Capstone Project 1 Choice

Timothy Baney

June 14th, 2017

Based on the feedback that I have received from my mentor, and on the Springboard community discussion boards, I have decided to go with the Airbnb Kaggle competition as my first capstone project. The problem I want to solve is a classification problem, where based on features of first time users, “What device they use, their age, time of day they go on the website”, what country they are most likely to find a room at, and travel to. The client in this case is Airbnb. By finding the best model that most accurately predicts a first-time user’s destination, Airbnb will use that model to enhance their recommendations to first time user in hopes of not only getting more traffic to their site, but also to get first time users to pick a room and check out more, hopefully increasing their revenue. The data I will use for this project is publicly available on Kaggle.com, <https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings/data>. I will acquire this data by downloading the csv files they provide to you, and reading them into Pandas data frame with the read\_csv method. After wrangling, and cleaning the data, my approach will begin by quantifying any categorical data, and looking for strong correlations between the features of first time users and the target, ‘destination country’. After finding what variables are most strongly correlated, I will use k-nearest-neighbors. I will be doing all my code in an IPython Notebook, and will transfer the important code, and graphs to PowerPoint.