CS506 Project Deliverable 0

Group Member: Jiaqi Zhao, Lingyan Jiang, Ruoqi Shi

Project Introduction:

Journey App, designed by Vibons, is a motivational app for users to easily jot down their daily ideas, pour in their daily emotions and receive educational articles about customized topics through notifications. Nevertheless, as every user can be inundated with so many app notifications, our client wants to find an optimal time to notify the users of the contents and to inspire them to read.

In a nutshell, our goal is to find the send time optimization for each user of the Journey App, based on the dataset given by our client. Send time optimization analyzes when is the most likely time for each recipient to open the notifications. Such an analysis can boost the open rates for our client and enhance engagement with all the users. More information about the "send time optimization" can be found here: https://blog.robly.com/the-science-behind-send-time-optimization/

The dataset is a raw CSV file and is given by our client. The dataset includes the following information: 'User Id', 'User Created At', 'Activation Date', 'Activity Date', 'Name', 'Content Type', 'Journey Name', 'Action', 'Duration', 'Device', 'Channel', 'Session Id' and 'Rating'.

Data source:

http://52.23.136.25/raw_data.csv

Specific questions:

1. What research did the team do about optimizing the sending time of the notifications?
2. During the data cleaning step, how do we fill in the missing data, and how do we reduce the outliers?
3. What data do we choose as the training data and what data do we abandon？
4. What algorithm do we choose to train the data, and what hyper-parameters and cross-validation do we choose?
5. After the testing process, what accuracy do our predictions react to, and how do we adjust the algorithm to meet clients' demanding click rate as more than 40%?
6. During the real users testing, what's the real click rate do we reach by using our algorithm?
7. How to improve our algorithm so that our clients can use it in the future with the increase of application functions and users?
8. After completing this project, what do we still need to improve in the project?

Approach:

Step 1: Clean data, split training, testing, and cross-validation datasets and analyze the data of raw_data.xlsx.
Step 2: List down the parameters which can be deciding factors whether the customer opens the notification or not.
Step 3: Generate a model that predicts optimal send time and maximize the hitting rate of each passage.