

# PREDICTING INSTACART'S CUSTOMERS BEHAVIORS

BY: MANAL ALQHATANI  
NADIA ALGHAMDI

## ABSTRACT

Instacart is an American company provides grocery shopping service where users can order their groceries through an app, and a freelance or aka “shopper” takes responsibility for fulfilling user orders. Thus, pretending that Instacart is our client who wants to improve its business; we are using machine learning classification algorithm to help them improve their business by do so.

Our objective/ business need: is to predict whether previous purchased will be reorder again. Machine learning classification will help us doing so.

## DESIGN

We fitted 4 classifiers models (KNN, Logistic Regression, Random Forest and Decision Tree) without sampling and again with some sampling techniques like (SMOTE & Random oversampling)

## DATA

The dataset we used obtained from [Keggle](https://www.kaggle.com/datasets/instacart/instacart-open-transaction-event-train) website; it contains 6 csv files having relational set of files describing user's orders over time

## ANALYSIS STEPS

- 1- We joined all related tables and create data frame that includes the target-label feature (reordered).
- 2- Applied EDA which gave us a whole picture of data; from it we knew that Banana is the most ordered item, the most popular department is produce, and Sunday is the busiest day in terms of number of orders.
- 3-Since the dataset is huge and consumed a lot of time and memory, we chose 1% of the data to be our sample which left us with (317410, 11)

4- Before fitting the model, we checked the imbalanced of the target feature. The distribution of target variable shows that nearly 63% of observations have class 1 and about 37% having class 0.

5- Applying some sampling techniques: we used SMOTE and Random oversampling. Both resample data equally.

7- We fitted classifiers also without any sampling methods.

## CONCLUSION AND SUMMARY

The result from all the 12 experiments we did, random forest with Random oversampling technique outperforms and predicted whether customers will have the same order in their next purchase based on their previous purchases with around 71% accuracy.

Recommendation: doing more experiments on our data might give us more accurate prediction!

### SMOTE

Model	Precision	Recall	F1	Accuracy
KNN	71.252	62.335	66.495	60.567
Logistic Regression	77.730	56.009	65.106	62.311
Descisiononn Tree	69.580	68.543	69.057	61.441
Random Forest	74.166	81.315	77.576	70.490

### Random Oversampling

Model	Precision	Recall	F1	Accuracy
KNN	70.719	63.203	66.750	60.472
Logistic Regression	77.637	56.329	65.288	62.399
Descisiononn Tree	69.633	69.633	69.633	61.874
Random Forest	74.679	80.275	77.376	70.531

### Without Sampling :

Model	Precision	Recall	F1	Accuracy
KNN	68.849	78.565	73.386	64.229
Logistic Regression	67.958	90.638	77.676	67.295
Descisiononn Tree	69.982	69.011	69.493	61.964
Random Forest	73.305	84.413	78.468	70.918

