

Wowshop Tweet Sentiment Analysis

**- Team: Tabitha Berum, Jacinta Chepkemoi, Christine Malinga, Ian Musau**

**- Course: DSC Hybrid**

https://trello.com/b/6NtdSMgt/group-2

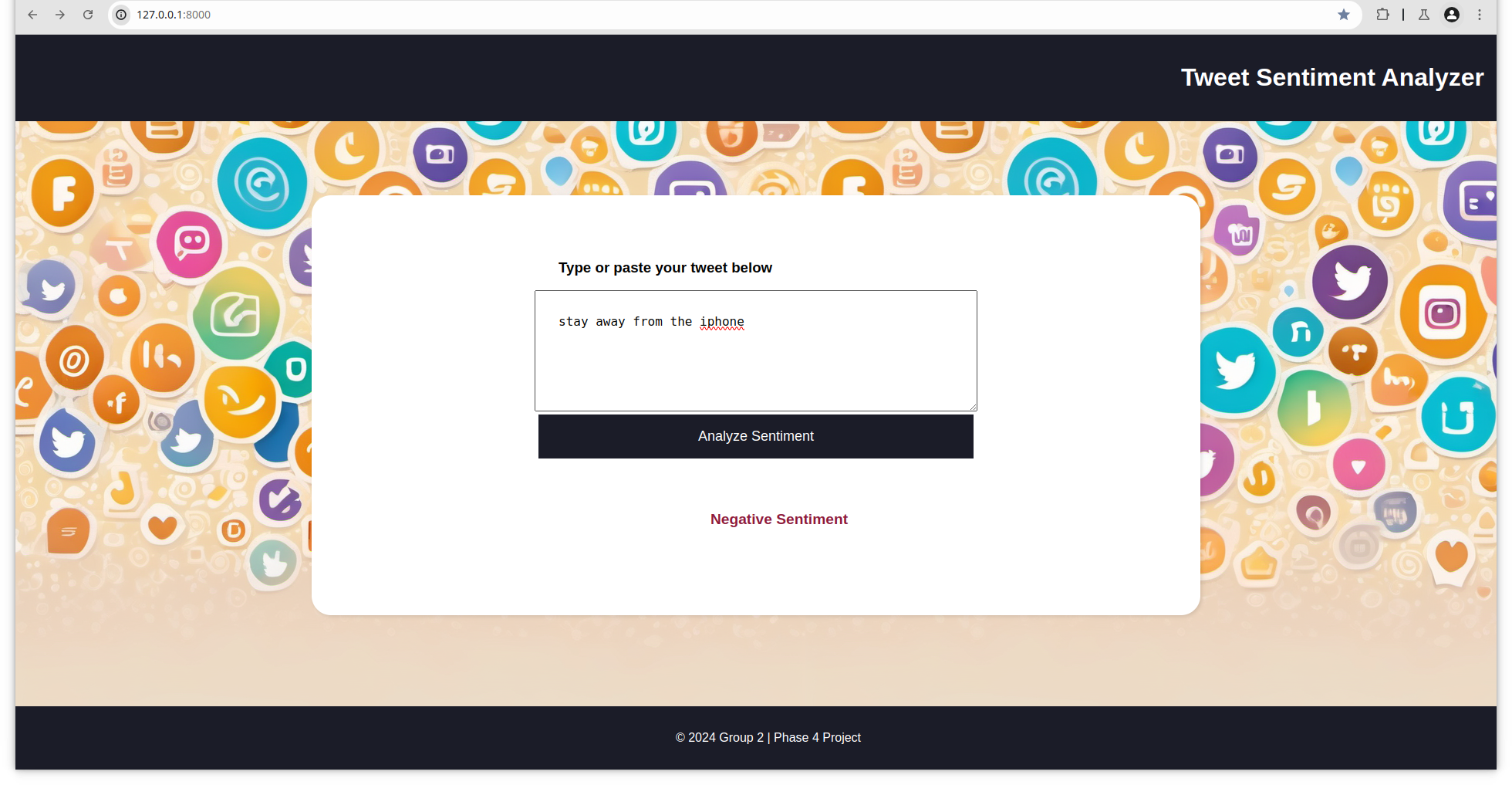
# Overview

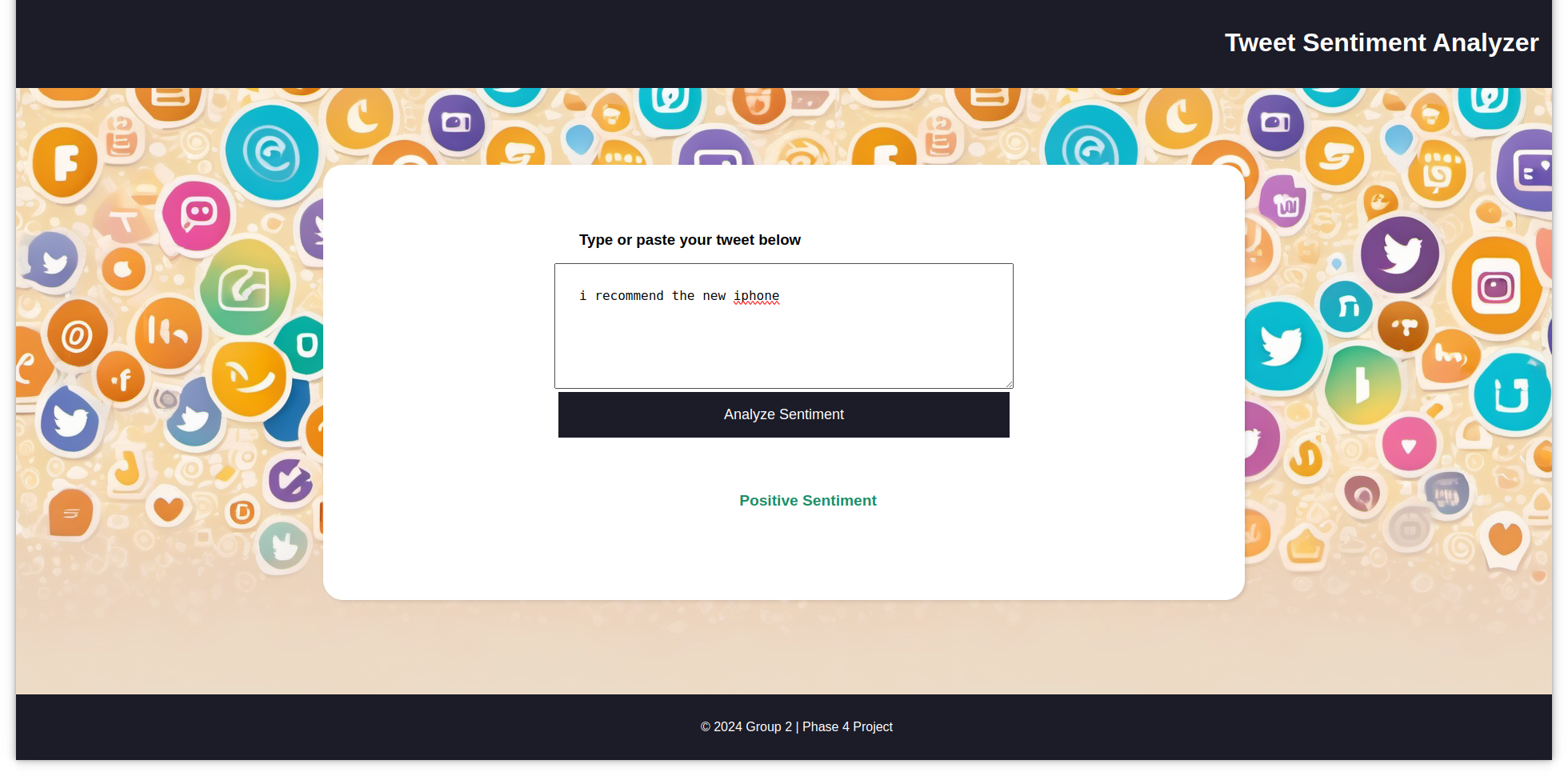
Wowshop, an e-commerce company, operates in a competitive market where customer satisfaction is crucial. Customers frequently tweet about their experiences with Wowshop, including order issues, product reviews, and customer service interactions. The company wants to enhance its customer service by quickly identifying and addressing customer issues mentioned on Twitter, improving overall service quality, and increasing customer retention.

They have tasked the team with creating a proof of concept that showcases tweet sentiment analysis. The team created a model that can read tweets and assess whether it is positive or otherwise. A simple web app was also developed to showcase the model at work. Upon delivery it will be assessed and the project can move forward.

Problems that surface during development mainly concern the small size of the dataset for training an effective and robust model. With a larger dataset, even better results can be achieved.

**Below are screenshots of the demo:**





# Business Problem

The aim of the project is to create a model that can answer:

* How do customers feel about a product?
* What is the consensus about the storefront?
* Which products are worth restocking and dropping?

# Data Understanding

The training dataset was sourced from the CrowdFlower and is stored in a csv.

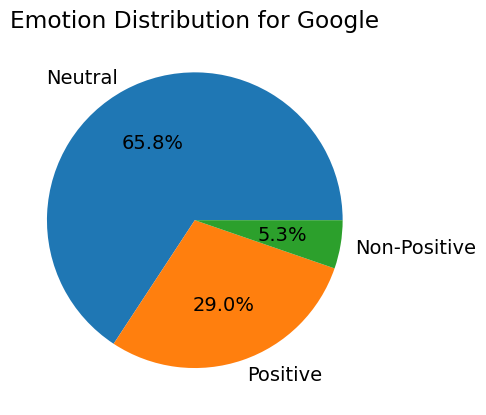
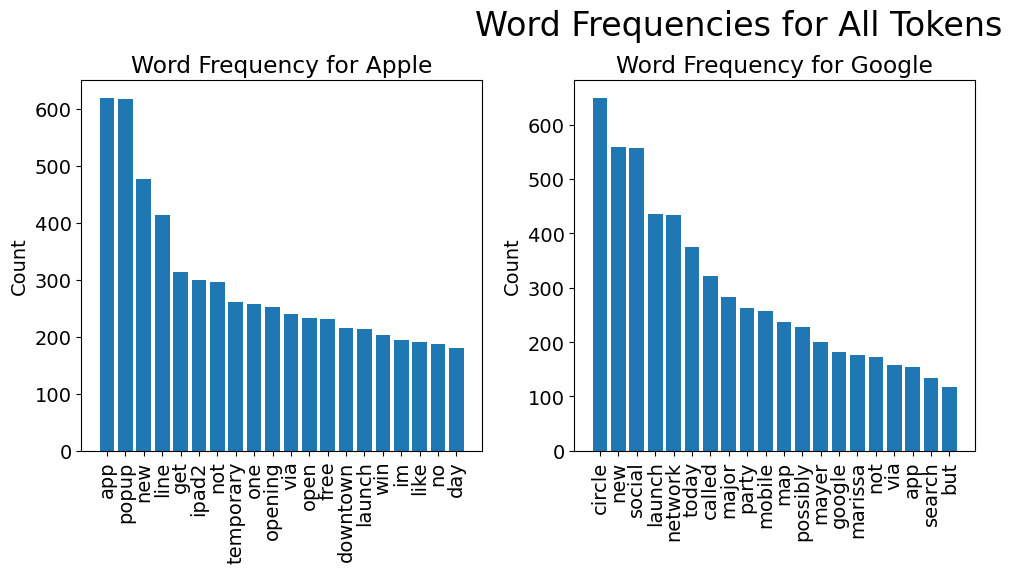
1. The dataset contains 9,092 tweets related to the products from Google and Apple.
2. The dataset has 3 features, namely the tweet, the associated product and the emoticon conveyed.

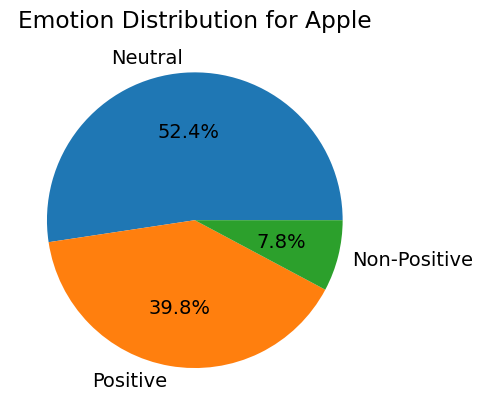
# Data Preparation and Analysis

This step of the analysis followed this sequence:

1. Tweets were categorized according to the brand discussed (eg. Google)
2. Incomplete records were dropped
3. Duplicates dropped
4. The format of the tweets’ associated emotion is cleaned
5. The tweets are tokenized and stop-words removed
6. The tokens are lemmatized and stemmed

# Visualizations

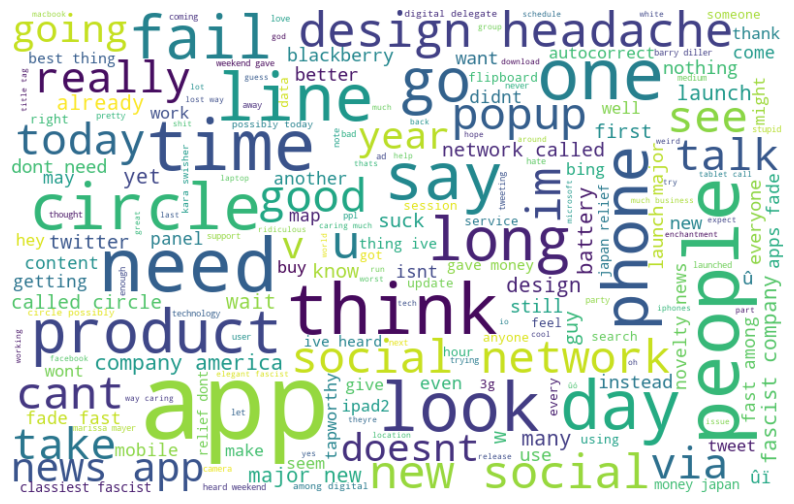




### Positive Sentiment Word Cloud



### Negative Sentiment Word Cloud



# Modeling

The modeling process involved trying various types of model and comparing their performance. Among those tried are the Multinomial Naive Bayes, LinearSVC, Logistic Regression and a Keras neural network.

During the development process, the Naive Bayes model displayed best performance with an accuracy of 87%. The next best was the LinearSVC performing around 85% accuracy.

The adjusting of parameters and tuning did not improve model performance and instead reduced accuracy. As a result a model with default parameters was selected to move forward.

# Deployment

The model was pickled and deployed to a webapp running on fastAPI and Uvicorn. Screenshots of the application are included above.

# Conclusion

The sentiment analysis model demonstrates strong performance in classifying positive sentiments but needs improvement in accurately identifying negative sentiments. Addressing the imbalance in the dataset could enhance overall performance. The substantial difference in performance between positive and negative sentiment necessitates improving classification for negative sentiment.

With an accuracy greater than 80% the model does well to identify tweet sentiment. It is worthwhile to pursue this further and use the analysis to collect customer opinion on products.