Twitter Sentiment Analysis

Volha Puzikava
June, 2023

Disclaimer

The described analyses fulfill educational purposes only. The hypothetical business case and the results of sentiment analysis should not be perceived as real customers' attitudes and served as a push for remedial actions, as they have not been approved by any professional media organization.

Overview

Sentiment analysis/ opinion mining

approach to identify the emotional tone behind a body of text and categorize pieces of writing as positive, negative or neutral.

Overview Cont'd

Sentiment analysis:

- ➔ understand how customers feel about brand
- ➔ provide insights to improve products and services
- ➔ make business more responsive to customer feedback
- ➔ react quickly to negative sentiment and turn it around
- ➔ monitor brand's reputation in real-time
- ➔ keep customers happy by always putting their feelings first

Overview Cont'd

This Project:

analyzes Twitter sentiments about Apple and Google products to better understand how people feel about them

Outline

1) Business Problem
2) Data Understanding
3) Part I: Supervised ML Algorithms:
    a) *Data Preparation and Exploration*
    b) *Data Modeling*
    c) *Model Evaluation*
4) Part II: Neural Networks:
    a) *Data Preparation and Exploration*
    b) *Data Modeling*
    c) *Model Evaluation*
5) Conclusions

Business Problem

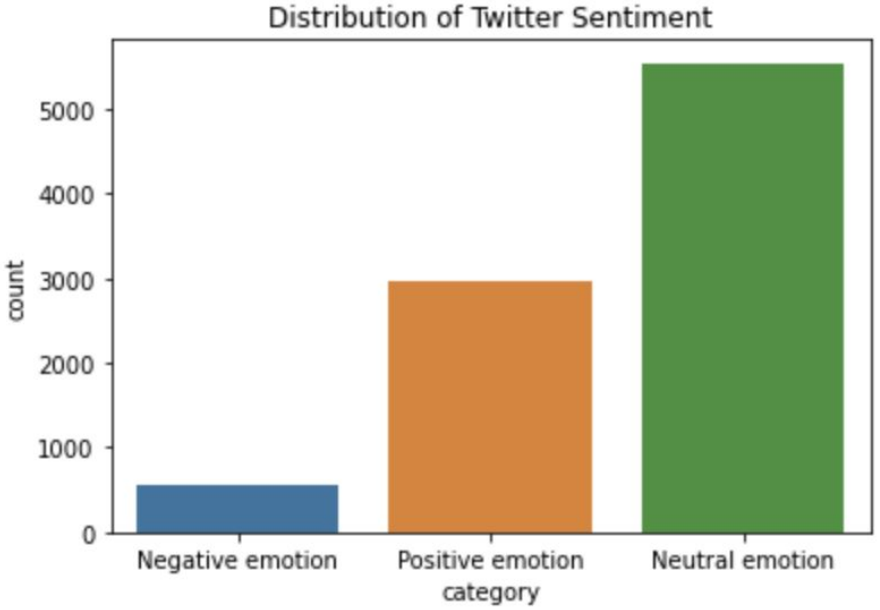
Tweeter Home Entertainment Group asked

- to analyze Twitter sentiment about Apple and Google products
  - ➔ to monitor brands
  - ➔ to understand customers needs

The main purpose

- to *build model* that could
  - ➔ rate the sentiment of a Tweet based on its content
  - ➔ give insights how people feel about products

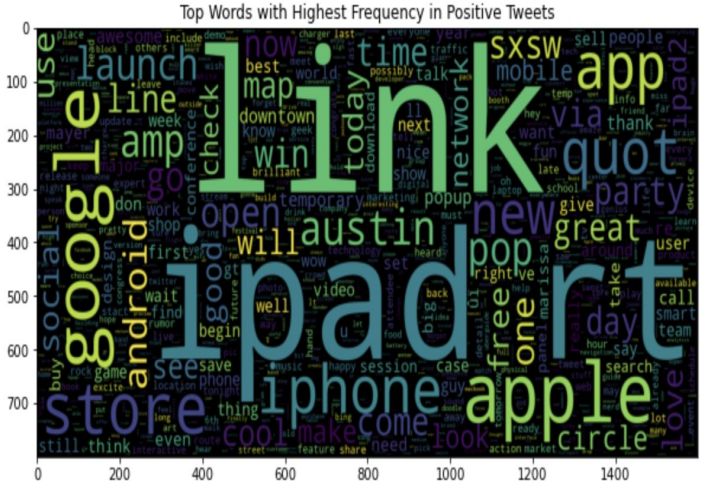
Data Understanding



- over 9,000 Tweets taken from CrowdFlower via data.world links
- *imbalanced multiclass classification problem*
- all classes equally important
- *evaluation metric*: model's ability to both capture Tweets and be accurate with those Tweets (F1 score)

Part I: Supervised ML Algorithms

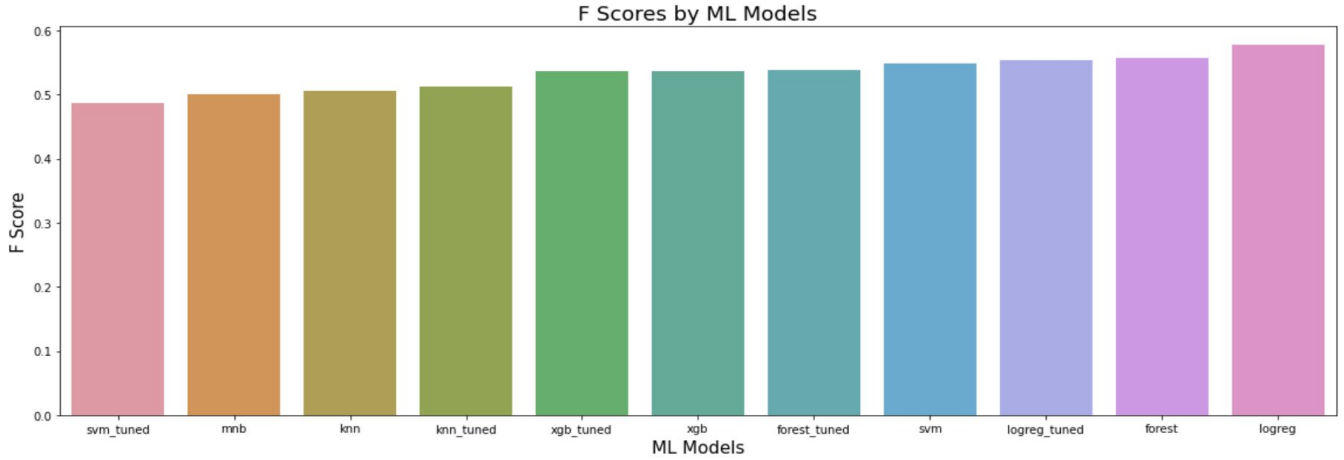
1. Text made *lowercase*, hashtags and @mentions removed, set of *tokens* generated
2. Distribution of *top 10 tokens* plotted for each category
3. Words transformed to *vectors*
4. Negative and positive categories *oversampled*



Supervised ML Algorithms Cont'd

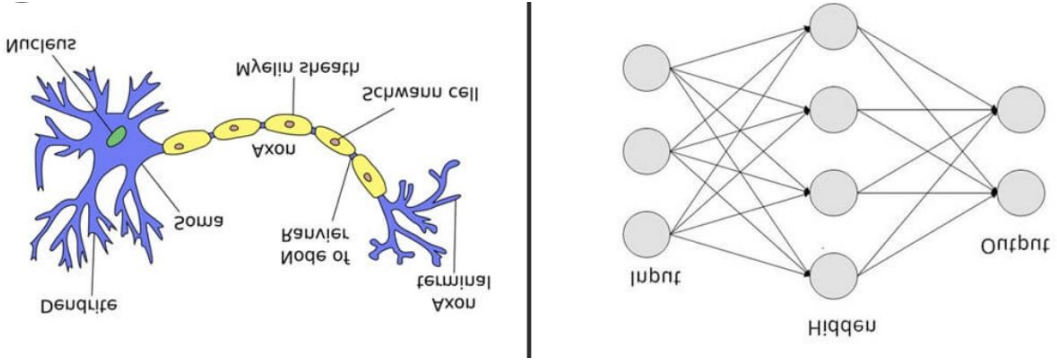
5. Different *ML algorithms* built

6. ***Best model***: F1 score = 58%



Part II: Neural Networks

1. Textual data transformed into *numerical representation*
2. Text reformatted into *matrix of vectors*
3. Descriptive categories converted into *integers product*
4. Data *rebalanced*



Neural Networks Cont'd

5. Various *neural networks models* tried out

6. ***Best model:*** F1 score = 65%

top-performing model

Conclusions

- ➔ **Best Model**: Neural Networks with F1 score of 65%
- ➔ Not the perfect result
- ➔ *Reason*: limited size of the dataset
- ➔ *Important*: quality and quantity of data

Thank you!

Email: helga.mikel@gmail.com
GitHub: @VolhaP87
LinkedIn: https://www.linkedin.com/in/volha-puzikava