Twitter-Sentiment-Analysis

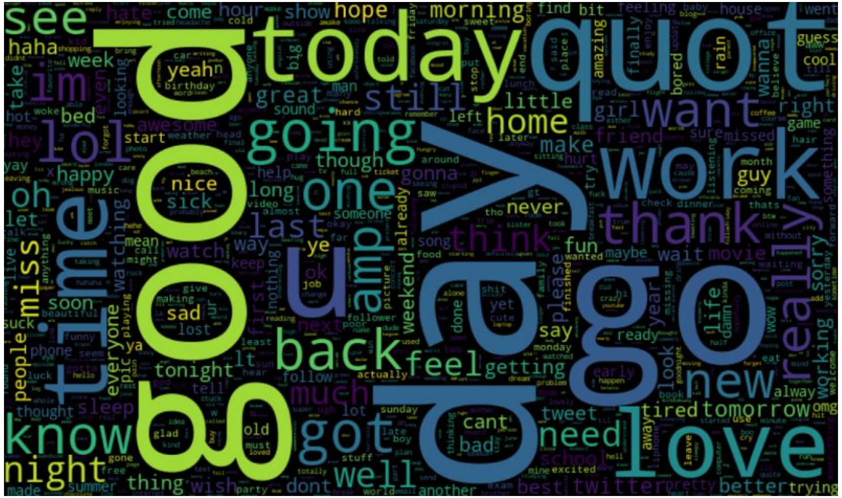
Group 08

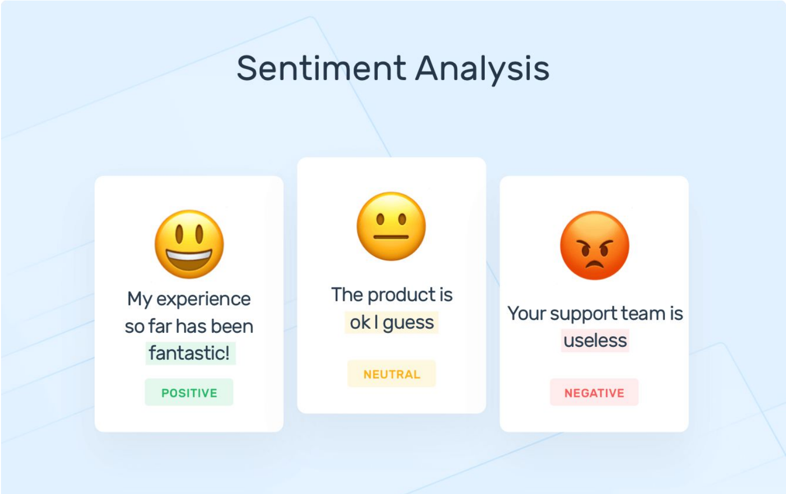
Rebecca Osborne, Sofía Pita, Samuel Maness, Nico Schmid

Github: https://github.com/hallowerised/Twitter-Sentiment-Analysis

What is Sentiment Analysis?

- It is the field of study that covers the use of natural language processing to identify the emotional tone behind a text.
- Another used word to describe it is **Opinion Mining**





Motivation

- Train a neural network to detect sentiments in tweets from given dataset (in English)
- Extract tweets from the Twitter API via search queries, users and hashtags by using **tweepy**.
- Classify these extracted tweets according to their sentiment

Twitter API

Dataset

- Kaggle - Sentiment140
- 1.6 million tweets labelled according to sentiment either negative or positive
- csv file contains
  - target - label 0 or 4 sentiment
  - ids - id of the tweet
  - date - date of the tweet
  - flag - query
  - user - user of this tweet
  - text - text of the tweet

Source: https://www.kaggle.com/datasets/kazanova/sentiment140

Data / Methods

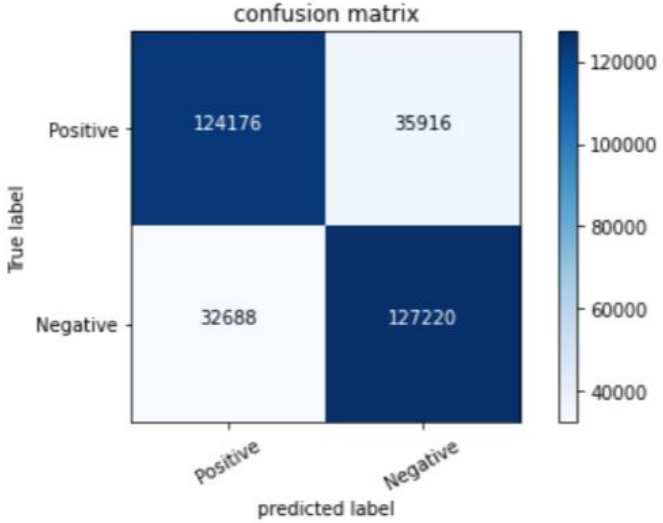
- Preprocessed / cleaned data
- Word Embeddings on text from tweets
- Neural Network model training with tensorflow and keras

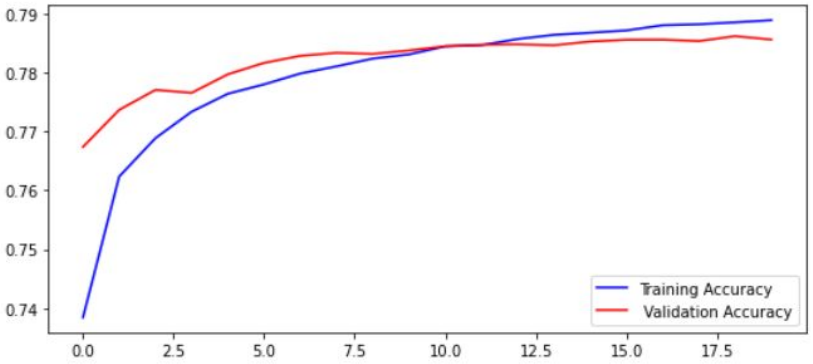
Layer (type)	Output Shape	Param #
input_15 (InputLayer)	[(None, 30)]	0
embedding (Embedding)	(None, 30, 300)	87217200
spatial_dropout1d_14 (Spatia	(None, 30, 300)	0
conv1d_14 (Conv1D)	(None, 26, 32)	48032
batch_normalization_4 (Batch	(None, 26, 32)	128
bidirectional_14 (Bidirectio	(None, 128)	49664
dense_42 (Dense)	(None, 256)	33024
dropout_14 (Dropout)	(None, 256)	0
dense_43 (Dense)	(None, 512)	131584
flatten_6 (Flatten)	(None, 512)	0
dense_44 (Dense)	(None, 1)	513

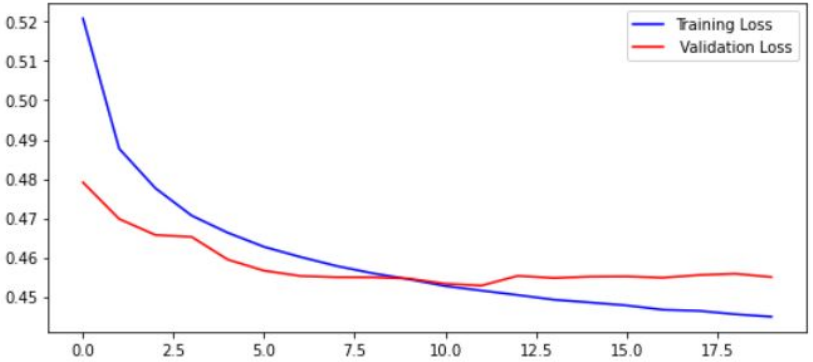
Total params: 87,480,145
Trainable params: 262,881
Non-trainable params: 87,217,264

Results

- Accuracy of our model
- Testing with tweets outside our dataset







Output Example

Our model tested on real tweets with search terms being “work”, “job”, “commute”

```
--------Tweets:


@matt_kercher when it rains it pours. Choose wisely!
1) Benefits
2) Work environment
3) Commute
4) Job satisfaction… https://t.co/CPvXDLoB0B
{'label': 'Neutral', 'score': 0.5478293299674988, 'elapsed_time': 0.047994136810302734}


@FreddyInSpace My new job is hybrid and my manager already thinks 3-day office weeks are too much.

Who is liking… https://t.co/FPDBZf4ye7
{'label': 'Neutral', 'score': 0.5062975883483887, 'elapsed_time': 0.055570125579833984}


Imagine hiring a writer and expecting them to work in office. Writing is the easiest job in the world and should ON… https://t.co/RZZWk499sE
{'label': 'Positive', 'score': 0.717910647392273, 'elapsed_time': 0.04293203353881836}


and if they are complaining about the transport, complain to the government (not necessary to the president right a… https://t.co/OBaU22NA78
{'label': 'Positive', 'score': 0.6574304699897766, 'elapsed_time': 0.04291200637817383}


i used to work in october so i had to commute every day men tagamo3 to there, work for 12-15 hrs a day, go back hom… https://t.co/B7jwSoPLEF
{'label': 'Negative', 'score': 0.33315762877464294, 'elapsed_time': 0.04356861114501953}


--------------------------------------------------------------------------------
There are 2 positive tweets, 1 negative tweets, and 2 neutral tweets
The average score of all the retrieved tweets is 0.552525132894516
The final grade is therefore neutral
```

Conclusion

- Classifying tweets using sentiment analysis
- Limitations
  - Training data is from 2009
  - Training data is only classified as positive or negative

Thank you for your attention!