Eun Koo 1003190254
Arjun Mittal 1003144457
Gaoxing Ding 1002975569

Word Count: 794

title???

Introduction

An interesting problem in today's housing market is the disruption caused by the internet for home sharing purposes. As consumers take the market into their own hands, away from licensed professionals, the problem of how to handle listing prices arises.

To address this, our group plans to offer pricing suggestion to new Airbnb hosts. Taking data samples from Seattle, we hope to create a classifier which can recommend which group of price categories a new property should fall into. The classifier will take into account important features of the house including the neighbourhood, amenities, and reviews if currently any exist.

Furthermore, we can extend this to help determine whether markets are currently over/under-valued. As it stands, hosts have autonomy of how to price their houses, which can lead to drastically different prices for the same neighbourhood - making this market quite difficult and time consuming for newcomers to tackle.

Due to the abundant information on how consumers price their houses, this makes the following problem a perfect for artificial intelligence. By taking a well established Airbnb market in Seattle, we hope to extend the pricing patterns found in this area to broader market, to account for prices which are uniform and inline with their true value.

Why machine learning, and not any other AI techniques? (-1)

Source of Data

Obtaining data from Kaggle datasets based on Airbnb data from Seattle [1]:

- Datasets contain both listing data and reviews/calendar data for separate listings

Processing review data:

- Collect available reviews for each listing in dataset
- Run through RNN to create label for each review before FC training with the listings through price categorization

What does this mean? where does the RNN come from? (-1)

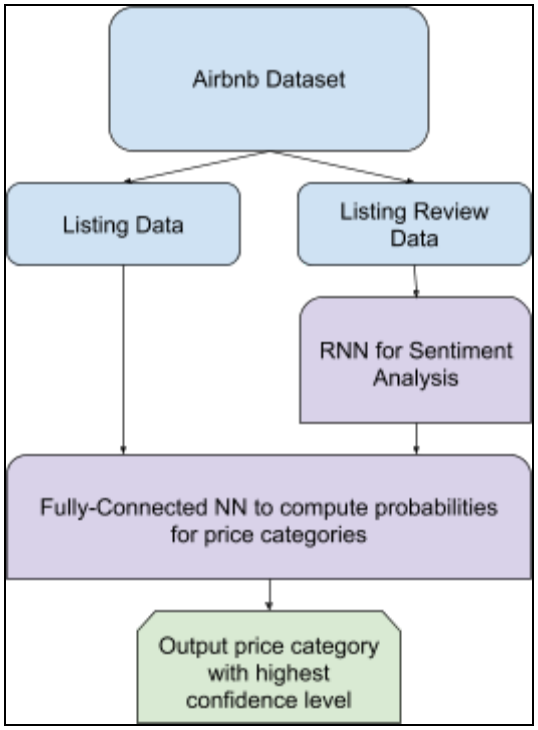
Cleaning listing data:

- Normalize values such as numbers of rooms/bathrooms, square feet, number of guests, location longitude and latitude
- Turn prices into categorical features based on different ranges (e.g. 0-50, 50-100) as labels for listing data

Eun Koo 1003190254
Arjun Mittal 1003144457
Gaoxing Ding 1002975569

Word Count: 794

Overall Structure of your software



1. Airbnb Dataset - grab the dataset and separate listing and associated review data
2. Listing Data, Listing Review Data - clean both datasets through normalization and creating categories for labels based on their price
3. RNN - model processes reviews of each listing and outputs labels to be used as a feature for the listing
4. Fully connected NN - model to take in features and output a price category based on confidence levels/softmax results

Extras: Heatmaps of Seattle/price indicator, Charts to show error on real listings on airbnb.com, Seeing if we can generalize model for different cities (Boston) [2]

"creating categories for labels based on their price" -- what does this mean? Are you creating on

Eun Koo 1003190254
Arjun Mittal 1003144457
Gaoxing Ding 1002975569

Word Count: 794

Plan

List of tasks chronologically planned for project:

Task	Estimated Date of Completion
Data collection and cleaning	March 10
Design, build, and test RNN	March 14
Build Fully Connected Layer NN	March 17
Train the network to achieve certain validation accuracy (hyperparameter tuning)	March 20
Evaluate the test set, create data visualizations/diagrams	March 26
Prepare for the demo and presentation	April 1

Missing deliverable - final report (-0.5)

Risks

1. Our dataset may contain improperly formatted reviews affecting sentiment analysis. There may also be missing, incorrect, or not fully representative data.
    a. A small number of reviews or a lack of any may skew the labels for some listings.
    b. New listings do not have any reviews so it is hard to measure how new listings are performing.
    c. Some reviews will be outdated, leading to inaccurate descriptions for listings that may have been renovated or improved from the past (or worsened).

2. The choice of model (parameters etc.) is a risk. It is possible that the model performs well on data from our validation set, but poorly on data from our test set. This portion will require research on different neural networks used in the past for similar problems - to find optimal parameters for training.

Eun Koo 1003190254
Arjun Mittal 1003144457
Gaoxing Ding 1002975569

Word Count: 794

Things to learn

1. Research on similar problems in the past (property speculation) and how to approach them with neural networks
2. Data cleaning for location values/category-grouping
3. Understand how standard RNN works and how to use it for sentiment analysis of reviews
4. Working with and processing giant datasets in a limited amount of time
5. How to summarize data using descriptive statistics and make the useful visualizations

This is a risk (-0.5)

Ethical Issues

One ethical issue is Airbnb's impact on local communities. Problems arising from the influx of tourists that have resulted from Airbnb's disruption have led to issues such as noise pollution, dissonance between tourists and locals within neighborhoods, and rise in housing costs. Moreover, the lack of safety regulations for fire and environmental safety, zoning and property use which are enforced upon hotels, has been the target of much controversy. [3]

Airbnb's impact on the local housing market has been the source of conflict as well. Investors can buy numerous apartments, driving up prices and manipulating the market to fit their needs. Pricing classifiers use neural networks to detect patterns and replicate them, which is why faulty data can leave the network prone to spreading misinformation, further damaging the market.

References

[1] Airbnb, "Seattle Airbnb Open Data," RSNA Pneumonia Detection Challenge | Kaggle, 26-Jun-2018. [Online]. Available: https://www.kaggle.com/airbnb/seattle. [Accessed: 25-Feb-2019].

[2] Airbnb, "Boston Airbnb Open Data," RSNA Pneumonia Detection Challenge | Kaggle, 17-Nov-2016. [Online]. Available: https://www.kaggle.com/airbnb/boston. [Accessed: 25-Feb-2019].

[3] "Is There An Ethical Problem with Using Airbnb? | The Mediterranean Traveller", *The Mediterranean Traveller*, 2019. [Online]. Available: https://www.themediterraneantraveller.com/airbnb-ethically/. [Accessed: 25- Feb- 2019].