**Airbnb User Bookings**

Synopsis

Under the supervision of: Submitted by:

Dr. D.P Sharma Sir Puranjay Guleria 169108110

Professor Samridhi Kohli 169108125



**Department of Information Technology**

MANIPAL UNIVERSITY JAIPUR

JAIPUR-303007

RAJASTHAN, INDIA

January/2019

**1.Objective of work**

The main objective of this project is to predict where will a new guest book their first travel experience.

New users on Airbnb can book a place to stay in 34,000+ cities across 190+ countries. By accurately predicting where a new user will book their first travel experience, Airbnb can share more personalized content with their community, decrease the average time to first booking, and better forecast demand.

**2.Motivation**

This project helps Airbnb to better predict their demand and take consequent informed decisions. Earlier a new user was overwhelmed with the various choices available for a perfect vacation or stay.

By predicting where a new user will book their first travel experience the company is better able to inform its users by sharing personalized content with their community. It will drastically decrease the time to first booking which will increases the company’s output and help them gain popularity among its user and an edge over their competitors in the market.

**3.Target Specifications if any**

Predicting where a new guest book their first travel experience.

**4.Functional Partitioning of project**

*4.1 Research and gaining knowledge*

Undertaking various courses and familiarizing ourselves with the working process of Data Science problems. Exposure and exploration of the Kaggle website, understanding kernels and datasets. Learning the pre-requisites: programming in Python, Pandas along with the Machine Learning algorithms and data visualization methods.

*4.2 Frequent Discussions and Guidance*

With frequent discussions with our mentor Mr. D.P Sharma Sir along with his guidance in the same will allow us to work in the right direction and take informed decisions.

*4.3Applying the knowledge gained*

After much exposure to this field and gaining the knowledge we will now apply our skills to real life problems and contribute to the society.

**5.Methodology**

*5.1 Understanding the dataset*

Given a list of users along with their demographics, web session records, and some summary statistics. All the users in this dataset are from the USA. There are 12 possible outcomes of the destination country: 'US', 'FR', 'CA', 'GB', 'ES', 'IT', 'PT', 'NL','DE', 'AU', 'NDF' (no destination found), and 'other'. The training and test sets are split by dates.

*5.2 Using the Kaggle platform*

In the test set, we will predict all the new users with first activities after 7/1/2014.In the sessions dataset, the data only dates back to 1/1/2014, while the user’s dataset dates back to 2010. Taking the help of Kaggle platform for testing out datasets as it is not feasible to have a large dataset say of 1TB to be stored in a local machine.

*5.3 Working on the dataset*

Using the dataset and studying various patterns of users’ first booking after signing up with Airbnb from different countries. Next plot out the observed and collected information. We can then apply various Machine Learning algorithms and calculate prediction scores. Finally choose the algorithm with the highest score to recommend users which are from that country the destinations which have been frequently used by travelers belonging to that region.

*5.4 Submitting our work on the Kaggle platform*

The result can now finally be uploaded on the platform and be used by Airbnb to better connect with their users.

**6.Tools required**

*6.1 Kaggle Kernels*

Kaggle is a platform for doing and sharing Data Science. Kaggle Kernels are essentially Jupyter notebooks in the browser that can be run right before your eyes, all free of charge. The processing power for the notebook comes from servers in the cloud, not our local machine allowing us to experience Data Science and Machine Learning without burning through the laptop’s battery and space.

*6.2 Anaconda*

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. Package versions are managed by the package management system conda. Allows us to develop Data Science projects using favorite IDEs, including Jupyter, JupyterLab, Spyder, and RStudio. Analyze data with scalability and performance with Dask, numpy, pandas, and Numba and visualize data with Matplotlib, Bokeh, Datashader, and Holoviews.

*6.3 Machine Learning algorithms*

Machine learning algorithms can be divided into 3 broad categories supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is useful in cases where a property (label) is available for a certain dataset (training set), but is missing and needs to be predicted for other instances. Unsupervised learning is useful in cases where the challenge is to discover implicit relationships in a given unlabeled dataset (items are not pre-assigned). Reinforcement learning falls between these two extremes there is some form of feedback available for each predictive step or action, but no precise label or error message. We will be using these kind of algorithms mainly supervised in order to predict the users’ favorite destination or recommended destinations for the user

*6.4 Python*

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. This language is one of the best language for applying Machine Learning algorithms.

*6.5 Dataset*

Airbnb will be providing us with the dataset, which would contain: Airbnb will be providing us with the dataset, which would contain

* train\_users.csv-the training set of users
* test\_users.csv-the test set of users
* sessions.csv-web sessions log for users
* countries.csv-summary statistics of destination countries in this dataset and their locations
* age\_gender\_bkts.csv-summary statistics of users' age group, gender, country of destination.
* sample\_submission.csv-correct format for submitting our predictions

**7.Work Schedule**

(a) January 2019

Enroll and start the course on Machine Learning using Kaggle. Start recapitulating basics of Python and its various libraries such as numpy, pandas etc.

(b) February 2019

End course and start analyzing the dataset

(c) March 2019

Start coding and implementing various algorithms for the prediction

(d) April 2019

Pick the final algorithm by trail and test and finish coding

(e) May 2017

Appropriate documentation and upload our solution

**8. References**

<https://www.freetutorials.eu/machine-learning-a-z-become-kaggle-master/>

<https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings>