TEAM 11 - INSTACART-MARKET-BASKET-ANALYSIS

**Problem**: The goal of the competition is to use this anonymized data on customer orders over time to predict which previously purchased products will be in a user’s next order.

* Decided to predict reorder.

**Team :**

* Ashwin Kumar
* Md. Fazil Chembra

**1. Describe your pipeline.**

* Read all provided .CSV file.
* Join all files to get data details.
* Identified independent factors
* Removed Rows with NaN values.
* Added 0 to days since prior order.
* Identified Reorder as dependent variable.
* Created Train Set and Test Set.
* Using RandomForestClassifier Accuracy Score: 70.22067939499132.
* Using DecisionTreeClassifier Accuracy Score: 64.30283494503678.
* Using DecisionTreeClassifier Accuracy Score: 66.22034878915612

**2. Potential shortcomings with your current pipeline**

* Data size was too high.
* Online solutions mention algorithm like XGBoost, LightGBM etc

**3. Possible improvements to your pipeline**