PREDICTING HOUSE PRICE USING MACHINE LEARNING

Phase 4 submission document

Project Title: House Price Prediction

Phase 4: Development Part II

ABSTRACT

The trend of the sudden drop or constant rising of housing prices has attracted interest from the researcher as well as many other interested people. There have been various research works that use different methods and techniques to address the question of the changing of house prices. This work considers the issue of changing house price as a classification problem and discuss machine learning techniques to predict whether house prices will rise or fall using available data. This work applies various feature selection techniques such as variance influence factor, Information value, principle component analysis, and data transformation techniques such as outlier and missing value treatment as well as different transformation techniques. The performance of the machine learning techniques is measured by the four parameters of accuracy, precision, specificity, and sensitivity. The work considers two discrete values 0 and 1 as respective classes. If the value of the class is 0 then we consider that the price of the house has decreased and if the value of the class is 1 then we consider that the price of the house has increased.

INTRODUCTION

Development of civilization is the foundation of the increase in demand for houses day by day. Accurate prediction of house prices has been always a fascination for buyers, sellers, and bankers also. Many researchers have already worked to unravel the mysteries of the prediction of house prices. Many theories have been given birth as a consequence of the research work contributed by various researchers all over the world. Some of these theories believe that the geographical location and culture of a particular area determine how the home prices will increase or decrease whereas other schools of thought emphasize the socio-economic conditions that largely play behind these house price rises.

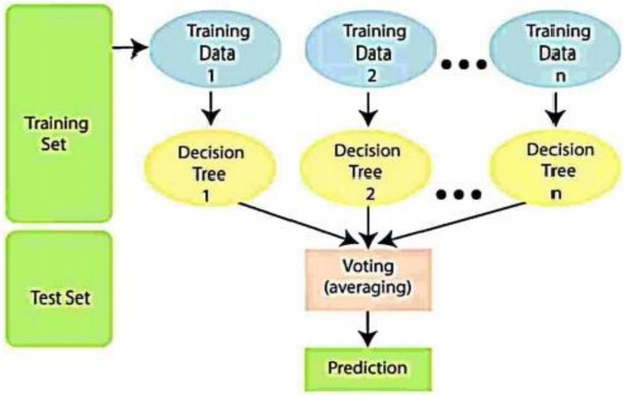
We all know that a house price is a number from some defined assortment, so obviously prediction of prices of houses is a regression task. To forecast house prices one person usually tries to locate similar properties in his or her neighborhood and based on collected data that person will try to predict the house price.

All these indicate that house price prediction is an emerging research area of regression that requires the knowledge of machine learning. This has motivated me to work in this domain.

Reale state appraisal is an integral part of the property buying process. Traditionally, the appraisal is performed by professional appraisers specially trained for real estate valuation.

MODELS FOR HOUSE PRICE PREDICTION

The Random Forest model proved to be the most appropriate model giving the highest value of the R-square and the minimum Root Mean Square Error Value. Hence, it can be concluded that the Random Forest Model is the most appropriate model for this dataset and should be used for predicting house prices.



Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

Use of Random Forest

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

Working Of Random Forest algorithm

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

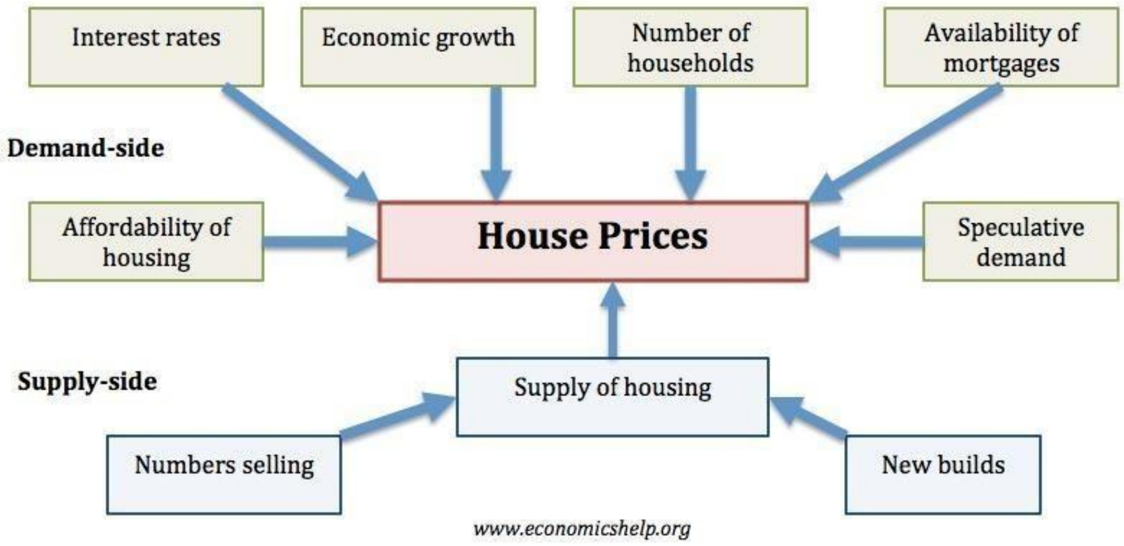
FEATURES OF HOUSE PRICE PREDICTION

House price prediction is a complex task that involves the use of various features to estimate the value of a property accurately. These features, also known as predictors or input variables, can vary depending on the specific machine learning model and dataset used, but here are some common features that are often considered when predicting house prices:

1. **Location**: The location of the property is a critical factor. Features related to location may include the neighborhood, proximity to schools, parks, public transportation, and the distance to amenities like shopping centers and hospitals.
2. **Property Size**: Features like the total area of the property (in square feet or square meters), the number of bedrooms, bathrooms, and the size of the yard can significantly impact the price.
3. **Property Age and Condition**: The age of the property and its overall condition, including recent renovations or updates, can affect its value.
4. **Amenities and Features**: Special features such as a swimming pool, garage, fireplace, central heating, air conditioning, or smart home systems can influence the price.
5. **Historical Sales Data**: Past sales data of similar properties in the area can be a useful feature for predicting the current property's price.
6. **Market Trends**: Current real estate market conditions, such as supply and demand, interest rates, and economic factors, can play a role in price prediction.
7. **Crime Rate**: Safety is a concern for potential homebuyers, so the local crime rate can be a relevant feature.
8. **School Quality**: Proximity to good schools and the overall quality of the education system in the area can be a significant factor for families.
9. **Transportation**: Access to public transportation and commuting options can affect property values.
10. **Local Amenities**: Proximity to parks, shopping centers, restaurants, and other local amenities can influence the price

FACTORS THAT AFFECT HOUSE PRICING

In order to predict house prices, first we have to understand the factors that affect house pricing.

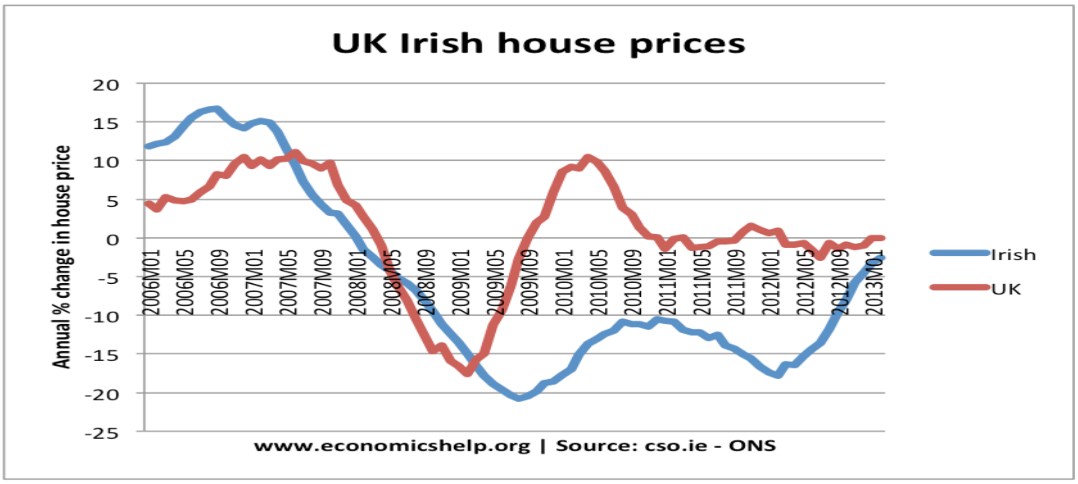


- **Economic growth.** Demand for housing is dependent upon income. With higher economic growth and rising incomes, people will be able to spend more on houses; this will increase demand and push up prices. In fact, demand for housing is often noted to be income elastic (luxury good); rising incomes leading to a bigger % of income being spent on houses. Similarly, in a recession, falling incomes will mean people can't afford to buy and those who lose their job may fall behind on their mortgage payments and end up with their home repossessed.

- **Unemployment.** Related to economic growth is unemployment. When unemployment is rising,fewer people will be able to afford a house. But, even the fear of unemployment may discouragepeople from entering the property market.

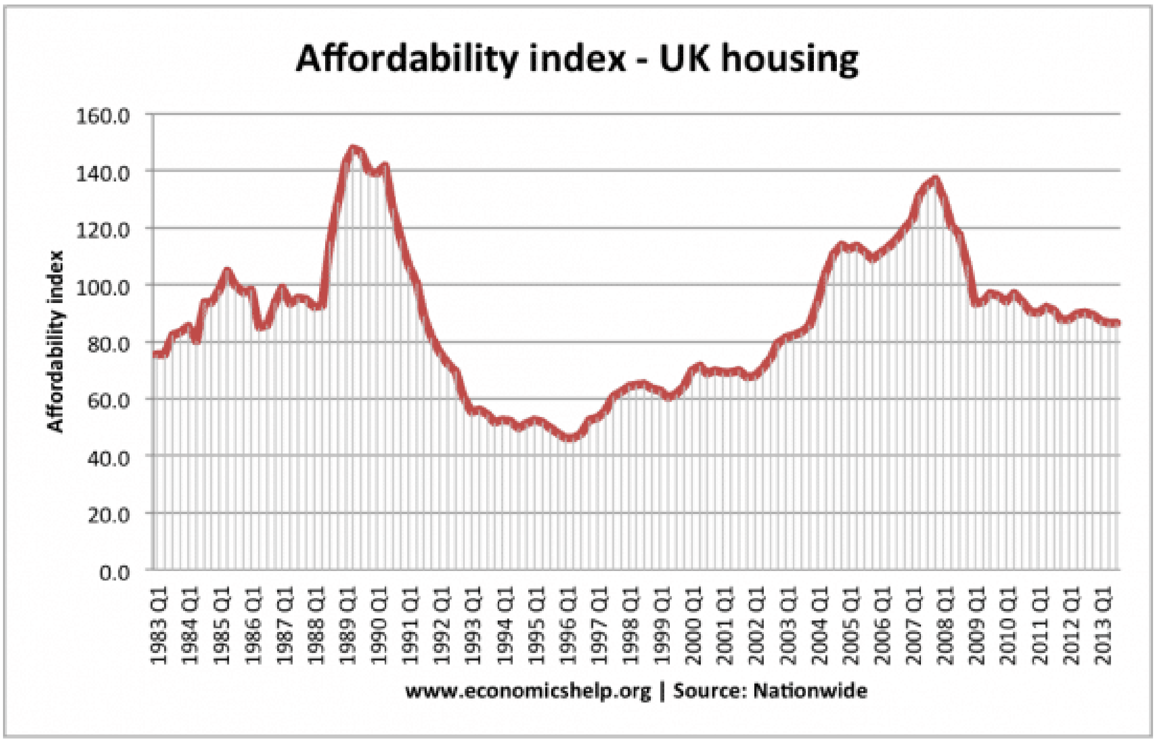
- **Interest rates.** Interest rates affect the cost of monthly mortgage payments. A period of high- interest rates will increase cost of mortgage payments and will cause lower demand for buying a house. High-interest rates make renting relatively more attractive compared to buying. Interest rates have a bigger effect if homeowners have large variable mortgages. For example, in 1990-92, the sharp rise in interest rates caused a very steep fall in UK house prices because many homeowners couldn’t afford the rise in interest rates.

- **Consumer confidence**. Confidence is important for determining whether people want to take therisk of taking out a mortgage. In particular expectations towards the housing market is important; if people fear house prices could fall, people will defer buying.
- **Mortgage availability**. In the boom years of 1996-2006, many banks were very keen to lend mortgages. They allowed people to borrow large income multiples (e.g. five times income). Also, banks required very low deposits (e.g. 100% mortgages). This ease of getting a mortgage meant that demand for housing increased as more people were now able to buy. However, since the credit crunch of 2007, banks and building societies struggled to raise funds for lending on the money markets. Therefore, they have tightened their lending criteria requiring a bigger deposit to buy a house. This has reduced the availability of mortgages and demand fell.
- **Supply**. A shortage of supply pushes up prices. Excess supply will cause prices to fall. For example, inthe Irish property boom of 1996-2006, an estimated 700,000 new houses were built. When the property market collapsed, the market was left with a fundamental oversupply. Vacancy rates reached 15%, and with supply greater than demand, prices fell.



By contrast, in the UK, housing supply fell behind demand. With a shortage, UK house prices didn't fall as much as in Ireland and soon recovered – despite the ongoing credit crunch. The supply of housing depends on existing stock and new house builds. Supply of housing tends to be quite inelastic because to get planning permission and build houses is a time-consuming process. Periods of rising house prices may not cause an equivalent rise in supply, especially in countries likethe UK, with limited land for home-building.

- **Affordability/house prices to earnings.** The ratio of house prices to earnings influences the demand. As house prices rise relative to income, you would expect fewer people to be able to afford. For example, in the 2007 boom, the ratio of house prices to income rose to 5. At this level, house prices were relatively expensive, and we saw a correction with house prices falling.



Another way of looking at the affordability of housing is to look at the percentage of take-home paythat is spent on mortgages. This takes into account both house prices, but mainly interest rates and the cost of monthly mortgage payments. In late 1989, we see housing become very unaffordable because of rising interest rates. This caused a sharp fall in prices in 1990-92.

EVALUATION OF HOUSE PRICE PREDICTION

The evaluation of a house price prediction model is crucial to determine its accuracy and performance. Several evaluation metrics and techniques can be used to assess the effectiveness of the model. Here are some common methods and metrics for evaluating house price prediction models: Mean Absolute Error (MAE): MAE measures the average absolute difference between the predicted house prices and the actual prices. A lower MAE indicates a better model fit.

1. Root Mean Square Error (RMSE): RMSE is similar to MAE but places more weight on larger errors. It is calculated by taking the square root of the average squared differences between predictions and actual values.
2. Mean Absolute Percentage Error (MAPE): MAPE expresses the prediction error as a percentage of the actual value. It's useful for understanding the relative magnitude of errors in the context of house prices.
3. R-squared (R^2) or Coefficient of Determination: R^2 measures the proportion of the variance in the dependent variable (house prices) that is explained by the model. A higher R^2 value indicates a better fit.
4. Cross-Validation: Use techniques like k-fold cross-validation to assess the model's performance across multiple subsets of the data. Cross-validation helps ensure that the model generalizes well to unseen data.
5. Residual Analysis: Examine the model's residuals (the differences between predicted and actual values) to check for patterns, such as heteroscedasticity (non-constant variance of residuals) and autocorrelation. Residual plots can reveal issues with the model's assumptions.
6. Outliers Detection: Identify and analyze outliers in the predicted vs. actual house prices. Extreme outliers can skew evaluation metrics and should be investigated.
7. Feature Importance: Assess the importance of each input feature in the model. This can help you understand which features have the most influence on house price predictions and may guide feature selection or engineering.
8. Domain Expert Review: Involve domain experts, such as real estate professionals, to assess the model's predictions. Their insights can help identify whether the model makes sense in a real-world context.
9. Business Metrics: Evaluate the model's performance in terms of business goals. For example, if you're building the model for a real estate company, assess how well it aligns with their objectives, such as maximizing profit or minimizing losses.

ADVANTAGE OF LSTM OVER OTHER MODELS

The LSTM model can be tuned for various parameters such as changing the number of LSTMlayers, adding dropout value or increasing the number of epochs.

Long Short Term Memory (LSTM)
LSTMs are widely used for sequence prediction problems and have proven to be extremely effective. The reason they work so well is because LSTM is able to store past information that is important, and forget the information that is not. LSTM has three gates:
The input gate: The input gate adds information to the cell state

The forget gate: It removes the information that is no longer required by the model. Theoutput gate: Output Gate at LSTM selects the information to be shown as output.

CONCLUSION

In conclusion, house price prediction is a complex and vital aspect of the real estate industry. It involves the analysis of numerous variables, including location, property features, economic trends, and market conditions. Several methods and techniques, such as regression analysis, machine learning algorithms, and deep learning models, can be employed to forecast house prices.

The accuracy of house price predictions depends on the quality and quantity of data, as well as the appropriateness of the chosen model. It is essential to continually update and refine the prediction models as new data becomes available and market conditions change.