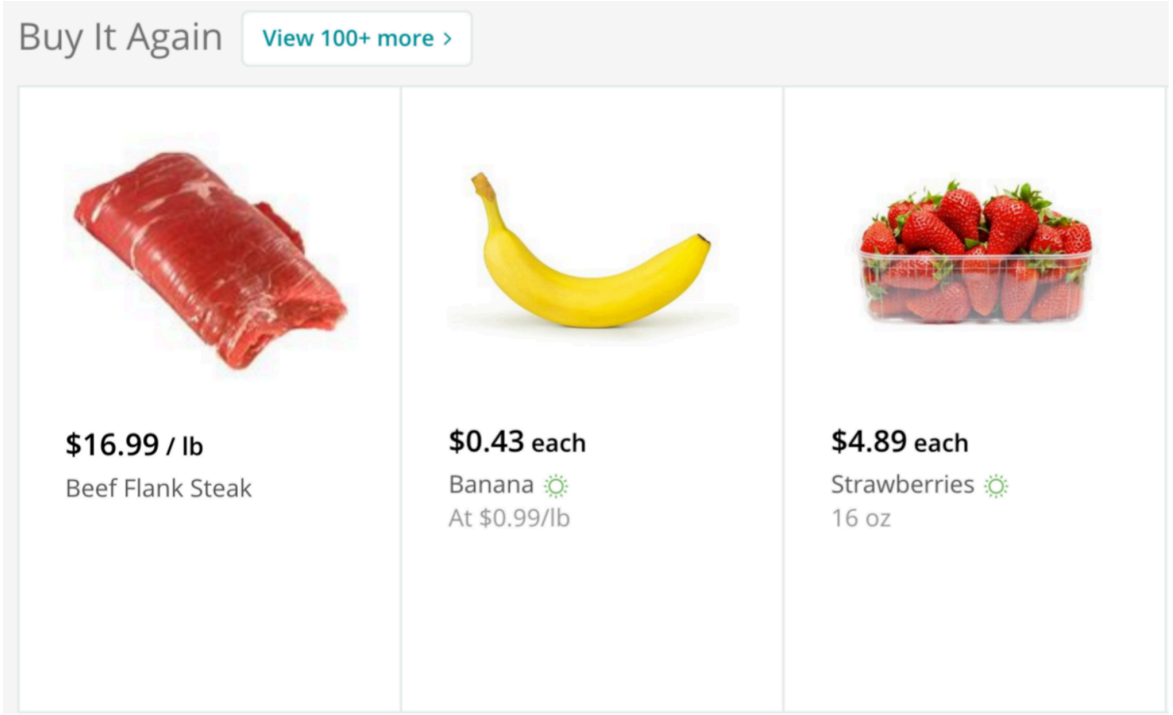
Instacart Reorder Predictions

Predicting which products a user will reorder

Intro

- **Data**: 8 million+ product orders
- **Goal**: Implement a classification model to predict product reorders
- **Uses**: Recommendation system



kaggle

Approach

Work with a subsample of data to **quickly** and **iteratively**:

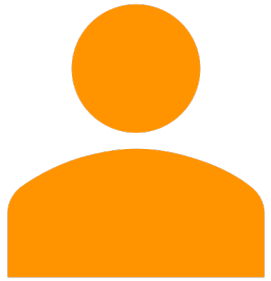
- Engineer features
- Train models
- Compare models

...Then **scale** up!

Features

Engineered a **variety** of features:

- Focus levels:
  - User
  - Product
  - User-Product
- Recency:
  - Recent behavior

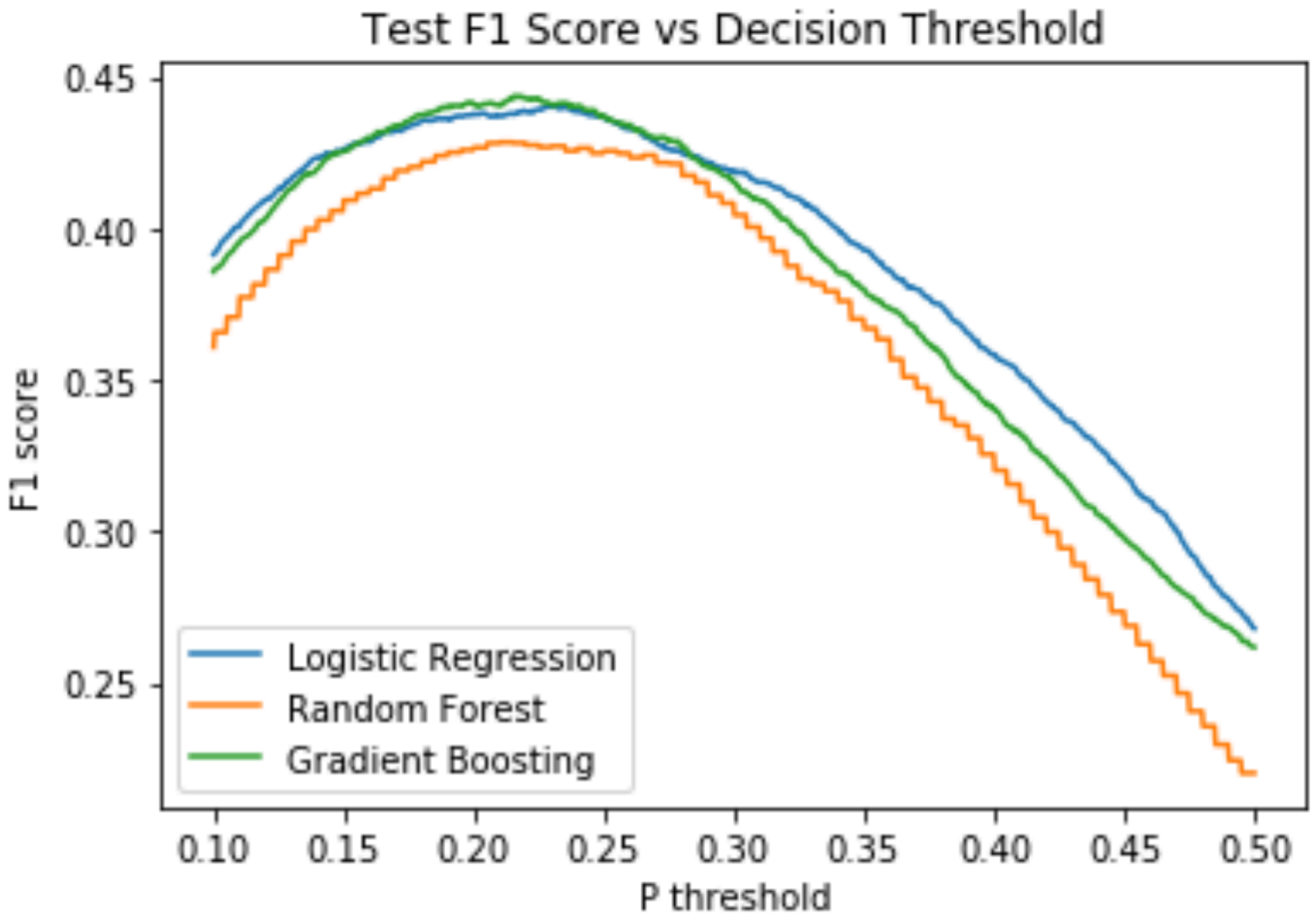


Average # of days since **user** made prior order

Probability of a **product** appearing in an order

Did **user** purchase the **product** in at least one of his or her previous three orders?

Model Selection



Precision and Recall matter, **optimize F1**!

All **three models** performed comparatively well

Gradient Boosting selected with best F1 test score

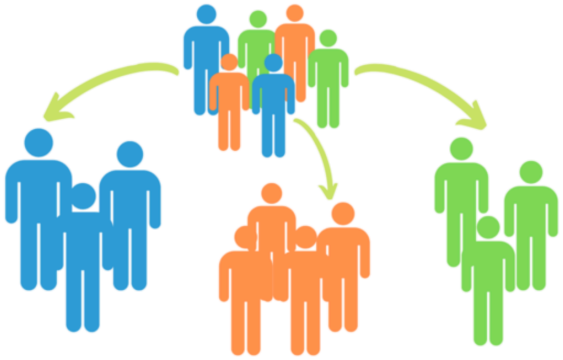
Model Tuning

- Max Depth = 8
- Row Sampling = 0.80
- Column Sampling = 0.70
- Learning Rate = 0.50
- Threshold = 0.220

Results

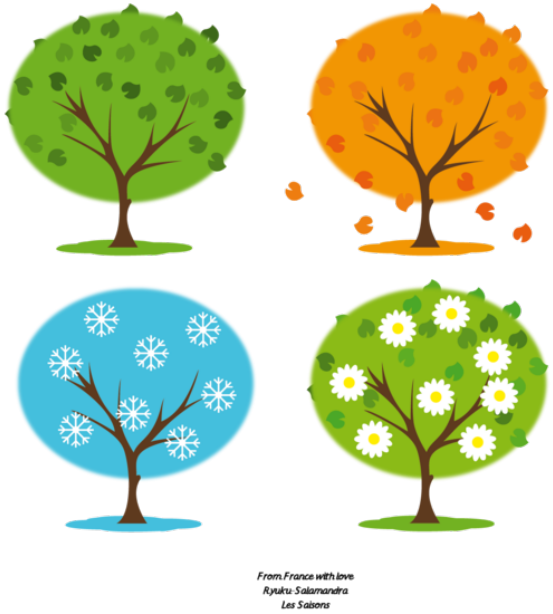
Metric	Test Score
Accuracy	0.871
Precision	0.393
Recall	0.499
F1	**0.440**

Future Improvements



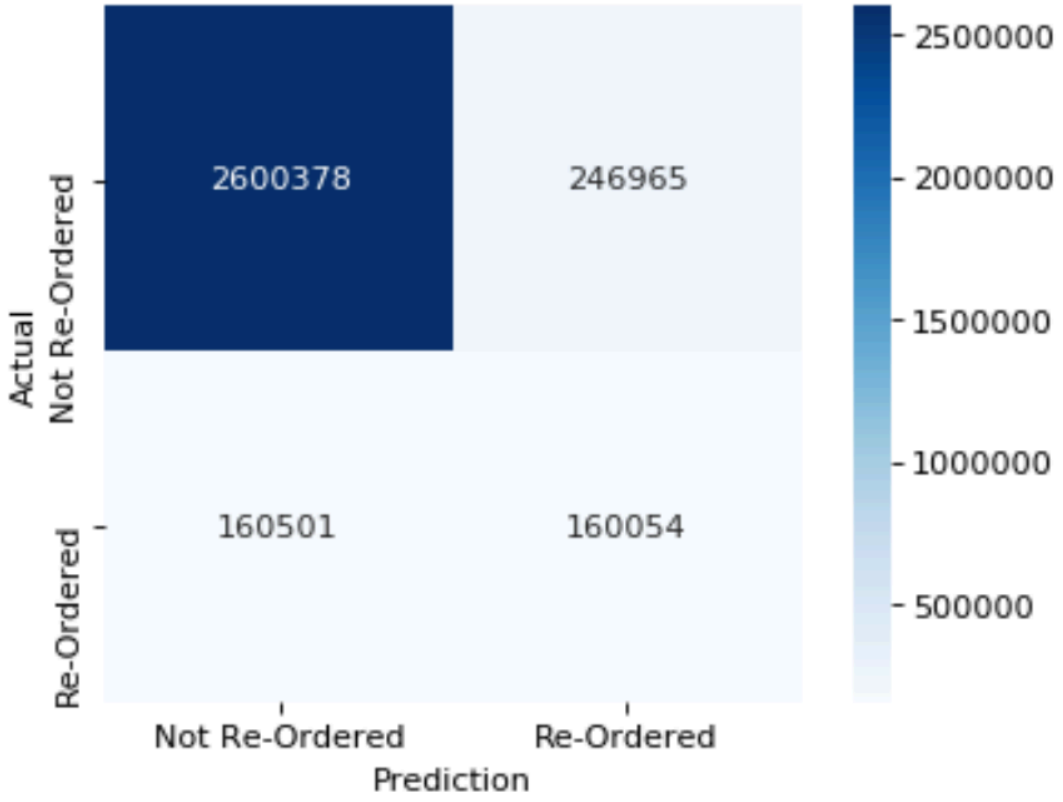
More and more features!

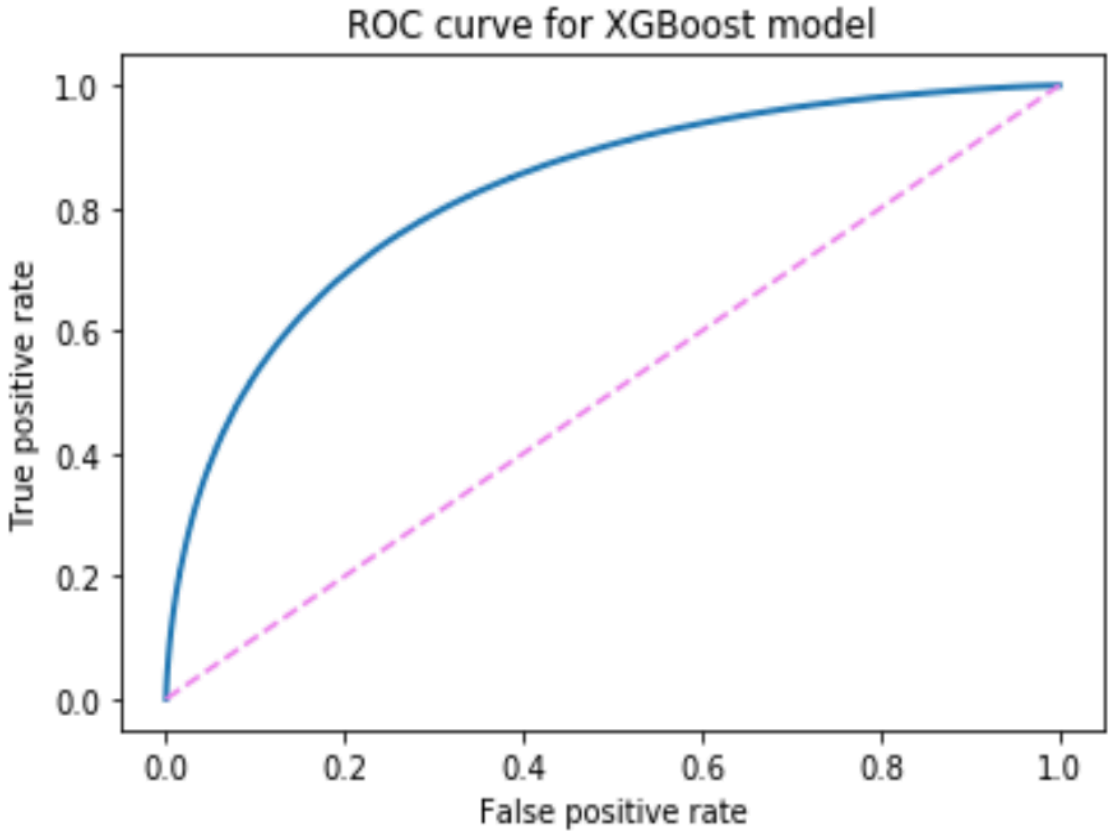
- Greater focus on recency of product orders
- Segmentation of users
- Seasonality of products and user behavior
- Organic or gluten-free items



Thank you!

Appendix


2600378
246965
160501
160054
Actual
Not Re-Ordered
Re-Ordered
Not Re-Ordered
Re-Ordered
Prediction
2500000
2000000
1500000
1000000
500000


ROC curve for XGBoost model
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

Credits

- *"The Instacart Online Grocery Shopping Dataset 2017", Accessed from https://www.instacart.com/datasets/grocery-shopping-2017 on 5/06/20*