


A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with blue dots. The lines are thin and grey, creating a mesh-like structure.

Improving Online Advertising via Classification

Kyle Ness, Data Scientist,
Online Analysis Co.

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with several nodes highlighted in blue.

Problem

Our client, Everything Meal Prep, wants better online ad targeting for their vegan and carnivore product lines

- Can a classification system accurately classify customer as vegan or carnivore?
- How much better is this than random advertising? (baseline)
- Insights into how to market these products (word choice)?

Approach

Roughly 1,050 posts were scraped from each of r/vegan and r/carnivore, assembled into datasets, concatenated.

- Title and description data / content
- Performed cleaning (duplicates, regex), pre-processing, EDA
- Baseline accuracy post-cleaning: still 50%
- Count vectorized data, fitted to Naive Bayes and Logistic Regression estimators

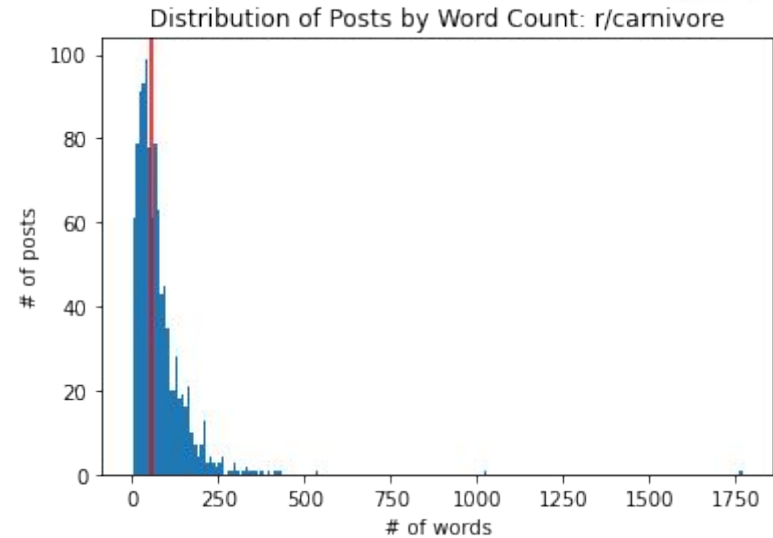
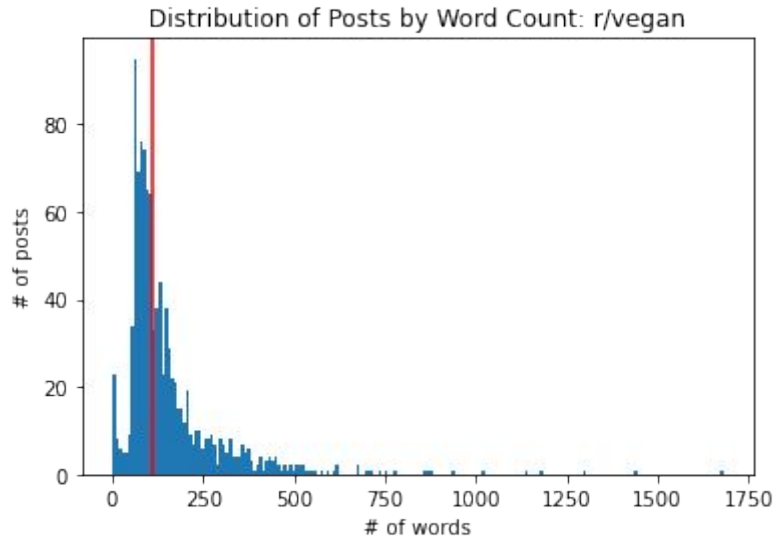
Cleaning

```
' 479 million goats/year [1](https://ourworldindata.org/meat-production)✂ 574 million sheep/year [1](https://ourworldindata.org/meat-production)✂ 656 million turkeys/year [1](https://ourworldindata.org/meat-production)✂ 69,000 million chickens/year [1](https://ourworldindata.org/meat-production)✂ 1,500 million pig s/year [1](https://ourworldindata.org/meat-production)✂ 302 million cows/year [1](https://ourworldindata.org/meat-production)✂ at least 900,000 million fish/y ear [2](http://fishcount.org.uk/published/std/fishcountstudy.pdf)✂ at least 300,000 whales and dolphins in fishing bycatch/year [3](https://www.fishforward.eu/en/project/by-catch/)✂ 250,000 endangered turtles in fishing bycatch/year [3](https://www.fishforward.eu/en/project/by-catch/)✂ salmon farms pass diseases to wild fish populations [4](http://www.eurocbc.org/Effects%20of%20Aquaculture%20on%20World%20Fish%20Supplies.htm)✂ carnivorous fish farms require even more wild fish feed [5](https://cdnscepub.com/doi/full/10.1139/cjfas-2016-0379)✂ 1/4 shark species threatened with extinction due to fishing [6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3897121/)✂ 77% agricultural land for only 18% calorie/37% protein supply [7](https://ourworldindata.org/global-land-for-agriculture)✂ livestock account for 31% of food emissions [8](https://ourworldindata.org/food-ghg-emissions)✂ twice as many emissions for livestock over crops for human c onsumption [8](https://ourworldindata.org/food-ghg-emissions)✂ only 6% of soy is grown for humans [9](https://www.ucsus.org/resources/soybeans) ✂ animal farm ers given legal exception to bestiality laws [10](https://www.researchgate.net/publication/320208999_How_Meat_Changed_Sex_The_Law_of_Interspecies_Intimacy_after_Industrial_Reproduction)✂ environmental racism of placing polluting animal farms in marginalized areas [11](https://europepmc.org/backend/ptpmcrender.fcgi?acc id=PMC1440786&blobtype=pdf)✂ 65% lactose intolerance worldwide [12](https://medlineplus.gov/genetics/condition/lactose-intolerance/#frequency)✂ right-wing ideologies predict greater acceptance of animal exploitation and more meat consumption [13](https://www.sciencedirect.com/science/article/abs/pii/S0191886914000 944) ✂ milk and ties to nazis [14](https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1373&context=asj) ✂ working in animal slaughter tied with PTSD, incre ased violent crimes [15](http://animalstudies.msu.edu/Slaughterhouses_and_Increased_Crime_Rates.pdf)✂ imported colonizer animals replacing native biodiversity [16](https://ourworldindata.org/what-are-drivers-deforestation) ✂ farmed honeybees endanger critical native pollinators and reduce species biodiversity [17](ht tps://www.wired.com/2015/04/youre-worrying-wrong-bees/)✂ human slavery in fishing supply chain [18](https://www.unodc.org/documents/human-trafficking/Issue_Pap er_-_TOC_in_the_Fishing_Industry.pdf)✂ amazon rainforest & indigenous homes destroyed for harvesting and feeding cows [19](https://www.amnesty.org/en/late st/news/2019/11/brazil-halt-illegal-cattle-farms-fuelling-amazon-rainforest-destruction/)✂ watchdominion.com [20](https://www.dominionmovement.com/watch) ✂ go vegan [21](https://veganbootcamp.org/).'
```

Virtually half of this post provides no value towards NLP classification.

In response: hyperlinks, new line chars, polls, parentheses, and other special characters were removed

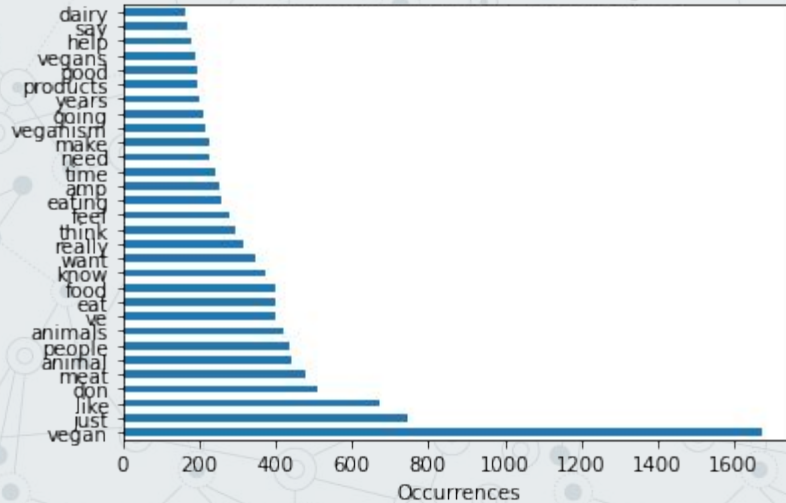
Insight: Distributions



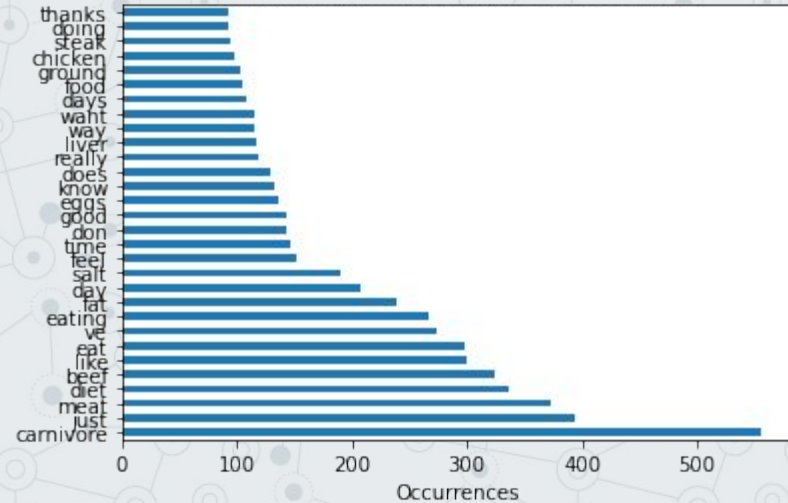
Takeaway: similar, but r/vegan usually has longer posts

Insight: Common Words

30 Most Common Words on r/vegan



30 Most Common Words on r/carnivore



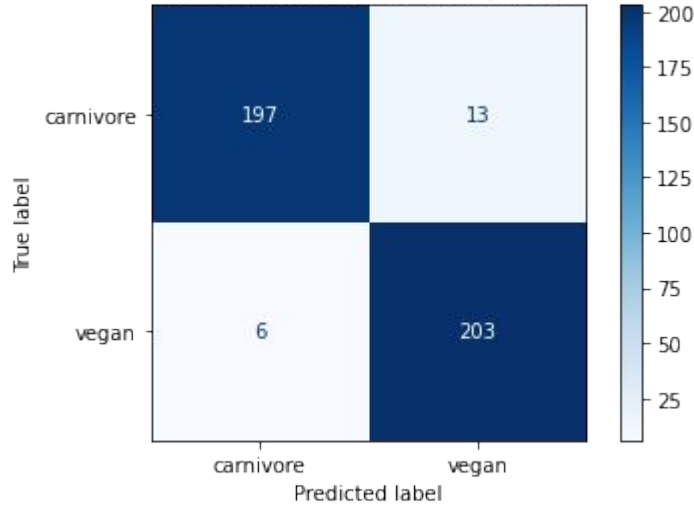
- Keywords / buzzwords that can be used in advertising
- Gives a hint of how the classification models will be discerning between the two subreddits
- Speculation: vegans = more conscientious?

Model Creation, Evaluation

- Going forward with Naive Bayes and Logistic Regression:
 - Tuning of transformer and estimator parameters via GridSearch
 - Best parameters for the estimators were all defaults
 - Models were evaluated based on accuracy, recall, specificity and precision.

Comparison

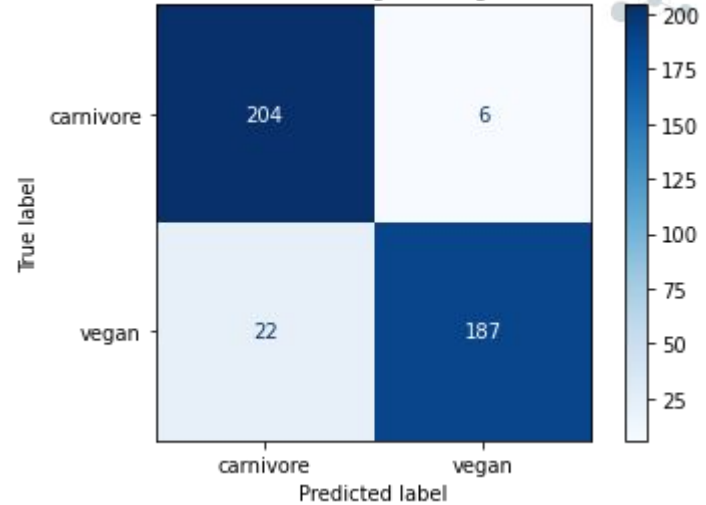
Confusion Matrix for Naive Bayes Model



Accuracy: 95.47%
Specificity: 93.81%
Precision: 93.98%
Recall: 97.13%

In this context,
these are most
important.

Confusion Matrix for Logistic Regression Model



Accuracy: 93.32%
Specificity: 97.14%
Precision: 96.89%
Recall: 89.47%

Conclusion, Recommendations

- Accurate classification of customers is indeed possible, and highly so!
 - Use the Naive Bayes model here
 - Keep in mind subreddit characteristics (post length, word choice, user personality) when deploying ads
 - Should benefit Everything Meal Prep greatly
 - Future: try different estimators, collect more data