**Project Report**

**PROJECT-TITLE: HOUSE PRICE PREDICTION USING REGRESSION MODELS**

**Bachelor of Computer Application**

**NAME: SAHIL KR. SINGH (23CSA2BC195)**

**NAME: VIRAJ KHATRI (23CSA2BC241)**

**NAME: HARSHIT KUMAR (23CSA2BC185)**

**NAME: KUNAL KUMAR (23CSA2BC211)**

**MAY,2025**

****

**VIVEKANANDA GLOBAL UNIVERSITY** **ACKNOWLEDGEMENT**

*I have taken this opportunity to express my gratitude and humble regards to the Vivekananda Global University to provide an opportunity to present a project on the “House Price Prediction Using Regression Models” which is a “Machine Learning” based project.*

*I would also be thankful to my project guide* ***Anshul Gupta*** *to help me in the completion of my project and the documentation.I have taken efforts in this project but the success of this project would not be possible without their support and encouragement.*

*I would like to thanks our Dean sir “Dr.RC TRIPATHI” to help us in providing all the necessary books and other stuffs as and when required .I show my gratitude to the authors whose books has been proved as the guide in the completion of my project I am also thankful to my classmates and friends who have encouraged me in the course of completion of the project.*

*Thanks*

|  |  |
| --- | --- |
| PROJECT GUIDE:  **Project Guide: Anshul Gupta** | SUBMITTED BY:  **NAME: SAHIL KR SINGH**  **NAME: VIRAJ KHATRI** |

**NAME: HARSHIT KUMAR**

**NAME: KUNAL KUMAR**

**Place: Jaipur**

**Date: 14-04-25**

**DECLARATION**

We hereby declare that this Project Report titled “House Price Prediction Using Regression Models” submitted by us and approved by our project guide **ANSHUL GUPTA**, to the Vivekananda Global University, Jaipur is a Bonafide work undertaken by us and it is not submitted to any other University or Institution for the award of any degree diploma / certificate or published any time before.

|  |  |
| --- | --- |
|  |  |
|  |  |
| **Project: Guide:**  **Anshul Gupta** |  |
|  |  |

**Table of Contents**

[1 Project Title 4](#_gjdgxs)

[2 Problem Statement 4](#_30j0zll)

[3 Project Description 4](#_1fob9te)

[3.1 Scope of the Work 4](#_3znysh7)

[3.2 Project Modules 4](#_2et92p0)

[3.3 Context Diagram (High Level) 4](#_tyjcwt)

[4 Implementation Methodology 4](#_3dy6vkm)

[5 Technologies to be used 4](#_1y810tw)

[5.1 Software Platform 4](#_4i7ojhp)

[5.2 Hardware Platform 5](#_2xcytpi)

[6 Advantages of this Project 5](#_3whwml4)

7 Screenshots and Output

[8 Conclusion 6](#_49x2ik5)

[9 References 6](#_2p2csry)

# Project Title

House Price Prediction Using Regression Models

# 2. Problem Statement

The real estate market is one of the most dynamic and important sectors of the economy. Buyers and sellers often face challenges in estimating the right price for a property due to the variability of influencing factors such as location, size, number of bedrooms, and more. Traditional methods of valuation are often subjective and inconsistent. Hence, there is a need for a data-driven approach that can accurately predict house prices using historical data and machine learning techniques.

# 3. Project Description

This project leverages machine learning regression techniques to predict house prices based on various features. The primary aim is to develop models that learn patterns from historical housing data and provide accurate price predictions for new, unseen data.

## 3.1 Scope of the Work

* To study the relationship between house features and their market prices.
* To build and compare multiple regression models including **Linear Regression** and **Decision Tree Regression**.
* To evaluate the models using performance metrics like **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)**, and **R² Score**.
* To provide a basic prediction interface or script that takes user input and gives estimated house prices.
* Optional: Visualize data insights and model predictions for better understanding.

## 3.2 Project Modules

**Data Collection**

* Use a dataset such as the "House Prices – Advanced Regression Techniques" from Kaggle.
* Common features: Lot Area, Year Built, Number of Rooms, Garage, Location, etc.

**Data Preprocessing**

* Handle missing values.
* Encode categorical variables using Label Encoding or One-Hot Encoding.
* Normalize/standardize numerical features.
* Split data into training and testing sets.

**Model Training**

* Train Linear Regression model.
* Train Decision Tree Regressor.
* Use cross-validation to prevent overfitting.

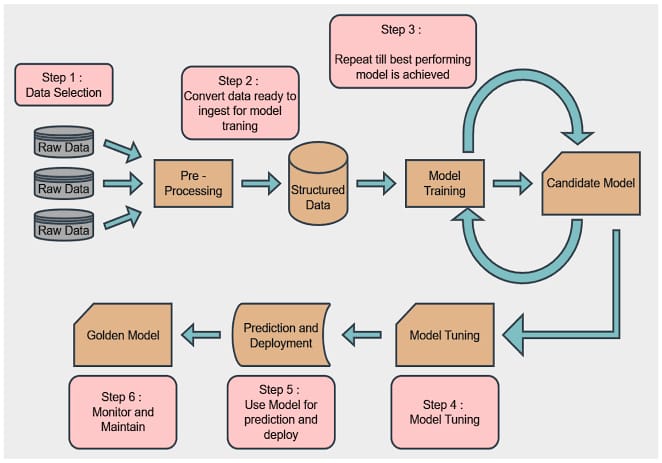
**Model Evaluation**

* Evaluate models using metrics:
  + Mean Absolute Error (MAE)
  + Mean Squared Error (MSE)
  + Root Mean Squared Error (RMSE)
  + R² Score
* Compare models to identify the best-performing one.

**Prediction Interface (Optional)**

* Simple CLI or GUI interface for users to input house features and get a price estimate.

## 3.3 Context Diagram (High Level)



**4. Implementation Methodology**

1. **Data Analysis**: Explore dataset using visualization libraries (e.g., Seaborn, Matplotlib, Pandas).
2. **Feature Engineering**: Select and transform relevant features.
3. **Model Building**:
   * Use scikit-learn to implement Linear Regression and Decision Tree models.
   * Train models on training data.
4. **Testing & Validation**: Evaluate model on test data and fine-tune parameters if needed.
5. **Deployment (Optional)**: Save the model using joblib or pickle and deploy it using a Flask/Django web interface or CLI app.

**5. Technologies to be Used**

**5.1 Software Platform**

* **Python 3.x**
* **Jupyter Notebook / VS Code**
* **Libraries:**
  + **pandas**
  + **numpy**
  + **matplotlib / seaborn (for visualization)**
  + **scikit-learn (for ML models)**
  + **joblib / pickle (for model saving)**

**5.2 Hardware Platform**

* **Any standard system with the following configuration:**
  + **Processor: Intel i3/i5 or equivalent**
  + **RAM: Minimum 4GB**
  + **Storage: Minimum 2GB free space**
  + **OS: Windows/Linux/Mac**

**6. Advantages Of this Project**

* Provides quick and accurate price estimates.
* Helps real estate agents and buyers make informed decisions.
* Reduces manual valuation errors.
* Easy to scale with more data for better accuracy.
* Can be enhanced with new features like location intelligence or satellite data.

**7.Conclusion**

This project demonstrates the use of machine learning techniques to solve a real-world problem—predicting house prices. With effective preprocessing and modeling, we can achieve reasonably accurate predictions. The comparison between Linear Regression and Decision Tree models shows that model selection depends on data characteristics and feature interactions.

**8.References**

Kaggle Dataset  
*House Prices - Advanced Regression Techniques*  
https://www.kaggle.com/c/house-prices-advanced-regression-techniques

Scikit-learn Documentation  
For implementing machine learning models like Linear Regression and Decision Tree Regressor  
https://scikit-learn.org/stable/

Python Data Science Handbook  
*Jake VanderPlas* – A comprehensive guide for data analysis and machine learning using Python  
https://jakevdp.github.io/PythonDataScienceHandbook/

Pandas Documentation  
For data preprocessing and manipulation  
https://pandas.pydata.org/

NumPy Documentation  
For numerical operations in Python  
<https://numpy.org/doc/>

Matplotlib and Seaborn  
For data visualization

* <https://matplotlib.org/>
* https://seaborn.pydata.org/

Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow  
*Aurélien Géron* – Excellent book for practical ML projects and model building

YouTube & MOOCs

* *Krish Naik*, *StatQuest with Josh Starmer*, and *freeCodeCamp* tutorials on regression models.
* [Coursera Machine Learning Course by Andrew Ng](https://www.coursera.org/learn/machine-learning)

**SOURCE CODE:-**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn.linear\_model import LinearRegression**

**from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score**

**import pandas as pd**

**# Load your dataset**

**data = pd.read\_csv("house\_prices.csv")**

**# Features and target**

**X = data[['Area', 'Bedrooms', 'Age']]**

**y = data['Price']**

**# Split the data**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2)**

**# Model training**

**model = LinearRegression()**

**model.fit(X\_train, y\_train)**

**# Prediction**

**y\_pred = model.predict(X\_test)**

**# Evaluation**

**print("MAE:", mean\_absolute\_error(y\_test, y\_pred))**

**print("MSE:", mean\_squared\_error(y\_test, y\_pred))**

**print("R² Score:", r2\_score(y\_test, y\_pred))**

**9. SCREENSHOT -**

