Shekhar2023


127.0.0.1
untitled folder Apple iCloud Yahoo Bing Google Wikipedia Facebook Twitter LinkedIn The Weather Channel Yelp TripAdvisor Gmail YouTube Maps

Predicted Score

shekhar, your predicted grade is B.
We wish you good luck in upcoming exams.

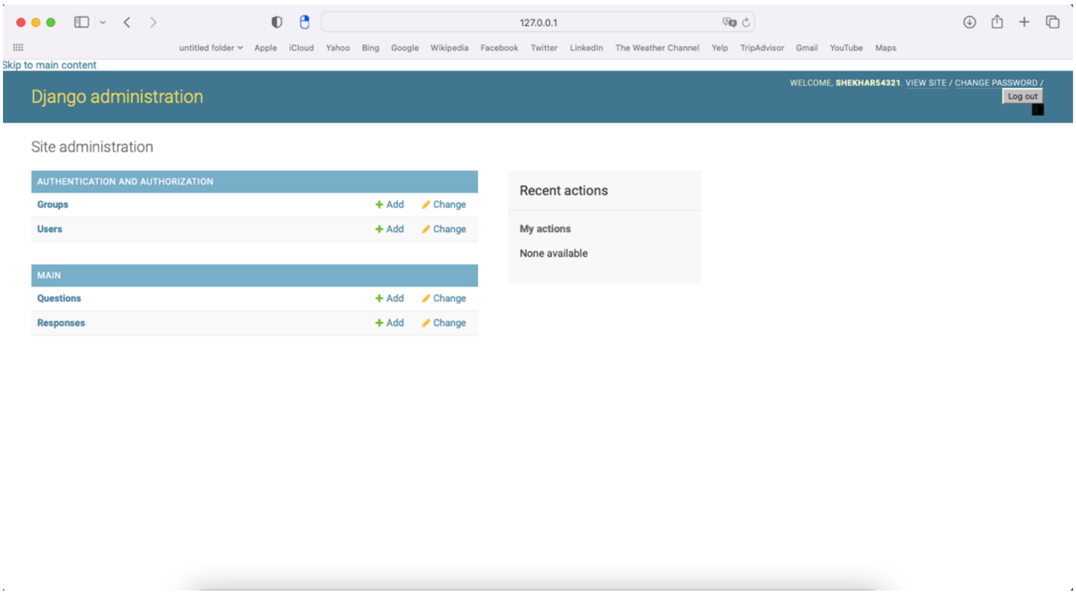
Copyright © Shekhar2023


127.0.0.1
untitled folder Apple iCloud Yahoo Bing Google Wikipedia Facebook Twitter LinkedIn The Weather Channel Yelp TripAdvisor Gmail YouTube Maps
Skip to main content
Log in

Django administration

Username: shekhar54321

Password: ••••••••


127.0.0.1
untitled folder Apple iCloud Yahoo Bing Google Wikipedia Facebook Twitter LinkedIn The Weather Channel Yelp TripAdvisor Gmail YouTube Maps
Skip to main content
WELCOME, SHEKHAR54321. VIEW SITE / CHANGE PASSWORD /
Log out

Django administration

Site administration

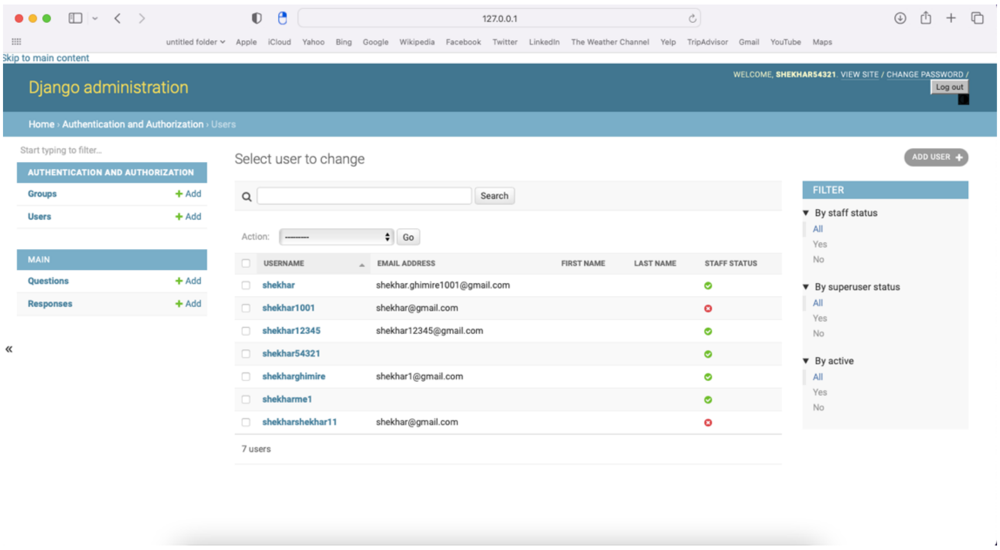
AUTHENTICATION AND AUTHORIZATION		
Groups	+ Add	Change
Users	+ Add	Change

MAIN		
Questions	+ Add	Change
Responses	+ Add	Change

Recent actions

My actions

None available


127.0.0.1
untitled folder Apple iCloud Yahoo Bing Google Wikipedia Facebook Twitter LinkedIn The Weather Channel Yelp TripAdvisor Gmail YouTube Maps
Skip to main content
WELCOME, SHEKHAR54321. VIEW SITE / CHANGE PASSWORD /
Log out
Home › Authentication and Authorization › Users
Start typing to filter...
AUTHENTICATION AND AUTHORIZATION
Groups + Add
Users + Add
MAIN
Questions + Add
Responses + Add
«
ADD USER +
Search
Action: ——— Go

Django administration

Select user to change

☐	USERNAME	EMAIL ADDRESS	FIRST NAME	LAST NAME	STAFF STATUS
☐	shekhar	shekhar.ghimire1001@gmail.com			✓
☐	shekhar1001	shekhar@gmail.com			✗
☐	shekhar12345	shekhar12345@gmail.com			✓
☐	shekhar54321				✓
☐	shekharghimire	shekhar1@gmail.com			✓
☐	shekharme1				✓
☐	shekharshekhar11	shekhar@gmail.com			✗

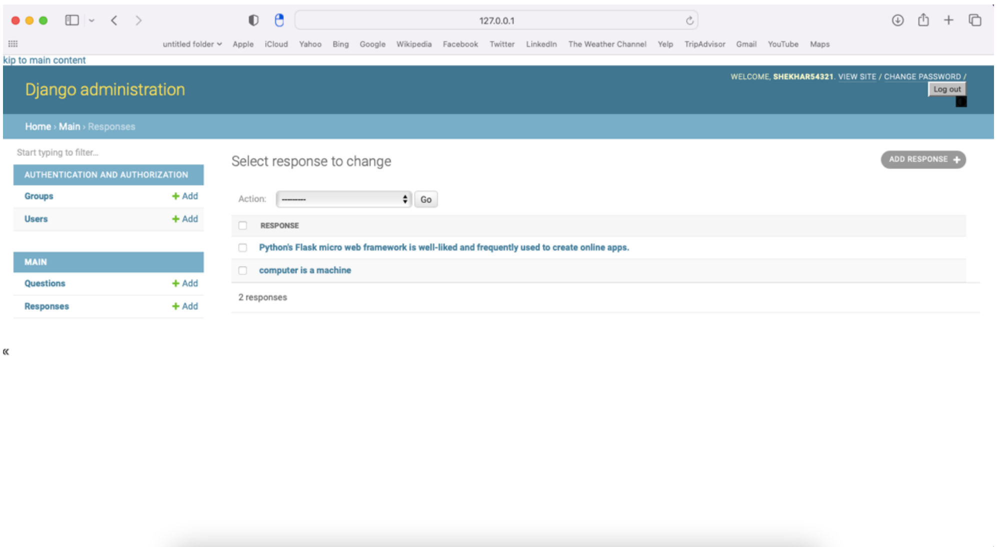
7 users

FILTER

▼ By staff status
All
Yes
No

▼ By superuser status
All
Yes
No

▼ By active
All
Yes
No


127.0.0.1
untitled folder Apple iCloud Yahoo Bing Google Wikipedia Facebook Twitter LinkedIn The Weather Channel Yelp TripAdvisor Gmail YouTube Maps
kip to main content
WELCOME, SHEKHAR54321. VIEW SITE / CHANGE PASSWORD /
Log out
Home › Main › Responses
Start typing to filter...
AUTHENTICATION AND AUTHORIZATION
Groups + Add
Users + Add
MAIN
Questions + Add
Responses + Add
«
ADD RESPONSE +
Action: ——— Go

Django administration

Select response to change

☐	RESPONSE
☐	Python's Flask micro web framework is well-liked and frequently used to create online apps.
☐	computer is a machine

2 responses