NSU MACHINE LEARNING

Predicting Airbnb Listing Prices of Popular Tourist Destinations in Europe

Moshiur Rahman
Shahariar Ifti
MD. Shahriar Hussain

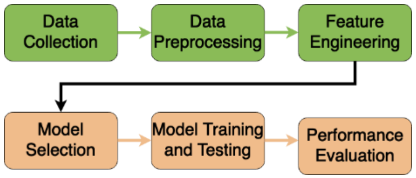
Motivation

This project presents a machine learning approach to predicting Airbnb listing prices in ten popular tourist destinations across Europe.

The goal of the project is to benefit individual hosts and travelers, as well as provide valuable insights for market analysis and strategic planning in the hospitality industry.

Several machine learning algorithms, including Linear Regression, Support Vector Regression, Decision Tree, Random Forest, and Gradient Boosting, are trained and evaluated to identify the optimal model.

System Design



The first phase is data collection, where we collect the required dataset from a verified source. Next, the second phase is data preprocessing, where we prepare the data for analysis. The third phase of the project is feature engineering, where we perform operations such as one hot encoding to transform categorical features to numerical features and feature scaling to ensure all the features have similar values. Then, the fourth phase is model selection where we choose appropriate models for the task. In the fifth phase, we split the dataset into training set and test set, and we train the selected models on the training set and test them on the test set. In the final phase of the project, we take the output values and evaluate the models using standard performance metrics.

Dataset and Features

The dataset is collected from Kaggle, and it is the dataset on which the work is based. However, we process the dataset according to our specific task.

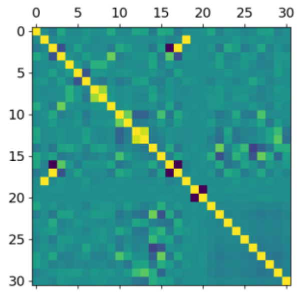
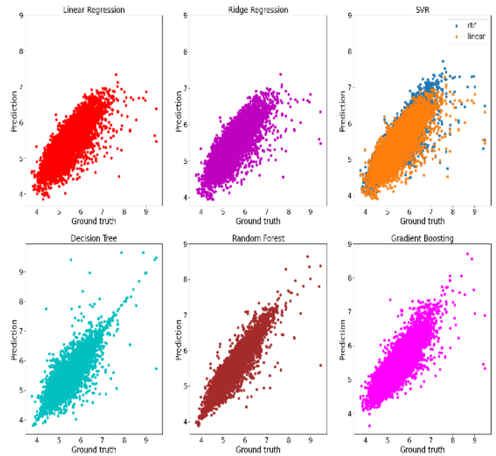


Fig. Correlation between the label and the features, as well as among the features themselves

Results



Performance Evaluation

		Train			Test		
Model	Hyperparameters	RMSE	MAE	R^2	RMSE	MAE	R^2
Linear Regression		277.060	78.523	0.664	301.513	83.227	0.656
Ridge Regression	$\alpha = 1.0$	277.417	78.863	0.661	302.037	83.639	0.653
SVR	kernel = 'rbf'	263.861	64.237	0.760	288.201	72.986	0.729
	kernel = 'linear'	278.984	78.605	0.658	303.824	83.459	0.649
Decision Tree		1.74×10^{-14}	7.38×10^{-16}	**1.0**	276.548	52.958	0.766
Random Forest	n_estimators = 100	152.909	20.313	0.979	227.065	51.915	**0.854**
Gradient Boosting	n_estimators = 100, learning_rate= 1.0	245.376	65.435	0.769	270.408	73.368	0.734

Discussion

The Random Forest algorithm emerged as the best-performing model for predicting Airbnb listing prices in popular tourist destinations across Europe, demonstrating superior accuracy and robustness in capturing the complex pricing dynamics. The model's high performance underscores its potential to provide valuable pricing insights to hosts and travelers alike.
For future work, we aim to experiment with neural networks, which may further enhance prediction accuracy by capturing more intricate patterns within the data. Additionally, we plan to scale the project globally, extending our analysis to a broader range of tourist destinations world- wide, thereby increasing the model's applicability and utility n the global hospitality market.

References

[1] K. Schmidt, "The 10 most visited cities in Europe," dw.com, 27-Sep-2023. [Online]. Available: https://www.dw.com/en/the-10-most-visitedcities-in-europe/g-66897729.
[2] L. D. Redman and L. Matthews, "The 10 best European cities to visit or live in - AFAR," AFAR Media, 05-Mar-2024. [Online]. Available: https://www.afar.com/magazine/europes-best-cities.
[3] K. Gyodi and L. Nawaro. "Determinants of Airbnb prices in European cities: A spatial econometrics approach," in Tourism Management, vol. 86, pp. 104319, 2021.
[4] A. Ahuja, A. Lahiri, and A. Das, "Predicting Airbnb rental prices using multiple feature modalities," arXiv.org, 13-Dec-2021. [Online]. Available: https://arxiv.org/abs/2112.06430.
[5] Y. Luo, X. Zhou, and Y. Zhou, "Predicting Airbnb listing price across different cities," 2019. [Online]. Available: https://www.semanticscholar.org/paper/PredictingAirbnb-Listing-Price-Across-Different-LuoZhou/5aea98236bcc318a71f42c91312c5f948717d686.
[6] P. R. Kalehbasti, L. Nikolenko, and H. Rezaei, "Airbnb price prediction using machine learning and sentiment analysis," in Lecture notes in computer science, 2021, pp. 173–184.
[7] P. Choudhary, A. Jain, and R. Baijal, "Unravelling Airbnb predicting price for new listing," arXiv.org, 25-May-2018. [Online]. Available: https://arxiv.org/abs/1805.12101.
[8] W. McKinney, 'Data Structures for Statistical Computing in Python', in Proceedings of the 9th Python in Science Conference, 2010, pp. 56–61.
[9] C. R. Harris et al., 'Array programming with NumPy', Nature, vol. 585, no. 7825, pp. 357–362, Sep. 2020.
[10] Pedregosa, F., et al. "Scikit-learn: Machine Learning in Python," in Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[11] J. Hunter. "Matplotlib: A 2D graphics environment," in Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.
[12] K. Gyodi and Ł. Nawaro, "Determinants of Airbnb prices in European cities: A spatial econometrics approach (Supplementary Material)". Zenodo, Jan. 13, 2021. doi: 10.5281/zenodo.4446043.
[13] A. Hoerl, R. Kennard. "Ridge regression: applications to nonorthogonal problems," in Technometrics, vol. 12, no. 1, pp. 69–82, 1970.
[14] C. M. Bishop, Pattern recognition and machine learning. Springer Verlag, 2006.
[15] L. Breiman. "Random forests," in Machine learning, vol. 45, pp. 5–32, 2001.
[16] J. Friedman. "Greedy function approximation: a gradient boosting machine," in Annals of statistics, pp. 1189–1232, 2001.