Artificial Intelligence & Machine Learning – Task 2 Report

1. Objective

The objective of this task is to perform feature engineering, apply feature scaling, train multiple regression models, optimize their performance, and compare them using appropriate evaluation metrics such as RMSE and R^2 score.

2. Dataset Description

The California Housing Dataset from scikit-learn is used for this task. It contains housing-related features such as median income, house age, average rooms, population, and location-based attributes. The target variable is the median house value.

3. Feature Engineering

Feature engineering includes inspecting data distributions, handling missing values, removing outliers if necessary, and selecting relevant features. Correlation analysis helps in understanding feature importance.

4. Feature Scaling

Feature scaling is applied using StandardScaler or MinMaxScaler to normalize numerical features. This step is crucial for models like Linear and Ridge Regression to ensure faster convergence and better performance.

5. Model Implementation

The following regression models are implemented and trained: Linear Regression, Ridge Regression, and Decision Tree Regressor. Each model is trained using the scaled training dataset.

6. Model Evaluation

Model performance is evaluated using Root Mean Square Error (RMSE) and R^2 Score. Lower RMSE and higher R^2 values indicate better model performance.

7. Performance Comparison

After evaluation, models are compared based on their metrics. Linear and Ridge Regression provide stable results, while Decision Tree may capture non-linear patterns but risks overfitting.

8. Conclusion

This task demonstrates the importance of preprocessing, feature scaling, and model comparison in machine learning workflows. Proper evaluation ensures the selection of the most suitable model for prediction tasks.