



NLP Project

By: Marissa Bush

Table of Contents

01.
Business
Problem

02.
Data

03.
Methods

04.
Findings

05.
Conclusions +
Future Work



01. Business Problem

This project determines which tweets can be correctly identified to contain either positive sentiments (“Positive emotion”) or negative sentiments (“Negative emotion”) using machine learning.

02. Data

Dataset used is from CrowdFlower via data.world.

People were asked to rate tweet sentiment as either positive, negative, or 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 crashy as this year's iPhone app. #sxsw"

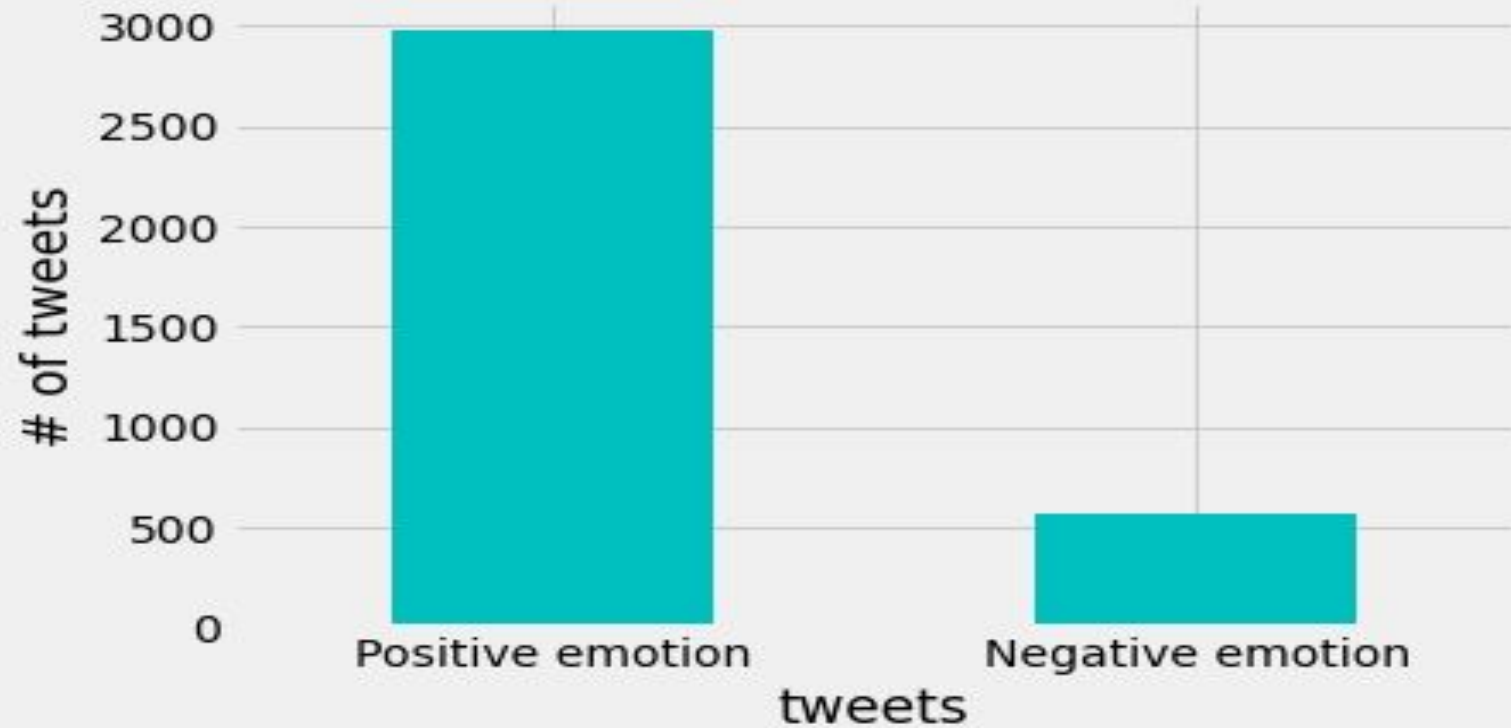
Data highlights

3 Columns, 9093 Rows:

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



Distribution of tweet sentiment





03. Methods

OSEMN method for data analysis



OSEMN method for data analysis

Obtain

Data collection

Model

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

Scrub + Explore

Cleaned data

Interpret

Chose best model

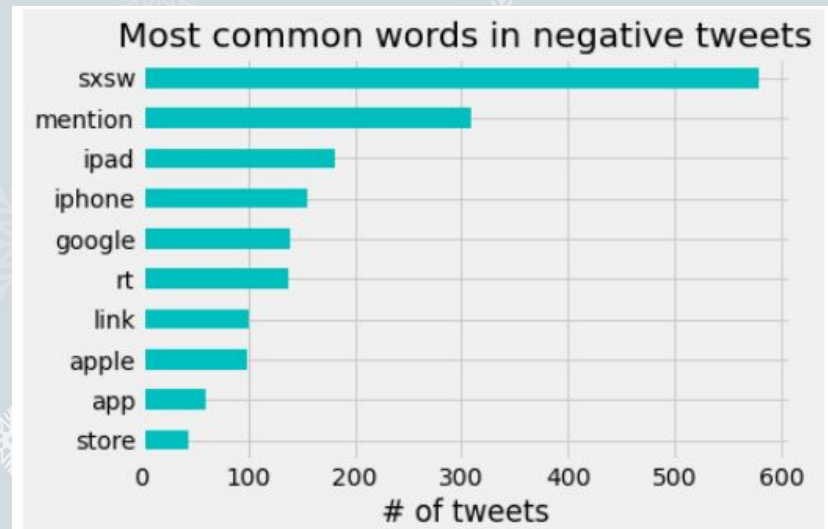
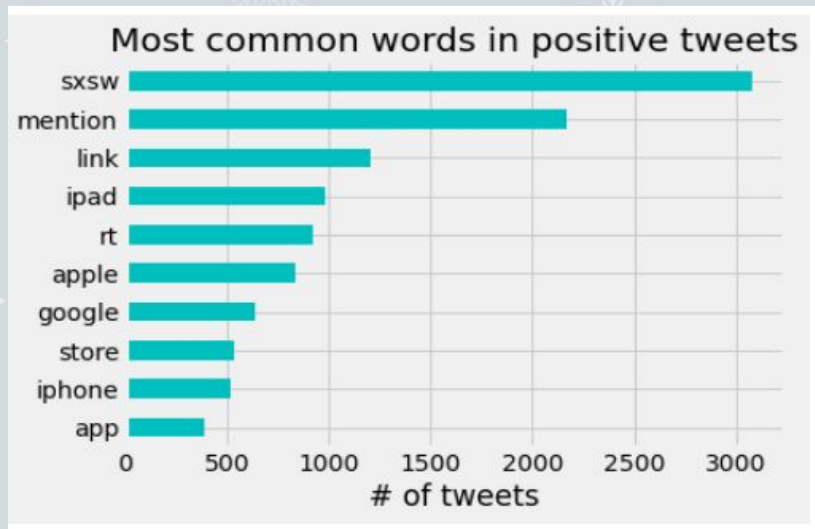
**@sxsw I hope this year's festival isn't as crashy as this
year's iPhone app. #sxsw**



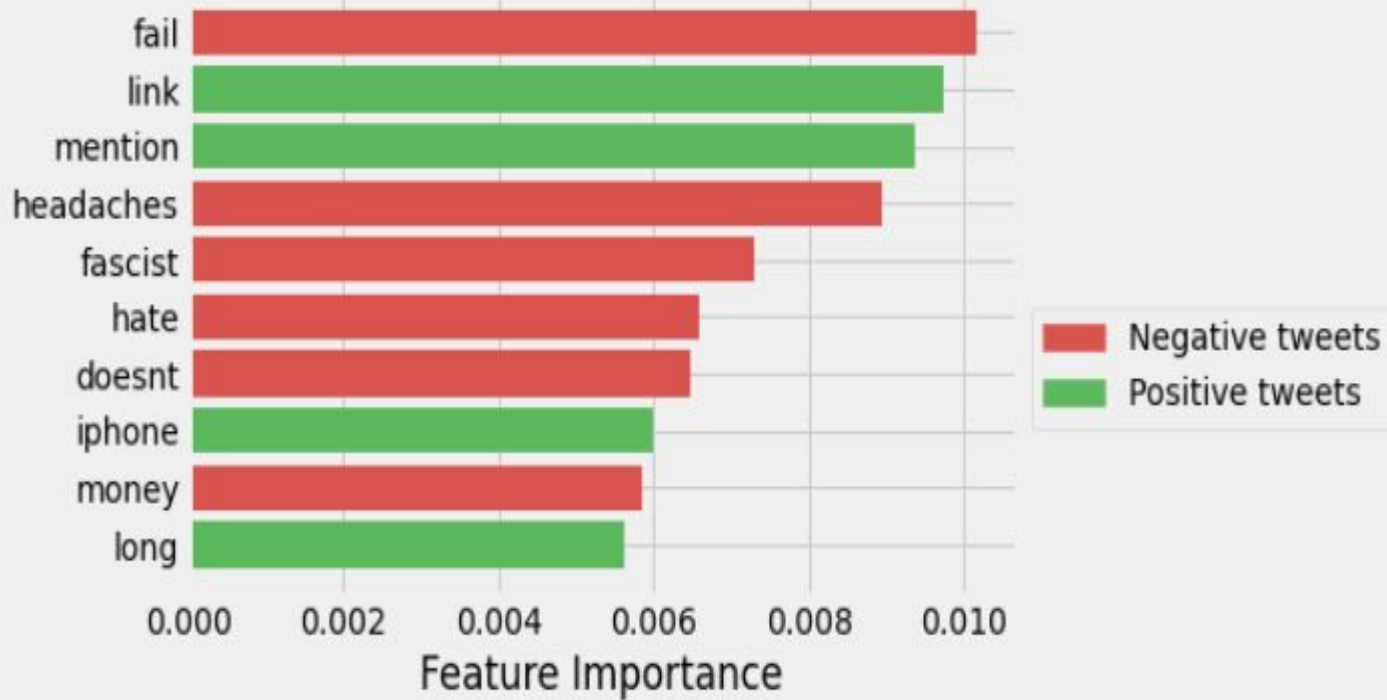
**sxsw hope years festival isnt crashy years iphone app
SXSW**



04. Findings



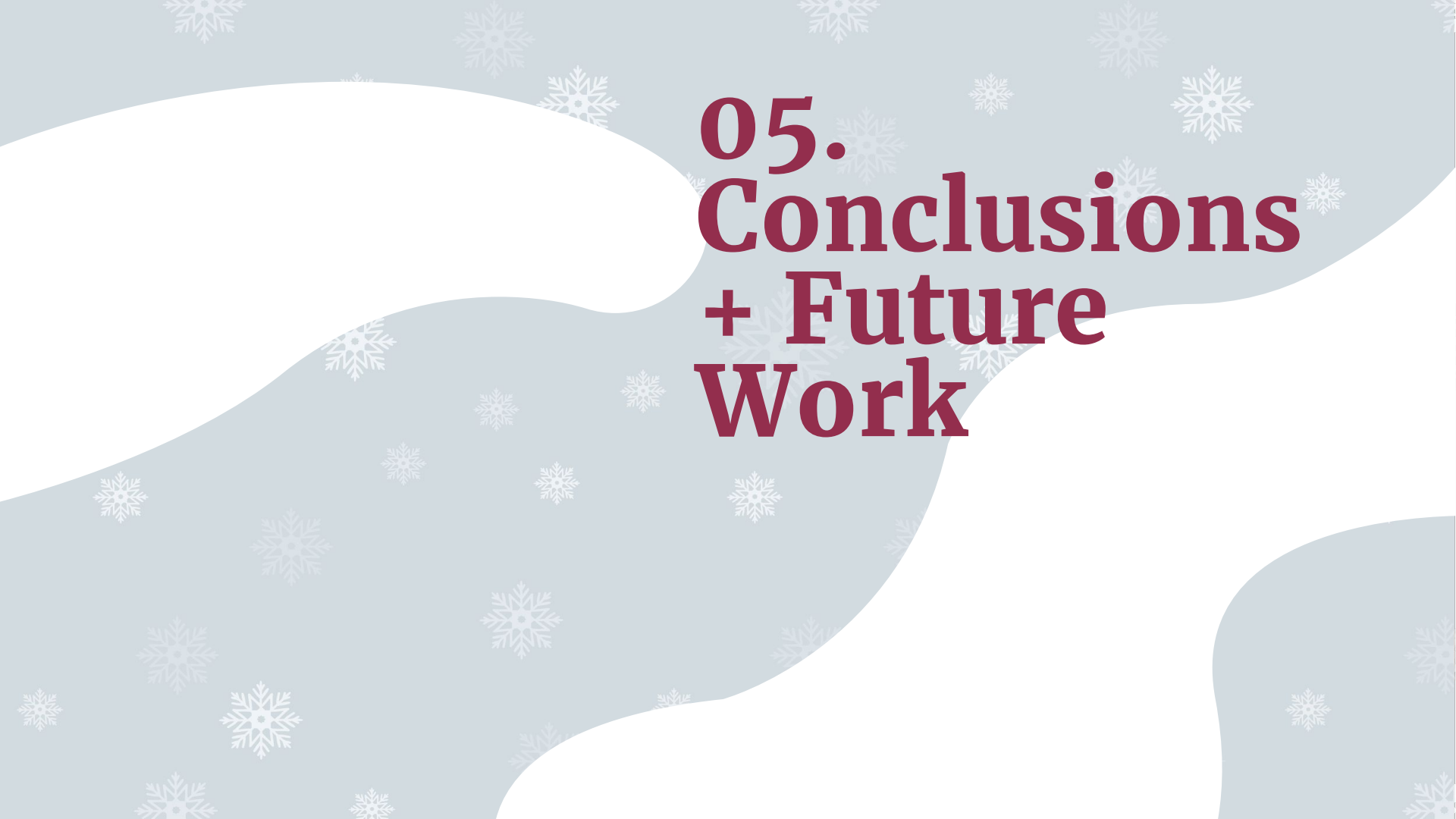
Most Important Words





Best Model

Random Forest Classifier
F1 Score: 0.93



05. Conclusions + Future Work



Conclusions + Future Work

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

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



Thank you!



Contact

Marissabush.02@gmail.com

https://github.com/Marissa841/phase_4_project

<https://marissa841.github.io/>