# Report: Implementation of Predictive Models for Exam Scores

## Objective

The primary objective of this analysis was to develop predictive models that estimate student exam scores based on two key factors:  
- Hours Studied  
- Attendance Percentage  
  
This predictive capability aims to identify at-risk students and allow teachers to intervene early, providing additional support to improve their academic performance.

## Data Overview

The dataset consisted of three key variables:  
1. Hours\_Studied: Number of hours a student studied.  
2. Attendance: Percentage of classes attended by a student.  
3. Exam\_Score: The target variable indicating the student's score in the exam.

Key observations:

- Hours\_Studied: The data followed a roughly normal distribution, with most students studying between 10 and 30 hours.

- Attendance: Attendance levels were diverse, ranging from 60% to 100%, with a significant peak at 100%.

- Exam\_Score: Exam scores were left-skewed, with most students scoring between 60 and 75.

## Methodology

The workflow included the following steps:  
1. Data Preprocessing:  
 - Relevant columns (Hours\_Studied, Attendance, Exam\_Score) were selected.  
 - The data was split into training, validation, and test sets (80%-10%-10%).  
2. Model Selection:  
 - Two predictive models were implemented:  
 - Decision Tree Regressor  
 - Linear Regression  
3. Model Evaluation:  
 - Metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Accuracy (percentage of predictions within ±5 of the actual score) were used to evaluate model performance.

## Results

### Decision Tree Regressor

The Decision Tree model was trained with a maximum depth of 5 to balance interpretability and overfitting.

Performance:

- Training Set:  
 - MSE: 7.18  
 - MAE: 1.62  
 - Accuracy: 98.49%  
- Validation Set:  
 - MSE: 5.48  
 - MAE: 1.55  
 - Accuracy: 99.24%  
- Test Set:  
 - MSE: 7.15  
 - MAE: 1.63  
 - Accuracy: 98.49%

The Decision Tree achieved high accuracy and low error across all datasets. Its interpretability allowed tracing the decision path for individual predictions, providing insight into how the model arrived at its predictions.

### Linear Regression

Linear Regression was implemented to model the relationship between hours studied, attendance, and exam scores assuming linearity.

Performance:

- Test Set:  
 - MSE: 5.81  
 - R²: 0.59

While the Linear Regression model had a slightly lower MSE on the test set compared to the Decision Tree, it struggled to capture non-linear relationships or interactions between features, resulting in lower overall accuracy.

### Example Predictions

Using both models, predictions were made for a student who studied 30 hours and had 0% attendance:  
- Decision Tree Prediction: 67.77  
- Linear Regression Prediction: 54.26  
  
The Decision Tree captured the interactions between study hours and attendance better, leading to a more plausible prediction.

## Conclusion

- Decision Tree: The preferred model due to its ability to model non-linear relationships and interactions between features, resulting in better accuracy and interpretability.  
- Linear Regression: While simpler, it was less effective due to its linear assumption.

The analysis highlights the importance of attendance and study hours in determining student performance. By focusing on these factors, teachers can identify and support at-risk students. However, exceptional performers (outliers) cannot be fully explained by the data provided, indicating the need for additional features such as prior knowledge, study techniques, or cognitive ability.