AI-201 LAB FINAL PROJECT
INSTRUCTOR: MR. AAMIR MAAROFI

Student Score Analyzer & Grade Predictor

AI-POWERED ACADEMIC PERFORMANCE ANALYSIS SYSTEM

TEAM MEMBERS:

Batool Binte Fazal (2024140)
Rida Syed (2024540)

INTRODUCTION

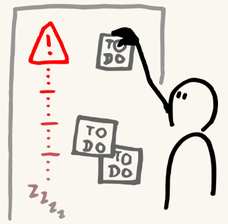
Project Overview

Problem

- Early identification of at-risk students
- Predicting performance before exams
- Manual dataset analysis is time-consuming

Solution

- Predicts scores with high accuracy
- Identifies at-risk students (Low/Medium/High risk)
- Analyzes study time, parental education, test prep impact
- One-click predictions & insights



Dataset

- Source: Kaggle - Students Performance in Exams
- Size: ~1000 records
- Demographics: gender, race/ethnicity, parental education

System Architecture and OOP Design

6 CORE CLASSES IMPLEMENTATION:

1. DataCleaner

- Removes null values and duplicates
- Encodes categorical variables using LabelEncoder
- Validates required columns
- Exception handling for invalid datasets

2. FeatureEngineer (NumPy)

- Performance Index:

PI = 0.5×math + 0.3×reading + 0.2×writing

- Z-score outlier detection
- Percentile ranking using np.percentile()
- Risk score normalization

3. ModelTrainer

- Trains 3 ML models with StandardScaler pipeline
- Train/test split (80/20)
- Model evaluation with multiple metrics

4. Visualizer

- Generates 4 Matplotlib plots with advanced styling
- 3D scatter using Axes3D, gradient coloring
- Statistical overlays and correlation analysis

5. Predictor

- Real-time predictions (score, grade, pass/fail, risk level)
- Grade conversion (A/B/C/D/F)

6. ModelPersistence

- Saves/loads models using Pickle
- Persistent scaler and feature names

LIBRARIES USED

MACHINE LEARNING MODELS & ADVANCED FEATURES

Three ML Models:

Linear Regression:

- Score prediction (0-100)
- Metrics: RMSE, R^2

Logistic Regression:

- Pass/Fail classification
- Accuracy, max_iter=1000

K-Nearest Neighbors:

- Risk assessment
- KNeighborsClassifier (k=5)

Advanced NumPy Operations:

- Vectorized computations for efficiency
- Z-score: identifies outliers where $|z| > 2$
- Percentile ranking: compares student to entire cohort
- Polynomial interactions: study_time × parental_education
- Weighted performance index using NumPy arrays

Data Visualization using Matplotlib:

- Histogram with gradient coloring
- Bar chart with error bars and statistical overlays
- Donut pie chart for demographics
- 3D scatter plot (Study Time × Parental Education × Score) with correlation statistics

Technology Stack and Results

Core Libraries:

- NumPy: Advanced numerical computations, vectorized operations
- Pandas: Data analysis, GroupBy operations, Q&A style analytics
- Matplotlib: 4 visualizations including 3D scatter (mpl_toolkits.mplot3d)
- Scikit-learn: ML models, StandardScaler, train_test_split
- Streamlit: Interactive web dashboard
- Pickle: Model persistence and reloading

Project Statistics:

- 📊 ~1000 students analyzed
- ⚙️ 12+ engineered features
- 🤖 3 ML models trained
- 📈 4 advanced visualizations
- 💾 Complete model persistence

Project Summary

Key features

- Complete ML Pipeline: Data cleaning → Feature engineering → Training → Prediction
- Advanced Analytics: Percentile ranking, z-score outliers, polynomial features
- Interactive Dashboard: Streamlit web app with CSV upload & real-time predictions
- Production Ready: Exception handling, model save/load, scalable OOP design

Real-World Impact:

This system enables educational institutions to make data-driven decisions, identify struggling students early, and provide timely interventions - transforming raw academic data into actionable insights that improve student outcomes.

AI-201 LAB FINAL PROJECT
INSTRUCTOR: MR. AAMIR MAAROFI

Thankyou!

TEAM MEMBERS:

Batool Binte Fazal (2024140)
Rida Syed (2024540)