Artificial Meat Review Predictions

Tamar Brand-Perez, Tiffany Price, Ben Tubbs, and Jose Santos

Team Members & Responsibilities

Tiffany Price:

Project Manager

Tamar Brand-Perez:

Database Lead

Ben Tubbs:

Machine Learning Lead

Jose Santos:

Github Lead

Executive Summary

Overview

The fake meat industry has recently experienced a dramatic positive shift in consumer interest and exponential growth. Our team wants to predict how consumer reviews affect ratings, which will assist store owners in their decision making process of which brand to sell.

Objectives

By analyzing the Amazon consumer reviews between 2017 - 2021, our machine learning tool will predict whether a non-rated review will be positive or negative.

Products Used

Technology

Tools

Language

Algorithms

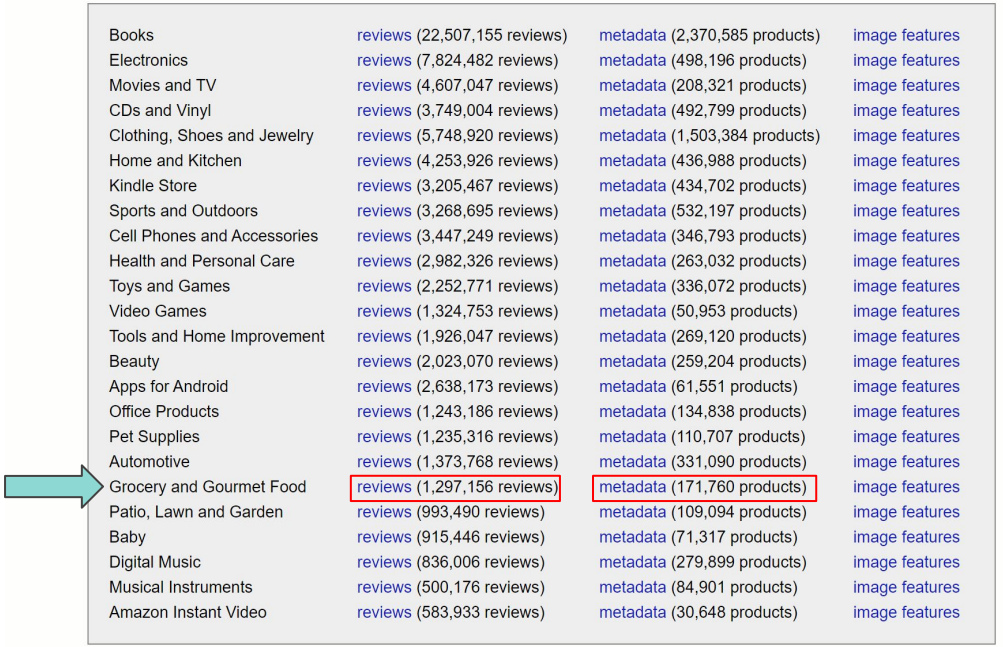
- Algorithm: Bag of Words Model
- Machine Learning: Naive Bayes Classifier

Data Exploration & Analysis

Database

Exploration

- Amazon product data
  - Reviews
    - *****/*
    - text
    - review length
  - Metadata
    - price



Books	reviews (22,507,155 reviews)	metadata (2,370,585 products)	image features
Electronics	reviews (7,824,482 reviews)	metadata (498,196 products)	image features
Movies and TV	reviews (4,607,047 reviews)	metadata (208,321 products)	image features
CDs and Vinyl	reviews (3,749,004 reviews)	metadata (492,799 products)	image features
Clothing, Shoes and Jewelry	reviews (5,748,920 reviews)	metadata (1,503,384 products)	image features
Home and Kitchen	reviews (4,253,926 reviews)	metadata (436,988 products)	image features
Kindle Store	reviews (3,205,467 reviews)	metadata (434,702 products)	image features
Sports and Outdoors	reviews (3,268,695 reviews)	metadata (532,197 products)	image features
Cell Phones and Accessories	reviews (3,447,249 reviews)	metadata (346,793 products)	image features
Health and Personal Care	reviews (2,982,326 reviews)	metadata (263,032 products)	image features
Toys and Games	reviews (2,252,771 reviews)	metadata (336,072 products)	image features
Video Games	reviews (1,324,753 reviews)	metadata (50,953 products)	image features
Tools and Home Improvement	reviews (1,926,047 reviews)	metadata (269,120 products)	image features
Beauty	reviews (2,023,070 reviews)	metadata (259,204 products)	image features
Apps for Android	reviews (2,638,173 reviews)	metadata (61,551 products)	image features
Office Products	reviews (1,243,186 reviews)	metadata (134,838 products)	image features
Pet Supplies	reviews (1,235,316 reviews)	metadata (110,707 products)	image features
Automotive	reviews (1,373,768 reviews)	metadata (331,090 products)	image features
Grocery and Gourmet Food	reviews (1,297,156 reviews)	metadata (171,760 products)	image features
Patio, Lawn and Garden	reviews (993,490 reviews)	metadata (109,094 products)	image features
Baby	reviews (915,446 reviews)	metadata (71,317 products)	image features
Digital Music	reviews (836,006 reviews)	metadata (279,899 products)	image features
Musical Instruments	reviews (500,176 reviews)	metadata (84,901 products)	image features
Amazon Instant Video	reviews (583,933 reviews)	metadata (30,648 products)	image features

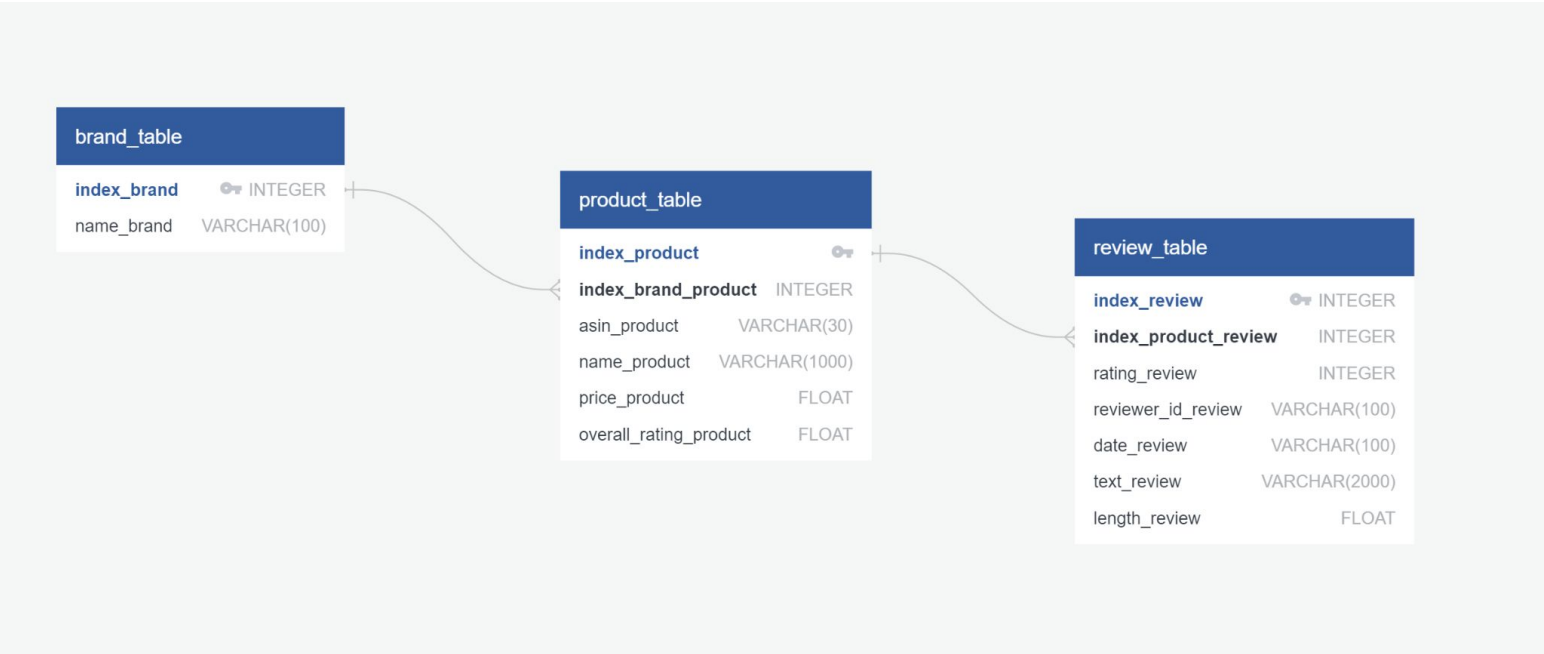
Exploration (cont'd)

- Amazon product data (cont'd)
  - scraping
    - *****/*
    - text
    - brand
    - product
    - price

Questions

- Inform store owners:
  - positive or negative review?
  - Frequent words reviewers associate with a brand.
  - price change over time?
  - rating change over time?
  -

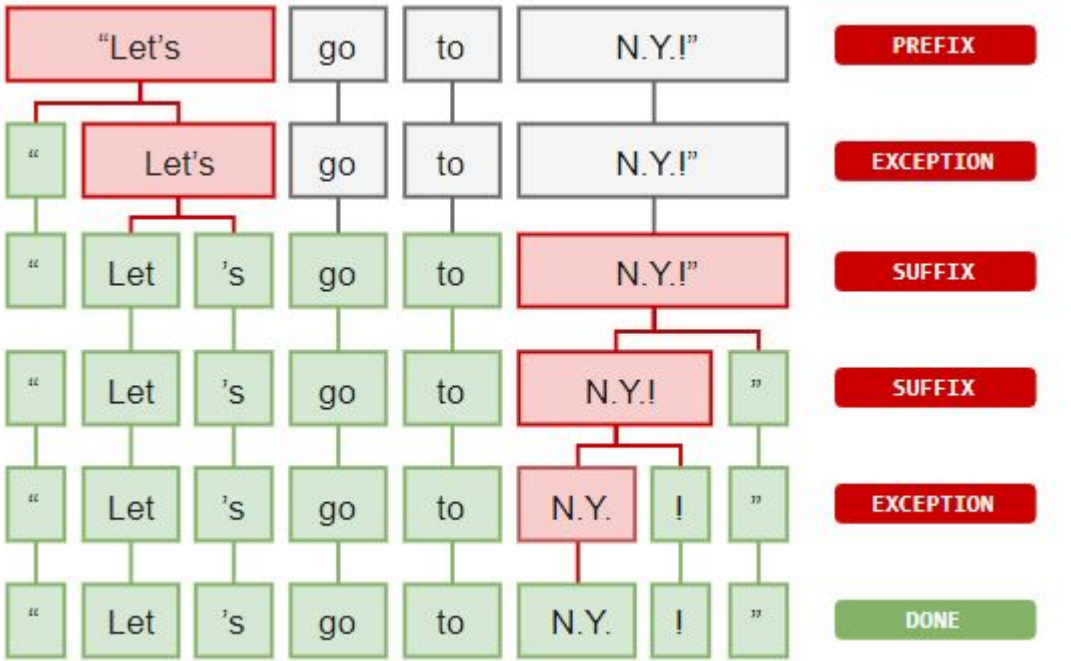
Database


brand_table
index_brand INTEGER
name_brand VARCHAR(100)
product_table
index_product
index_brand_product INTEGER
asin_product VARCHAR(30)
name_product VARCHAR(1000)
price_product FLOAT
overall_rating_product FLOAT
review_table
index_review INTEGER
index_product_review INTEGER
rating_review INTEGER
reviewer_id_review VARCHAR(100)
date_review VARCHAR(100)
text_review VARCHAR(2000)
length_review FLOAT

Machine Learning

Preprocessing and Normalization

- Tokenization
- Casing
- Removing Non Alphanumerics
- Length
- Stop Words
- Lemmatization



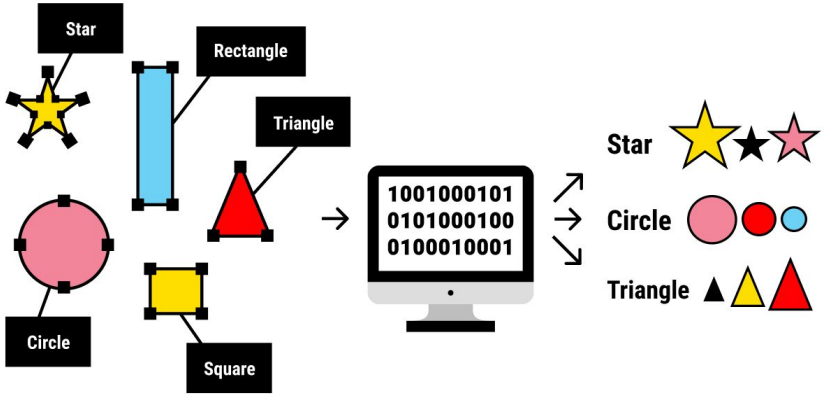
Cleaning

	raw_word	cleaned_word
0	..trouble..	trouble
1	trouble<	trouble
2	trouble!	trouble
3	<a>trouble	trouble
4	1.trouble	trouble

	original_word	lemmatized_word
0	trouble	trouble
1	troubling	trouble
2	troubled	trouble
3	troubles	trouble

	original_word	lemmatized_word
0	goose	goose
1	geese	goose

Featureset



	Mono-grams	Bi-grams	Tri-grams
Phrases Extracted (crazy good, stone cold crazy)	crazy (2) cold good	crazy good cold crazy stone cold	stone cold crazy

	the	red	dog	cat	eats	food
1. the red dog →	1	1	1	0	0	0
2. cat eats dog →	0	0	1	1	1	0
3. dog eats food →	0	0	1	0	1	1
4. red cat eats →	0	1	0	1	1	0

Naive Bayes

P(Queen | Face) =
P(Face | Queen) * P(Queen) / P(Face)

P(Face | Queen) is 1
P(Queen) is 4/52 = 1/13
P(Face) is 12/52 = 3/13

P(Queen | Face) = 1 * (1/13) / (3/13)
= (1/13) * (13/3)
= 1/3

(Correct since 4 Queens out of 12 Face cards = 4/12 or 1/3)

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|A) \cdot P(A)}{P(B)}$$

Naive Bayes and NLP

P(Positive | Fake meat is delicious) =

P(Fake meat is delicious | Positive) * P(Positive) / P(Fake meat is delicious)

P(Positive | fake) * P(Positive | meat) * P(Positive | delicious)

We do the same for the negative.

P(Positive | Fake meat is delicious) > P(Negative | Fake meat is delicious)?

Most Informative Features

Most informative Features (Lemmas)									
lic	great	tin	eat	received	dis	rec	purchased	receive	flavor
perfect	purchase	rip	tasty	using	amazing	bag	ill	thought	order
pack	tho	ordered	love	though	candy	pic	good	thank	red
ate	always	ash	came	deliver	end	expected	far	fresh	know
low	pleasant	rat	ship	sure	tasted	total	warn	warning	wow
buy	go	per	get	ick	thin	product	ive	although	believe
chocolate	cola	fan	huge	late	looked	met	probably	rob	salt
take	want	us	absolutely	ten	ice	expect	item	read	taste
ant	use	absolute	favorite	happy	live	te	big	thanks	an
advertised	bar	bitter	boy	can	chew	cook	cooky	covered	delivery

Most Informative Features II

Feature	Sentiment	Certainty
Lic	Positive	17.0%
Great	Positive	12.9%
Tin	Negative	9.7%
Eat	Positive	9.7%
Received	Negative	9.4%
Dis	Negative	8.4%
Rec	Negative	8.4%
Purchased	Negative	8.3%
Receive	Negative	7.6%
Flavor	Negative	6.3%
Perfect	Positive	5.7%
Purchase	Negative	5.4%
Rip	Negative	5.0%
Tasty	Positive	5.0%
Using	Positive	5.0%

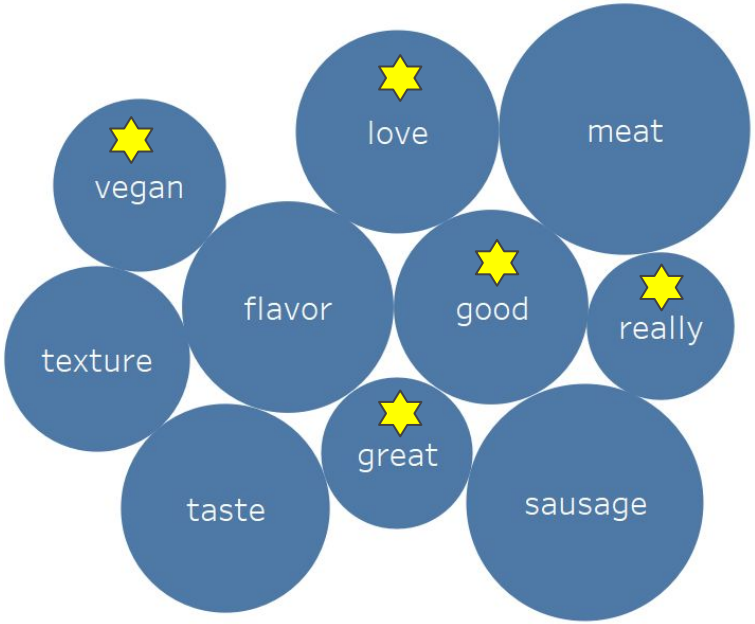
Dashboard

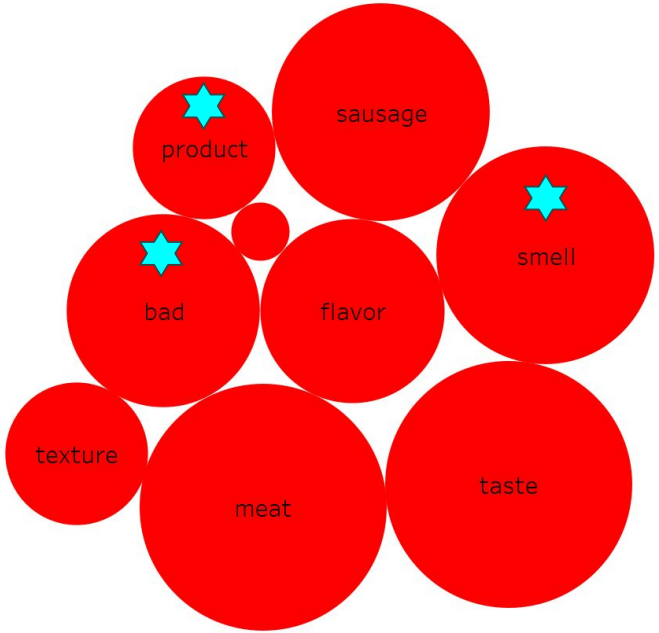
Fake Meat Products Reviewed

Brand Name	Product Name
Beyond Meat	Beyond Meat Beyond Breakfast Sausage Plant-Based Breakfast Patties, Classic 7.4 oz
	Beyond Meat Beyond Sausage Plant-Based Dinner Sausage Links, Brat Original 14 oz
	Beyond Meat from PlantBased Frozen oz lb. Package, Ground Beef Substitute, 16 Ounce
Boca	Boca Original Vegan Non-GMO Soy Chik'n Veggie Nuggets (10 oz Pouch)
	Boca Original Vegan Spicy Non-GMO Soy Chik'n Veggie Patties (4 Count)
Gardein	Gardein Gluten-Free Ultimate Plant-Based Beefless Ground Crumbles, Vegan, Frozen, 13.7 oz.
	Gardein Sliced Italian Plant-Based Saus'age, Vegan, Frozen, 9 oz.
	Gardein, Burger Beefless Ultimate, 12 Ounce
Quorn	Quorn Foods Meatless Grounds, Vegetarian, Frozen, 12 Oz
	Quorn Foods Meatless Nuggets, Vegetarian, Frozen, 10.6 Oz
	Quorn, Meat-Free Meatballs, 10.6 oz (Frozen)
Tofurky	Tofurky Deli Slices Oven Roasted
	Tofurky, Deli Slices, Hickory Smoked, 5.5 oz

Fake Meat Brand Sentiment

BEYOND MEAT®



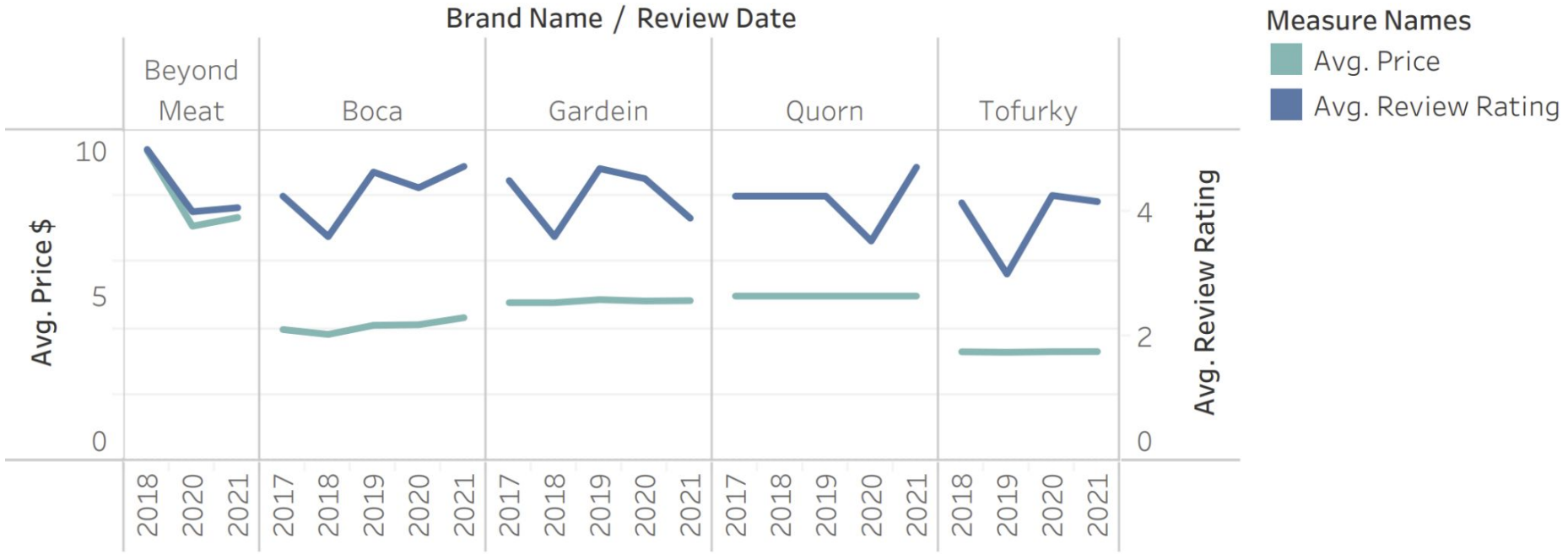


Vegan, love, good, really, great

Product, bad, smell

Keyword frequency by brand in positive reviews

Brand Name	flavor	good	great	love	meat	really	sausage	taste	texture	vegan
Beyond Meat	36%	25%	22%	33%	29%	22%	61%	28%	33%	21%
Boca	6%	12%	11%	6%	2%	9%		9%	1%	10%
Gardein	32%	23%	25%	24%	35%	18%	39%	27%	37%	28%
Quorn	11%	16%	18%	19%	20%	29%		24%	18%	10%
Tofurky	15%	23%	24%	18%	14%	21%		12%	11%	32%



Review v. Price by Brand per Year

Conclusion

Results

- 80.### % accuracy from ML model
- Limitations

In Summary

Recommendation for Future Analysis

- Interactive visualizations
- Larger data set

What Our Team Would Have Done Differently

- Improve the algorithm to make it more than 80% accurate
- Algorithm does not consider the order

Demo

https://project-group-3.herokuapp.com/

Project Home

Welcome to our NLP model display website!

Interactive Dashboard

Things to include from the module

Content: The presentation should tell a cohesive story about the project and include the following:

- ~~Selected topic~~
- ~~Reason the topic was selected~~
- ~~Description of the source of data~~
- ~~Questions the team hopes to answer with the data~~
- ~~Description of the data exploration phase of the project~~
- Description of the analysis phase of the project
- ~~Technologies, languages, tools, and algorithms used throughout the project~~
- Result of analysis
- ~~Recommendation for future analysis~~
- ~~Anything the team would have done differently~~

Live Presentation

Requirements for the live presentation follow:

- All team members present in equal proportions.
- The team demonstrates the dashboard's real-time interactivity.
- The presentation falls within any time limits provided by the instructor.
- The submission includes speaker notes, flashcards, or a video of the presentation rehearsal.

Purpose & Sourcing

- Amazon Grocery and Gourmet Food data set: (http://deepyeti.ucsd.edu/jianmo/amazon/index.html).
- NLP was used to train a sentiment classifier.
- Goals:
  - Predict user sentiment for artificial meat products.
  - Assist store owners' with their analysis.

Sentiment specific keywords

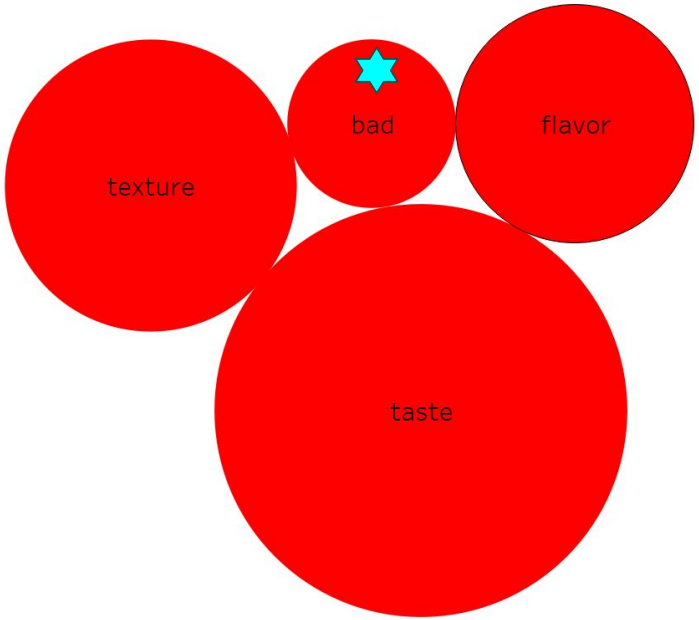
	+ KEYWORDS	- KEYWORDS
BEYOND MEAT	Vegan, love, good, really, great	Product, bad, smell
BOCA	Vegan, love, good, really, great, meat	Bad
gardein	Vegan, love, good, really, great, texture, sausage, flavor	Bad
Quorn	Vegan, love, good, really, great	Bad, product, food
Tofurky FAMILY OWNED SINCE 1980	Vegan, love, good, really, great, taste, meat	Bad, smell, food

Keyword frequency by brand in negative reviews

Brand Name	bad	flavor	food	meat	people	product	sausage	smell	taste	texture
Beyond Meat	42%	42%		72%	7%	35%	100%	82%	36%	40%
Boca	4%	8%							12%	20%
Gardein	8%		65%	12%	13%	18%			36%	13%
Quorn	23%	33%	12%	16%	53%	47%		6%	8%	27%
Tofurky	23%	17%	24%		27%			12%	8%	

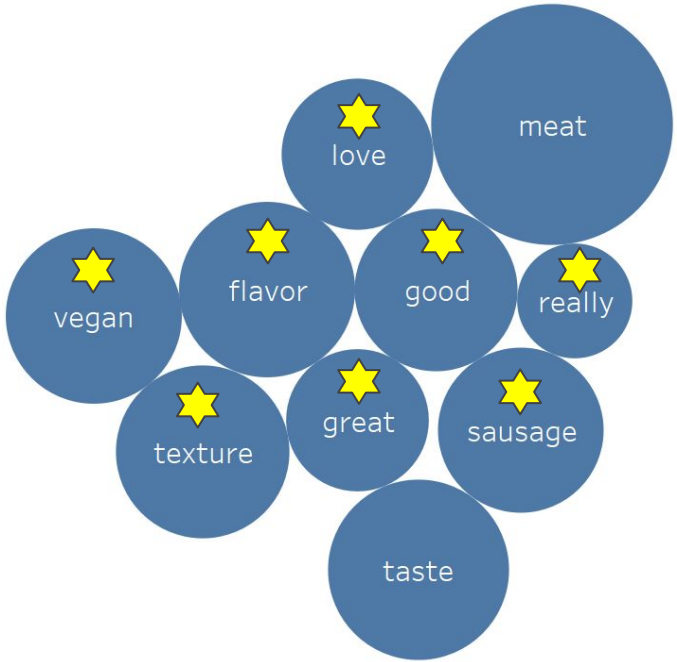
Fake Meat Brand Sentiment

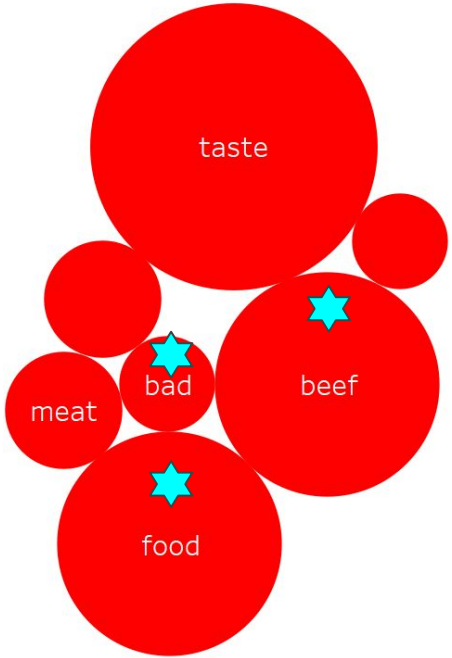
Vegan, love, good, really, great, meat



bad

Fake Meat Brand Sentiment

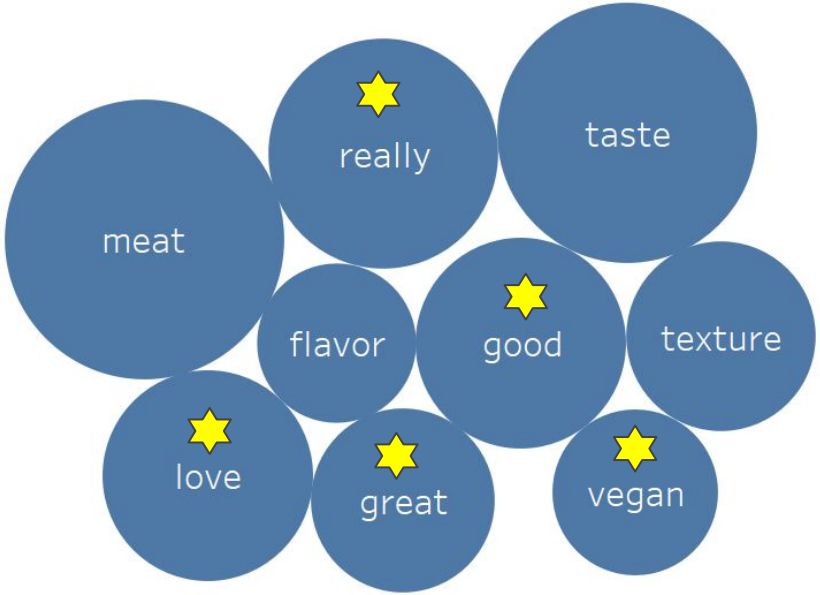


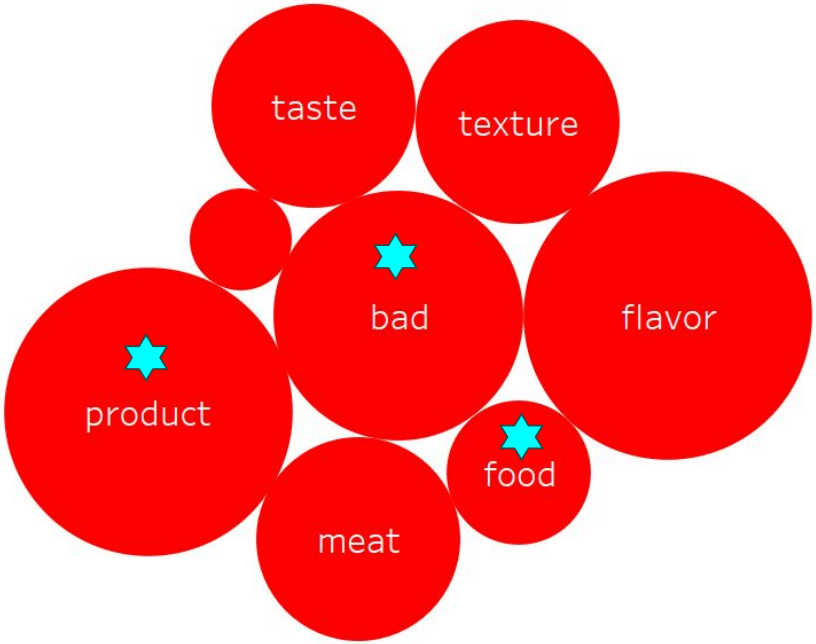


Vegan, love, good, really, great, texture, sausage, flavor

bad

Fake Meat Brand Sentiment

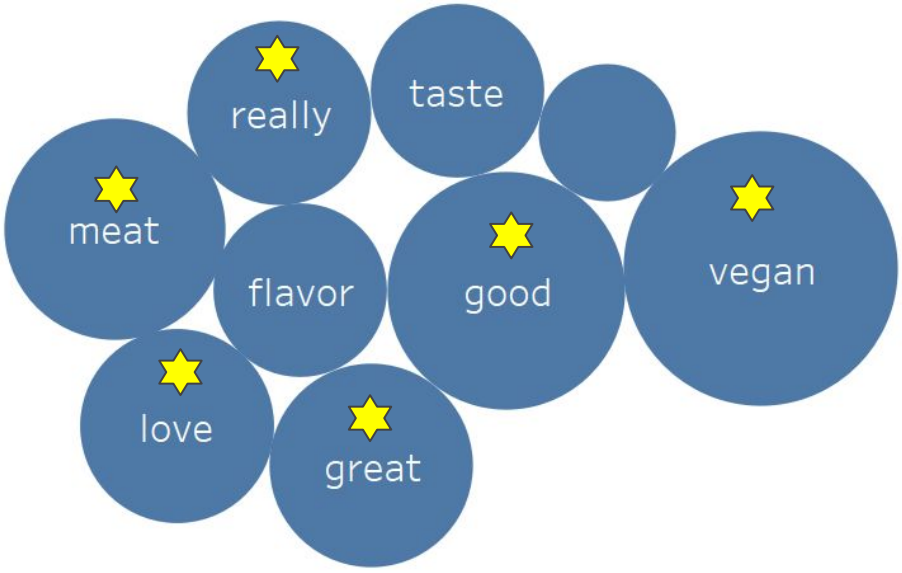


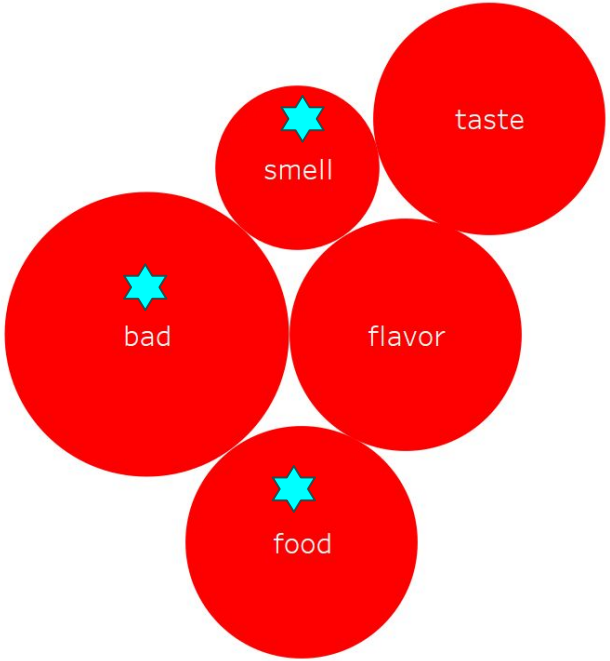


Vegan, love, good, really, great

Bad, product, food

Fake Meat Brand Sentiment



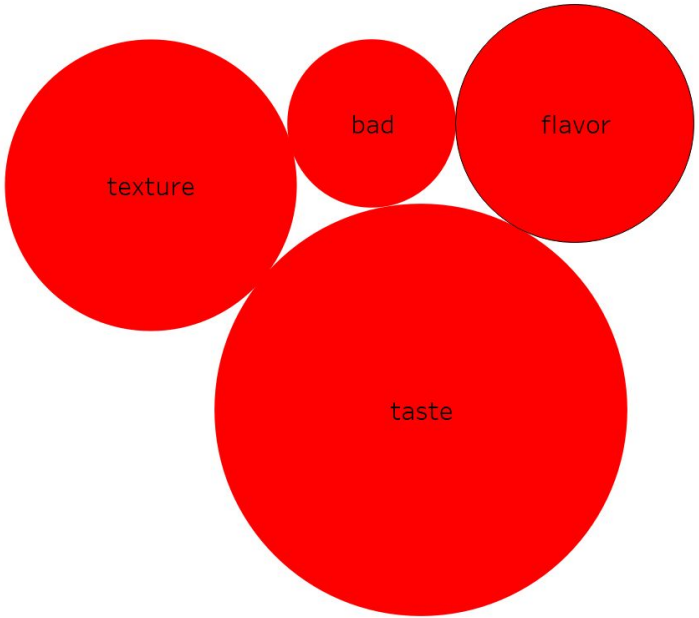


Vegan, love, good, really, great, taste, meat

Bad, smell, food

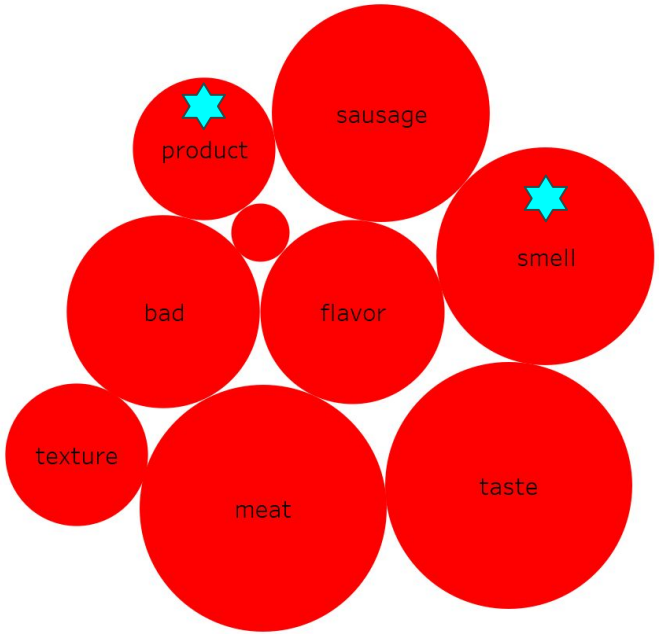
Fake Meat Brand Sentiment

BOCA



bad

BEYOND MEAT



Product, bad, smell