

## EMOTION DETECTION IN TWITTER POSTS

By Melissa Rodriguez

## SOCIAL MEDIA IN OUR DAILY LIFE

<https://www.bls.gov/news.release/atus.t12.htm>  
<https://www.broadbandssearch.net/blog/average-daily-time-on-social-media>



26 years 5 months



8 years 4 months



6 years 8 months



3 years 7 months

For the average time spent in a lifetime, social media is the third activity where time is most used.

In average, people spend 2 hours 23 minutes a day in Social media.

## **EMOTION DETECTION ON SOCIAL MEDIA**

- **Mental Health**
- **Public Safety**
- **Customer Service**
- **Business Strategies**



Emotion detection on social media can be used to:

- Identify mental health issues
- Identify threats to public safety
- Improve customer service
- Define business strategies



## EMOTION VS. SENTIMENT

**Sentiment** is a general thought, feeling, or sense.

**Emotion** is a person's internal state of being and involuntary physiological response to an object or a situation, based on or tied to physical state and sensory data.

There are differences between sentiment and emotion.

The diagram shows three circular icons representing sentiment polarity. The first icon, labeled 'NEGATIVE', is orange with a red border and features a black angry face emoji. The second icon, labeled 'NEUTRAL', is light blue with a blue border and features a black neutral face emoji with a single dash for an eye. The third icon, labeled 'POSITIVE', is yellow with a yellow border and features a black smiling face emoji with two dashes for eyes.

<https://www.newgenapps.com/blog/the-secret-way-of-measuring-customer-emotions-social-media-sentiment-analysis>

<https://owins.com/cdnarr/14435906>

## SENTIMENT ANALYSIS

- Informs user's reaction polarity
- Fails to describe the exact feelings of the customers and the intensity of their reaction

Sentiment Analysis describes polarity from texts:

- Positive
- Negative
- Neutral

## EMOTION ANALYSIS

- An advanced form of sentiment analysis
- Analyzes human emotions
- Tries to identify the proper emotion from context



<https://www.newgenapps.com/blog/the-secret-way-of-measuring-customer-emotions-social-media-sentiment-analysis>

<https://owips.com/clipart-14435986>

Emotion Analysis is a deeper level of sentiment analysis

## EKMAN'S EMOTION MODEL



Anger



Disgust



Fear



Sadness



Happiness



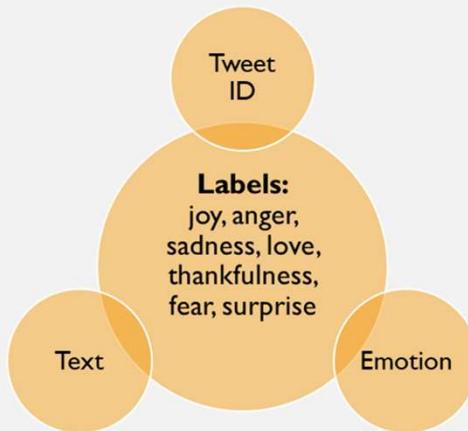
Surprise

- There are two different categories for emotion models: Discrete and Dimensional.
- For the purposes of this work we will work with a discrete model: Ekman's Emotion Model which states that there are six basic emotions:
  - Anger
  - Disgust
  - Fear
  - Sadness
  - Happiness
  - Surprise
- These are classified as discrete because they can be distinguished by a person's facial expression and biological processes.

## TWITTER EMOTION DATASET

- [Twitter\\_Emotion\\_SocialCom\\_Wang](#): contains tweet id and label.
- Text was extracted from Twitter
- Attributes: Id, Text, Label
- 1,339,794 observations

\* [http://knoesis.org/sites/default/files/wenbo\\_socialcom\\_2012\\_0.pdf](http://knoesis.org/sites/default/files/wenbo_socialcom_2012_0.pdf)



- Dataset was downloaded from link provided
- The classification process was done by assigning to each emotion, a set of words. These words were used as a hashtag criteria to collect and classify the tweets
- The tweets were reviewed to confirm classification was correct.
- Paper related to the referenced research:  
[http://knoesis.org/sites/default/files/wenbo\\_socialcom\\_2012\\_0.pdf](http://knoesis.org/sites/default/files/wenbo_socialcom_2012_0.pdf)
- Link to download dataset(only id and label): <http://knoesis.org/projects/emotion>

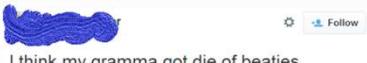
## METHODOLOGY

- 
- ```
graph TD; A[Twitter Emotion Data] --> B[Emotion Data Subset]; B --> C[Classification Model]; C --> D[Train deep learning neural network model]; D --> E[Test model]; E --> F[Evaluate and adjust model]; F --> G[Under sampling]; G --> H[Data Preprocessing]; H --> I[Exploratory Data Analysis (EDA)]; I --> J[Merge and load to a database]; J --> K[Extract Twitter Posts Texts]; K --> L[Download Tweet_Id with Label]
```
- The diagram illustrates a methodology flow. It starts with 'Twitter Emotion Data' (orange box), which leads to 'Emotion Data Subset' (orange box). From 'Emotion Data Subset', it moves to 'Classification Model' (orange box). The 'Classification Model' then leads to a series of steps: 'Train deep learning neural network model', 'Test model', and 'Evaluate and adjust model'. A feedback loop returns from 'Evaluate and adjust model' back to 'Emotion Data Subset'. Additionally, there is a feedback loop from 'Classification Model' back to 'Twitter Emotion Data'.
- Download Tweet\_Id with Label
  - Extract Twitter Posts Texts
  - Merge and load to a database
- Exploratory Data Analysis (EDA)
  - Data Preprocessing
  - Under sampling
- Train deep learning neural network model
  - Test model
  - Evaluate and adjust model

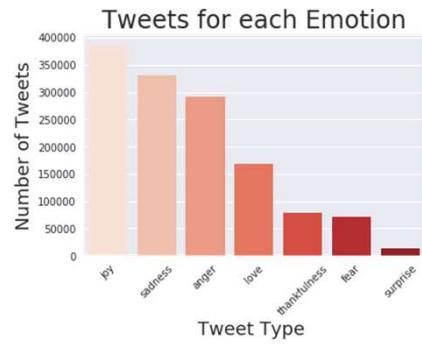
Twitter  
Emotion Data

Emotion Data  
Subset

Classification  
Model

## CHALLENGES

<https://www.thepoke.co.uk/2014/05/21/the-25-worst-best-spelling-mistakes-on-twitter/>

### Challenges for Emotion Detection using Natural Processing Languages:

- Lack of enough labeled data for model training
- Not enough labeled data for each emotion
- Sarcasm
- Misspelling
- Slang
- Emojis
- Text length

#### **MODEL PERFORMANCE:**

- Baseline accuracy is 28% (lowest score expected for this dataset.)
- Model Accuracy = 63%
- Cohen Kappa Score = 0.57 (moderate agreement)
- 17% of the test data was not classified

#### **REMARKS**

- Model Accuracy is 35% higher than the baseline.
- Cohen Kappa Score value falls under the “Moderate Agreement” level.
- The model results are good with room for improvement.

<https://towardsdatascience.com/interpretation-of-kappa-values-2acd1ca7b18f>

| Original Text                                                                                                                        | Actual  | Predicted      |
|--------------------------------------------------------------------------------------------------------------------------------------|---------|----------------|
| I'm about to delete and unfollow people who refuse to use proper grammar and proper spelling #notrocketscience #annoyed #canhandleit | Anger   | <b>Anger</b>   |
| so stressed, i can barely breathe #nervous #terrified #abouttocracy                                                                  | Fear    | <b>Fear</b>    |
| I'm ready to give up on life. I'm tired of all the disappointment that surrounds me. #depressed                                      | Sadness | <b>Sadness</b> |
| i love my life. i love my family #happy                                                                                              | Joy     | <b>Joy</b>     |

| Original Text                                                                                                                     | Actual  | Predicted      |
|-----------------------------------------------------------------------------------------------------------------------------------|---------|----------------|
| Still no access to my articles. Want to burn the world down. #Rage                                                                | Anger   | <b>Sadness</b> |
| Just think about all the times we have shared together... #depressed                                                              | Sadness | <b>Anger</b>   |
| It felt like Christmas day cause the family are over for tea and I am watching Christmas movies #excited                          | Joy     | <b>Sadness</b> |
| Work related tensions and stress and continuous thinking about it gave me further stress and tensions and finally severe #anxiety | Fear    | <b>Joy</b>     |

## CLASSIFICATION EXAMPLES

## RECOMMENDATIONS



Bigger dataset



Quality of annotations



Oversampling

- Given more time and resources, model performance can be improved by having bigger datasets with quality assurance applied to annotations. This represents a big challenge since it is dependent on exhaustive manual work.
- Would like to try oversampling for those classes that were not able to use due to unbalanced data.
- Gather more data not only from Twitter but from other social media portals.
- Explore the impact on model when using emojis as a separate feature.

# *Thank You!*

// FLATIRON SCHOOL