Forecasting Changes in House Prices in the United States

Modupe Olayinka | C00293417

Abstract

Despite the desire for urban living, the market for all resources is growing more competitive. Rent, food, water, and electricity are unaffordable for a middle-class household. Living costs have increased significantly due to inflation making a comeback and the 8% price increase in the US housing market. The cost of apartments in the city is increasing and estimating the price of a home is terrifying. The housing market is a crucial indicator of how well a country's economy is performing, and house price forecasting is a crucial aspect of real estate. We are motivated to predict changes in US property prices. To find models that apply to home buyers and sellers, machine learning techniques are utilized to evaluate historical real estate transaction data from the Federal Reserve and house price data from Zillow. To predict future home prices, we will assist in reasonably forecasting house values. This data will be merged and combined, and a random forest model will be trained using it. The model will forecast whether future home prices will rise or fall. Backtesting will be used to quantify the error before adding additional predictors to strengthen our model.

Introduction

For years, house price prediction has been a popular research problem because traditional house price prediction based on cost and sale price comparison does not meet accepted standards and certification process. An accurate house price prediction is critical for stakeholders in the real estate industry, which is growing rapidly in many countries, such as homeowners, customers, and estate agents.

According to [2], the US housing market increased by 7.4% year over year in September. Simultaneously, the number of homes sold fell 22.8% while the number of homes for sale increased 2.6%. According to [3,] housing markets have a positive impact on a country's currency, which is an important scale in the national economy. Homeowners will buy goods such as furniture and household equipment for their homes, and homebuilders or contractors will buy raw materials to build houses to meet house demand, indicating the economic wave effect caused by the new house supply.

In this study, we will use data from the US Federal Reserve as well as Zillow house price data. The goal of this project is to assist individuals in choosing the ideal period to invest in real estate in Louisiana. We will forecast whether home prices will rise or decline in the upcoming quarter.

Methods

Data Collection and Exploration

Data sets from the US Federal Reserve and Zillow were used in this study. The data set from the US Federal Reserve contained information about inflation and mortgage interest rates and it was collected with 2681 entries. The Zillow dataset was collected with 5 features for homes from various suburbs in the United States. Data manipulation and exploratory data analysis were conducted to uncover underlying trends, understand the features and identify inconsistencies to prepare the data sets for machine learning approaches. Additional features were also added to the data set through feature engineering to improve how well our model performed.

The dataset was divided into two; training and testing because a supervised learning algorithm was going to be employed. For the training data set, 80 % of the dataset was used to train the model while the remaining 20 % was used for testing the model on unseen data.

Random Forest Classifier

After cleaning and exploring the housing data set, we developed a Random Forest classifier model to predict house prices. Random forest is one of the ensembles learning methods used for classification and regression. It was implemented using one of the popular machine learning libraries in Python; Scikit learn. Random Forest classifier was used to average the outcomes of multiple decision trees which were applied to several subsets of our datasets to improve our classification accuracy. Instead of relying on one decision tree, the random forest uses the branch from each tree to figure out the final performance based on the majority votes. This helped us forecast if house prices will go up in the next quarter or not given economic indicators.

To evaluate our model's performance, an accuracy score will be generated to get the ratio of the number of correct predictions to the total number of input samples.

Results and Discussion

To understand the dataset better, data exploration was carried out. Fig. 1 shows the distribution of the houses price over time. It shows the sharp increase of house prices with time with houses prices above 400k as of 2022. This is due to issues like inflation, demand for houses etc. We know that in inflation, everything generally gets expensive. We decided to carry out an analysis without inflation being a factor that affects housing prices.

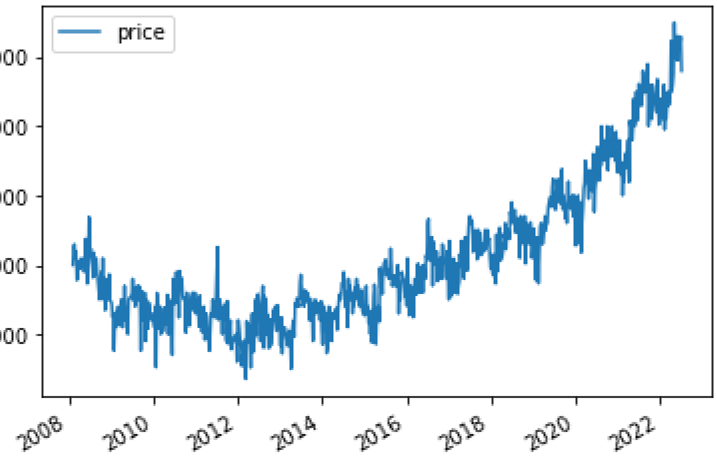

Fig 1. Housing prices against time

In Fig 2 we see that there is also a change in house prices but not on the same magnitude as with inflation being part of the factors that predict house prices. In 2014 we see the average price around $95000, and in 2022 it goes up to an average of $13500

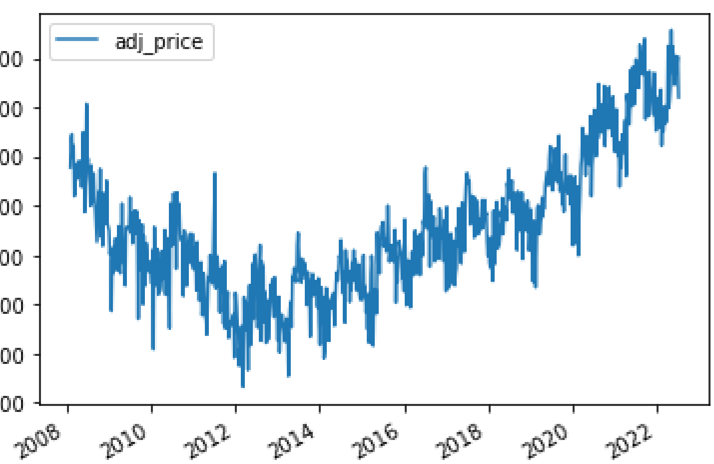

Fig. 2 Housing prices against time excluding inflation

Testing the Proposed Model

After training the model with the training dataset, the next phase of the study is to test the predictive prowess of the model. This was achieved by removing the missing values for the next quarter's column, setting up our target variable “change” which is house prices going up or down for the next quarter and listing our predictor variables. The predicted and actual house prices were then combined, and the difference was computed.

Performance Evaluation of the Proposed Model

After training and testing the model, accuracy evaluation metrics were used to get the performance of the model. The accuracy score of our model was 0.63

```
accuracy
                                                                Python
0.6341991341991342
```

To improve our model, we added extra information to the model by adding new variables that show the recent trend in house prices. This improved our accuracy score to 0.75.

```
accuracy
                                                                Python
0.7532467532467533
```

After getting the performance of the model a scatter plot was generated to show the diagnostics between the actual value and the predicted value from the model. This is shown in Fig.3. with green being the predicted value and it was correct and red the predicted value was incorrect. An appreciable performance of the correct predicted prices over the incorrect predicted values could be as a result of the K-fold cross-validation and Coupling effect of multiple regressions. The k-fold cross-validation approach is a good way to find a good bias-variance trade- off. This approach is used by Stacking Regression to determine the generalization efficiency of each variable model.

Various regression methods can complement one another. The second stacking level will learn, and correctly forecast house prices based on the first stacking level's pre-estimated prices.

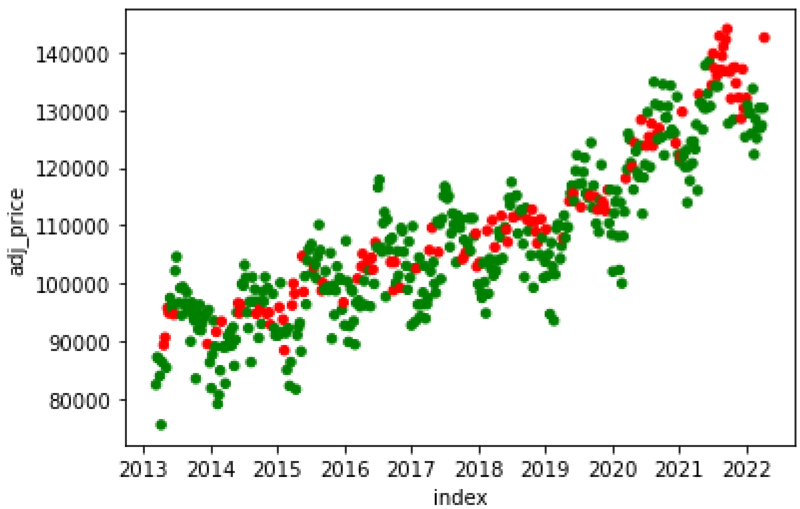

Fig. 3 Scatter Plot Real vs Predicted

Conclusion

Living expenses have risen dramatically due to inflation rearing its ugly head once more and the 8% price increase in the US home market. It is therefore important that mechanisms are created to forecast the potential change in house prices. Landowners, real estate investors, and policymakers may use house price prediction to calculate the valuation of a home and the acceptable sale price. This will assist potential buyers in determining the ideal period to invest in real estate for a home. This project has shown how important economic indicators are in influencing whether house prices increase over time. Random Forest Classifier has proved to be a great machine learning technique in forecasting house prices after evaluating the accuracy score. Future work should include exploring other machine learning models to predict house prices.

References

1. V. Limsombunchai, “House price prediction: hedonic price model vs. artificial neural network”, New Zealand agricultural and resource economics society conference, 2004, pp. 25–26.
2. https://www.redfin.com/us-housing-market#news
3. S. Temür, M. Akgün, and G. Temür, “Predicting Housing Sales in Turkey Using Arima, Lstm and Hybrid Models,” J. Bus. Econ. Manag., vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190
4. A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande and G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique", Procedia Comput. Sci., vol. 199, pp. 806-813, 2021
5. C. H. R. Madhuri, G. Anuradha and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study", 6th IEEE Int. Conf. &&&Smart Struct. Syst. ICSSS 2019, pp. 1-5, 2019
6. S. A. Basher and P. Sadorsky, "Forecasting Bitcoin price direction with random forests: How important are interest rates inflation and market volatility?", Mach. Learn. with Appl., vol. 9, pp. 100355, 2022
7. Zhou, T., Clapp, J. M., & Lu-andrews, R. (2021). Is the behavior of sellers with expected gains and losses relevant to cycles in house prices? Journal of Housing Economics, 52(May 2020), 101750. https://doi.org/10.1016/j.jhe.2021.101750