Sarcasm Detection in Tweets

CS 725 : Introduction to Machine Learning

Kalpesh Patil - 130040019
Ashwin Bhat - 13D070006
Navjot Singh - 130110071
Yash Bhalgat - 13D070014

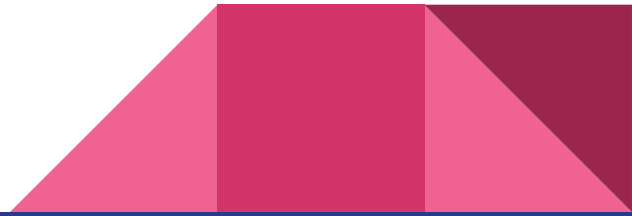
Introduction

Sarcasm is defined as a cutting, often ironic remark intended to express contempt or ridicule.

Sarcasm detection is the task of correctly labeling the text as 'sarcastic' or 'non-sarcastic'.

Challenging because :-

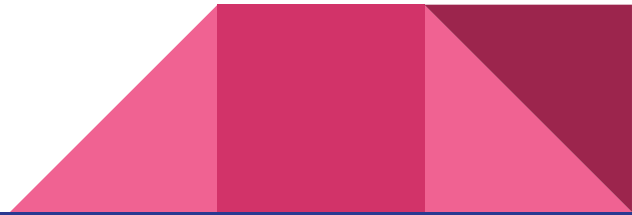
1) Lack of intonation
2) Lack of facial expressions



Data

- Querying the Streaming API using keywords #sarcasm and other sentiment tweets, filtering out non-English tweets and re-tweets

- Stream data using Twitter
  - https://github.com/guyz/twitter-sentiment-dataset
- Sentiment dataset for Sarcasm detection
  - https://github.com/dmitryvinn/twitter-sarcasm-measurement

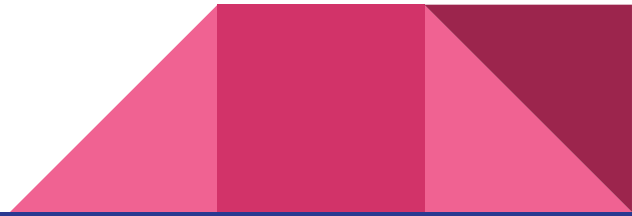


Feature Engineering

- Lexical Features
- Pragmatic Features
- Explicit Incongruity

Lexical Features

- **Unigrams** are used to extract lexical features from the text



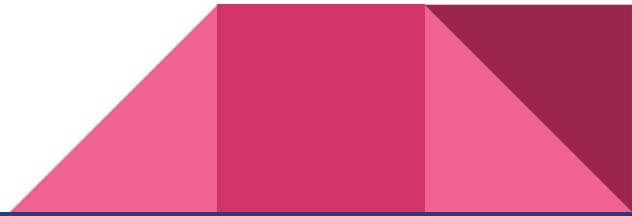
Pragmatic Features

- Number of capitalizations
- Number of emoticons
- Number of slang laughter words like “lol”
- Number of punctuation marks

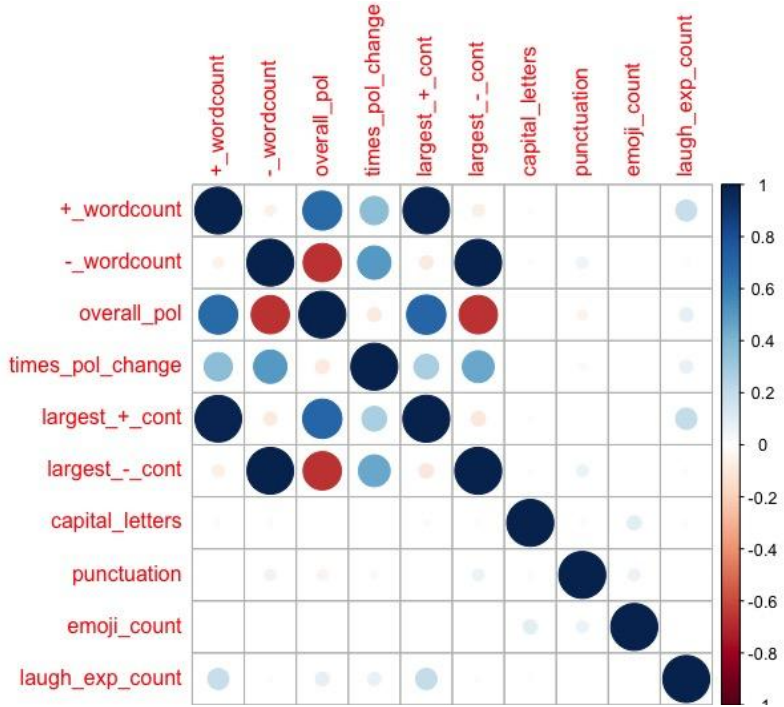
Explicit Incongruity

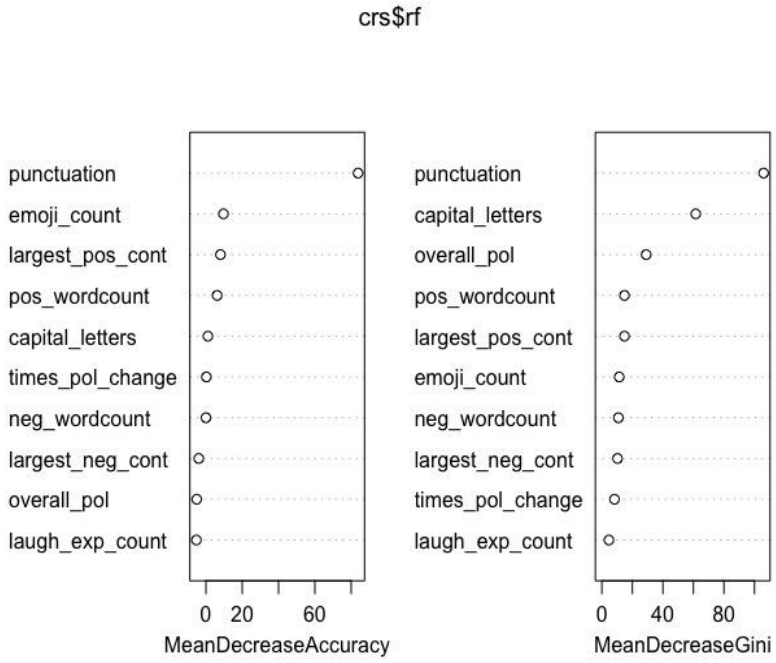
Linguistic theory - positive sentiment contrasted with negative situation

- Number of sentiment incongruities
- Number of words with positive or negative polarities
- Largest positive/negative polarity word sequence
- Overall lexical polarity of the input



Feature Analysis





Evaluation

Classifier Models tried →

- Random Forest
- Neural Network
- SVM

Random Forest

Actual \ Predicted	0	1
0	0.44	0.13
1	0.20	0.23

- Accuracy - 0.67
- Precision - 0.62
- Recall - 0.53
- F-score - 0.59

Neural Network

Actual \ Predicted	0	1
0	0.20	0.19
1	0.09	0.52

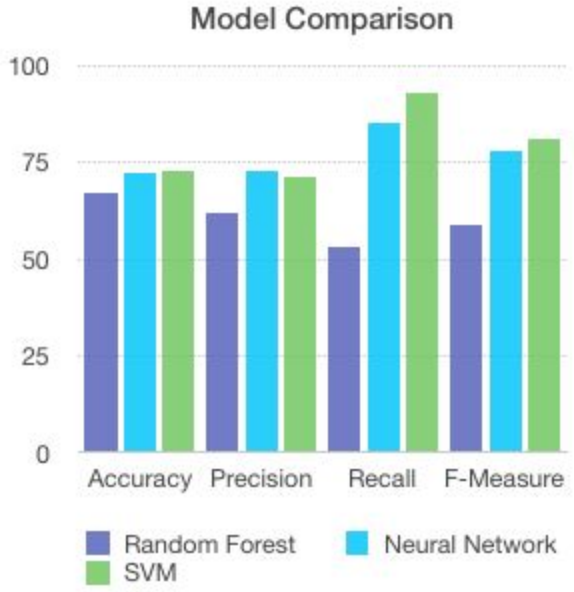
- Accuracy - 0.72
- Precision - 0.73
- Recall - 0.85
- F-score - 0.78

SVM

Actual \ Predicted	0	1
0	0.15	0.23
1	0.04	0.58

- Accuracy - 0.73
- Precision - 0.71
- Recall - 0.93
- F-score - 0.81

Comparison of Models



Conclusion

SVM model performs the best!

LibSVM could handle high dimensionality of features as compared to Neural Networks

Features like punctuation and capitalization are important since they reduce the impurity function

References

- LibSVM : A Library for Support Vector machines
- Rattle : A GUI for Data Mining using R
- Senti-Strength Tool : automatic sentiment analysis of word corpus.
- Keras: Deep Learning library for Theano and TensorFlow

References

- Cliche, M. The sarcasm detector, 2014.
- Bharti, Santosh Kumar, Korra Sathya Babu, and Sanjay Kumar Jena. "Parsing-based sarcasm sentiment recognition in Twitter data."
- Joshi, Aditya, Vinita Sharma, and Pushpak Bhattacharyya. "Harnessing context incongruity for sarcasm detection."

Thank You