# *PREDICTING STUDENT PERFORMANCE USING MACHINE LEARNING*

INTRODUCTION TO MACHINE LEARNING (BIT4333)

Presented by:

CAROLLISA WADIWILL - 202501010307

# *EXECUTIVE SUMMARY*

- Goal: Predict student exam scores using regression models.
- Dataset: Student demographics, behavior, academic info.
- Models: Linear Regression, SVR, XGBoost, CatBoost.
- Best Model: Linear Regression – MSE = 3.256, $R^2$ = 0.770
- Key predictors: Attendance, Hours Studied, Previous Scores
- Streamlit app demonstrates interactive score predictions.

# *PROBLEM STATEMENT*

1. Understand key factors affecting student performance.
2. Predict exam scores accurately using ML regression.
3. Demonstrate interactive simulation of predictions for various scenarios.

# DATASET OVERVIEW

- Key Features:
- Hours_Studied, Attendance, Previous_Scores
- Motivation_Level, Parental_Involvement, Sleep_Hours
- Optional Features: Access_to_Resources, Extracurricular Activities, Tutoring, Internet Access, Teacher Quality, Peer Influence, Family Income, School Type, Physical Activity
- Source: Kaggle – Student Performance Factors

# METHODOLOGY

## Data Preprocessing

- Fill missing values (median/mode)
- One-hot encoding for categorical features
- Standardization for numeric features

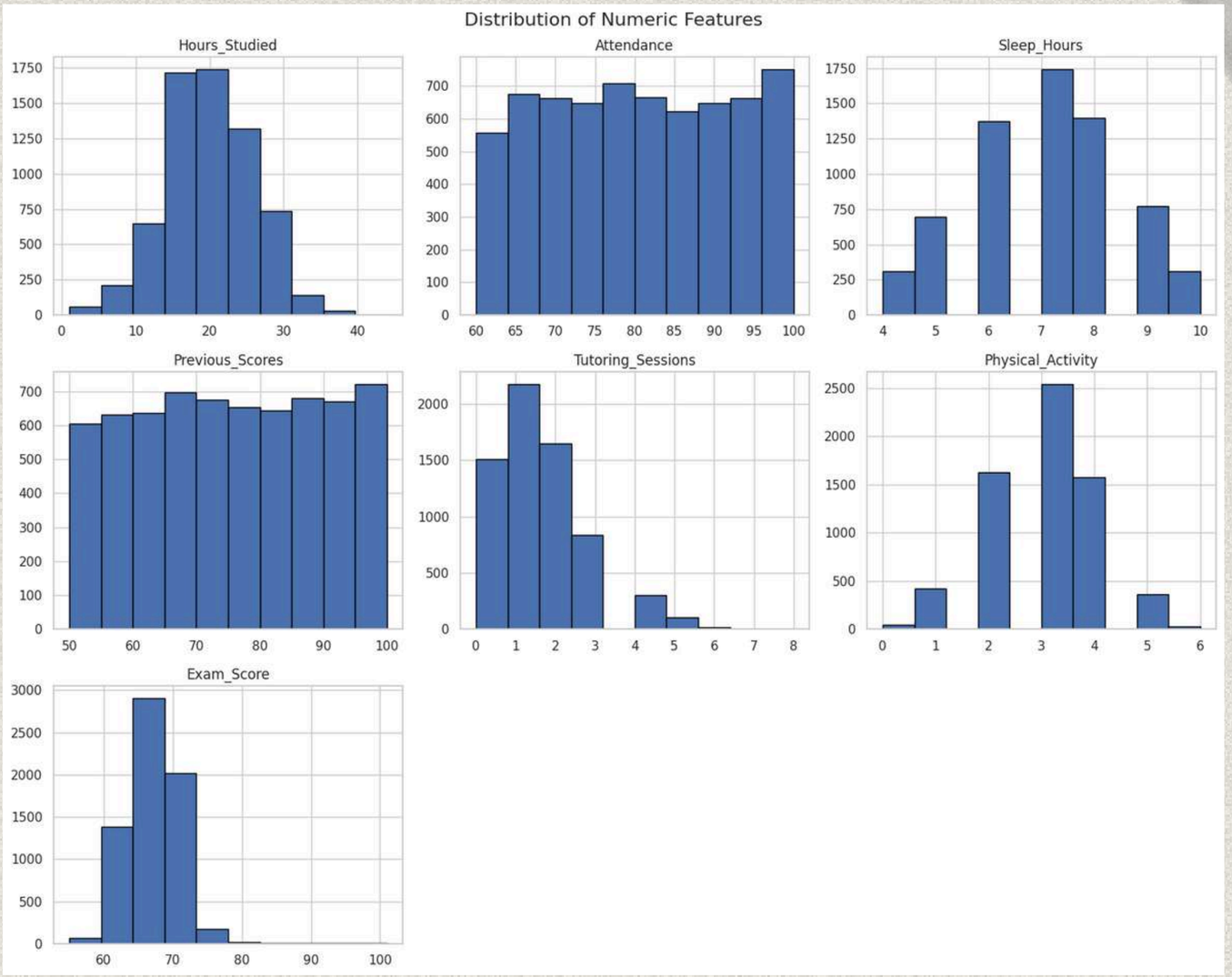
## Feature Selection

Main features: Hours_Studied, Attendance, Previous_Scores, Motivation_Level, Parental_Involvement, Sleep_Hours

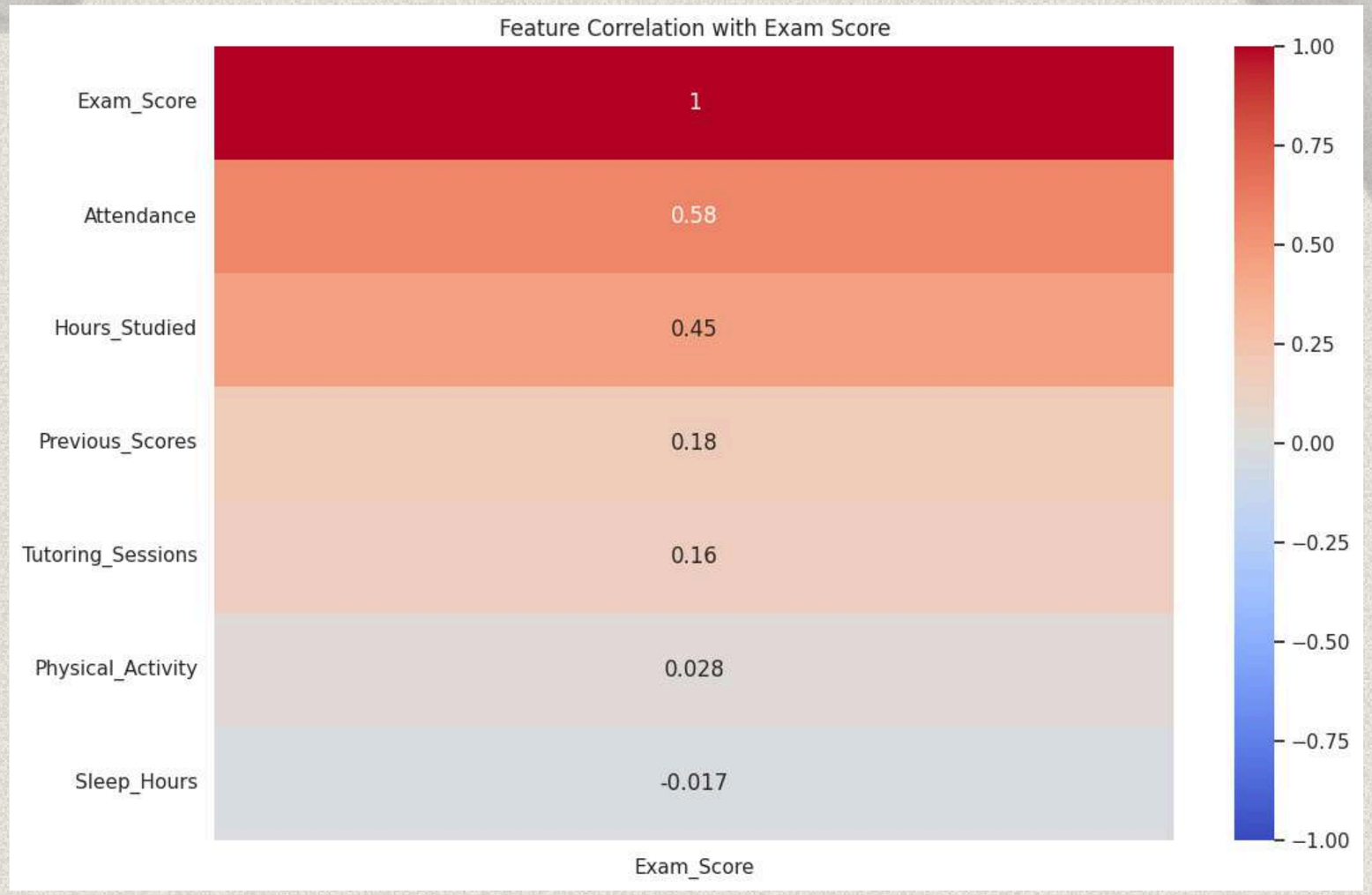
## Model Development

- Linear Regression, SVR, XGBoost Regressor, CatBoost Regressor
- Best: Linear Regression (MSE = 3.256, $R^2$ = 0.770)

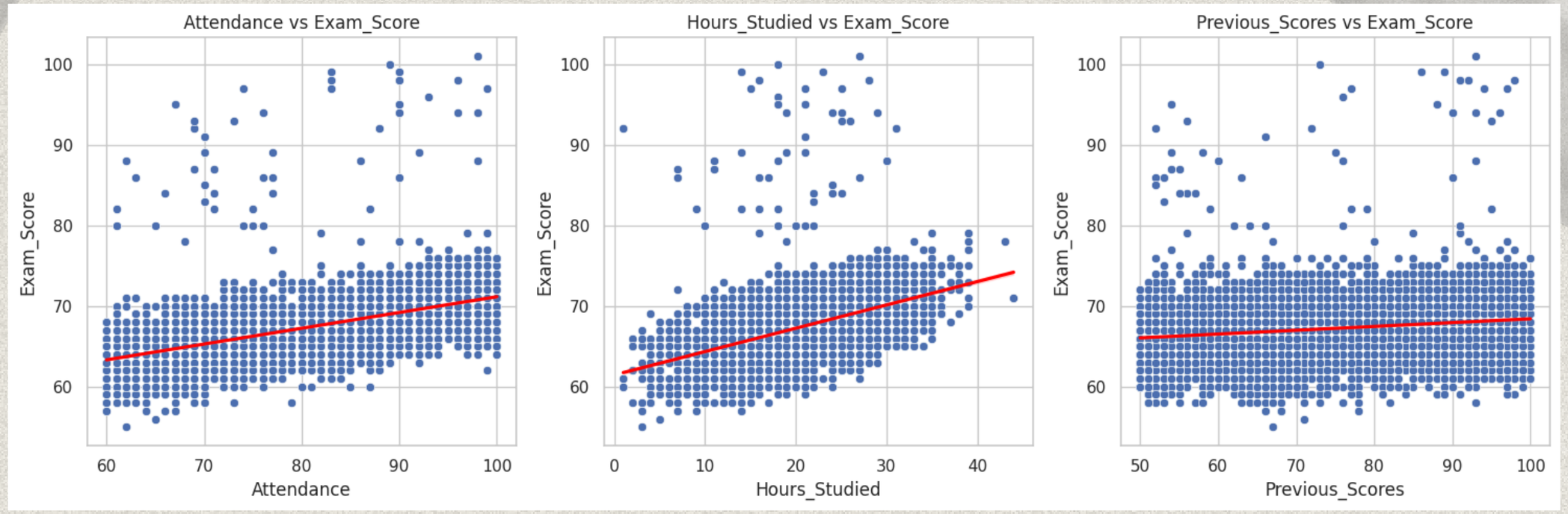
# VISUALISATION - HISTOGRAM



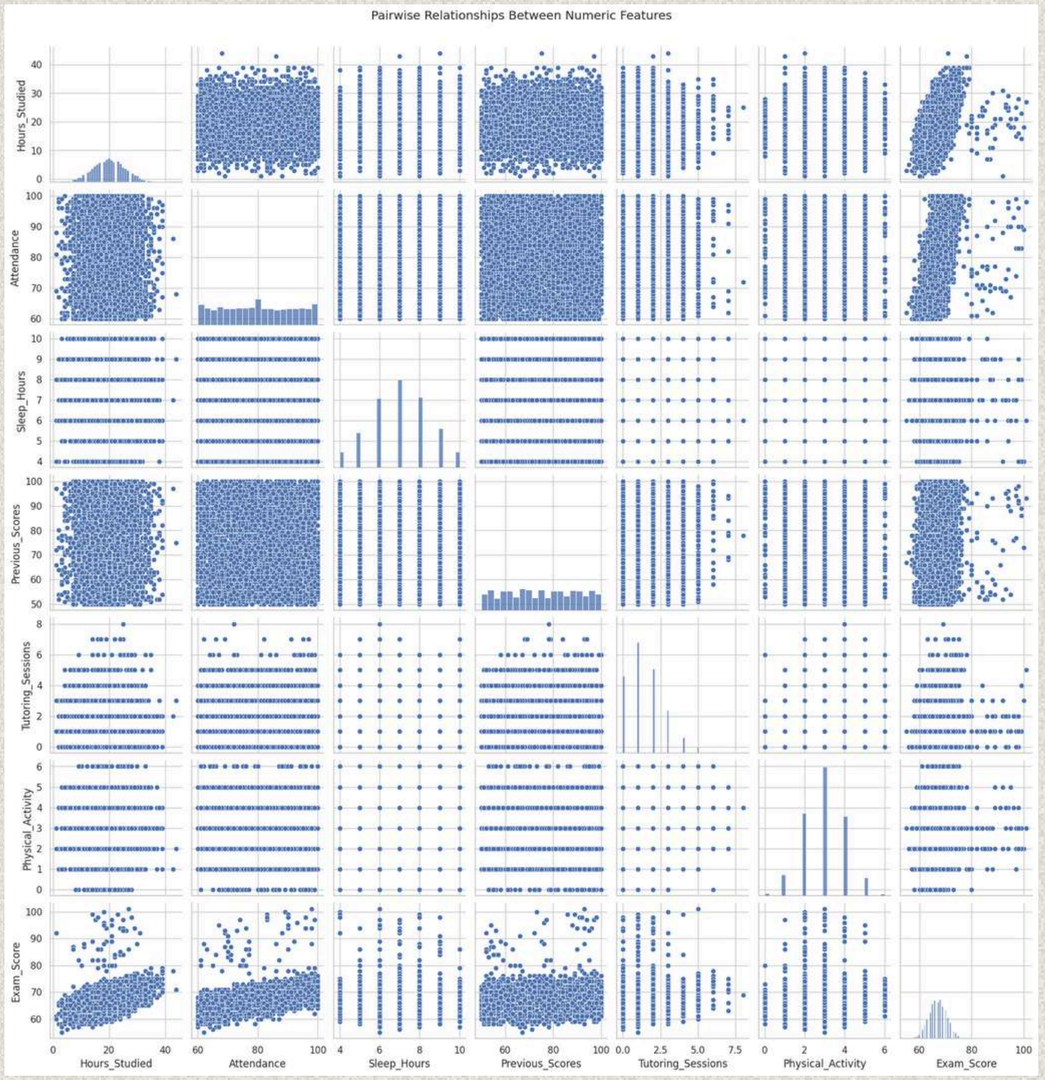
# VISUALISATION - CORRELATION HEATMAP



# VISUALISATION - SCATTERPLOT



# VISUALISATION - PAIRPLOT



# MODEL EVALUATION

| MODEL | MSE | R2 |
|---|---|---|
| | | |
| Linear Regression | 3.256 | 0.770 |
| SVR | 3.380 | 0.761 |
| XGBoost | 4.614 | 0.674 |
| CatBoost | 3.734 | 0.736 |

## TOP FEATURES

- Attendance – 2.29
- Hours_Studied – 1.76
- Previous_Scores – 0.71

# STREAMLIT APP

- Interactive exam score predictions
- Input main + optional features
- Features encoded & scaled to match training
- Shows predicted Exam_Score in real-time

# EXAMPLE INPUT

- Hours_Studied: 20
- Attendance: 85
- Previous_Scores: 70
- Motivation_Level: Medium
- Parental_Involvement: High
- Sleep_Hours: 7

# *CONCLUSION*

- Linear Regression is effective and interpretable.
- Key predictors: Attendance, Hours Studied, Previous Scores
- Streamlit app demonstrates interactive, hypothetical predictions
- Model helps identify patterns and potential interventions (demo purposes)

# ACKNOWLEDGEMENT


- Thank Sir Nazmirul Izzad Bin Nassir for guidance and support
- Gratitude to peers and dataset contributors

# *Thank You*

For Your Attention