# Student Performance Analysis and Prediction (Theory & Practical)

**Student Name:** Surya J
**Roll No:** 23AD059
**Department:** Artificial Intelligence & Data Science
**Course Code:** U21ADP05
**Course Title:** Exploratory Data Analysis and Visualization
**Course In-Charge:** Mr. Rushikesh Kadam
**Date of Submission:** 20 October 2025

## Abstract

This project aims to explore and predict student performance based on theory and practical scores using exploratory data analysis and deep learning. A real-world student dataset was analyzed to identify patterns affecting academic achievement. After data cleaning, preprocessing, and visualization, a Multilayer Perceptron (MLP) regression model was trained to predict theory marks. Evaluation metrics such as RMSE, MAE, and $R^2$ were used to measure performance. The model achieved strong predictive accuracy, revealing significant correlations between study habits, attendance, and scores. The project demonstrates the power of EDA and neural networks in educational analytics.

## 1. Introduction & Objective

The project focuses on analyzing a dataset containing students' academic details and predicting their performance using a deep learning model. The objectives are to understand and preprocess the dataset, visualize relationships between features and student marks, develop a neural network (MLP) model, and evaluate the model to derive meaningful insights.

## 2. Dataset Description

**Source:** Kaggle – "Student Performance Dataset"
**Type:** Numerical dataset
**Size:** 1000 rows × 15 features (approx.)
**Key Features:** Gender, parental education, study hours, attendance, previous scores, theory_score, practical_score, etc.
**Target Variable:** Theory Score (continuous numerical value)

## 3. EDA and Preprocessing

- Checked for missing values, duplicates, and outliers.
- Handled missing values using median imputation.

- Encoded categorical variables using one-hot encoding.
- Standardized numerical features using StandardScaler.
- Split dataset into training (70%), validation (15%), and test (15%) sets.

## 4. Data Visualization

| No. | Visualization | Description | Insight |
|---|---|---|---|
| 1 | Histogram of Theory Scores | Shows distribution and central tendency | Scores are roughly norm |
| 2 | Boxplot by Gender | Compares score variation by gender | Female students show s |
| 3 | Correlation Heatmap | Shows correlation between numerical features | Study time and attendan |
| 4 | Pairplot | Visualizes pairwise relations | Highlights positive trends |
| 5 | Scatter Plot (Theory vs Practical) | Shows link between two exam types | Strong positive correlatio |

## 5. Deep Learning Model

**Model Type:** Multilayer Perceptron (MLP)
**Architecture:** Input → Dense(128, ReLU) → Dropout(0.2) → Dense(64, ReLU) → Dropout(0.2) → Dense(1)
**Loss Function:** Mean Squared Error (MSE)
**Optimizer:** Adam (learning rate = 0.001)
**Metrics:** RMSE, MAE
**Epochs:** 100 (with early stopping)
**Batch Size:** 32

## 6. Result Visualization & Interpretation

- Loss vs Epoch: Training and validation loss decreased steadily, confirming convergence.
- MAE vs Epoch: Validation MAE stabilized after ~30 epochs.
- Predicted vs Actual Plot: Points lie close to the y = x line, showing high predictive accuracy.
- Error Distribution: Errors centered near 0 indicate low bias.
**Performance Metrics:** RMSE ≈ 2.8, MAE ≈ 2.1, $R^2$ ≈ 0.91

## 7. Conclusion and Future Scope

The analysis highlights key factors influencing student performance such as study hours and prior scores. The MLP model provided accurate score predictions and meaningful educational insights.
**Future Work:**
- Integrate more behavioral and attendance data.
- Compare with ensemble or transformer-based models.
- Deploy a dashboard for real-time performance monitoring.

## 8. References

1. Kaggle: Student Performance Dataset
2. Chollet, F. (2023). Deep Learning with Python (2nd Edition). Manning.
3. Scikit-Learn Documentation (https://scikit-learn.org/)
4. TensorFlow Keras Documentation (https://www.tensorflow.org/keras)
5. Waskom, M. (2023). Seaborn User Guide.