

Project Proposal for DATS 6202 Group 8

The problem we are solving in our final project is to predict first destination for Airbnb users. As one of our group members have used Airbnb when she first came to the U.S., we are wondering to know what country of other people would likely to book as their first destination in Airbnb.

Dataset “Airbnb New Users Booking” is from Kaggle. There are 213451 observations, fourteen features and one target in train dataset. Its sample size is large enough to train a deep network.

For this project, we would like to use Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine these four deep networks. Most of them would be standard form of the network. We would found the reference material in Github and relevant book. And we did not use any common framework for network like keras, tensorflow and caffe.

We will measure the performances of networks by confusion matrix on test data, which split from train, and score in Kaggle of real test data.

Timeline:

07.26 Meet with team and search for datasets in Kaggle.

07.28 Meet with team, search for more data and decide topic

08.01 Project proposal, exploring data

08.03 Data exploration and data preprocessing

08.09-08.09 More exploration, preprocessing and model fitting