

# Text Analysis Report

## Overview

This report outlines the implementation of sentiment analysis on food reviews using a Multinomial Naive Bayes model. The dataset utilized for this analysis is the Amazon Fine Food Reviews dataset, which consists of reviews labeled as either positive or negative. The objective of this analysis is to classify reviews based on their sentiment.

## Dataset

The dataset consists of two main columns:

- **review:** Contains the text of the food reviews.
- **sentiment:** Indicates whether the review is positive or negative.

The dataset was loaded from a CSV file, and the sentiment labels were preprocessed to convert them into binary format:

- **1** for positive reviews
- **0** for negative reviews

## Data Preprocessing

1. **Loading the Dataset:** The dataset was loaded using pandas from a CSV file located at E:\AI models\Food Reviews\Amazon Fine Food Reviews.csv.

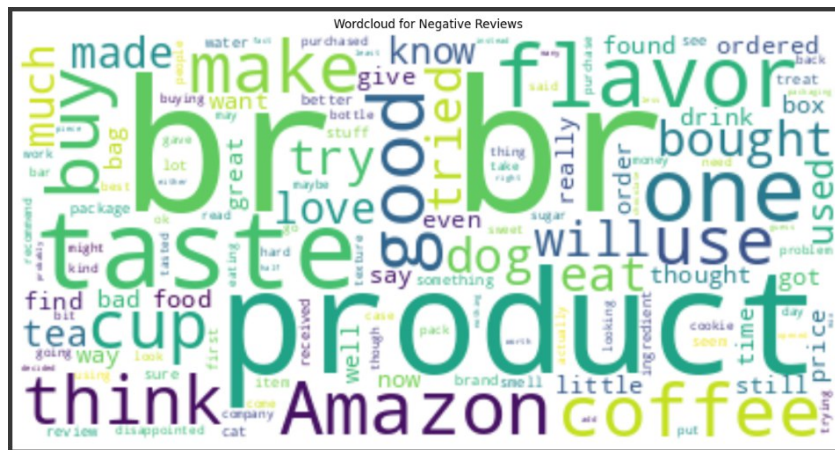
### 2. Cloud Visualizations:

The following word clouds illustrate the most frequent words in the dataset before any text preprocessing. The first cloud represents positive words, while the second cloud displays negative words. These visualizations highlight the themes and sentiments expressed in the reviews.

- **Positive Words Cloud**



- **Negative Words Cloud**



3. **Label Encoding:** The sentiment labels were transformed from categorical values ('positive' and 'negative') to numerical values (1 and 0) using the map function.
4. **Splitting the Data:** The dataset was split into training and testing sets using train\_test\_split, with 80% of the data used for training and 20% for testing. This stratified split ensures that both positive and negative reviews are represented proportionally in both sets.

## Text Vectorization

To convert the text data into a format suitable for machine learning models, we used the Count Vectorization technique. This method helps in transforming the reviews into numerical features while considering the frequency of words relative to the entire dataset. The vectorization process involved:

- Fitting the Count Vectorizer on the training data.
- Transforming both the training and testing datasets

into feature matrices, utilizing a vocabulary derived from the Amazon Fine Food Reviews dataset.

## Model Training

A Multinomial Naive Bayes model was trained on the features derived from the Count Vectorization of the training set. This model was chosen due to its effectiveness in handling text classification tasks, particularly when dealing with word frequency. The training process involved fitting the model to the training data (features and labels) obtained from the Amazon Fine Food Reviews dataset.

## Predictions and Evaluation

After training the model, it was used to make predictions on the testing set. The performance of the model was evaluated using the following metrics:

- **Accuracy:** The proportion of correctly predicted reviews out of the total reviews in the test set.
- **Classification Report:** This report includes precision, recall, and F1-score for both classes (positive and negative), providing a comprehensive view of the model's performance, based on the predictions made by the Multinomial Naive Bayes model.

## Results

The following results were obtained after evaluating the model:

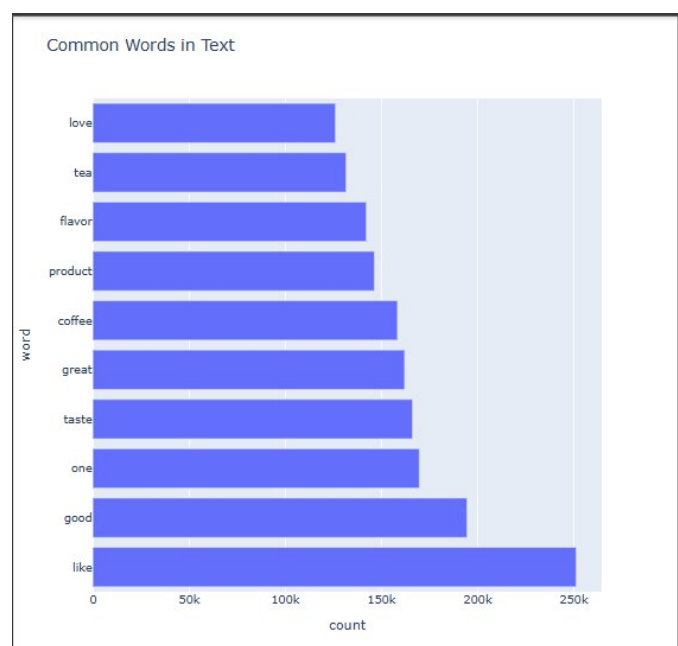
- **Accuracy:** 88%
- **Classification Report:**

```
Accuracy: 0.88
Classification Report:
              precision    recall  f1-score   support

     0         0.75      0.66      0.70      24906
     1         0.91      0.94      0.92      88785

 accuracy          0.88      0.88      0.88      113691
 macro avg         0.83      0.80      0.81      113691
 weighted avg      0.87      0.88      0.87      113691
```

“This chart shows the frequency of the most common words in the reviews, providing insight into key themes in customer opinions.”



## Model Saving

The trained model was saved to a file using the joblib library, enabling easy loading and usage in future applications without the need to retrain the model.

## Conclusion

The Multinomial Naive Bayes model effectively classified food reviews into positive and negative sentiments. The results indicate that the model performs well in distinguishing between the two classes, with an accuracy of 0.88.

Future work may involve exploring more complex models, such as neural networks or ensemble methods, to further improve accuracy and performance.