Capstone Project Summary

Building a Machine Learning Model to Enhance Business Opportunities

Central European University, Business Analytics MSc

Chun-Hua, Hsu

June 2023

Contents

1 Introduction

The client of my capstone project (the company) is an internet company devoted to assisting users in finding service providers that fit their needs best. The company has extensive website user data collected through Google Analytics 4 (GA4) and aims to create more business opportunities by categorizing website users into meaningful groups. With the patterns of each group, the company can develop customized marketing strategies and offer unique content to website users, boosting conversion.

2 Data Engineering

The raw data is user data from GA4, including behavior (an event), devices, and location information, such as clicking on a link, page loading, device category, and country. Though events and their parameters are stored as structs, I can still transform users' data into practical datasets with SQL commands in BigQuery. I mainly computed the number of times triggered and the average time spent on events by each user.

3 Model Building

The main goal of this project is to create a customer segmentation model with interpretable results and provide possible business recommendations. BigQuery ML supports only one customer segmentation model, the K-Means model. I created multiple models in different clusters, the K value, with assorted hyperparameter tunings and normalized datasets with different selections for event variables. Subsequently, I determined the optimal K value for each model setup and evaluated the performance of the remaining models through the Davies–Bouldin index and the Mean Squared Distance, respectively, which are in-built metrics of BigQuery ML.

4 Model Interpretations and Findings

The best model has four clusters with distinctive characteristics. Users in cluster A seem to enjoy exploring and reading content on the website. Members of cluster B usually enter the main page and stay there without other actions. However, those grouped in cluster C visit the website with a clear purpose and find their matches swiftly after having minimal interactions. This group is the most valuable user. Cluster D users tend to browse all contents and pages and participate in the interactions on the website as much as possible before connecting to the provider. Furthermore, the cluster composition distribution between users who use mobile and those who use desktop is the same.

5 Recommendations

In order to improve the conversion, the company could actively provide users with additional incentives, such as Q&As, questionnaires, surveys, and personalized content depending on the features of the cluster.

Breaking down each event into more elaborate components as variables might be an improvement. Alternatively, Given the limitations of the K-Means model, the company could consider creating different customer segmentation models in another platform.

6 Summary

Enhancing the conversion of website users is one of the significant operational objectives for the company. With data from GA4, I employed query commands and machine learning features in BigQuery to construct multiple customer segmentation models. I evaluated those models through the metrics of BigQuery. Additionally, I made interpretations of the cluster patterns of the best model and provided recommendations from a business perspective. This allows the company to understand its website users better and adopt effective business strategies to achieve its goals.

7 Learning Outcomes

During this project, I gained experience with Google's cloud computing services. With direct access to the most recent and authentic business data, I built, optimized, evaluated, and interpreted an unsupervised machine learning model, which was only briefly covered in the classroom. Through this process, I also learned how to use SQL to handle complex data structures and visualize data properly in Looker Studio.