Sentiment Classification of tweets about Apple and Google Google Products using NLP

Members;

Brenda Kinoti, John Karanja, Victor Mawira, Michelle Mwendwa, Loise Mbago, Stephen Gathai.

Identifying Key Drivers of Customer Satisfaction

In today's competitive market, businesses need to constantly monitor their their customers' satisfaction to stay ahead. Our project aims to use machine machine learning to identify the key drivers of customer satisfaction, using using online reviews and ratings. This information can help businesses make make data-driven decisions to improve their products and services.

Project Objectives

1

Sentiment Analysis

To classify tweets about Apple and Google products products as positive, negative, or neutral.

2

Natural Language Processing Techniques

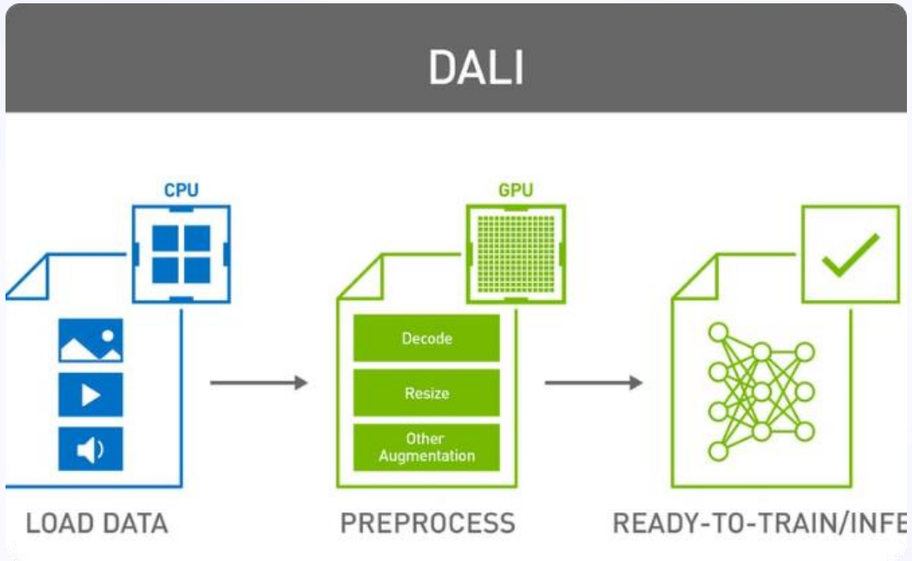
To implement various NLP NLP preprocessing steps, steps, such as tokenization, tokenization, stopword removal, and text normalization, to prepare prepare the tweet data for for effective analysis and and modeling.

3

Prediction System

To create a user-friendly friendly prediction system system that allows users to users to input a tweet, and and the system should provide a sentiment rating rating based on the content of the tweet.

Data collection and preprocessing



Data Collection

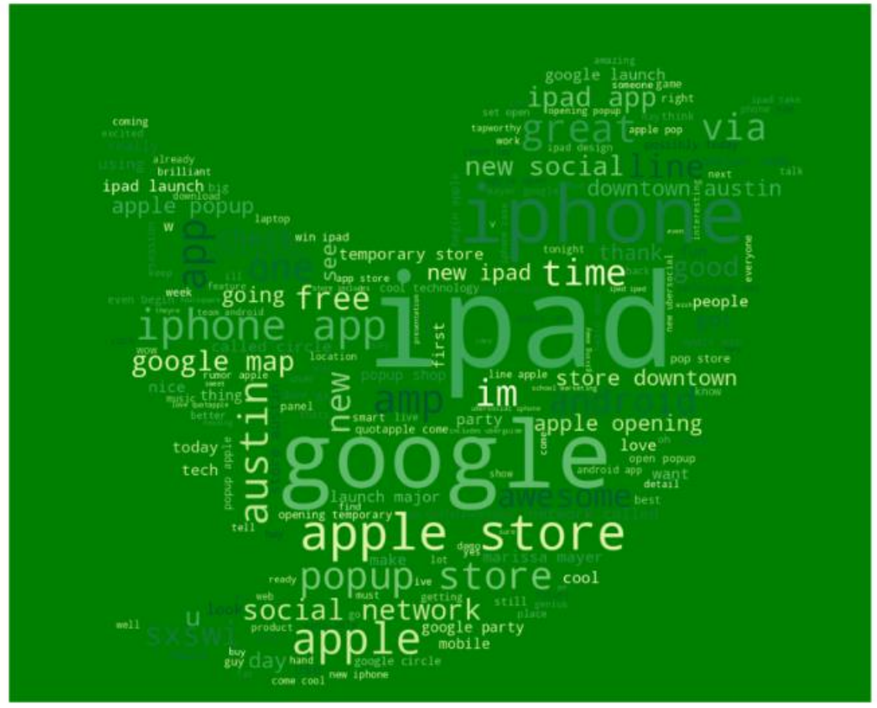
The dataset contains 9092 tweets obtained from from CrowdFlower Open Source Datasets. The tweets tweets are are about either Apple or Google or any of any of their respective products and services.

Data Preprocessing

This included tokenization, stemming, lemmatization, removal of stop words from the tweets. Filling of missing values to prepare the tweets for further analysis.

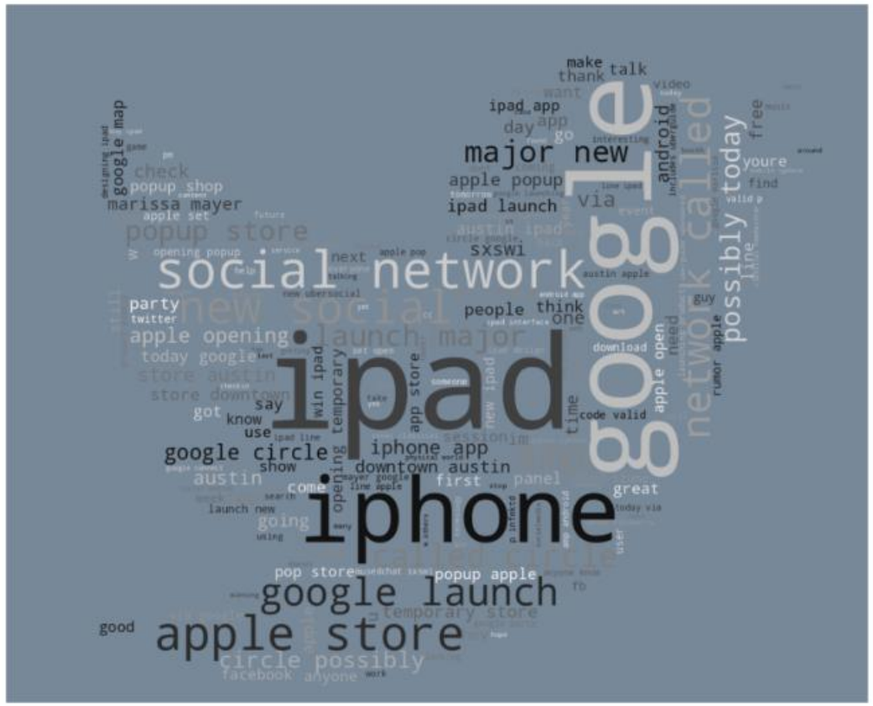
EDA Analysis for Positive, Negative and Neutral Sentiments

Positive Sentiment



Negative Sentiment

Neutral Sentiments



Comments

1. We can see that a lot of people are excited about a new and temporary pop-up store downtown, ipad 2 launch, Google's Marissa Mayer and a new social network called circle.

2. People were unhappy about the new design design for ipad, and also on the iphone battery. The new social circle also has some some negative feedback.

3. Neutral emotions about the social network circle as well as the pop up stores.

Modelling results and evaluation

Model	Accuracy	Precision (weighted avg)	Recall (weighted avg)	F1-Score (weighted avg)
Logistic Regression	0.65	0.62	0.65	0.62
Naive Bayes	0.65	0.62	0.65	0.61
XG Boost	0.66	0.65	0.66	0.63
Random Forest	0.67	0.66	0.67	0.65

- Modelling process:
  - Train/test Split
  - TF-IDF Vectorizer
  - Building the classifiers
  - Evaluation
- Highest performing model was Random Forest with an overall 67% accuracy in classifying tweet sentiment
- The model is better at classifying negative tweets than positive tweets

Recommendations

Brand/Product Awareness

Prioritize marketing and investment strategies around hp products like the iPad, Apple, and iPad/iPhone app slightly mentioned.

Sentiment Analysis

Leverage the predominantly neutral and positive sentiment for positive investment and marketing decisions. Address negative sentiment to uphold brand reputation.

Frequent Terms

Utilize frequently mentioned mentioned terms highlighted highlighted in the word cloud cloud for targeted marketing marketing and resonating with the audience..

Data-driven Decision Making

Utilize the cleaned and transformed dataset for advanced analyses like trend trend analysis, topic modelling, and sentiment analysis for customer insights. insights.

Engagement Strategy

Develop strategies to engage engage neutral segments, converting them into positive positive brand advocates for for increased influence.

Data-driven Decision Making

Establish ongoing feedback mechanisms using real-time user feedback to enhance products and responsiveness to customer needs.

Next Steps/Future work

Develop a program that can automatically grab tweets for analysis

Build a model to evaluate posts/comments on other forms of social media

Incorporate trending data to see how opinions/brands of products change over time

Train and test models that incorporate engagement on social media eg retweets

Thank You!!!