Using Machine Learning for Airbnb Price Prediction

Tuong D. Vu
University of Trento
January 14, 2021

Abstract

Airbnb has been increasingly gaining popularity since 2008 as an economical and convenient way for tourists. Setting the price right significantly contributes to the financial success of Airbnb hosts. This project uses a sample of 50,000 accommodation rental offers in New York City, USA, to predict Airbnb properties' rental price by using several machine learning algorithms. We first perform feature engineering and exploratory data analysis to get insights from the data. We then implement and compare the prediction performance of four different models: multiple linear regression, ridge regression, lasso regression, and xgboost. The results highlight that : (a) XGBoost gives the most satisfying prediction results of rental prices based on mean squared error and R2, (b) Room type, the number of people a listing can accommodate, and location features are the most important features to predict the rental price.

Keywords: predictive modelling,Airbnb, Machine Learning, Regression, Boosting

References

1. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

2. John W Tukey. *Exploratory data analysis*, volume 2. Reading, MA, 1977.

3. Dan Wang and Juan L Nicolau. Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on airbnb. com. *International Journal of Hospitality Management*, 62:120–131, 2017.

4. Inside Airbnb. http://insideairbnb.com/get-the-data.html, 2019. [Online; accessed 04-Dec-2019].

5. Qiang Ye, Rob Law, and Bin Gu. The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1):180–182, 2009.

6. Sherwin Rosen. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55, 1974.

7. Franklin A Graybill. *Theory and application of the linear model*, volume 183. Duxbury press North Scituate, MA, 1976.

8. Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

9. Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

10. Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. CRC press, 2015.

11. Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

12. Frank E Harrell Jr. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.

13. Joel Grus. *Data science from scratch: first principles with python*. O'Reilly Media, 2019.

14. Max Kuhn and Kjell Johnson. *Applied predictive modeling*, volume 26. Springer, 2013.

15. Didrik Nielsen. Tree boosting with xgboost-why does xgboost win" every" machine learning competition? Master's thesis, NTNU, 2016.

16. Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

17. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

18. Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

19. Eyal Ert, Aliza Fleischer, and Nathan Magen. Trust and reputation in the sharing economy: The role of personal photos in airbnb. *Tourism management*, 55:62–73, 2016.

20. Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013.

21. Michael H Kutner, Christopher J Nachtsheim, John Neter, and William Li. *Applied linear statistical models*, volume 5. McGraw-Hill Irwin New York, 2005.

22. Duckworth. How Does Airbnb Price Breakdown Work? https://bnbduck.com/how-does-airbnb-price-breakdown-work/, 2018.

23. Olivia Zaleski. Inside Airbnb's Battle to Stay Private. https://www.bloomberg.com/news/articles/2018-02-06/inside-airbnb-s-battle-to-stay-private, 2018.

24. Craig Smith. 105 Airbnb Statistics and Facts (2020) | By the Numbers. https://expandedramblings.com/index.php/airbnb-statistics/, 2020.

25. Airbnb. Airbnb About Us. https://news.airbnb.com/about-us/, 2020.

26. Airbnb Statistics. https://ipropertymanagement.com/research/airbnb-statistics , 2020.

27. Ben Monty and Mark Skidmore. Hedonic pricing and willingness to pay for bed and breakfast amenities in southeast wisconsin. *Journal of Travel Research*, 42(2):195–199, 2003.

28. Ana Portolan. Impact of the attributes of private tourist accommodation facilities onto prices: A hedonic price approach. *European Journal of Tourism Research*, 6(1):74, 2013.

29. Dominik Gutt and Philipp Herrmann. Sharing means caring? hosts' price reaction to rating visibility. *ECIS 2015 Research-in-Progress Papers*, page 54, 2015.

30. Tapio Ikkala and Airi Lampinen. Defining the price of hospitality: networked hospitality exchange via airbnb. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 173–176, 2014.

31. Jun Li, Antonio Moreno, and Dennis Zhang. Pros vs joes: Agent pricing behavior in the sharing economy. *Ross School of Business Paper*, (1298), 2016.

32. Venoo Kakar, Julisa Franco, Joel Voelz, and Julia Wu. Effects of host race information on airbnb listing prices in san francisco. 2016.

33. Yang Yang, Noah J Mueller, and Robertico R Croes. Market accessibility and hotel prices in the caribbean: The moderating effect of quality-signaling factors. *Tourism Management*, 56:40–51, 2016.

34. Yuan Cai, Yongbo Zhou, Noel Scott, and Jianyu Ma. Price determinants of airbnb listings: evidence from hong kong. *Tourism Analysis*, 24(2):227–242, 2019.

35. Emily Tang and Kunal Sangani. Neighborhood and price prediction for san francisco airbnb listings. http://cs229.stanford.edu/proj2015/236$_report.pdf$, 2015. $[Online]$.

36. Yang Li, Quan Pan, Tao Yang, and Lantian Guo. Reasonable price recommendation on airbnb using multi-scale clustering. In *2016 35th Chinese Control Conference (CCC)*, pages 7038–7041. IEEE, 2016.

37. Pouya Rezazadeh Kalehbasti, Liubov Nikolenko, and Hoormazd Rezaei. Airbnb price prediction using machine learning and sentiment analysis. *arXiv preprint arXiv:1907.12665*, 2019.

38. Yong Chen and Karen Xie. Consumer valuation of airbnb listings: a hedonic pricing approach. *International journal of contemporary hospitality management*, 2017.

39. Chris Gibbs, Daniel Guttentag, Ulrike Gretzel, Lan Yao, and Jym Morton. Use of dynamic pricing strategies by airbnb hosts. *International Journal of Contemporary Hospitality Management*, 30(1):2–20, 2018.

40. Zhihua Zhang, Rachel JC Chen, Lee D Han, and Lu Yang. Key factors affecting the price of airbnb listings: A geographically weighted approach. *Sustainability*, 9(9):1635, 2017.

41. V Raul Perez-Sanchez, Leticia Serrano-Estrada, Pablo Marti, and Raul-Tomas Mora-Garcia. The what, where, and why of airbnb price determinants. *Sustainability*, 10(12):4596, 2018.

42. Beatriz Benitez-Aurioles. Why are flexible booking policies priced negatively? *Tourism Management*, 67:312–325, 2018.

43. Chirag Shah. *A Hands-On Introduction to Data Science*. Cambridge University Press, 2020.