AI-Based Student Performance Prediction System

Internship Program – Artificial Intelligence / Machine Learning (AI/ML)
Internship Organization: SkillOrbit
Duration: 3 Months (First Half Project Submission)

Submitted By:

Jayesh Manish Patel
Parul Institute Of Technology

1. Introduction to Artificial Intelligence

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think, learn, and make decisions. AI enables systems to perform tasks such as problem-solving, pattern recognition, and decision-making. AI is widely applied in healthcare, finance, education, automation, and cybersecurity. In the education sector, AI can help analyze student performance and provide insights for improvement.

2. Introduction to Machine Learning

Machine Learning (ML) is a subset of AI that enables systems to learn from data without explicit programming. It identifies patterns in historical data and makes predictions. There are three major types of ML: Supervised Learning, Unsupervised Learning, and Reinforcement Learning. This project uses Supervised Learning because the output (Pass/Fail) is known during training.

3. Problem Statement

Educational institutions face challenges in predicting student performance manually. Manual analysis is time-consuming and may not provide early warning signals for weak students. There is a need for an automated AI-based system that predicts student performance based on academic parameters.

4. Objectives of the Project

The primary objective of this project is to build a machine learning model that predicts whether a student will pass or fail. Additional objectives include understanding the ML pipeline, performing data preprocessing, training a model, evaluating performance, and visualizing academic data.

5. Dataset Description

The dataset used in this project contains academic information such as Study Hours, Attendance Percentage, Previous Exam Score, and Assignment Marks. The target variable is Result (Pass/Fail). The dataset was created in CSV format and contains sufficient records for model training and testing.

6. Data Preprocessing

Data preprocessing is a crucial step in machine learning. In this project, categorical values such as Pass/Fail were converted into numerical form (1 for Pass, 0 for Fail). The dataset was split into training and testing sets using an 80:20 ratio to ensure unbiased model evaluation.

7. Machine Learning Algorithm – Logistic Regression

Logistic Regression is a supervised learning algorithm used for binary classification problems. It calculates the probability of a binary outcome using a sigmoid function. It is simple, efficient, and suitable for beginner-level AI/ML projects.

8. Model Training Process

The dataset was divided into training (80%) and testing (20%) subsets. The Logistic Regression model was trained on the training dataset and then evaluated using the testing dataset. The model learned relationships between academic features and the final result.

9. Model Evaluation

The performance of the model was evaluated using Accuracy Score, Confusion Matrix, and Classification Report. Accuracy measures the proportion of correct predictions. The confusion matrix helps identify correct and incorrect classifications.

10. Data Visualization

Data visualization techniques were used to understand patterns and relationships in the dataset. Graphs such as Study Hours vs Previous Score and Attendance vs Result were plotted using Matplotlib. Visualization helps in interpreting the model’s behavior.

11. Results and Discussion

The AI model successfully predicts whether a student will pass or fail based on input academic parameters. The model achieved satisfactory accuracy, demonstrating that machine learning can effectively assist in educational analytics.

12. Advantages of the System

The system reduces manual effort, provides early prediction of student performance, and supports decision-making for teachers and institutions.

13. Applications

This system can be used in schools, colleges, online education platforms, and academic counseling systems to monitor and predict student outcomes.

14. Conclusion

This project demonstrates the practical implementation of Artificial Intelligence and Machine Learning in the education sector. The Logistic Regression model effectively classifies student performance and provides a foundation for advanced AI-based educational systems.

15. Future Scope

In the future, this system can be enhanced by adding more parameters such as behavioral data, using advanced models like Decision Trees or Random Forest, and deploying the system as a web application.