PHASE 1 ASSIGNMENT

PROJECT TITLE: PREDICTING HOUSE PRICES USING MACHINE LEARNING

PROBLEM DEFINITION: The problem is to predict house prices using machine learning techniques. The objective is to develop a model that accurately predicts the prices of houses based on a set of features such as location, square footage, number of bedrooms and bathrooms, and other relevant factors. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

GITHUB LINK:

https://github.com/achu031122/Predicting-House-Prices-using-Machine-Learning.git

DOCUMENT:

Predicting House Prices Using Machine Learning

Introduction

The real estate market is one of the most dynamic and complex sectors of the economy. Predicting house prices accurately is crucial for buyers, sellers, and real estate professionals. Machine learning (ML) provides a powerful toolset for analyzing data and making predictions. In this document, we will outline the steps to build a machine learning model for predicting house prices.

1. Understanding the Problem

1.1 Problem Statement

The goal is to develop a model that can predict the selling price of a house based on its features. This involves understanding the relevant factors that influence house prices, gathering data, and building a predictive model.

1.2 Data Collection

Collect a comprehensive dataset containing information about houses, including features such as square footage, number of bedrooms, location, etc. Historical sales data can be a valuable resource.

2. Exploratory Data Analysis (EDA)

2.1 Data Exploration

Examine the dataset's structure and dimensions.

Identify missing values and outliers.

Explore the distribution of the target variable (house prices) and features.

2.2 Feature Engineering

Create new features that might enhance the predictive power of the model.

Handle missing values through imputation or removal.

Convert categorical variables into numerical representations.

3. Data Preprocessing

3.1 Scaling

Normalize or standardize numerical features to bring them to a similar scale. This ensures that no single feature dominates the learning process.

3.2 Splitting the Data

Divide the dataset into training and testing sets. The training set is used to train the model, and the testing set evaluates its performance.

4. Model Selection

4.1 Choose Algorithms

Select regression algorithms suitable for predicting continuous variables. Common choices include Linear Regression, Decision Trees, Random Forests, and Gradient Boosting.

4.2 Model Training

Train the selected models using the training dataset.

5. Model Evaluation

5.1 Metrics

Choose appropriate evaluation metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE) to assess the model's performance.

5.2 Validation

Use the testing dataset to evaluate the model's performance on unseen data.

6. Hyperparameter Tuning

Optimize the model's hyperparameters to improve its performance. This can be done using techniques like grid search or random search.

7. Deployment

Once satisfied with the model's performance, deploy it for making predictions on new data. This could involve integrating the model into a web application or a real estate platform.

8. Monitoring and Maintenance

Continuously monitor the model's performance and update it if necessary. Real estate markets can change, and models may need periodic retraining.

Conclusion

Building a machine learning model to predict house prices involves a systematic process of understanding the problem, preparing the data, selecting and training models, and evaluating their performance. By following these steps, you can create an effective predictive tool for the real estate market.

SUBMITTED BY,

STUDENT REG NO: 711221104003

NAAN MUDHALVAN: au711221104003