Message level twitter sentiment analysis

Mentor: Mr. Sandip Modha

Presented By:
Ankita Mungalpara (201811020)
Himadri Pandya (201811029)
DAIICT, Gandhinagar

Outline

- Introduction
- Problem Statement
- Literature Survey
- Proposed Approach
- Results
- Conclusion
- Future Work

Introduction

- Sentiment Analysis in Twitter
- English tweets
- Message-level sentiment classification
- Use Long Short-Term Memory (LSTM)
- Word embeddings pre-trained on a big collection of Twitter messages
- A text processing tool suitable for social network messages for tokenization, word normalization, segmentation and spell correction

Problem statement

Given a message classify whether the message expresses positive, negative, or neutral sentiment (3-point scale)

Literature Survey

DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis

Authors:

- Christos Baziotis
- Nikos Pelekis
- Christos Doulkeridis

(From University of Piraeus - Data Science Lab Piraeus, Greece)

Proposed Approach

- Tweet Preprocessing and Cleaning using RE, NLTK and ekphrasis
- Tweet Visualization
- Try different combinations of various feature extraction techniques with various Machine learning and deep learning models
- Compare results

Tweet Preprocessing Using RE

Removing Twitter Handles (@user)

Original tweet

@solz_b He's a true Niners fan, he brought it up in a interview during his 2nd season. :D

He's a true Niners fan, he brought it up in a interview during his 2nd season. :D

Tweet after removing handle

Tweet Preprocessing Using RE

Removing Punctuations, Numbers and Special Characters

Tweet after removing handle

He's a true Niners fan, he brought it up in a interview during his 2nd season. :D

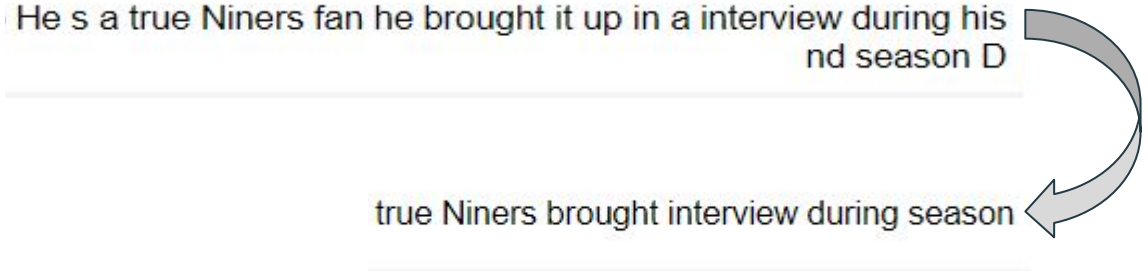
He s a true Niners fan he brought it up in a interview during his nd season D

Tweet after removing punctuation, numbers and special characters

Tweet Preprocessing Using RE

Removing Short Words

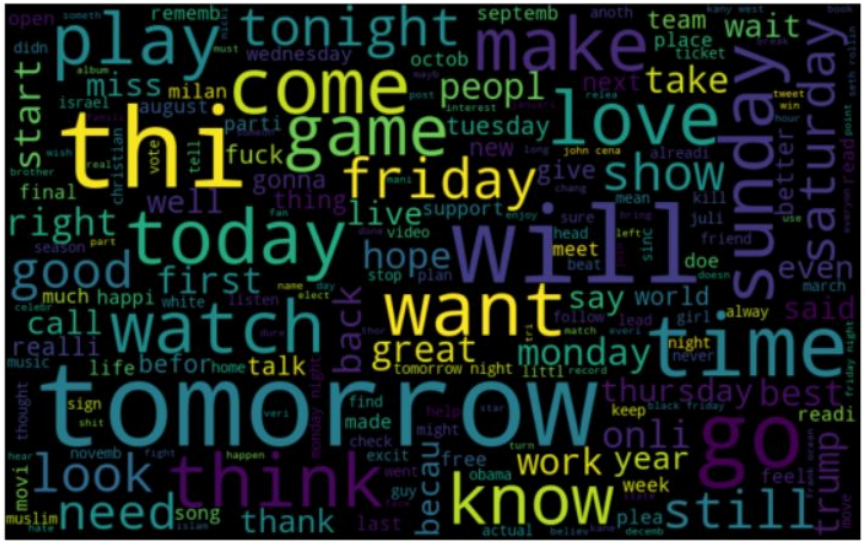
Tweet after removing punctuation, numbers and special characters



Tweet after removing short words

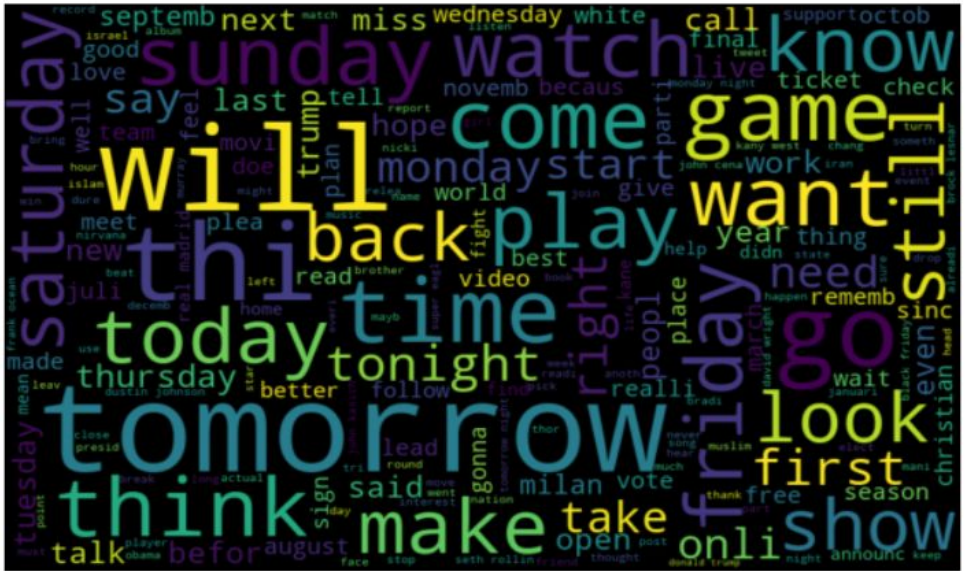
Tweet Visualization

Understanding Common Words Used in the Tweets



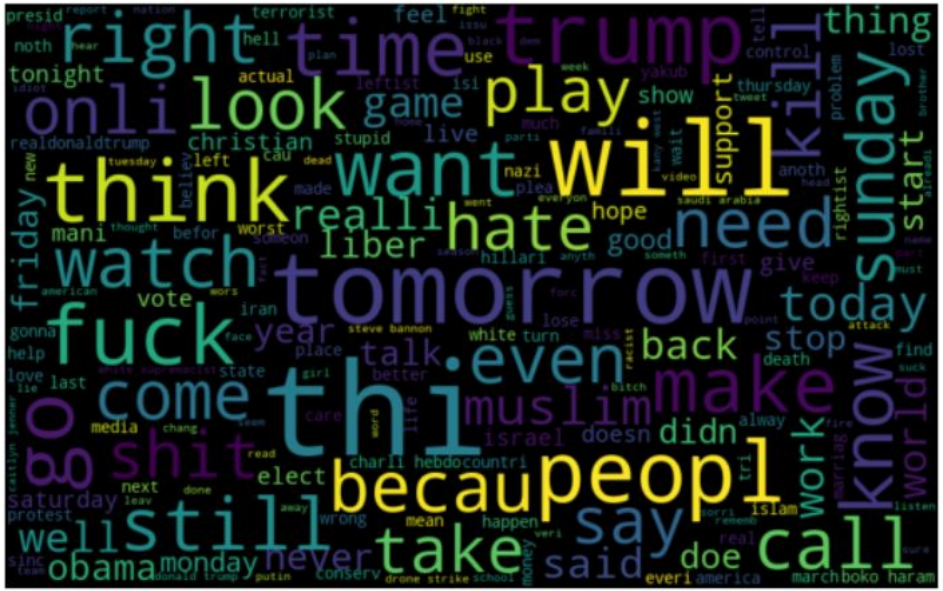
Tweet Visualization

Words in Neutral Tweets



Tweet Visualization

Words in Negative Tweets

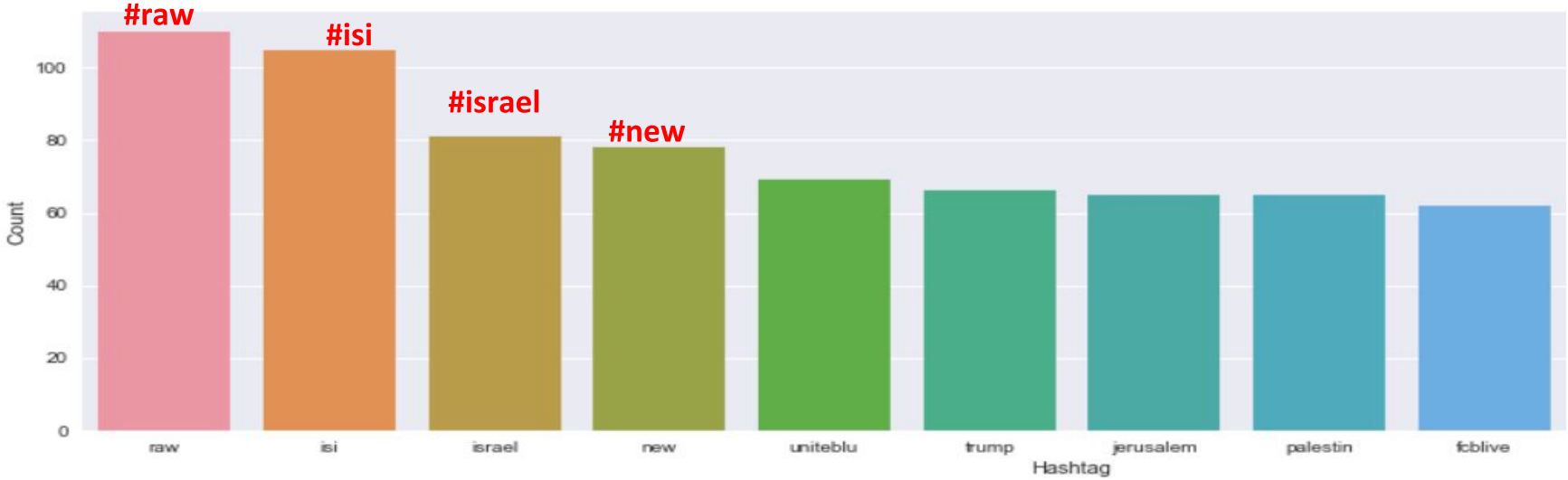


Tweet Visualization

Words in Positive Tweets

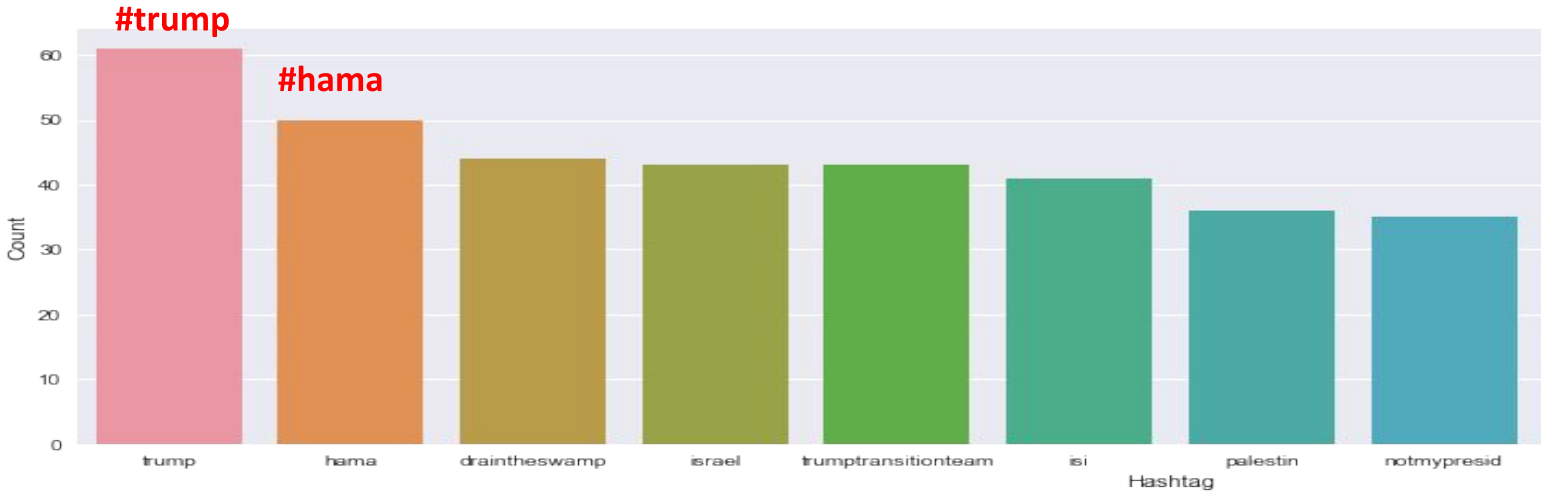
Tweet Visualization

Understanding Impact of Hashtags on Tweet Sentiment (Neutral Tweet)



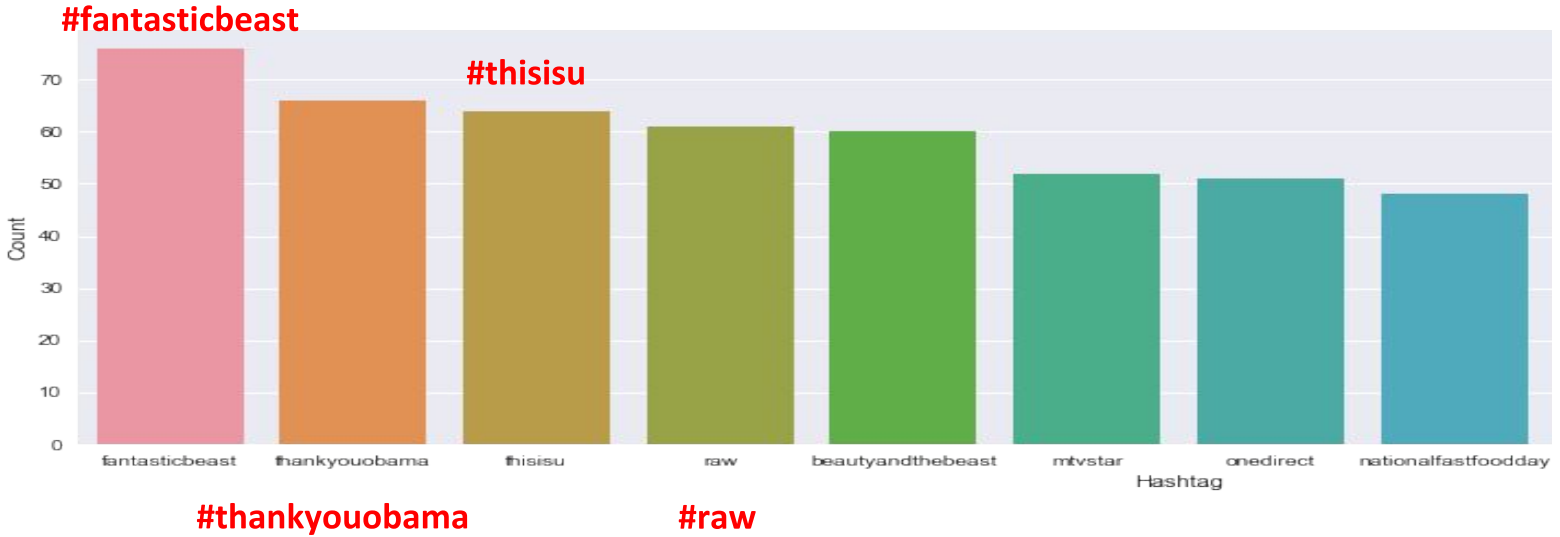
Tweet Visualization

Understanding Impact of Hashtags on Tweet Sentiment (Negative Tweet)



Tweet Visualization

Understanding Impact of Hashtags on Tweet Sentiment (Positive Tweet)



Our Results

Model	Recall	F1 Score
Logistic Regression		
CountVector Features	0.55	0.55
TF-IDF Features	0.53	0.52
Word2Vec Features	**0.59**	**0.59**
Support Vector Machine		
CountVector Features	0.54	0.54
TF-IDF Features	0.54	0.55
Word2Vec Features	**0.56**	**0.57**

Model	Recall	F1 Score
Naive Bayes Classification		
CountVector Features	0.58	0.56
TF-IDF Features	**0.58**	**0.57**
Stochastic Gradient Descent		
CountVector Features	0.55	0.55
TF-IDF Features	0.54	0.53
Word2Vec Features	**0.58**	**0.58**

Model	Recall	F1 score
NLTK with GloVe		
Naive Bayesian	0.33	.22
Logistic Regression	0.57	0.58
SVM	0.55	0.57
Stochastic Gradient Descent	0.56	0.57
Adaboost	0.49	0.50
Ekphrasis with GloVe		
Naive Bayesian	0.37	0.30
Logistic Regression	**0.61**	**0.62**
SVM	0.60	0.62
Stochastic Gradient Descent	0.59	0.61
Adaboost	0.52	0.53

Using LSTM

Model	Recall	F1-Score
RNN with ekphrasis and GloVe		
Attention	0.66	0.66
Regular	**0.69**	**0.68**

Our Results

Predictions On Tweets

Tweet	Predicted Sentiment
I am going to school	Neutral
Nice to meet you	Positive
I love you	Positive
We are not going for the movie	Negative
The food of that restaurant was not good at all	Negative
We went for tracking	Neutral

Thank You