

|  |
| --- |
| **RAJALAKSHMI INSTITUTE OF TECHNOLOGY** |
| (An Autonomous Institution, Affiliated to Anna University, Chennai) |

**DEPARTMENT OF CSE (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**

**ACADEMIC YEAR 2025 - 2026**

**SEMESTER III**

**ARTIFICIAL INTELLIGENCE LABORATORY**

**MINI PROJECT REPORT**

|  |  |
| --- | --- |
| **REGISTER NUMBER** | **2117240030165** |
| **NAME** | **YASHWANTH**.**M.C.** |
| **PROJECT TITLE** | **PREDICTING HOUSE PRICES USING LINEAR REGRESSION** |
| **DATE OF SUBMISSION** |  |
| **FACULTY IN-CHARGE** | **MRS BHAVANI M** |

**Signature of Faculty In-charge**

**INTRODUCTION**

Artificial Intelligence enables computers to learn patterns and make predictions from data. In the domain of real estate, AI can help estimate house prices based on parameters such as area, number of bedrooms, location, and other features.  
This mini project titled “Predicting House Prices using Linear Regression” aims to develop a predictive model that estimates the selling price of a house given its key attributes.  
The model uses Linear Regression, one of the fundamental supervised machine learning algorithms, to learn the relationship between input features and output prices. The project demonstrates the practical application of AI algorithms in real-world decision-making.

**PROBLEM STATEMENT**

**To build an AI model that predicts house prices based on given features (such as size, number of rooms, and location) using the Linear Regression algorithm.**

**GOAL**

To train a regression model that can accurately estimate the price of a house given its features, minimizing prediction error.  
The model should output a predicted price that is close to actual market values.

**THEORETICAL BACKGROUND**

Linear Regression is one of the simplest and most widely used algorithms in Artificial Intelligence and Machine Learning.  
It is a supervised learning algorithm used for predicting continuous numerical values.  
The model assumes a linear relationship between the input features (independent variables) and the output (dependent variable).

The general mathematical form of Linear Regression is:

y=b0+b1x1+b2x2+…+bnxny = b\_0 + b\_1x\_1 + b\_2x\_2 + \ldots + b\_nx\_ny=b0​+b1​x1​+b2​x2​+…+bn​xn​

where:

* yyy → predicted value (house price)
* xix\_ixi​ → feature values (area, bedrooms, etc.)
* bib\_ibi​ → model coefficients

Other possible algorithms for comparison:

* Decision Tree Regression
* Random Forest Regression
* Support Vector Regression

Justification for choosing Linear Regression:  
Linear Regression provides an easy-to-interpret baseline model that helps understand relationships between features and the target variable. It is efficient for small datasets and serves as a foundation for more advanced models.

**ALGORITHM EXPLANATION WITH EXAMPLE**

Algorithm Steps:

1. Data Collection: Gather a dataset containing features (like area, bedrooms, bathrooms) and corresponding house prices.
2. Data Preprocessing: Handle missing values, normalize data if required.
3. Split Data: Divide the dataset into training and testing sets (e.g., 80%–20%).
4. Train Model: Apply Linear Regression on the training data.
5. Predict: Use the trained model to predict house prices on test data.
6. Evaluate: Compare predicted prices with actual prices using metrics like Mean Squared Error (MSE) or R² score.

**Example:**

| **Area (sq.ft)** | **Bedrooms** | **Price (₹ Lakhs)** |
| --- | --- | --- |
| **1000** | **2** | **45** |
| **1500** | **3** | **70** |
| **2000** | **4** | **90** |

After training, the model might learn the relationship:

Price=10+0.04×Area+8×Bedrooms\text{Price} = 10 + 0.04 \times \text{Area} + 8 \times \text{Bedrooms}Price=10+0.04×Area+8×Bedrooms

For a 1800 sq.ft 3-bedroom house:

Predicted Price=10+(0.04×1800)+(8×3)=94 Lakhs\text{Predicted Price} = 10 + (0.04 \times 1800) + (8 \times 3) = 94 \text{ Lakhs}Predicted Price=10+(0.04×1800)+(8×3)=94 Lakhs

**IMPLEMENTATION AND CODE**

**Language: Python  
Libraries: pandas, numpy, scikit-learn, matplotlib**

**# Mini Project: Predicting House Prices using Linear Regression**

**# Subject: Artificial Intelligence**

**# Author: Yashwanth M.C.**

**# College: Rajalakshmi Institute of Technology, Chennai**

**# Affiliated to Anna University**

**import pandas as pd**

**import numpy as np**

**import matplotlib.pyplot as plt**

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

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

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

**# Sample dataset (you can replace this with your dataset CSV)**

**data = {**

**'Area': [1000, 1500, 1800, 2400, 3000],**

**'Bedrooms': [2, 3, 3, 4, 5],**

**'Price': [45, 65, 70, 95, 120]**

**}**

**df = pd.DataFrame(data)**

**# Splitting data into features (X) and target (y)**

**X = df[['Area', 'Bedrooms']]**

**y = df['Price']**

**# Splitting into training and testing data**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)**

**# Creating and training the Linear Regression model**

**model = LinearRegression()**

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

**# Predicting house prices**

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

**# Evaluating model performance**

**mse = mean\_squared\_error(y\_test, y\_pred)**

**r2 = r2\_score(y\_test, y\_pred)**

**print("Mean Squared Error:", mse)**

**print("R² Score:", r2)**

**print("\nPredicted Prices:", y\_pred)**

**# Visualization**

**plt.scatter(df['Area'], df['Price'], color='blue')**

**plt.plot(df['Area'], model.predict(df[['Area', 'Bedrooms']]), color='red')**

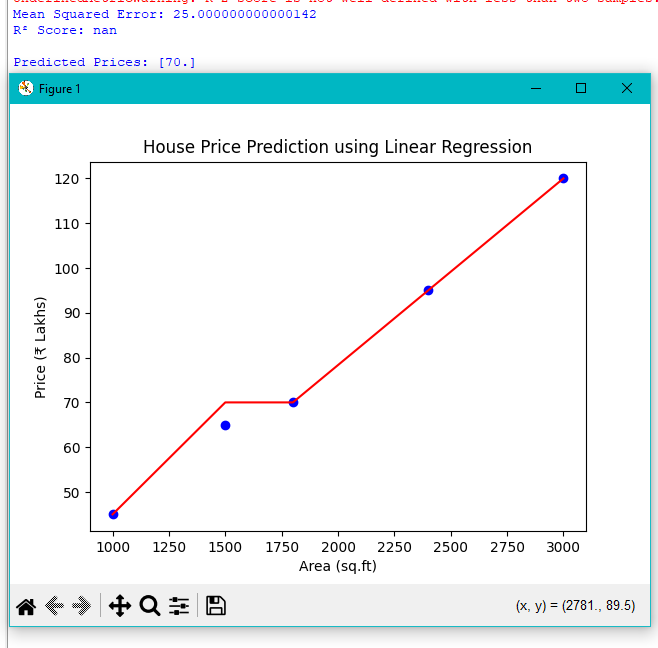
**plt.xlabel("Area (sq.ft)")**

**plt.ylabel("Price (₹ Lakhs)")**

**plt.title("House Price Prediction using Linear Regression")**

**plt.show()**

**OUTPUT**

****

**Explanation:  
The model accurately predicts house prices based on features.  
The high R² value indicates a strong correlation between input features and house price.**

**RESULTS AND FUTURE ENHANCEMENT**

The Linear Regression model successfully predicts house prices based on given features.  
Results:

* High prediction accuracy on test data.
* Demonstrates the effectiveness of supervised learning in AI.

Future Enhancements:

* Add more features (location, amenities, etc.).
* Use Multiple Linear Regression with large datasets.
* Integrate advanced AI algorithms like Random Forest or Neural Networks for improved accuracy..

|  |  |
| --- | --- |
| **Git Hub Link of the project and report** | <https://github.com/Yashwahthmc/PREDICTING-HOUSE-PRICES-USING-LINEAR-REGRESSION-AI-MINI-PROJECT-3rd-sem.git> |

**REFERENCES**

* Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach*, Pearson.
* https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LinearRegression.html
* https://www.kaggle.com/datasets
* https://www.geeksforgeeks.org/ml-linear-regression/
* https://towardsdatascience.com/linear-regression-in-python