

Analyzing Twitter's Trending Topics

Virgilia Antonucci

Music · Trending



BETTER THAN REVENGE

30.5K posts

Trending in United States



Courtney

Trending with [Justine](#), [Justine](#)

Entertainment · Trending



Fassbender

2,224 posts

Politics · Trending



Jill Stein

3,591 posts

Trending



The Social Network

4,656 posts

Data

Emotion Classification NLP

Tweets classified based on 4 emotions - joy, sadness, anger and fear.

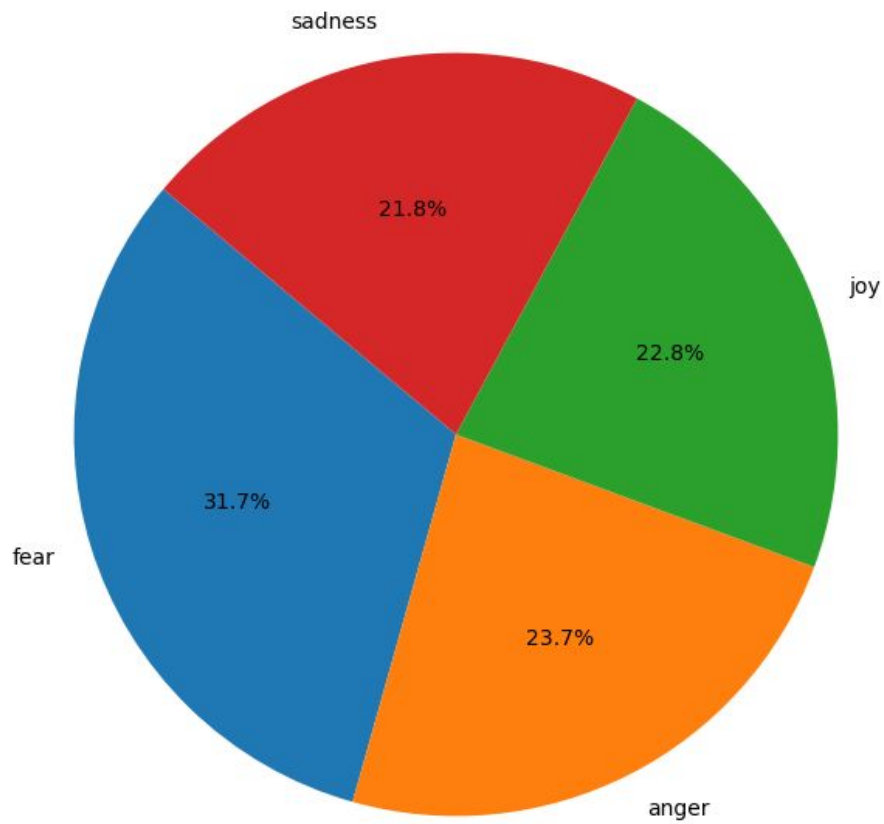


Steps

- Lowercasing
- Removing Mentions
- Tokenizing
- Stemming
- Vectorizing

We end up with two sets of data. One is tweets with hashtags and one is tweets without.

Distribution of Sentiments in Tweets

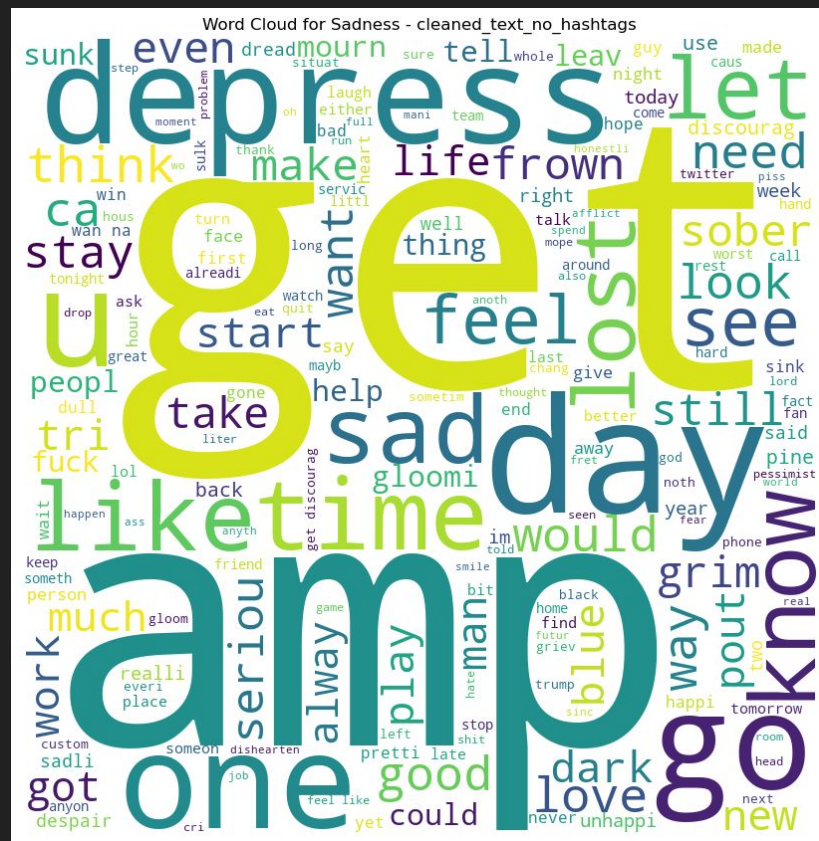
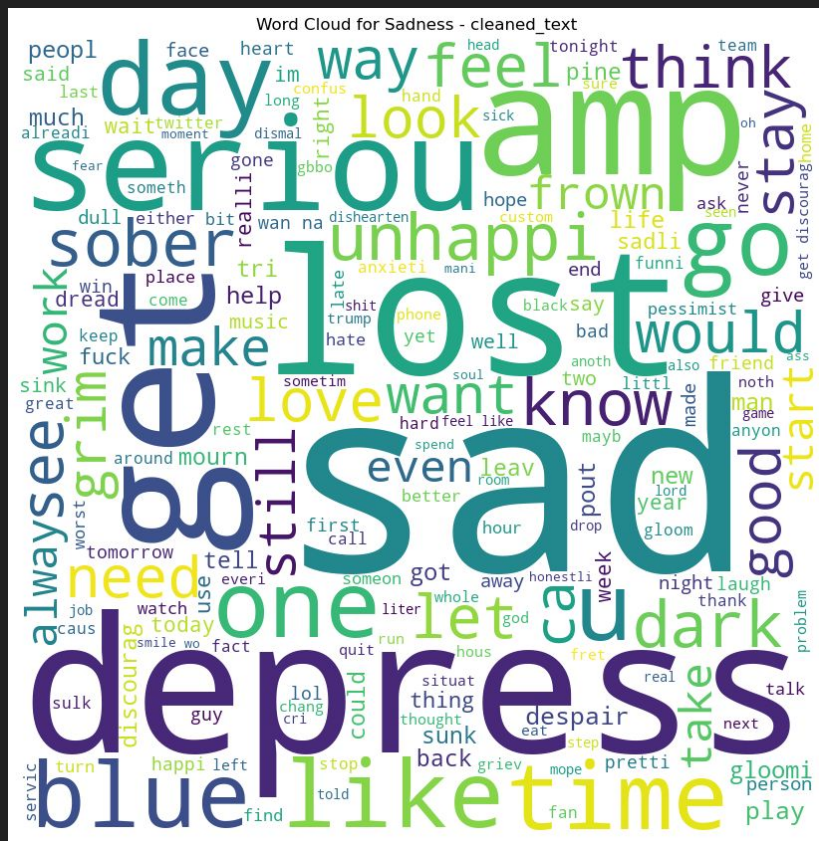


Word Cloud for Anger - cleaned_text




Word Cloud for Anger - cleaned_text_no_hashtags





[illegible]

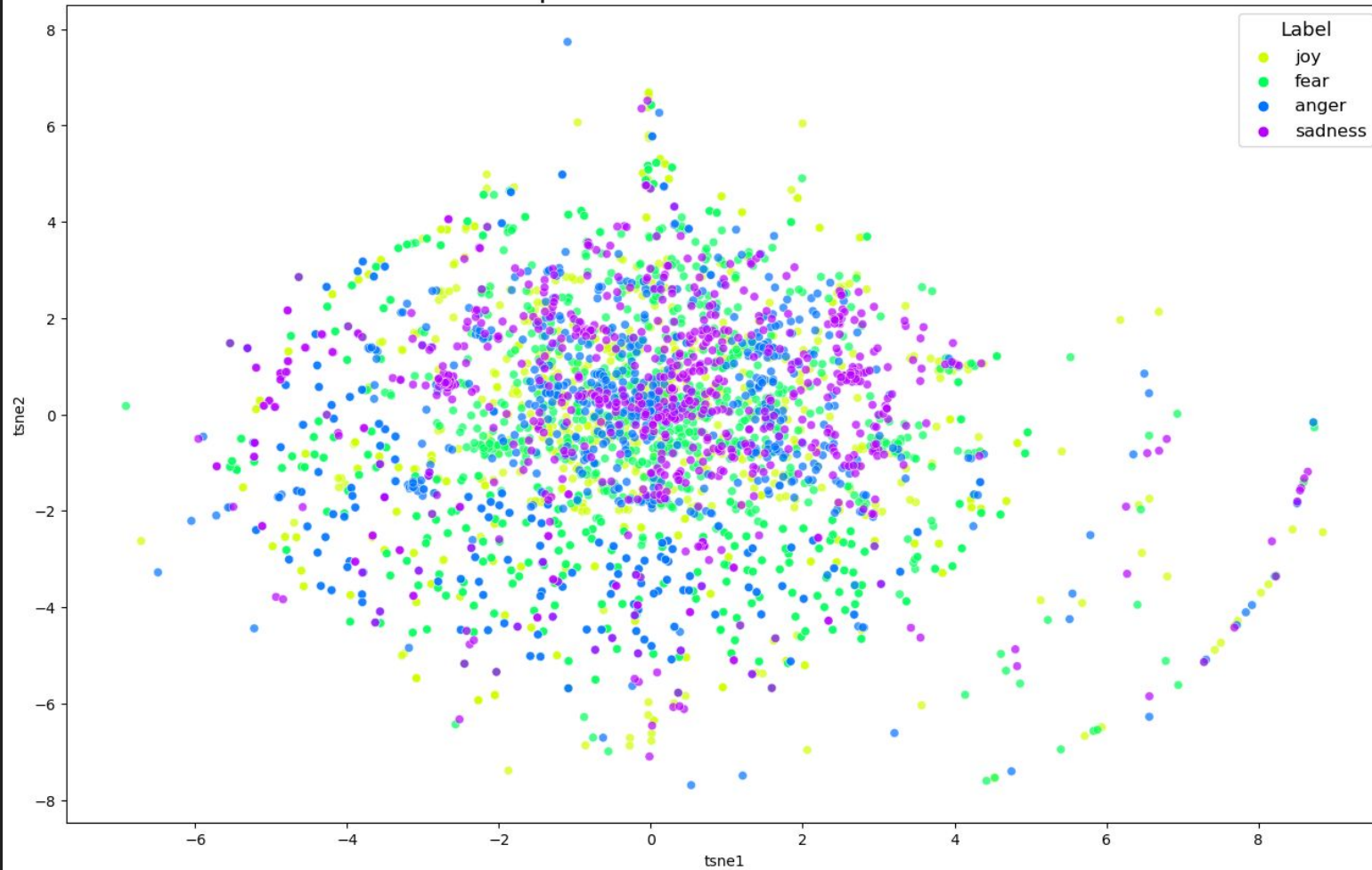
Word Cloud for Joy - cleaned_text



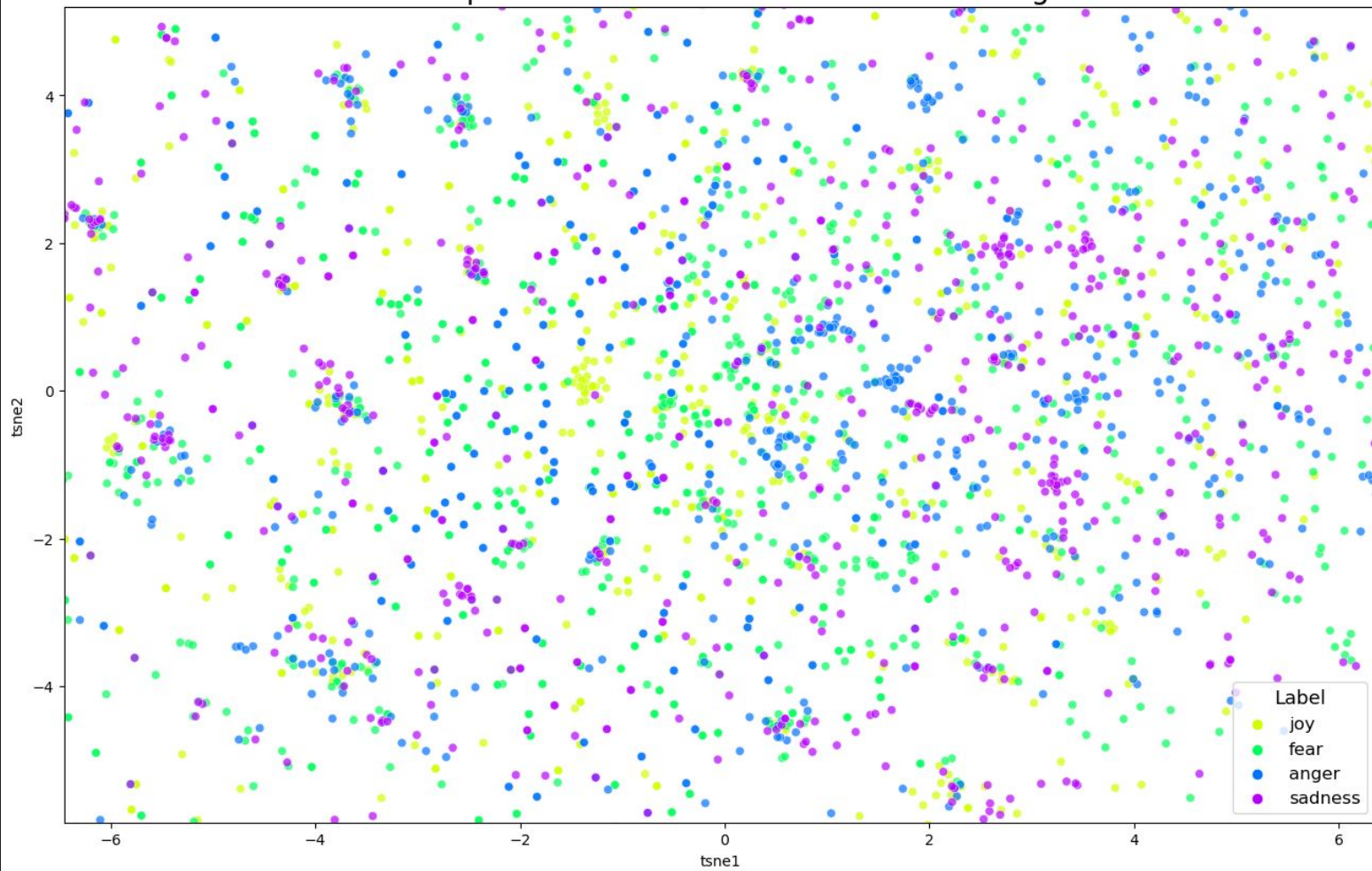
Word Cloud for Joy - cleaned_text

The word cloud features a variety of words related to joy and happiness. The most prominent words are 'smile', 'cheer', 'happier', 'watch', 'love', 'day', 'rejoice', 'laugh', 'like', 'amazing', 'broadcast', 'live', 'music', 'optimism', 'good', 'play', 'today', 'need', 'got', 'feel', 'thing', 'get', 'u', 'want', 'could', 'remini', 'quot', 'night', 'pleas', 'litt', 'show', 'morn', 'home', 'realli', 'got', 'hilar', 'god', 'world', 'actual', 'heyday', 'everyon', 'never', 'someth', 'go', 'always', 'laught', 'joy', 'great', 'work', 'gbbo', 'may', 'hearti', 'ask', 'breezi', 'broadcast', 'music', 'beauti', 'bright', 'call', 'still', 'know', 'life', 'right', 'year', 'mayb', 'see', 'peopl', 'chirp', 'cheeri', 'glad', 'look', 'rejoic', 'thought', 'lot', 'best', 'done', 'one', 'time', 'thank', 'hope', 'fun', 'watch', 'music', 'talk', 'week', 'start', 'jovial', 'laugh', 'like', 'amazing', 'broadcast', 'hilar', 'sparkl', 'happy', 'birthd', 'ever', 'everi', 'come', 'tear', 'back', 'face', 'next', 'give', 'much', 'let', 'wish', 'even', 'follow', 'break', 'way', 'fuck', 'glee', 'anim', 'well', 'joyou', 'lol', 'new', 'chang', 'somem

t-SNE plot of Labels for Cleaned Texts



t-SNE plot of Labels for Texts without Hashtags



Models Used

- Multinomial Naive Bayes
- SVM
- XGBoost
- Random Forest
- LSTM (Long Short-Term Memory): recurrent neural network (RNN) architecture widely used in Deep Learning

Support Vector Machine with GridSearch

```
Classification Report:
              precision    recall  f1-score   support

     anger      0.86      0.85      0.86       154
       fear      0.85      0.91      0.88       223
        joy      0.92      0.92      0.92       185
     sadness      0.82      0.76      0.79       161

 accuracy      0.87      0.87      0.87       723
  macro avg      0.86      0.86      0.86       723
 weighted avg      0.87      0.87      0.87       723
```

```
-----
Fitting 5 folds for each of 36 candidates, totalling 180 fits
Best parameters for Texts without Hashtags:
{'svm_C': 10, 'svm_kernel': 'linear', 'tfidf_max_df': 0.5, 'tfidf_ngram_range': (1, 2)}
Results for Texts without Hashtags:
Accuracy: 0.7427385892116183
```

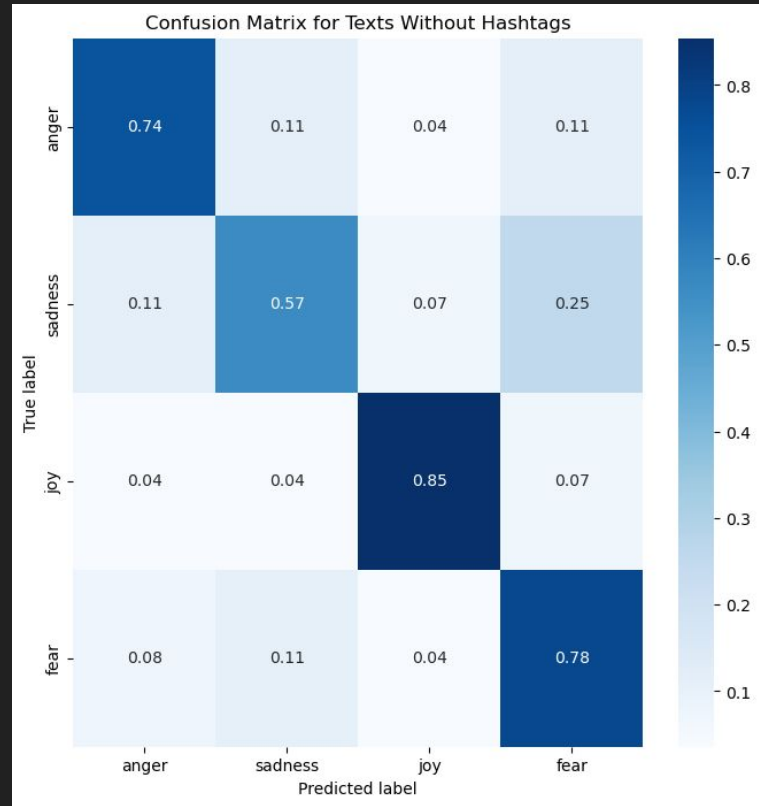
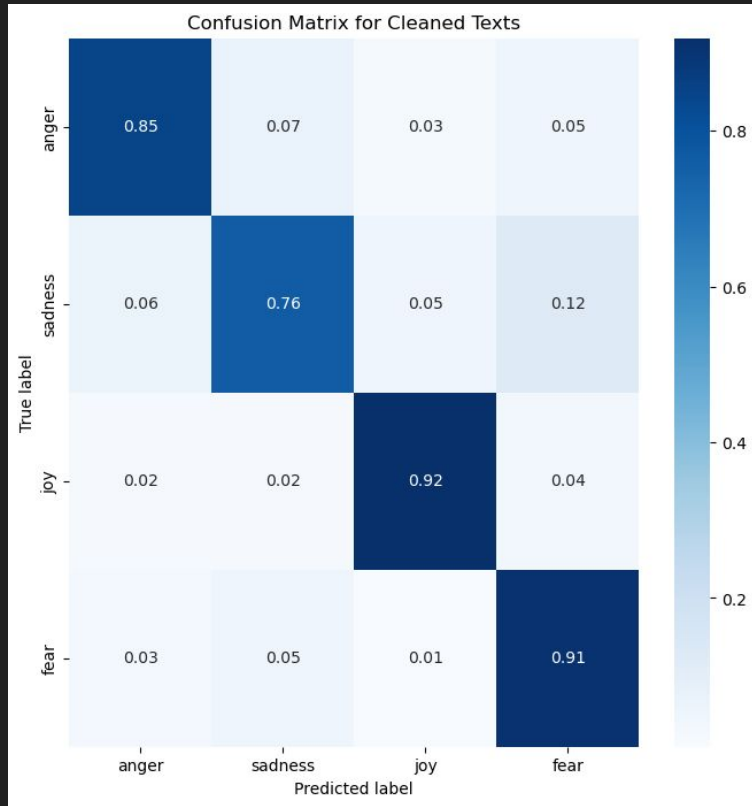
```
Classification Report:
              precision    recall  f1-score   support

     anger      0.73      0.74      0.74       154
       fear      0.71      0.78      0.74       223
        joy      0.86      0.85      0.86       185
     sadness      0.65      0.57      0.61       161

 accuracy      0.74      0.73      0.74       723
  macro avg      0.74      0.73      0.74       723
 weighted avg      0.74      0.74      0.74       723
```

With
Hashtags

Without
Hashtags



Conclusion:

Hashtags are a must!

Next Step

- Expand Sentiment analysis to more emotions to handle more nuance

Thank you.

[LinkedIn](#) [GitHub](#)