Phase-1 Submission Template

Student Name: Sanjay K

Register Number: 731123106042

Institution: Government college of engineering,Erode

Department: Electronics and Communication Engineering

Date of Submission: 02/05/2025

1.Problem Statement

Buying or selling a house can be stressful, especially when it comes to setting the right price. There are so many factors that affect house prices—location, number of rooms, square footage, and more and it's hard to consider them all manually. This project focuses on using data science and machine learning to build a smart system that can predict house prices accurately. With the help of regression models, we aim to bring more confidence and data-driven decision-making to the real estate world.

2.Objectives of the Project

Our main goal is to build a reliable model that can forecast house prices based on key features. Here's what we aim to do:

- Use smart regression techniques to improve prediction accuracy.
- Understand which features influence price the most.
- Experiment with different models
- Present the predictions and insights in an easy-to-understand way.
- Optionally, build a simple website or app where users can try the model.

3. Scope of the Project

- **Data Focus**: Use the House Prices dataset from Kaggle or any other websites.
- **Modeling Approach**: Apply and compare various regression models.
- **Data Handling**: Perform data cleaning, handle missing values, and engineer features to improve model accuracy.
- **Evaluation**: Measure model performance using metrics to determine prediction accuracy.
- **Tools and Output**: Use Python for development, with optional deployment of a simple Website or app for user interaction.

4. Data Sources

We're using the House Prices dataset from Kaggle or any other website. It's a well-known dataset with detailed information on home sales. This dataset is static and publicly available, which makes it ideal for training and testing machine learning models.

5. High-Level Methodology

- **Data Collection** – Obtain the dataset and load it into the working environment.
- **Data Cleaning** – Clean the dataset by handling missing values and correcting inconsistencies.
- **Exploratory Data Analysis (EDA)** – Analyze patterns, trends, and relationships between variables.
- **Feature Engineering** – Modify or create new features to improve predictive power.
- **Model Building** – Train multiple predictive models using the prepared data.
- **Model Evaluation** – Evaluate performance using appropriate accuracy metrics.

- **Visualization & Interpretation** – Present findings through charts, graphs, and summaries.

- **Deployment** – Create a simple interactive tool for users to explore predictions.

6.Tools and Technologies

Programming Language: Python

Development Environment: Google Colab,R And VS code

Libraries:

1. Pandas -Data manipulation and analysis.
2. Numpy-Numerical operations and array manipulation.
3. Scipy-Scientific and technical computing.
4. OpenCV-Computer vision and image processing.
5. Geopandas- Handling spatial/geographic data.

6. Matplotlib-Data visualization.

7.SeaBorn-Statistical data visualization built on top of Matplotlib.

8.Plotly- Interactive data visualization

9.Folium- Python library for creating interactive maps.

101.Scikit learn - machine learning library for classical ML algorithms

11.Tensor flow - deep learning and numerical computation framework

And etc….

Optional Tools for Deployment - Streamlit and etc…

7.Team Members and Roles

ROLE	TEAM MEMBERS
Data collection	Kayalvizhi V
Data preprocessing	Bafina A
Evaluation	Catherin Jersha J S
Model &Visualization	Sanjay K