# Forecasting House Prices Using Smart Regression Techniques in Data Science

Student name: S. Anusiya

Register number :620123106005

Institution: AVS Engineering college

Department :ECE

Date of submission :17/05/2025

Github repository link:

https://github.com/Anusiya1903/Anusiya-s-naan-mudhalvan-project-.git

## 1)Problem Statement

Accurately forecasting house prices is a significant challenge due to various influencing factors such as location, size, amenities, and market conditions. Traditional methods often fail to capture complex relationships within data. This project aims to develop a robust and intelligent regression-based model to predict house prices using advanced data science techniques.

## 2)Abstract

This project explores predictive modeling techniques to forecast house prices based on various features such as size, location, number of rooms, and other relevant variables. By utilizing smart regression algorithms like Linear Regression, Decision Trees, Random Forests, Gradient


Boosting, and ensemble methods, we aim to enhance prediction accuracy. The system is implemented using Python and deployed using a web-based interface for real-time predictions.


## 3)System Requirements

**Hardware:**

- 8GB RAM or higher

- 2GHz processor or higher

- 500GB HDD or SSD


**Software:**

- Python 3.x

- Jupyter Notebook/VS Code

- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, flask/streamlit

- Web browser


## 4)Objective

To develop a data-driven model capable of accurately predicting house prices using regression techniques, and deploy it through a user-friendly interface for practical usage.


## 5)Flowchart of Project Workflow

Start → Data Collection → Data Preprocessing → Exploratory Data Analysis → Feature Engineering → Model Building (Regression Techniques) → Model Evaluation → Model Deployment (Web App) → Future Enhancements → End

## 6)Dataset Description

- **Source:** Kaggle or real estate datasets

- **Features:**


 - `LotArea`, `YearBuilt`, `BedroomAbvGr`, `FullBath`, `GarageCars`, `TotalBsmtSF`, etc.

 - `SalePrice` (Target)


## 7)Data Preprocessing

- Handling missing values

- Encoding categorical variables (Label/One-Hot Encoding)

- Outlier detection and treatment

- Feature scaling (Standardization/Normalization)


## 8)Exploratory Data Analysis (EDA)

- Visualizing relationships (scatter plots, histograms, box plots)

- Correlation heatmap to understand relationships

- Trend analysis over time/location


## 9)Feature Engineering

- Creating new features (e.g., `HouseAge = YrSold - YearBuilt`)

- Removing irrelevant or low-importance features

- Feature selection using statistical tests or model-based importance


## 10)Model Building

- **Algorithms Used:**

 - Linear Regression

 - Ridge/Lasso Regression

- Decision Tree Regression

- Random Forest Regression


- Gradient Boosting/XGBoost

- Train-Test Split or Cross-Validation

- Hyperparameter tuning (GridSearchCV, RandomizedSearchCV)


## 11)Model Evaluation

- Metrics:

- Mean Absolute Error (MAE)

- Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)

- $R^2$ Score

- Comparison of model performance


## 12)Deployment

- Using Flask or Streamlit to create a web interface

- Input features through a form

- Display predicted house price

- Host on local server or cloud (Heroku, AWS, etc.)


## 13)Source Code

- Python scripts (.py/.ipynb)

- Web interface code (HTML/CSS for Flask or Streamlit script)

- GitHub repository (recommended)


## 14)Future Scope

- Integration with real-time real estate APIs

- Inclusion of satellite imagery or geospatial data

- Use of deep learning models for better predictions

- Mobile application integration

**15)Team members and roles:**

  **1) J.Ayisha banu: Data cleaning and documentation.**

  **2) S.Anusiya: EDA and problem objective.**

  **3) M.Dharani: Feature engineering and reporting.**

  **4) M.Kaviya: Model development and visualization of results.**