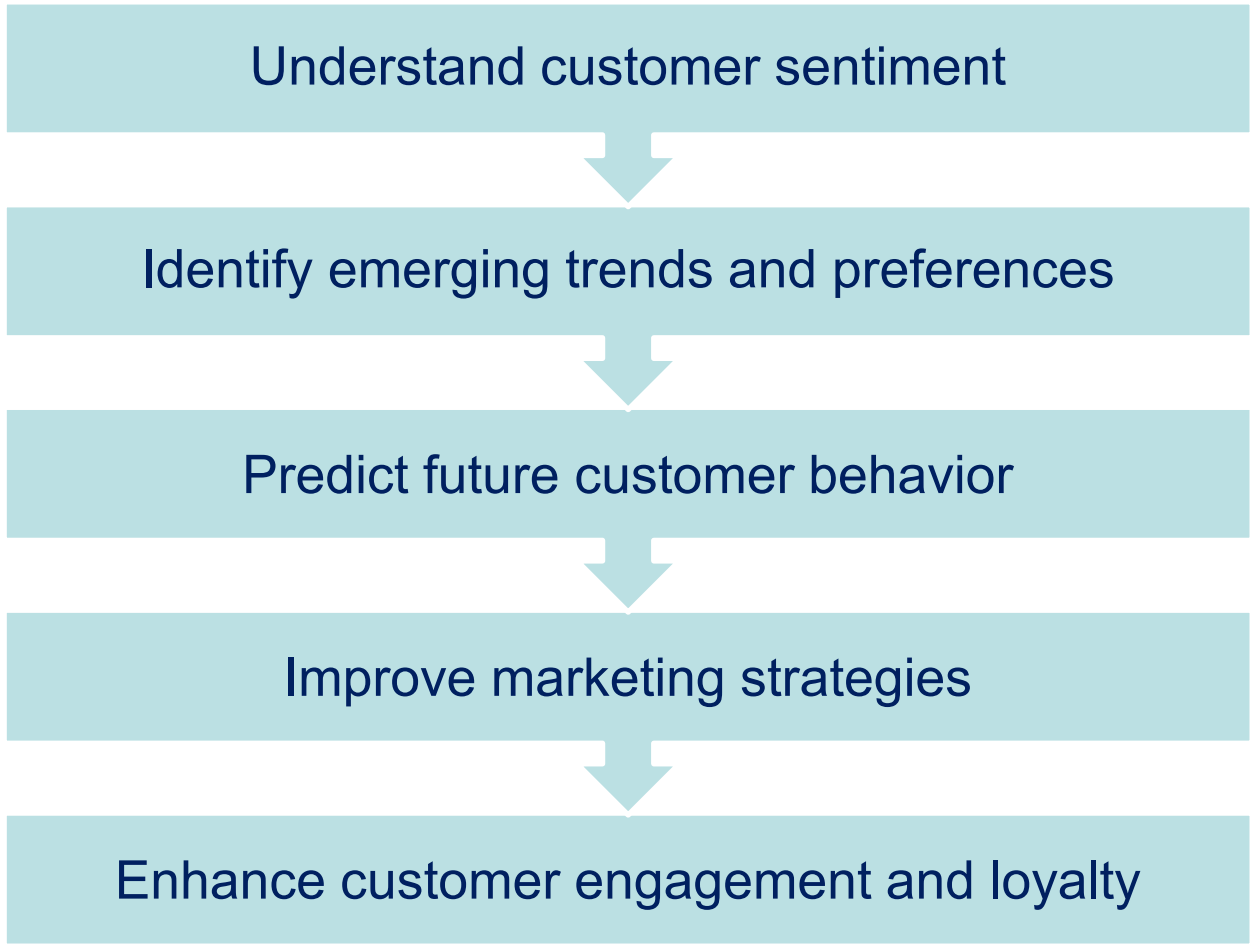
Brand Sentiment Analysis of Twitter Posts

Jonathan Agustin Fernando Calderon Juliet Lawton

Video: https://vimeo.com/843997462/2f1e66d926?share=copy

Businesses increasingly leverage social media to gain key insights on market performance.

Why Use AI in Business?



Questions

How can we detect when a Twitter post is referencing a brand?

How can we accurately classify the sentiment of Twitter posts?

How can we create a tool that does both?

Project Goals

Develop a Brand Sentiment Analyzer tool

Build a Brand Classifier that can accurately detect brand names in tweets

Build a Sentiment Classifier that can accurately classify the sentiment of a tweet

Explore key course topics

Project Overview

Explore Sentiment140 dataset

Create additional data

Preprocess the data

Train and test different models

Select highest performers to create Brand Sentiment Analyzer

Datasets

Sentiment140

Surge AI

Brands Chosen

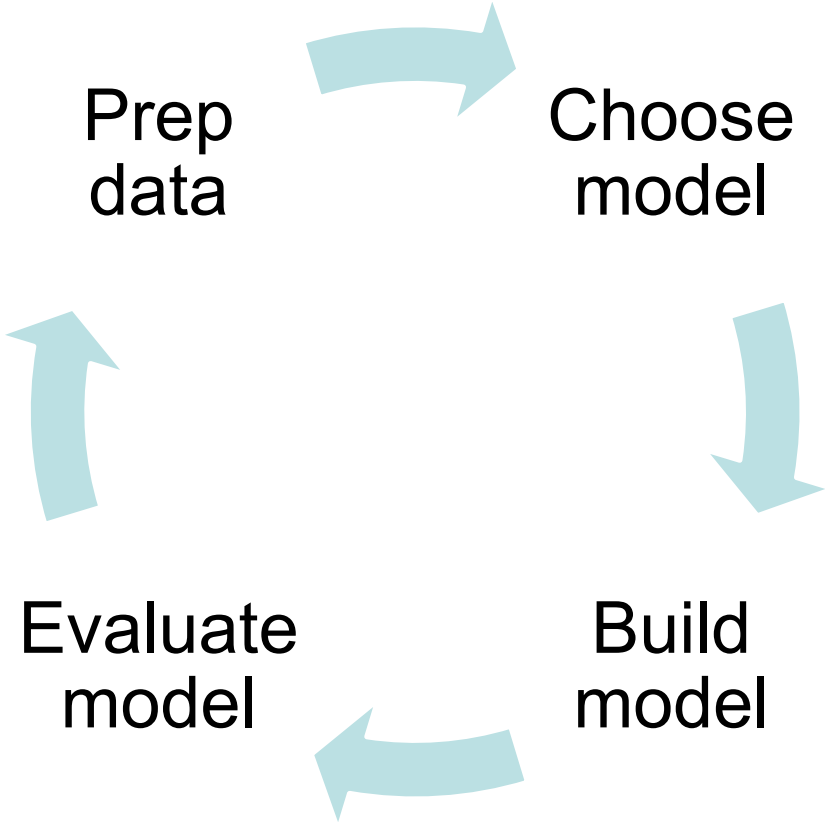
1. Facebook
2. Google
3. Apple
4. Starbucks
5. Disney
6. Walmart
7. Target
8. Microsoft
9. Amazon
10. Sony

Datasets

Sentiment140

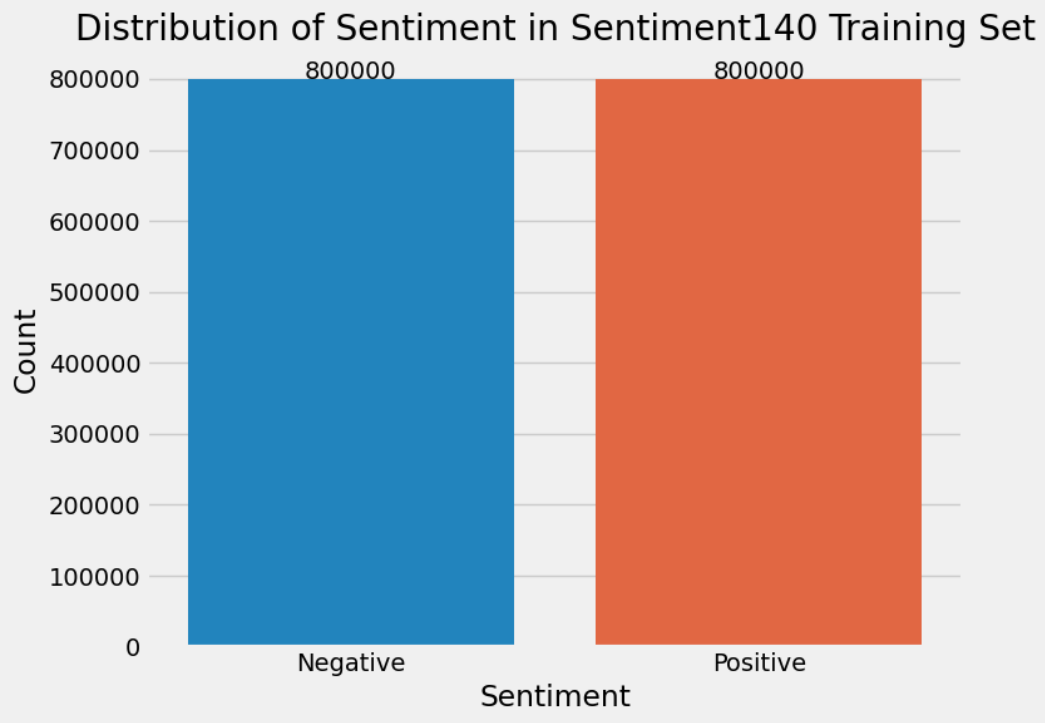
~~Surge AI~~

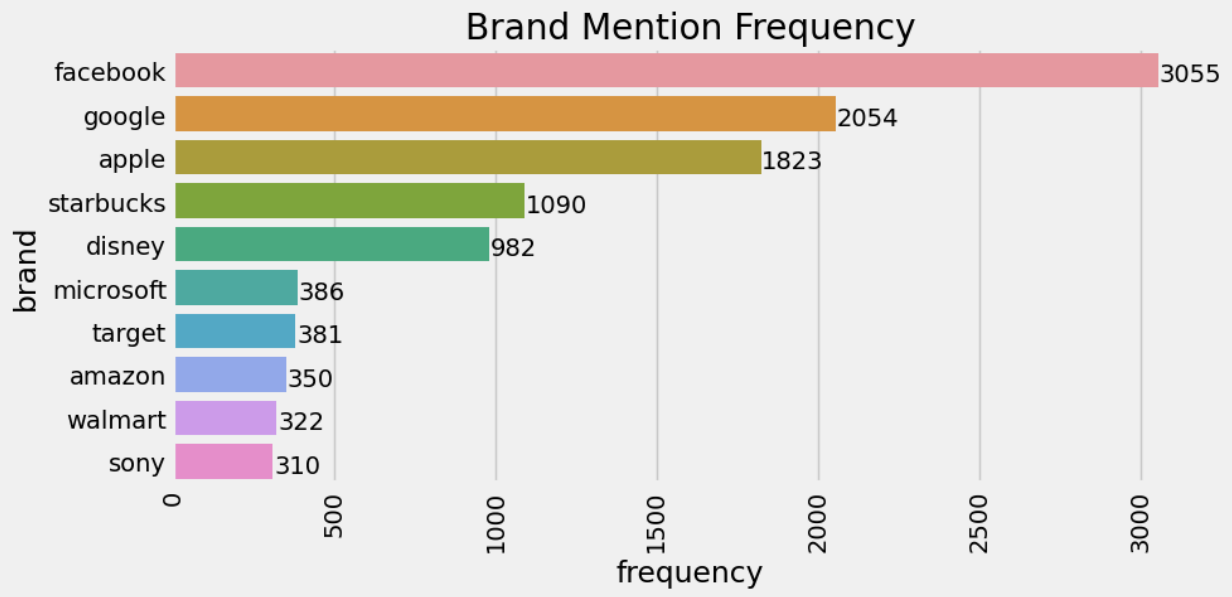
Experimental Design



First Stage: Assess

Exploratory Data Analysis





Preprocessing

Clean URLs and @mentions

Tokenize text

Remove stop words and punctuation

Reduce tokens to stems

I LOVE @Health4UandPets u guys r the best!!

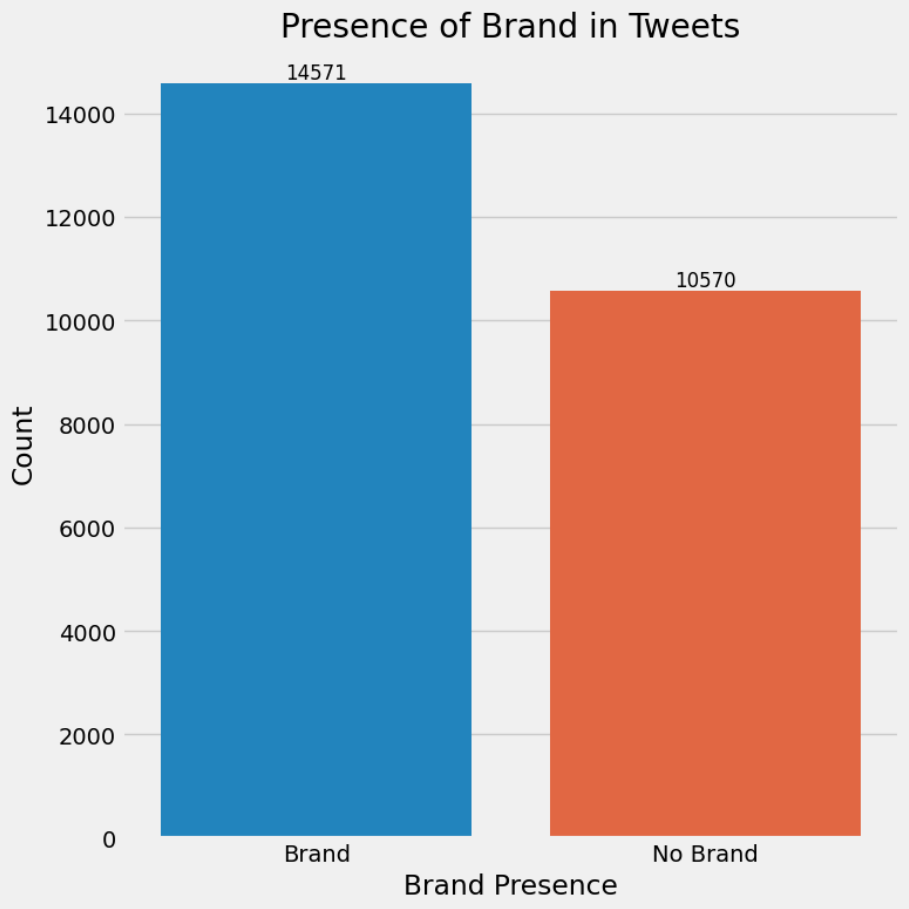
['i', 'love', 'u', 'guy', 'r', 'best']

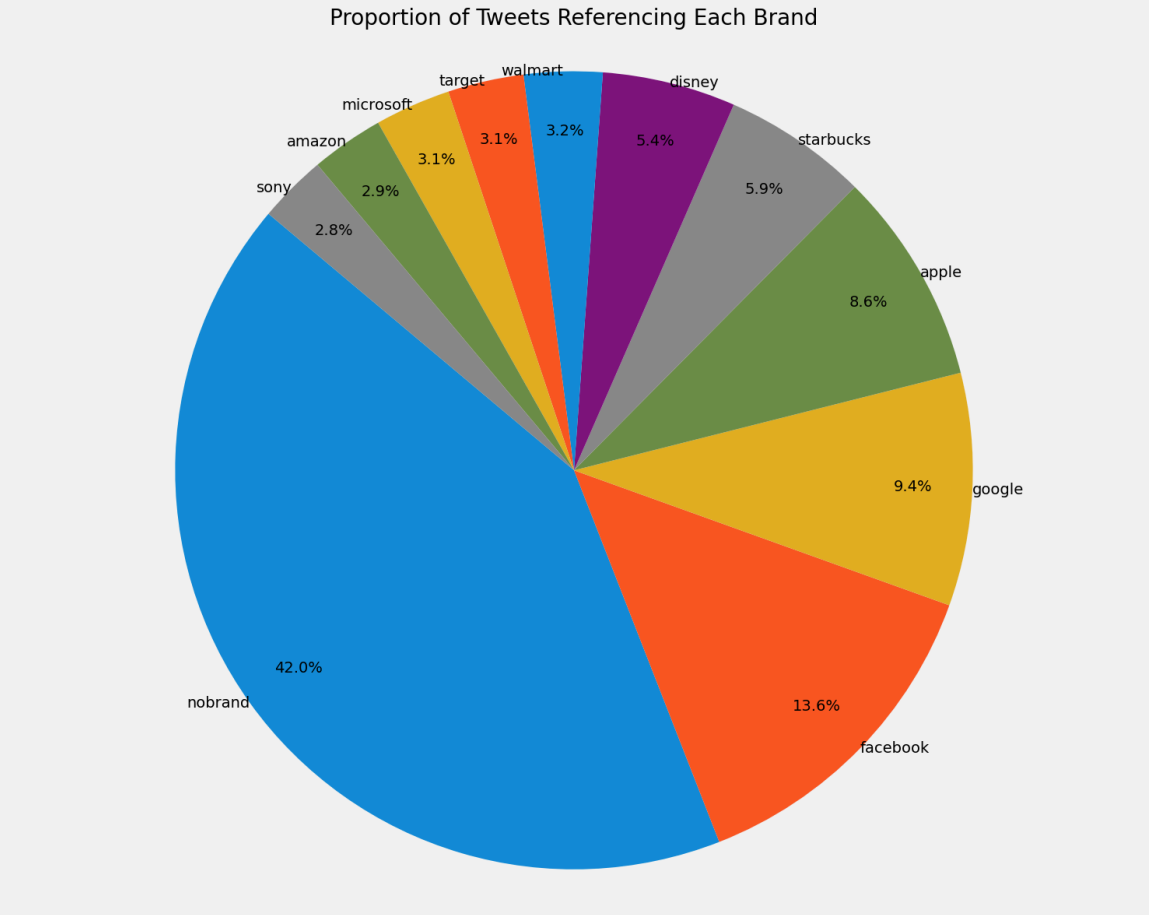
Creating Data for the Brand Classifier

tweet	sentiment
@ashman01 My only complaint about Facebook is they've changed it so much	0

tweet	sentiment	brand
@ashman01 My only complaint about Facebook is they've changed it so much	0	facebook

Creating Data for the Brand Classifier





Generating Synthetic Data

```
def create_positive_prompt(self, brand: str, count: int) -> str:
        """
        Create a positive prompt.

        Args:
            brand (str): The brand name.
            count (int): The count of the data to be generated.

        Returns:
            str: The created prompt.
        """
        logging.info(f"Creating positive prompt for brand: {brand}")
        return f"""
            Create some random data that only mention the brand {brand} in the format:
            "tweet"|||brand|||sentiment
            where:
                tweet is a Twitter post,
                brand is {brand}, and
                sentiment is 1 indicating positive sentiment of the tweet.

            EXAMPLES (with other brands)
                "My Taco Bell was great guys"|||taco bell|||1
                "Big thanks to our friends at Hattiesburg Coca-Cola"|||coca cola|||1

            AIM: Try to come up with a diverse dataset of tweets that mention brand
            {brand} and are positive.

            Create {count} data points.
            """
```

Second Stage: Develop

Machine Learning Algorithms Considered

Multinomial Naïve Bayes

Logistic Regression

Linear Support Vector Classifier

Transformers

Vectorization Methods

Count Vectorizer

Term Frequency-Inverse Document Frequency (TF-IDF)

Word2Vec

Doc2Vec

Universal Sentence Encoder (USE)

Brand Classifier

Framework: Brand Classifier objects were designed for each model, defining their vectorization methods.

Algorithms: Multinomial Naïve Bayes, Linear SVC, Logistic Regression

Sentiment Classifier

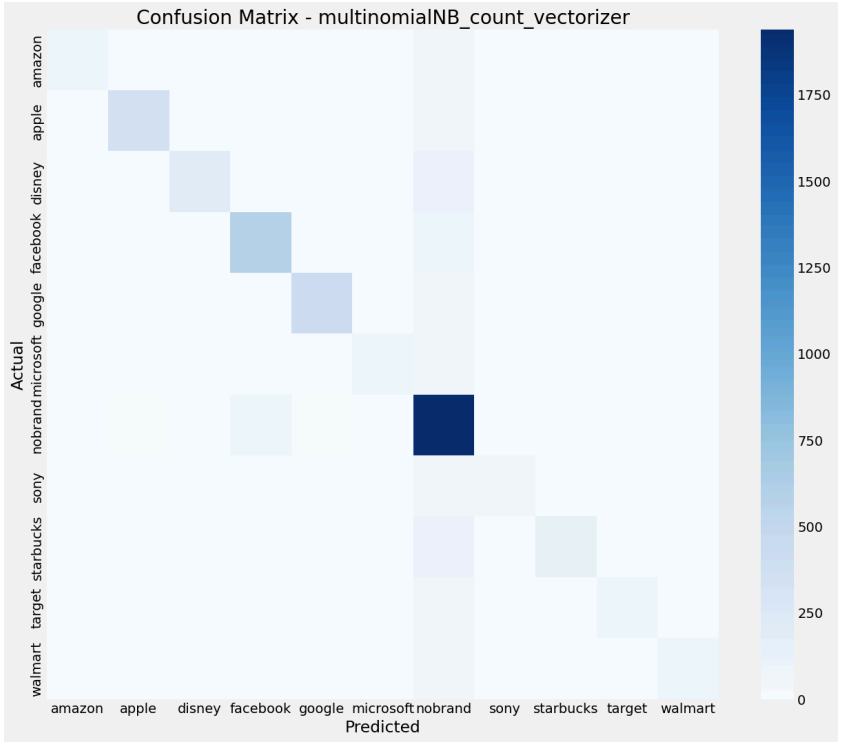
Framework: Sentiment Classifier objects were designed for each model, defining their vectorization methods.

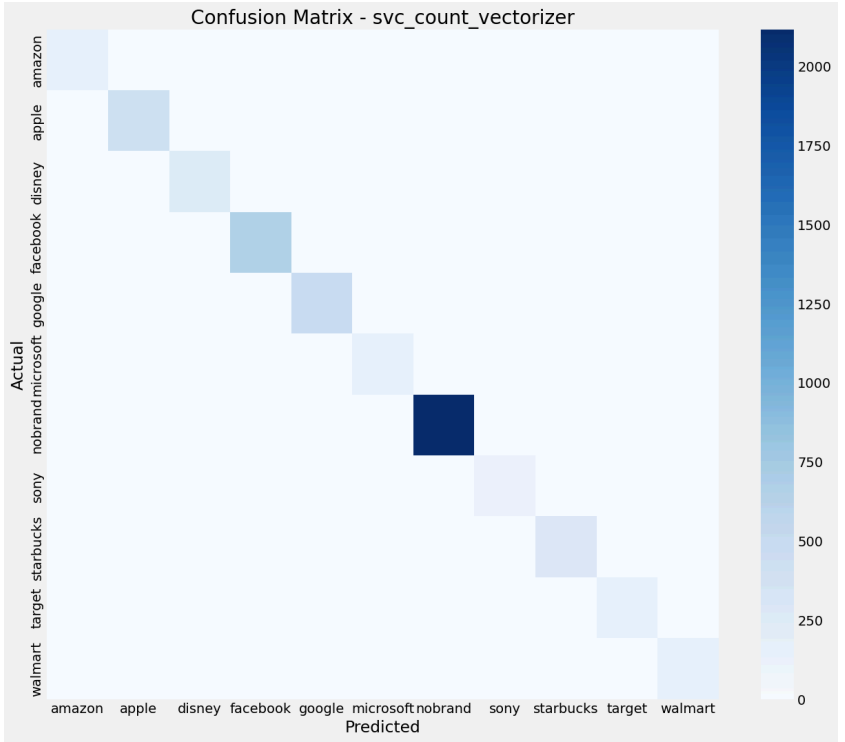
Algorithm: All models used Logistic Regression due to its effectiveness in binary classification.

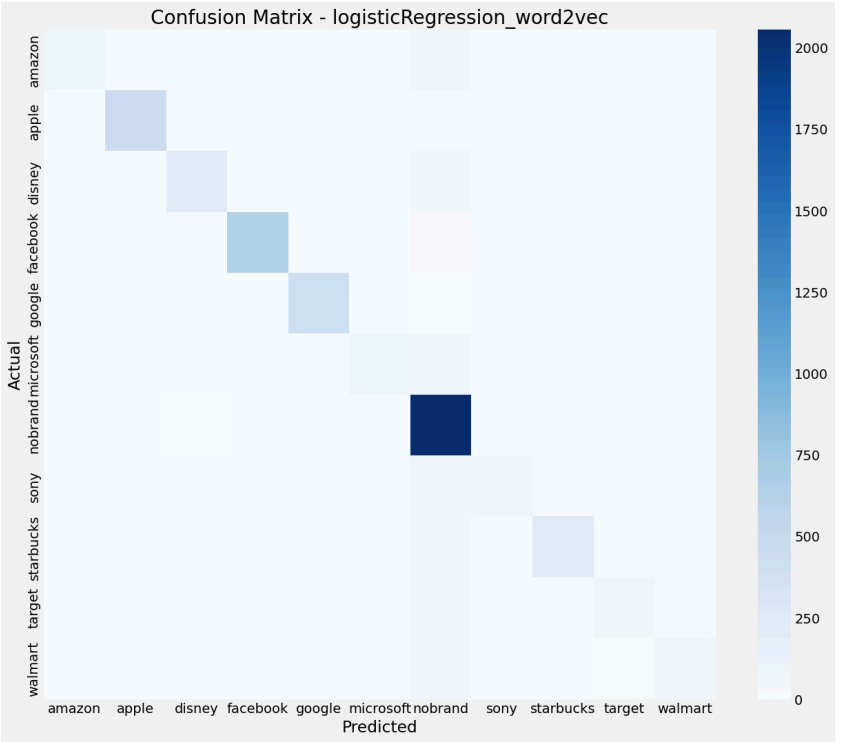
Brand Sentiment Analyzer

Third Stage: Evaluate

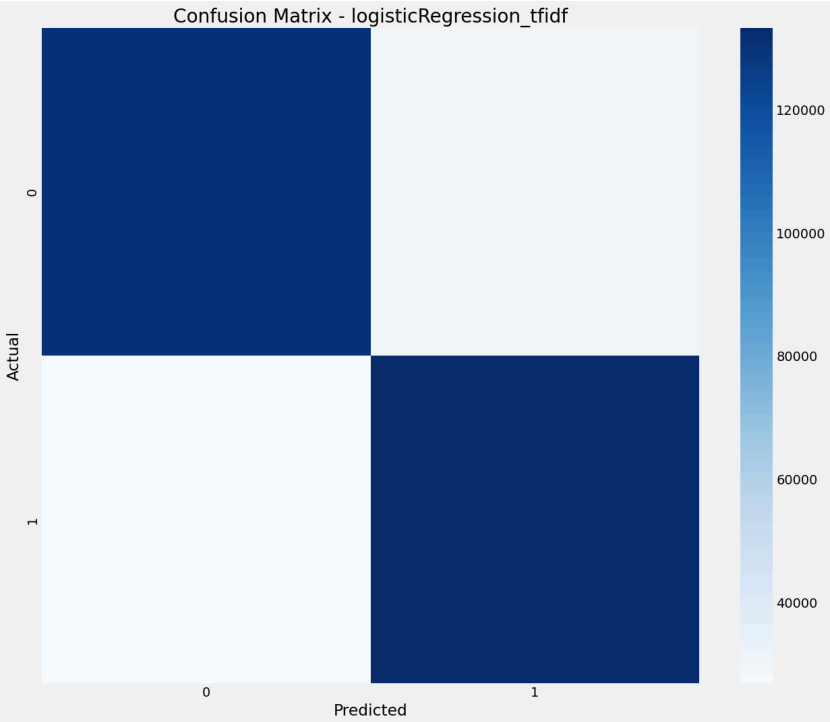
Evaluating Brand Classifiers

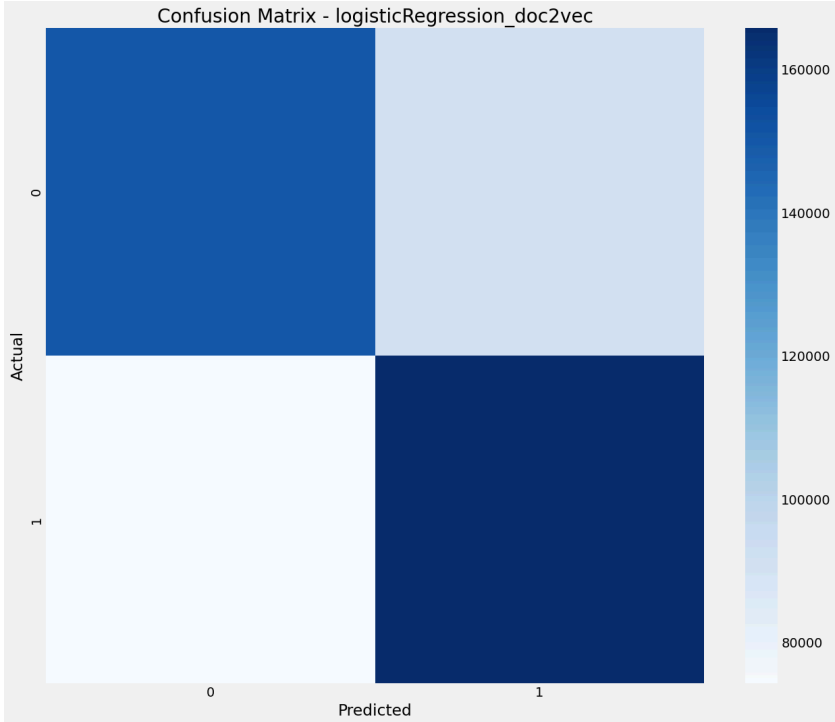


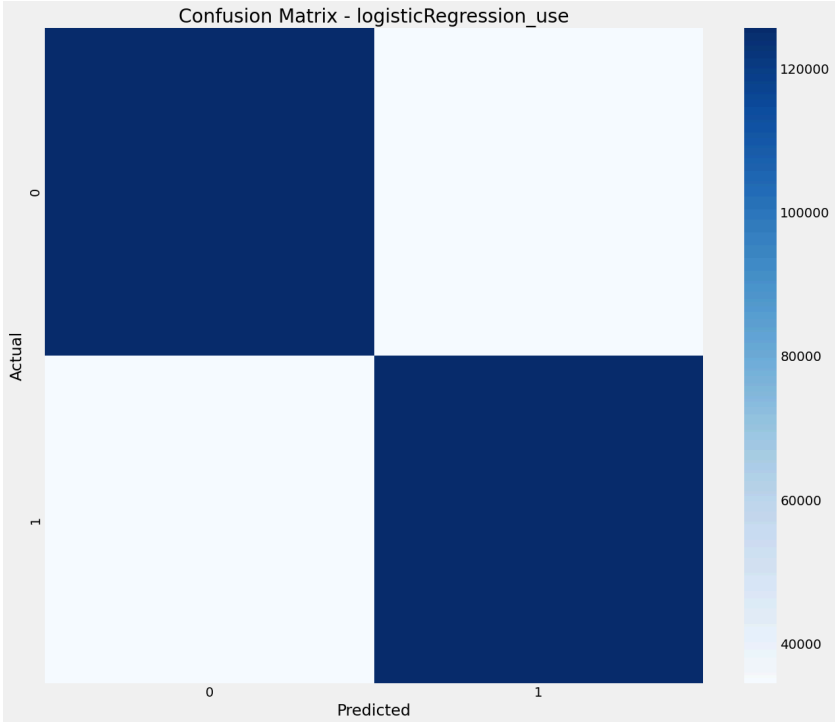




Evaluating Sentiment Classifiers







Evaluating Brand Sentiment Analyzer

Highest Performer for Brand Classification

Linear SVC with Count Vectorization

Highest Performer for Sentiment Classification

Logistic Regression with TF-IDF Vectorization

Challenges & Limitations

Ambiguity in language

Insufficient data on certain brands

Large dataset size

Computational resources and processing time

Future Work

Fine-tune model parameters

Test on larger and more diverse datasets

Develop a multimodal model

Explore the impact of data duplication on model performance

Use Large Language Models (LLMs)

Conclusion

1. Developed a robust brand sentiment analyzer
2. Identified areas for future work
3. Importance of aligning model evaluation with project goals

Acknowledgements

Advisor: Professor Andrew Van Benschoten

Director: Dr. Ebrahim Tarshizi

Program: Applied Artificial Intelligence Program
at the University of San Diego