Sentiment Analysis on Restaurant Reviews

Group 9

TABLE OF CONTENTS

01

OVERVIEW

Project Introduction

Overview

We are implementing Sentiment Analysis and Machine Learning techniques to determine the probability of a restaurant receiving positive, negative or neutral reviews.

Additionally, we have gone a step further and applied Prompt Engineering methods to identify whether a restaurant fulfills the criteria of being awarded a Michelin star or not.

Objective

○ Identifying underlying patterns and factors contributing to a restaurant's recognition and success

○ Demonstrating the capabilities of NLP in real-world applications, specifically in the culinary and hospitality industry

02

METHODOLOGY

DATA DESCRIPTION

Data Input: Restaurant Reviews & Business Dataset (Source: Yelp)

Features:

- The Yelp dataset comprises over 6 million reviews from more than 150,000 businesses and was collected for academic and NLP research.
- We have extracted 100,000 records from the dataset to use in our project.
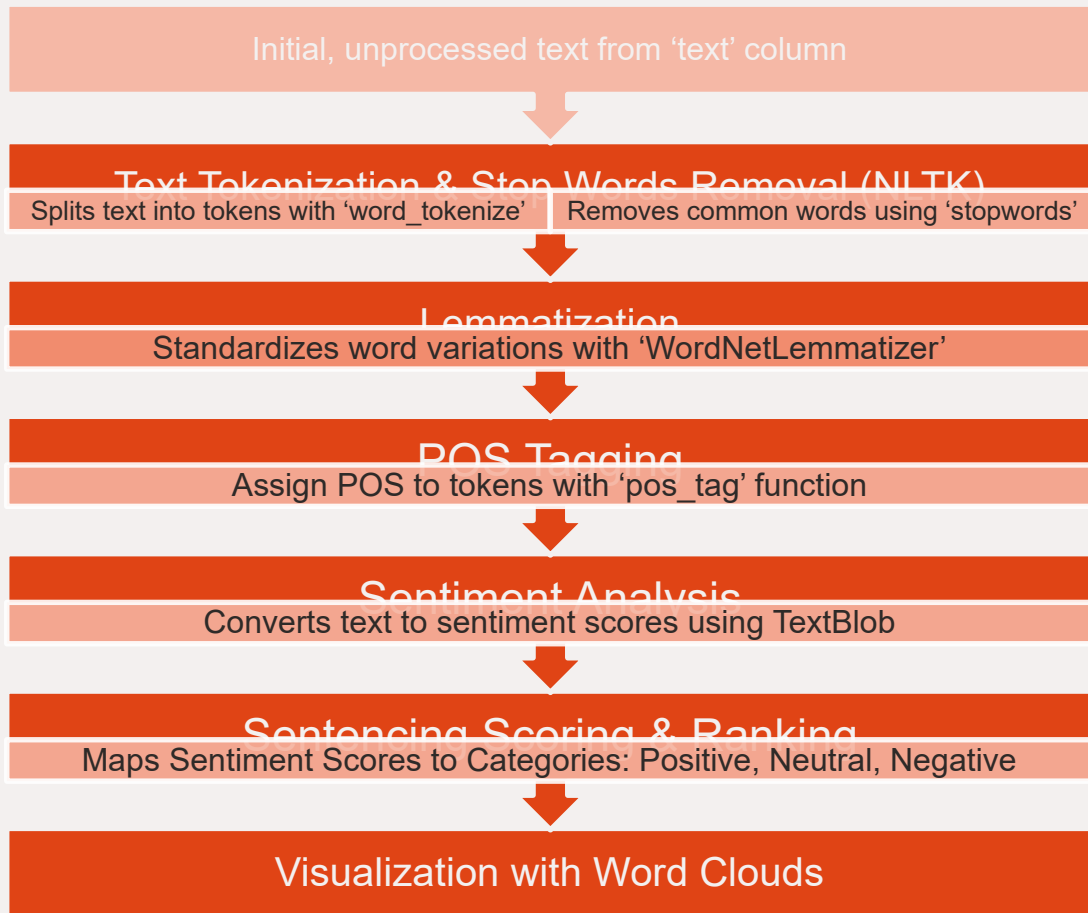
6,990,280 reviews

150,346 businesses

200,100 pictures

11 metropolitan areas

DATA PREPARATION

Initial, unprocessed text from 'text' column

Text Tokenization & Stop Words Removal (NLTK)

Splits text into tokens with 'word_tokenize'   Removes common words using 'stopwords'

Lemmatization

Standardizes word variations with 'WordNetLemmatizer'

POS Tagging

Assign POS to tokens with 'pos_tag' function

Sentiment Analysis

Converts text to sentiment scores using TextBlob

Sentencing Scoring & Ranking

Maps Sentiment Scores to Categories: Positive, Neutral, Negative

Visualization with Word Clouds

03
MODEL DETAILS

MODEL DETAILS

Data Split:
- Training: 95%, Testing: 5%
- Transforming: TF-IDF Vectorizer

Models
- Bernoulli Naïve Bayes Classifier
- Logistic Regressions
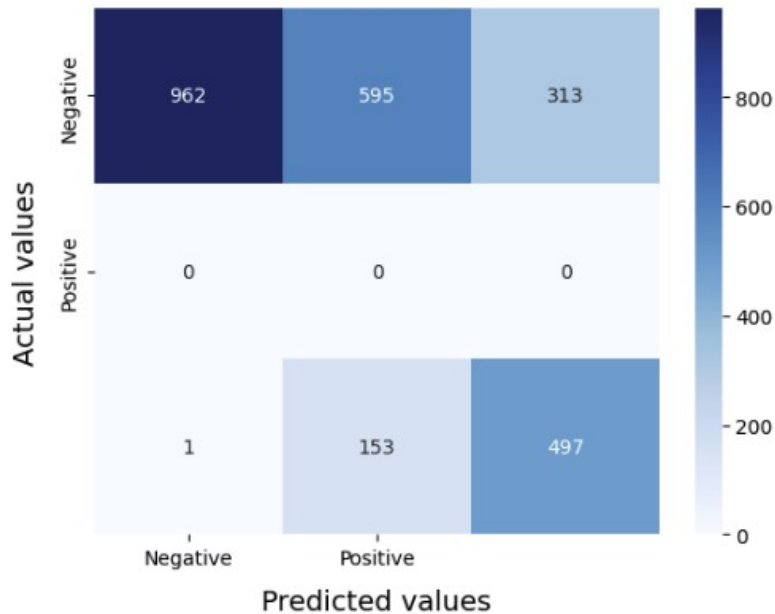
Model Evaluations
- Accuracy Score
- Confusion Matrix with Plot
- ROC-AUC Curve

```python
BNBmodel = BernoulliNB()
BNBmodel.fit(X_train, y_train)
model_Evaluate(BNBmodel, X_test, y_test)
```

```python
LRmodel = LogisticRegression()
LRmodel.fit(X_train, y_train)
model_Evaluate(LRmodel, X_test, y_test)
```
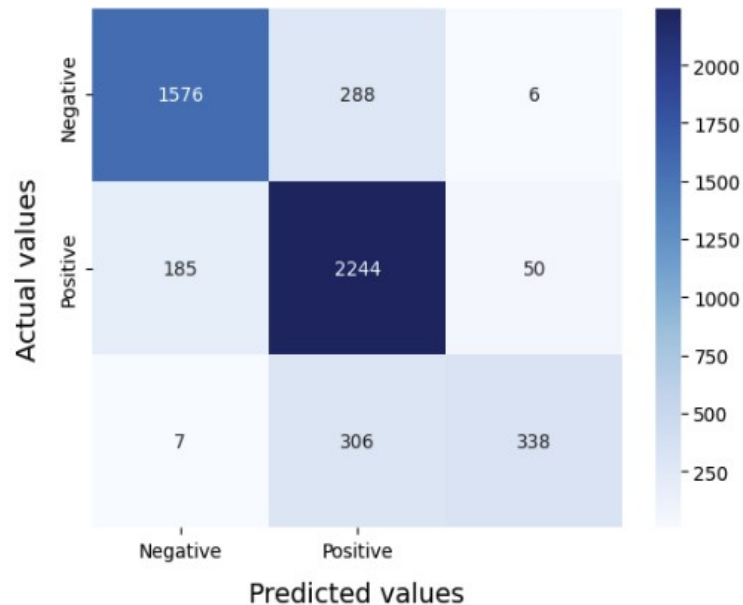
MODEL RESULTS

Bernoulli Naive Bayes Classifier
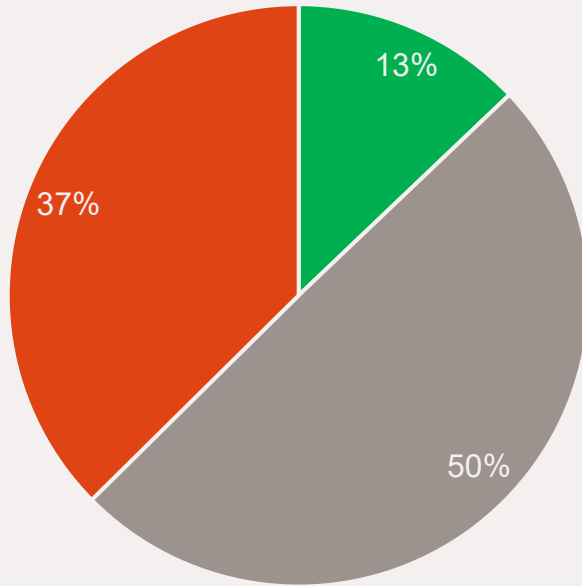
Logistic Regression

04

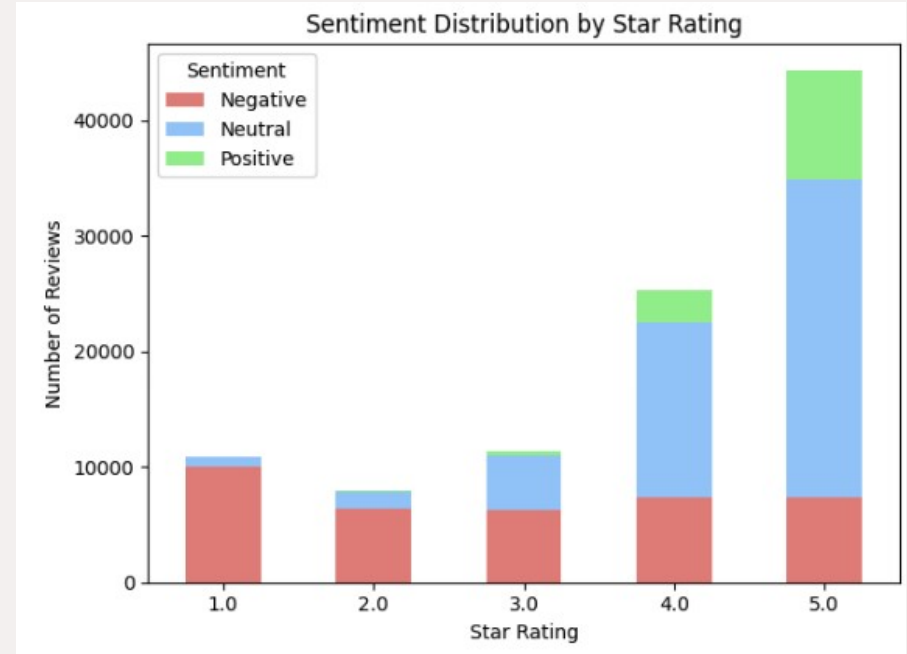## ANALYSIS

INSIGHTS & RESULTS

Sentiment Distribution by sentiment scores

Sentiment distribution by Star Ratings

INSIGHTS & RESULTS

Positive Tweets

Negative Tweets
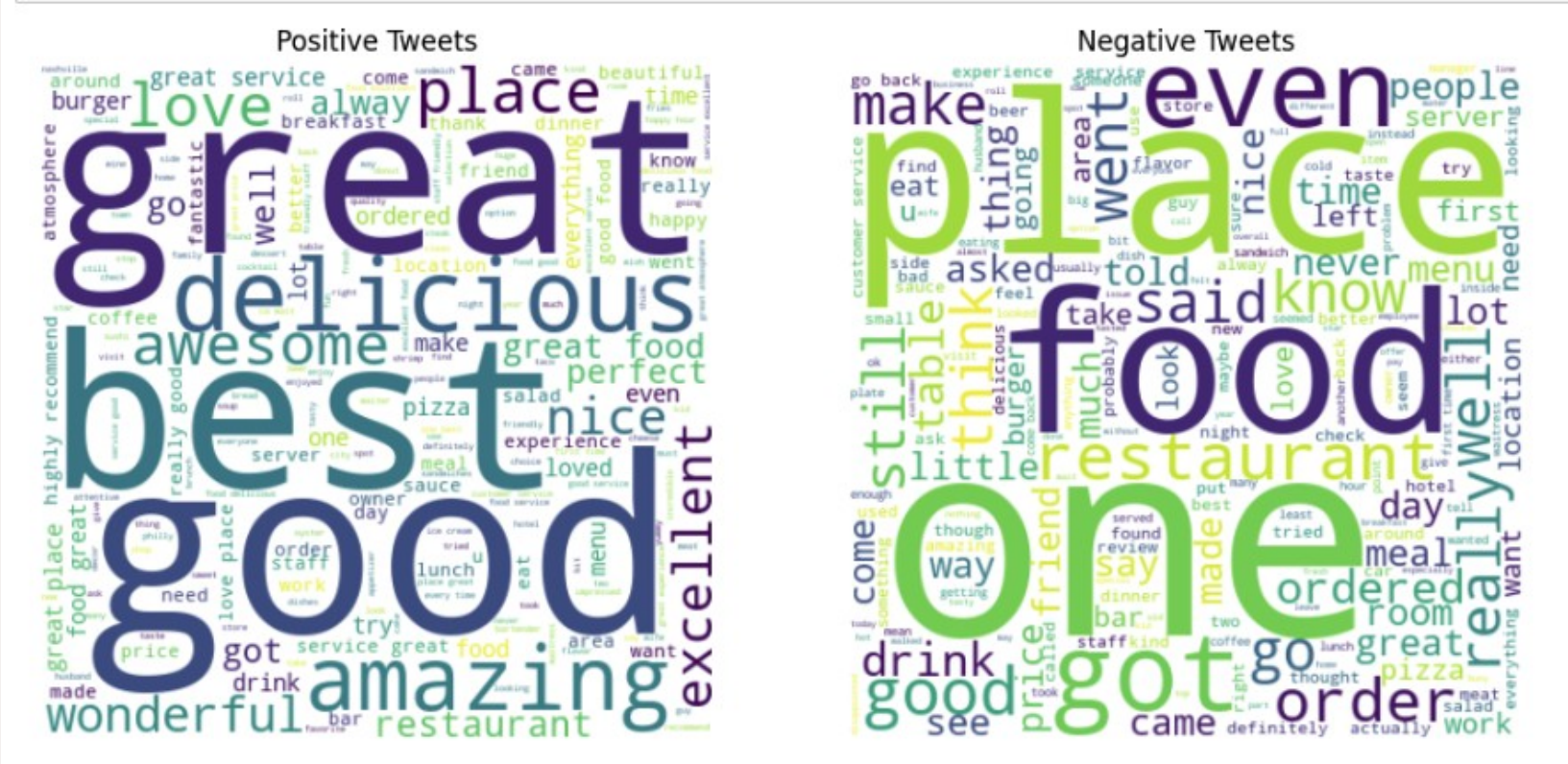
BONUS

LLM Prompt Engineering

```python
response = client.chat.completions.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are a Michelin food critic."},
        {"role": "user", "content": f"{review}"},
        {"role": "assistant", "content":
        """
        Evaluate the review based on the Michelin star criteria. If a criterion is not explicitly me
        1. Quality of product (fresh, seasonal ingredients): Return 'Not Mentioned' if not found in
        2. Mastery of flavor and cooking techniques: Return 'Not Mentioned' if not found in document
        3. Personality of the chef in the cuisine: Return 'Not Mentioned' if not found in document.
        4. Value for money: Return 'Not Mentioned' if not found in document.
        5. Consistency of food: Return 'Not Mentioned' if not found in document.
        Output your analysis in JSON format.
        """},
        {"role": "user", "content": "Extract the Michelin star evaluation."}
```
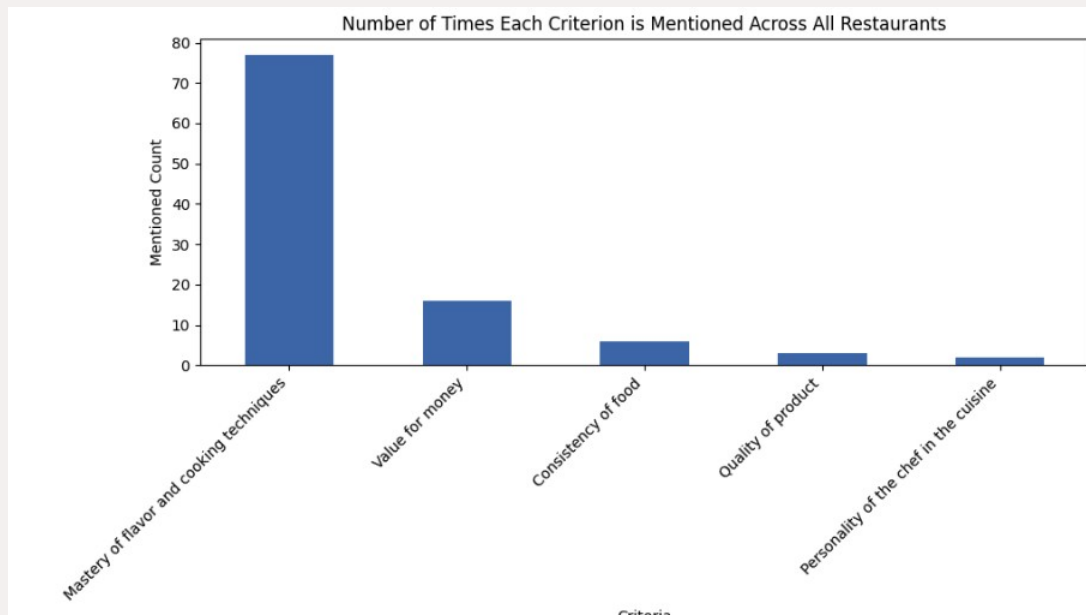
- Create a pre-defined criteria list of 'what it takes to earn a Michelin star'
- Use open-source LLM model and applied engineering for the first 750 rows

Number of Times Each Criterion is Mentioned Across All Restaurants

CONCLUSION

1. Classifying the type and the ratio of reviews that each restaurant would receive w.r.t how many stars it has can be used to improve their services and quickly analyze where they're lacking compared to their competitors

2. Average sentiment distribution gives an overall picture of how the restaurant is doing and what they overall opinion is of the restaurant. We can see a direct correlation between higher rated restaurants and the number of positive reviews they receive

3. Restaurants can see what their most frequent compliments/complaints are and take business actions accordingly to get another Michelin Star

4. We were able to see that restaurants with lesser number of Michelin stars had higher number of negative reviews indicating that they need to improve their performance

THANKS!

Any questions?