Summary:

Sarcasm is another problem that faces the users of social media often. The sarcasm is the usage of the words in a different way such as insulting someone with positive words.

Proposed method:

according to Edwin Lunando and Ayu Purwarianti in their paper “Indonesian Social Media Sentiment Analysis with Sarcasm Detection” they proposed a model that consists of three main parts as follows :

1-PreProcessing

2-Feature Extraction

3-Classification

Explaining Proposed Model:

First we are going to talk about the PreProcessing Part. Mainly in their model the preprocessing part consist of the word correction from informal word to formal word or to correct misspelled word and to convert the numeric characters used in the text into alphabetic character.

Second the Feature Extraction. They used the Uni-gram , Negativity , Number of interjection words , Question Words. They used the Uni-gram because they found that it is more suitable for the Indonesian language because the Indonesian language mainly consists of informal words , they used the negation to calculate the weight of the word as when the word like “not” is places in the text it is going to change the weight of the following text. Then they calculated the negativity of the words according to the public negativity of a certain topic. Then they used the number of interjection words , they find these words because according to their calculations they found from every 100 text that have interjection word there is 20 text that is classified as sarcasm. Last feature extraction method is the Question word , when they find that the text has question word so this sentence is neutral has no weight in the text.

Third part the classification. Their classification model consists of two parts the first part is the part that they classify the text in to three classes that it is positive or negative or neutral. Then the second part that they classify the positive text to find whether it is an opinion or neutral and if it is an opinion it is positive or negative.

Datasets:

They collected the dataset manually from Twitter.

Results:

they used 3 models

Naive Bayes : 76.5%

Maximum entropy: 76.7 %

Support Vector Machine: 77.3%