Food Classifier

Identifying vegetarian food by their nutrients

Why do we need a meat classifier?

- Provides a better understanding of the market.
- Identifies food containing meat for vegetarians or people with allergies.

Which evaluation metric should be used? What is a good score?

- Recall should be used to minimize incorrect vegetarian labels.
- A recall of at least .9 is ok for internal statistics.
- At least .95 for vegetarians.

How were labels identified?

- Food categories that should mostly be vegetarian or mostly contain meat were labeled 0 or 1 respectively.
- Uncertain food categories were not used in the modeling process.

Veg Labels:

Cheese
Nut & Seed Butters
Pasta by Shape & Type

Meat Labels:

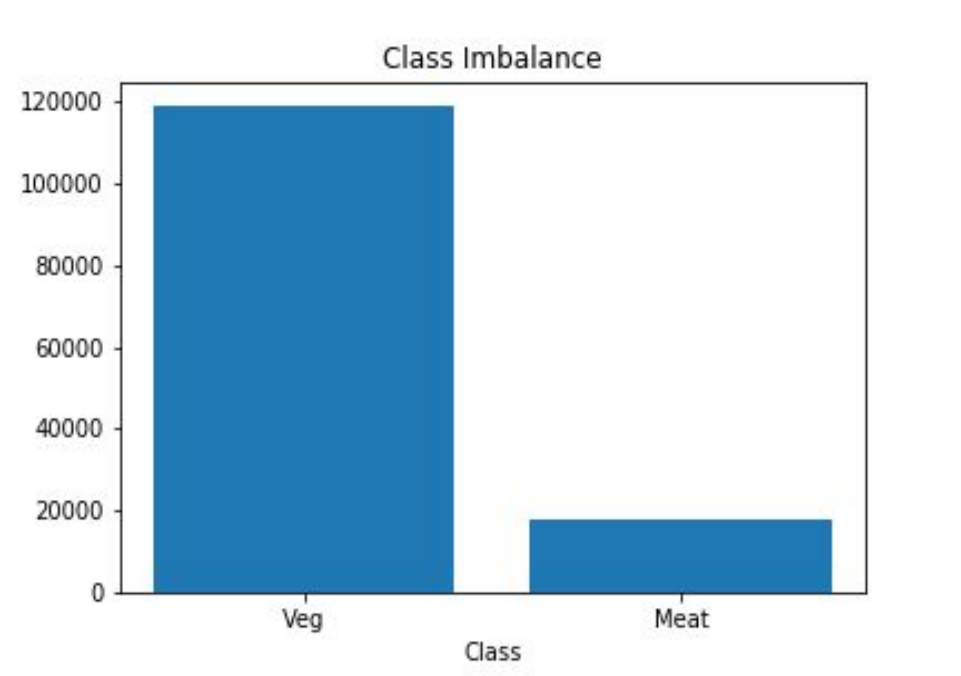
Pepperoni, Salami & Cold Cuts
Frozen Fish & Seafood
Other Meats

No Label:

Sweets
Soups, Sauces, and Gravies
Fast Foods

Data Preprocessing

- Removed impossible data before splitting.
- Removed nutrients with higher than .8 correlation.
- Downsampled vegetarian foods to match meat sample size.



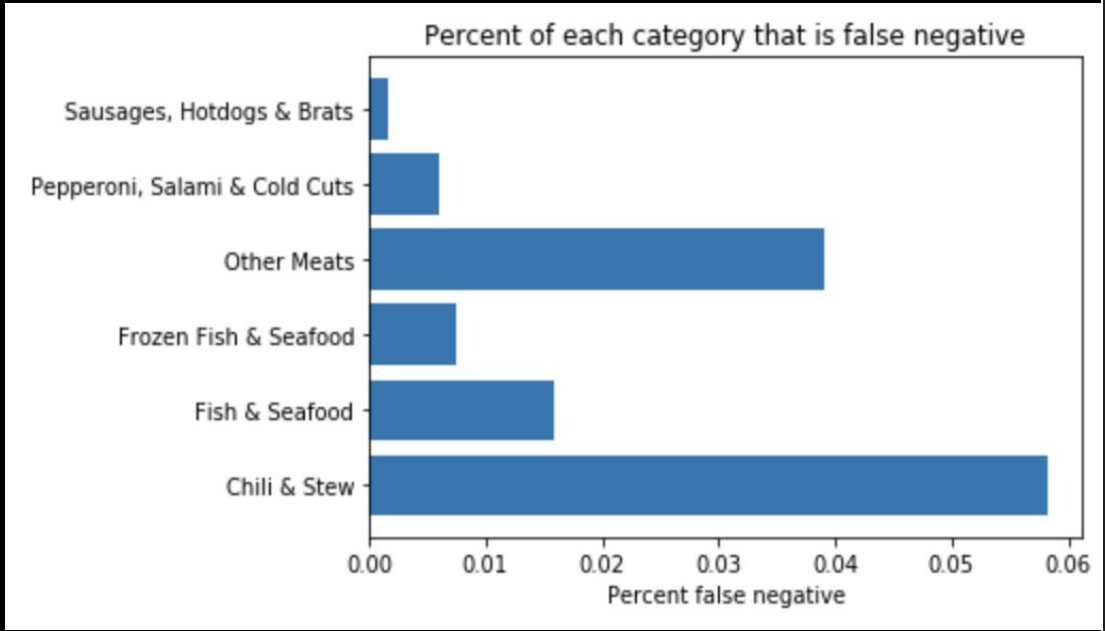
Random Forest Results

- Recall of .988.
- Accuracy of .99.
- Many errors were came from the original labeling process.

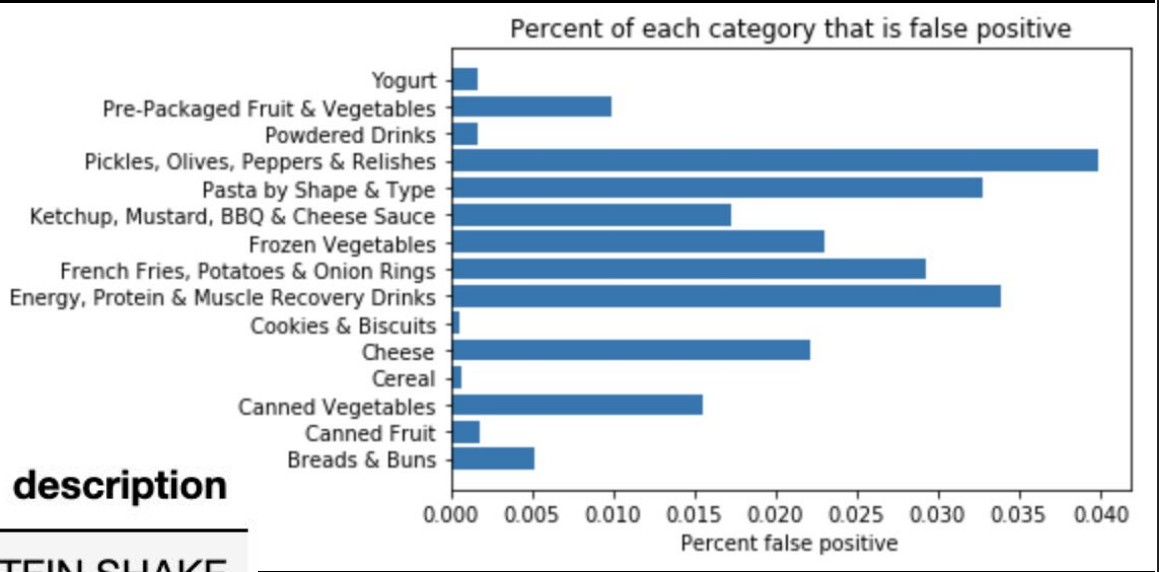
False Negatives

Examples:

prediction	description
0	CHILI SAUCE HOT
0	BPLLYWOOD BURGERS
0	SHRIMP CASHEW KORMA
0	CHILI SAUCE



False Positives



Examples:

prediction	description
1	PURE WHEY PROTEIN SHAKE
1	EGG SALAD
1	WHEAT LOAVES, WHEAT
1	ORIGINAL BEEF TOASTED RAVIOLI

Conclusions

- Our model fit very well on certain labels.
- May be generalizable.
- Many of the model's errors came from assumptions about labels, not the model being wrong.
- Might need a model focused on eggs.