**PHASE – I**

**PROBLEM DEFINITION AND DESIGN THINKING**

|  |  |
| --- | --- |
| **DATE** | 26-09-2003 |
| **TEAM ID / TEAM NAME** | Proj\_224021\_Team­\_1 |
| **PROJECT NAME** | Prediction House Price using Machine Learning |
| **STUDENT NAME WITH ID** | Murugan M - au610521104056  Monika S - au610521104055  Yuvaraj M - au610521104102  Sathesh PC - au6105211040082  Rahul Ganesh B - au610521104306 |

**Prediction House Price using Machine learning**

**Problem Definition and Design Thinking**

**INTRODUCTION**

The project at hand involves the application of machine learning techniques to tackle the common and critical problem of predicting house prices. Accurate house price predictions have substantial implications for various stakeholders within the real estate industry, including prospective buyers, sellers, and investors. By developing a robust model capable of forecasting house prices accurately, this project aims to provide valuable insights into the real estate market.

**Problem Definition**

The problem at hand is to predict house prices using machine learning techniques. The primary objective is to develop a model that accurately forecasts house prices based on various features such as location, square footage, number of bedrooms and bathrooms, and other relevant factors. This project encompasses multiple critical stages, including data preprocessing, feature engineering, model selection, model training, and model evaluation.

**Key Components of the Problem:**

**1. Target Variable: House Prices**

The ultimate goal is to predict the prices of houses with precision.

**2. Features:**

Location: The geographical area or neighborhood where the house is located.

Square Footage: The total area of the house in square feet.

Number of Bedrooms: The count of bedrooms in the house.

Number of Bathrooms: The count of bathrooms in the house.

Other relevant factors: Additional features that can influence house prices, such as amenities, year built, condition, and proximity to facilities.

**Design Thinking**

In this section, we will outline the steps and approach we will follow to tackle the problem of predicting house prices:

**1. Data Source**

Data Collection: We will commence by sourcing a dataset containing comprehensive information about houses. This dataset should include features like location, square footage, number of bedrooms and bathrooms, and corresponding house prices. Ensuring that our data source is reliable and comprehensive is crucial for building an accurate model.

**2. Data Preprocessing**

Data Cleaning: Before analysis, we will ensure that the data is devoid of errors and inconsistencies. This involves handling missing values, outliers, and any data quality issues.

Feature Engineering: We will create new features or transform existing ones to capture valuable information. For instance, we may calculate the price per square foot or derive the age of the house from the year built.

Feature Scaling: Numerical features may be on different scales; therefore, we will normalize or standardize them to ensure they contribute equally to the model.

**3. Feature Selection**

Exploratory Data Analysis (EDA): We will explore the dataset to gain insights into the significance of each feature. EDA will help us understand how different features correlate with house prices.

Feature Importance: Based on EDA and domain knowledge, we will select the most relevant features for predicting house prices. This step is crucial for model efficiency and interpretability.

**4. Model Selection**

Algorithm Choice: We will choose an appropriate regression algorithm suited for predicting house prices. Options include Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor.

Hyperparameter Tuning: To optimize the model's performance, we will fine-tune its hyperparameters.

**5. Model Training**

Data Splitting: We will divide the dataset into training and testing sets to assess the model's performance.

Model Training: Using the selected model, we will train it on the preprocessed training data to capture patterns and relationships.

**6. Evaluation**

Metric Selection: We will select evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R²) to measure the model's accuracy.

Model Evaluation: We will assess the model's performance on the testing dataset using the chosen evaluation metrics.

**7. Documentation and Reporting**

Throughout the project, maintaining clear documentation is vital. We will document each step, decisions made, and results obtained. This documentation will enhance reproducibility and facilitate future reference.

**8. Conclusion**

By adhering to this structured design thinking approach, we aim to construct a reliable and accurate house price prediction model.

This approach ensures that we systematically address data quality, feature relevance, and model performance. Subsequent project phases will involve model refinement, deployment, and continuous monitoring and maintenance.

This document provides an overview of our understanding of the problem and the proposed approach for solving it.

It serves as a roadmap for our project, guiding us through the various stages of data preparation, modeling, and evaluation.