instacart

Order Predictions

Lukas Wadya
May 8, 2019

__35%__ of what consumers purchase on Amazon and __75%__ of what they watch on Netflix come from __product recommendations__

McKinsey & Company

Let's Predict!

- Help users quickly reorder their favorite products
- Give users a nudge when they might be forgetting something they are likely to want
- Potentially increase sales
- Leverage the data we are already collecting

Start Small...

Work locally with subsets of data to quickly and iteratively...

Engineer Features → Train Models → Compare Models

...Then Go Big

Scale up after choosing models and setting hyper-parameters

SQL Database

Train on AWS EC2

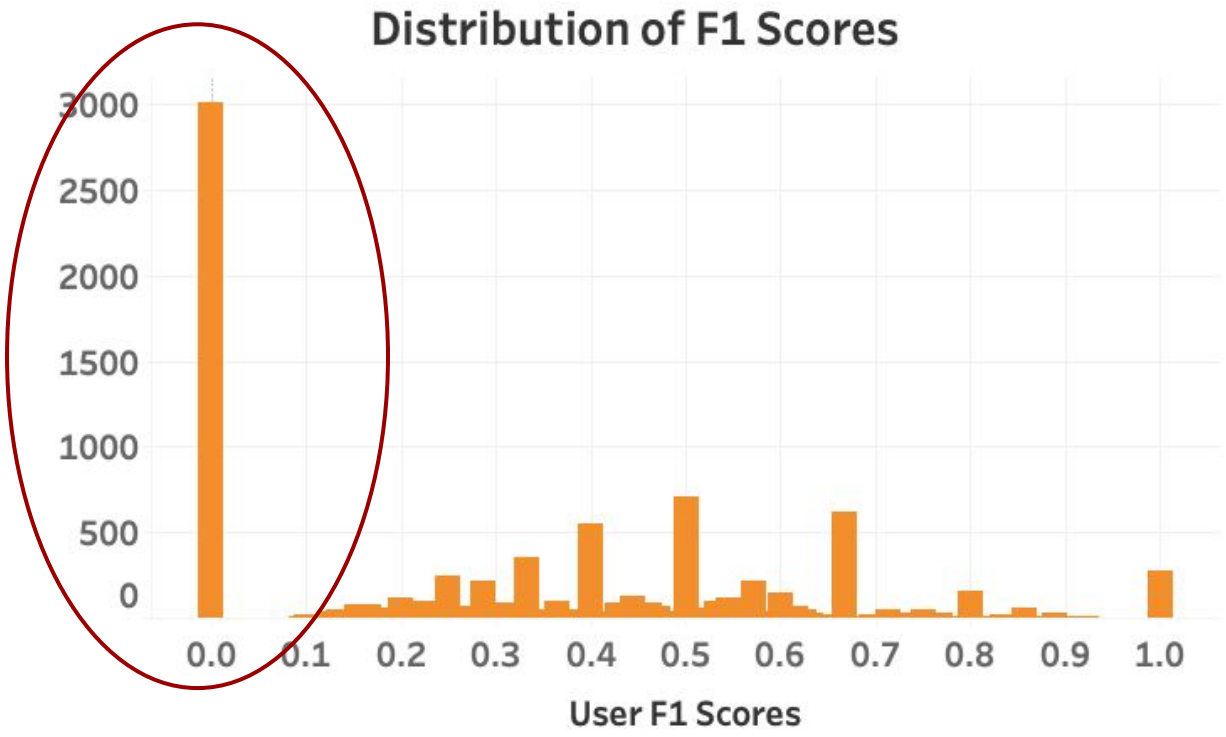
Finalize Locally

Model Scores

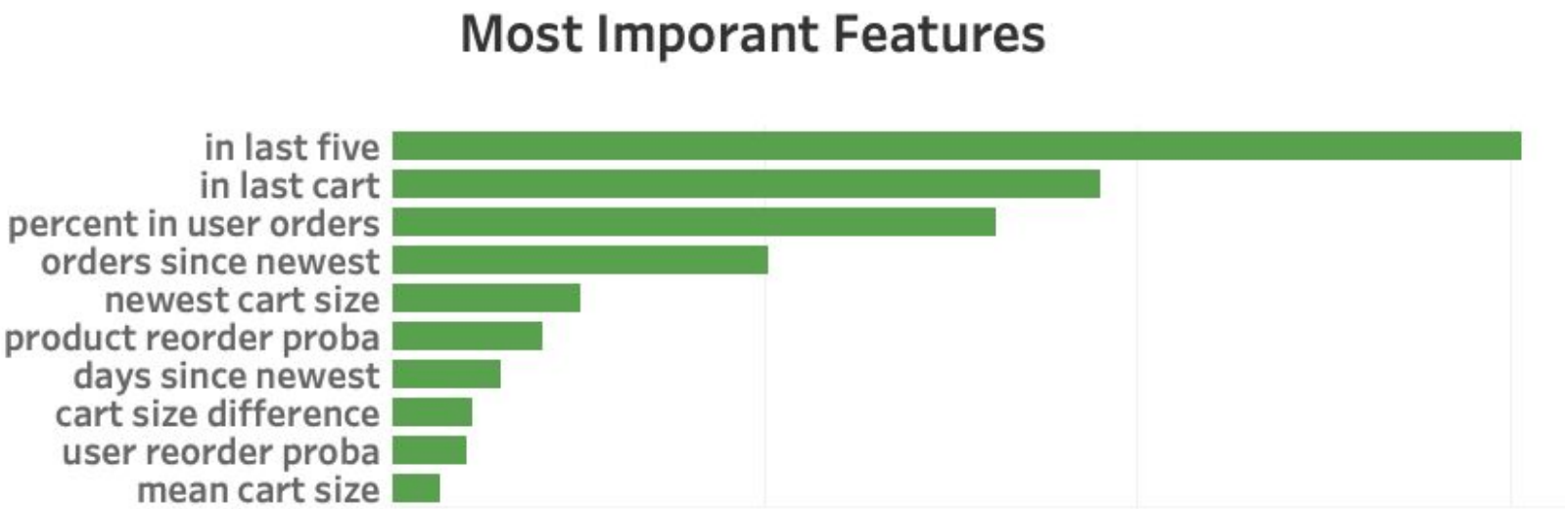
F1 Scores	Logistic Regression	**Gradient Boosting**
Aggregate Test Set	0.446	**0.459**
Average Per User	0.365	**0.367**

Sample Prediction

Actual Order	Predicted Order	Prediction Probability
Frozen Whole Strawberries	Frozen Whole Strawberries	0.642
Sparkling Water	Sparkling Water	0.206
Mixed Fruit Fruit Snacks	Bag of Organic Bananas	0.125
	Seedless Cucumbers	0.101
	Mixed Fruit Fruit Snacks	0.074


Distribution of F1 Scores
3000
2500
2000
1500
1000
500
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
User F1 Scores

Features


Most Imporant Features
in last five
in last cart
percent in user orders
orders since newest
newest cart size
product reorder proba
days since newest
cart size difference
user reorder proba
mean cart size

Conclusions

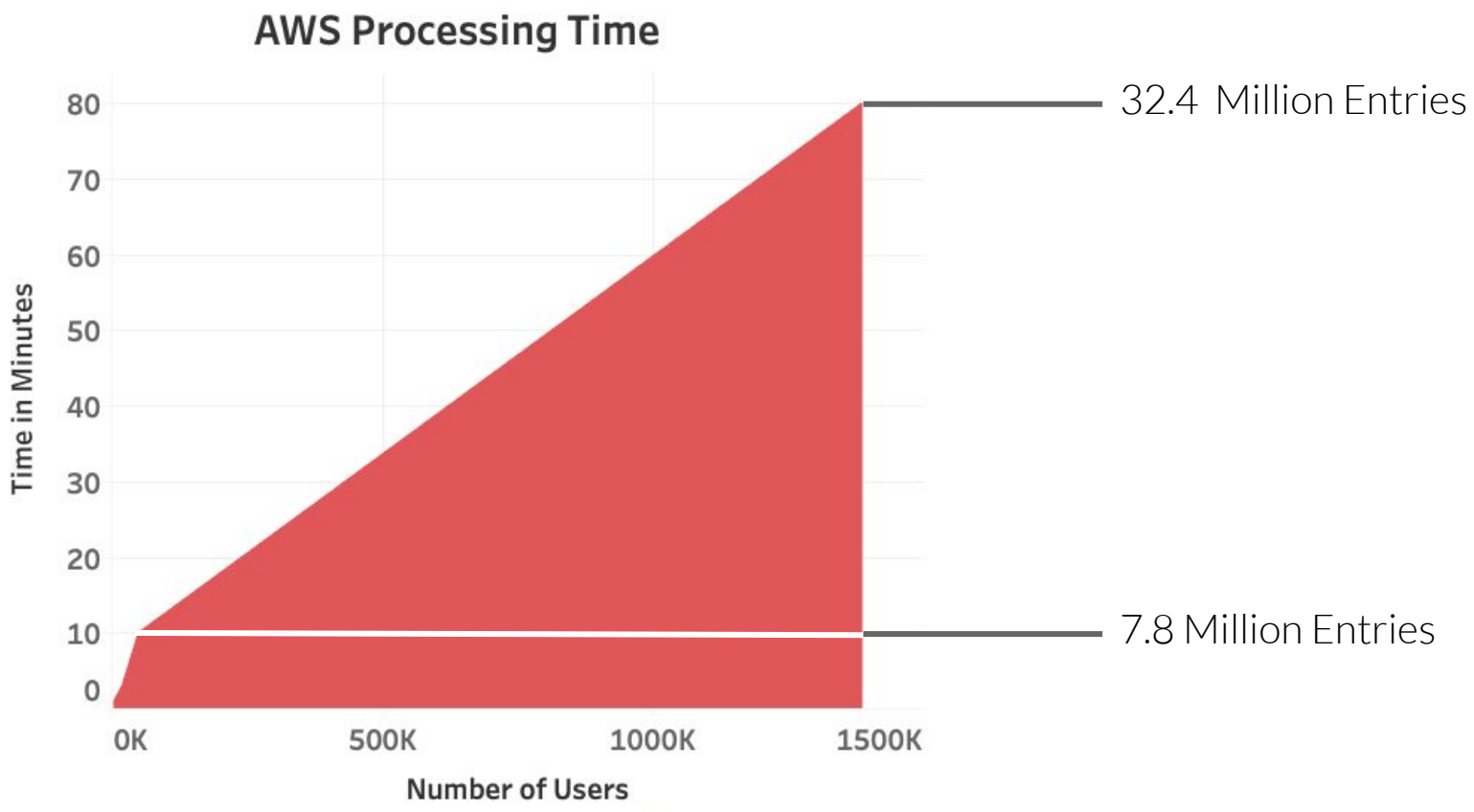
- Gradient Boosting Model is likely the best choice
- For a very simple model, focus on just the most recent orders
- Adding more data improved all models, but particularly the Gradient Boosting Model

Future Improvements

- Features, features, and more features!
- Consider the effects of close substitutes
- Adjust classification threshold on a per-user basis
- Improve project AWS pipeline

Thank You!

Appendix


AWS Processing Time
Time in Minutes
80
70
60
50
40
30
20
10
0
0K
500K
1000K
1500K
Number of Users
32.4 Million Entries
7.8 Million Entries