Project Proposal: House Price Prediction Using Machine Learning

Yehonatan Fridman I.D 207689647

Hagay Knorovich I.D 318813367

Introduction

The objective of this project is to develop a predictive model to accurately estimate house prices utilizing various features such as size, location, and maintenance condition. This endeavor will employ multiple machine learning algorithms to extract and analyze patterns from the data, ultimately enhancing the prediction accuracy.

Dataset

The "Housing Prices Dataset" dataset available on Kaggle ([link](https://www.kaggle.com/datasets/yasserh/housing-prices-dataset)) has been selected for this project. It includes a comprehensive set of features related to housing that affect pricing, offering a robust framework for analyzing and applying different predictive models.

Objectives

1. Conduct data analysis and preprocessing to handle missing values and encode categorical variables appropriately.
2. Implement four distinct machine learning algorithms to construct predictive models.
3. Critically evaluate and compare the models' performances using suitable metrics.

Methodology

The following machine learning algorithms will be utilized:

1. Linear Regression: A basic yet powerful regression technique that assumes a linear relationship between the dependent and independent variables.
2. Decision Trees: A versatile, non-linear approach that splits data into branches to make predictions based on variable values.
3. Support Vector Regression (SVR): A robust regression technique that utilizes support vectors to predict continuous outcomes effectively.
4. K-Nearest Neighbors (KNN): A straightforward, non-parametric learning technique that predicts values based on the closest data points in the feature space.

Expected Outcome

The project aims to discern the most effective machine learning algorithm for predicting house prices from the given dataset. We will assess the models based on their accuracy, computational requirements, and overall robustness.

Tools and Technologies

* Programming Language: Python
* Libraries: Pandas, NumPy, scikit-learn, matplotlib, seaborn

Conclusion

This project endeavors to augment understanding and application of machine learning algorithms in practical scenarios such as housing price predictions. Comparing multiple algorithms will reveal their relative strengths and weaknesses, providing valuable insights for future applications.