The Lowest Form of Wit: Identifying Sarcasm in Social Media

Vivian Hsu, Saachi Jain

Data and Preprocessing

Our dataset consists of English tweets obtained using the Google add-on Twitter Archiver. We downloaded tweets from November 10th to December 3rd.

Referencing Liebrecht et. al and González-Ibáñez et. al, we obtained sarcastic tweets by getting tweets with the hashtag "#sarcastic." We further assumed that any tweets expressing "emotional" hashtags such as "happy," "joy," "lucky," "sad," "angry," and "disappointed" were non-sarcastic tweets expressing positive or negative sentiment.

To clean the data, we took out all hashtags, all links to other websites, and all tags to other accounts (tokens beginning with "@"). After cleaning, if a tweet has fewer than three tokens left, we took it out of our data set. After processing, we had **26,206** sarcastic tweets and **101,361** non-sarcastic tweets.

Features

Our features included the frequency of unigrams (lemmatized, with punctuation as their own grams), and bigrams (punctuation stripped). Each gram was its own feature.

We further had three features on the overall context of the sentence. We had a feature for the number of words in all caps (greater than 1 letter). Referencing Dr. Matthieu Cliche's work, we used *pattern.en* to split the tweet into two chunks: words before and including the verb phrase and words after verb phrases, and found the differences in the polarity of the sentiment using *TextBlob*. We then used *TextBlob* to find the subjectivity of the entire tweet.

Feature Name	# of Features
Unigrams	8,223
Bigrams	20,790
Other Features	3
Total Features	**29,014**

Results and Analysis

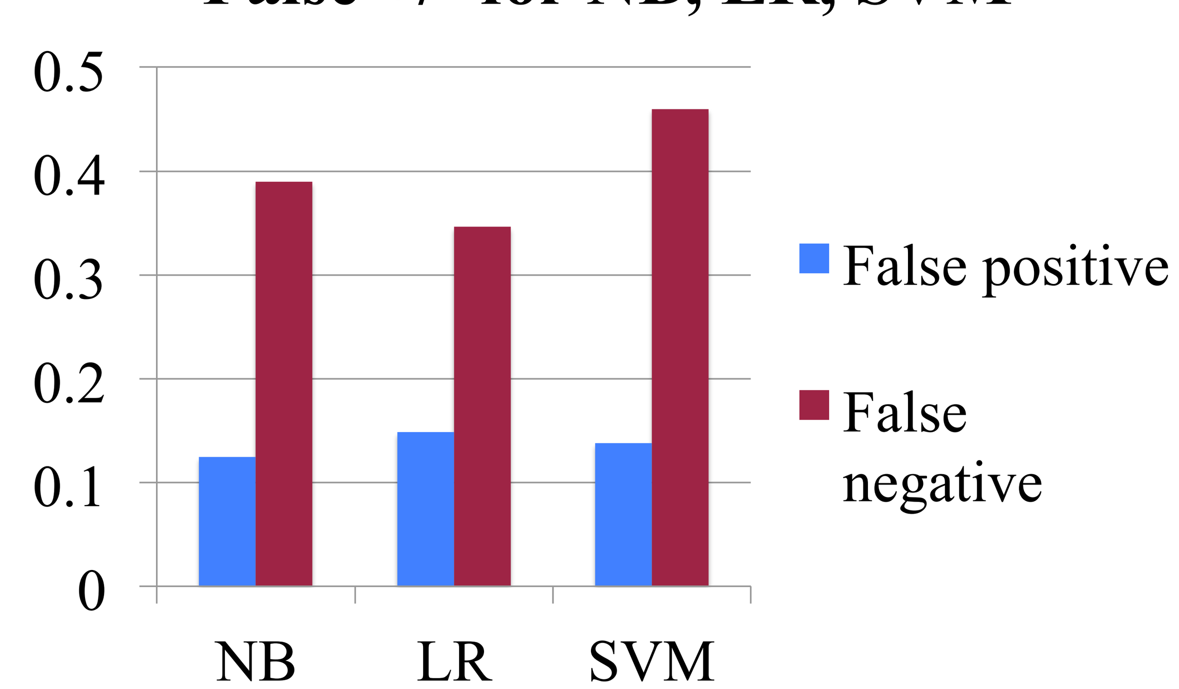
We used sklearn's implementation of Multinomial Naïve Bayes, Logistic Regression, and Linear Support Vector Classification.

We partitioned 70% of our tweets into a training set, and 30% into the testing set. We then ran each of the three models.

To see that our models have learned from the features, we also ran the three models where each tweet had only one feature, randomly assigned to be 0 or 1.

Model Name	Accuracy	Random
Multinomial NB	0.8250	0.7843
Logistic Regression	0.8247	0.7843
Linear SVM	0.7986	0.7843

False +/- for NB, LR, SVM

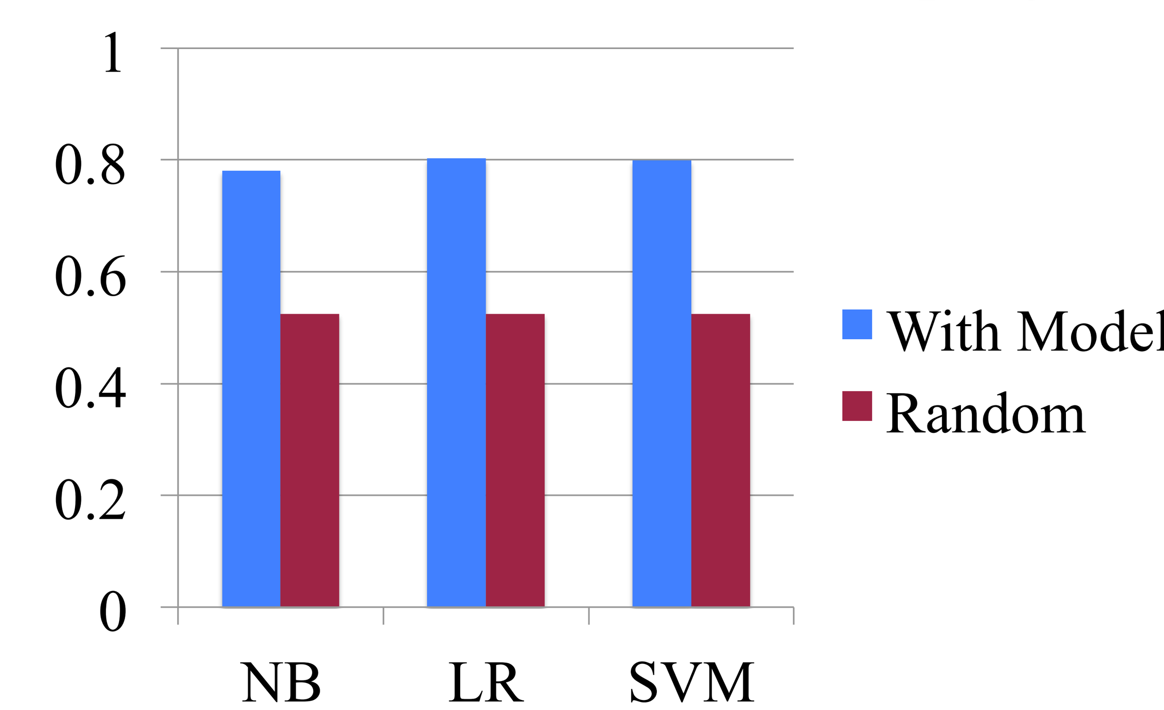


To provide a robust calculation of each model's efficacy we performed 4-fold cross validation using sklearn's cross validation module on each of the three models.

Model Name	Accuracy
Multinomial NB	0.8341
Logistic Regression	0.8475
Linear SVM	0.8284

To further explore the effects of an unbalanced training set, we oversampled the sarcastic tweets such that we had each sarcastic tweet occur 4 times in the dataset. Thus, we had about an equal number of sarcastic and not sarcastic tweets in the data set.

NB, LR, SVM with Oversampling



To estimate how much each type of feature contributed to the accuracy, we performed single feature calculations. Below is averaged accuracies across all three models.

Feature Name	Accuracy
Unigrams	0.8092
Bigrams	0.8138
Capitalization	0.7818
Sentiment Split	0.7817
Subjectivity	0.7843

Discussion

Logistic Regression was best at predicting whether a tweet is sarcastic. Of our features, bigrams were most significant in performing a correct estimate.

Currently there are some limitations to our data set. We had no neutral non-sarcastic tweets, because of the difficulty of annotating a general pool of tweets without the (relatively rare) "#sarcastic" hashtag. Similarly, it was impossible to find a concrete statistic on how many tweets out of a general tweet set would be sarcastic, so our percentage of sarcastic tweets in our test set may be higher than in real life. In future work, we would like to expand our data set to a larger and more representative set of tweets.

We will further explore more accurate classifiers for predicting and evaluating unbalanced data.

Sarcasm detection can have several applications in security, sales, and health. A sarcasm detector can not only help people interpret others' writings, but can also help writers avoid being misunderstood.

References

Cliche, Mathieu, Ph.D. *The Sarcasm Detector*.

Gonzalez-Ibanez, Roberto, Smaranda Muresan and Nina Wacholder. "Identifying Sarcasm in Twitter: A Closer Look."

Liebrecht, Christine, Florian Kunneman, and Antal Van Den Bosch. "The Perfect Solution for Detecting Sarcasm in Tweets #not."