



# Sentiment Analysis of Twitter Data w/Machine Learning

Robert Cauvy



# The Business Problem

People are increasingly expressing their thoughts on social media platforms and those conversations can severely affect public which impacts sales. It is critical for brands to stay informed of how they are perceived and quickly adapt to shifting trends.



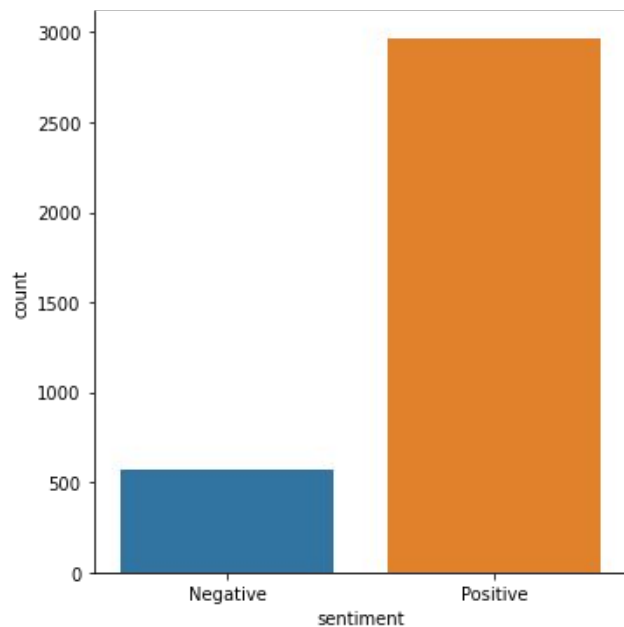
# Business Objectives

- Obtain actionable insights from public opinion via twitter data Apple can use to stay ahead of its competition
  - What led the participants to negative sentiments about their brand or products
  - What led the participants to negative sentiments about their brand or products
  - What did sentiments did participants have towards competing brands or products
- Develop Machine Learning models that can be used to gather future public feedback



# The Data

- +9,000 tweets captured from events held by Apple and Google at the 2011 SXSW Conference
- The tweets were labeled by humans into one of the following categories:
  - Negative Sentiment
  - Positive Sentiment
  - Neutral Sentiment
  - Sentiment is Unknown



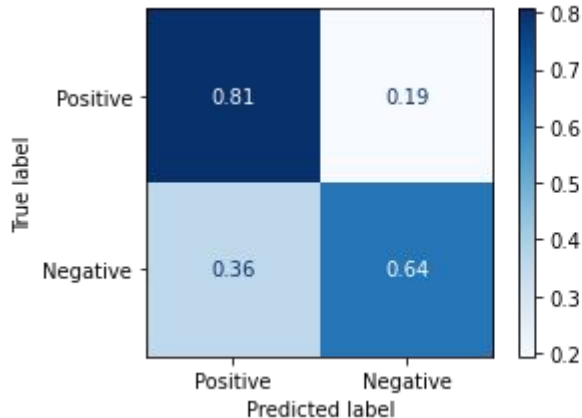
# Methods

- Trained basic Machine Learning models to labeled twitter data
  - Classifiers include:
    - Naive Bayes
      - Vanilla, GridSearchCV, Random OverSample
    - Logistic Regression
      - Vanilla, Balanced Class Weight, GridSearchCV
    - Random Forest
      - Vanilla, Balanced Class Weight, GridSearchCV
- These classifiers were selected for interpretability and to serve as a baseline for tuning more complex models

# Model Results

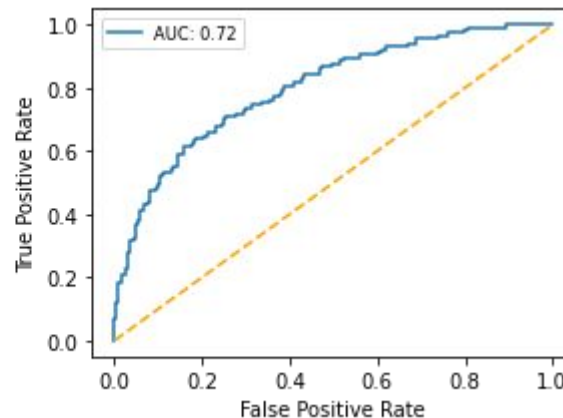
## Logistic Regression Model

- Correctly predicted 64% of the Negative Tweets
- Correctly predicted 81% of the Positive Tweets
- Least overfit

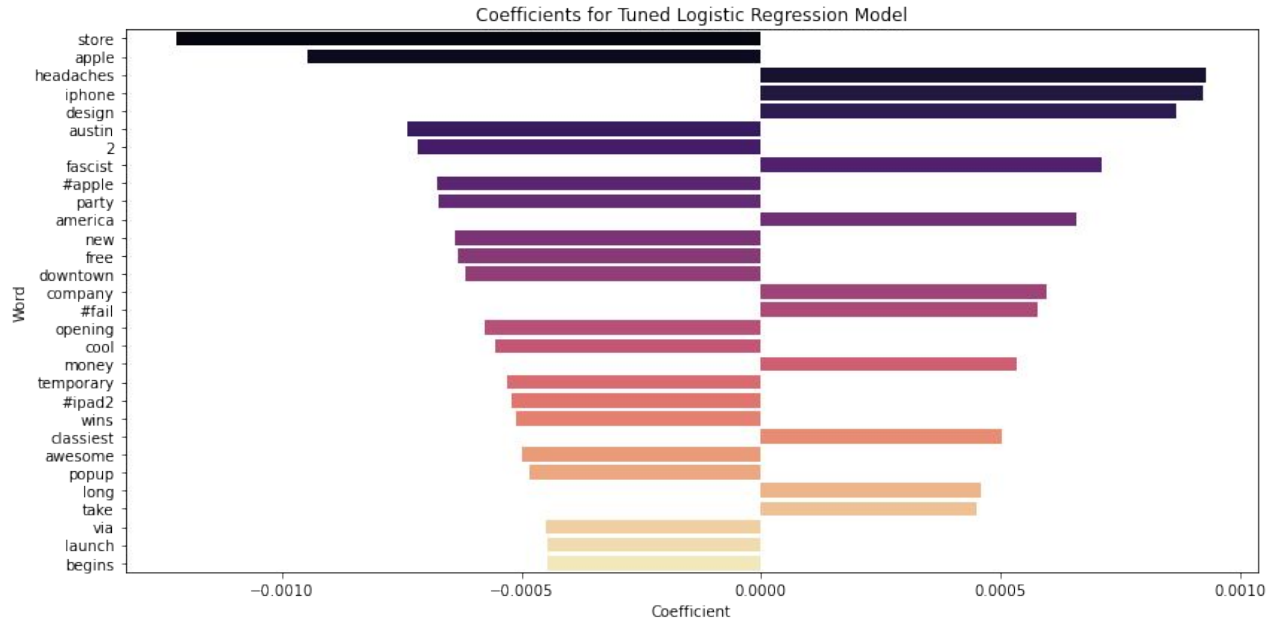


## Naive Bayes Model

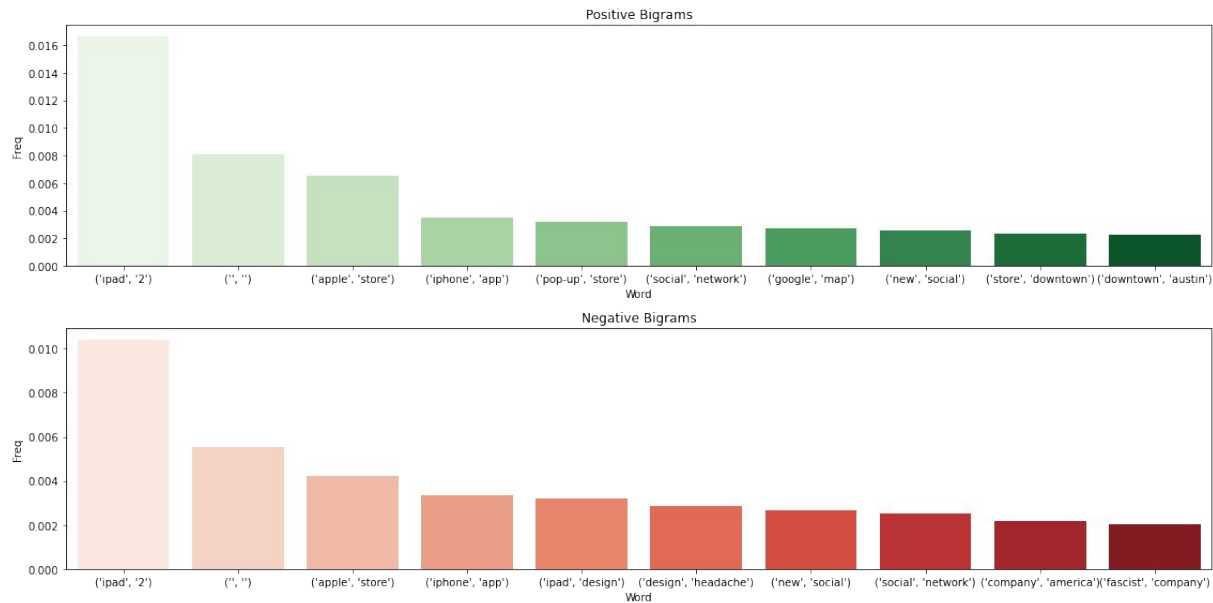
- Correctly predicted 68% of the Negative Tweets
- Correctly predicted 85% of the Positive Tweets
- Overfitting to the training data



# Most Important Features



# Bigram Results

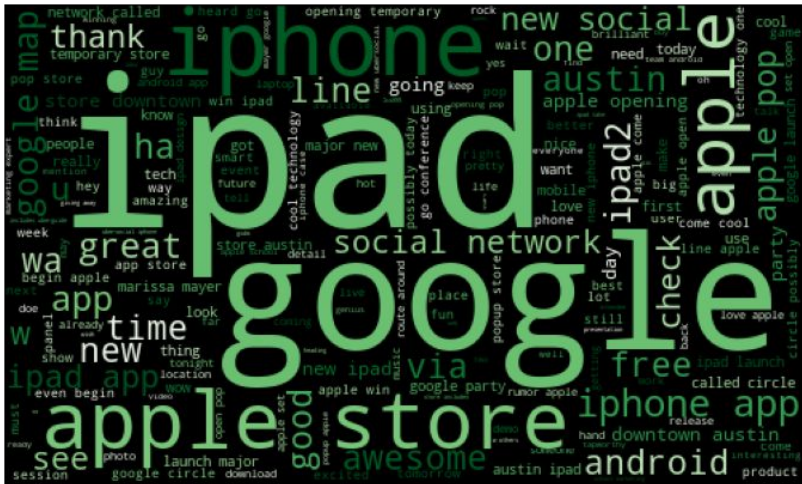




# Insights From Negative Tweets



# Insights From Positive Tweets



# WordCloud Comparison

## Positive

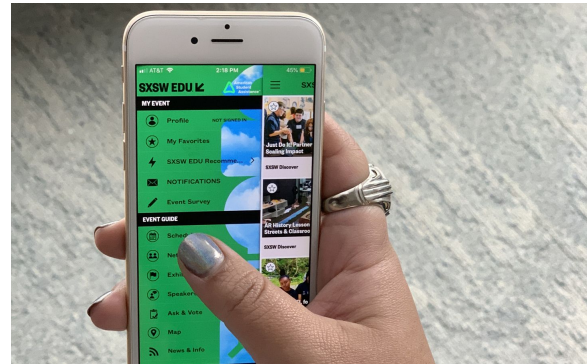


## Negative



# Conclusions and Recommendations

- Machine Learning models can monitor public sentiment
- Can track threats and weakness of their competition
- Product Designs Insights
  - Improve iPhone battery
  - Retool the iPad design
  - Ensure apps in the app store are functioning properly
- Marketing Insights
  - Pop-up stores generate a lot of excitement
  - Parties and events and giveaways should be further looked into





# Next steps

- Gather more data
- Develop a multiclass model
- Test more complex preprocessing techniques and different classification models



# Thank you!

For questions or comments please contacts:

Robert Cauvy

[rcauvy@gmail.com](mailto:rcauvy@gmail.com)

<https://github.com/rcauvy/product-sentiment-analysis-nlp>

