

Predictive Shopping

with  instacart

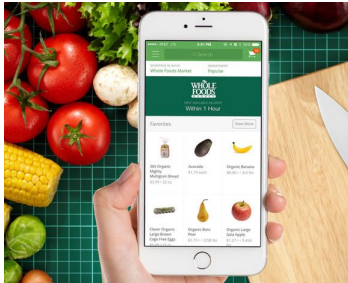
Stella Moretti



The Business Case for Predictive Commerce

For consumers

Quicker, smoother shopping experiences



For retailers

Personalized marketing strategies → conversion





instacart

Groceries Delivered From Local Stores

Instacart Best Sellers



\$0.33 each

Banana 🌱
At \$0.539/lb



\$2.49 each

Organic Bananas
3 lbs



\$5.99 each

Organic Strawberries 🌱
16 oz



\$6.99 each

365 Organic Baby Spinach
16 oz



\$2.99 each

Organic Hass Avocado 🌱



\$2.79 each

Organic Green Avocados
🌱



\$0.59 each

Lemon 🌱



\$3.99 each

Strawberries 🌱
16 oz



\$0.69 each

Lime 🌱



\$5.49

Organic Valley Whole Milk
64 fl oz

Top 25 Late Night Purchases



\$1.25 ~~\$1.49~~
\$0.24 off
 Edy's/Dreyer's Cookie
 Dough Grand Ice Cream
 5.8 fl oz



\$1.25 ~~\$1.99~~
\$0.74 off
 Häagen-Dazs Dulce de
 Leche Ice Cream
 3.6 fl oz



\$1.25 ~~\$1.99~~
\$0.74 off
 Häagen-Dazs Caramel
 Cone Ice Cream
 3.6 fl oz



\$1.25 ~~\$1.99~~
\$0.74 off
 Häagen-Dazs Chocolate
 Ice Cream Cup
 3.6 fl oz



\$6.49
 Edy's/Dreyer's Vanilla
 Bean Grand Ice Cream
 1.5 qt



\$4.99 ~~\$5.99~~
\$1.00 off
 Häagen-Dazs Caramel
 Cone Ice Cream
 1 pt



\$6.49
 Edy's/Dreyer's Strawberry
 Grand Ice Cream
 1.5 qt



\$6.49
 Edy's/Dreyer's Butter
 Pecan Blast Grand Ice
 Cream
 1.5 qt



\$6.49
 Edy's/Dreyer's Coffee
 Grand Ice Cream
 1.5 qt



\$6.49
 Edy's/Dreyer's Caramel
 Delight Grand Ice Cream
 1.5



\$6.49
 Edy's/Dreyer's Spumoni
 Ice Cream
 1.5 qt



\$5.99
 Ben & Jerry's Truffle
 Ice Cream
 1 pt



\$5.99
 Snickers Ice Cream Bars
 6 ct



\$6.29
 Friendly's Vanilla Ice
 Cream
 1.5 qt



\$3.99 ~~\$6.49~~
\$2.50 off
 Breyers Vanilla Ice Cream
 48 oz



\$6.49
 Edy's/Dreyer's Espresso
 Chip Grand Ice Cream
 1.5 qt



\$5.99
 365 Vanilla Ice Cream
 1.75 qt



\$4.99
 Ben & Jerry's Pumpkin
 Cheesecake Ice Cream
 1 pt



\$7.69
 Van Leeuwen Ice Cream
 Ice Cream Mint Chip
 16 oz



\$7.69
 Van Leeuwen Strawberry
 Ice Cream
 1 pt



\$10.99
 Van Leeuwen Ice Cream
 Ice Cream Coffee Crunch
 16 oz



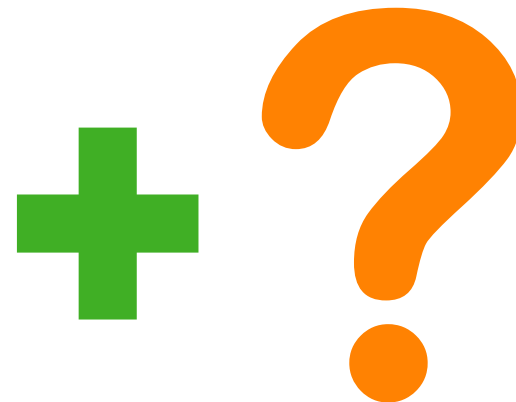
\$6.99
 Van Leeuwen Artisan
 Cookies & Cream Ice
 Cream
 1 pt



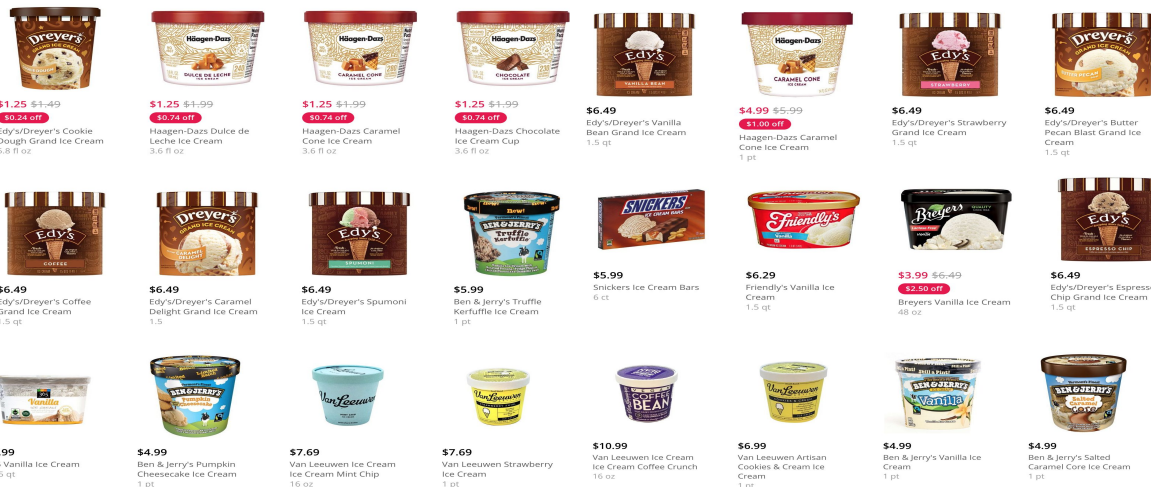
\$4.99
 Ben & Jerry's Vanilla Ice
 Cream
 1 pt



\$4.99
 Ben & Jerry's Salted
 Caramel Cone Ice Cream
 1 pt



Top 25 Late Night Purchases



\$7.99

Amy's Pizza Cheese
13 oz

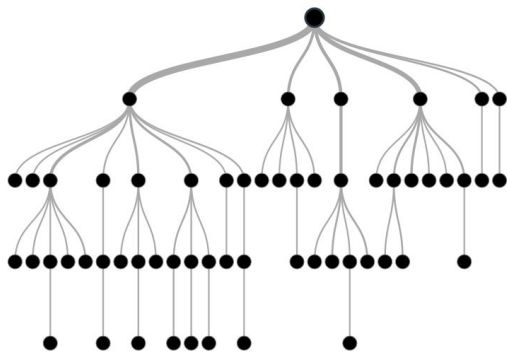
Predicting a repeat purchase



Key features:

- Product preference
- Purchase history
- Schedule

Random Forest Classification



Significant feature engineering

Binary target

GridSearchCV to tune model

Comparable results yielded by:

- Decision tree classifier
- XGBoost (gradient boosted)

Cross-validated Test Scores

85%

Accuracy

89%

Precision

85%

Recall

86%





F1

Confusion Matrix

True label

did not purchase

purchased

did not purchase	1146104 	202203 
purchased	41266 	266932 

did not purchase

purchased

Predicted label

false positives





Sample Basket Analysis

Instacart Customer # 55137




Product	Customer's Previous Orders									Total times purchased	Mean days between orders	Mean add to cart sequence	10th Order	
	1	2	3	4	5	6	7	8	9				Predicted	Actual
Fabric Softener Dryer Sheet										2	12	2	0	0
Crumbled Bacon										1	22	4	1	1
Steak Strips										1	35	3	0	0
Leaf Spinach										3	27	1	0	0
Shredded Mexican Blend Cheese										2	30	5	1	1
Dry Roasted Macadamias										4	28	2	1	1
Organic Guacamole										3	20	3	1	1
Creamy Almond Butter										4	28	2	1	1
Himalayan Pink Salt										1	89	3	0	0
Sparkling Mineral Water										2	9	2	0	0


Applications


Dynamic Product Recommendations

Buy It Again [View 100+ more >](#)



\$16.99 / lb
Beef Flank Steak



\$0.43 each
Banana 
At \$0.99/lb

Targeted Discounts



Stella Moretti <stella.moretti@gmail.com>

Shhh... We've got a secret!

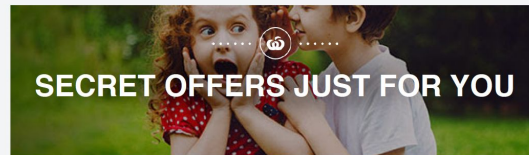
1 message

Woolworths Rewards <contacts@email.woolworthsrewards.com.au>
Reply-To: No Reply <reply-fe9611737163017d74-1410333_HTML-106021430-6389533-85@email.woolworthsrewards.com.au>
To: stella.moretti@gmail.com

25 August 2017 at 01:05



Member number: 9344757256404 | Start earning points today



Stock availability may vary by store. While stocks last.

Hi Stella, as one of our valued members, we're giving you the chance to earn bonus points with secret offers on products you like to shop for.

These secret offers are so secret,



Thank you

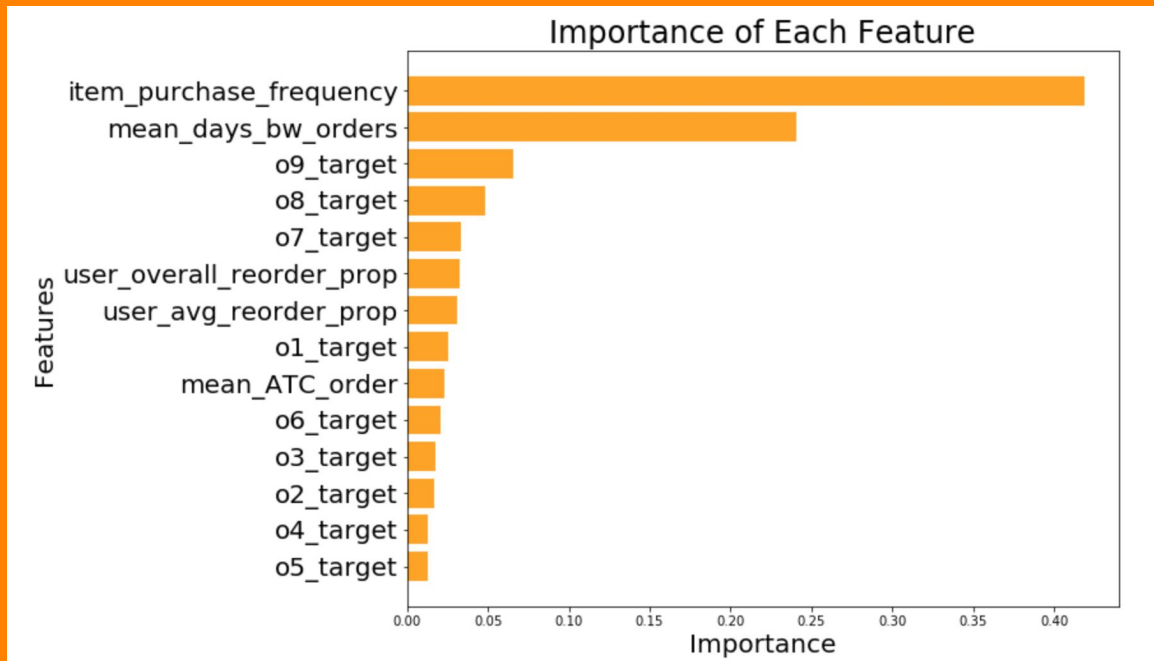
“The Instacart Online Grocery Shopping Dataset 2017”

Accessed October 2017 from:

<https://www.instacart.com/datasets/grocery-shopping-2017>

Back Up

Features Importance



RF Test / Train Scores

Test scores:

	precision	recall	f1-score	support
0.0	0.97	0.85	0.90	1348307
1.0	0.57	0.87	0.69	308198
avg / total	0.89	0.85	0.86	1656505

Accuracy: 0.852342733647

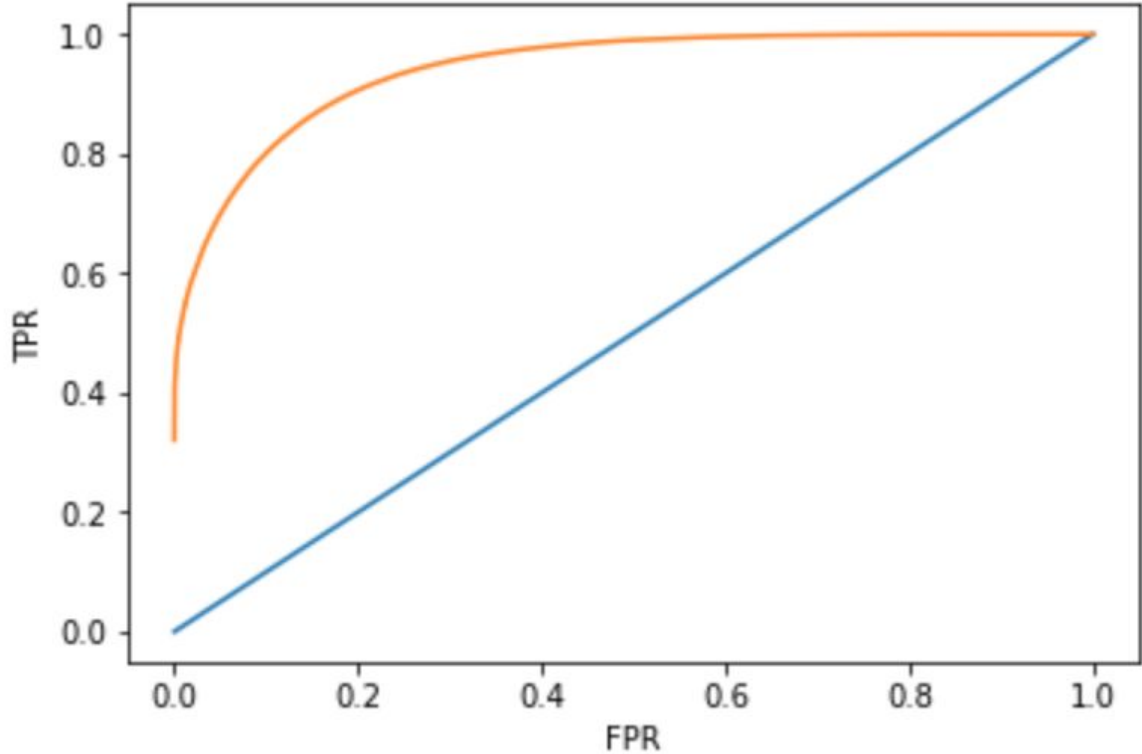
Cohen Kappa Score: 0.594830345576

Training scores:

	precision	recall	f1-score	support
0.0	0.88	0.86	0.87	720506
1.0	0.86	0.88	0.87	720506
avg / total	0.87	0.87	0.87	1441012

Receiver Operating Characteristic (ROC) Curve

FPR = False Positive Rate
TPR = True Positive Rate



Decision Tree Classifier

Test scores:

	precision	recall	f1-score	support
0.0	0.96	0.85	0.90	1348307
1.0	0.56	0.86	0.68	308198
avg / total	0.89	0.85	0.86	1656505

Accuracy: 0.848694087854

Cohen Kappa Score: 0.586570362734

Training scores:

	precision	recall	f1-score	support
0.0	0.86	0.85	0.85	720506
1.0	0.85	0.86	0.86	720506
avg / total	0.86	0.86	0.86	1441012

XG Boost Classifier

- Boosting is an ensemble technique where new models are added to correct the errors made by existing models.
- Uses gradient descent to minimize the loss when adding new models

Test scores:

	precision	recall	f1-score	support
0.0	0.93	0.88	0.90	1348307
1.0	0.56	0.71	0.63	308198
avg / total	0.86	0.84	0.85	1656505

Accuracy: 0.844240735766

Cohen Kappa Score: 0.531911230374

Training scores:

	precision	recall	f1-score	support
0.0	0.75	0.88	0.81	720506
1.0	0.85	0.71	0.77	720506
avg / total	0.80	0.79	0.79	1441012

Why Random Forest?

- High accuracy
- Fast
- Non-homogenous features
- Easy to tune to prevent overfit
 - Minimum sample split = 75
 - Minimum impurity decrease = 10-5
- Implicit feature selection

Extensions

- Ensemble classifier models
- Further feature engineering
- Find similar customers to make new product recommendations



Use data to predict:

- Will a customer return?
- Which products will they purchase again?
- Which products might they try for the first time?
- How long until their next purchase?
- Which shoppers have similar tastes?