

# Predicting Sentiment of Ingredients from Online Food Recipes

AUTHOR  
Seyoung Lee

## 1 Introduction

As the reliance on social networking platforms for food education grows, the use of online resources is likely shaping both taste preferences and perceptions of food (Reau 2013). Reau’s research emphasizes that exposure to food-related content on social media can influence dietary choices by highlighting trends, popular ingredients, and user-generated recommendations, which impact what individuals perceive as desirable or normative eating behaviors. In addition to this reliance, an increasing number of researchers have found that extended use of screen-media devices such as television has previously been linked to unhealthy dietary behaviors among children and adolescents. This includes an increased tendency to consume sugary and fatty foods, as well as a decreased intake of fruits and vegetables, which are determining factors in the onset of overweight and obesity (Sina et al. 2021).

The study examining the effect of digital media use on taste preferences revealed that the dietary behaviors and taste preferences of children and adolescents were influenced by increased exposure to digital media. Specifically, prolonged television viewing was linked to heightened sweet preferences across all groups. Moreover, a positive trend was observed between increased smartphone use and high sweet taste preferences in all participants, particularly in young children (Sina et al. 2021). Hence, this paper explores the sentiment expressed in comments on the popular recipe-sharing platform, allrecipes.com, and correlates it with the ingredients of different recipes. This correlation aims to offer insights into taste preferences and potentially identify popular flavor profiles (Allrecipes 2023).

Simultaneously, our aim is to investigate the trends reflected in these ingredients and explore how perceptions evolve based on the nutritional information provided in these recipes. By delving into the nutritional aspects, we seek to gain insight into the public’s interests and preferences regarding health-conscious eating habits, ingredient awareness, and dietary choices. This comprehensive analysis allows us to uncover nuanced patterns in culinary preferences and potentially identify emerging trends in nutritional awareness among consumers.

By analyzing shifts in sentiment by ingredients, researchers can identify food categories that are gaining popularity due to perceived tastiness, health benefits, or other factors. Hence, this paper aims to explore the variations in sentiment by ingredients by exploring the following research questions:

- Q1: Is there a significant correlation between individual ingredients and the sentiment of recipes in online comments?
- Q2: Is there a significant correlation between ingredient pairs and sentiment of recipes in online comments?
- Q3: Can we predict the popularity and sentiment of recipes based on their nutritional content?

### 1.1 Exploratory Data Analysis

For each recipe, details encompassing the number of reviews, comments, average sentiment scores, and nutritional information, including metrics like calories, fat, carbohydrates, and protein content measured in grams, were compiled.

We analyzed a total of 940 recipes sourced from allrecipes.com, a prominent food blog comprising a vast collection of user-submitted recipes across various cuisines and dietary preferences. Allrecipes.com is a leading platform in Cooking & Recipes, reaching over 100 million total visits (“Allrecipes.com competitors - top sites like allrecipes.com | similarweb” n.d.). Our dataset of recipes contains a total of 9,437,608 tokens from 101,181 individual comments. Each comment’s sentiment was analyzed using the Textblob library, which assigns a sentiment score on a scale from -1 to 1. A score of -1 indicates the most negative sentiment, 0 indicates neutrality, and +1 represents the most positive sentiment (Loria n.d.). The following figure illustrates the distribution of sentiment scores across the comments.

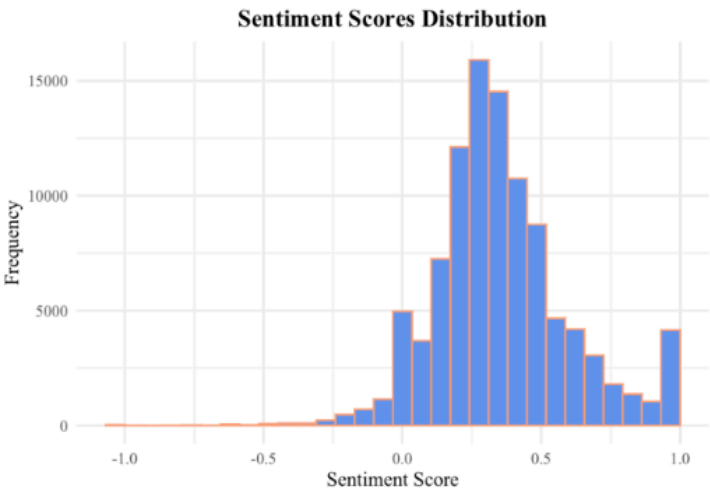


Figure 1: Distribution of Average Sentiment Scores

The sentiment scores display a clear left skew. Even after normalizing the histogram (Figure 1), the distribution remains left-skewed, indicating an inherent positivity bias in the sentiment data, and suggesting a general tendency toward positive sentiment in food recipe reviews.

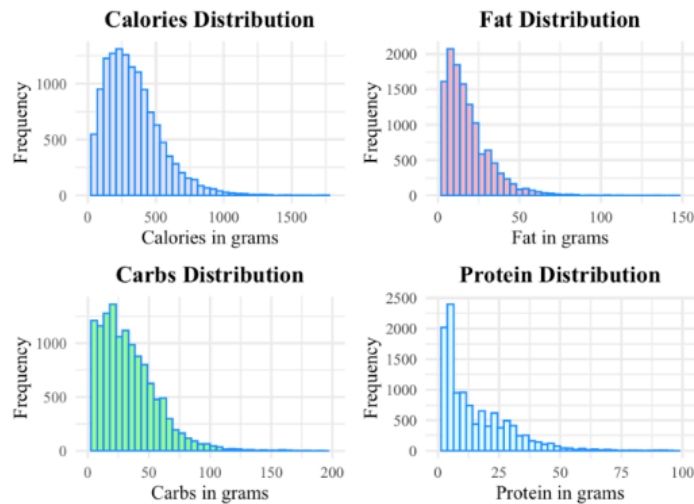


Figure 2: Nutrition Summaries Distribution

In Figure 2, we observe the distribution of nutritional attributes across all dishes. The distinct right skew in all distributions indicates a trend towards higher values and can also suggest the presence of numerous outliers within the higher range of the distribution. This suggests that certain recipes tend to exhibit elevated nutritional content in terms of calories, fat, carbohydrates, and proteins, while the majority fall within typical ranges.

## 2 Methods

### 2.0.1 Phase 1: Individual Ingredient Sentiment

We opted to employ a simple linear regression model with individual ingredients as the predictor variable and the average sentiment as the outcome variable. This method aligns with the research objective of examining the relationship between specific ingredients and the sentiment expressed in recipe comments.

In this context, each ingredient serves as an independent variable, which allows us to assess the impact of individual ingredients on the overall sentiment of recipe comments. We can identify which ingredients tend to have more of a positive or negative sentiment from reviewers by analyzing the regression coefficients and statistical significance associated with each ingredient.

Most Statistically Significant Ingredients on Sentiment			
The six most significant p-values are presented.			
Ingredient	Estimate	Confidence Interval	P-value
white sugar	0.143	0.03, 0.25	0.012
chili powder	-0.533	-0.96, -0.10	0.015
lemon juice	0.360	0.03, 0.69	0.031
ground beef	-0.342	-0.67, -0.02	0.040
sour cream	-0.262	-0.53, 0.01	0.057
ground cumin	-0.333	-0.68, 0.01	0.060

Table 1: Ingredient sentiment estimates

When examining the sentiment estimates of individual ingredients, we find a total of four ingredients that yield statistically significant p-values: white sugar, chili powder, lemon juice, and ground beef. Interestingly, white sugar and lemon juice, typically associated with sweeter recipes and particularly in baking, exhibit statistically positive sentiment (white sugar: 95% CI[0.03, 0.25],  $p=0.012$ ; lemon juice: 95% CI[0.03, 0.69],  $p=0.031$ ). Conversely, ingredients associated with different taste preferences such as umami or spiciness demonstrate statistically significant negative sentiment (chili powder: 95% CI[-0.96, -0.10],  $p=0.015$ ; ground beef: 95% CI[-0.67, -0.02],  $p=0.040$ ).

### 2.0.2 Phase 2: Ingredient Pair Sentiment

As only a small fraction of individual ingredients demonstrate significance, exploring groups, or in this case, ingredient pairs, may offer more significant sentiment outcomes. When analyzing individual ingredients, it may be difficult to isolate a single ingredient that yields statistically significant sentiment results. Hence, investigating the sentiment expressed by ingredient pairs can be more relevant to the goal of predicting sentiment by ingredients.

Statistically Significant Ingredient Pairs on Sentiment			
The 10 most significant p-values are presented.			
Ingredient	Estimate	Confidence Interval	P-value
chili powder, salt	-0.757	-1.3, -0.23	0.005
ground black pepper, ground cumin	-0.819	-1.4, -0.23	0.007
ground black pepper, sour cream	-0.580	-1.1, -0.07	0.025
lemon juice, white sugar	0.526	0.05, 1.0	0.031
chili powder, ground cumin	-0.609	-1.2, -0.04	0.035
cayenne pepper, olive oil	-0.599	-1.2, -0.03	0.039
butter, cornstarch	0.623	0.03, 1.2	0.039
vegetable oil, white sugar	0.332	0.00, 0.66	0.047
green bell pepper, onion	-0.538	-1.1, 0.01	0.053
ground black pepper, onion	-0.319	-0.65, 0.01	0.056

Table 2: Ingredient pair sentiment estimates

Table 2 reveals that the first nine ingredient pairs exhibit significance, while only four of the individual ingredients exhibited significance. This suggests that combinations of ingredients provide a more nuanced perspective on the popularity or public sentiment towards a recipe. Consistent with earlier findings, ingredient pairs commonly linked with baking (e.g., lemon juice-white sugar, butter-cornstarch, vegetable oil-white sugar) tend to evoke positive sentiment. Conversely, pairs associated with savory or spicier dishes (e.g., chili powder-salt, ground black pepper-ground cumin, ground black pepper-sour cream, chili powder-ground cumin, cayenne pepper-olive oil) yield more negative sentiment.

### 2.0.3 Phase 3: Popularity and Sentiment by Nutritional Content

Having potentially identified a pattern regarding the sentiment associated with specific taste preferences in recipes, we extend our analysis to explore the individual nutritional profiles of these recipes. By examining the recipes' nutritional content, we aim to uncover patterns in the recipes that garner the most popularity. In this context, we utilize the number of reviews as a metric for popularity and consider the sentiment of each dish, utilizing the average sentiment scores, to find potential correlations.

Because there is potential collinearity between the different nutritional attributes in the recipes, we opted to use a generalized additive model (GAM) to explore the relationship between nutritional content and sentiment/popularity. GAMs can accommodate for more complex and non-linear patterns without explicit transformations or model assumptions. More specifically, it can handle interactions between the predictors, while multivariate linear models do not account for collinearity.

GAM Regressing # of Reviews and Sentiment on Nutrition				
Nutritional Content	Sum Squared (Reviews)	P-Value (Reviews)	Sum Squared (Sentiment)	P-Value (Sentiment)
Carbs	2.539876e+06	0.00696	0.00127	0.76857
Calories	2.548550e+04	0.78686	0.00728	0.48129
Fat	6.741522e+05	0.16434	0.06312	0.03833
Protein	8.413360e+00	0.99608	0.00113	0.78145

Table 3: Generalized Additive Model Estimates and Significance

Table 3 examines the different generalized additive models that regresses the number of reviews (as a metric of popularity) and average sentiment on nutritional content: `Number of Reviews ~ Carbs + Calories + Fat + Protein` and `Average Sentiment Score ~ Calories + Fat + Protein + Carbs`. In our analysis, we find that among the predictor variables considered, only `Carbs` yielded significance in the generalized additive model predicting the number of reviews. In the GAM model predicting average sentiment, only `Fat` exhibited significance as a predictor.

The statistical significance of `Carbs` in predicting the number of reviews suggests that variations in carbohydrate content have a discernible impact on the number of reviews a recipe receives. This supports our earlier expectations that baking recipes tend to elicit more positive reactions in both individual ingredients and ingredient pairs compared to other types of recipes. We also see that `Fat` content is significant in predicting sentiment.

Additionally, the sum squared estimates offer insights into the smoothness of the relationships captured by each predictor variable in the GAM models. In other words, the sum squared values tend to decrease on the regression with the number of reviews. This suggests a more linear approximation between the outcome variable, the number of reviews, and nutritional content. On the other hand, sentiment exhibits a more complex, non-linear relationship with nutritional content.

## 3 Results

During the exploratory phase of our research, we noted significant skewness in both sentiment and nutritional content. Sentiment tended to exhibit a pronounced left skew, while nutritional aspects such as carbs, fat, protein, and calories displayed a notable right skew. This skewness shows that linear regression may not be a reliable method to assess the relationship between sentiment and nutritional content for Q3. Therefore, alternative modeling approaches are used to approach the relationship between sentiment and nutritional content.

Addressing the research question 1, it becomes apparent that only a handful of ingredients exhibit a significant association with sentiment. The ingredients white sugar and lemon juice are linked to positive sentiment, while ingredients typically associated with spicier or umami taste preferences, such as chili powder and ground beef, are generally associated with negative sentiment.

In response to research question 2, we observe a greater number of significant associations in ingredient pairs compared to those identified in individual ingredients. This suggests that predicting sentiment becomes more reliable when considering ingredient combinations. Additionally, we can infer that sentiment prediction tends to be more favorable when analyzing groups with a larger number of ingredients. Moreover, there appears to be a stronger positive association towards sweet taste preferences, especially ingredients commonly used in baking.

To investigate our results outlined in questions 1 and 2, which propose that distinct taste preferences correlate with significantly varied sentiment, we also examine the nutritional composition of each recipe. Using a generalized additive model regression to analyze the influence of individual nutritional components —calories, fat, carbs, and protein in grams—on both the number of reviews and sentiment, we identify statistically significant effects of different types of nutritional content on either the number of reviews or sentiment. Specifically, we observe that carbohydrate in grams impacts the number of reviews, while fat content in grams affects the average sentiment.

## 4 Discussion

To further examine the takeaways from the study, we revisit the study revealing the associations between children's television viewing habits and their food and beverage preferences (Sina et al. 2021). The study primarily highlights positive correlations between digital media usage and preferences for sweet, fatty, and salty foods, alongside negative associations with preferences for bitter foods. This underscores the influence of digital media exposure on taste preferences.

Although our research did not directly measure digital media usage, we utilized a popular online recipe sharing platform to gauge the sentiment and popularity of various recipes. Initial analyses of ingredients and sentiment revealed a positive sentiment towards commonly used baking ingredients and a negative sentiment towards spicy ingredients. Further investigation into sentiment and review counts showed that carbohydrate content impacts recipe popularity, while fat content affects recipe sentiment. Additionally, sentiment and popularity were specifically influenced by carbohydrate and fat content rather than calories and protein, suggesting a potential link between these findings and taste preferences observed in the digital media intake study.

### 4.0.1 Limitations

One evident limitation in our data is the skewness present in our variable distribution. This skewness affects our ability to investigate linear relationships and direct correlations between variables effectively. Moreover, we cannot establish causal relationships or determine which predictors are most strongly correlated with specific response variables. Exploring correlations among variables can illustrate the true relationships between sentiment and ingredients, or sentiment and nutritional content.

Additionally, we mainly focused on nutritional facts for recipe assessment, which can overlook the categorizations in food, such as snacks, desserts, healthy, vegan, etc. Exploring these categories can provide deeper insights into how perceptions of different food categories vary among individuals.

### 4.0.2 Future work

In the case of our analysis, this research could be improved by utilizing a more nuanced regression analysis, as the relationship between sentiment, the number of reviews, and nutritional facts may be complex and non-parametric. Furthermore, our analysis was limited to individual ingredients and ingredient pairs. Our comparison revealed a greater number of statistically significant sentiment estimates with ingredient pairs. Examining groups of ingredients, rather than solely focusing on pairs, could provide deeper insights into perceptions of different food categories. Additionally, in exploring the impact of increased digital media use, it could be beneficial to investigate other social media platforms, such as TikTok or Instagram, that have growing food and beverage content.

## 5 References

- Allrecipes. 2023. "About Us." <https://www.allrecipes.com/about-us-6648102>.  
"Allrecipes.com competitors - top sites like allrecipes.com | similarweb." n.d. <https://www.similarweb.com/website/allrecipes.com/competitors/>; SimilarWeb. Loria, S. n.d. "Simplified Text Processing." <https://textblob.readthedocs.io/en/dev/>; TextBlob.  
Reau, Brenda J. 2013. "Social Media Has the Opportunity to Influence Food Preferences." [https://www.canr.msu.edu/news/social\\_media\\_has\\_the\\_opportunity\\_to\\_influence\\_food\\_preferences#:~:text=Market%20research%20shows%20that%20co](https://www.canr.msu.edu/news/social_media_has_the_opportunity_to_influence_food_preferences#:~:text=Market%20research%20shows%20that%20co)  
Sina, E., C. Buck, W. Ahrens, S. De Henauw, H. Jilani, L. Lissner, D. Molnár, et al. 2021. "Digital Media Use in Association with Sensory Taste Preferences in European Children and Adolescents-Results from the i.family Study." *Foods* 10 (2): 377. <https://doi.org/10.3390/foods10020377>.