***Applied Machine learning***

***(Timetable Clash Detection)***

*A*

*Project Report*

*Submitted for Applied Machine Learning*

*By*

*Nishkarsh Gupta*

***Sap ID:500124595***

***Batch: 03***

*To*

*Er. Kushagra Shukla*



*SCHOOL OF COMPUTER SCIENCE*

*AIML CLUSTER*

*UNIVERSITY OF PETROLEUM AND ENERGY STUDIES*

## Problem Statement

Build an optimized logistic regression model for timetable clash detection and further clash predictions by pre-processing real-world data and tuning hyperparameters for reliable decision-making

## Introduction

In academic institutions, scheduling classes without conflicts is a complex task. Timetable clashes can lead to resource overutilization, student inconvenience, and administrative challenges. This project aims to leverage Machine Learning (ML) techniques to detect potential timetable clashes, thereby assisting in efficient schedule planning.

## Objectives

* Develop a machine learning model to predict timetable clashes.
* Utilize classification algorithms to identify potential conflicts.
* Evaluate model performance using appropriate metrics.
* Provide a tool that can assist in proactive timetable management.

## Motivation

Efficient scheduling is crucial in educational institutions to ensure optimal utilization of resources such as classrooms and instructors. Manual scheduling processes are often time-consuming and prone to errors, leading to clashes that disrupt the academic workflow. By implementing an automated system using Machine Learning, institutions can proactively identify and resolve scheduling conflicts, enhancing operational efficiency and improving the overall educational experience.

## Methodology

### Data Collection:



To develop an effective model for detecting timetable clashes, we primarily gathered data directly from students. This data encompassed:

* **Course Enrollments:** Information about the courses each student registered for.
* **Instructor Assignments:** Details of instructors assigned to each course.
* **Room Allocations:** Data on the rooms allocated for each class.
* **Class Schedules:** Specific days and time slots for each class session.

The data was collected through surveys and academic records, ensuring a comprehensive understanding of the existing scheduling patterns and potential conflicts experienced by students.

To augment the dataset and simulate a wider range of scheduling scenarios, we generated additional synthetic data entries. These synthetic entries were created based on the patterns and structures observed in the collected student data. This approach allowed us to:

* **Enhance Dataset Diversity:** Introduce a variety of scheduling combinations to train the model effectively.
* **Simulate Potential Clashes:** Create hypothetical scenarios that could lead to timetable conflicts, aiding the model in learning to identify such instances.
* **Balance the Dataset:** Ensure an equitable distribution of clash and non-clash instances, which is crucial for training a reliable classification model.

By combining real-world data with thoughtfully generated synthetic data, we assembled a robust dataset of approximately 200 entries. This dataset provided a solid foundation for training and evaluating our machine learning model aimed at detecting timetable clashes.

### 2. Data Preprocessing

* **Time Conversion:** Start and end times were converted to minutes for numerical processing.
* **Duration Calculation:** Duration of each class was computed.
* **Encoding:** Categorical variables were one-hot encoded to convert them into numerical format.

### 3. Feature Engineering

Features selected for model training included:

* One-hot encoded Course\_ID, Instructor\_ID, Room\_ID, Day
* Start\_Minutes
* End\_Minutes
* Duration

### 4. Model Selection

A Logistic Regression model was chosen for its simplicity and effectiveness in binary classification tasks.

### 5. Model Training

The dataset was split into training and testing sets (80% training, 20% testing). The Logistic Regression model was trained on the training set.

### 6. Model Evaluation

Model performance was evaluated using:

* Confusion Matrix
* Classification Report (Precision, Recall, F1-Score)

## Algorithm Used

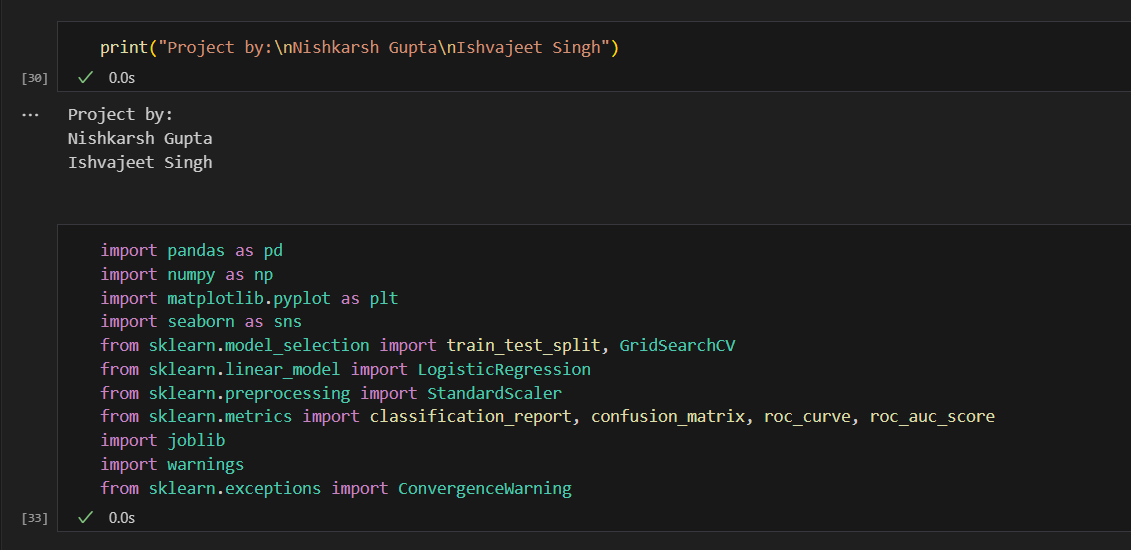
**Logistic Regression**

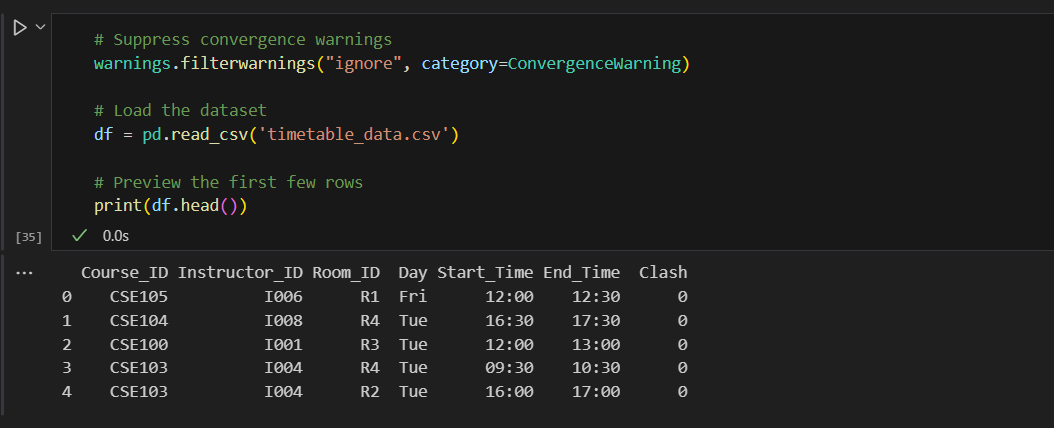
Logistic Regression is a statistical model used for binary classification tasks. It estimates the probability that a given input point belongs to a certain class. The model uses the logistic function to map predicted values to probabilities between 0 and 1.

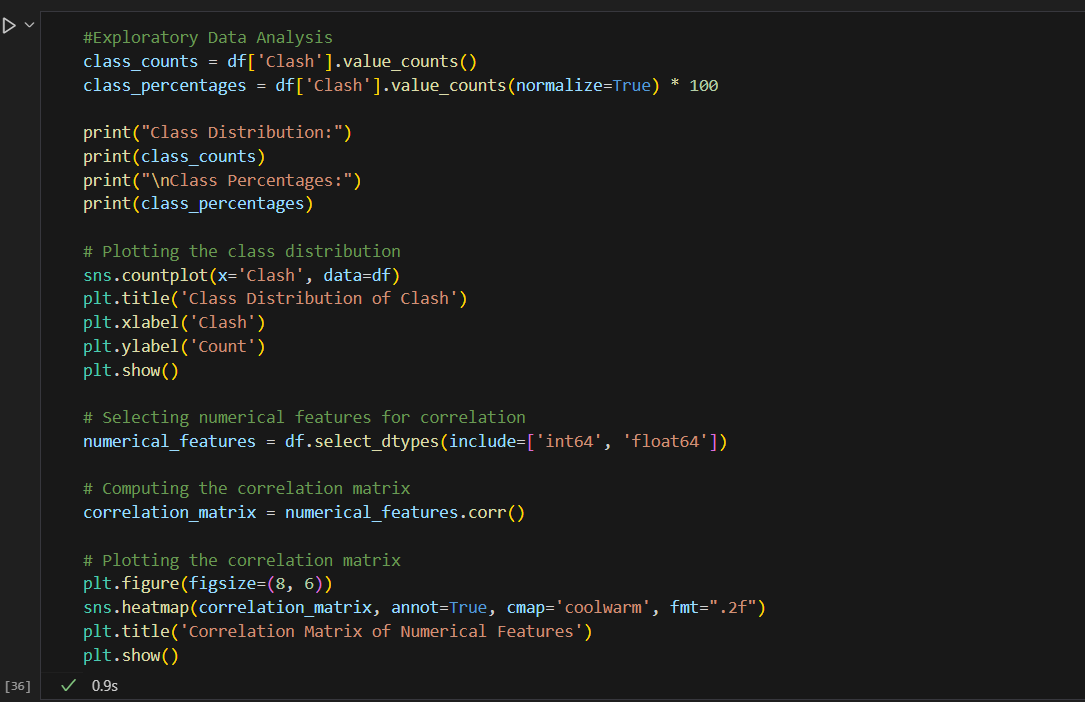
**Why Logistic Regression?**

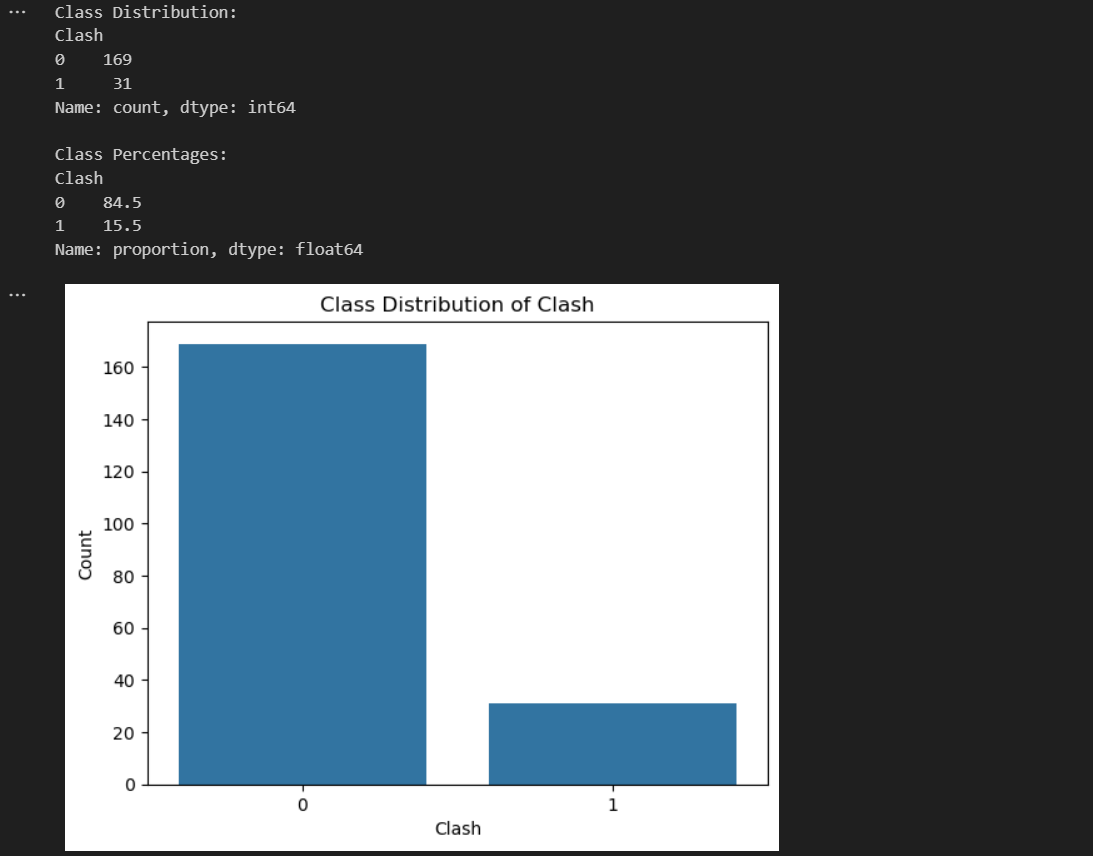
* **Interpretability:** The model provides coefficients that indicate the impact of each feature on the prediction.
* **Efficiency:** It is computationally less intensive compared to more complex models.
* **Performance:** Suitable for linearly separable data and provides good baseline performance.

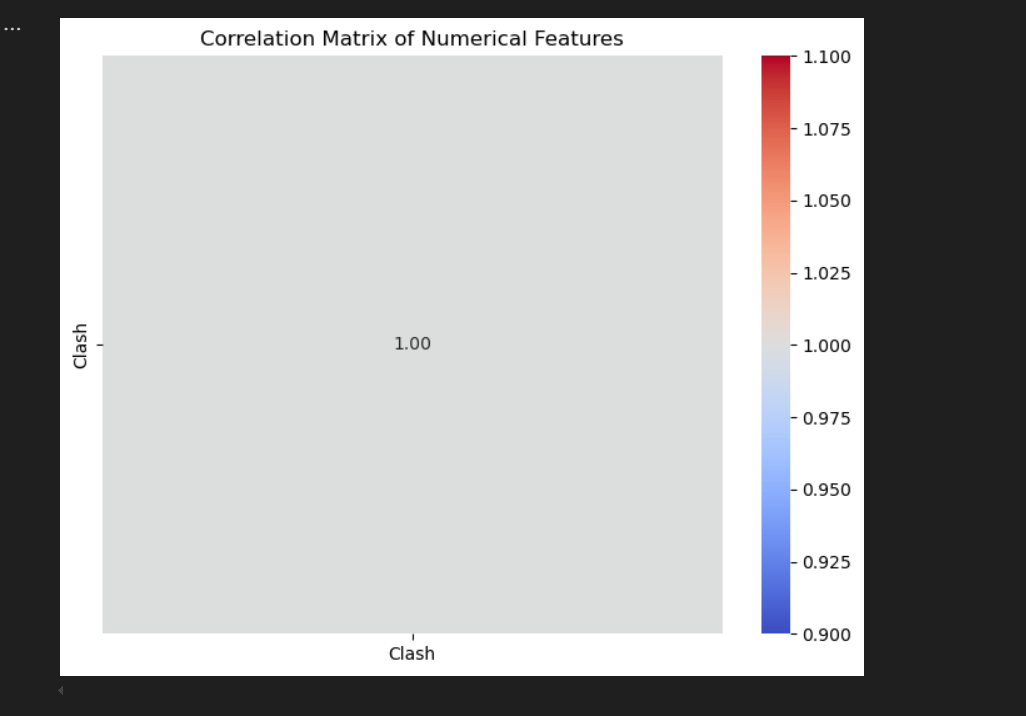
# Code:





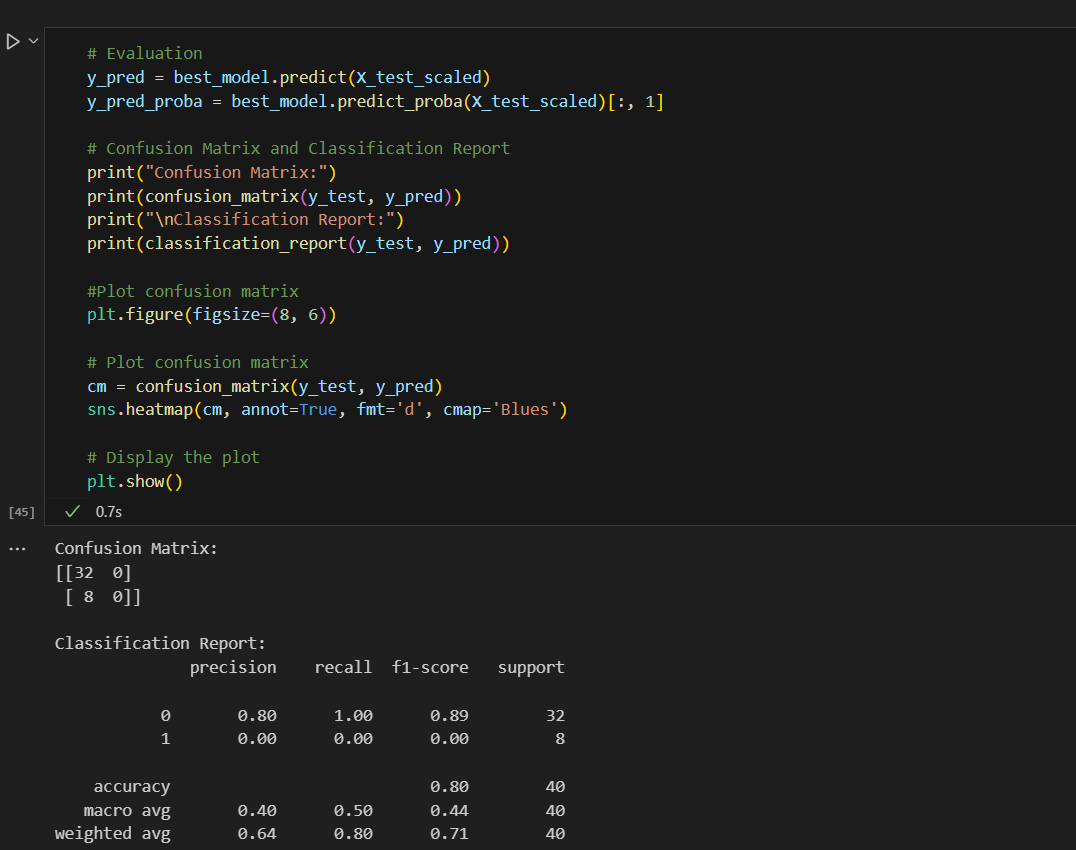


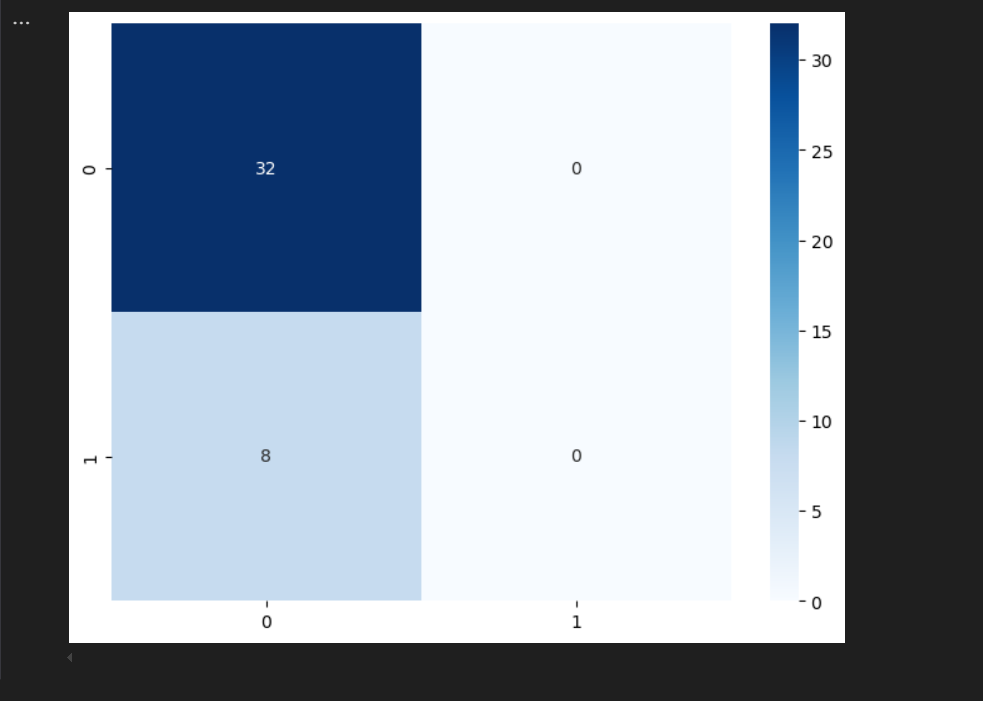


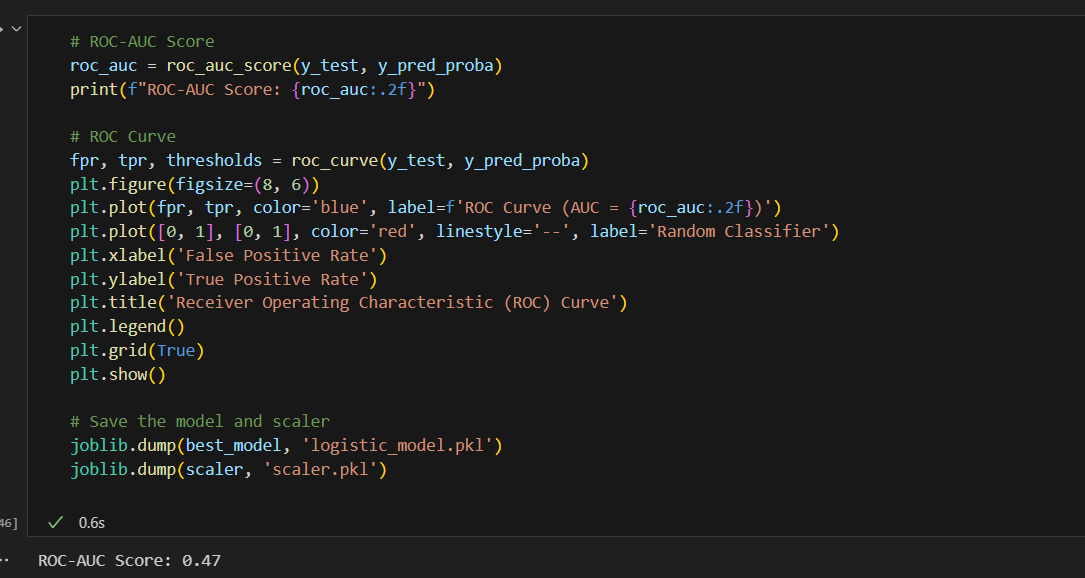


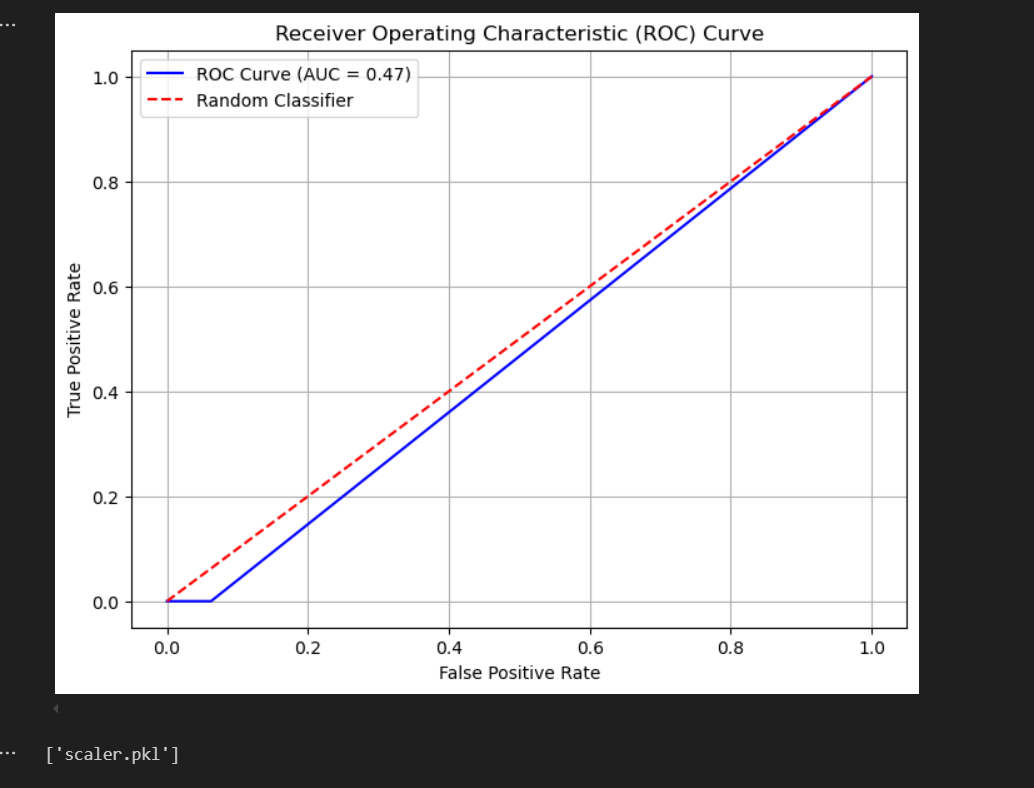


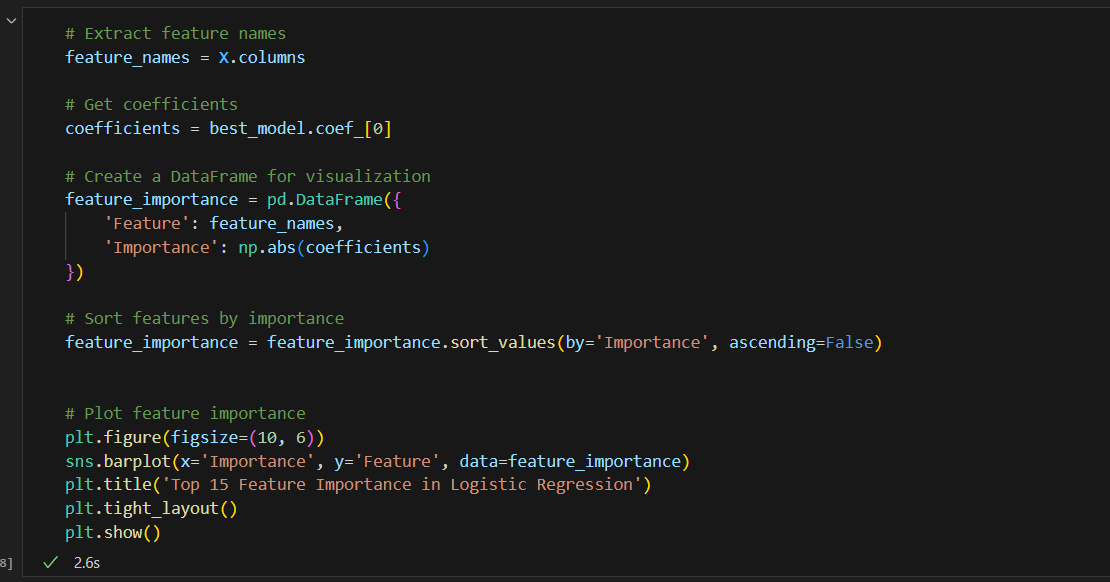


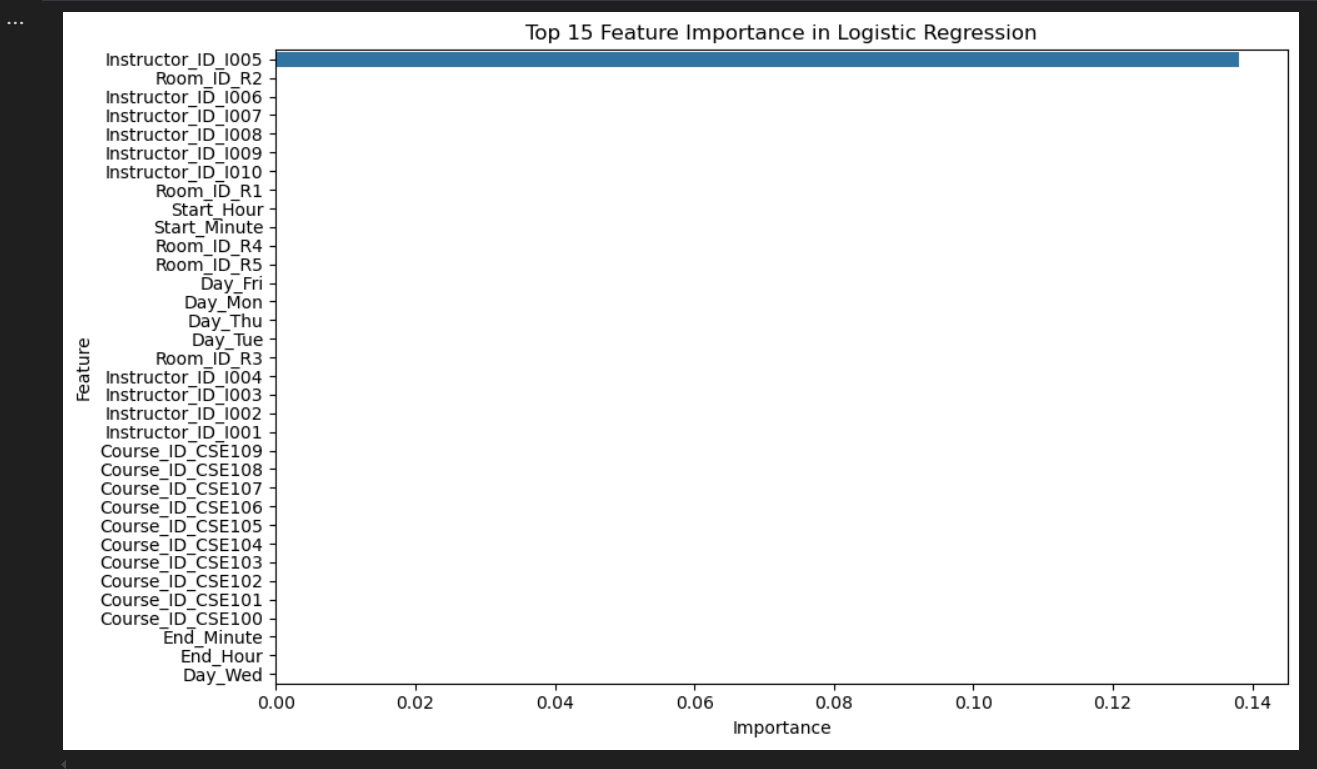




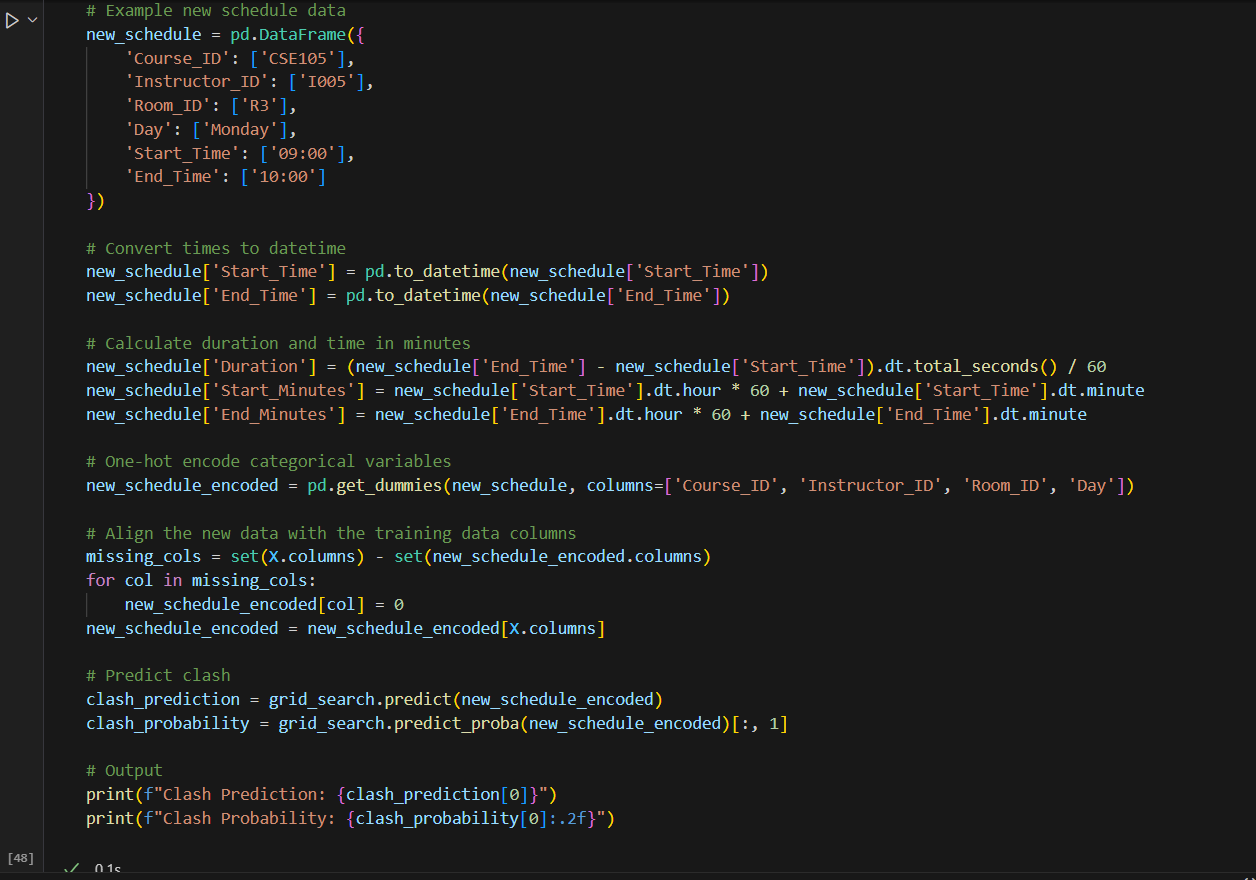


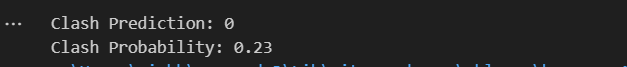






## #Code to depict the prediction of new class schedule





# Requirements

* Pandas
* Scikit-learn
* Joblib
* matplotlib

## Conclusion

This project successfully demonstrated the application of machine learning techniques to detect timetable clashes within a college setting. By collecting data primarily from students and supplementing it with synthetically generated entries, we created a comprehensive dataset that reflects real-world scheduling scenarios.​

Utilizing a Logistic Regression model, we achieved effective classification of potential scheduling conflicts. The model's performance metrics indicate its reliability in identifying clashes, thereby offering a valuable tool for academic institutions to enhance their scheduling processes.