Week 4: Student Pass/Fail Classification Report

1. Objective:

The goal of this project was to predict whether a student would pass or fail based on three key features: **Hours Studied**, **Attendance Percentage**, and **Previous Exam Scores**.

2. Models Evaluated:

I trained and compared three different classification algorithms to determine which provided the best balance of accuracy and clarity:

- **Logistic Regression:** Used as a baseline model to establish linear relationships between study habits and passing.
- **K-Nearest Neighbors (KNN):** An instance-based learner that categorizes students based on how similar their data is to others.
- **Random Forest:** An ensemble method that uses multiple decision trees to improve prediction strength.

3. Performance Metrics:

Each model was evaluated using the following metrics:

- **Accuracy:** Measures the overall percentage of correct predictions.
- **F1-Score:** Used to ensure the model performs well even if the number of "Pass" and "Fail" cases is not equal.
- **Confusion Matrix:** Provided a detailed breakdown of True Positives (correctly predicted passes) and False Positives (incorrectly predicted passes).

4. Key Insights: Complexity vs. Interpretability:

- **Interpretability:** Logistic Regression was the most interpretable model. It clearly showed how much each additional hour of study increased the probability of passing.
- **Complexity:** Random Forest was the most complex and accurate model. While it captured non-linear patterns (like a student passing despite low attendance due

to high scores), it is harder to explain exactly "why" it made a specific decision compared to a simple line.