PROJECT REPORT

The Sentiment Scout: Aspect based Sentiment Analysis of Customer Experience

Anvitha Hiriadka

MS in Information Systems

College of Engineering, Northeastern University

April 2024

_____

Abstract: In today's competitive landscape, understanding customer sentiment is paramount across diverse industries. However, the abundance of unstructured customer reviews presents a significant challenge for businesses seeking actionable insights. Traditional manual analysis methods prove time-consuming and biased.

This project addresses these challenges by implementing Sentiment Analysis to automate sentiment understanding and aspect extraction. Through meticulous building blocks including data pre-processing, feature engineering, and model training, we evaluated three models.

Our analysis reveals that the Multinomial Naive Bayes model emerged as the top performer, achieving an accuracy of 79.34%. This research underscores the importance of sentiment analysis in informing business strategies and provides practical insights for organizations aiming to leverage customer feedback for enhanced decision-making and customer satisfaction.

PROBLEM SYNOPSIS

The customer's perception is your reality. – Kate Zabriskie

In a world where consumer opinions shape the success of businesses, there is a silent battleground across diverse industries. In the realm of business, from the sizzle of fast-food chains to the runway of fashion empires, understanding customer sentiment is paramount. The challenge lies in deciphering the large unstructured customer reviews scattered across different platforms. Businesses are inundated with

an abundance of reviews ranging from glowing endorsements to harsh criticisms, making it difficult to extract actionable insights.

Traditional methods of manually analyzing the reviews are time consuming and often prone to bias. By implementing the Sentiment Analysis, we aim to automate the process of understanding customer sentiment along with the aspects.

APPLICATIONS

1) Transforming this customer feedback into a strategic asset that can empower businesses to thrive in an increasingly competitive market landscape.
2) This can help businesses focus and improve on the aspects where they failed to gain positive sentiment from the customers.
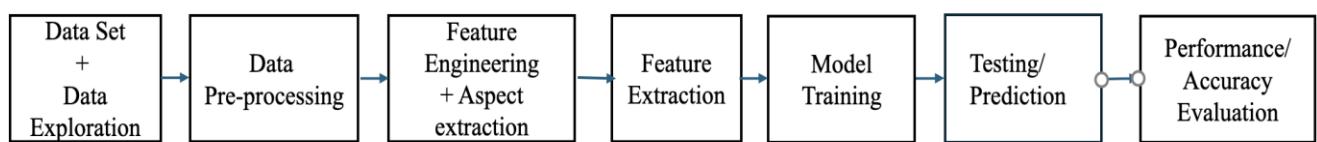
SCOPE AND OBJECTIVE

This project is about sentiment analysis of customer reviews about products or service purchased. Considering the data set of reviews about a restaurant in the United States of America to train the machine learning model to analyze and identify the sentiment present in the textual content.

Customers may review a service or product based on different aspects or factors. The review could be about the staff, quality of service, price, ambience or cleanliness. In this project, we will analyze the customer reviews and find the sentiment polarity of each review and classify it into different classes i.e., positive, neutral and negative towards the aspect or factor mentioned in the review (aspect towards which the sentiment was expressed)

The model is trained using the reviews, aspects and it's associated sentiments, this model can then be generalized against unseen textual data like reviews or feebacks given by the customers.

BUILDING BLOCK OF THE MODEL

Data Set + Data Exploration → Data Pre-processing → Feature Engineering + Aspect extraction → Feature Extraction → Model Training → Testing/ Prediction → Performance/ Accuracy Evaluation

Model Design Description

Data Set

The dataset used here is of reviews of Mc Donald's Stores in the United States of America, scraped from Google Reviews.

This extensive dataset comprises 33,396 reviews, offering a rich repository of customer experiences and opinions regarding different McDonald's locations nationwide. Within this dataset, you will find a comprehensive array of information including reviewer ID, store name, category, store address, latitude, longitude, rating count, review time, as well as detailed reviews and ratings. This data set serves as a valuable resource for gaining deep insights into customer sentiments and perceptions, enabling comprehensive analysis and informed decision-making for various stakeholders within the organization.

Key Features in data set:

reviewer_id: Unique identifier for each reviewer (anonymized)

store_name: Name of the McDonald's store

category: Category or type of the store

store_address: Address of the store

latitude: Latitude coordinate of the store's location

longitude: Longitude coordinate of the store's location

rating_count: Number of ratings/reviews for the store

review_time: Timestamp of the review

review: Textual content of the review

rating: Rating provided by the reviewer

Data Exploration

The data set contains ten different features, the most useful features among these which are relevant for the sentiment analysis are 'review' and 'rating'. This would facilitate generating a meaningful insight called 'sentiment polarity'.

The data used here is mostly textual content which requires natural language processing before it is passed to the machine learning model. The only feature that

is considered here in this case which is discrete in nature is the feature 'rating'. This feature 'rating' is later used to derive sentiments based on the threshold set.

Visualizing the distribution of ratings given by the customers to the restaurant using barplot. This helps us understand how ratings are distributed across different categories (e.g., 1 star, 2 stars, 3 stars, 4 stars, 5 stars).
The ratings distribution in data set is as shown in Figure 1.
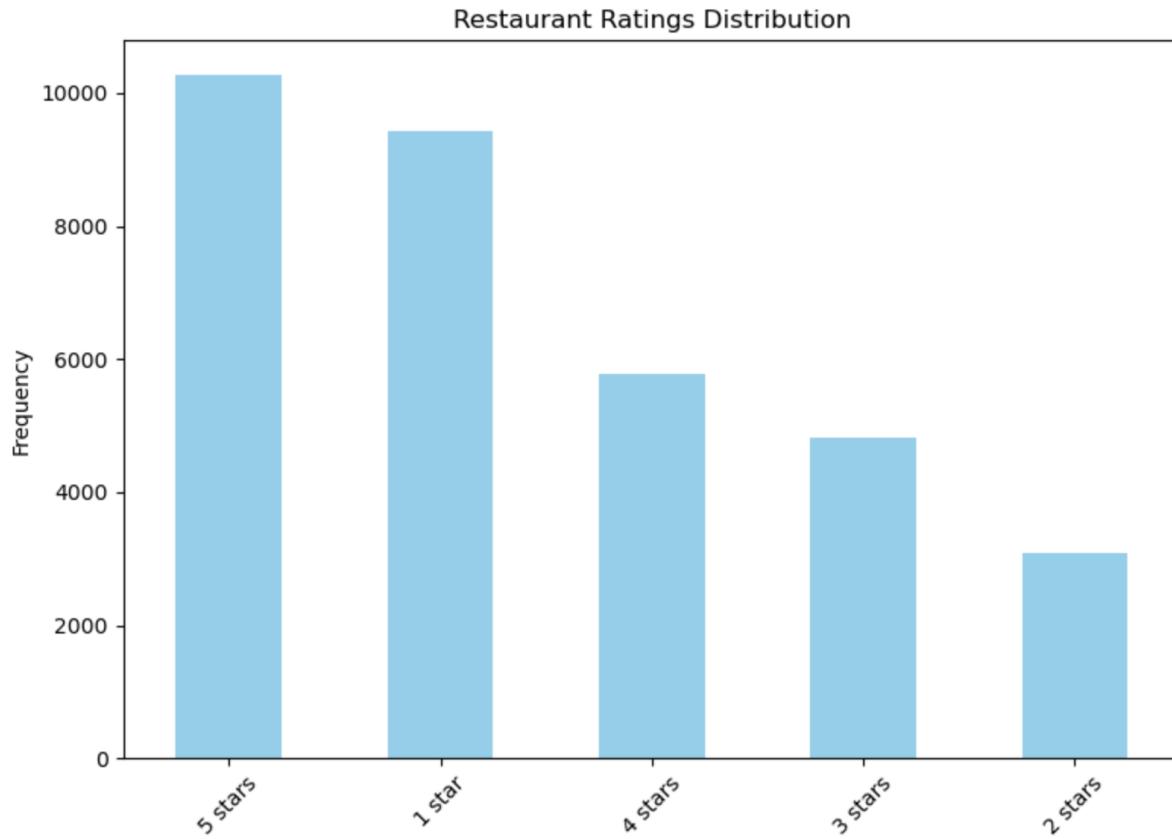
Figure 1. Ratings Distribution of Restaurant

Data Preprocessing

Standardization is a very essential part of data pre-processing. In this case of processing the reviews of restaurant, there is high possibility for the content of the review to be diverse and noisy. It may contain spelling errors, abbreviations, punctuation, emoticons, capitalization. These affect the text data quality.

From observing the 'review' feature in the dataframe, we can conclude that are many noises and irrelevant characters in the reviews. The dataset must be cleaned and pre-processed before it is sent to the machine learning model. Therefore, Standardization needs to be performed for the text content [Reviews].

In this context, we will use some of the popular toolkit and libraries for further pre-processing steps.

Toolkits and Libraries used:

1) Natural Language ToolKit (NLTK)

Natural Language ToolKit (NLTK) platform provides comprehensive python libraries for Natural Language Processing (NLP), which offers libraries for text preprocessing such as lemmatization, and POS Tagging.

2) spaCy

spaCy is another fast Python library used for NLP (Natural Language Processing) that provides a pipeline approach for text preprocessing, supporting tasks like tokenization, Stop Word removal, lemmatization and part-of-speech tagging.

Below is the list of text pre-processing done on the 'Review' feature of the dataframe using above mentioned tools.

1. Data cleaning (Removing duplicate and Missing values)
2. Case folding
3. Noise removal – Removed using Regular Expression
4. Stop-word removal
5. Tokenization
6. Lemmatization

1. Data cleaning

Data cleaning is performed by checking for duplicate and missing values in the data frame.

2. Case folding/ Lowercasing

Applying case folding for the Text in the review to lower case ensure consistency in text processing and help avoid issues related to case sensitivity during analysis and modelling.

Converted the review feature content to lowercase.

3. Noise Removal

Removing unwanted characters such as punctuation, irrelevant characters, emoticons, extra white spaces, etc. Using regular expressions.

4. Stop word removal

Using pre-trained model(en_core_web_sm) from spaCy Library.

Removing stop words, which are common words that occur frequently in the sentence, and it does not hold much semantic meaning (Example: "The", "is", "of", etc.).

5. Tokenization

Choosing the spacy method, as it provides high performance tokenization. Here the entire review is broken down into words (Tokens).

6. Lemmatization

Performing Lemmatization to reduce the words into their dictionary forms or lemma. This step normalizes the text further.

Using NLTK's(Natural Language Toolkit) WordNetLemmatizer() for lemmatization

Feature Engineering

This step involves deriving target variables (sentiment) and a feature variable(aspects).

Target variables 'sentiment' is derived from the feature 'review' and the target variable 'rating_sentiment' is derived from feature 'rating'. This will provide us with target variables having sentiment polarity classes positive, neutral and negative.

Aspect extraction is done for extracting top aspects from each review. Identifying the most common aspects or features discussed in positive, negative, and neutral reviews.

• Sentiment Classification [Target Variables]

This is a crucial step here in Data Pre-processing as we are deriving the sentiment class for every review sample.

Deriving two target variables 'sentiment' and 'rating_sentiment' which will be passed independantly to the model in different cases later.

'Sentiment' Target Variable:

Deriving the Sentiments for each review text in the data frame, using Natural Language Processing Tool kit's (NLTK) pre-trained model sentiment analyzer called "Vader".

- VADER belongs to a kind of sentiment analysis that depends on lexicons of sentiment-related words.

The result of this sentiment analyzer is a sentiment polarity score, which has four measurements i.e.

- Positive

- Neagtive

- Neutral

- Compound

Here, Positive, Negative, Neutral addresses the extent of the content of review that falls into these classifications.

The compound score is an aggregated score that combines the positive, negative, and neutral scores to provide an overall sentiment score for the review. It's a normalized score that ranges between -1 (extremely negative) and 1 (extremely positive), with values closer to -1 indicating negative sentiment, values closer to 1 indicating positive sentiment, and values around 0 indicating neutral sentiment.

Therefore, after we obtain the sentiment score, we will map these scores to the sentiment label based on the threshold set. In our case here, we will use compound value to conclude the "Sentiment" class for each review using the compound component of the sentiment score.

- **Rating sentiment:**

This is rating_sentiment target variable derived from the rating feature. Converting the 5 star ratings to the sentiment polarity Marking three classes of sentiment from rating considering the following range:

Positive: 4–5-star rating

Neutral: 3-star rating

Negative: 2-1-star rating

Aspect Extraction

Extraction of Aspects from the review is done using spaCy Library's pre-trained model 'en_core_web_sm'.

We are using POS tagging method to get the aspects from the review. Based on the tags given to the words such as "NOUN" and "PROPN", extracting the aspect from

the review sample. A new feature variable "aspect" is formed using these extracted aspects for each sample.

The top 20 aspects extracted from the reviews is as shown in Figure2.
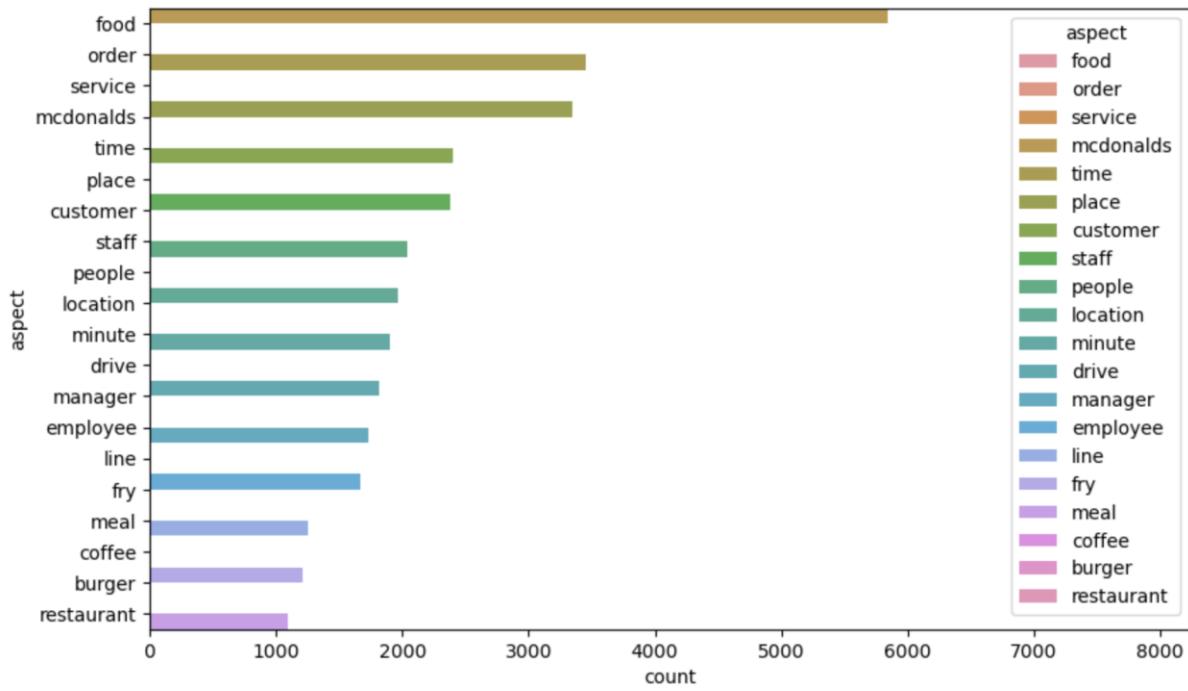

Figure 2. Top 20 Aspects extracted from the reviews

The below three figures (Figure 3, Figure 4, Figure 5) show all the aspects for sentiments positive, neutral and negative sentiments with their counts appearing in reviews.

	aspect	count
0	food	3143
1	service	2955
2	mcdonalds	2483
3	place	1519
4	order	1418
5	staff	1060
6	time	901
7	location	705
8	love	624
9	customer	576
10	people	508
11	area	438
12	coffee	421
13	restaurant	394
14	breakfast	390
15	employee	387
16	drive	366
17	kid	349
18	fry	336
19	experience	330

Figure 3. Aspects count with positive sentiment polarity

	aspect	count
0	food	1136
1	mcdonalds	811
2	order	760
3	service	642
4	time	368
5	place	356
6	staff	296
7	people	290
8	customer	226
9	location	209
10	drive	209
11	fry	158
12	employee	155
13	lot	149
14	area	149
15	meal	147
16	line	133
17	minute	132
18	burger	129
19	restaurant	128

Figure 4. Aspects count with neutral sentiment polarity

	aspect	count
0	food	3143
1	service	2955
2	mcdonalds	2483
3	place	1519
4	order	1418
5	staff	1060
6	time	901
7	location	705
8	love	624
9	customer	576
10	people	508
11	area	438
12	coffee	421
13	restaurant	394
14	breakfast	390
15	employee	387
16	drive	366
17	kid	349
18	fry	336
19	experience	330

Figure 5. Aspects count with negative sentiment polarity

Feature Extraction

Converting the text data of review into numerical feature vectors that can be used as input to the classifier. In this project, we are using a common technique TF-IDF vectorization.

It transforms text documents into numerical vectors based on the frequency of terms (TF) and their inverse document frequency (IDF) before passsing into the model.

Model Selection

The two different machine Learning predictive models are selected for Aspect Based Sentiment Analysis.

Multinomial Naive Bayes Classifier:

Multinomial Naive Bayes is known for its simplicity and computational efficiency. It is particularly effective when dealing with large datasets and high-dimensional feature spaces, which is often the case in text classification tasks like sentiment analysis. The model's fast training and prediction times make it a practical choice.

Support Vector Machines:

Support Vector Machines are good at handling complicated patterns. With kernel functions, they can understand relationships between words and sentiments that aren't simple and linear.

Model Training

Multinomial Naive Bayes Classifier

Multinomial Naive Bayes is a solid choice for aspect-based sentiment analysis due to its efficiency and effectiveness with text data. This classifier is particularly well-suited for tasks involving large volumes of text, like reviews, where the frequency of words matters. It's adept at handling the varied vocabulary found in reviews and can efficiently process the occurrence counts of words associated with different aspects.

In aspect-based sentiment analysis, the goal is to identify and analyze sentiment expressed towards specific aspects or features mentioned in the reviews. By considering the occurrence of words associated with both positive and negative sentiments within each aspect, the model can effectively classify the sentiment polarity for each aspect mentioned in the review.

Two different cases are considered for Multinomial Naive Bayes Classifier

> **Model1 - Case1** MNB Classifier trained using the 'sentiment' (sentiment derived from review) target variable.

Model2 - Case2 MNB Classifier trained using the 'sentiment' (sentiment derived from rating) target variable.

Support Vector Machines

Support Vector Machines (SVMs) are a type of supervised learning algorithm commonly used for sentiment analysis. They work by finding the hyperplane that best separates different sentiment classes in a high-dimensional feature space. SVMs are effective for text classification tasks, as they can handle both linearly and non-linearly separable data using different kernel functions. They are known for their ability to generalize well to new data and handle overfitting. SVMs are popular for sentiment analysis due to their effectiveness in learning complex decision boundaries from text data.

Capable of handling complex patterns and understanding non-linear relationships between words and sentiment.

Utilizes kernel functions for enhanced analysis.

Testing and Prediction

The model is tested with 20% of the Pre-processed data after training.

Testing is performed for various types of input reviews for three different sentiment polarities 'positive', 'negative' and 'neutral'.

Performance/Accuracy Evaluation

The performance and Accuracy of three different models are evaluated and below are the results.

Model	Accuracy	Precision	Recall	F1-Score
Multinominal Naive Bayes Model 1 Case1	75.51%	Negative = 0.90 Neutral = 0.99 Positive = 0.71	Negative = 0.61 Neutral = 0.28 Positive = 0.99	Negative = 0.73 Neutral = 0.44 Positive = 0.82
Multinominal Naive Bayes Model 2 Case1	79.34%	Negative = 0.76 Neutral = 0.98 Positive = 0.81	Negative = 0.90 Neutral = 0.19 Positive = 0.90	Negative = 0.82 Neutral = 0.32 Positive = 0.85
Support Vector Machines	81.29%	Negative = 0.55 Neutral = 0.55 Positive = 0.81	Negative = 0.80 Neutral = 0.19 Positive = 0.77	Negative = 0.69 Neutral = 0.28 Positive = 0.79

Inference:

From the above score we can infer that :

1) **Multinominal Naive Bayes Model 1 Case1 [Target variable as 'sentiment' derived from reviews] :**
- **Accuracy:**
    - Overall, the model correctly predicts 75.51% of the instances in the dataset.
- **Precision:**
    - The model has high precision for the Neutral class (0.99), indicating that when it predicts an instance as Neutral, it is correct 99% of the time.
    - It also has high precision for the Negative class (0.90), indicating a high accuracy of predictions for the Negative class.
    - However, the precision for the Positive class is comparatively lower at 0.71, suggesting that some instances predicted as Positive might be incorrect.

- **Recall:**
  - The model has a high recall for the Positive class (0.99), indicating that it effectively identifies most actual Positive instances.
  - Recall for the Negative class is moderate at 0.61, implying that the model identifies about 61% of actual Negative instances.
  - Recall for the Neutral class is the lowest at 0.28, indicating that the model misses a significant portion of actual Neutral instances.
- **F1-score:**
  - The F1-score balances precision and recall. The Positive class has the highest F1-score (0.82), reflecting a good balance between precision and recall for this class.
  - The Negative class also has a reasonably high F1-score (0.73), indicating a decent balance between precision and recall.
  - However, the Neutral class has the lowest F1-score (0.44), indicating that the precision and recall for this class are not well-balanced.

Overall, while the model performs well in predicting the Positive class, it struggles with the Neutral class, which is evident from its lower precision, recall, and F1-score for that class.

2) **Multinominal Naive Bayes Model 2 Case2 [Target variable as 'sentiment' derived from ratings] :**

Accuracy:

The model correctly predicts 79.34% of the instances in the dataset.

Precision:

The model has high precision for both the Neutral class (0.98) and the Positive class (0.81), indicating high accuracy for these classes.

Precision for the Negative class is slightly lower at 0.76, but still reasonably high.

- **Recall:**
  - Recall for the Negative and Positive classes is high (0.90), indicating that the model effectively identifies most actual instances for these classes.

- o However, recall for the Neutral class is notably lower at 0.19, suggesting that the model misses a significant portion of actual Neutral instances.
- **F1-score:**
  - o The F1-scores for the Negative and Positive classes are relatively balanced, with scores of 0.82 and 0.85 respectively.
  - o However, the Neutral class has a much lower F1-score of 0.32, indicating an imbalance between precision and recall for this class.

3) Support Vector Machines Model3 [Target variable as 'sentiment' derived from ratings] :

- **Accuracy:**
  - o The model correctly predicts 81.29% of the instances in the dataset.
- **Precision:**
  - o Precision for the Negative and Neutral classes is the same, both at 0.55.
  - o Precision for the Positive class is higher at 0.81, indicating a higher accuracy   for this class.
- **Recall:**
  - o Recall for the Negative class is relatively high at 0.80, indicating that the model effectively identifies a significant portion of actual Negative instances.
  - o Recall for the Positive class is also relatively high at 0.77.
  - o However, recall for the Neutral class is notably lower at 0.19, suggesting that the model misses a large portion of actual Neutral instances.
- **F1-score:**
  - o The F1-score for the Negative class is 0.69, indicating a reasonably balanced performance between precision and recall for this class.
  - o The F1-score for the Positive class is 0.79, indicating a good balance between precision and recall.
  - o However, the F1-score for the Neutral class is low at 0.28, indicating an imbalance between precision and recall for this class.

Therefore, considering all the above factors, selecting Model 2 - Case2 Multinomial Naive Bayes for the loading into the streamlit Web Application as it has a good accuracy, precision, recall when compared to other models. Inspite, of Support

Vector Machines showing an accuracy of 81.29%, it does not have a good precision towards Negative and Positive instances.

IMPROVEMENT

- Improving the model by finding out more refined aspects from the reviews.
- Enhancing the model further, by implementing a new classifier from scratch using the same bayes theorem by including other features along with reviews for having more accurate and precise model to develop the advanced level sentiment analysis.

WEB APPLICATION

The Web Application for the Sentiment Scout is developed using Streamlit.

URL : sentiment-scout-application

Github : Application code for sentiment scout github link

Figure 6. Web Application for Sentiment Scout

ACKNOWLEDGMENTS

CITATION

1) Dataset: Elgiriyewithana, N. (2023) 'McDonald's Store Reviews'. Data Set Link

2) Geron,A.(2017). Hands-On Machine Learning with Scikit-Learn & TensorFlow(1st Ed.).O'REILLY

3) Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

4) I. Perikos and I. Hatzilygeroudis, "Aspect based sentiment analysis in social media with classifier ensembles," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 2017, pp. 273-278, doi: 10.1109/ICIS.2017.7960005.

5) Syahrul Mubarok, M., A., & Dwi Aldhi, M. (Directors). (2017). Aspect-based Sentiment Analysis to Review Products Using Naïve Bayes .

6) NLTK :: Sample usage for sentiment. (n.d.-b). https://www.nltk.org/howto/sentiment.html

7) Pai, A. (2024b, March 17). What is Tokenization in NLP? Here's All You Need To Know. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/

8) GeeksforGeeks. (2024, January 3). Removing stop words with NLTK in Python. GeeksforGeeks. https://www.geeksforgeeks.org/removing-stop-words-nltk-python/#removing-stop-words-with-sklearn

9) NLTK :: nltk.stem.wordnet module. (n.d.). https://www.nltk.org/api/nltk.stem.wordnet.html

10) GeeksforGeeks. (2024b, March 21). Sklearn Feature Extraction with TF-IDF. GeeksforGeeks. https://www.geeksforgeeks.org/sklearn-feature-extraction-with-tf-idf/

11)   Kanika. (2024, April 23). Collecting Feedback Using 1 to 5 Rating Scale Surveys. Measure Customer Satisfaction using Surveys with 1 to 5 Rating Scale. https://www.zonkafeedback.com/blog/collecting-feedback-with-1-to-5-rating-scale-survey

12)   Image for web application generated using Adobe Firefly. (April 16, 2024).