Project Report: Student Result Predictor

Course: AI & ML

Name: SHIVANSH GUPTA

Registration Number: 25BCY10066

---

1. Introduction

For my AI/ML project, I have built a **Student Pass/Fail Predictor**. In college, students often worry about their final results based on their internal marks and attendance. I wanted to create a simple Machine Learning model that can predict if a student will **Pass** or **Fail** based on their daily habits and exam scores.

2. Problem Statement

The goal is to use historical data of students (like their study hours, attendance percentage, and internal marks) to predict the **"Final Result"**. Since the result can only be one of two things—**Pass** or **Fail**—this is a **Binary Classification** problem.

3. Tools and Libraries Used

I used the **Python** programming language because it is easy to learn and has good libraries for Machine Learning.

- **Pandas:** I used this to load and read the data from the CSV file (`student_data.csv`).
- **Scikit-learn (sklearn):** This is the main library I used to build the Logistic Regression model.
- **VS Code:** The code editor used to write and run the script.

4. Algorithm Used: Logistic Regression

I chose the **Logistic Regression** algorithm for this project. Even though the name says "Regression," it is actually used for **Classification** tasks where the output is categorical (Yes/No, True/False, 0/1).

- **Why I chose it:** My target variable is "Final Result," which has only two classes (Pass or Fail). Logistic Regression is the standard algorithm for this type of binary problem.
- **How it works:** It draws a line (decision boundary) that separates the students who passed from the students who failed.

5. How the Code Works (Step-by-Step)

My code follows a standard Machine Learning pipeline:

1. **Data Loading:** First, I created a dataset (`student_data.csv`) with columns like *Attendance_Pct*, *Internal_Test1_Mark*, and *Study_Hours*.
2. **Preprocessing:** I separated the features (inputs) from the target (output).
    - Input (X): Attendance, Marks, Study Hours.
    - Output (y): Final Result (where 1 = Pass, 0 = Fail).
3. **Splitting the Data:** I used the function `train_test_split` to divide my data into two parts:
    - **Training Data (80%):** Used to teach the model.
    - **Testing Data (20%):** Used to check if the model is correct.
4. **Training:** I used `model.fit()` to train the Logistic Regression model on the training data. This is where the machine "learns" the pattern.
5. **Testing:** Finally, I used `model.predict()` on the test data to see if it could correctly guess the results.

6. Results

After running the code, the model achieved an accuracy of **1.0 (100%)** on the test data. I also tested it with a custom input for a new student (e.g., 79% attendance, 8 hours study), and the model correctly predicted that the student would **Pass**.

7. Conclusion

This project helped me understand how Machine Learning works in real life. I learned how to load data, split it, and use a basic algorithm like Logistic Regression to solve a classification problem. It demonstrates that simple academic habits like studying and attending classes are strong indicators of the final result.