NLP Project

By: Marissa Bush

Table of Contents

01. Business Problem

Apple is hosting a tech conference and they have asked for a machine learning model to determine whether tweets have a positive or negative sentiment when introducing a new product.

02. Data

3 Columns, 9093 Rows:

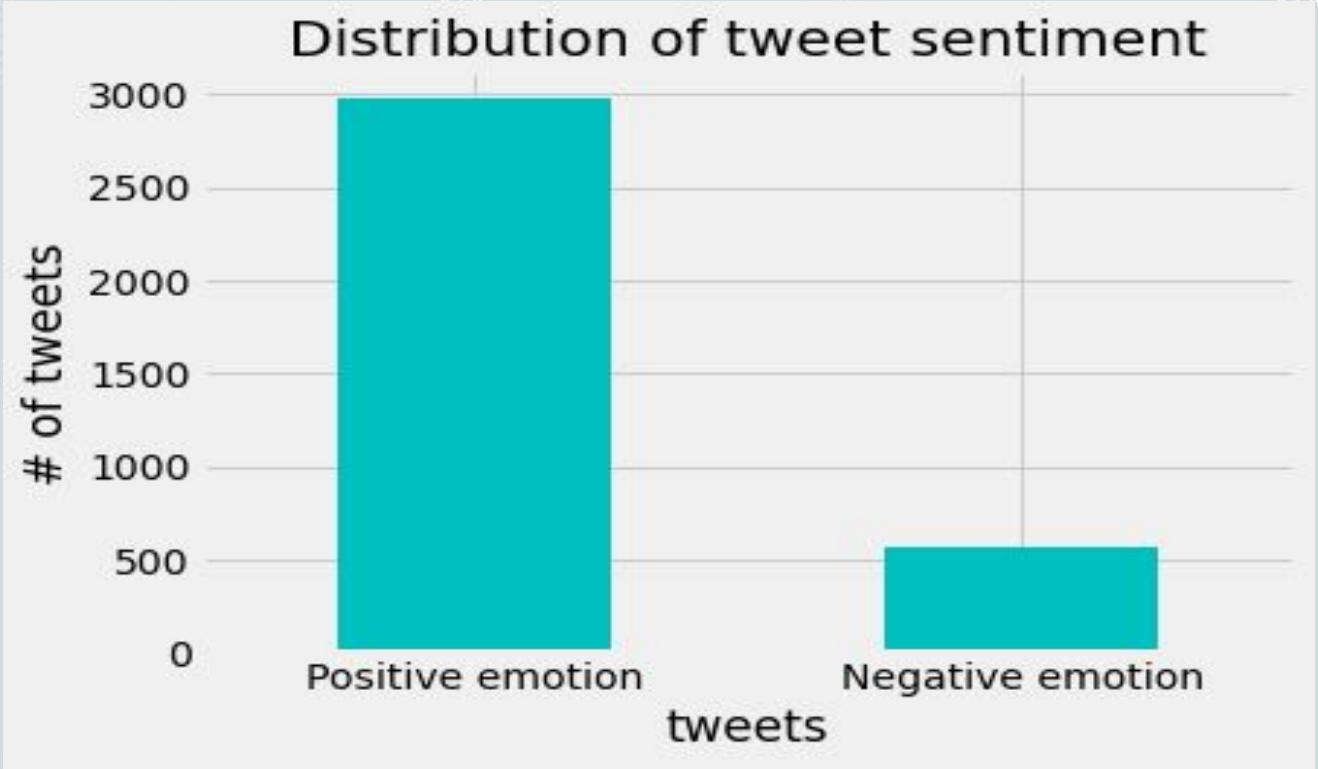
- Over 9000 tweets were marked as either positive, negative, or neutral.
- Target column presented an imbalanced class problem. Most of the dataset contained tweets that were neutral.

Positive tweet example

“Beautifully smart and simple idea RT @madebymany @thenextweb wrote about our #hollergram iPad app for #sxsw”

Negative tweet example

"@sxsw I hope this year's festival isn't as crazy as this year's iPhone app. #sxsw"


Distribution of tweet sentiment
3000
2500
2000
1500
1000
500
0
# of tweets
Positive emotion
Negative emotion
tweets

03. Methods

OSEMN method for data analysis

OSEMN method for data analysis

Obtain

Data collection

Scrub + Explore

Cleaned data

Model

Iterative modeling with logistic regression, decision tree classifiers, and random forest classifiers.

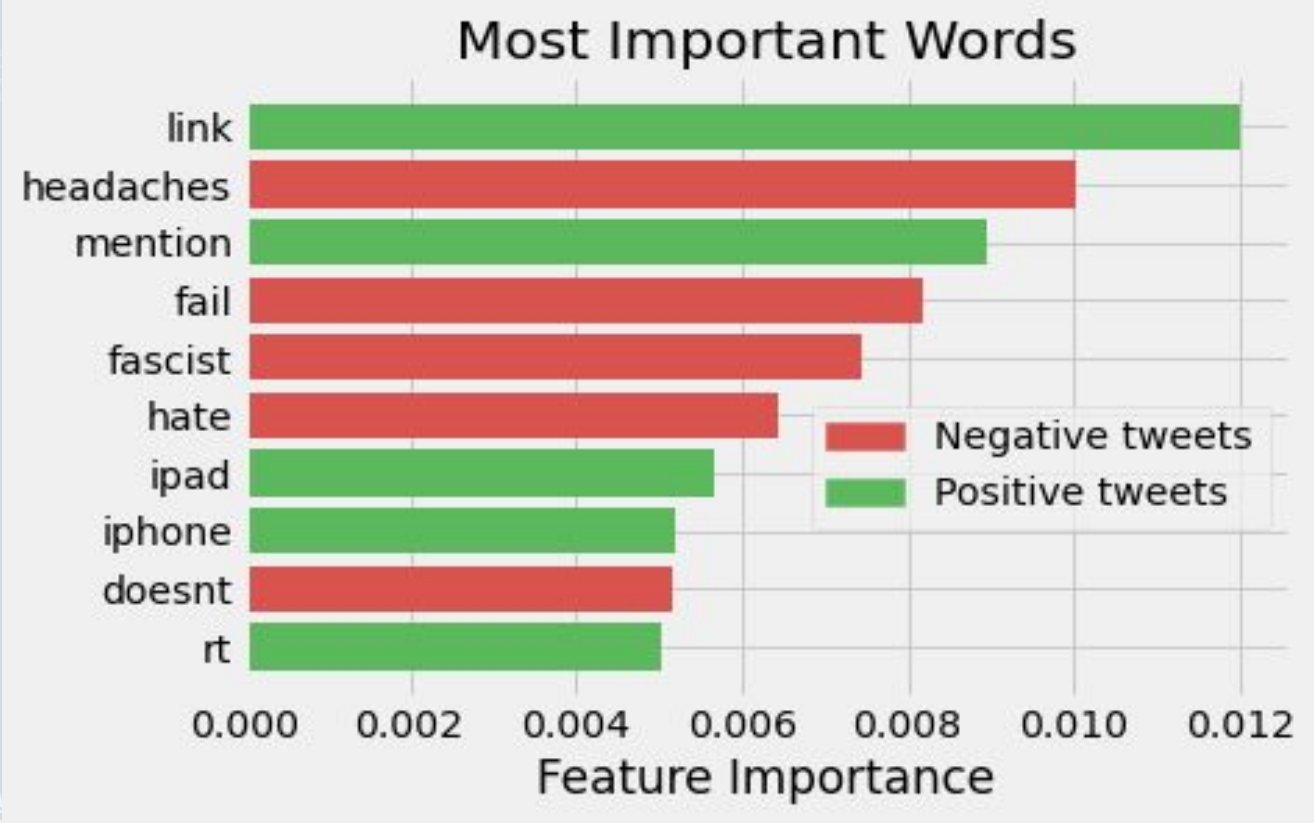
Interpret

Choose best model

@sxsw I hope this year's festival isn't as crashy as this year's iPhone app. #sxsw

sxsw hope years festival isn't crashy years iphone app sxsw

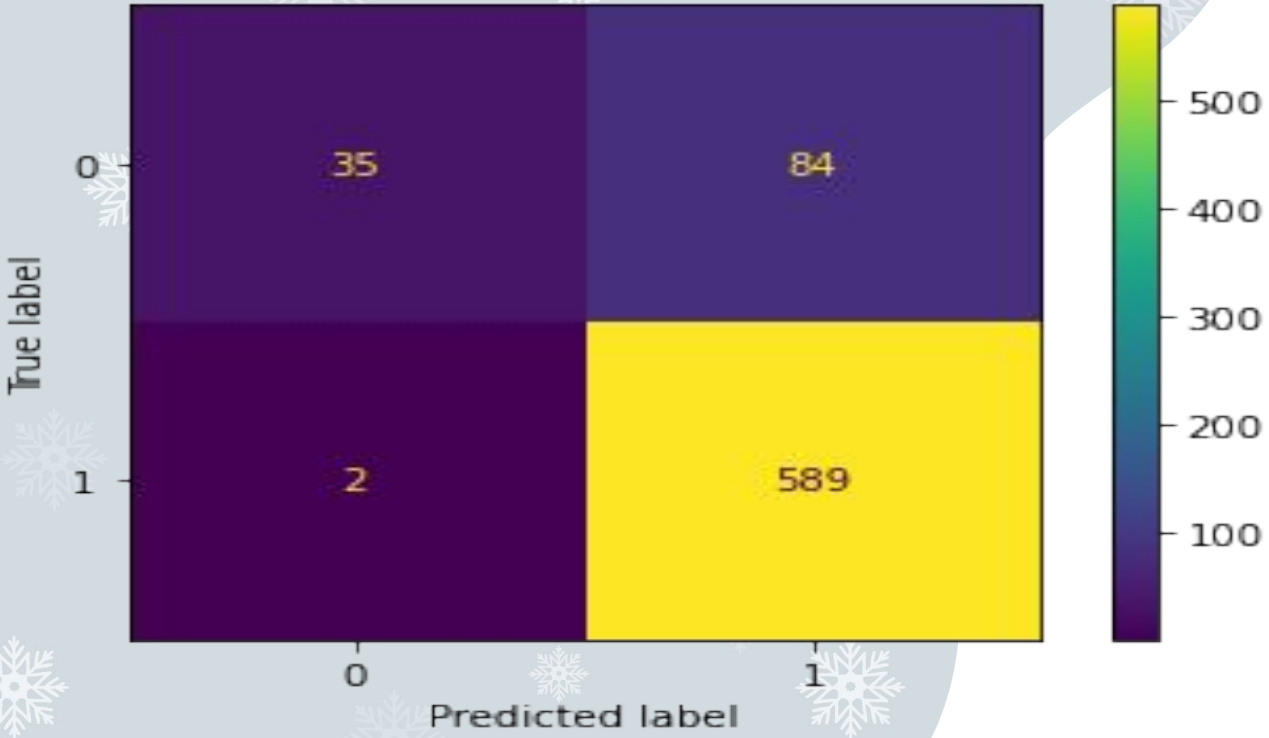
04. Findings


Most Important Words
link
headaches
mention
fail
fascist
hate
ipad
iphone
doesnt
rt
Negative tweets
Positive tweets
0.000
0.002
0.004
0.006
0.008
0.010
0.012
Feature Importance

Best Model

Random Forest Classifier
F1 Score: 0.933

Confusion Matrix of Random Forest Classifier



05. Conclusions

Conclusions

- Using machine learning, we can predict whether a tweet is positive or negative.
- Negative tweets have strongly negative words.
- Positive tweets have less distinctive words that indicate positivity.

6. Future Work

Future Work

For future work, I would look at the labels of the tweets. There are some tweets that appeared to be incorrectly labeled in the original dataset.

Thank you!

Contact

Marissa Bush.02@gmail.com

https://github.com/Marissa841/phase_4_project

https://marissa841.github.io/