

Real Estate Trend Prediction with Artificial Intelligence Techniques

Authors : Sophia Liang Zhou

Abstract : For investors, businesses, consumers, and governments, an accurate assessment of future housing prices is crucial to critical decisions in resource allocation, policy formation, and investment strategies. Previous studies are contradictory about macroeconomic determinants of housing price and largely focused on one or two areas using point prediction. This study aims to develop data-driven models to accurately predict future housing market trends in different markets. This work studied five different metropolitan areas representing different market trends and compared three-time lagging situations: no lag, 6-month lag, and 12-month lag. Linear regression (LR), random forest (RF), and artificial neural network (ANN) were employed to model the real estate price using datasets with S&P/Case-Shiller home price index and 12 demographic and macroeconomic features, such as gross domestic product (GDP), resident population, personal income, etc. in five metropolitan areas: Boston, Dallas, New York, Chicago, and San Francisco. The data from March 2005 to December 2018 were collected from the Federal Reserve Bank, FBI, and Freddie Mac. In the original data, some factors are monthly, some quarterly, and some yearly. Thus, two methods to compensate missing values, backfill or interpolation, were compared. The models were evaluated by accuracy, mean absolute error, and root mean square error. The LR and ANN models outperformed the RF model due to RF's inherent limitations. Both ANN and LR methods generated predictive models with high accuracy (> 95%). It was found that personal income, GDP, population, and measures of debt consistently appeared as the most important factors. It also showed that technique to compensate missing values in the dataset and implementation of time lag can have a significant influence on the model performance and require further investigation. The best performing models varied for each area, but the backfilled 12-month lag LR models and the interpolated no lag ANN models showed the best stable performance overall, with accuracies > 95% for each city. This study reveals the influence of input variables in different markets. It also provides evidence to support future studies to identify the optimal time lag and data imputing methods for establishing accurate predictive models.

Keywords : linear regression, random forest, artificial neural network, real estate price prediction

Conference Title : ICDSAA 2021 : International Conference on Data Science and Advanced Analytics

Conference Location : Montreal, Canada

Conference Dates : June 14-15, 2021