TWITTER SENTIMENT ANALYSIS

SULEYMAN QAYUM

BUSINESS UNDERSTANDING

❖ Gallup, a global analytics and advice firm, plans to conduct a large-scale survey regarding public sentiment towards the political parties in the United States.

❖ In order to achieve this, they require a model that can rate the sentiment of a Tweet based on its content.

❖ Once training of the model has been completed, it is to be deployed as a sentiment analysis tool for that receives data from the Twitter API.

DATA UNDERSTANDING

❖ Dataset is a collection of 1,600,000 tweets, all of which are labelled with a corresponding sentiment.

❖ Dataset was relabeled using *VADER* Sentiment Analyzer

    ❖ *Improved accuracy of labels*

    ❖ *Addition of NEUTRAL sentiment label*
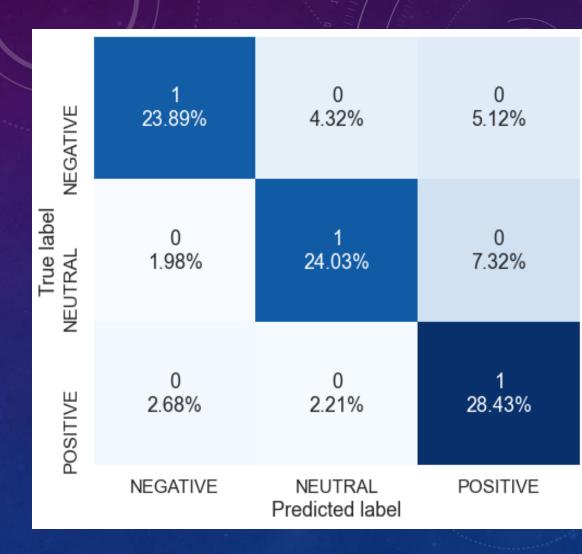
DATA PREPARATION

❖ Data was processed extensively

- *Removal of twitter handles*
- *Removal of web addresses*
- *Removal of special characters, numbers, & accented characters*
- *Removal of commonly used words*
- *Lemmatization*

RESULTS

❖ Overall, the model predicts sentiment well:

- *Accuracy = %78*
- *Tends to predict POSITIVE sentiment*
- *Generalizes well*

LIMITATIONS & NEXT STEPS

- ❖ Fine-Grained Sentiment Analysis
- ❖ Emotion Detection
- ❖ Identification of Domain-Specific Tweets
- ❖ Utilization of Emoticons

THANK YOU

Email: sqayum33@gmail.com

Github: @sqayum