

House Price Prediction Using Machine Learning

Authors

Parth Tandon, Abhay Pratap Singh Under the guidance of : Dr. Vijaya Choudhary

Department of Computer Science and Engineering

IILM University, Greater Noida, India

Abstract

The real estate market involves complex decision-making processes, including accurate house price estimation, financial planning, and loan approval analysis. Traditional methods of evaluating property prices and financial feasibility are often time-consuming and prone to errors. With the advancement of machine learning and data analysis, it is possible to develop intelligent systems that provide accurate predictions and financial insights.

This paper presents a House Price Prediction using Machine Learning which supports systems that integrates multiple machine learning models (Classification ,Regression etc) and financial calculations into a single web application. The system uses regression techniques to predict house prices based on features such as area, number of rooms, city, and state. In addition, a classification model is implemented to predict loan eligibility based on financial parameters such as income, credit score, and employment status.

The system also includes an EMI calculator and a rent-versus-buy analysis module, which helps users compare the long-term cost of renting and buying a property. These features enable users to make informed and data-driven decisions. The application is developed using Python, Flask, and Scikit-learn, providing a user-friendly interface for interaction.

The proposed system demonstrates the practical application of machine learning in real estate and financial planning, offering a comprehensive solution for property price prediction and decision support. Buying a House is the most important decision in a person's life therefore our model will help them with this. Additionally, our model also helps to find the best EMI and if Renting could be a better decision.

1. Introduction

The real estate sector plays an important role in the economy, and buying a house is one of the most important financial decisions for individuals. Determining the correct price of a property and understanding the financial impact of buying a house requires careful analysis of various factors such as location, area, number of rooms, market trends, and personal financial condition. Manual estimation of house prices and financial planning is difficult and often leads to incorrect decisions.

With the advancement of machine learning and data science, it has become possible to analyze large datasets and make accurate predictions based on historical data. Machine learning algorithms can be used to estimate house prices, predict loan approval, and analyze financial outcomes. These technologies can help users make better and more informed decisions in the real estate market.

In addition to price prediction, buying a house also involves financial considerations such as loan eligibility, EMI calculation, and long-term cost comparison between renting and buying. Most existing real estate platforms only provide property listings and do not offer intelligent decision-making tools. Therefore, there is a need for a system that can combine price prediction, financial analysis, and loan evaluation in a single application.

This project aims to develop an AI-based web application that predicts house prices using machine learning, checks loan eligibility using classification algorithms, calculates EMI, and performs rent-versus-buy analysis. The system integrates multiple models and financial calculations to provide a complete decision support platform for users who want to buy or rent a house.

The proposed system demonstrates the practical use of machine learning in real-world applications and provides a user-friendly interface for accurate and data-driven real estate decision making.

2. Literature Review

The application of machine learning in real estate and financial decision-making has gained significant attention in recent years. Various researchers have proposed models for house price prediction and loan approval systems using different machine learning techniques.

Several studies have focused on **house price prediction using regression algorithms**. For instance, Harrison and Rubinfeld [1] analyzed housing data using regression techniques to identify key factors affecting property prices. Similarly, Kok et al. [2] explored the use of machine learning models in real estate valuation and found that advanced algorithms can improve prediction accuracy compared to traditional methods. Research has shown that ensemble models such as Random Forest and Gradient Boosting outperform basic regression models due to their ability to capture complex relationships in data.

In the domain of **loan approval prediction**, classification algorithms such as Logistic Regression, Support Vector Machines, and Random Forest have been widely used. Breiman [3] introduced the Random Forest algorithm, which is known for its robustness and high accuracy in classification problems. Studies indicate that Random Forest performs better in financial prediction tasks due to its ability to handle non-linear data and reduce overfitting.

Further research has explored **financial decision support systems** that combine multiple analytical components. For example, Chen et al. [4] proposed systems integrating predictive models with financial calculations to assist users in making better investment decisions. These systems demonstrate the importance of combining machine learning with financial modeling.

However, most existing systems focus on a **single task**, such as house price prediction or loan approval. Very few studies integrate multiple components like price prediction, loan eligibility, EMI calculation, and rent-versus-buy analysis into a unified system. The proposed system addresses this gap by combining regression, classification, and financial modeling into a single web-based application, providing a comprehensive solution for real estate decision-making.

3. Methodology

The proposed system is developed using machine learning techniques, financial calculations and web applications. The methodology used in this model follows the following steps :

1.Data Collection:

Data is the most important first step. We have collected a Kaggle Dataset and scraped sites like Housing.com / MagicBricks.com for latest data.

2.Data Processing:

After collecting the data two major steps are done :

1. Removing duplicate and null values
2. Using encoding techniques

3.House Price Prediction Model:

The dataset is divided into two parts : Testing and Training set which uses Linear Regression model.

4.Loan Eligibility Prediction Model:

In this step we used two models: Logistic Regression and Random Forest. The model is better results were selected.

5.EMI Calculator:

Some mathematical and financial formulas were used in this step.

Tools and Techniques :

1. Python
2. Pandas and NumPy
3. Scikit – Learn
4. Flask
5. Html / CSS
6. Pickle

3.8 System Architecture

The proposed system is designed as a web-based application that integrates machine learning models with financial calculation modules to provide a complete real estate decision support system. The architecture follows a modular approach, where each component performs a specific task such as price prediction, loan eligibility analysis, EMI calculation, and rent-versus- buying comparison.

1. Overall Architecture

The system consists of three main layers:

- User Interface Layer:
Developed using HTML and CSS, this layer allows users to input property details and financial information such as area, location, income, credit score, and loan parameters.
- Application Layer (Flask Backend):
This layer processes user inputs, communicates with machine learning models, performs financial calculations, and generates outputs.
- Model Layer:
This layer contains trained machine learning models:
 - Regression model for house price prediction
 - Classification model for loan eligibility prediction

2. Input Module

The system accepts user inputs in four categories:

- Property Details: Area, BHK, city, and state
- Loan Details: Down payment (%), interest rate, and tenure
- Financial Details: Income, credit score, employment status, and dependents

- Rent Details: Monthly rent, rent growth rate, and comparison duration

3. Machine learning models generate predictions

4. Financial calculations are performed

5. Results are displayed on the web interface

3. Processing Module

The backend processes the input data through the following steps:

1. **Data Encoding:**
Categorical variables such as city and state are converted into numerical format using encoding techniques.
2. **House Price Prediction:**
The regression model predicts the house price based on user inputs.
3. **EMI Calculation:**
The system calculates the monthly EMI, total interest, and loan amount using financial formulas.
4. **Loan Eligibility Prediction:**
The classification model evaluates the probability of loan approval based on financial details.
5. **Rent vs Buy Analysis:**
The system compares the total cost of renting and buying over a specified period and provides a recommendation.

4. Output Module

The system generates the following outputs:

- Predicted house price
- Loan amount and EMI
- Total interest payable
- Loan approval probability and status
- Total rent cost
- Net buying cost and future property value
- Final recommendation (Rent or Buy)

5. System Workflow

The workflow of the system is as follows:

1. User enters property and financial details
2. Data is processed and encoded

6. Integration

All modules are integrated into a single Flask-based web application, ensuring smooth interaction between the user interface, machine learning models, and financial calculation components.

4. Results and Discussion

The proposed AI-based real estate decision support system was successfully implemented and tested using various input combinations. The system integrates house price prediction, loan eligibility prediction, EMI calculation, and rent-versus-buy analysis into a single web-based application.

1. House Price Prediction Results

The regression model was able to predict house prices based on input features such as area, BHK, city, and state. The predicted prices were generated in Lakhs and later converted into Crores for better readability. The results were found to be reasonable and consistent with expected real estate trends.

2. Loan Eligibility Prediction Results

The classification model (Random Forest) was used to predict loan approval. The system provided both:

- Loan status (Approved/Rejected)
- Approval probability (%)

The model produced reliable predictions based on user financial inputs such as income, credit score, employment status, and dependents. For example, a lower credit score or insufficient income resulted in lower approval probability.

3. EMI and Financial Calculation Results

The EMI module calculated monthly installment and total interest accurately using standard financial formulas. The correction of unit conversion (Lakhs to Rupees) ensured that EMI values were realistic and aligned with actual banking calculations.

4. Rent vs Buy Analysis

The system compared the long-term cost of renting and buying based on:

1. Use of Real-Time Data

The current system uses static datasets for training. In future, real-time data from real estate platforms can be integrated using APIs to provide more accurate and up-to-date predictions.

2. Improved Machine Learning Models

More advanced algorithms such as Gradient Boosting, XGBoost, or Deep Learning models can be used to improve the accuracy of house price and loan prediction.

3. Inclusion of More Features

Additional features such as location ratings, nearby facilities, crime rate, transportation access, and market trends can be included to enhance prediction quality.

4. Graphical Visualization

Future versions of the system can include graphs and charts, such as EMI breakdown, rent vs buy comparison graphs, and price trend analysis, to improve user understanding.

5. Break-Even Analysis

A break-even point feature can be added to determine the exact time when buying becomes more beneficial than renting.

6. Mobile Application Development

The system can be extended into a mobile application for better accessibility and user convenience.

7. User Authentication and Personalization

Adding user login functionality can allow saving user preferences, previous searches, and personalized recommendations.

8. Integration with Banking Systems

The system can be connected with banking APIs to provide real-time loan approval and interest rates.

[5] J. Han, M. Kamber, J. Pei, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2011.

[6] T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning," Springer, 2009.

[7] Kaggle, "Indian Housing," Available: <https://www.kaggle.com>

Project Guide Signature :

References

- [1] Abigail Bola Adetunji, Oluwatobi Noah Akande, Funmilola, Alaba Ajala, Ololade Oyewo, Yetunde Faith Akande, Gbenle Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique, 2022.
- [2] Dr. M. Thamarai and Dr. S P. Malarvizhi, "House Price Prediction Modeling Using Machine Learning," Journal of Machine Learning Research, 2020.
- [3] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei, "Housing Price Prediction via Improved Machine Learning Techniques," 2020.
- [4] I. H. Witten, E. Frank, M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques," 2016.