

Where's the Beef?



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



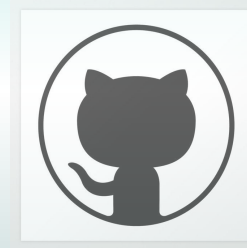
The Artificial Meat Industry

- **Security**
- **Safety**
- **Sustainability**
- **\$4.3 billion Industry**





Technology, Tools & Language Used



Bag of Words Model



Machine Learning: Naive Bayes Classifier



Data

Data Source

- Amazon product data
 - Me ~~X~~ ata
- Amazon scraped data
- Reddit scraped data



Alternatives to Meat





Data content

| Source | Amount | Star rating | Review Text | Reviewer | Date | Brand | Product | Price | asin | len |
|----------------|-------------|-------------|-------------|----------|------|-------|---------|-------|------|-----|
| UCSD Amazon | > 1 million | ★ | ✓ | ✓ | ✓ | | | | ✓ | ✓ |
| UCSD metadata | 116 | ★ | | | | ✓ | ✓ | ✓ | ✓ | |
| Scraped Amazon | 1073 | ★ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Scraped Reddit | 19 | | ✓ | | | | | | | |

Data Cleaning

- Dropped columns
- Added length column
- Dropped duplicates
- 1 and 5 stars
- Trimmed the 5 stars



Database

| groceries_review_table | |
|------------------------|---------------|
| index_g_review | INTEGER |
| rating_g_review | INTEGER |
| reviewer_id_g_review | VARCHAR(100) |
| date_g_review | VARCHAR(100) |
| text_g_review | VARCHAR(2000) |
| length_g_review | FLOAT |

| brand_table | |
|-------------|--------------|
| index_brand | INTEGER |
| name_brand | VARCHAR(100) |

| product_table | |
|------------------------|---------------|
| index_product | INTEGER |
| 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 |

| reddit_thread | |
|---------------|---------------|
| title | VARCHAR(100) |
| comment | VARCHAR(2000) |





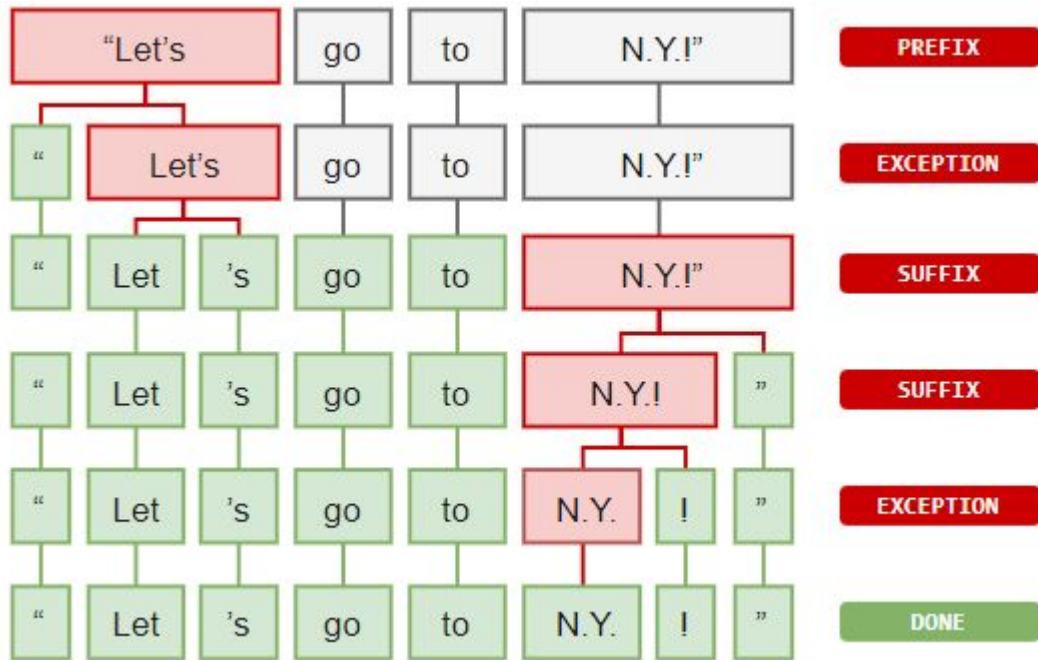
Questions

- How do people feel about fake meat products
 - Positive or negative review?
 - Frequent words reviewers associate with a brand.
 - Price change over time?
 - Rating change over time?
- Inform grocery store owners

Machine Learning

Preprocessing and Normalization

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

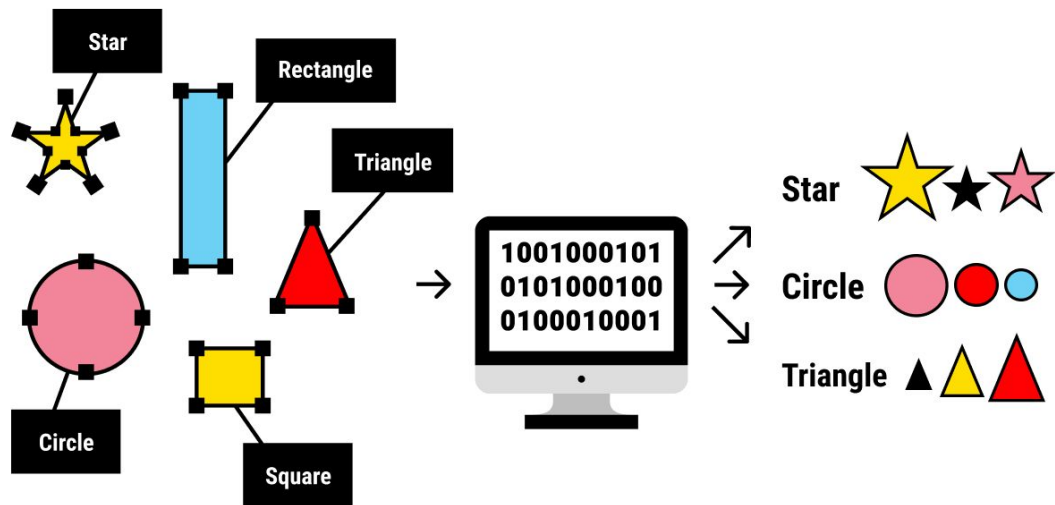




| | original_word | lemmatized_word |
|---|---------------|-----------------|
| 0 | trouble | trouble |
| 1 | troubling | trouble |
| 2 | troubled | trouble |
| 3 | troubles | trouble |

| | original_word | lemmatized_word |
|---|---------------|-----------------|
| 0 | goose | goose |
| 1 | geese | goose |

Feature Set



1. the red dog →

2. cat eats dog →

3. dog eats food →

4. red cat eats →

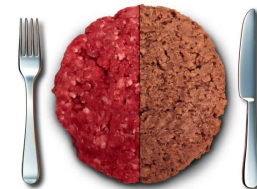
| | the | red | dog | cat | eats | food |
|--------------------|-----|-----|-----|-----|------|------|
| 1. the red dog → | 1 | 1 | 1 | 0 | 0 | 0 |
| 2. cat eats dog → | 0 | 0 | 1 | 1 | 1 | 0 |
| 3. dog eats food → | 0 | 0 | 1 | 0 | 1 | 1 |
| 4. red cat eats → | 0 | 1 | 0 | 1 | 1 | 0 |

Most Informative Features

| 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% |

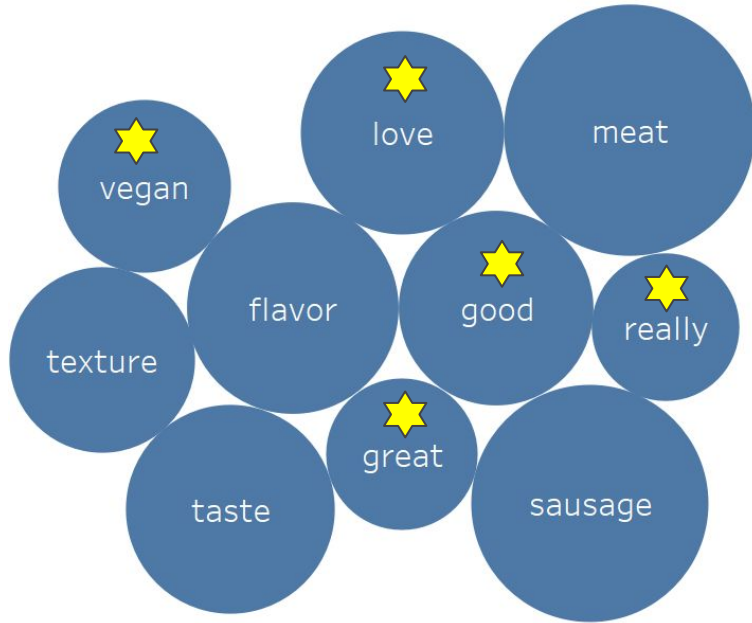
Visualizations

Fake Meat Products Reviewed

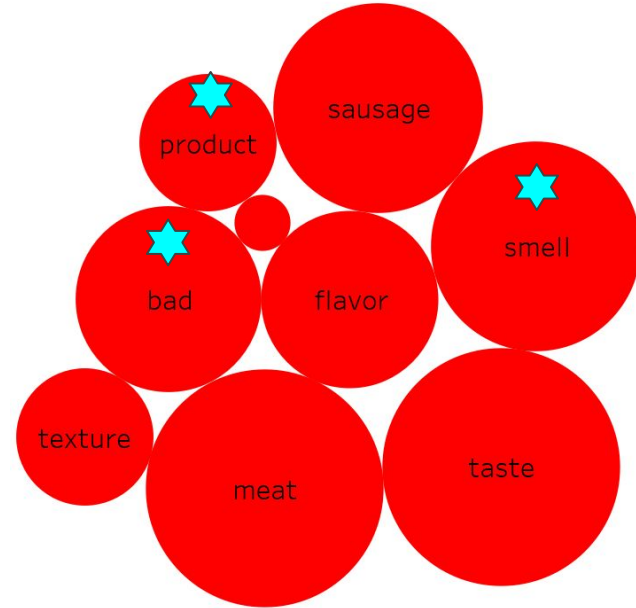
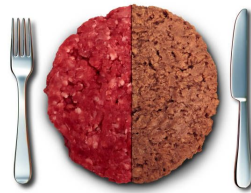


| Brand Name | Product Name |
|-------------|---|
| Beyond Meat | Beyond Meat Beyond Breakfast Sausage Plant-Based Breakfast Patties, Classic 7.4 oz Beyond Meat Beyond Sausage Plant-Based Dinner Sausage Links, Brat Original 14 oz Beyond Meat from PlantBased Frozen oz lb. Package, Ground Beef Substitute, 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 |

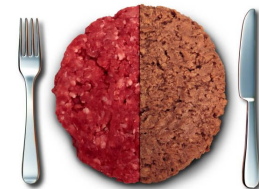
Fake Meat Brand Sentiment








★ Vegan, love, good, really, great

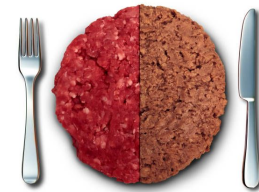


★ Product, bad, smell

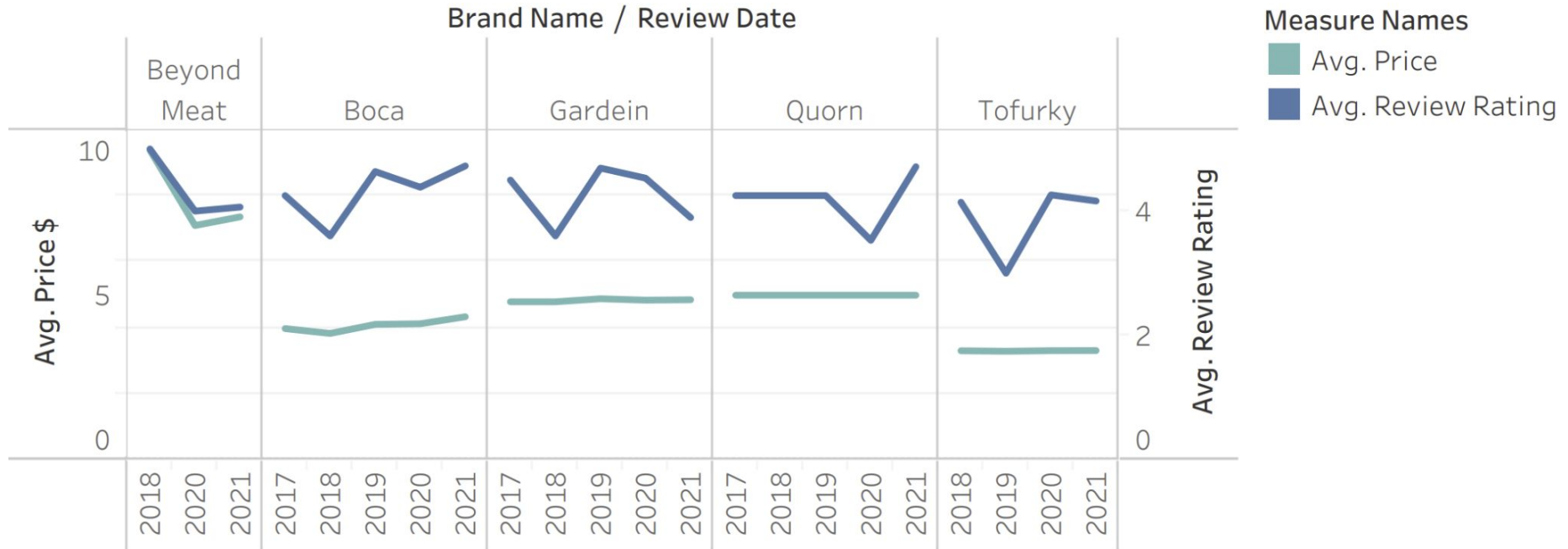


Keyword frequency by brand in **positive** reviews

| | Brand Name | flavor | good | great | love | meat | real | really | taste | texture | vegan |
|--|-------------|--------|------|-------|------|------|------|--------|-------|---------|-------|
|  | Beyond Meat | 36% | 25% | 22% | 33% | 29% | 27% | 22% | 28% | 33% | 21% |
|  | Boca | 6% | 12% | 11% | 6% | 2% | 10% | 9% | 9% | 1% | 10% |
|  | Gardein | 32% | 23% | 25% | 24% | 35% | 33% | 18% | 27% | 37% | 28% |
|  | Quorn | 11% | 16% | 18% | 19% | 20% | 22% | 29% | 24% | 18% | 10% |
|  | Tofurky | 15% | 23% | 24% | 18% | 14% | 8% | 21% | 12% | 11% | 32% |



Review v. Price by Brand per Year





Summary

Results

- ML model accuracy is 82%
- Stakeholders' prediction of customer sentiment
- Specific words and brand association

Limitations

- Rushed time frame
- Limited initial data set
- Algorithm does not consider the order of words

Future Analysis

- Interactive visualizations
- Larger data set
- Location data



Ways to Improve Project

- Search for a more accurate algorithm
- Have algorithm consider word order
- Host website on server

Interactive Website Demo

Project Home

Welcome to our NLP model display website!



Interactive Dashboard