

Will an AirBnB listing receive a 5-star Rating?

Young Jeong

May 8th, 2019

Project Overview

Often AirBnB hosts wonder about making their listings perfect; achieving a high (5*) ratings can open the door for them to become what's called "SuperHost". SuperHosts gain higher search exposures, leading to a boost in reservation counts and income; according to AirBnB, SuperHosts earn around 50% more per month than their non-SuperHost counterparts. Having been an AirBnB host myself, I decided to explore on the relationship between a listing and the rating. Classification modelling was used to predict whether an AirBnB listing received a 5-star rating or not based on its information and what amenities it provided. With this model, I also made a Web app.

Insideairbnb.com is an independent research group that has curated data on AirBnB listings from major cities all over the world. I used their data, along with Walkscore.com's Walkscore and Bikescore data to determine what features attributed to the 5-Star ratings.

I ran a number of Classification models on 44 different features:

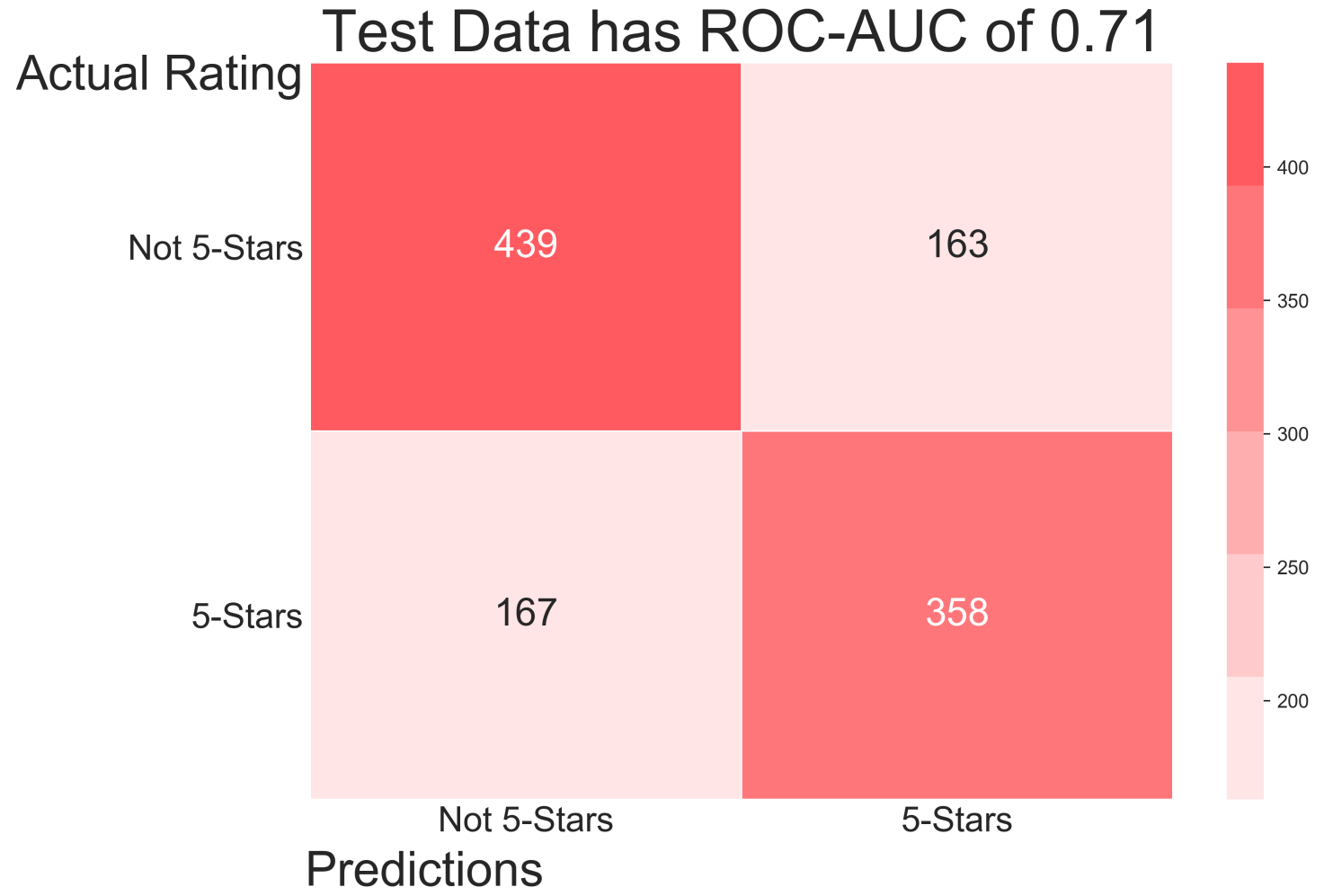
- K-Nearest Neighbors
- Logistic Regression
- Support Vector Machine
- CART Models
 - Decision Trees
 - Random Forest
 - Gradient Boosting
 - Adaptive Boosting
 - Catboost
 - XG Boosting

The metric I used to select the best model was the ROC-AUC score (Receiver Operating Characteristics Area Under the Curve). The best models using the ROC-AUC as the metric were the boosting CART models, with the XGBoost scoring 0.7891, followed by Catboost at 0.7837, Gradient Boosting at 0.7795, and Adaptive Boosting at 0.7757. Even though I could have chosen XGBoost as my model, the difference in metric

performance wasn't enough to make a decision here. Therefore, I also chose to look at time performance (in terms of test predictions).

Comparing the time performance between the 4 Boosting Models, Gradient Boosting performed the fastest at 0.8168 seconds. It was 3 times faster than next fastest model (Adaptive boost at 2.39 seconds), and more than 8 times faster than XGBoost (6.84 seconds). Considering product implementation, I chose the Gradient Boost as the model of choice for its excellent timing performance.

Using the model produced I made predictions on the holdout (test) set. The ROC-AUC score for the test set was 0.71, with the following confusion matrix:



Product: 5-Star Predictor

With the model that I created, I was able to put together a web app using Flask, that lets user enters information about their (potential) listing and spits out the probability of it receiving a 5-star rating and a recommendation on which amenity the user could add that would boost that probability the highest. The product is intended for potential new AirBnB hosts that want to see how their listings will fare in ratings and

what they could do to improve the chances. See Appendix B for the screenshot.

Tools Used

- API: Google Geocoding, Walkscore.com
- Data Storage: Pandas, pickle
- Data Visualization: seaborn, matplotlib
- Data Analysis: Scikit-learn, XGBoost, Catboost, Adaboost
- Cloud Computing: Amazon EC2
- Presentation: Google Slides

Even though I practiced using PostgreSQL, I didn't think the data was big enough to warrant running in a SQL format. To challenge myself with SQL knowledge, I completed a series of SQL challenges provided.

Data

The data was obtained by the various methods from following resources:

1. Insideairbnb.com: Inside Airbnb is an independent research site that visualizes Airbnb listings for major cities around the world. From here, I was able to download csv file that contained listings information from all of Seattle.
2. Walkscore.com: Using Walkscore's API, I extracted walkability and bikeability scores of the listings, to add to my data as the features.
3. Google Geocode API: Since the listings from Inside Airbnb only had lat/long codes, and Walkscore API required physical address, I used the Google Geocode API to extract exact addresses.

The data were accumulated and then combined to create a single dataset of 5632 total listings. The features used were accumulated are listed in the Appendix.

Conclusions & Improvements

With the AUC-ROC score of 0.71, I can definitely make some improvements, both with the model and the web application product. Here are some major ones that I want to make:

1. Description NLP - Some of the features that I ended up not using were some descriptions about the listings that the hosts entered (location, host interaction, description of the room/house itself, etc.) I want to run an NLP and determine what are some of the keywords that lead to a 5-star rating. This could be implemented on both the model and the app.
2. Zillow/Redfin Integration - Because my app is intended for potential AirBnB hosts, I want to integrate it

with Zillow and/or Redfin so that a person who wants to buy a property in the hopes of turning it into an AirBnB income property can predict their 5-star rating probability on the app without having to extract information separately. I see a lot of potential usage if this is implemented.

Appendix A: Features Used

Features	Type	Description
neighborhood_group_cleansed	Str	Neighborhood in Seattle
room_type	Str	Type of room offered (Room or whole house)
bedrooms	Int	Number of bedrooms
bathrooms	Float	Number of bathrooms
beds	Int	Number of beds
price	Int	Nightly price of the listings (in Dollars)
security_deposit	Int	Security Deposit required (in Dollars)
cleaning_fee	Int	Cleaning Fee charged (in Dollars)
extra_people	Int	Charge for having extra people (in Dollars)
minimum_nights	Int	Minimum nights required for a reservation
calendar_updated	Int	how often the host updates the availability calendar
availability_30	Int	How many days are available for reservation in the next 30 Days
availability_90	Int	How many days are available for reservation in the next 90 Days
cancellation_policy	Bool	Is the cancellation policy flexible or not? (True if yes)
instant_bookable	Bool	Is the listing instantly bookable without host's review? (True if yes)
Bikescore	Int	A listing's Bikescore
Walkscore	Int	A listing's Walkscore
24-hour check-in	Bool	Does the listing provide A/C? (1 if yes, 0 if no)

BBQ grill	Bool	Does the listing provide a BBQ Grill? (1 if yes, 0 if no)
Bed linens	Bool	Does the listing provide extra Bed linens? (1 if yes, 0 if no)
Cable TV	Bool	Does the listing provide cable TV? (1 if yes, 0 if no)
Coffee Maker	Bool	Does the listing provide a coffee maker? (1 if yes, 0 if no)
Dishwasher	Bool	Does the listing provide a dishwasher? (1 if yes, 0 if no)
Elevator	Bool	Is there an elevator access for the listing? (1 if yes, 0 if no)
Extra pillows and blankets	Bool	Does the listing provide extra pillows/blankets? (1 if yes, 0 if no)
Family/kid friendly	Bool	Is the listing family and/or kid-friendly? (1 if yes, 0 if no)
Fire extinguisher	Bool	Does the listing provide a fire extinguisher? (1 if yes, 0 if no)
First aid kit	Bool	Does the listing provide a first aid kit? (1 if yes, 0 if no)
Gym	Bool	Does the listing provide gym access? (1 if yes, 0 if no)
Indoor fireplace	Bool	Does the listing provide indoor fireplace? (1 if yes, 0 if no)
Internet	Bool	Does the listing provide internet? (1 if yes, 0 if no)
Keypad	Bool	Does the listing provide a keypad entry access? (1 if yes, 0 if no)
Lock on bedroom door	Bool	Does the listing have locks on its bedroom(s)? (1 if yes, 0 if no)
Lockbox	Bool	Does the listing provide a lockbox for access? (1 if yes, 0 if no)
Long term stays allowed	Bool	Does the listing allow for long term stay? (1 if yes, 0 if no)
Luggage dropoff allowed	Bool	Is luggage dropoff allowed prior to reservation? (1 if yes, 0 if no)
Pack 'n play/travel crib	Bool	Does the listing provide Pack'n Play or travel crib? (1 if yes, 0 if no)
Patio or balcony	Bool	Does the listing have a patio or balcony? (1 if yes, 0 if no)

Pets	Bool	Does the listing allow pets? (1 if yes, 0 if no)
Private entrance	Bool	Does the listing have a private entrance? (1 if yes, 0 if no)
Private living room	Bool	Does the listing have a private living room? (1 if yes, 0 if no)
Refrigerator	Bool	Does the listing come with a refrigerator? (1 if yes, 0 if no)
Safety card	Bool	Does the listing provide a safety card? (1 if yes, 0 if no)
self check-in	Bool	Does the listing have self check-in? (1 if yes, 0 if no)

Appendix B: Flask App Screenshot

Created With:

5_Intepret_Data

This app is created with Flask, with a bit of scikit-learn magic.

Created By:

This app is created by Young Jeong. You can check his other work on [github](#) or connect with him on [LinkedIn](#).

Prediction:

Your probability of getting a 5-star ratings is **66.8%**.

Recommendation:

Add this amenity: **Dishwasher**
This will boost your probability to: **73.5%**

Tell me about your space!

...and I will tell you how likely it will receive a 5-star rating on Airbnb

Where in Seattle?

West Seattle

Bedrooms

1

Price Per Night

Price: \$125

Minimum # of Nights?

Nights: 2

Extra Person Charge

Extra Charge: \$0

Type of Room?

Private Room

Bathrooms

1

Security Deposit

Security Deposit: \$400

Available days in the next 30?

Days: 0

Select All Amenities:

☒ 24-hour check-in

☐ Cable TV

☒ Extra pillows and blankets

☐ Gym

☐ Lock on bedroom door

☐ Pack 'n Play/travel crib

☐ Private living room

☐ Air conditioning

☐ Coffee maker

☐ Family/kid friendly

☐ Indoor fireplace

☐ Lockbox

☒ Patio or balcony

☐ Refrigerator

☒ BBQ grill

☐ Dishwasher

☐ Fire extinguisher

☐ Internet

☐ Long term stays allowed

☐ Pets allowed

☐ Safety card

☒ Bed linens

☐ Elevator

☐ First aid kit

☐ Keypad

☐ Luggage dropoff allowed

☐ Private entrance

☒ Self check-in

Is it Instantly Bookable?

Yes

Beds

1

Cleaning Fee

Cleaning Fee: \$0

Available days in the next 90?

Days: 0

Click for Ratings!