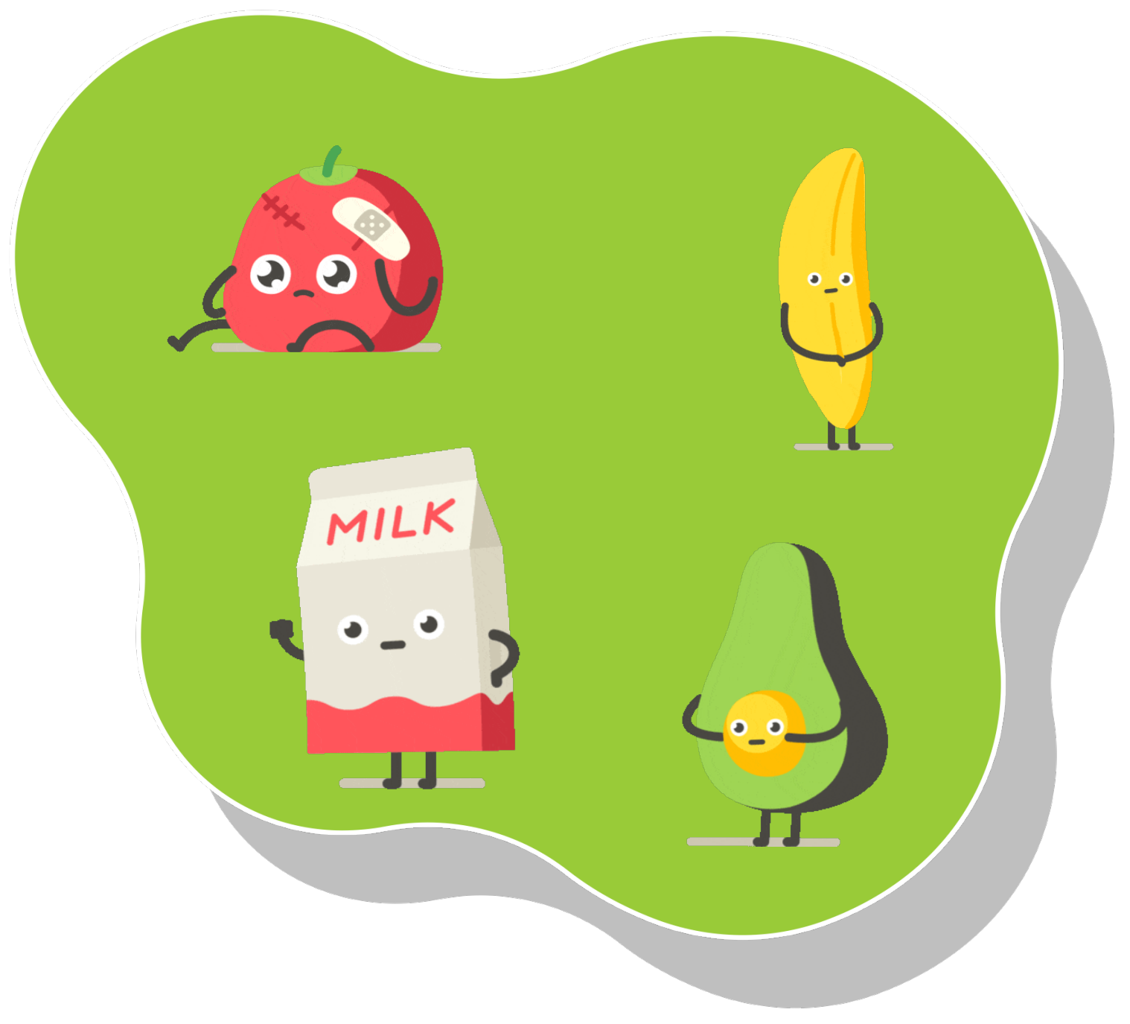
# The Moody Foodies

Georgia Tech - CSE 6242 Fall 2021

**Ibukun Ajayi • Monica Cheng • Justin Huang • Yang Luo • Justin Nand •**

*jnand@gatech.edu*



## In a Nutshell

The Moody Foodies have developed a data driven application that will allow users to enter a known food item and generate a set of recommendations with similar flavor profiles. Additionally, an included nutritional optimizer can select the best single alternative given a user's dietary constraints.

## The Incentive

60% of American adults have one or more diet-related chronic diseases according to the USDA. While much research has been dedicated to designing new food intake patterns which optimize the nutrient content, individual's eating habit may still be a major barrier to adapt new eating patterns. Thus, we attempt to combine nutrition optimization with food recommendation to provide more accessible, desirable, and healthy alternatives.

## Who would care?

Health-conscious individuals and those with specific dietary constraints who want to stay true to their taste preferences and enjoy meals with a particular flavor profile.

## Methods

### Data Gathering & Processing

Our project used two main datasets of 50000+ records. The first dataset includes recipes with annotated nutritional data, a subset of the Recipe 1M+ dataset [1]. The second dataset contains the molecular flavor profiles of known food ingredients from FlavorDB [2]. We used the ingredient names as the common key to join both datasets. The ingredient list/names in both datasets were tokenized in a consistent way before the joining operation as part of the data preparation stage.

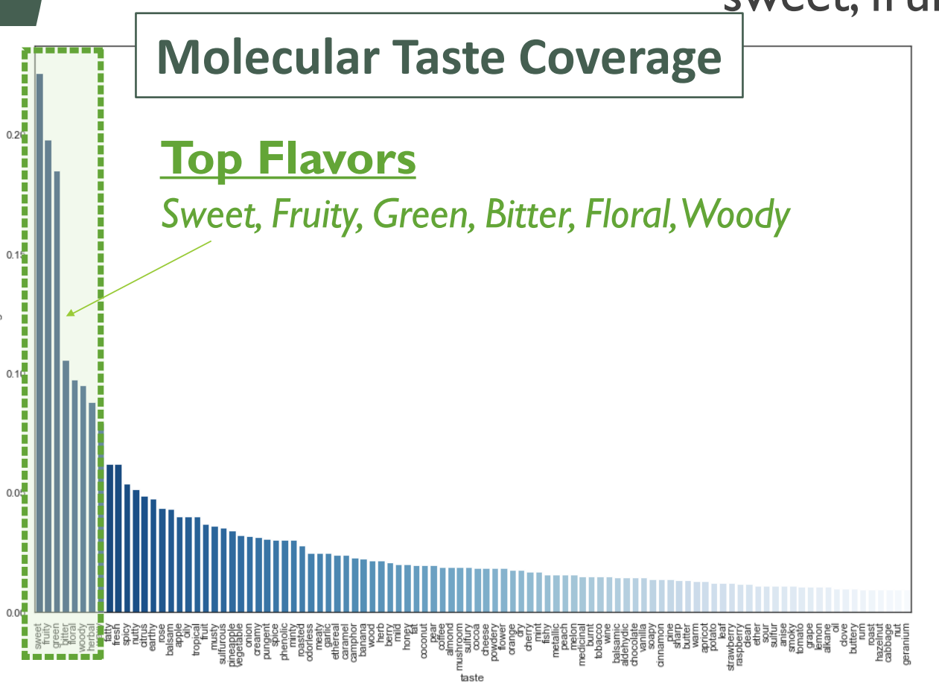
The resulting dataset is a consolidated recipe-flavor-molecule list that serve as input to the doc2vec model. This consists of records with one document per recipe and its transformed molecular data consisting of a list of flavor molecules from each component ingredient forming a single unique set with each molecule's presence recorded by its PubChem ID.

### Data Exploration

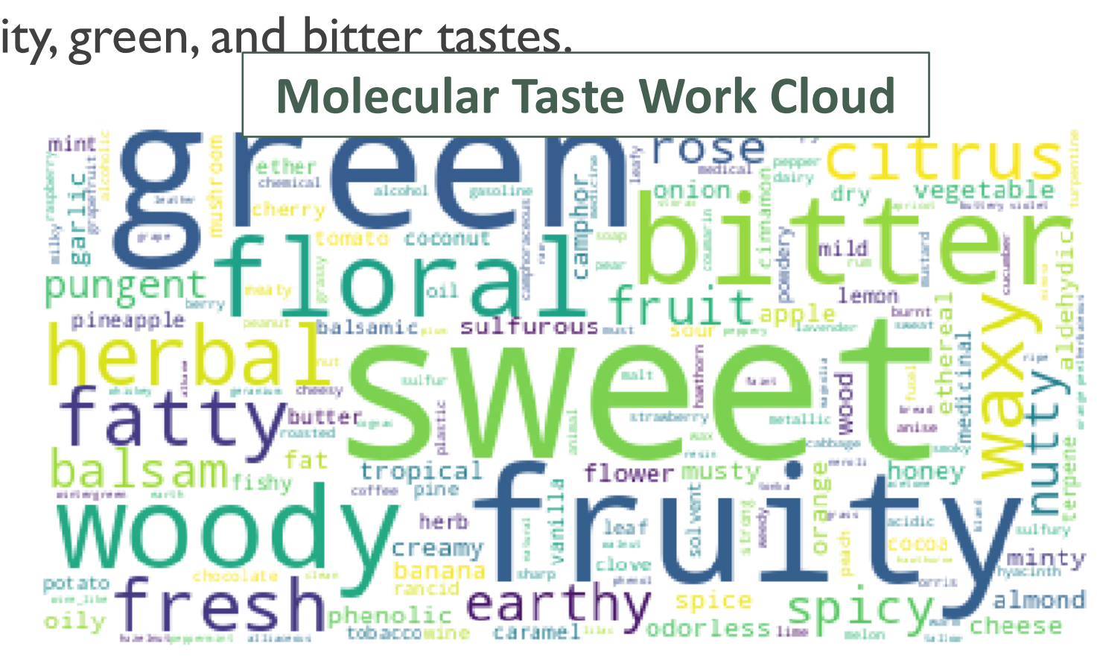
After computing the embeddings with doc2vec, exploratory analysis of the "flavor space" was done by applying "Uniform Manifold Approximation and Projection" (UMAP) to reduce the dimensionality of the space to three dimensions. The resulting points were then clustered using the "Density-Based Spatial Clustering of Applications with Noise" (DBSCAN) algorithm. We examined emergent relationships and normalized for unique tastes/notes by applying a Term Frequency – Inverse Document Frequency (TF-IDF) transformation.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

We also explored the spread of the taste annotations of the flavor molecule in our dataset. As shown below, the database is skewed towards a few common tastes i.e.. sweet, fruity, green, and bitter tastes.



Molecular Taste Coverage



Molecular Taste Work Cloud

The investigation of the dataset also revealed some quality issues such as inconsistencies in recipes records and nutritional values. Our methodology minimizes the impact of this data noise on the project output. However, more work will be required to address this challenge in future work.

### Nutrition Optimizer

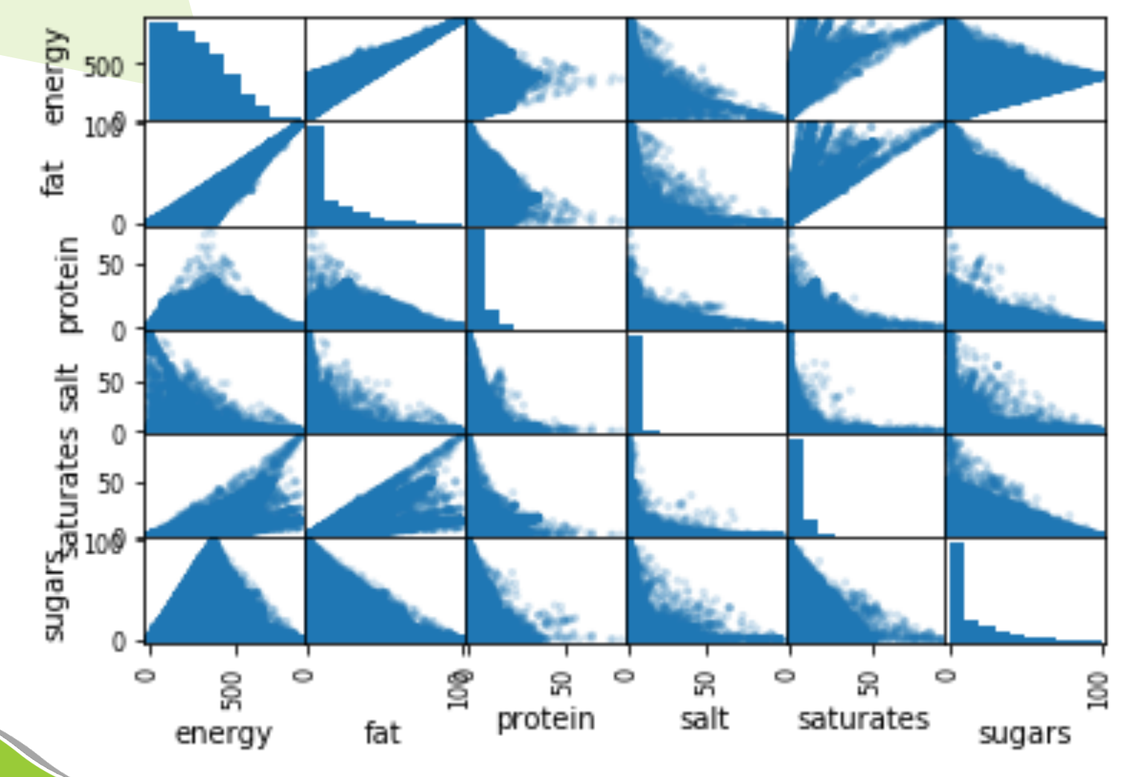
To provide a more refined nutritional selection, we implemented a process that would optimize the output recommendation based on users preferred nutritional preference. We approached this by following the daily nutrition consumption recommendation (Daily Value, DV) from FDA as the standard value and calculate the percent daily value (%DV) for each nutrient per 100g food. DVs are the recommended amounts of nutrients to consume or not to exceed each day, while %DV is how much a nutrient in a single serving of an individual packaged food or dietary supplement contributes to one's daily diet.

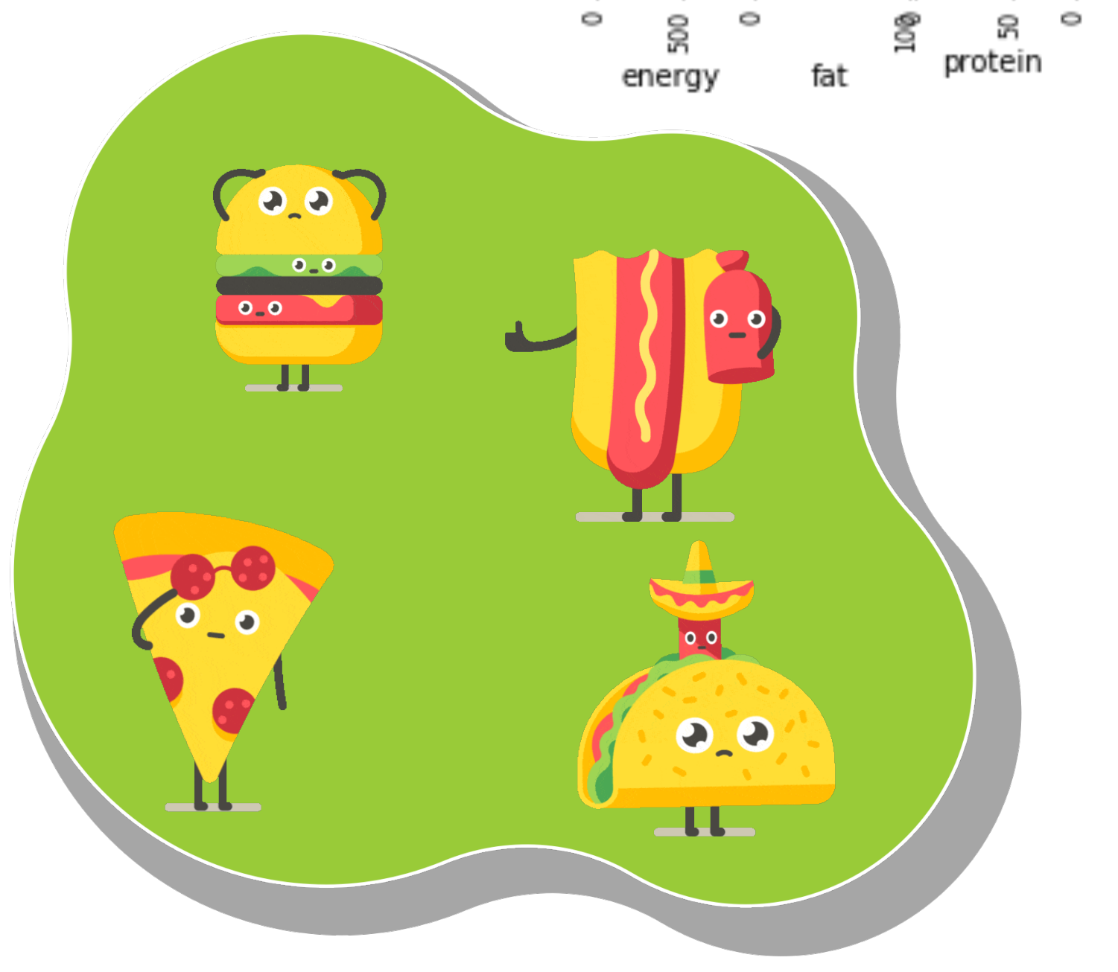
$$\%\mathrm{DV} = \frac{nutrition\ value\ per\ 100g\ of\ the\ food}{daily\ nutrition\ recommend\ value\ (DV)}$$

Based on the %DV, we compute a single healthiness score for each as:

$$\mathrm{H} = c_1\%DV_{Fat} + c_2\%DV_{protien} + c_3\%DV_{salt} + c_4\%DV_{saturate} + c_5\%DV_{sugar}$$

*Matrix Plot showing distribution of Nutritional Values*

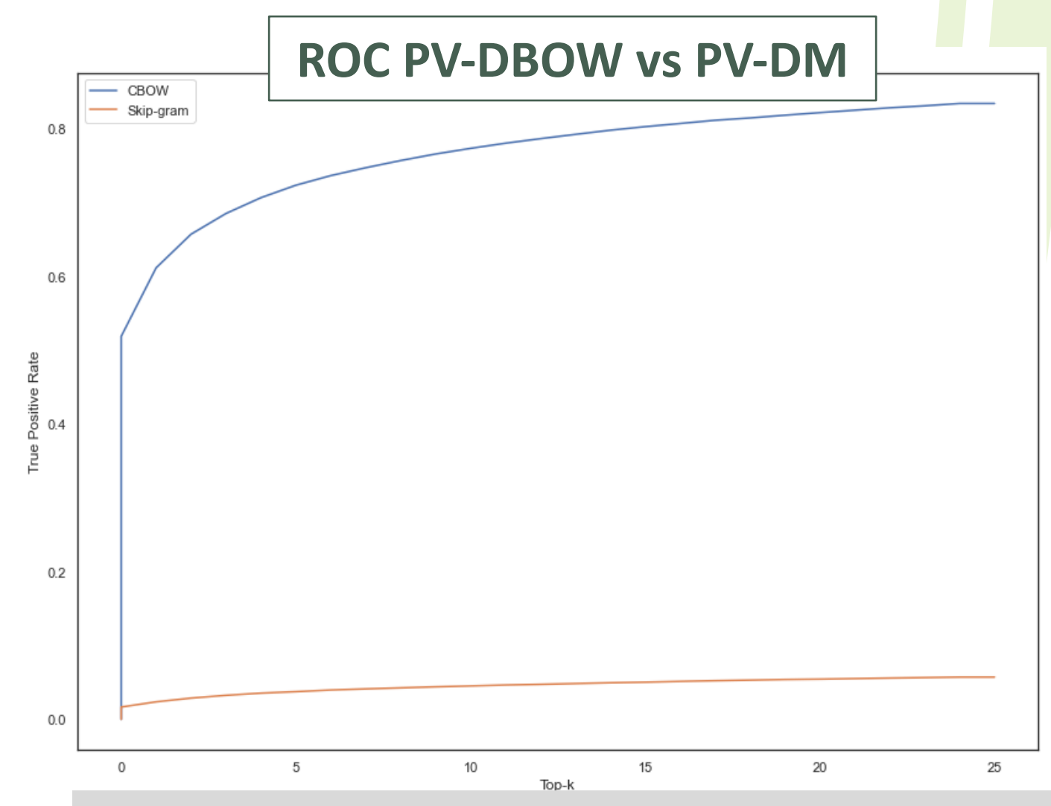




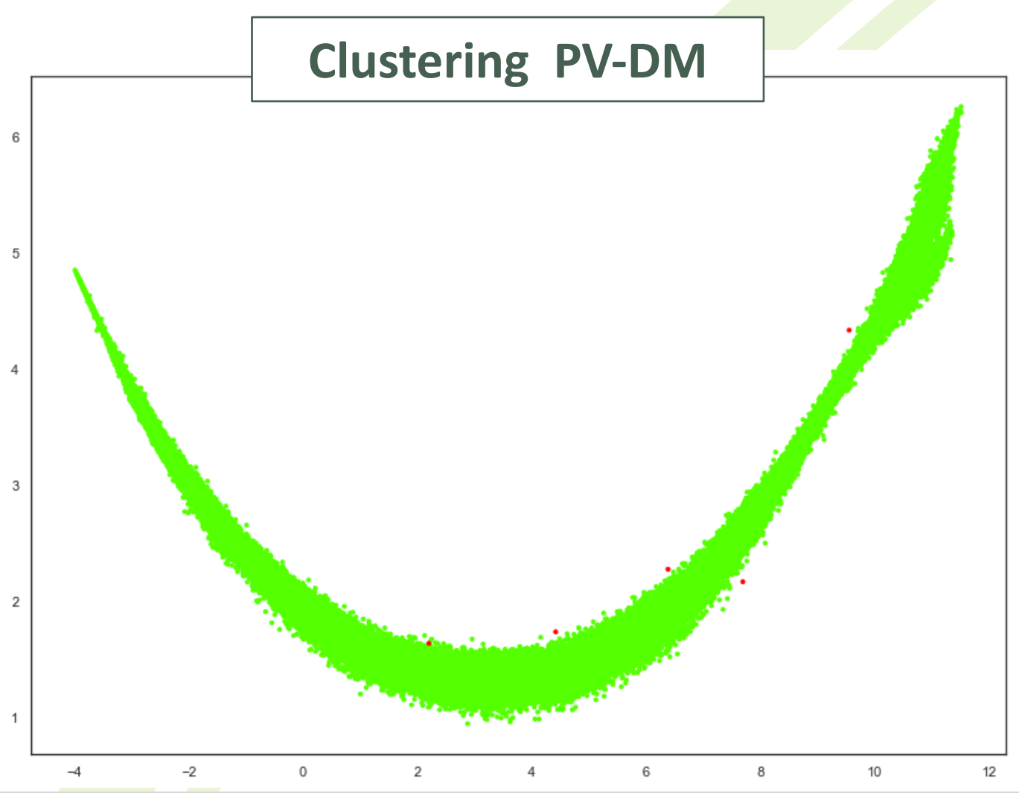
## Experiments

We evaluated different methods of string matching based off of Levenshtein distance. Our best method was able to find correct string matches at roughly 80%. Additionally, when extended to the top 3 matches, the method was able to find the correct string match 87% of the time, which suggests a high potential for improvement by additional string preprocessing.

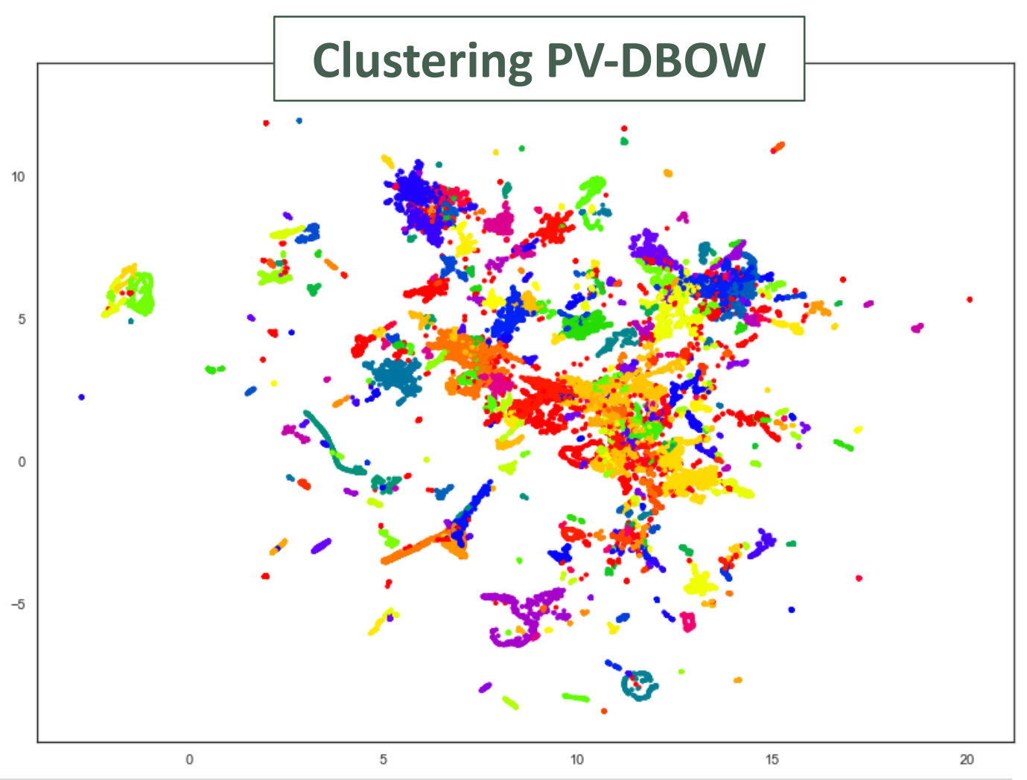
We also evaluated and trained two different word embedding models; the PV-DBOW model and the PV-DM model. PV-DBOW model predicts a single word from a fixed window size of context words, whereas PV-DM model predicts several context words from a single input word. We built and trained both models, and their embeddings qualitatively compared both to determine which "flavor space" most accurately positions similar tasting recipes near one another. We benchmarked both approaches on their accuracy and speed, choosing the technique that provided the highest quality result. Shown below are the performance results and graphical representation of the results from both models. Based on these results, the **PV-DBOW** model was selected for our work.



ROC PV-DBOW vs PV-DM



Clustering PV-DM



Clustering PV-DBOW

### Benchmark

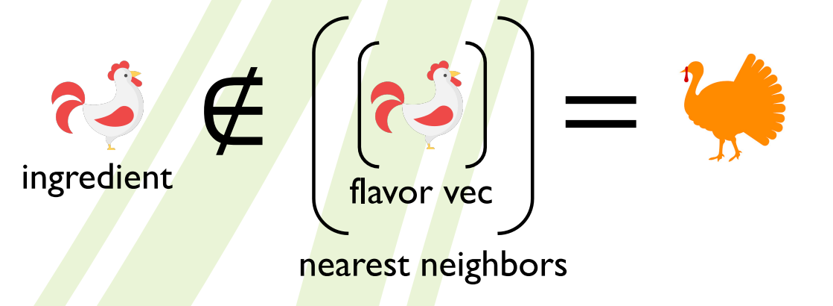
| | PV-DM | PV-DBOW |
|---|---|---|
| Accuracy | ❑ 6% in top 25 | ❑ 83% in top 25 |
| Speed | ❑ 55 CPU hours | ❑ 6 CPU mins |
| Cluster Analysis | ❑ 2 Clusters | ❑ 437 Clusters |

## Discussions and Findings

We conducted further deep dive into the model's predictions and our findings demonstrated the model's conceptual understanding of "flavor" based on 3 "computational cooking" tasks.

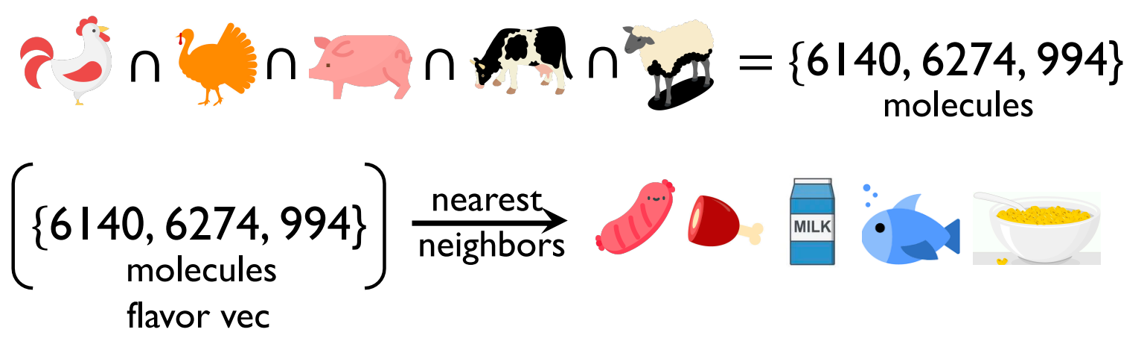
### "Tastes like Chicken"

Nearest neighbors of the flavor vector for the ingredient chicken was searched in the flavor space while explicitly excluding entries that listed chicken amongst their ingredients. The results identified Turkey as a suitable alternative.



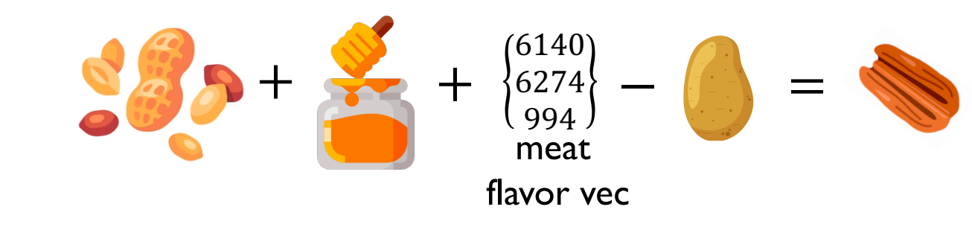
### What is "meat" flavor?

The intersection of the molecular composition of Chicken, Turkey, Pork, Beef, and Lamb identified three key molecules. A flavor vector was computed using only those three molecules, and nearest neighbors located. The resulted in many types of meat tasting foods: sausage, ham, buttermilk, whitefish, walrus, and a macaroni.



### Flavor math

This task attempted to "cook" a food based on vector math of desired flavor components. A query vector was computed using the flavor vectors of peanut, honey, meat flavor, and potato. The results surfaced the suggestion of "Honey-Glazed Pecans"

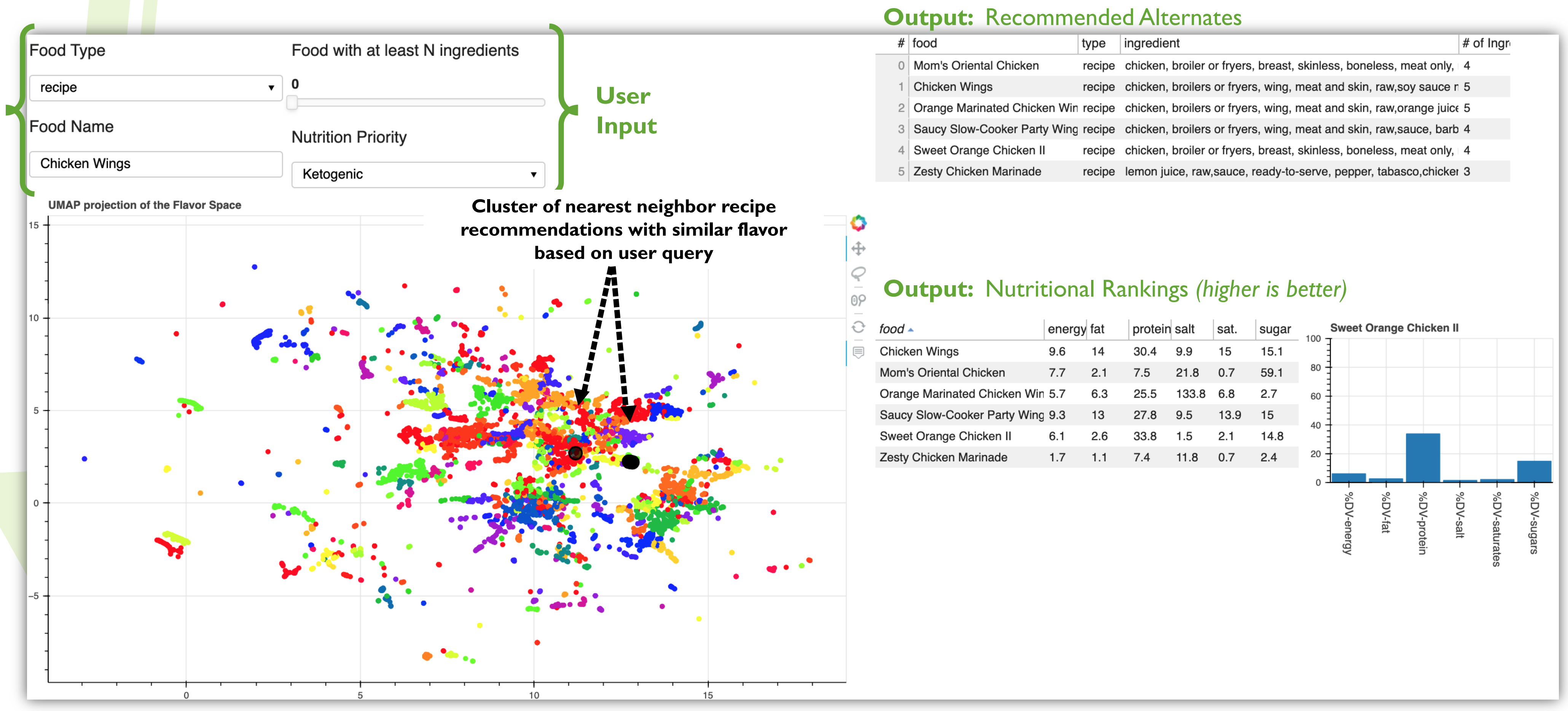


Despite the presence of some data noise, the embedding model was able to provide meaningful results to non-trivial flavor queries, indicating that the selected model was successful in learning some conceptual aspects of flavor. While many recommendations are intuitively identified as similar in flavor by a human observer, the most interesting results were those that were counter to human intuition. For example, upon investigating the molecular components of macaroni in the "meat flavor" task we noticed that the food item contained all three shared flavor molecules. So, we can infer that a craving for meat might be satiated by macaroni, a non-meat.

Further examination of why the PV-DM model failed by comparison, indicates that embeddings computed based on the context prediction task tended to produce embedding vectors that include information about the general probability of a molecule occurring. This effectively encoded a non-zero value for every molecule, whereas some molecules may not be present in a food. The dense and "complete" nature of the hidden layer produced embeddings in an almost uniform manifold, as observed after dimensional reduction and clustering. The resulting PV-DM space was tightly packed and lost its descriptive power to differentiate individual flavors from one another.

## User Experience

We developed a highly interactive user interface using ***Bokeh widgets*** that will allow users enter any food item of their choice and further specify nutritional preference. Our tool will generate a set of recommended healthy options of meals and recipes with a similar flavor profile. The nutritional optimizer selects the "best" option form these recommendations as a healthy alternative. As shown below, the tool will also present a graphical representation of how closely related the recommended alternative are, relative to the user's query input.



**User Input**

Food Type: recipe ▾

Food with at least N ingredients: 0

Food Name: Chicken Wings

Nutrition Priority: Ketogenic ▾

UMAP projection of the Flavor Space

**Cluster of nearest neighbor recipe recommendations with similar flavor based on user query**

**Output:** Recommended Alternates

| # | food | type | ingredient | # of Ingr |
|---|---|---|---|---|
| 0 | Mom's Oriental Chicken | recipe | chicken, broiler or fryers, breast, skinless, boneless, meat only, | 4 |
| 1 | Chicken Wings | recipe | chicken, broilers or fryers, wing, meat and skin, raw,soy sauce r | 5 |
| 2 | Orange Marinated Chicken Win | recipe | chicken, broilers or fryers, wing, meat and skin, raw,orange juice | 5 |
| 3 | Saucy Slow-Cooker Party Wing | recipe | chicken, broilers or fryers, wing, meat and skin, raw,sauce, barb | 4 |
| 4 | Sweet Orange Chicken II | recipe | chicken, broiler or fryers, breast, skinless, boneless, meat only, | 4 |
| 5 | Zesty Chicken Marinade | recipe | lemon juice, raw,sauce, ready-to-serve, pepper, tabasco,chicker | 3 |

**Output:** Nutritional Rankings *(higher is better)*

| food ▴ | energy | fat | protein | salt | sat. | sugar |
|---|---|---|---|---|---|---|
| Chicken Wings | 9.6 | 14 | 30.4 | 9.9 | 15 | 15.1 |
| Mom's Oriental Chicken | 7.7 | 2.1 | 7.5 | 21.8 | 0.7 | 59.1 |
| Orange Marinated Chicken Win | 5.7 | 6.3 | 25.5 | 133.8 | 6.8 | 2.7 |
| Saucy Slow-Cooker Party Wing | 9.3 | 13 | 27.8 | 9.5 | 13.9 | 15 |
| Sweet Orange Chicken II | 6.1 | 2.6 | 33.8 | 1.5 | 2.1 | 14.8 |
| Zesty Chicken Marinade | 1.7 | 1.1 | 7.4 | 11.8 | 0.7 | 2.4 |

Sweet Orange Chicken II

## Conclusion

This project leverages advances in NLP's ability to surface high level concepts that generalize relationships between words in a document corpus for understanding the psychological experience of "taste" using the molecular basis of flavor. We developed a model that was robust and able to effectively capture the conceptual idea of "taste" in terms of the molecular flavor composition of food. The most compelling feature of our model is ability to the determine and provide taste recommendations in cases where there are no shared ingredients between the query and its suggested alternatives – exemplified by the "tastes like chicken" test. Overall, this work presents an innovation method for determining healthy diet alternatives without compromising on the experience of eating. Further work might investigate the model's suggestion of flavor analogs.