

Hotdog or not Hotdog

Brian Hoepfl and Ryan Huckleberry





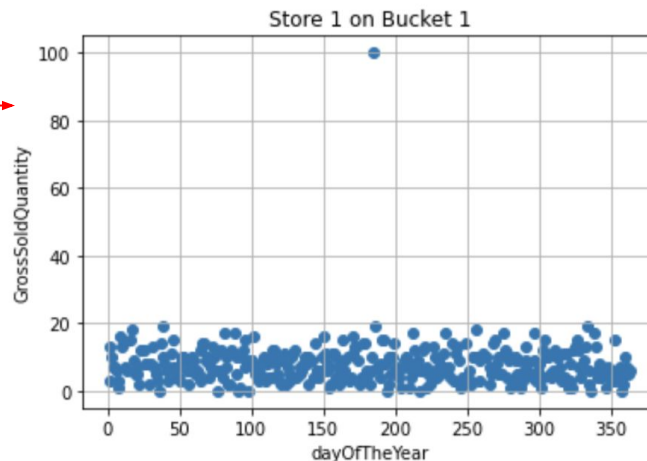
Data Preprocessing



Data Exploration & Data Cleaning

1. Many features are all constant when conditioned on StoreNumber
 - a. Cash/Credit Site, EBT Site, Loyalty Site, ExtraMile Site, CoBrand, Alcohol, Carwash, Food Service, City, State
2. Outliers on days 184, 185 due to July 4th!
 - a. Display of Store 1 Bucket 1 Outlier →
3. Stores 1, 2, 3 more similar than store 4
4. Integer typed categorical variables

*Minimal cleaning because of clean nature of data





Feature Selection and Generation

- Feature Selection
 - Removed many features that were either dependent on the store or the same for all of the training data (Ex: Tx same for all data)
- Feature Generation
 - We also generated some features. It was difficult to determine given the extent of the data. (i.e. There was no year or specific location)
 - Days of the week were very useful (discussed more in feature importance)
 - Other parameters such as seasons were not useful for model



Modeling Process





Models Applied

Different Models Applied:

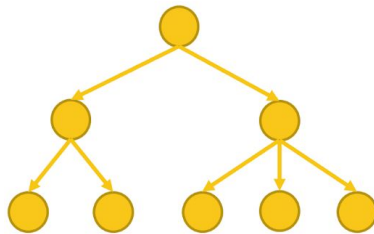
- Naive Bayes
- Random Forests
- Linear Regression
- Dense Neural Network

Random Forest Regression

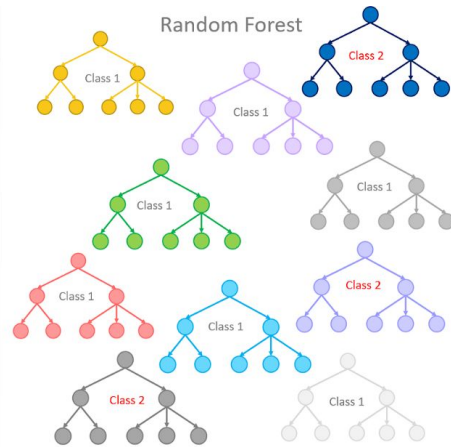
Trained model by changing parameters such as:

- Number of estimators
- Max_features
- Max_depth

Single Decision Tree



Random Forest



Feature Importance Analysis



	Feature	Importance Mean	Importance
0	dayOfWeek	0.700363	0.7003628586759241 +/- 0.0313354731247596
1	StoreNumber	0.674502	0.6745016112909391 +/- 0.03329376371024005
2	3HourBucket	0.263259	0.26325937354220524 +/- 0.0203095298503114
3	dayOfTheYear	0.017331	0.01733132537176749 +/- 0.002206075216880444
4	season	0.008720	0.008719934560932074 +/- 0.002868334634141871



Outcomes

Outcomes from the Random Forest Model:

- RMSE Validation 80/20 split: 9.85674
- RMSE Validation 70/30 split: 9.670871

Used better validation split (70/30):

- RMSE Scoring: 3.868947

Random Forest Regression 70/30

```
R^2 Training Score: 0.87
OOB Score: 0.81
Validation Score: 0.80
```