



Predicting Student Performance Using Machine Learning

Harnessing the power of data science to unlock academic potential and transform educational outcomes through intelligent predictions.

Developed BY – Baishali Debnath

Guide - Himanshu Souda



Why Predict Student Performance?



Early Intervention

Identifying at-risk students early enables timely support and targeted assistance before challenges become insurmountable.



Personalised Support

Data-driven predictions allow educators to tailor interventions specifically to individual student needs and learning patterns.



Strategic Planning

Proactive academic planning helps institutions allocate resources effectively, ensuring support reaches those who need it most.

Dataset & Key Features

Core Predictive Variables

Our model analyses multiple factors that influence academic success:

- **Study Hours:** Weekly time dedicated to coursework
- **Attendance Rate:** Class participation percentage
- **Sleep Patterns:** Average nightly rest duration
- **Previous Scores:** Historical academic performance
- **Extracurricular Activities:** Engagement beyond academics

Target Outcome

Performance Score – A continuous variable representing predicted academic achievement.

Data Source: Curated from public educational datasets available on platforms such as Kaggle, ensuring diverse and representative samples.

Data Preprocessing with Pandas & NumPy



01

Data Cleaning

Identifying and handling missing values, removing duplicate entries to ensure data integrity and reliability.

02

Categorical Encoding

Converting categorical variables into numerical format (e.g., Extracurricular Activities: Yes/No → 1/0) for model compatibility.

03

Feature Scaling

Applying StandardScaler to normalise features, ensuring uniformity across different measurement scales and improving model performance.

Clean, well-structured data forms the foundation of accurate machine learning predictions and reliable insights.

Model Evaluation & Insights

±5

Average Prediction Error

Performance scores predicted within 5 points
of actual results

0.85

R² Score

Model explains 85% of variance in student
performance

3.2

Mean Absolute Error

Average absolute deviation from true
performance values

What This Means

The **Mean Absolute Error (MAE)** measures the average magnitude of prediction errors, providing a clear picture of model accuracy in practical terms.

Practical Application

The **R² Score** reveals how well our model captures the relationships between study habits, attendance, and academic outcomes, validating its reliability.

🎓 Student Performance Prediction

Enter the student's details to predict the final score:

Study Hours
2

Attendance Percentage
60

Previous Grade
55

Family Support
yes

Extracurricular Activities
yes

Predict Final Grade

🎯 Predicted Final Score: 53.04

Streamlit Web App: Real-Time Predictions

Intuitive Input Interface

Users enter their data through interactive sliders and dropdowns for study hours, attendance percentage, sleep duration, previous scores, and extracurricular involvement.

Instant Predictions

The app processes inputs in real-time, delivering immediate performance score predictions with dynamic visualisations and clear result displays.

User-Friendly Design

Clean, responsive interface makes complex machine learning accessible to students and educators without technical expertise.

Behind the Scenes: App Architecture



User Input

Student data entered via Streamlit forms



Preprocessing

Data scaled using saved StandardScaler



Model Prediction

Trained Linear Regression model generates score



Display Output

Prediction visualised dynamically

Backend Infrastructure

Pre-trained model and scaler objects are serialised using **Pickle**, ensuring consistency between training and deployment environments whilst maintaining prediction accuracy.

Frontend Experience

Streamlit handles all user interactions, from input collection to result presentation, creating a seamless bridge between complex algorithms and user-friendly design.

Impact & Future Enhancements

Student Empowerment

Students can self-assess their study habits and make informed adjustments to optimise their academic performance.

Enhanced Features

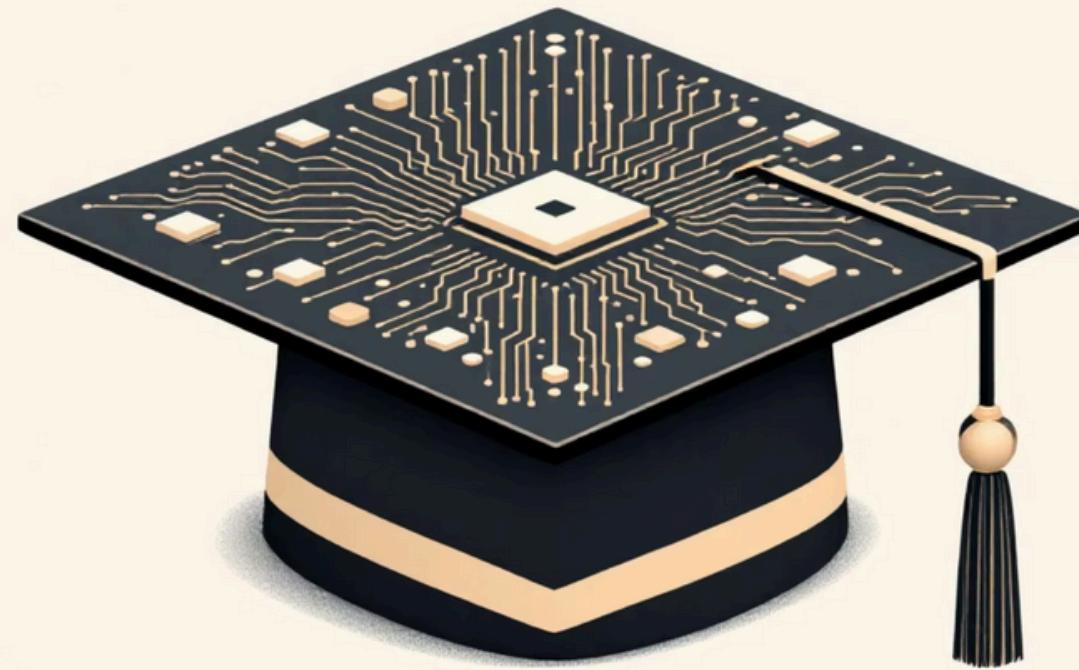
Future versions will integrate psychological factors, course difficulty ratings, and motivation metrics for richer predictions.

Continuous Improvement

Plans include cloud deployment with automated model retraining based on new data and user feedback loops.

- ❑ **Long-term Vision:** Creating an adaptive learning ecosystem where predictions become more accurate over time, evolving with educational trends and individual student journeys.

Conclusion: Data-Driven Student Success



Transforming Raw Data

Machine learning converts complex educational data into clear, actionable insights that drive meaningful improvements.

Bridging Complexity

Streamlit creates an accessible interface that brings sophisticated predictive models to everyone in the educational community.

Unlocking Potential

When technology and education unite, we create opportunities for every student to reach their fullest academic potential.

"The future of education lies in personalised, data-informed approaches that recognise and nurture individual student capabilities."



Access the Project:

GitHub Repository -

[https://github.com/baishalid03/Student Performance Prediction](https://github.com/baishalid03/Student_Performance_Prediction)