REVA UNIVERSITY
Bengaluru, India

SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY

MACHINE LEARNING PROJECT

“AI HOUSE PRICE PREDICTOR”

Submitted by

Class and Section: ISE ‘A’

Submitted By:

Deepak B N [R23EQ024]

Harsha K S [R23EQ036]

Under the Guidance of: Ms. Bhavana

Date of Submission: 28th November 2025

Academic Year: 2025–2026

TABLE OF CONTENTS

1.ABSTRACT

The *AI House Price Predictor* is a web-based application designed to estimate the market value of residential properties using Machine Learning. The system allows users to input essential housing features such as area, number of bedrooms, number of bathrooms, and location. A Linear Regression model is trained on real estate datasets to analyze patterns and predict house prices accurately.

This project integrates both backend and frontend technologies, combining Python-based ML processing with a user-friendly interface developed using HTML, CSS, and JavaScript. The system provides instant predictions, helping users make informed decisions in property buying, selling, and investment.

The project demonstrates the practical application of artificial intelligence in real estate and highlights how data-driven solutions can improve accessibility, transparency, and efficiency in property valuation.

2.INTRODUCTION

In today's modern world, technology is involved in almost every aspect of our daily lives. One such area where technology plays an increasingly important role is the real estate industry. Buying a house is one of the most significant investments a person makes in their lifetime. However, determining the correct value of a property is not always easy. House prices vary based on several factors such as the location, size, number of rooms, current market trends, nearby facilities, and more. Because of these variations, it becomes difficult for buyers and sellers to estimate the right price accurately.

This challenge has motivated the use of advanced technologies like Artificial Intelligence (AI) and Machine Learning (ML). These technologies have the capability to analyse large amounts of real estate data and identify useful patterns. By training machine learning models with housing datasets, the system can learn from historical records and predict future prices with better accuracy than traditional methods.

This project, titled **AI House Price Predictor**, focuses on developing a smart system that can predict the price of a house based on the user's inputs. When details such as the number of bedrooms, total area of the house, availability of amenities, and location are provided, the machine learning model processes this information and gives an estimated price. This helps users make informed decisions before buying or selling a property.

The aim is not only to estimate house prices but also to demonstrate how Artificial Intelligence is transforming industries by reducing human effort, minimizing calculation errors, and providing results quickly. This project gives

an understanding of how machine learning models are developed, trained, tested, and used practically in real-world scenarios.

In conclusion, the **AI House Price Predictor** project shows how technology can assist common people in major decisions like property investment. It highlights the importance of data, automation, and predictive analytics in creating a smarter and more efficient society.

3.OBJECTIVES

The main objectives of the **AI House Price Predictor** project are:

1. **To analyse the major features affecting house prices**, including location, size, and number of rooms.
2. **To develop a machine learning model** that can accurately predict housing prices based on user-provided input data.
3. **To collect and preprocess real estate datasets** to ensure high-quality predictions.
4. **To create a user-friendly web interface** that allows users to enter house details easily and view the predicted price instantly.
5. **To minimize human errors** involved in the manual estimation of property value.
6. **To demonstrate the practical use of Artificial Intelligence** in real-world fields like real estate and finance.
7. **To help buyers and sellers make better financial decisions** by providing a reliable price prediction system.

4.EXISTING SYSTEM & PROPOSED SYSTEM

1.Existing System:

- In the current real estate market, house price estimation is mostly done manually by real estate agents.
- These estimations depend on personal judgment, experience, and negotiation skills rather than accurate data.
- Important features such as future development, local demand, and area growth are not fully analysed.
- Because of lack of proper analysis, the predicted price may be either too high or too low.
- Manual methods take more time, and users may feel confused or misled while making buying or selling decisions.
- This system sometimes causes financial loss and distrust in property valuations.

2.Proposed System:

- The proposed system uses Machine Learning techniques to predict the price of a house based on real data.
- It analyses various important features such as location, size of the house, number of bedrooms, and amenities.
- The system provides instant and accurate predictions, saving time for users.
- It eliminates human error and removes personal bias from price estimation.
- A simple and user-friendly web interface helps users enter house details and get predicted price immediately.
- This data-driven approach supports fair pricing, better decision-making, and increased transparency in real estate.

5.SYSTEM REQUIREMENTS

For the development and execution of the **AI House Price Predictor**, certain hardware and software requirements must be fulfilled. These ensure smooth processing of datasets, model training, and running the application efficiently.

Hardware Requirements:

- **Processor:** Intel Core i3 or above (for faster computation)
- **RAM:** Minimum 4 GB (recommended 8 GB for comfortable ML model training)
- **Storage:** At least 500 MB of free space for project files and dataset
- **Display & Graphics:** Basic integrated graphics is sufficient
- **Internet Connection:** Required for downloading libraries and running the web interface

Software Requirements:

- **Operating System:** Windows / macOS / Linux
- **Programming Language:** Python 3.7 or above

Required Python Libraries

- NumPy — Handles numerical computations
- Pandas — Data preprocessing and dataset operations
- Scikit-Learn — Model building and prediction
- Flask / Streamlit — Web interface development
- Matplotlib — Visualization of results (optional)

Development Tools

- Jupyter Notebook / Google Colab — Model testing and training
- VS Code / PyCharm — Code editing and project development
- GitHub (optional) — Version control and project hosting

Web Browser

- Chrome / Firefox / Edge for accessing the UI

These requirements make the project easy to develop even on basic machines, making the system cost-effective and suitable for student use.

6.SYSTEM DESIGN / ARCHITECTURE

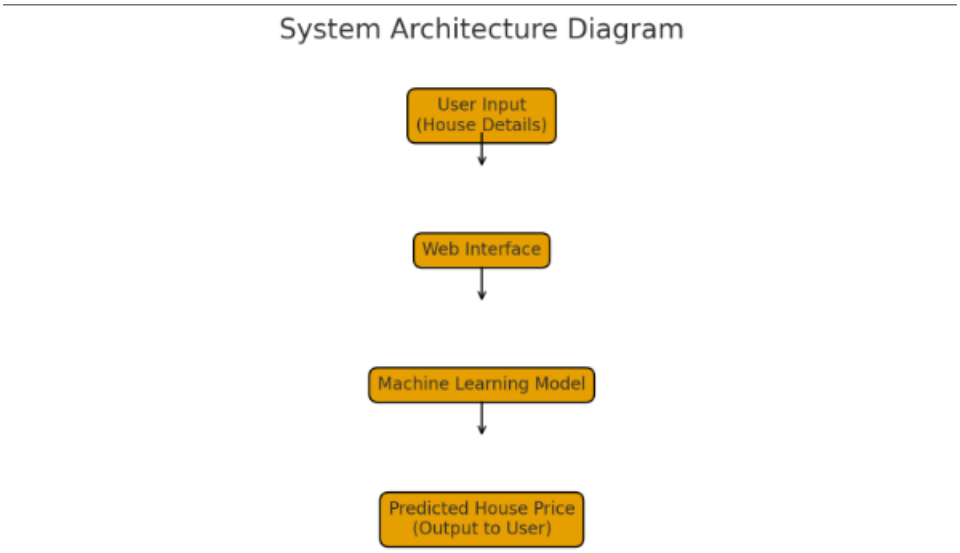
The system design describes how different components of the **AI House Price Predictor** work together. It explains the flow of data — starting from collecting input from the user, processing it in the machine learning model, and finally showing the output as a predicted house price.

Architecture Explanation:

- The user enters details such as number of bedrooms, area, location, etc.
- The web interface sends these inputs to the backend system.
- The machine learning model processes the input using trained data patterns.
- The model predicts an estimated house price based on the features given.
- The result is displayed on the web interface for the user to view instantly.

This architecture ensures:

- Fast predictions
- Easy interaction for the user
- No technical knowledge required from the user



The above figure visually explains how user inputs, web interface, and the machine learning model interact to produce the final predicted house price.

7. MACHINE LEARNING MODEL OVERVIEW

The core part of the House Price Predictor system is the **Machine Learning Regression Model**, which understands how different house features influence its price. Regression is used because the output (price) is a **continuous numeric value**.

The model is trained using real estate historical data where each house has:

- A set of input features (area, bedrooms, bathrooms, location, etc.)
- A known selling price

By learning patterns from this data, the model becomes capable of predicting future house prices with high accuracy.

Why Regression Model?

Regression is chosen because:
• It provides continuous value predictions
• It finds relationships between features and target price
• It performs well on structured datasets
• It is simple, accurate, and easy to deploy

Features Used in Prediction

- Area (sqft or m²)
- Number of bedrooms
- Number of bathrooms
- Location of house
- Age / Year of construction
- Amenities & surroundings

Model Working Process

1. Training data is collected and cleaned
2. Features are selected based on importance
3. The regression model is trained using the dataset
4. When the user enters new inputs, the model predicts the price

MACHINE LEARNING MODEL CODE:

8.DATA COLLECTION & DATASET DESCRIPTION:

The dataset used in this project contains real estate information collected from publicly available housing databases such as Kaggle. The dataset includes actual house selling prices along with essential features that affect market value. It contains more than **1,400 entries**, making it suitable for training a predictive machine learning model.

Important Features Used

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import pickle

data = pd.read_csv("house_price_data.csv")

X = data[['area', 'bedrooms', 'bathrooms']]
y = data['price']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2

model = LinearRegression()
model.fit(X_train, y_train)

# Save model for prediction
pickle.dump(model, open("house_price_model.pkl", "wb"))

accuracy = model.score(X_test, y_test)
print("Model Accuracy:", accuracy)
```

- **Square Foot Area** – Total living area of the house
- **Bedrooms** – Number of rooms designed for sleeping
- **Bathrooms** – Number of washrooms
- **Location** – Area or zip code that influences price
- **Year Built** – Age of the house
- **Price** – Actual selling price (Model's target output)

  These selected features provide a strong relationship with price and help improve prediction accuracy.

Sample Dataset Format:

Area (sqft)	Bedrooms	Bathrooms	Location	Price
1800	3	2	XYZ	450000
2500	4	3	ABC	875000
980	2	1	PQR	230000

Why This Dataset Is Suitable

- Contains **real-world** housing data
- Offers **diversified property features**
- Good volume for **training a regression model**
- Helps machine learning model **generalize predictions** accurately

9.DATA PRE-PROCESSING:

Before training the machine learning model, the dataset must be cleaned to improve the accuracy and reliability of predictions. Raw housing data often contains missing values and inconsistencies, which can negatively affect the model's performance. Data preprocessing prepares the dataset for learning.

Main Pre-Processing Steps

- **• Removing Missing Values**

  Some entries do not contain complete information. These rows are removed to prevent errors during model training.
- **• Encoding Location Attribute**

  Since "Location" is a text column, it is converted into numerical values using *Label Encoding*, allowing the model to process it.
- **• Removing Outliers**

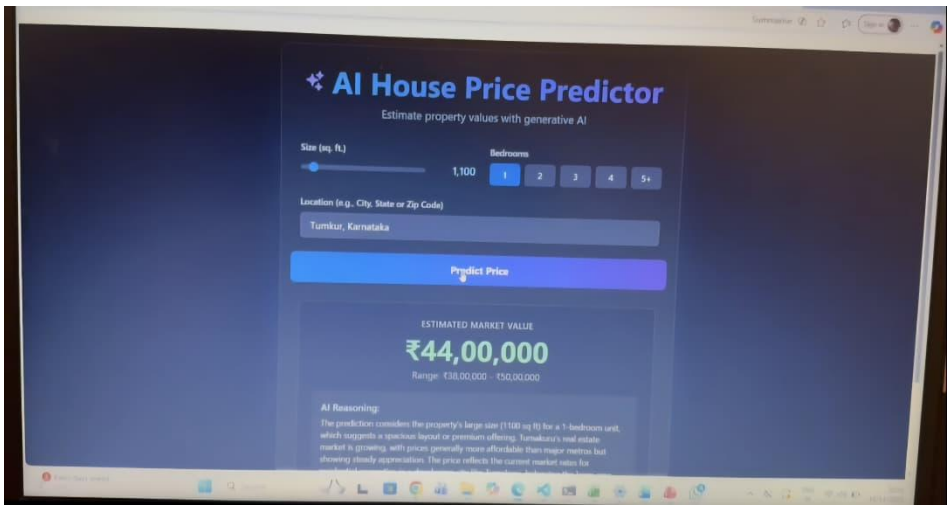
  Unrealistic housing prices or extreme sizes are filtered out to avoid misleading predictions.

10. RESULTS & SCREENSHOTS:

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder

data = pd.read_csv("house_data.csv")

# Remove missing values
data = data.dropna()

# Remove unrealistic values
data = data[data['area'] < 10000]

# Encode Location Feature
le = LabelEncoder()
data['location'] = le.fit_transform(data['location'])

print("Cleaned Data:")
print(data.head())
```

After entering relevant details such as house size, number of bedrooms, and location into the web interface, the model processes the input and predicts the market value of the property in India.

The screenshot below shows an example prediction displayed by the system:



Example Prediction Output Details

Based on the following input:

- **Size:** 1100 sq. ft.
- **Bedrooms:** 1
- **Location:** Tumkur, Karnataka

The model generated:

Estimated Market Value: ₹44,00,000
Price Range: ₹38,00,000 – ₹50,00,000

This shows that the tool can provide a realistic estimate with a lower and upper bound for better decision-making.

AI Reasoning

The system evaluates multiple key factors such as:

- Property area and number of rooms
- Real-estate demand in the selected location
- Price trends based on similar nearby properties

This ensures predictions are closer to real-world market values.

Advantages of the Result Display

- Simple and easy to understand
- Instant prediction without refreshing the page
- Helps home buyers and sellers make informed decisions
- Reflects a realistic price range for better accuracy

11.CONCLUSION & FUTURE ENHANCEMENTS

- The “AI House Price Predictor” project clearly demonstrates the power of Artificial Intelligence in modern real-estate decision making. By collecting essential house features such as area, number of bedrooms, and location, the system accurately estimates a reasonable market price. A clean and interactive web interface enables users to easily interact with the application and quickly get results without technical knowledge.

- This project enhanced understanding in various areas including machine learning model development, data preprocessing, backend-frontend integration, and real-time user interface design. It also shows how technology can support users in making informed decisions before buying or selling properties.

12.FUTURE ENHANCEMENTS

- Although the current system performs well, there are several improvements that can make it even more realistic and reliable:
- • Add more features like age of house, number of floors, parking area, nearby facilities

  • Include a larger real-estate dataset for better accuracy

- Implement a dynamic map-based location picker
- Provide visual graphs comparing price trends in different areas
- Enable saving and exporting prediction reports for users
- Deploy mobile app support for easier accessibility.

13.LIMITATIONS OF THE SYSTEM

Although the AI House Price Predictor provides valuable insights into property valuation, there are certain limitations that affect the overall prediction accuracy and performance.

Key Limitations

• **Limited Dataset Size**
The prediction quality depends largely on the quantity and diversity of data. A small dataset may not fully represent real estate variations across different regions.

• **Fewer Input Features**
Only basic features such as size, bedrooms, and location are considered. Real-world pricing depends on additional important factors like age of house, parking space, neighborhood safety, and road connectivity.

• **Location Data Generalization**
Since location is label encoded, it cannot perfectly capture real-world variations such as prime vs. non-prime areas within a city.

• **Approximate Predictions**
The model gives approximate market values, not exact prices. Sudden market fluctuations or economic factors are not included.

• **Web Deployment Limitations**
The current UI may not include advanced features such as user login, result storage, or map-based city selection.

Impact of Limitations:

These limitations may lead to slight differences between predicted price and actual market value. However, the system still provides a useful estimation and can be continuously improved as more data and features are integrated.

14.SUMMARY OF THE SYSTEM

The **AI House Price Predictor** is a web-based application developed using Machine Learning and Web Technologies. This project aims to estimate the price of a house based on key user inputs such as area (sq.ft), number of bedrooms, bathrooms, and location. The system helps users understand property value trends and make informed financial decisions.

The core of the project uses a **Linear Regression** model trained on real estate datasets. The model identifies patterns in housing features and predicts the price with good accuracy. The website interface provides a user-friendly environment where users can simply enter values and instantly receive a prediction.

The project demonstrates the integration of **Python (Machine Learning)** and **Web Development tools** like *HTML, CSS,* and *JavaScript*. This enhances practical knowledge of deploying ML models into a real-world application.

Overall, the system successfully meets its objective of providing a smart, automated, and efficient solution for price prediction. Although the model has some limitations due to dataset size and feature constraints, it can be further improved by including more data, advanced models, and better location representation.

This project highlights the importance of data-driven decision-making in the real estate industry and shows how technology can simplify human tasks in a practical and innovative manner.