Student Performance Prediction Using Machine Learning

AI-Driven Data Engineering Mini Project

Explore how machine learning transforms academic assessment through data-driven insights

GitHub Repository

Real-World Motivation

Many institutions mandate 85% attendance, but physical presence doesn't guarantee learning. Students can attend classes without genuine engagement or comprehension.

Traditional evaluation methods are time-consuming and inherently subjective. Instructors need objective, data-driven tools to identify at-risk students early.

The solution: An AI-powered prediction system that analyzes actual academic indicators rather than just attendance records.

Key Insight

Presence ≠ Performance

Learning requires engagement, not just attendance

Problem Statement

Prediction Goal

Classify students as high performers or those requiring intervention based on comprehensive academic metrics

Challenge

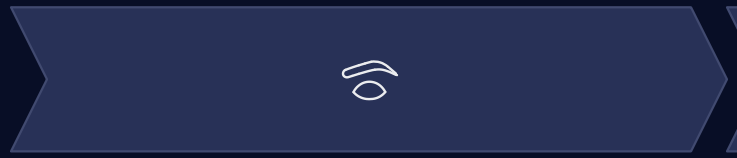
Attendance requirements make attendance an unreliable indicator of actual learning and comprehension

Focus

Identify meaningful academic performance indicators that reflect true student understanding and progress

Human-Like Decision Logic

Experienced educators naturally evaluate student performance through multiple data points. Machine learning automates this sophisticated human judgment using quantitative analysis.

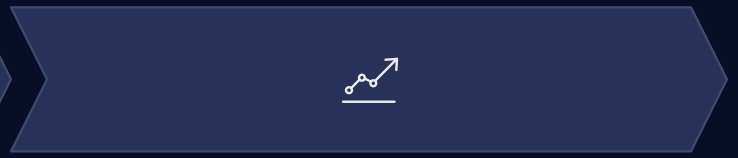


Internal Exams

Assess knowledge retention and test-taking ability

Assignments & Labs

Measure practical application and hands-on skills



Performance Trends

Track improvement or decline over semester duration

ML models learn to weigh each factor appropriately, replicating expert instructor intuition through data analysis.

Student Data Considered

Primary Indicators

High Weight Factors

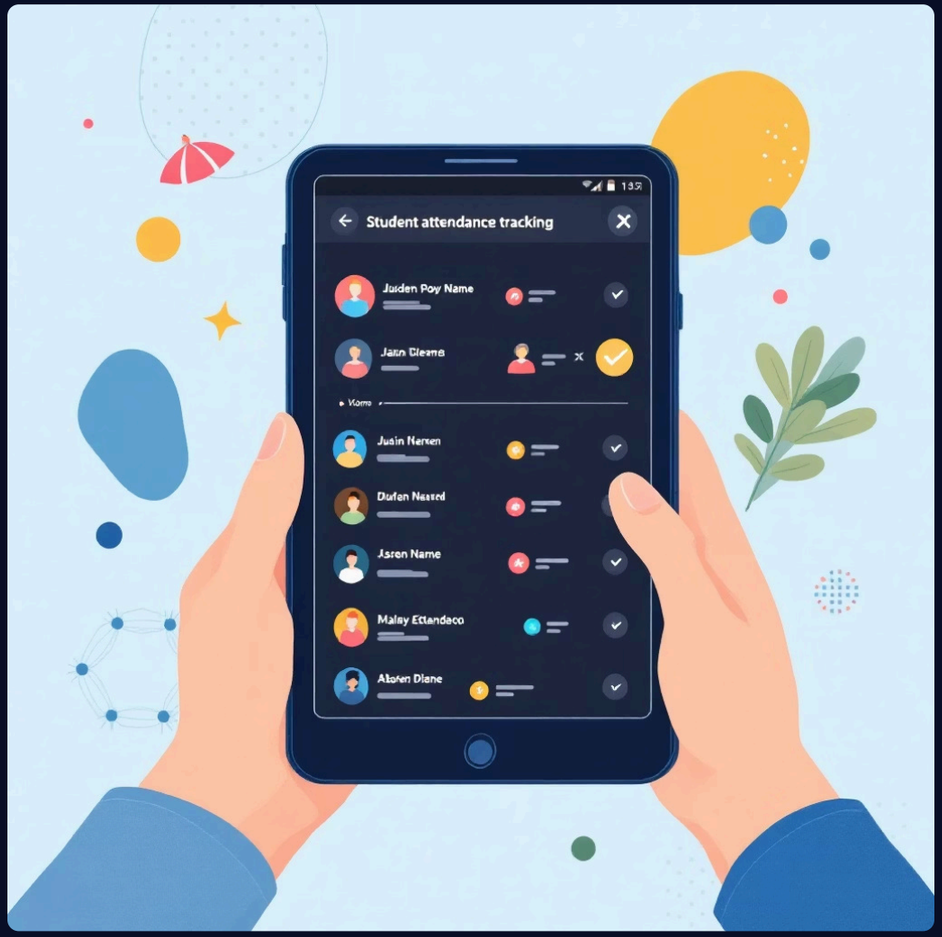
- Internal Assessment Marks
- Quiz Scores
- Assignment Submission Status
- Lab / Practical Marks
- Previous Semester Performance

These metrics directly reflect academic achievement and understanding

Supporting Indicator

Low Weight Factor

- Attendance Percentage



Attendance provides context but carries minimal predictive weight due to mandatory requirements

Tools & Technologies Used

Python & Environment

Core programming language with Jupyter Notebook for interactive development and experimentation

Data Engineering

Pandas for data collection, cleaning, and preprocessing. NumPy for efficient numerical computations

ML Framework

Scikit-learn for model training, evaluation, and prediction algorithms

Visualization

Matplotlib for performance tracking and results visualization

Data Engineering Workflow

Data Collection

Gather academic records using Pandas data structures

Data Cleaning

Remove inconsistencies and handle missing values

Feature Selection

Identify and extract relevant academic indicators

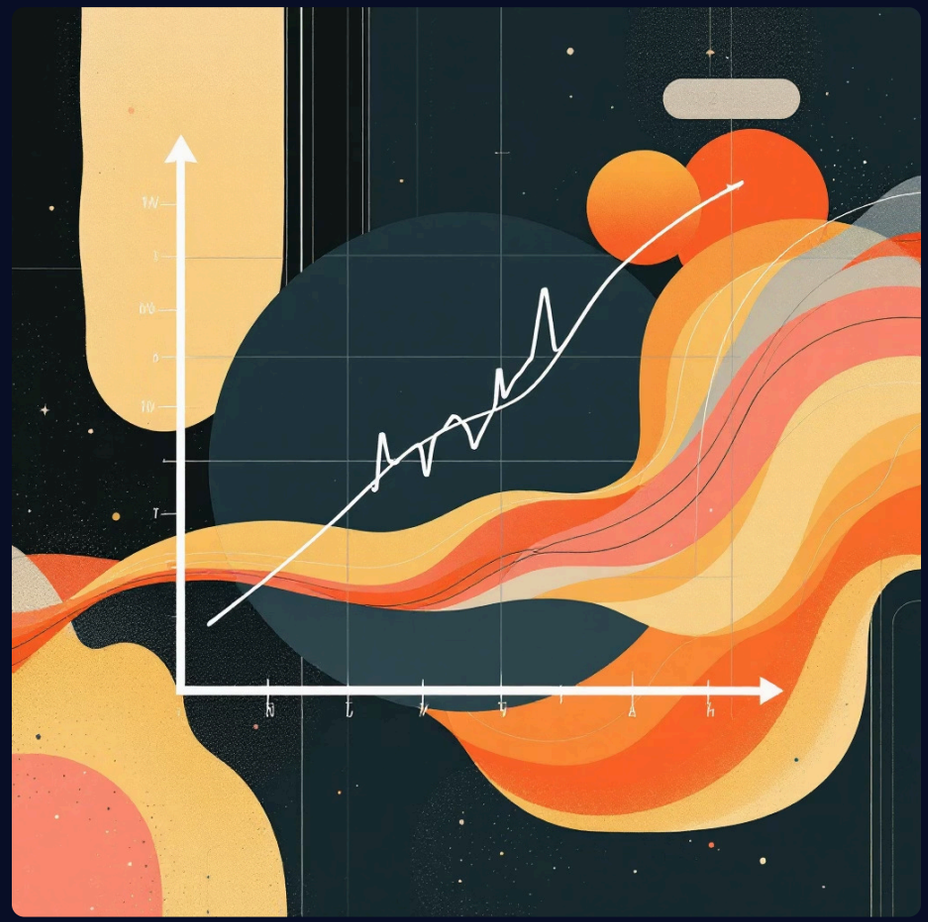
Model Training

Train ML algorithm using historical performance data

Prediction

Generate performance scores for current students

Machine Learning Model



Linear Regression Approach

Supervised Learning

Trained on historical data with known outcomes

Continuous Output

Predicts numerical final score percentages

Efficiency

Fast execution with minimal computational overhead

Automatic Weighting

Learns importance of each factor from data

Performance Categorization

Predicted scores determine intervention strategy and support level needed

Predicted Score	Category
≥ 75%	Super Performer
50% – 74%	Average Performer
< 50%	Underperformer

Key Advantage: Categorization reflects actual learning outcomes rather than attendance compliance, enabling fair and targeted academic intervention.

Conclusion

Attendance ≠ Learning Quality

Mandatory attendance policies cannot accurately measure comprehension or academic achievement

AI-Driven Prediction Works

Data engineering and ML enable realistic, objective performance forecasting

Efficient Implementation

Tools like Pandas and Scikit-learn ensure rapid development and deployment

Early Intervention Enabled

System identifies at-risk students before semester end, allowing timely support

Future Impact: Scalable framework that adapts to different institutions and academic programs