# **PROJECT REPORT**

# AI Student Performance Predictor

**Name: Dev Ojha (E23CSEU2071) Batch-70**

**Course title: Artificial Intelligence and Machine Learning**

**Course code: CSET301**

**Name of the Instructor: Dr. Yajnaseni Dash**

**Name of the Institution: Bennett University**

## **Abstract**

The AI Student Performance Predictor is a machine learning-based web application designed to analyze student data and predict academic outcomes. Utilizing a dataset sourced from Kaggle, this project leverages classification algorithms to determine whether a student is likely to pass or fail based on demographic and academic indicators. This tool aims to assist educators in identifying at-risk students and making data-driven decisions to support academic success.

## **1. Introduction**

Predicting student performance is a critical task in educational settings, enabling timely intervention and support. With the advent of machine learning, it is now possible to build predictive models that offer valuable insights from existing data. This project presents a Flask-based web application that uses a trained classification model to predict student outcomes based on input features.

## **2. Dataset Description**

The dataset used for this project is sourced from Kaggle's "Students Performance in Exams" and contains 1000 rows with the following features:

* **gender**: male, female
* **race/ethnicity**: Group A-E
* **parental level of education**: various education levels
* **lunch**: standard, free/reduced
* **test preparation course**: none, completed
* **math score**: 0-100
* **reading score**: 0-100
* **writing score**: 0-100

An additional column "pass/fail" was derived by calculating the average of the three scores and setting a threshold.

## **3. Methodology**

### **3.1 Data Preprocessing**

* Categorical features were label encoded.
* A new binary target variable "pass/fail" was created.
* Data was scaled using StandardScaler.

### **3.2 Model Selection and Training**

* Model: Logistic Regression
* Performance evaluated using accuracy score.
* Trained model and scaler were serialized using joblib.

## **4. Web Application Architecture**

The application is built using Flask and follows a modular structure:

### **4.1 Folder Structure**

AI\_Student\_Predictor/

├── app/

│ ├── app.py

│ └── templates/

│ └── index.html

├── code/

│ ├── main.py

│ ├── best\_model.pkl

│ └── scaler.pkl

├── dataset/

│ └── student\_performance.csv

### **4.2 HTML Interface**

The index.html form collects input data and displays prediction results dynamically using Flask's Jinja templating engine.

## **5. Results and Discussion**

The logistic regression model performed effectively in predicting student performance. It identified patterns based on both demographic and academic performance features. Though simple, it achieved satisfactory accuracy and interpretability.

## **6. Conclusion**

This project demonstrates a practical application of machine learning in education. The predictor is a lightweight, user-friendly web tool that can be expanded with more advanced models and larger datasets.

## **7. Future Work**

* Integration with larger, real-time datasets.
* Model enhancement using ensemble techniques.
* UI improvements with interactive visualizations.
* Deployment to a cloud server for wider accessibility.

## **8. References**

* Kaggle Dataset:<https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>
* scikit-learn Documentation:<https://scikit-learn.org/stable/>
* Flask Documentation:<https://flask.palletsprojects.com/>