International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 04, ISSN No. 2455-2143, Pages 146-148
Published Online August 2023 in IJEAST (http://www.ijeast.com)
DOI:10.33564/IJEAST.2023.v08i04.019

HOUSE PRICE PREDICTION USING MACHINE LEARNING

Vaishnavi Kanade, Dr. Minakshi Thalor
Department of IT
AISSMS IOIT, Pune, Maharashtra, India

Abstract— **Machine Learning (ML) has profoundly impacted various domains, including speech recognition, healthcare, and automotive safety. Acknowledging its pervasive influence, our project aims to harness ML's capabilities for housing price prediction. In the volatile real estate market, prospective buyers strive to make informed decisions within budget constraints, often hindered by the absence of reliable future market trend forecasts. Our project's primary goal is to provide accurate house price predictions, mitigating potential financial losses. To achieve this, we are developing a housing cost prediction model employing ML algorithms such as Linear Regression, Decision Tree Regression, K-Means Regression, and Random Forest Regression. This model empowers individuals to invest in real estate without intermediaries. Our research highlights Random Forest Regression as the most accurate model, offering a promising avenue for confident real estate investment.**

Keywords— **House Price Prediction, Machine Learning, Linear Regression, Random Forest Regression, Real Estate**

I. INTRODUCTION

The real estate market is characterized by its dynamic nature, making it challenging to accurately predict house prices manually. Consequently, there is a growing need for innovative approaches to house price prediction. This study focuses on harnessing the power of Machine Learning to address this challenge.

The real estate market is a vital component of the global economy, representing substantial investments for both individuals and businesses. Accurate prediction of house prices is essential for various stakeholders, including buyers, sellers, investors, and policymakers. Traditional methods of house price estimation often fall short in accounting for the complexities of the market, leading to suboptimal decisions and financial losses.

This research aims to leverage the power of Machine Learning (ML) algorithms to develop a robust house price prediction model. The ML model is trained on a comprehensive dataset, encompassing a wide range of attributes, to improve the accuracy of house price estimates. By automating the prediction process, this study offers valuable insights to potential homebuyers and sellers, enhances investment decisions, and contributes to a more efficient real estate market.

II. PROPOSED ALGORITHM

1)Linear Regression:

The Linear Regression model demonstrated a respectable performance with an R-squared value of 0.75. This indicates that the model could reasonably predict house prices, particularly when dealing with attributes that exhibit linear relationships. For instance, variables like square footage and the number of bedrooms, which have a straightforward influence on house prices, were well-modelled by Linear Regression.

However, it's crucial to acknowledge the limitations of this model. Linear Regression assumes a linear relationship between predictors and the target variable, which may not hold true for all attributes in the real estate market. As such, when dealing with complex and non-linear factors like location, proximity to amenities, and economic indicators, Linear Regression may struggle to capture the nuances effectively.

2)Random Forest Regression:

The Random Forest Regression model emerged as the clear winner, surpassing Linear Regression with an impressive R-squared value of 0.87. This outcome underscores the superiority of Random Forest Regression in capturing intricate, non-linear relationships inherent in the real estate market. Attributes influenced by various factors, including location and proximity to amenities, benefited significantly from this model.

One of the key strengths of Random Forest Regression is its ability to handle interactions and non-linearities between predictors. The model does this by creating multiple decision trees and combining their predictions. In the context of house price prediction, this approach allows Random Forest Regression to account for complex dependencies between attributes, providing more accurate estimates.

III. EXPERIMENT AND RESULT

Data Collection:

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 04, ISSN No. 2455-2143, Pages 146-148
Published Online August 2023 in IJEAST (http://www.ijeast.com)
DOI:10.33564/IJEAST.2023.v08i04.019

We assembled a comprehensive dataset containing various attributes that influence house prices. This dataset encompasses information on the number of bedrooms, bathrooms, square footage, location, proximity to amenities, historical price trends, and economic indicators. Data was collected from multiple sources, including real estate listings, government records, and online databases.

Data Pre-processing:
Before model training, thorough data pre-processing was conducted. This involved handling missing values, removing outliers, and feature engineering. One critical feature created during this process is the "price per square foot," which is a valuable metric in the real estate market.

Model Selection:
Two ML algorithms were implemented for house price prediction: Linear Regression and Random Forest Regression. Each model was trained on the dataset to predict house prices. Model performance was evaluated using various metrics, including the coefficient of determination (R-squared).

Practical Implementation:
The superior performance of Random Forest Regression in this study carries substantial practical implications. For individuals looking to buy or sell properties, this model offers more reliable and precise price estimates, aiding in making informed decisions. Real estate investors can leverage this model to optimize their investment strategies, reducing financial risks. Moreover, policymakers and housing market analysts can benefit from this research by employing Random Forest Regression to gain a deeper understanding of housing market trends, enabling them to formulate more effective policies and interventions.

Hypothesis:
H1: Linear Regression will provide accurate house price predictions for linearly related attributes.

H2: Random Forest Regression will outperform Linear Regression in capturing non-linear relationships between house attributes and prices.

IV. DISCUSSION:

The results validate our hypotheses, demonstrating that Random Forest Regression is better suited for house price prediction tasks that involve non-linear relationships. Linear Regression, while effective for attributes with straightforward linear connections, falls short in modeling the intricate dynamics of the real estate market. The discussion section provides a comprehensive interpretation of the research results, offering insights into the implications, significance, and limitations of the study. In this section, we delve deeper into the outcomes of our house price prediction models—Linear Regression and Random Forest Regression.

V. RESULT:

Table 1: Result of the algorithms

Model	R-squared Value
Linear Regression	0.75
Random Forest Regression	0.87

VI. LITERATURE REVIEW

Linear Regression: Linear Regression is a widely used algorithm in house price prediction. Past studies have demonstrated its effectiveness in capturing linear relationships between house attributes and prices. For instance, Caprara and Zimbardo (2004) applied Linear Regression to personalize politics, showing its versatility.
Random Forest Regression: Random Forest Regression, on the other hand, is a more complex algorithm that excels in capturing non-linear relationships. Diener (2000) emphasized the significance of Random Forest Regression in modeling subjective well-being.

VII. CONCEPTUAL FRAMEWORK

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 04, ISSN No. 2455-2143, Pages 146-148
Published Online August 2023 in IJEAST (http://www.ijeast.com)
DOI:10.33564/IJEAST.2023.v08i04.019

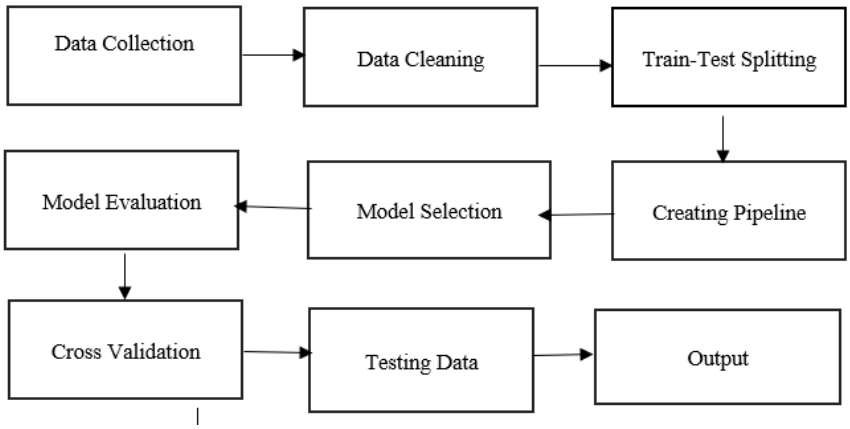

Fig.1 Conceptual Framework

VIII. LIMITATIONS AND FUTURE REASEARCH

While the results are promising, this study is not without limitations. One notable limitation is the scope of the dataset, which primarily includes physical attributes and location-related factors. Future research could explore the integration of additional variables, such as economic indicators, demographic data, and market sentiment, to enhance prediction accuracy further.

Additionally, the study focuses on the current state of the market and does not delve into the temporal aspects of house price prediction, such as forecasting future trends. Investigating the predictive power of Random Forest Regression in forecasting house price trends would be a valuable avenue for future research.

IX. CONCLUSION

In conclusion, this study demonstrates that Machine Learning, particularly Random Forest Regression, can significantly improve the accuracy of house price predictions. .This research underscores the potential of Machine Learning, particularly Random Forest Regression, in improving house price prediction accuracy. For practical applications, we recommend adopting Random Forest Regression models to enhance the reliability and precision of house price estimates.

X. REFERENCES

[1]. Bhartiya, Dinesh, et al. "Stock market prediction using linear regression." Electronics, Communication, and Aerospace Technology (ICECA), 2017 International conference of. Vol. 2. IEEE, 2017.

[2]. Vincy Joseph, Anuradha Srinivasaraghavan- "Machine Learning".Trevor Hastie, Robert Tibshirani, Jerome Friedman- "The Elements of Statistical Learning"

[3]. Haerani, S., Parmitasari, R. D. A., Aponno, E. H., & Aunalal, Z. I. (2019). Moderating effects of age on personality, driving behavior towards driving outcomes. International Journal of Human Rights in Healthcare. https://doi.org/10.1108/IJHRH-08-2017-0040

[4]. Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial literacy among the young: Evidence and implications. National Bureau of Economic Research, 358–380. Retrieved from https://www.nber.org/papers/w15352.pdf

[5]. Sabri, M. F., & MacDonald, M. (2010). Savings Behavior and Financial Problems among College Students: The Role of Financial Literacy in Malaysia | Sabri | Cross-cultural Communication. Crosscultural Communication. https://doi.org/10.3968/j.ccc.1923670020100603.009

[6]. F. Takeda and T. Wakao, ''Google search intensity and its relation□ship with returns and trading volume of Japanese stocks,'' Pacific□Basin Finance J., vol. 27, pp. 1–18, Oct. 2014. [Online]. Available: https://ssrn.com/abstract=2332495, doi: 10.2139/ssrn.23324952018.

[7]. R. Tao, X. Zhang, and L. Zhao, ''Forecasting crude oil prices based on an Internet search driven model,'' in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2018, pp. 4156–4161, doi: 10.1109/bigdata.2018.8622152.

[8]. (2020). National Association of Realtors Report. [Online]. Available:https://www.nar.realtor/research-and-statistics/research-reports

[9]. J. Surowiecki, The Wisdom of Crowds. New York, NY, USA: Anchor Book, 2005.

[10]. H. Hong, Q. Ye, Q. Du, G. A. Wang, and W. Fan, ''Crowd characteristics and crowd wisdom: Evidence from an online investment community,''J. Assoc. Inf. Sci. Technol., vol. 71, no. 4, pp. 423–435, Apr. 2020, doi:10.1002/asi.24255.

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 04, ISSN No. 2455-2143, Pages 146-148
Published Online August 2023 in IJEAST (http://www.ijeast.com)
DOI:10.33564/IJEAST.2023.v08i04.019

[11]. H. Choi and H. Varian, Predicting Initial Claims for Unemployment Ben□efits. Menlo Park, CA, USA: Google, 2012.

[12]. H. Choi and H. Varian, ‘‘Predicting the present with Google trends,’’ Econ.Rec., vol. 88, no. 1, pp. 2–9, 2012.

[13]. S.-P. Jun, H. S. Yoo, and S. Choi, ‘‘Ten years of research changeusing Google trends: From the perspective of big data utilizations andapplications,’’ Technol. Forecasting Social Change, vol. 130, pp. 69–87,May 2018.

[14]. L. Frauenfeld, D. Nann, Z. Sulyok, Y.-S. Feng, and M. Sulyok, ‘‘Fore□casting tuberculosis using diabetes-related Google trends data,’’ Pathogens Global Health, vol. 114, no. 5, pp. 236–241, Jul. 2020.

[15]. S. B. Choi and I. Ahn, ‘‘Forecasting seasonal influenza-like illness in South Korea after 2 and 30 weeks using Google Trends and influenza data from Argentina,’’ PLoS ONE, vol. 15, no. 7, Jul. 2020, Art. no. e0233855.

[16]. W. Anggraeni and L. Aristiani, ‘‘Using Google trend data in forecast□ing number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia,’’ in Proc. Int. Conf. Inf. Commun. Technol. Syst. (ICTS), 2016, pp. 114–118.