Analyzing Consumer Sentiment using Twitter

Pharoah Evelyn

Overview

This notebook aims to analyze sentiment from Twitter data, and select a model that can best predict future sentiment.

- Run Natural Language Processing to discover insights
- Machine Learning with Scikit-Learn, Keras

Outline

- Business Problem
- Data
- Methods Used
- NLP Analysis

- Conclusions
- Recommendations
- Next Steps

Business Problem

Tech news company Engadget analyzes trends to provide coverage on their content platform.

Mining sentiment on Twitter can help Engadget quickly gauge user opinion.

Having deep understanding of general consensuses will help their recommendation process.

Data

SXSW Event

Tweets surrounding new Apple & Google products

Comparing consumer sentiment between products

9093 samples in total

Methods Used

Phase 1 - NLP

- Exploratory Data Analysis
- Natural Language Processing
- Word Clouds

Phase 2 - Modeling

- Addressing Class Imbalances
- Scikit Learn
- Keras

NLP Analysis - Apple

Apple Words

Apple Words without Products

NLP Analysis - Apple

Positive Word Cloud

Negaitive Word Cloud

NLP Analysis - Google

Google Words

Google Words without Products

NLP Analysis - Google

Positive Word Cloud

Negaitive Word Cloud

Conclusions

NLP Findings

Apple

- We see much excitement for a popup store & its location downtown, which is selling the latest products.
- Unfortunately, we can see that Apple users are experiencing issues with their battery life.
- Some users also do not approve of Apple's new product design. Also, it seems some people are calling Apple fascist.

NLP Findings

Google

- There is much buzz for Google Circle (a social network) in this dataset, along with Google Maps.
- We also see much mention of Marissa Mayer, a prolific businesswoman.
- Some users are experiencing interface issues in the same breath.

NLP Findings

What can we gain from this information?

- We can perceive a level of mixed reception since negative sentiments show discontent with the launch of products, new features & new designs of either company.
- However positive sentiment shows lots of excitement about the products & services nonetheless, as well as peripheral features such as the event location & stores
    - Keep in mind the data was overwhelmingly positive

Conclusions

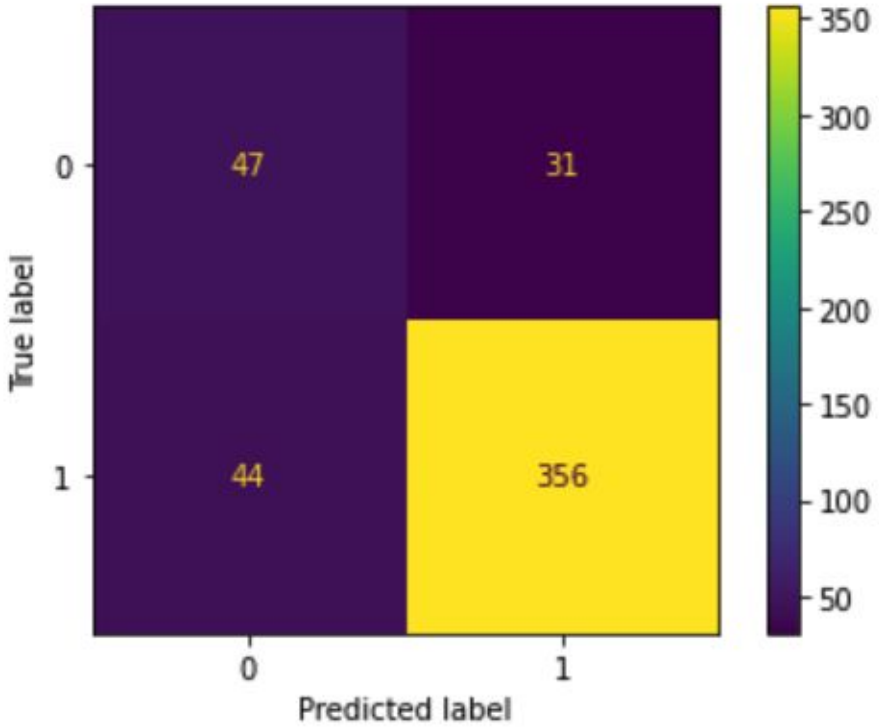
The best model is a Keras Model with Parameter Tuning.

- The confusion matrix reports a highest correct classification of negative sentiment.
- The Scikit-Learn models all performed well on the accuracy, but most of their accuracy scores don’t perform better than either Keras model.
    - The Linear SVC model has 86.82% accuracy, just above the tuned Keras model.
- The Logistic Regression model is the second-best performing model for predicting negative sentiment, with 44 correct predictions.

Output

Interpretation of churn with a Tuned SVM:

- **47 True Negatives:** Tweets with negative sentiment
- **31 False Positives:** Predicted tweets with positive sentiment, but they're negative
- **44 False Negatives:** Predicted tweets with negative sentiment, but they're positive
- **356 True Positives:** Tweets with positive sentiment



Recommendations

Usage for our model:

- Use this model to predict Twitter sentiment for future events, or even with a live API
- Use this model to predict if negative tweets are about Apple or Google
    - You can do this by splitting the original dataframe only to include negative tweets and have the products be the target column
- Build a web app using techniques in this notebook to predict twitter sentiment

Recommendations

- To further increase model performance, we could implement a grid search on all models.
- Utilize a live API to collect tweets and store in a database for future activities.

Next Steps

- Engadget can publish various news articles targeted to their audience about the different findings this notebook uncovers
    - Apple articles targeted at Apple enthusiasts
    - Google articles targeted at Google enthusiasts
    - Negative or positive press about this particular SXSW as a whole
- Run similar analysis on future events or tech releases. Thus providing timely updates to enthusiasts & avid readers.

Questions?