House rate prediction with the help of Supervised Regression model

A Minor Project Synopsisis Submitted in partial fulfillment for the award of degree of Bachelor of Technology in Information Technology

Submitted to

Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P.)

Submitted By

Bhavesh Kulkarni (0827CD201017)
Harsh Garg (0827CD201023)
Ashwani Shrivstav (0827CD201013)
Mahendra Chourasiya (0827CD201031)

Under the guidance of
Prof. Pawan Makhija

Department of Information Technology
Acropolis Institute of Technology & Research, Indore

Jan-June 2023

Acropolis Institute of Technology & Research, Indore
Department of Information Technology

CERTIFICATE

This is to certify that the work embodies in this synopsis entitled "House rate prediction with the help of Supervised Regression model" being submitted by Bhavesh Kulkarni (0827CD201017), Harsh Garg (0827CD201023), Ashwani Shrivstav (0827CD201013) and Mahendra Chourasiya (0827CD201031) for partial fulfillment of the requirement for the award of Bachelors of Technology in Information Technology discipline to Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P.) during the academic year 2023-24 is a record of bonafide piece of work carried out under the supervision of Prof. Pawan Makhija in the Department of Information Technology of Acropolis Institute of Technology & Research, Indore

Supervisor
Prof. Pawan Makhija

Head of the Department
Prof. Prashant Lakkadwala
Dept. of IT

Contents

1. Abstract

House Price Index (HPI) is commonly used to estimate the changes in housing price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price. As a result, to explore various impacts of features on prediction methods, our project will provide both traditional and advanced machine learning approaches to investigate the difference among several advanced models. This project will also comprehensively validate multiple techniques in model implementation on regression and provide an optimistic result for housing price prediction.
Technology used: Machine learning algorithms linear regression, random forests are used to train the model on a historical dataset of house prices and predict the future price.

2. Introduction of the Project (1 paragraph)

To build this model, a historical dataset of house prices and relevant features will be collected and pre-processed. The pre-processed data will then be used to train and test different machine learning algorithms such as linear regression and random forest. The best performing algorithm will be selected based on performance metrics such as accuracy and mean squared error.
The final model will be evaluated on a test set of data and its performance will be compared to other existing models in the literature. In addition, the model will be validated on real-world data to demonstrate its practical applications.

3. Objectives of the Project (100 words)

The objective of this house rate prediction project in machine learning is to:

Develop a predictive model to estimate the future value of a house based on various factors such as location, size, age, and surrounding amenities.

Explore and analyze the relationship between the different features and the target variable (house price) using data visualization techniques.

Train and evaluate different machine learning algorithms to determine the best performing model for the task of house price prediction.

The ultimate goal of this project is to provide a reliable and accurate tool for estimating house prices that can be used by a wide range of stakeholders, including real estate agents, investors, homeowners, and government agencies.

4. Literature/ Existing System Survey (200 words)

There are several existing systems for house price prediction, including online platforms such as Zillow and Redfin, as well as real estate appraisal software and services. These systems typically use a combination of machine learning algorithms and feature engineering to predict house prices and provide insights into the real estate market.
Zillow: Zillow is an online real estate platform that provides information and tools related to buying, selling, and renting homes. One of its key features is Zestimate, which is a house price prediction tool that uses machine learning algorithms to estimate the value of a property.
Redfin: Redfin is another online real estate platform that provides information and tools related to buying, selling, and renting homes.

"A comparison of regression techniques for house price prediction" by Sun et al. (2019) evaluates the performance of different regression techniques, including linear regression, ridge regression, lasso regression, decision trees, and random forests, on a dataset of house price prediction. The authors also compare the results of these models with traditional statistical methods.

"An empirical comparison of machine learning algorithms for house price prediction" by Kim et al. (2020) compares the performance of various machine learning algorithms, including linear regression, decision trees, random forests, support vector machines, and neural networks, on a dataset of house price prediction. The paper also investigates the impact of feature selection on the prediction accuracy of these models.

5. Hardware and Software Requirements

The hardware and software requirements for house rate prediction in machine learning are as follows:

Hardware:
A personal computer or laptop with a minimum of 8 GB of RAM.
A modern GPU is not necessary, but can be useful for accelerating the training of deep learning models if any deep learning is used in future.

Software:
An operating system, such as Windows, MacOS, or Linux
A programming language such as Python with packages for data analysis, visualization, and machine learning
A software environment for data analysis and machine learning, such as Jupyter Notebook.
Packages for machine learning, such as scikit-learn.
Packages for data visualization, such as Matplotlib and Seaborn.
Packages for data manipulation and preprocessing, such as pandas and numpy.

6. Problem Statement

Given a set of features such as location, size, number of bedrooms and bathrooms, age of the property, and other relevant factors, predict the value or price of a house in a particular real estate market. The goal of this project is to build a machine learning model that can accurately estimate the value of a property based on these features. This model can then be used to provide valuable insights into the real estate market, support decision-making for buyers, sellers, and real estate professionals, and help real estate appraisers in their assessments.

7. Proposed Solution (200 words)

There are certain steps for solution:

1. Data Collection: Collect a dataset of house prices and their corresponding features such as location, size, age, number of rooms, and others. This data can be obtained from real estate websites, government databases, or through surveys.
2. Data Preprocessing: Clean and preprocess the data to remove any irrelevant or missing values, and to ensure that it is in a format suitable for analysis.
3. Feature Engineering: Create new features from the existing data that may impact the house rate. For example, the proximity to public transportation, the availability of amenities nearby, or the crime rate in the area.

4. Exploratory Data Analysis: Analyze the data to understand the relationship between the features and the house rate. Use visualization techniques to identify trends and patterns.
5. Model Selection: Select a suitable machine learning algorithm to predict house prices. Common algorithms used for regression problems include linear regression, decision trees and random forest.
6. Model Training: Train the selected algorithm on the preprocessed and feature-engineered data.
7. Model Evaluation: Evaluate the performance of the trained model using appropriate metrics root mean squared error.
8. Model Deployment: In this step we deploy the model for testing. After that we test the model by the test set.

8. Expected Project Outcome (100-150 words)

Predictive Model: A machine learning model that can predict the house prices based on various features such as location, size, age, number of rooms, and others.

Model Performance: The model should have an acceptable level of accuracy, as measured by performance metrics such as mean squared error and root mean squared error.

Insights and Recommendations: The model should provide insights into the factors that impact house prices, such as the location, size, age, number of rooms, and others. These insights can be used to make recommendations for future house purchases or investments.

Deployment: The model should be deployed as a web service or integrated into an existing application to allow users to predict house prices based on their inputs.

9. Limitations and Future Scope of the Project (150 words)

Limitations:

1. Data Quality: The quality and relevance of the data collected can greatly impact the accuracy of the model. The model may be limited by the availability and quality of data.
2. Algorithm Selection: The choice of machine learning algorithm can greatly impact the performance of the model. The model may be limited by the limitations of the selected algorithm.
3. Model Performance: The model's performance may be limited by the size and complexity of the data, as well as the number of features used.

Future Scopes:

1. Improved Data Collection: In the future, collecting more comprehensive and relevant data could lead to a more accurate model.
2. Advanced Algorithms: Using advanced machine learning algorithms, such as deep learning, may lead to improved performance for the model.
3. Model Ensemble: Combining multiple models using ensemble methods may lead to a more robust and accurate model.

10. Conclusion (100-150 words)

In conclusion, house rate prediction using machine learning can provide valuable insights into the factors that impact house prices and can lead to the development of a predictive model that can be used to estimate the price of a house based on various features.

The results of the project will depend on the quality and relevance of the data collected, the choice of machine learning algorithm, and the performance of the model.

However, regardless of the specific results, this type of project can provide valuable experience in data analysis, machine learning, and real-world problem solving. It can also serve as a foundation for future work in the field of real estate analysis and prediction.

In the future, improved data collection, advanced algorithms, and integration with other data sources can lead to more accurate models and better insights into the real estate market.

11. References

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer Science & Business Media.
2. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112). New York: Springer.
3. Geron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.
4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.