Phase-1

Student Name: Reena.R

Register Number: 410723104066

Institution: Dhanalakshmi college of engineering

Department: Computer science and engineerng

Date of Submission: 30.04.2025

1.Problem Statement

The predicting house prices is a Little challenge due to the complexity and House prices depend on a Location, size, number of rooms, near schools Or market and availablity of transport. It is important for Financial descsion making, market efficiency and real estate business strategy by regression model

2.Objectives of the Project

To predict the future or current market value of residential properties based on historical data

3.Scope of the Project

To build a predictive model that accurately forecasts house prices using smart regression techniques in data science and the house price is predict by the historical data and with the key feaures of the surrounding in present and future . Limitation are depend on inconsistent data and lack of real time market example some natural disaster may affct the price . Tools and technologies used are Programming language, libraries and notebooks

4.Data Sources

Ames housing dataset which contains various features of the residential properties in ames ,lowa and USA and by using regession model predicting and analysis the House price. Source: (Kaggle) and Street

map are used to know the features of the properties how enrich in environment, easy availability of transport and spitting distance of school, college, market and parks. The data source are public and dynamic

(source : https://www.kaggle.com/datasets/yasserh/housing-prices-dataset)

5.High-Level Methodology

- **Data Collection** – Gather the historical data on house sales: Prices, size (sq ft), location, no of bedrooms/bathroom,years of built ,amenities, etc

- **Data Cleaning** – Handle missing values, outliers, inconsistencies like price per sq ft, age of house, renovation index, distance to city center,etc

- **Exploratory Data Analysis(EDA)** – Analyze distribution, correlations, patterns.Visualize important relationships (e.g., price vs. size, price vs. location).Detect multicollinearity using heatmaps or Variance Inflation Factor (VIF).

- **Feature Engineering** – Use smart methods to select powerful predictors:Correlation analysis, Recursive Feature Elimination (RFE), LASSO regularization for automatic feature selection, Tree-based feature importance

- **Model Building** – List Baseline Models: Start with Linear Regression,Ridge, and Lasso Regression.
Advanced Models: Tree-based models: Random Forest, XGBoost, LightGBM, CatBoost.
Ensemble techniques: Stacking, Blending.
Neural Networks for tabular data (deep regression).

- **Model Evaluation** – Evaluate using appropriate metrics:
Root Mean Squared Error (RMSE) (preferred for prices),Mean Absolute Error

(MAE), R^2 score,Analyze residuals

- **Visualization & Interpretation** – Use SHAP values or LIMEto explain how features influence price prediction. matplotlib,seaborn are used to represent the data which are collected graphically like graph, bar chart, histogram, scatter plot

- **Deployment** – Deploy model in production if needed: API, batch scoring, dashboard integration.Monitor model drift and retrain periodically as the housing market changes.

b.Tools and Technologies

- **Programming Language** – Python is a main programmming language

- **Notebook/IDE** – The pltform we used to work in are Google Colab, Jupyter Notebook, VS Code

- **Libraries** – The key libraries are use for data processing, visualization, and modeling are pandas, numpy, seaborn, matplotlib, scikit-learn, TensorFlow

- **Optional Toolsfor Deployment** – Some toools or frameworks that might use for deployment Are Streamlit,Flask and fast API

- **Version Control** - Git/GitHub

7.Team Members and Roles

Name	Roles	Responsibility
Shamila.D	Team leader	Preprocessing, data collection and model evolution
Saranya.S	Team member	Data cleaning and EDA
Sharumathi.M	Team member	Model building and visualization
Reena.R	Team member	Developer and presentation