Twitter Sentiment Analysis

Natural Language Processing

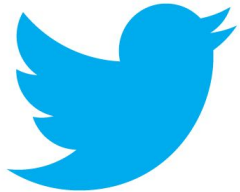
Allison Gao
Meaghan Ross
Raylin Soriano

BUSINESS PROBLEM

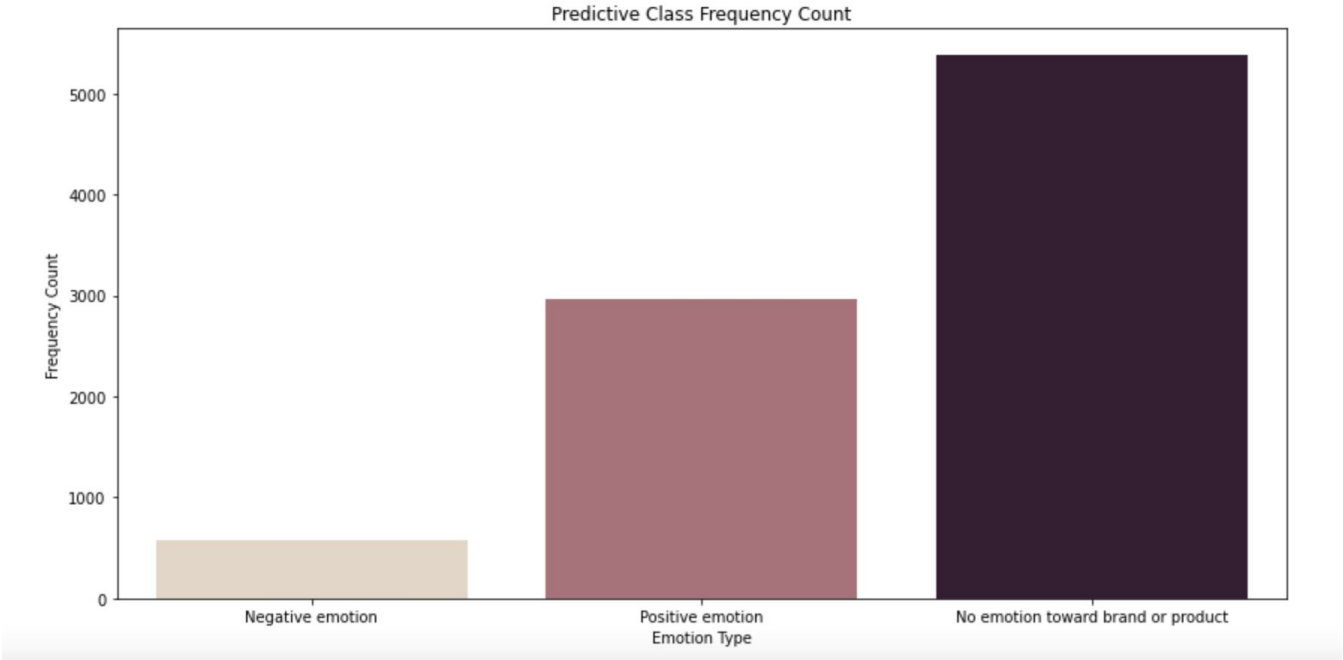
- **Client:** Samsung
- **Purpose of Analysis:** to provide recommendations to Samsung seeking to purchase a consumer sentiment tool that detects product sentiment based on Tweets.

DATA

- **Data Source:** the CrowdFlower via data.world
- 9094 Tweets
- 3 columns:
  - **A.** Text
  - **B.** Company each Tweet was directed at
  - **C.** Human rated emotion sentiment



EMOTION SENTIMENT



- Imbalanced Predictive Class

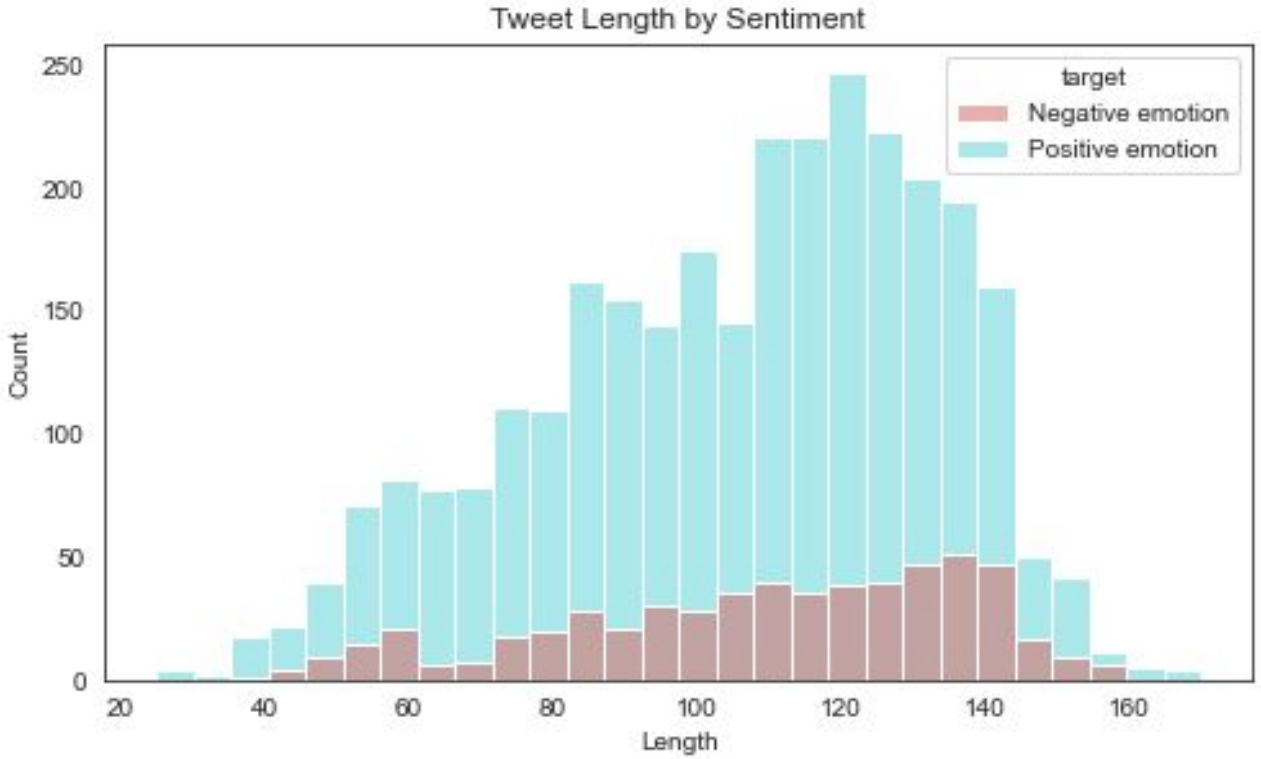
EDA

Uniquely Positive vs Negative Tweet Contents

Positive Tweets

Negative Tweets

Tweet Length Distribution



MODEL PERFORMANCE

Score Type	Baseline	Multinomial Naive Bayes	Logistic Regression	Random Forest	XGBoost
Accuracy Score	67%	65%	61%	67%	67%
Macro Precision Score	65%	76%	51%	68%	59%

Preprocessing

Tweet Tokenizer

.@wesley83 I have a 3G iPhone . After 3 hrs tweeting at #RISE_Austin , it was dead ! I need to upgrade . Plugin stations at #SXSW

Remove Stopwords

@wesley83 3g iphone . 3 hrs tweeting #rise_austin , dead ! need upgrade . plugin stations #sxsw .

Lemmatization

@wesley83 3g iphone . 3 hr tweet #rise_austin , dead ! need upgrade . plugin station #sxsw

Modeling

TF-IDF Vectorizer

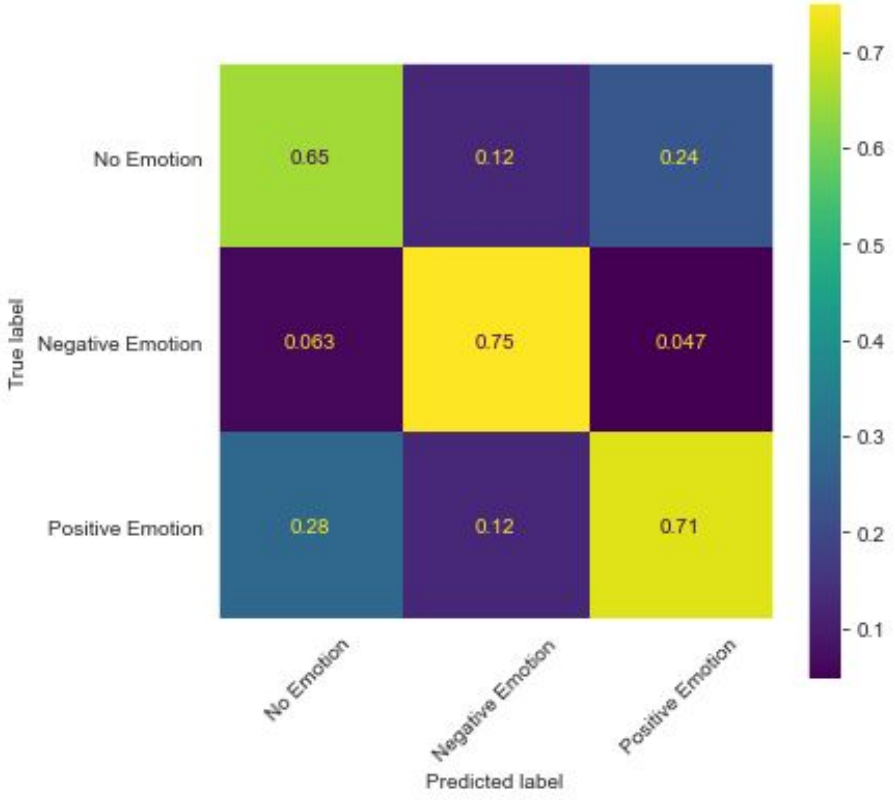
SMOTE

Multinomial Naive Bayes

- Positive Emotion
- No Emotion
- Negative Emotion

MODEL RESULTS

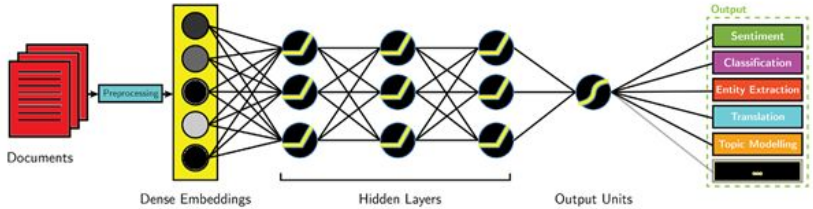
- **Best model:** Multinomial Naive Bayes
- **Accuracy Score:** 66%
- **Macro Precision Score:** 70%



CONCLUSIONS

Multinomial Naive Bayes with a TF-IDF Vectorizer and SMOTE is the best model to detect tweet sentiment.

FUTURE WORK



1. Utilize deep-learning such as Word2Vec create models that have stronger predictive powers.

2. Include tweets from the “I can’t tell” class to better prepare the model for unseen data.

THANK YOU!

Allison Gao

Email:

allison.gao@nyu.edu

GitHub:

@allisongao4015

Meaghan Ross

Email:

mer423@nyu.edu

GitHub:

@mross715

Raylin Soriano

Email:

sorianoraylin@gmail.com

GitHub:

@raylinsoriano