

Crimson Kim

CP193 Capstone Project Brief

Major in Computational Sciences (Data Science and AI)

25th September 2021

Abstract

For many businesses, customer repurchase rate is one of the keymetrics for their successes. Thus, discovering which customer features -- their behaviors (i.e. which day of the week do they use the service?) and characteristics (i.e. gender, age) -- contribute to the highest repurchase rate, could deliver very useful insights to businesses. For example, if a particular customer feature, say “their location” proved to be very important in affecting their repurchase rate, business leaders can concentrate more of their budgets, resources and efforts to take more advantage of this “customer location” aspect. This paper focuses on discovering user features that highly contribute to their repurchase rate, through fitting user behavior and characteristics dataset into machine learning models.

Summary of Work Completed So Far

I have decided on the “Whys and Hows” of the project, and the remaining work is to actually carry out the plan. The most important part initially, the data collection and processing part, is ready; I have prepared the main data set with around 50,000 rows (each reflecting a unique userID) and 83 columns (each reflecting a characteristic or behavioral data of the users). I have also carried out a business-level analysis of the data set already, over the Summer, and the results have actually been applied to leadership decision-making at my company. I have also created the links necessary for my project, as suggested by the Week 1 activity of CP193.

Plans for the Deadlines

My style of working is not to necessarily plan out the specific tasks I will be doing every week, but to dedicate a sufficient amount of work every week, and make sure the deliverables are met. I am

confident that if I follow through with expectations that are set together with Prof. Watson every week, I will be able to achieve the necessary requirements. Until Fall Week 8, I will focus on finalizing the data preparation, as well as the remaining logistical setups that are needed for the project. I find the detailed plannings recommended as default templates by the Handbook does not work well with me.

Aspirational HCs and LOs

Please reference the HC and LO tab from this [sheet](#)

Capstone Project Brief cover sheet

Name: Crimson Kim	
Project title: Customer Repurchase Rate Predictions	
Abstract: <p>For many businesses, customer repurchase rate is one of the keymetrics for their successes. Thus, discovering which customer features -- their behaviors (i.e. which day of the week do they use the service?) and characteristics (i.e. gender, age) -- contribute to the highest repurchase rate, could deliver very useful insights to businesses. For example, if a particular customer feature, say “their location” proved to be very important in affecting their repurchase rate, business leaders can concentrate more of their budgets, resources and efforts to take more advantage of this “customer location” aspect. This paper focuses on discovering user features that highly contribute to their repurchase rate, through fitting user behavior and characteristics dataset into machine learning models.</p>	
I have read/completed the following and reached out about any questions:	
<u>CK</u> I have read the Capstone Handbook <u>CK</u> I have read the HC Changes for AY2021/2022 <u>CK</u> I have updated my Capstone information in this sheet using this form (by Sept 10) <u>CK</u> I understand the systems/processes that my group will be using	
Links to working files	
Main folder(s)	https://drive.google.com/drive/u/0/folders/1V-uwR9QtucRpxDftNiw4_IGEAmwmq4Do
Paper / writeup (e.g., Google Doc, Overleaf)	https://docs.google.com/document/d/1jf3Sgxe7Q0AMM1jPhS3FV9EsAqxLCz2CDE6Bx7w0xiw/edit
Slides, images	
Data and analysis	Waiting for final confirmation by my company for data share policy

GitHub	https://github.com/crimson-kim/capstone
Group's project LO rubrics sheet [guidelines for custom LOs]	
My HC & LO plans, applications, annotations [template]	https://docs.google.com/spreadsheets/d/1ouLFmNeJXh23pai3mSMEecA1DLCpHkiwRg7_K-cNQoE/edit#gid=1427465626
Planning & progress tracking workspace (Notion, etc.)	https://www.notion.so/Crimson-Kim-Capstone-Project-bded7adbc56245fba0efe080a68c57b5
Other <i>(please specify and add, edit, or delete rows according to your project)</i> [Tools and systems for project management]	

In the time since my proposal was approved...

Summary of new progress and outcomes [weekly goals & accountability template]	<ul style="list-style-type: none"> I have processed and polished my data set more
Brief reflection on my progress and process so far	<ul style="list-style-type: none"> I have already carried out the initial version of the project, at my work, so I feel confident on how I can improve on the project and make it more academically rigorous. I need final confirmation from the Vice President, on the data share policy.

For this submission...

Request for feedback on these specific areas	<ul style="list-style-type: none">
Potential faculty for my Capstone committee (4-10)	Brief justification → Refer to the faculty expertise and interests sheet
Prof. Watson	Advisor, General Project Management

Prof. Diamond	Insights on machine learning and data science
Prof. Sheth	Insights on the application of the technology to businesses
Prof. Sterne	Insights into the algorithm developed in my system