Intro to Machine Learning Project Proposal: Airbnb Rental, Crime Rate, and the Neighborhood

Meng Wei, Zhen Yuan, Thomas Zhang

January 2021

1 Introduction

Airbnb is a thriving online rental market place that has a huge impact on both the tourism industry and the daily life of the local neighborhood. We are interested in the relationship between the Airbnb rental rate, criminal rate, and the local amenities. We will study the Chicago's Airbnb market in 2019. Our first goal is to use machine learning to predict the rental prices based on its own features, the criminal rate, and the number and type of amenities around it vicinity. On top of that, if possible, we will also like to see the reverse impact of Airbnb on the local crime rate and numbers of amenities. Xu et al. 2017 [2] and 2019 [3] showed a negative impact of Airbnb on local crime rate . Basuroy et al. 2020 showed an extremely weak increase in restaurant revenue driven by Airbnb [1]. By putting both criminal rate and local amenities on the same page, we hope to combine and compare both the positive and negative economic externalities brought by the Airbnb Industry on the local neighborhood.

As for division of labor, Meng Wei will be responsible for the data collection and cleaning. Zhen Yuan will be running the algorithems. Thomas Zhang will be drafting the proposals, presentation, and final report.

2 Data Set

We have pin-pointed three different data sets, each on Airbnb rental data, criminal history, and the neighborhood distribution of amenities, representatively. We plan to limit the scope of data to the Chicago area in 2018 in order to avoid the Covid-19 impact on the tourism industry, and as well as to keep a manageable workload. The Airbnb data set comes from Inside Airbnb [1], a site sourcing publicly available information from the Airbnb site. The data contains information on each listing and their reviews. Important variables include its longitude and latitude, room features, and price. The criminal history data comes

[1] http://insideairbnb.com/get-the-data.html

from the Chicago Data Portal [2]. This data set contains each crime recorded since 2001 and is published by the City of Chicago and Chicago Police Department. Important variables include the location of each crime and the type of the crime (burglary, murder, etc.). The data set for neighborhood amenities comes from Foursquare's Places API[3], or potentially Google Maps API. Important variables include each store location and their category, such as food, park, museum, and such.

3 Methodology

We will be using multiple supervised algorithms to conduct the task. For a continuous result, we will be using regressions including Ridge, Lasso, Support vector machine regression, and Random Forest regression. For a classification task that yields ranges, we will be using Naive Bayes classifier, logistic regression, Support Vector Machine classification, and Random Forest classification (Yu et al. 2016) [4]. The pricing buckets could be below $100, $100 to $200, $200 to $300, and above $300. To avoid over-fitting, we will apply cross-validation and calculate the feature important scores to select the most important features. Besides, Lasso and Ridge regressions will also contribute to reducing model complexity. Depending on our results, we can compare each algorithm and select the best to present, or include in the report all the results along with proper visualization for the reader to compare.

4 Discussion

After we construct the model, we will like to play with it to extract useful implications. For example, if we can extract the effect of Airbnb on crime rate, we can potentially plug the crime increase back to price prediction model as a mean to measure how much people value the crime impact. We can do the same in evaluating the impact of Airbnb on neighborhood amenities, given there is any substantial impact. We are then able compare the negative and positive externality of Airbnb in terms of consumer valued dollars and determine the direction of the total impact. In this way, we will present a more balance and comprehensive view of how Airbnb affects the local community, not leaning to any side, by accounting for both crime and business bringing.

[2] https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2
[3] https://developer.foursquare.com/docs/places-api/

References

[1] Suman Basuroy, Yongseok Kim, and Davide Proserpio. Estimating the impact of airbnb on the local economy: Evidence from the restaurant industry. *Available at SSRN 3516983*, 2020.

[2] Yu-Hua Xu, Jin-won KIM, and Lori Pennington-Gray. Explore the spatial relationship between airbnb rental and crime. 2017.

[3] Yu-Hua Xu, Lori Pennington-Gray, and Jinwon Kim. The sharing economy: A geographically weighted regression approach to examine crime and the shared lodging sector. *Journal of travel research*, 58(7):1193–1208, 2019.

[4] Hujia Yu and Jiafu Wu. Real estate price prediction with regression and classification. *CS229 (Machine Learning) Final Project Reports*, 2016.