Sentiment Analysis

Presenters :

-
- Edwin Korir
-
- Maryan Daud
-
- Sarah Njoroge

Last Update :Thursday, 22 June 2023

Table of Content

Sentiment Safari

- Purpose Statement Project
  - Objective Statement
  - Business Understanding
  - Methodology
- Data Story Telling
  - Results and Insights
    - EDA(graphs)
    - Model Result
- Conclusion
  - Recommendations
  - Challenges and Limitations
  - Next Steps

Objective Statement

- **Our project aims to analyze customer sentiments on social media platforms, specifically focusing on Twitter.**
- **The objective is to understand the overall sentiment towards various**  **Apple and**  **Google products and brands mentioned in the tweets to inform business on decisions to take, strategies and deliver customer satisfaction..**

Data Understanding

Datasets:

- **We collected a comprehensive dataset from Data World of customer tweets related to different products, including popular brands like Apple, Google, and more**  **.**
- **The dataset contains 9000 rows and 3 columns of tweets, allowing for a detailed sentiment analysis.**

Social Media Platform:

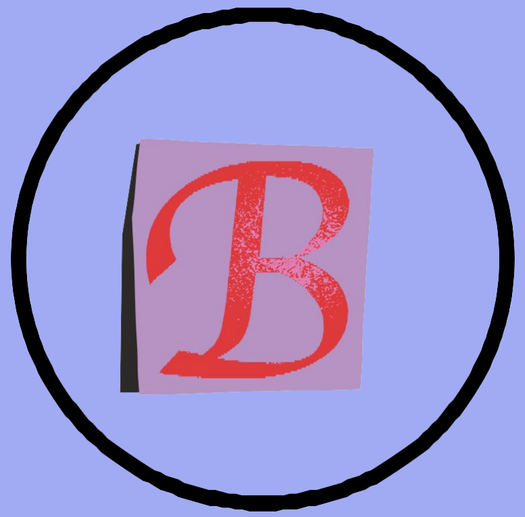
- **Data World was chosen as the primary platform for data collection.**
- **Twitter provides a rich source of real-time customer opinions and feedback, making it ideal for sentiment analysis.**

Methodology

Data Collection

Data Source

Dataset was from **Data World** with customer reviews and opinions from Twitter



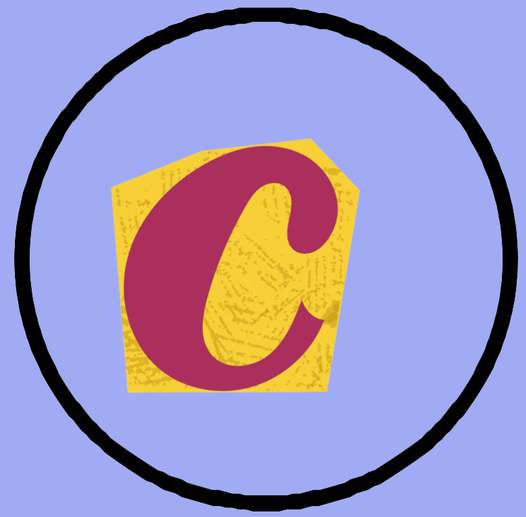
Data Preprocessing

EDA

1.removing duplicates2. handling missing

Pre-processing for NLP model

- Converted the tweet text to lowercase
- Removed html tags
- Removed the Url
- Expanded the contractions
- Removed the punctuations
- Tokenized
- Removed stopwords
- Lemmatized the tweet



Sentiment Classification

Feature extraction(CountVectorizer)

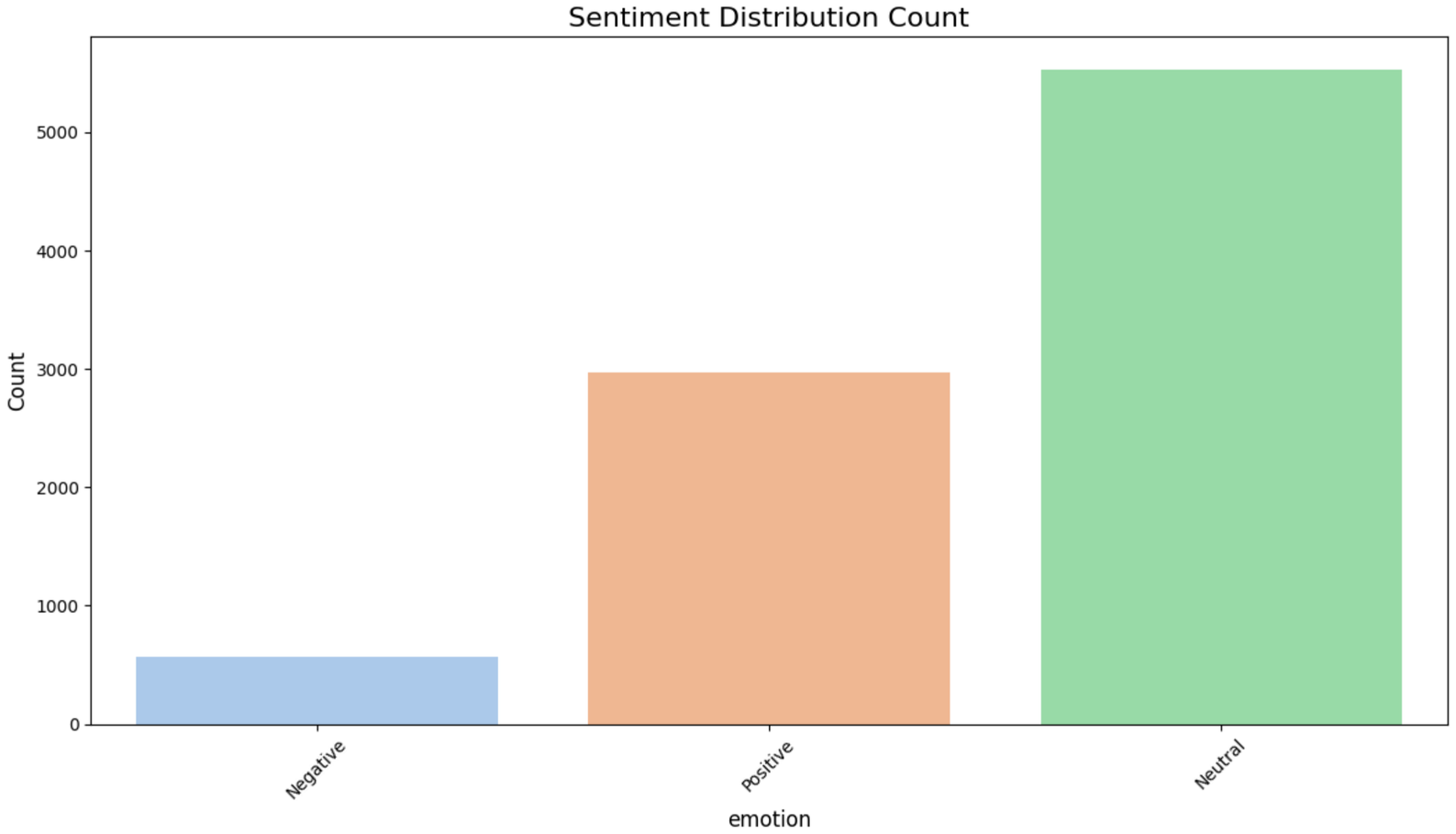
Machine learning algorithms

- Logistic Regression
- MultinomialNB
- XGBoost

Evaluation metrics

- Accuracy

Results and Insights



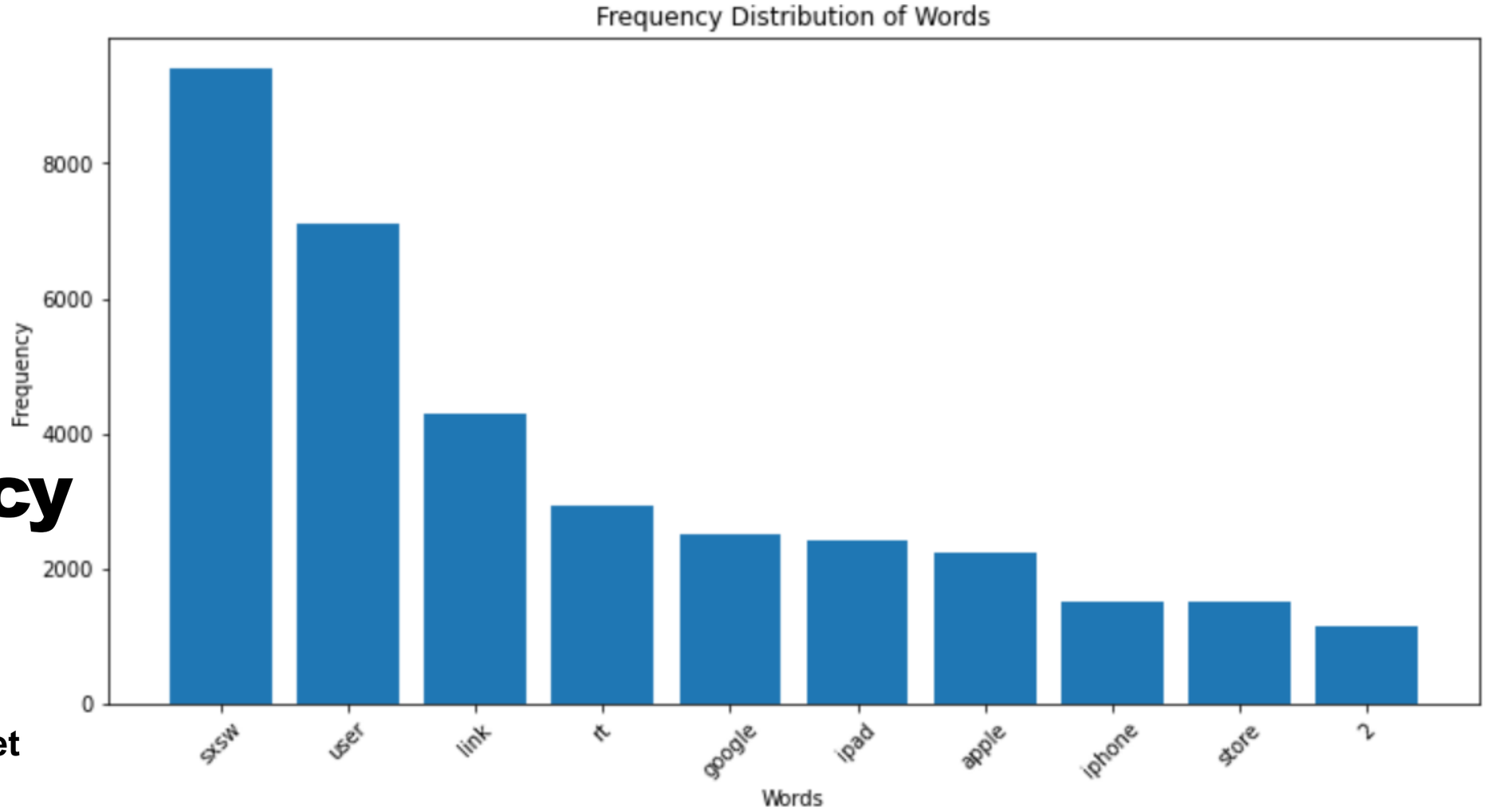
Observation:

From the graph: No emotion towards brand description had the highest value count

Words Frequency

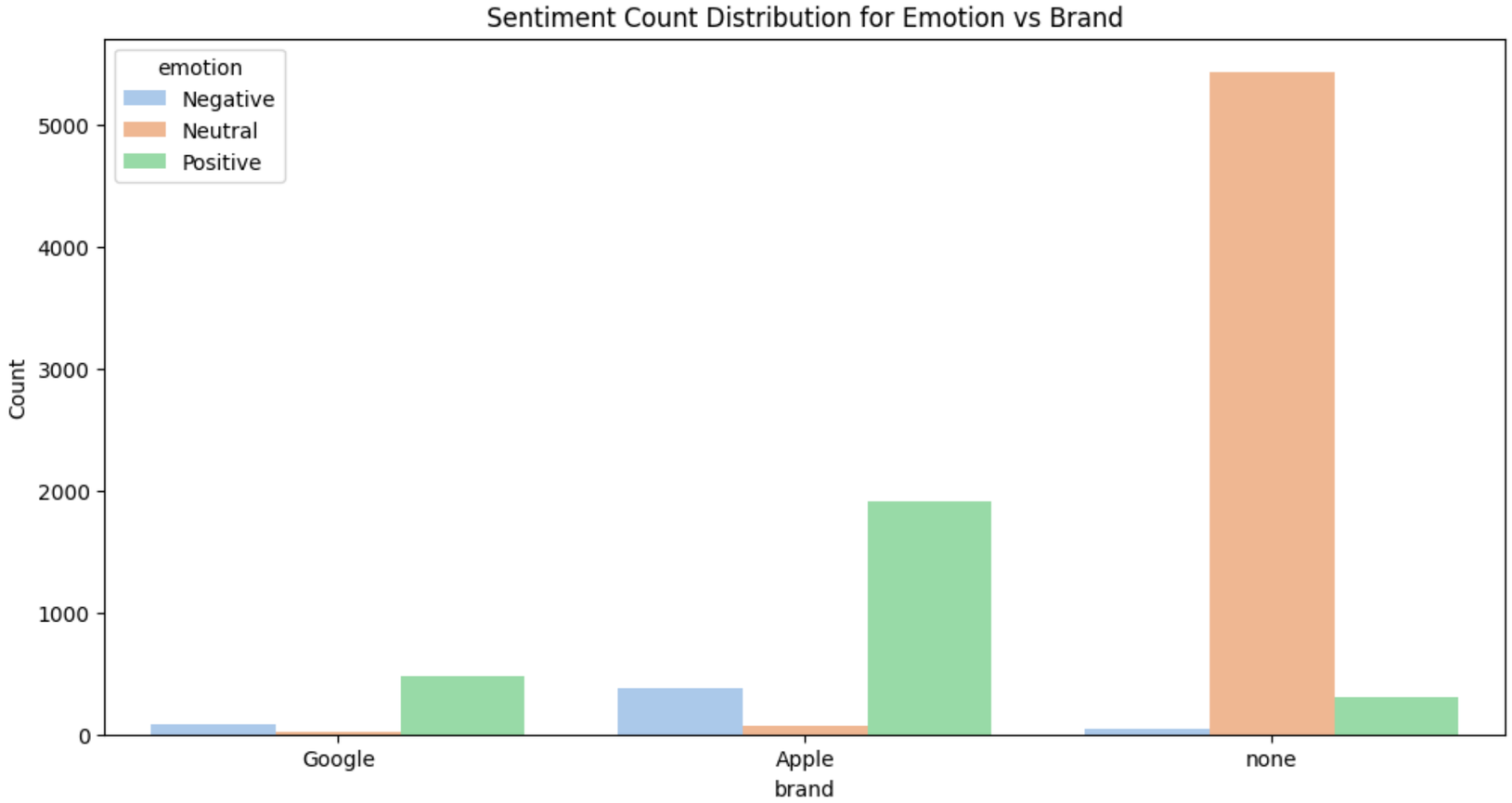
Observation

- Venue (south by south west) specific words such as sxsw
- Twitter specific words(re-tweet such as rt
- Brand or product specific words such as iphone



Brand and Emotion Relation

- None categorized brands has the highest count with list positive
- Apple had the highest positive compared to Google



Model Results

Binary		
Model	Train Accuracy	Test Accuracy
Logistic Regression	96%	90%
MultiClass		
Model	Train Accuracy	Test Accuracy
MultinomialNB	80%	65%
XGBoost	78%	68%

Recommendation

We recommend that there be more customer engagement.

Probably check on this areas;

- **Churn ratio ; rate at which customers discontinue their relationship with a product company within a given time period.**

- **Social media influencers; through brand or product endorsement**

- **Customer feedback -;The brands can introduce a rating system to accurately capture the sentiments of their customers**

Challenges and Limitations

- Limited Dataset Size: The dataset used for sentiment analysis is relatively small
- Class Imbalance Issue: The dataset suffers from class imbalance
- Language Ambiguity, Sarcasm Detection and data quality issues

Next Steps

1. In our future work, we plan to explore advanced techniques such as incorporating attention mechanisms, using ensemble methods
2. By considering these evaluation metrics, addressing limitations, and planning for future improvements, we aim to develop a robust NLP sentiment analysis solution that effectively captures ambiguite sentiments in text data.
3. Looking for a better dataset


thank
you