Alireza Mohammadshafie individual group. Project Proposal

Introduction

This project focuses on developing a reliable machine learning model to predict housing prices in California using diverse property and demographic features. Accurate housing price prediction helps buyers, sellers, and policymakers make informed decisions.

Objective

Our primary goal is to create a predictive model that estimates median house prices with high accuracy. We aim to explore data, engineer meaningful features, select appropriate algorithms, and fine-tune models to optimize performance.

Methods

We will utilize the California Housing Dataset from scikit-learn, performing thorough data exploration to identify patterns, handle missing values, and remove outliers if necessary. Feature engineering will include creating new variables and applying scaling techniques to improve model input quality.

Multiple regression algorithms will be implemented and evaluated using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared. Model hyperparameters will be tuned methodically to enhance accuracy. The best-performing model will be selected after comparing results on test data.

Evaluation

Model performance will be assessed with standard regression metrics to ensure robust and generalized predictions. The final model's ability to accurately predict housing prices on unseen data will be demonstrated and analyzed.

References

1-https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/

2-https://www.geeksforgeeks.org/what-is-regression-analysis/