Application of Machine Learning Models to Predict Real Estate Values

Bùi Đức Công
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
student497@ictu.edu.vn

***Abstract*—This undergraduate thesis explores the application of machine learning models to predict real estate values. The author discusses various machine learning models, including logistic regression, random forests, and K-Nearest Neighbors, and their evaluation. The work highlights the author's learning experience in Python programming, data processing, and understanding of machine learning theories. The thesis concludes with the limitations of the current model, such as limited data and a basic interface, and suggests future improvements. The author applied logistic regression, random forests, and K-Nearest Neighbors to predict real estate values, but specific performance metrics are not provided. The models were trained using a dataset of Vietnamese real estate transactions, and the author learned Python programming and data processing techniques during the process.**

***Index Terms*—machine learning, real estate, logistic regression, random forests, K-Nearest Neighbors, Python**

I. Introduction

The thesis introduces the problem of predicting real estate values and provides a background on real estate, market characteristics, and types of real estate. It also outlines the importance of machine learning in this context, focusing on logistic regression, random forests, and K-Nearest Neighbors. The author explains the significance of real estate valuation in the Vietnamese market, where accurate predictions can help investors and homeowners make informed decisions. The introduction also highlights the relevance of machine learning models in real estate, emphasizing their ability to handle complex data and provide accurate predictions [1]. Earlier ICTU student work on related designs is included here for context [5]. Related published studies provide context for the subject and methods [6], [7], [8], [9], [10]. These contextual works were selected for this English summary; they are not asserted to appear in the original thesis bibliography.

II. Methods

The author uses Python for programming and employs machine learning libraries such as Scikit-Learn, Google Colaboratory, and TensorFlow. The models are evaluated using classification reports and grid search for tuning. The thesis does not report specific numerical results but describes the process of model training and evaluation. The author trained logistic regression, random forests, and K-Nearest Neighbors models using a dataset of Vietnamese real estate transactions. The dataset included features such as location, size, age of the property, and market trends. The author used grid search to optimize the hyperparameters of the models, but no specific performance metrics are provided [1].

Supervisor: Nguyễn Tuấn Anh, Thai Nguyen University of Information and Communication Technology (anhnt@ictu.edu.vn).

III. Results

The author describes the process of learning Python, data processing, and understanding machine learning theories. The models were tested using logistic regression, random forests, and K-Nearest Neighbors, but no specific performance metrics are provided. The author mentions that the models were trained using a dataset of Vietnamese real estate transactions, but the exact number of data points or the performance of the models is not reported. The author also notes that the models were tested using classification reports, but the results are not detailed in the thesis [1].

IV. Discussion

The author discusses the advantages of the models, such as high accuracy and ease of use, but also highlights the limitations, including the small dataset size and the need for further optimization. The author plans to continue learning and improving the model in the future. The author suggests that the models could be improved by using a larger and more diverse dataset, as well as by incorporating additional features such as economic indicators and demographic data. The author also mentions the need for a more sophisticated user interface to make the models more accessible to users [1].

V. Conclusion

The thesis concludes with the author's reflections on the learning experience and the need for further data collection and model optimization to improve the accuracy of the predictions. The author emphasizes the importance of continued learning and improvement in the field of machine learning for real estate valuation. The author also suggests that the models could be applied in practical scenarios, such as real estate investment and property management, but no specific applications are provided [1].

References

[1] Bùi Đức Công, ”Application of Machine Learning Models to Predict Real Estate Values,” internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] Jose, J. (2022). Introduction to Time Series Analysis and its Applications. Christ University, Bangalore

[3] W. C. Apgar, with R. Peng and J. Olson. Recent Trends in Real Rents. Working Paper W87-5, Joint Center for Housing Studies of Harvard University, 1987

[4] ”Grid Search for model tuning”, Medium, 2022. [Online]. Available: https://towardsdatascience.com/grid-search-for-model-tuning-3319b259367e. [Accessed: 10- Jan- 2022]

[5] ICTU Student 91, ”Information Provision System for Real Estate,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2009.

[6] Nishant ., ”Real Estate Price Prediction Using Machine Learning,” International Journal for Research in Applied Science and Engineering Technology, 2023. doi:10.22214/ijraset.2023.53166.

[7] Zhihang Liu, ”Real Estate Price Prediction based on Supervised Machine Learning Scenarios,” Highlights in Science, Engineering and Technology, 2023. doi:10.54097/hset.v39i.6637.

[8] Tansu Alkan et al., ”Using machine learning algorithms for predicting real estate values in tourism centers,” Soft Computing, 2022. doi:10.1007/s00500-022-07579-7.

[9] Cihan Çılgın et al., ”Machine learning methods for prediction real estate sales prices in Turkey,” Revista de la construcción, 2023. doi:10.7764/rdlc.22.1.163.

[10] Ping-Feng Pai et al., ”Using Machine Learning Models and Actual Transaction Data for Predicting Real Estate Prices,” Applied Sciences, 2020. doi:10.3390/app10175832.