Real Estate Price Prediction

Vishal Pukale, Ujwal Mahakalkar, Atharva Bahadure

Prof. Shobha Bamane

Vishal Pukale, Ujwal Mahakalkar, Atharva Bahadure Students of Department of Computer Science, ISBM College of Engineering, Pune, Maharashtra, India

Professor, Department of Computer Science, ISBM College of Engineering, Pune, Maharashtra, India

ABSTRACT

This study focuses on the development of a house price prediction model using machine learning techniques. We analyse various features, including property size, location, and amenities, to forecast accurate property values. The model's performance is assessed through rigorous evaluation and validation techniques. The findings contribute to better-informed real estate decisions and market insights. Employing a comparative analysis of various algorithms, including Linear Regression, Random Forest, and Gradient Boosting, we aim to identify the most effective approach. Our methodology emphasizes feature engineering, cross-validation, and hyperparameter tuning to optimize model performance.

Keywords: ML, Code repositories, Web Applications

INTRODUCTION

Forecasting property values can present a complex undertaking due to the myriad of considerations that go into purchasing a home. These factors encompass aspects like the property's architectural design, the number of rooms, the kitchen, available parking, and outdoor spaces. Frequently, individuals may lack awareness of the influential factors impacting property prices. Nevertheless, with the aid of machine learning, we have the capability to identify an ideal home and provide precise price forecasts. The prediction of real estate prices is a critical aspect of the housing market, influencing both buyers and sellers. This technological innovation exhibits the potential to evolve into a highly popular social application within the Indian market.

MOTIVATION

Imagine relocating to a new city, where you have little to no knowledge about the area or the city itself. The real estate broker you have hired may not always have your best interests at heart and could provide you with property prices that are far from accurate. Even if you are planning to build a new house, an ill-informed builder may end up exceeding your budget. However, by utilizing a reliable house price prediction model, constructing a budget estimate becomes a more manageable task.

1. Investment Confidence: Boost your confidence when making real estate investments.
2. Savings and Budgeting: Facilitate better savings and budget planning for your   housing needs.
3. Selling Strategy: Develop effective strategies when it's time to sell your property.
4. Risk Reduction: Minimize the risk of overpaying or encountering financial setbacks.

TOOLS & LIBRARIES

Tools
a. Python
b. Jupiter Notebook
c. Flask
d. HTML
e. CSS
f. JS
g. Heroku
h. GitHub

Libraries
a. Pandas
b. biscuit Learn
c. canopy
d. Seaborn
e. Matplotlib

ALGORITHMS AND ITS WORKFLOW

Linear Regression: A simple and interpretable algorithm that models the linear relationship between features and house prices. Provides coefficients that can be interpreted to understand the impact of each feature on the house price.
Calculates the best-fitting line that minimizes the difference between predicted and actual house prices also coefficients represent the weights of each feature, indicating their impact on the predicted price.

Decision Trees: Effective for capturing complex relationships in real estate data, offering interpretability and visualization. intuitive and easy to visualize, aiding in understanding the decision-making process. Makes predictions by traversing the tree from the root to a leaf node, where each leaf corresponds to a predicted house price.

Random Forest: An ensemble of decision trees, providing improved accuracy and robustness by mitigating overfitting. Robust to outliers and noise in the data, making it suitable for real-world datasets. Constructs multiple decision trees using different subsets of the data and features.
Averages predictions from individual trees to reduce overfitting and enhance generalization for robust price predictions.

Gradient Boosting: Sequentially builds a strong predictive model, suitable for refining predictions and handling hard to bend patterns. Often achieves high accuracy and is less resulting to less overfit compared to individual decision trees. Builds an ensemble of weak models sequentially, with each new model correcting errors made by the previous ones.

The final prediction is a combination of predictions from all models, resulting in a strong and accurate model.

Support Vector Machines (SVM): Effective for high-dimensional datasets, offering flexibility in capturing intricate relationships. Can handle non-linear patterns. Finds the hyperplane that maximally separates houses of different prices in a high-dimensional space.

Also utilizes kernels to map data into a higher-dimensional space for capturing complex relationships.

Neural Networks: Deep learning models that can uncover non-linear patterns and adapt to diverse real estate market dynamics. Require large amounts of data for training and are capable of modelling complex, non-linear relationships in real estate markets. Consists of layers of interconnected nodes (neurons) that learn hierarchical representations of features.

Adjusts weights during training to minimize the difference between predicted and actual house prices through forward and backward propagation.

CHARACTERISTICS

Data Collection: To start, we gather a lot of information about properties. This includes details like the number of bedrooms, bathrooms, square footage, location, and more. We also collect historical sales data.

Features: These are like the ingredients in a recipe. We select specific details from the data that we think will affect the price of a property, such as the neighbourhood's safety, schools, nearby amenities, and property age.

Machine Learning: We use a computer program to analyse the data and find patterns. Think of it like teaching a robot to recognize what factors make a property expensive or affordable.

Training: We feed our program lots of data, both the details of properties and their actual sale prices. The program learns from this information to make predictions.
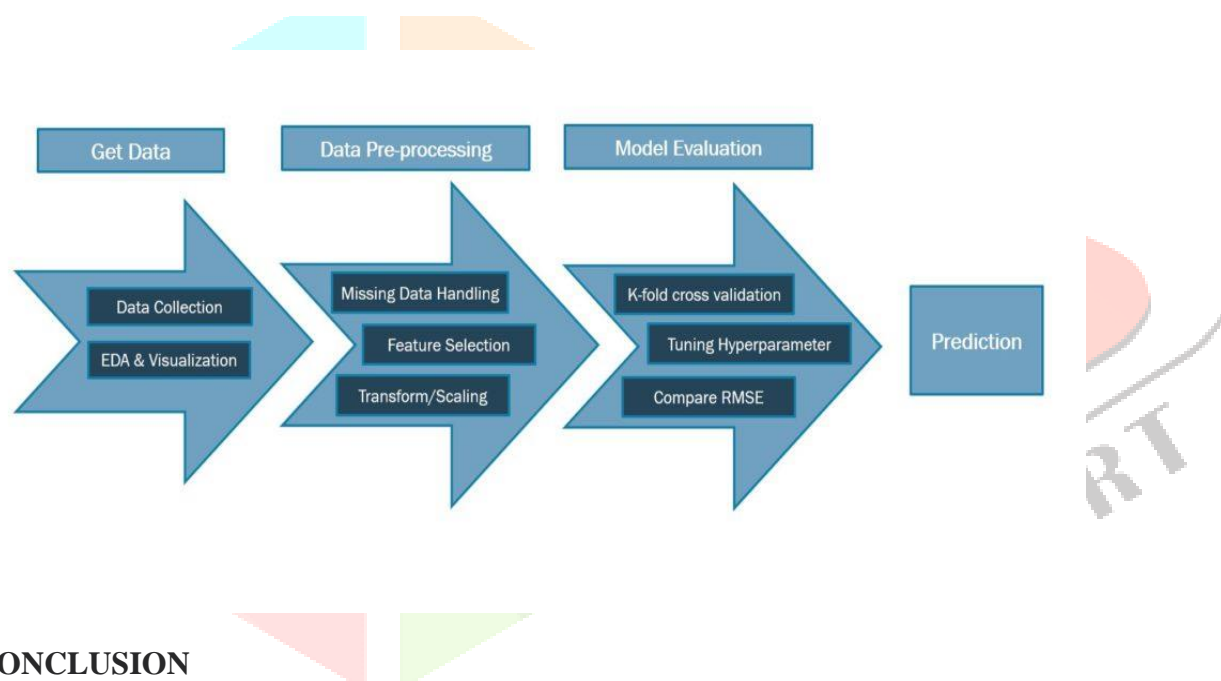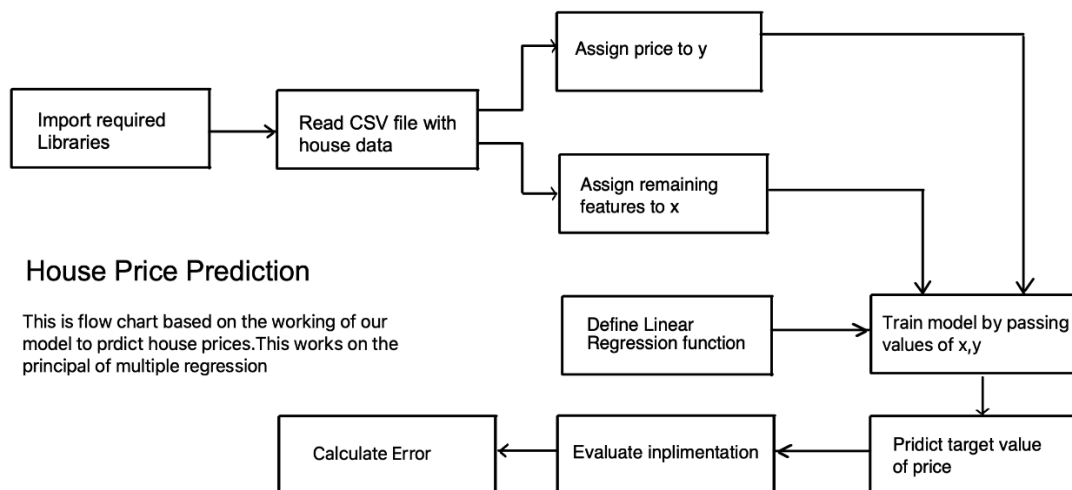
Algorithm: This is like the secret sauce of our recipe. It is the set of rules our program follows to make predictions. There are many algorithms, and we choose the one that works best for our data.

Testing: We need to check if our program is good at predicting prices. So, we give it some new, unseen data, and see how close its predictions are to the actual sale prices.

Accuracy: We measure how accurate our program is at predicting prices. If it is very close to the actual prices, it is doing a good job. If it is not close, we might need to make our program smarter.

User-Friendly Interface: Finally, we create an easy-to-use tool for people who want to predict property prices. It could be a website or an app where you input property details, and it gives you an estimate.

FLOWCHART

House Price Prediction

This is flow chart based on the working of our model to prdict house prices.This works on the principal of multiple regression

CONCLUSION

In conclusion, our real estate price prediction project has provided valuable insights into the complex world of property valuation. Using cutting-edge technology and extensive data analysis, we have achieved a better understanding of what factors influence real estate prices and how to make more accurate predictions.

Our model has demonstrated the potential to help both buyers and sellers in making informed decisions regarding real estate transactions. By leveraging historical data and advanced algorithms, we have created a tool that can estimate property values with greater precision.

However, it is important to remember that real estate is a dynamic market, influenced by numerous external factors such as economic conditions, government policies, and local developments. While our model provides a helpful framework, it is not infallible and should be used in conjunction with expert advice and market knowledge.

In the end, our project serves as a step forward in the quest for more accurate real estate price predictions. It has the potential to be a valuable resource for those looking to navigate the real estate market, offering insights that can aid in making more informed and strategic decisions. But it should always be used as one piece of the puzzle, alongside human expertise, and a broader understanding of the ever-changing real estate landscape.

REFERENCES

Scikit-Learn: A powerful Python library for machine learning and predictive modelling. scikit-learn.org

Pandas: Essential for data manipulation and analysis. pandas.pydata.org

Matplotlib and Seaborn: For data visualization. matplotlib.org and seaborn.pydata.org

"House Prices: Advanced Regression Techniques" Kaggle Competition: This Kaggle competition offers a dataset and code examples for predicting house prices. Competition Link

Towards Data Science on Medium: Offers articles on data science and machine learning. medium.com/towards-data-science

The Danh Phan, "Housing Price Prediction using Machine Learning Algorithms: The Case of Melbourne City, Australia," 2018

Chat-GPT : https://chat.openai.com