Twitter Sentiment Analysis with Natural Language Processing

Musa Irshad

Objectives & Data Analysis

- Dataset obtained from CrowdFlower via data.world.Data Source: Brands and Product Emotions - dataset by crowdflower | data.world
- Over 8,721 Tweets rated by human raters for sentiment: positive, negative, or neutral.
- Tweets primarily from the South by Southwest conference in 2011.
- Focus on Google and Apple products.
- Target engineered into two classes: positive sentiment and negative sentiment.
- Emphasis on positive sentiment due to its correlation with sales and return on investment.
- Our objective is to come up with predictive model which has high accuracy of determining positive sentiment tweets.
- Come up with investment options for GoldenGroup from the positive tweets so they can confidently invest in products and services accordingly.

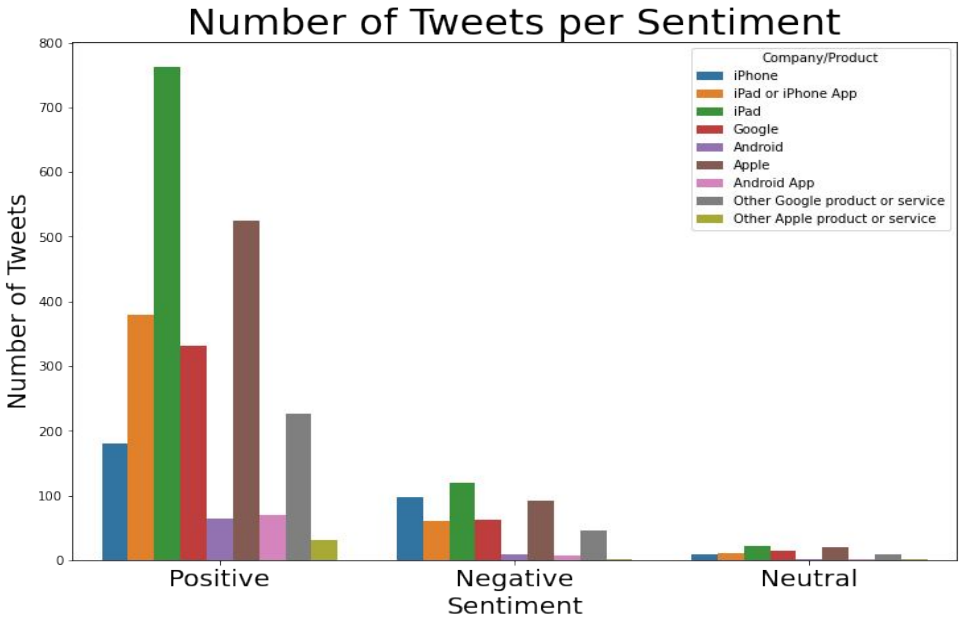
Business Overview

- GoldenGroup wants a predictive model for analyzing recent tweets about technology products.
- The goal is to determine if people have positive sentiments towards different brands.
- The model should be applicable to various brands, aiding in investment decisions.
- GoldenGroup prioritizes positive sentiment as an indicator of potential consumer interest.
- Specifically, they seek insights on positive tweets related to Apple and Google products.

Pre- EDA

Data:

- Positive 2869
- Negative 545



Positives for Apple

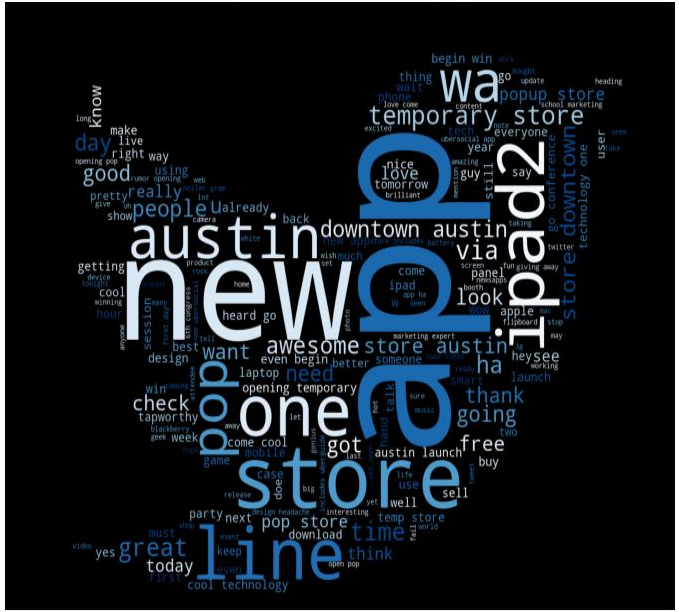
Tokens mentioned:

Ipad: 984

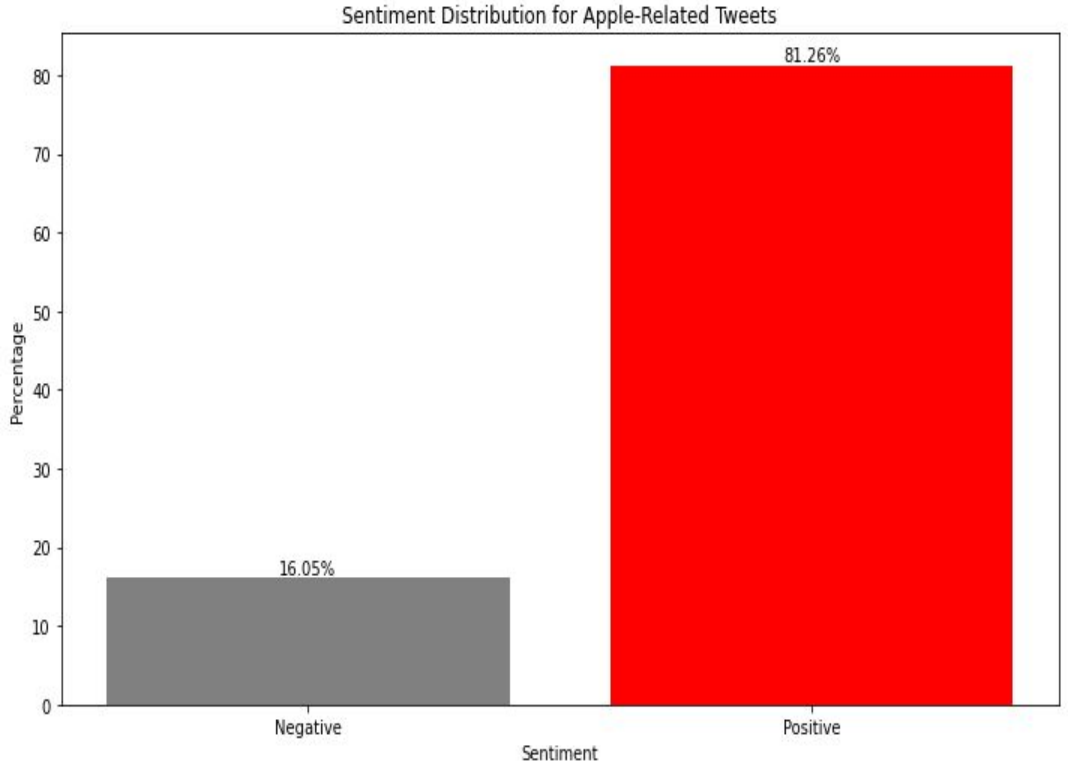
Apple :700

Iphone:450

case :240



Percentage of Tweet by Sentiment(Apple)



Positives for Google

Tokens mentioned:

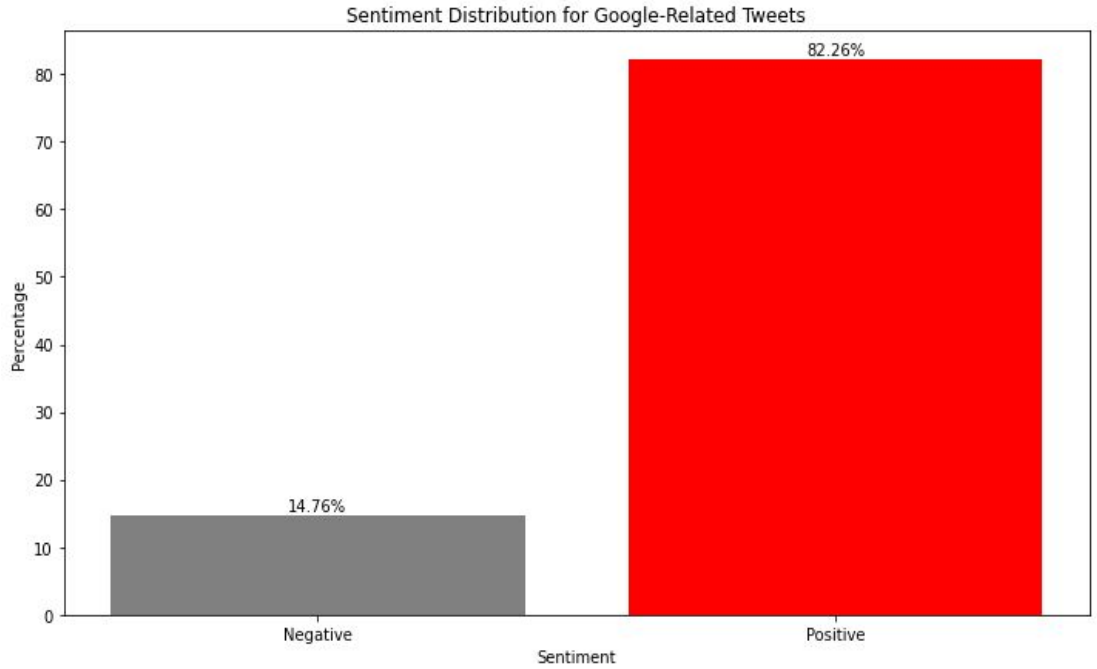
Google: 585

Android : 152

Party: 149

New: 352

Percentage of tweet by Sentiment(Google)

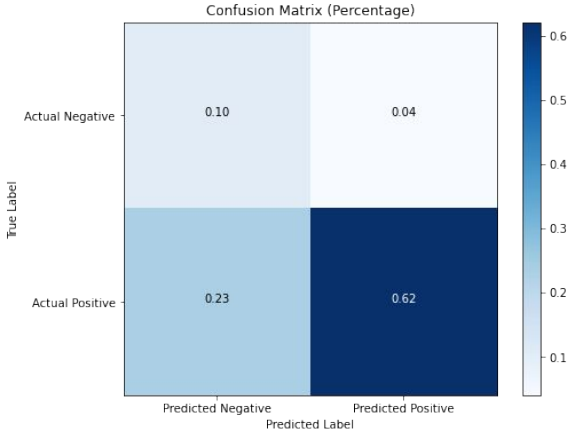


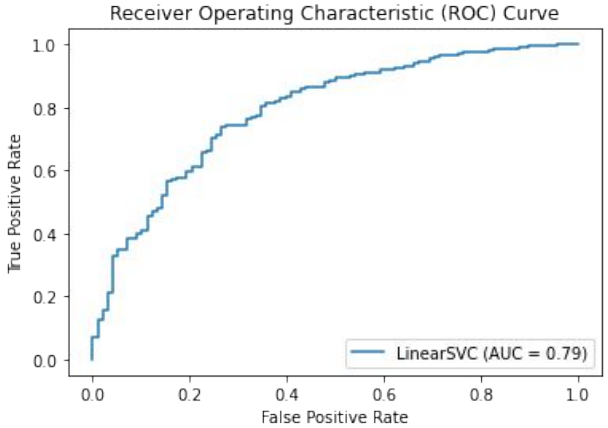
Binary Classification Model

Methods

- We used Smote for imbalanced class
- We used TFIDFVectorizer to prepare the text data for machine learning by converting it into a numerical format (TF-IDF matrix) where each text document is represented as a vector of TF-IDF features.
- We used GridsearchCv to find the best hyperparameters for tuning the model

SVC model





Recall score for positive sentiment: 1.0

Training Score: 0.8362701908957415

Testing Score: 0.856093979441997

Precision score of : 94

Recommendations

- We recommend Svc tuned model which has a Recall score of 1.0 for positive sentiment tweet and accuracy score and precision score of 85.6.
- One of the most used words was 'New', 'Case' for Apple, 'phone' and 'tablet' for google so GoldenGroup can invest in Accessories company for Iphone, Ipad and google products.
- GoldenGroup also has an opportunity of investing in event planner companies for Apple and Google as they both have positive feedback from users about events

Musa Irshad

https://www.linkedin.com/musairshad