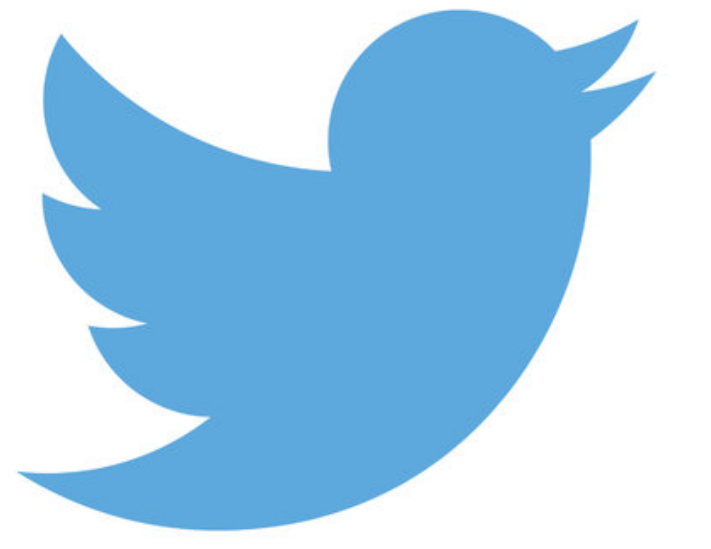


# Twitter Sentiment Analysis

Yingwen Ren, Wei Lai, Chuxuan Zhang, Ronghao Xu  
Department Computer science, SFU



## Objectives

According the study of Natural language processing, We provide sentiment analysis of twitter to achieve these objects:

- 1. Provide correct sentiment analysis for twitter to get their labels.
- 2. Use deep learning algorithm for an effective classification based on twitter condition and labels.
- 3. Improve the current twitter analysis model by increasing the accuracy and efficiency.

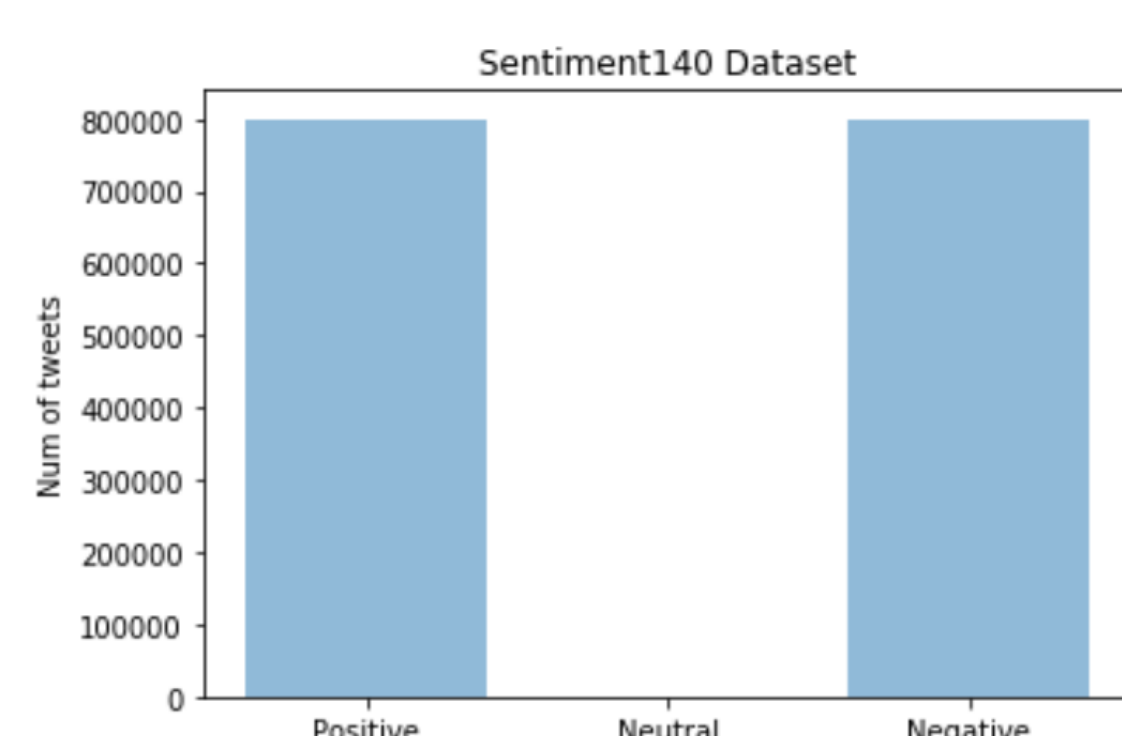
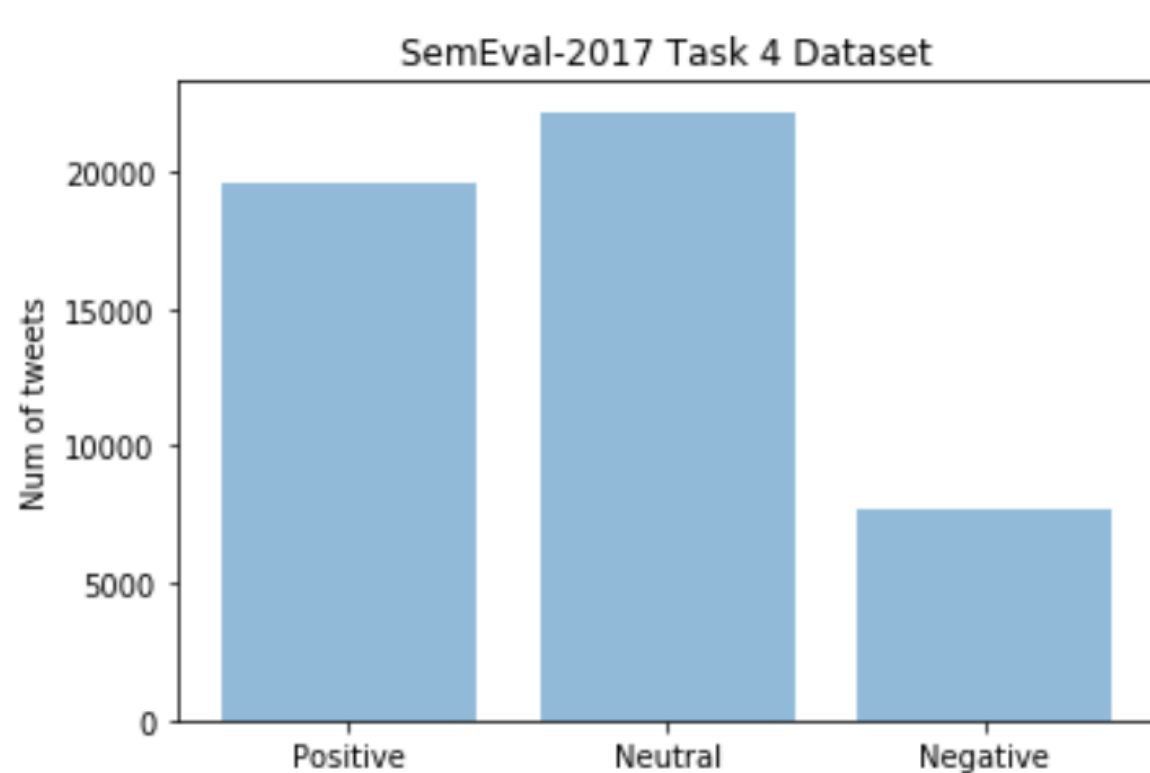
## Introduction

Twitter today has become a very popular communication tool among Internet users. Millions of messages are appearing daily, some of them could become valuable sources of people's opinions and sentiments. Such data can be efficiently used for marketing or social studies. We use different feature sets and machine learning classifiers to determine the best combination for sentiment analysis of twitter.

Sentiment analysis helps to collect efficient information by converting dissatisfied customers into promoters by analyzing this huge volume of opinions. In this way, Sentiment Analysis can be useful to understand how the mood and tendency of the public.

## Data Set

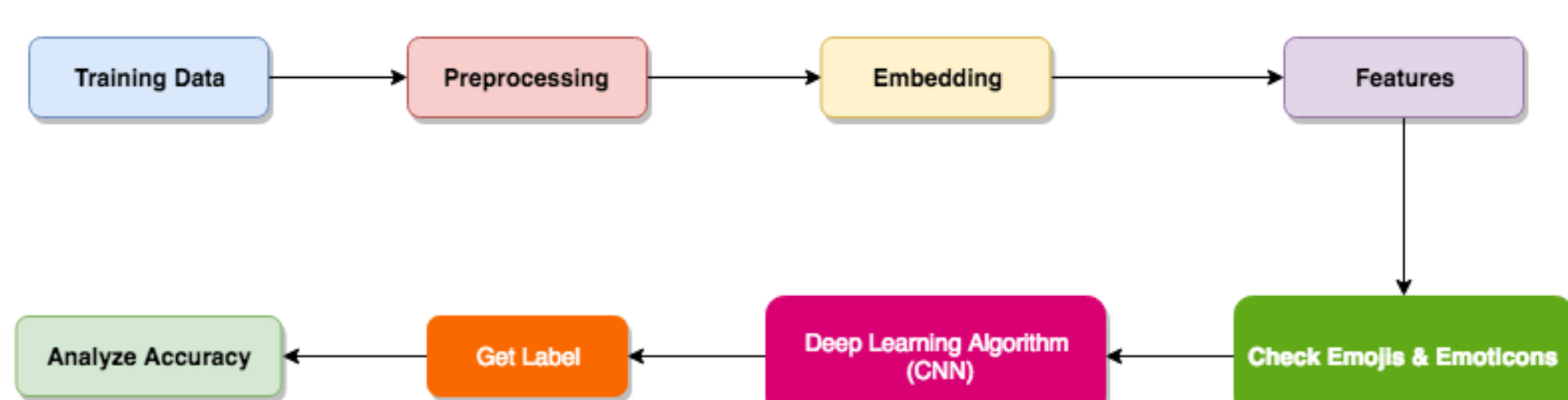
- SemEval-2017, International Workshop on Semantic Evaluation Sponsored by SIGLEX.
  - ▷ Task 4: Sentiment Analysis in Twitter
  - ▷ Subtask A. Message Polarity Classification: Given a message, classify whether the message is of positive, negative, or neutral sentiment.



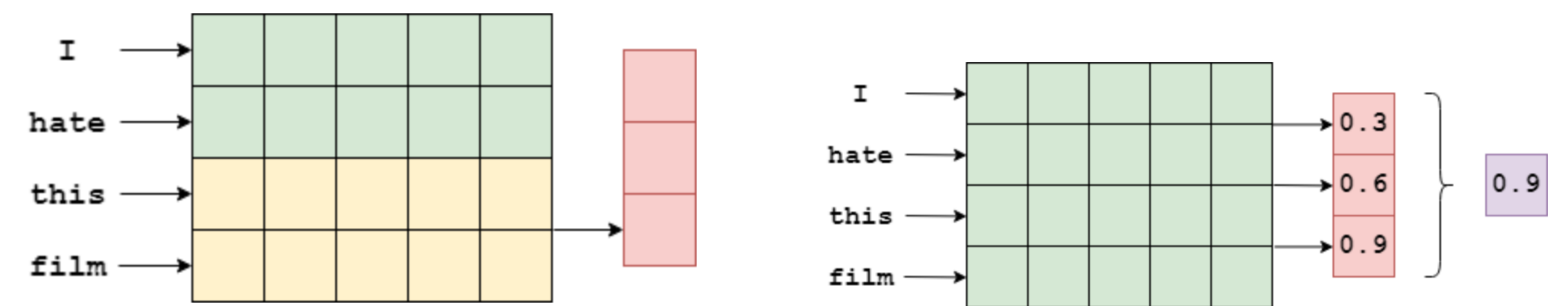
- Sentiment140 - A Twitter Sentiment Analysis Tool
  - ▷ CSV data with emoticons removed

## Proposed Method

- Baseline:
  - ▷ Feature extraction with spaCy
  - ▷ Logistic regression + n-gram
  - ▷ Naïve Bayes model + n-gram
  - ▷ SVC Linear model + n-gram
- Basic method:
  - ▷ Emoji Emoticon Detect:
    - ▶ Very effective for twitter sentiment analysis. (Wieslaw Wolny, 2016)
  - ▷ Deep Learning: CNN, Convolutional Neural Network
    - ▶ Glove: Global Vectors for Word Representation
    - ▶ Pre-trained word vector.
  - ▷ Deep Learning LSTM, Long short-term memory



## CNN model

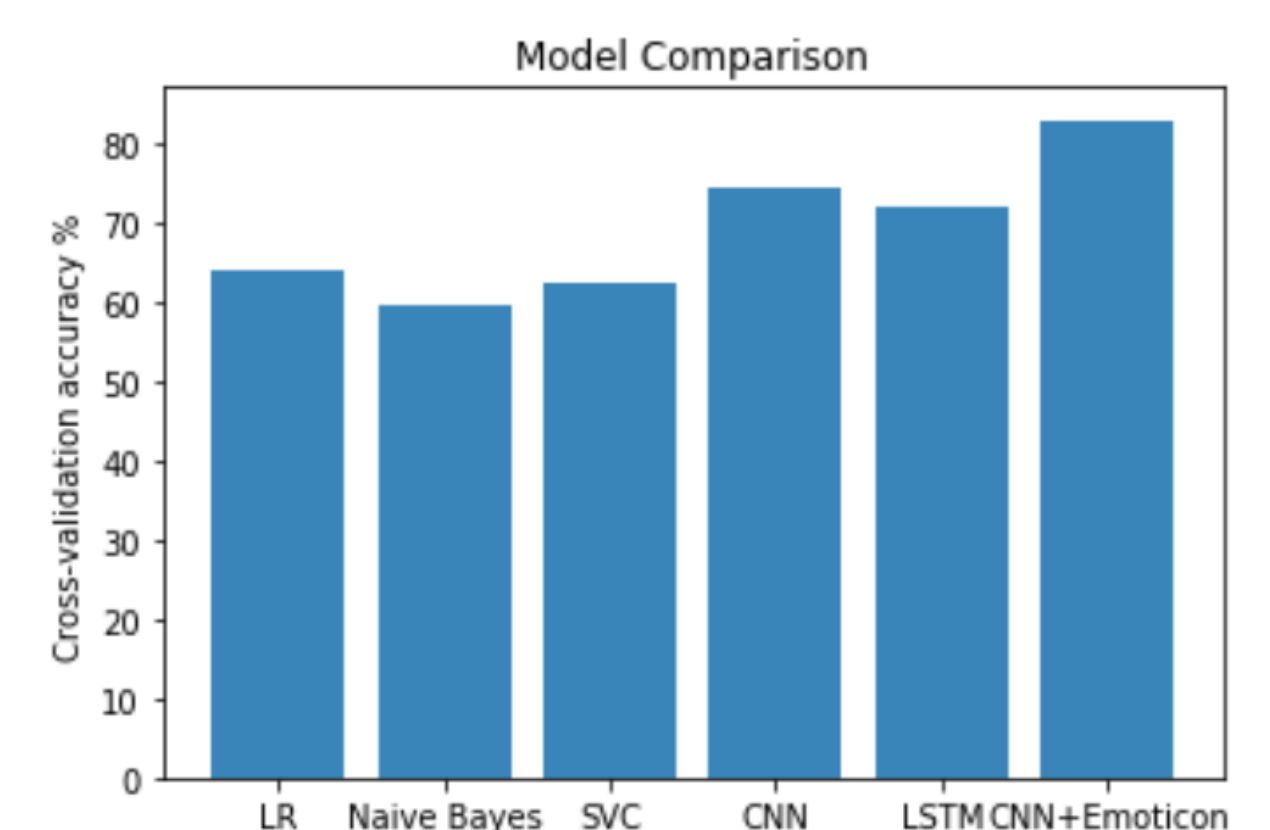
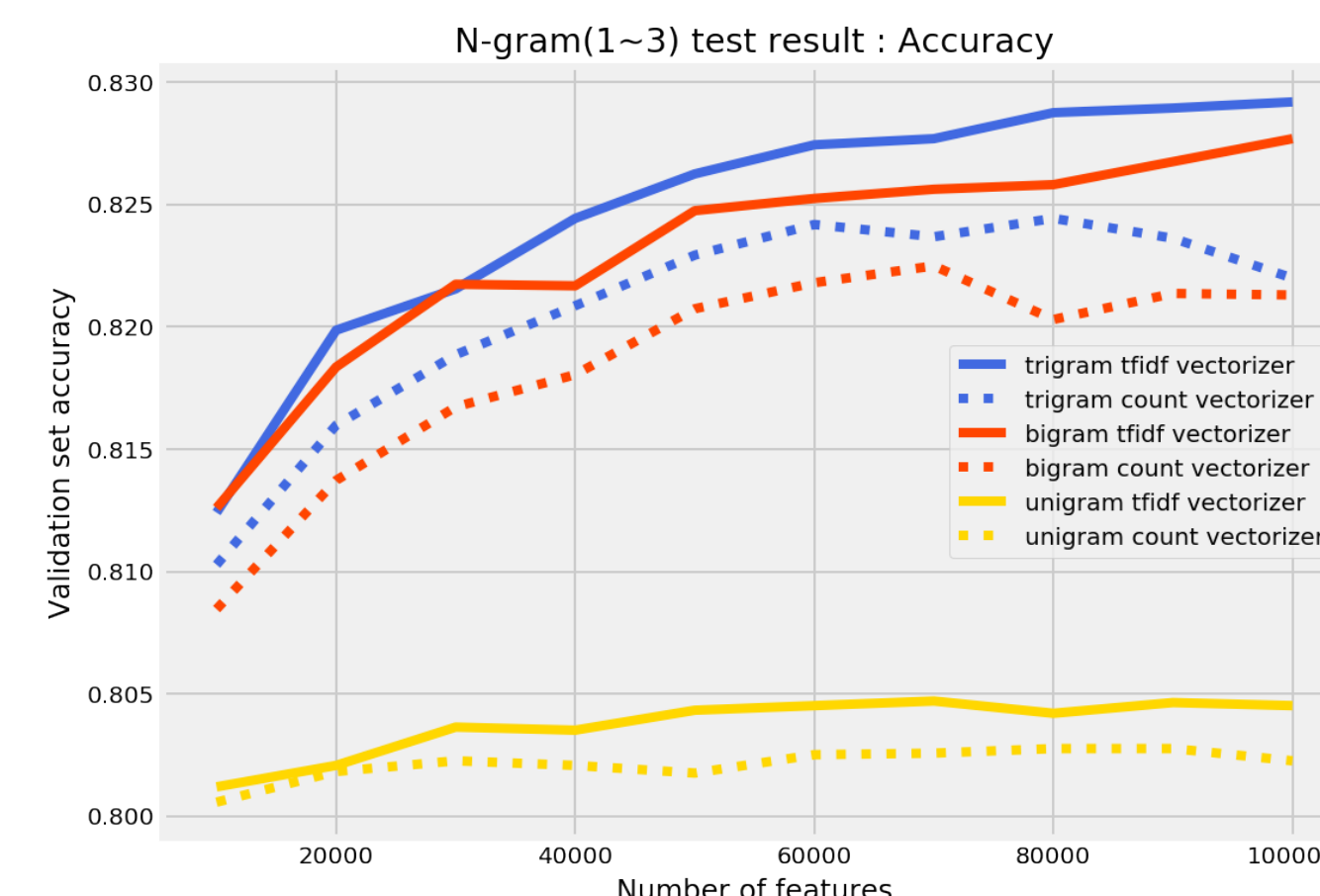


- use a filter that is  $n * \text{embdim}$
- moves "down" the image to cover the next bi-gram and another output
- use pooling on the output of the convolutional layers
- the maximum value is the "most important" feature

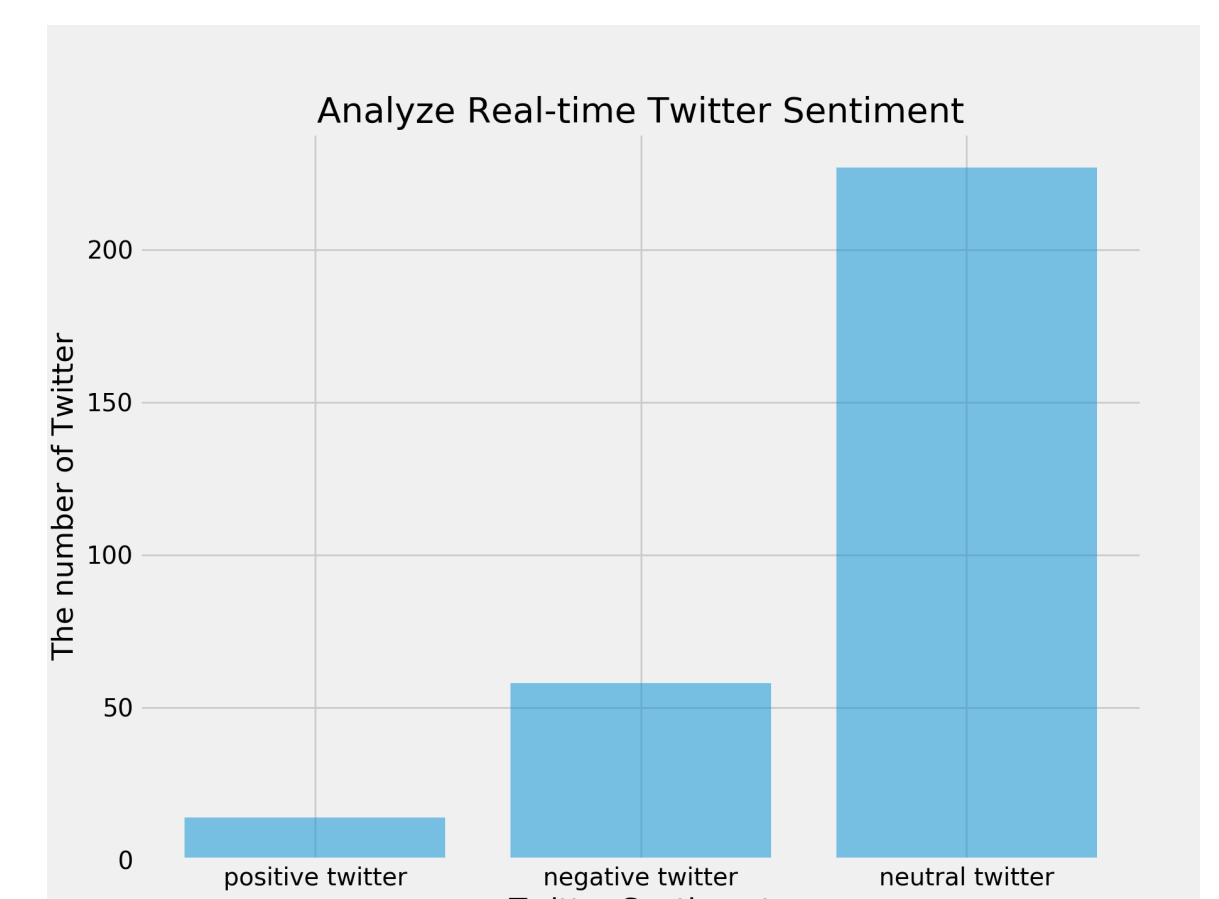
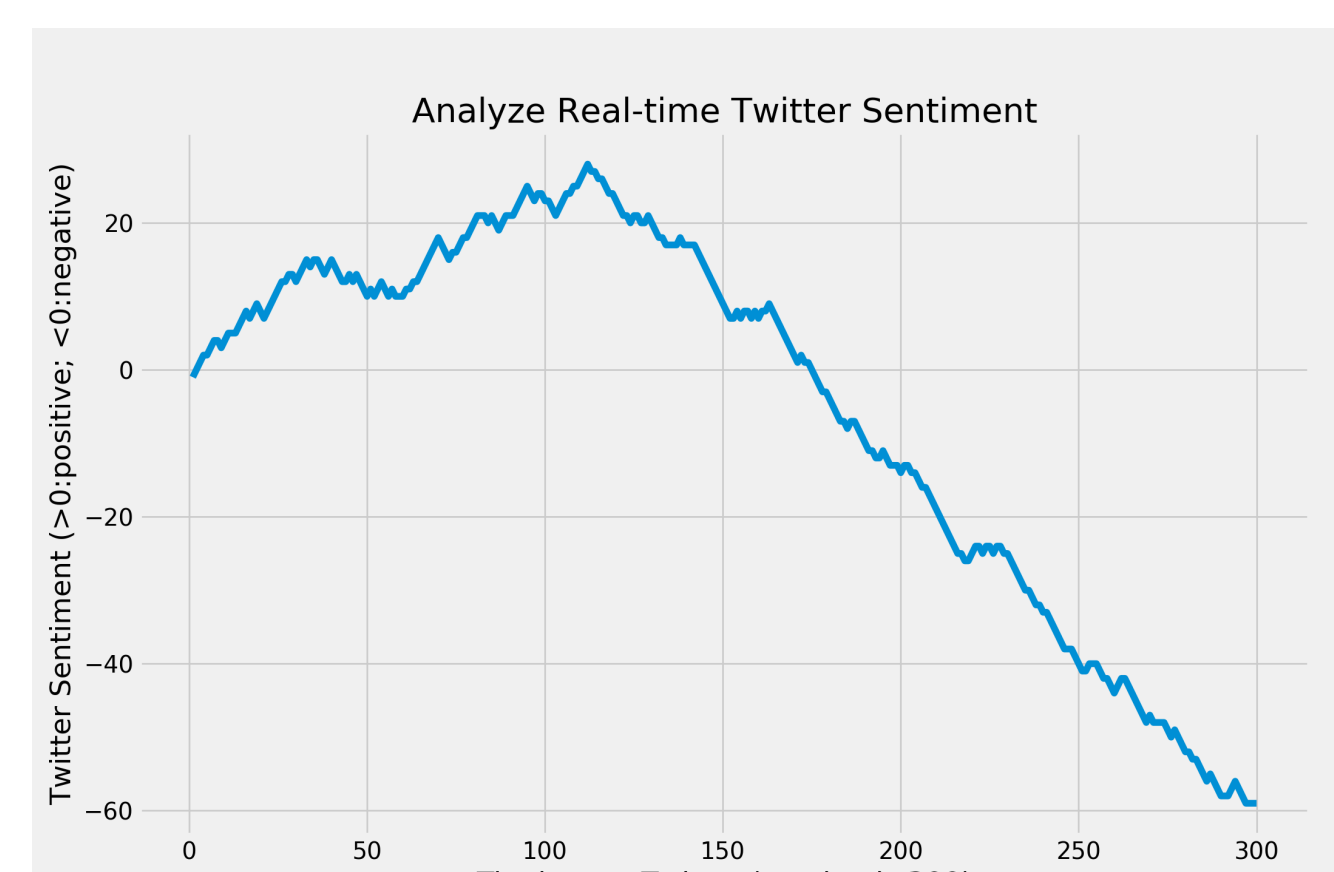
## Experiments

Experiment Process:

- Training data: Stanford sentiment 140 dataset, which contains 1.6 million tweets extracted from twitter API with emoticons removed + SemEval-2017 Task 4 dataset, which contains 49466 tweets in total.
- Task: The task is detecting whether a piece of tweet expresses a POSITIVE, a NEGATIVE, or a NEUTRAL sentiment. The sentiment can be general or about a specific topic, e.g., a person, a product, or an event.
- Testing data: We did a 5-fold cross-validation for testing the accuracy of our models.
- We applied several lexical based methods with n-gram to do the sentiment classification (SVC linear, logistic regression, Naïve Bayes)
- We improved the baseline by adding the detector for both emoji and emoticon.
- We also improved the baseline by applying the CNN and LSTM neural network models to improve the performance.
- We also tried to experiment with phrasing modeling and feature reduction method (PCA and chi-squared).



## Example



These two graphs shows the live sentiment analysis results of the tweets about Donald Trump. The latest 300 related tweets were included.(CNN+emoji/emotion detect).The result shows that the Neutral tweet amount is larger than the negative. The number of negative tweets is larger than positive.

## Reference

- 1. Wieslaw Wolny. 2016. Sentiment analysis of twitter data using emoticons and emoji ideograms.
- 2. Alec Go, Lei Huang, Richa Bhayani, 2009, Twitter sentiment Analysis.
- 3.Efthymios Kouloumpis, Theresa Wilson, Johnna Moore, 2011, Twitter sentiment analysis: The good the bad and the omg!.