

## **Project 3: Formal Proposal**

### **Problem Statement**

Instacart is a grocery ordering and delivery app that aims to make it easy to fill your refrigerator and pantry with your personal favorites when you need them. The Instacart app allows you to browse thousands of products from your favorite stores, from groceries and alcohol to home essentials and more. After selecting the products you want, personal shoppers will pick up those products for you as well as deliver them to you. Kaggle has challenged the data science community to use anonymized data on customer orders over time to predict which previously purchased products will be in a user's next order.

### **Objective**

The goal of this project is to develop a model capable of predicting whether an order is reordered or not. The obvious client here is Instacart, as it will be the one profiting from the predictive model. The users of Instacart will also benefit from the project, as the app would be more skilled at recommending products to users. Therefore, this is a classification problem, but it is also a recommendation system.

### **Exploratory Data Analysis:**

Analyze the orders file, conduct analysis on product and orders, and any other relevant files in the dataset.

### **Feature Engineering**

Calculate and state the relevant user features (such as reorder ratio, total number of orders of each user), product features, and order features.

### **Preprocessing**

Split data into training and testing set.

### **Modeling and Parameters**

The goal of this work was to construct a model capable of predicting whether or not a product would be reordered. Therefore, this is a classification problem. I used the following classification models for prediction:

- Logistic Regression
- Gradient Boosting
- Random Forest

### **Model Selection Metrics:**

The optimal model was selected based on both primary and secondary performance metrics. The primary performance metric was the model's f1 score, which is calculated from the precision and recall.

**Conclusion and future work:**

Draw any appropriate conclusions based on findings, and see if its possible to develop a model capable of predicting whether an order is ordered. After construction of the models, determine which model is the best based on selection criteria above.

State any potential ideas for future projects based on the findings.