

A still life composition featuring walnuts, cherry tomatoes, and a watermelon slice on a light pink background. In the top left, there are several walnuts, some whole and some cracked open to reveal the nutmeat. A string of red cherry tomatoes is draped along the left side. In the bottom left, a triangular slice of watermelon with a green rind and red flesh is visible.

# WebScraping and Classification using Natural Language processing

Vala Rahmani  
July 2019



## Goal

“With the rise of environmental and nutrition awareness, plant-based products have emerged that target the health conscious communities. Looking at the Vegan and Paleo subreddits we are interested in knowing what are some of the topics these groups, are interested in so we can gear our content towards them. This will be achieved by using CountVectorizer and Logistic Regression and success of our model will be evaluated by the accuracy score and f1\_score.”



[illegible]



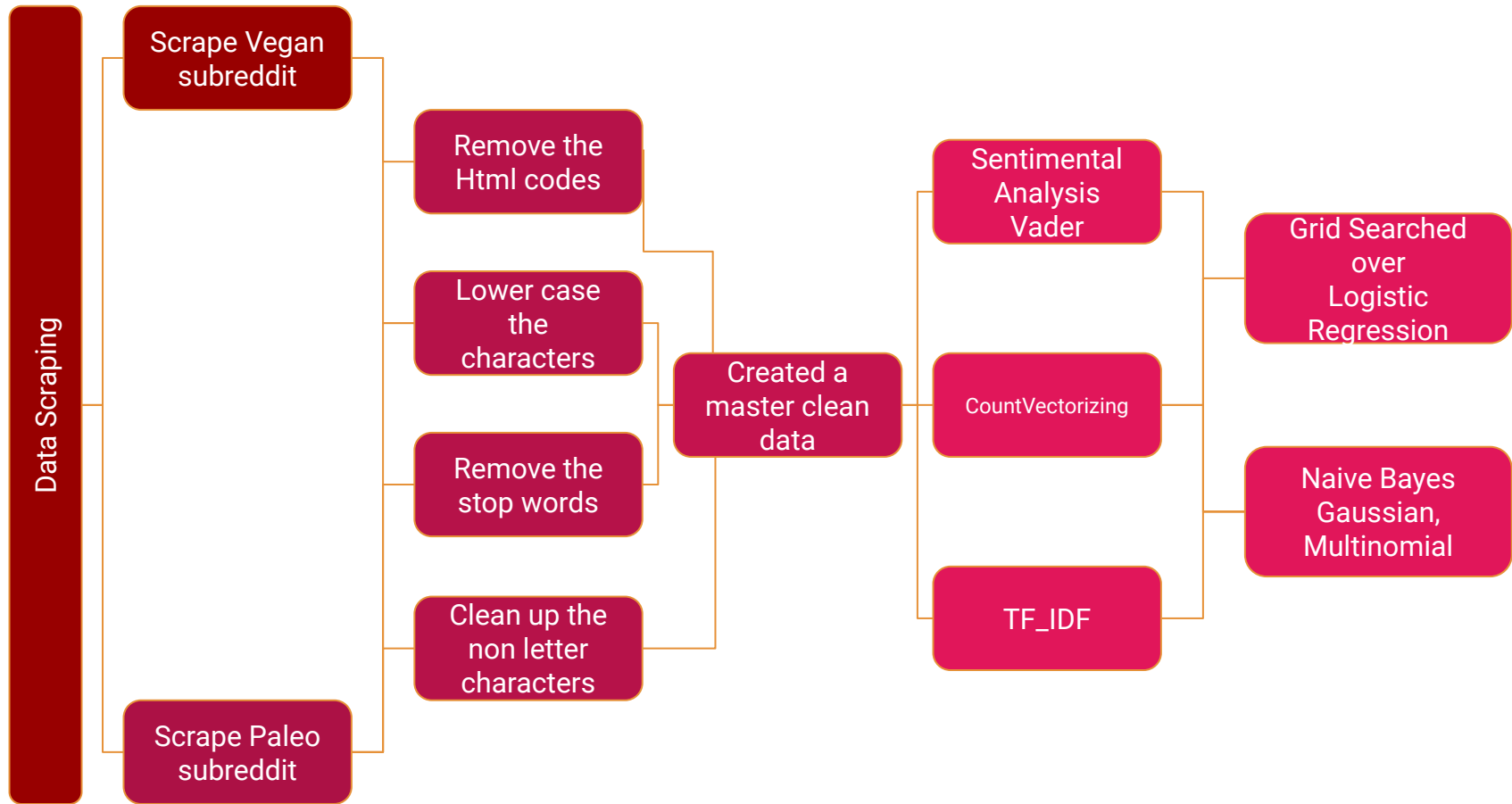
# Data Summary

**Vegan:** Doesn't use animal products such as meat, dairy and eggs

**Paleo:** Eats mainly fish, meat, vegetables, but not grains or dairy product

Values	Vegan	Paleo	Total
Number of posts	973	667	1640
Median number of comments	6	9	6
Members	345K	141K	486K







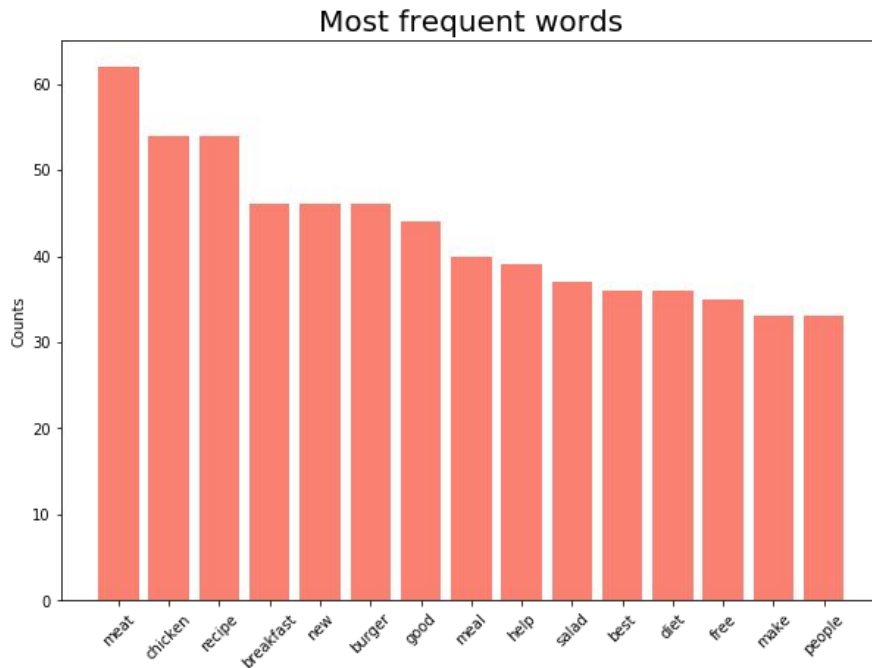
## Summary

**Accuracy score of: 0.77**

**F1\_score of :0.662**

**Baseline: 0.6088**

Final model:  
Used countvectorizer sentimental  
analysis and fit over logistic  
regression





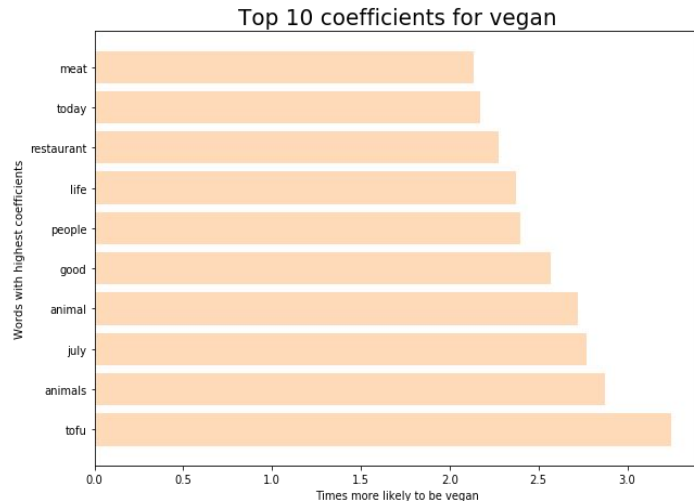
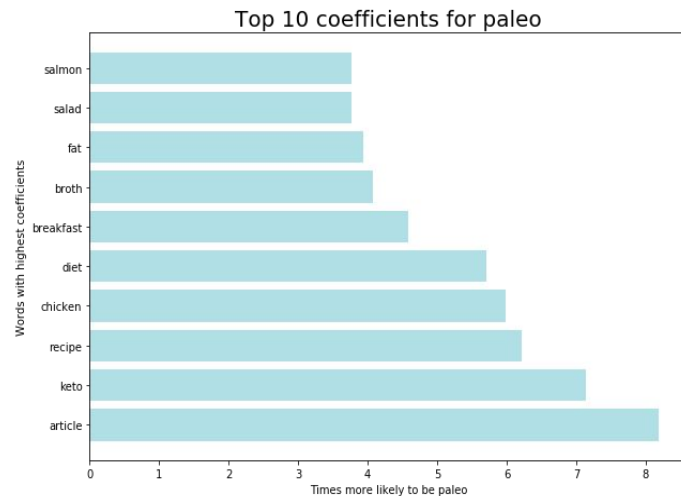
# Findings

***Makers of 'meatless meatballs' and 'vegan bacon' sue over Mississippi labeling law***

***The makers of 'meatless meatballs' and 'vegan bacon' sue over a Mississippi labeling law***

***Mississippi is forbidding grocery stores from calling veggie burgers “veggie burgers”: Proponents of the law say it’s necessary to avoid confusing consumers. That...doesn’t sound right.***

***As if it wasn’t already hard enough to have access to new and exciting vegan foods in South Mississippi, now this...***





## **Link to the project on github**

<https://github.com/valarn?tab=repositories>