ST. FRANCIS XAVIER
UNIVERSITY

CSCI - 525: Machine Learning Design

Dr. Jacob Levman

Phase 2 Report

Khan, Salal Ali
202307216 - x2023flb@stfx.ca

Hussain, Moayadeldin
202407556 - x2024ghl@stfx.ca

Javed, Muhammad
202303808 - x2023dvh@stfx.ca

Oct 3, 2024

Introduction

We are currently analyzing Airbnb's 2023 New York City dataset and are using machine learning to predict property prices. The real estate industry is critical for economic development and society progress, reflecting the aspirations of people, families and the overall social health of the region [1]. In the past, knowing property prices relied heavily on human experience and conventional methods. However, these approaches usually did not manage to success to account for the complexity and non-linear relationships of housing market data. With machine learning advancements, the scene has changed with the ability to extract precious information from large volumes of data [2]. Addressing the challenges, predicting house prices accurately remains demanding due to the multitude of influencing factors, as location, size, house quality, condition, etc. [3]. We are looking forward in our work to acquire knowledge from the literature of different models used and address the existing challenges in such problem.

One of the most interesting research work we found on utilizing AI in Airbnb datasets price prediction was founded in [4] by Stanford Researchers. They used plenty of models to address their target, including Ridge Regression, K-means Clustering, Support Vector Regression, Neural Netowrks and Gradient Boosting Tree Ensemble. They also used Feature Selection and Sentiment Analysis. Sentiment Analysis refers to the application of natural language processing to identify and classify subjective opinions in source materials (e.g., a document or a sentence) [5]. However, this topic is beyond our scope in our Project 1 at the course becuase it depends on transforming the words and sentences to a different feature space using a technique called word embeddings, as Word2Vec [6]. Moreover, the work in [7] focused also on Predicting Airbnb Listing Price with multiple models, utilizing Boston Airbnb open data set from Kaggle, which includes 3, 585 listings between 2016 and 2017. They used Linear Regression, K-nearest neighbor regression and Gradient Boosting regression. The authors in [3] depended on using Linear Regression, and Random Forests. In [2], Linear Regression, Random Forest Regression and Gradient Boosting Regression.

In [4], their experiments involved Mean absolute error (MAE), mean squared error (MSE) and R2 score were used to evaluate their training models. Among the models tested, Support Vector Regression performed the best and produced 69% R2 score, 0.147 MSE on test split, and 0.2132 on MAE. In [3], the Random Forest Regressor model had high accuracy store with 97.3% for training set and 82.3& for the testing set, in comparison with the linear

regression model with only 79.4% and 58.9% for training and testing. Hence, the authors concluded that the Random Regression model is the best model for study. In [2], they used MAE, RMSE and R2 Score, mostly similar to work in [4], as performance metrics. As we analyzed the works, we find that R2 scores for example in this work exceeds that in [4] for two models of the three they used, achieving 0.822 and 0.828 with both Random Forest and Gradient Boosting repsectively.

We hypothesize that the use of open source machine learning software applied to New York AirBnb's 2023 dataset may produce useful technology for House Prices prediction and analysis. We hypothesize that applying different Machine Learning, and Preprocessing technqiues, according also to our discussions with Dr. Jacob, will yield to improvments in property price and minimum nights.

References

1. Li, Chenxi. (2024). House price prediction using machine learning. Applied and Computational Engineering. 53. 225-237. 10.54254/2755-2721/53/20241426.

2. el Mouna, Lale & Hassan, Silkan & Haynf, Youssef & Nann, Mohamedade & Koumetio Tekouabou, Cédric Stéphane. (2023). A Comparative Study of Urban House Price Prediction using Machine Learning Algorithms. E3S Web of Conferences. 418. 10.1051/e3sconf/202341803001.

3. Maloku, Fatbardha. (2024). House Price Prediction Using Machine Learning and Artificial Intelligence. Journal of Artificial Intelligence & Cloud Computing. Volume 3(4): 1- 10. 1-10. 10.47363/JAICC/2024(3)357.

4. Luo, Tiejian & Chen, Su & Xu, Guandong & Zhou, Jia. (2013). Sentiment Analysis. 10.1007/978-1-4614-7202-5_4.

5. Mikolov, Tomas & Kai, Chen & Corrado,Greg and Dean, Jeffrey. (2013). Efficient Estimation of Word Representations in Vector Space. https://arxiv.org/abs/1301.3

6. Wang, Haoqian. (2023). Predicting Airbnb Listing Price with Different models. Highlights in Science, Engineering and Technology. 47. 79-86. 10.54097/hset.v47i.8169.