# Emotion Detection from Tweets

Didem Demirağ

Anisa Halimi

Nora von Thenen

# Problem Description

- Twitter users post valuable information every day which contain their opinions about certain topics

- It is of great importance to analyze this information on a big scale

- Our aim is to label certain emotions to a user's post

# Problem Description

| Tweet | Emotion |
|---|---|
| Enjoy the little things in life! | Happy |
| In the end you have to be your own hero because everyone else is too busy trying to save themselves. | Sad |
| Tomorrow will be a better day! | Hopeful |
| Bought 5 things from whole foods and it cost $230… | Complainable |
| Can't wait to work again! | Complainable |

Table 1: Examples of emotions in tweets

# Motivation

- Market researchers and companies

- Political campaigns

- People's reactions in a crisis

- Keeping track of the emotional state of a patient with a certain disease

- Define certain psychological disorders

- Sociologist can infer the life quality of a population.

# Selected Emotions Based on Ekman's Model
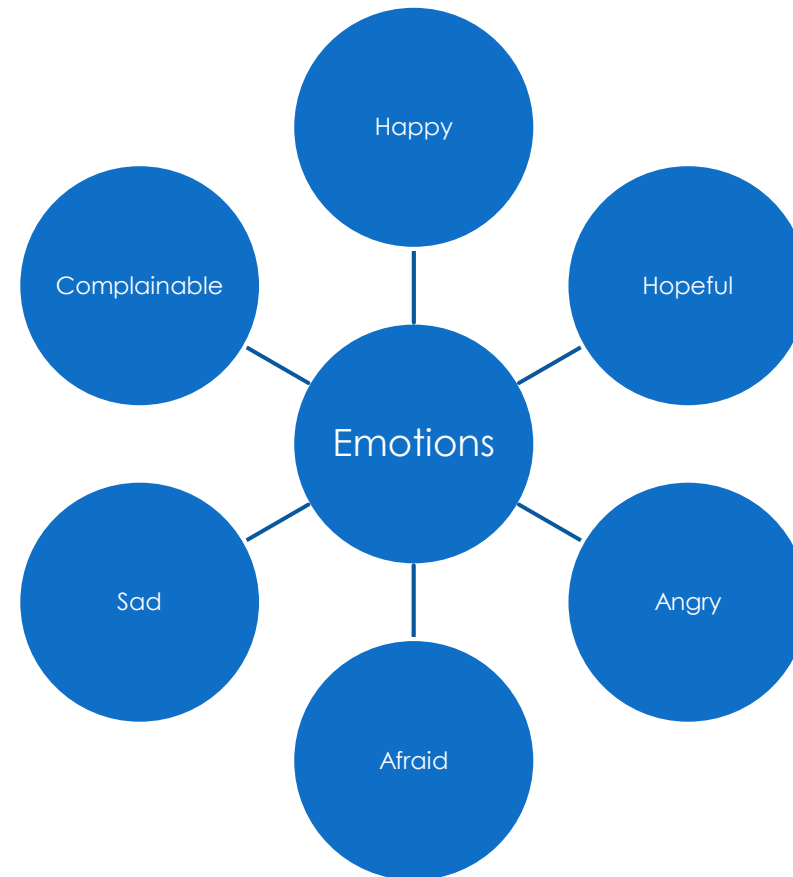
# Methodology

**Classify the tweets into three groups**
- Use 'Natural Language Toolkit' of Python

**Process tweets**
- Change hashtags into normal words
- Apply stemming and filter out stop words

**Use keyword lists to assign a more specific emotion**
- Find the keyword list with the most matches within the tweet

# Methodology

- After labeling:

  - Use Variance Threshold as feature reduction algorithm with different thresholds

  - Use different machine learning algorithms to train and test our dataset
    - Decision Tree
    - KNN (K-Nearest Neighbor)

  - Compute the recall and precision rates to measure the accuracy of the algorithms.

  - Compare these algorithms to decide which works most efficiently.

# Data

| Datasets | Number of tweets |
|---|---|
| positive labeled tweets | 2.949 |
| negative labeled tweets | 3.293 |
| neutral labeled tweets | 6.353 |
| total | 12.595 |

Table 2: Dataset size

# Results - pre-labeled Tweets

| Tweet | Label |
|---|---|
| Enjoy the little things in life! | positive |
| In the end you have to be your own hero because everyone else is too busy trying to save themselves | negative |
| Tomorrow will be a better day | positive |
| Bought 5 things from whole foods and it cost $230… | negative |
| Can't wait to work again! | negative |

Table 3: Some examples of pre-labeled tweets
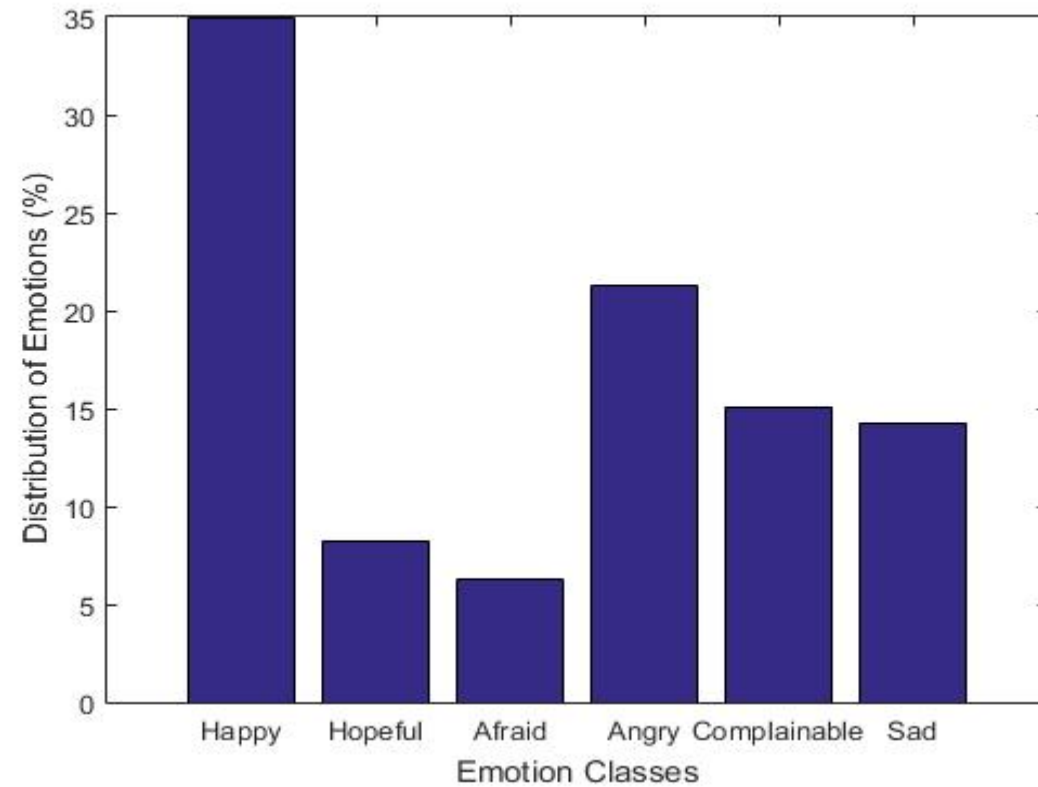
# Distribution of Emotions



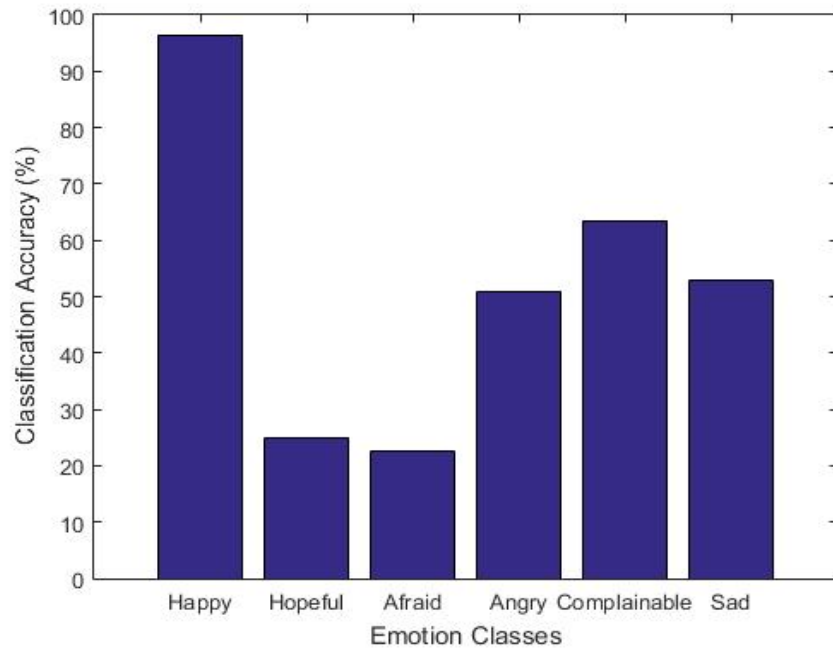Figure 1: Distribution of the emotions in the collected data

# KNN



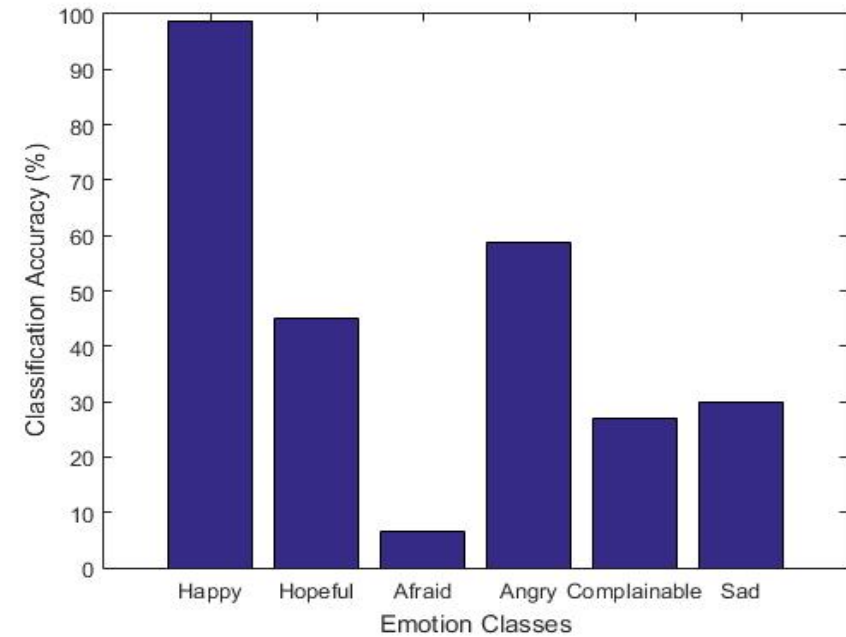Figure 2: The accuracy of KNN using unigrams with threshold 0.01



Figure 3: The accuracy of KNN using unigrams with threshold 0.008
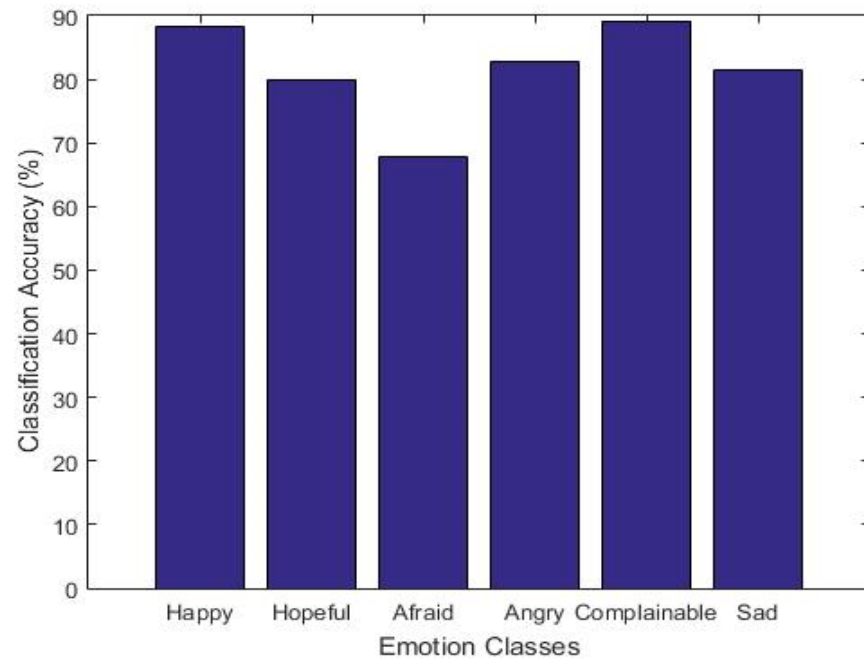
# Decision Tree



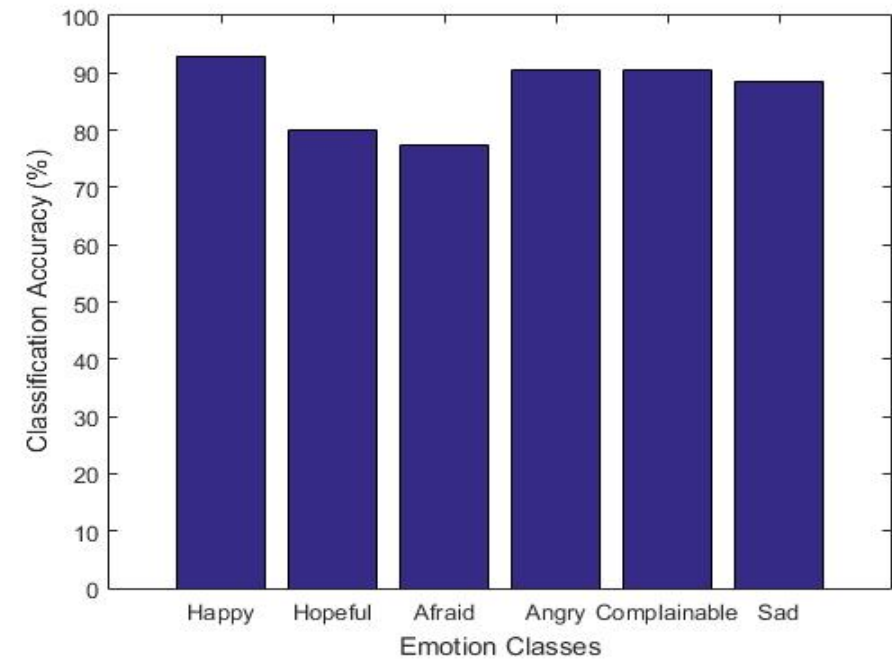Figure 4: The accuracy of Decision Tree using unigrams with threshold 0.01



Figure 5: The accuracy of Decision Tree using unigrams with threshold 0.008

# Recall and Precision Values

| | **KNN** | **Decision Tree** | **KNN** | **Decision Tree** |
|---|---|---|---|---|
| Threshold | 0.01 | | 0.008 | |
| Precision | 80.51 | 88.09 | 76.03 | 89.98 |
| Recall | 51.9 | 81.56 | 44.33 | 86.65 |
| F-measure | 63.11 | 84.7 | 56 | 88.28 |

Table 4: Precision, recall and F-measure of KNN and Decision Tree

# Conclusions

- Tweets are grouped into six emotions automatically

- Variance Threshold was found to be the most suitable for feature reduction

- Decision tree provides better accuracy than KNN

# References

[1] M. Hasan, E. Rundensteiner, and E. Agu. Emotex: Detecting emotions in twitter messages. In *ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conference*, pages 27–31, May 2014.

[2] P. Ekman. Basic emotions. Handbook of Cognition and Emotion, 98:45-60, 1999.