Data Science Institute of Australia

Capstone Project - Maximizing Customer Retention at Instacart

Start Date: 08/06/2019

Epic	Predicting Customer Repeat Purchases
Project owner	Mai Anh Ly
Project status	Development
Developers	Mai Anh Ly
Client	Ravi Gupta (Instacart CFO & COO) David Hahn (Instacart Chief Product Officer)
Stakeholders	Instacart UX lead Instacart web/mobile development lead Data warehousing/engineering lead End users Retail partners
Design complete	23/06/2019
Code complete	25/07/2019
Testing complete	30/07/2019
Release date	15/08/2019

Background

Instacart is a grocery delivery service that connects customers to grocers via a web/mobile app. Customers cite convenience as the primary reason for preferring Instacart vs. in-person shopping. Repeat customers also tend to repurchase products from their ordering history rather than via product search. Instacart will retain more customers and increase revenue over their competitors by capitalizing on these behaviours.

We can improve the Instacart shopping experience by predicting items a customer will repurchase and implementing a 'repurchase items' page on the app. The Instacart service will be more convenient to use, and encourage customers to make shopping with Instacart part of their routine.

Objectives

Objective	Metric
Predict products customers will repurchase	Accuracy of predictions is more than 75%
Improve customer experience with app/site	Majority of customers use new 'repurchase items' page
Improve customer retention	Monthly customer retention increases

Assumptions

Instacart business model:

- Instacart takes 20% of order value as revenue
- Month-to-month customer base is at least 200,000

Instacart customers:

- An average customer makes 3 orders in a month with each order worth $100
- Majority of customers access service from the app
- Client has addressed customer and retail partner privacy concerns

Technical constraints and considerations:

- Instacart stores purchase data of all customers who have made at least 1 order
- Products have been ordered by multiple users at multiple retailers
- Modelling is performed on customers who have made at least 3 orders
- Each order, customer and product in Instacart's database has a unique ID number
- Coding, testing and deployment can be done in the cloud

User Stories and Requirements

Ref #	Name	User Story	Functional Requirements
1	Candidate products on single page	As a customer, I want to easily repurchase products without wasting my time navigating through multiple pages and choices	- Product carousel - Option to remove products - 'Add all to cart' button
2	Record number of correct predictions	As a retail partner, I want to see which of my products are popular with customers so I know where to maximise profit	- Record result of each prediction in database - Ability to aggregate value of correct predictions by product and retailer
3	Record end user clicks and time spent on feature	As a UX designer, I want to know how customers will respond to the 'repurchase item' feature so I can judge its effectiveness	- Elements associated with feature logs clickstream data - Site/app logs time spent on feature page
4	Compatible with existing back-end infrastructure	As a data engineer, I need to know the volume of predictions the model will make per user so I can manage the infrastructure between the server and the front-end app/site	- All code made available in company repository
5	Compatible with existing front-end infrastructure	As a web/mobile developer, I need to know what content is supplied by the new feature so I can code the app/site layout appropriately	

Open Questions

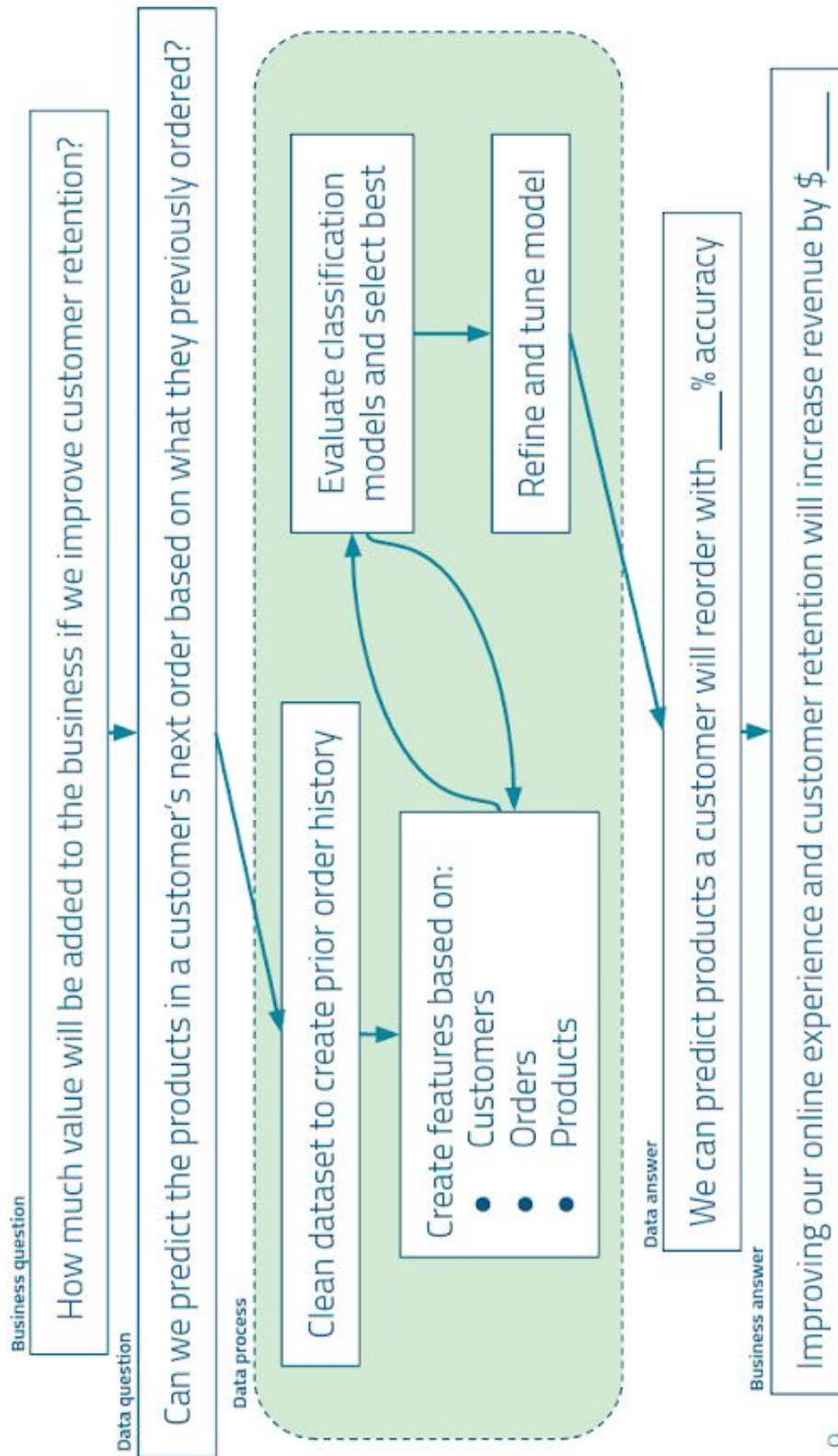| Question	Answer	Date Answered
How does Instacart differ from online groceries?	Instacart operates a web/mobile app where customers order from Instacart retail partners, which are then fulfilled and delivered by Instacart personal shoppers. Retail partners handle all inventory management.	25/06/2019
How was retailer and user privacy and confidentiality dealt with?	No retailer data was provided, user ID numbers have been anonymised and no personal information was released	23/06/2019
How do we determine if a predictive model has an effect on customer retention?	Conduct an A/B test where model has been implemented in one case and measure customer retention/churn over time	23/06/2019

Out of Scope

- UX design of "repurchase items" feature
- Creation of server infrastructure linking model to Instacart webpage/app
- Creation of dashboard for monitoring model data stream
- A/B testing of "repurchase items" feature

Risks

Risks	Likelihood	Impact	Migation
Memory limitations due to processing a very large production dataset	High	High	Test code using free Google Compute Engine runtime via Colab notebook Use low-bit datatypes to conserve memory
Customer will churn if feature orders products they do not want	Medium	Medium	Refine model to have recall score of 80% or higher Limit predictions to high probability of repurchase Include tutorial elements on first use of feature Add ability to remove items

User interaction and design

Business question

How much value will be added to the business if we improve customer retention?

Data question

Can we predict the products in a customer's next order based on what they previously ordered?

Data process

Clean dataset to create prior order history

Create features based on:
- Customers
- Orders
- Products

Evaluate classification models and select best

Refine and tune model

Data answer

We can predict products a customer will reorder with ____ % accuracy

Business answer

Improving our online experience and customer retention will increase revenue by $____

Figure 1. Project flowchart

Tasks

Task	Name	Due date	% Time allocated
Acquire data	Mai Anh Ly	25/06/2019	5
Clean data	Mai Anh Ly	27/06/2019	10
Create features for predicting repurchases	Mai Anh Ly	18/07/2019	30
Evaluate accuracy of classification models using all features/predictors	Mai Anh Ly	21/07/2019	15
Tune hyperparameters of best evaluated model	Mai Anh Ly	26/07/2019	20
Perform feature selection of tuned model	Mai Anh Ly	27/07/2019	10
Use tuned model to make final repurchasing predictions	Mai Anh Ly	28/07/2019	5
Complete documentation	Mai Anh Ly	31/07/2019	5

Results and Conclusions

We have built a model that predicts customer repurchases with 79% accuracy. This is a 4% improvement on the project objective for predictive accuracy. The model will also be correct in 81% of cases where it predicts an item repurchase. This will decrease the risk of the customer being shown items they do not want, or are not likely to repurchase.

We have yet to determine if the customer repurchase predictive model will improve the Instacart online experience. Subsequently, we do not yet know if the model will improve customer retention. However, we are confident the end-user model interface will add to the convenient shopping experience valued by Instacart customers. We estimate under the assumptions outlined in this document that a customer retention increase of 1% maintained over a year will deliver $1.4 million in revenue. Therefore, a strategy that will increase customer retention is worth pursuing.

Key insights

Our analysis from data cleaning and model development offers the following insights into customer preferences and behaviour:
- Customers tend to place more items they have purchased before in their cart the more recent the orders were made
- The top 10 repurchased items are all from the fresh fruits and vegetables aisle
- Six of the top 10 repurchased items contain the word "organic" in their name
- Item repurchases are strongly influenced by the number of purchases a customer has made from the same aisle as the target item

Next steps & recommendations

1. Work with data engineers to deploy model and determine the frequency of live data input into model pipeline.
2. Work with UX and web/mobile development teams to develop "repurchase items" feature based on model predictions.
3. Conduct an A/B test and measure customer churn from month-to-month.
4. Based on A/B test results, make a decision on implementing feature site-wide.

Supplementary recommendations

1. As any changes to an app/site interface risks customer churn, we suggest holding a social media marketing campaign publicising an upcoming "repurchase items" feature. This may encourage customers to feel positively inclined towards the feature.
2. Our analysis of customer orders revealed that organic produce are the most popular repurchases among the customer base. We suggest working with retail partners to capitalize on this preference.