**Naïve Bayes Twitter Sentiment Analysis**

**Introduction**

Sentiment Analysis, simply put, analyzes a String and tells whether or not that String has a positive or negative connotation attached to it.

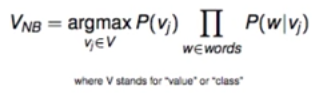
This topic deserves attention because many companies use sentiment analysis tools on the web to track the public opinion of their products, services, or organization. Such tools could help businesses adjust and target markets where they do not have a strong foothold. In the end, knowing what tools to use and how to apply them could make or break a business.

Using a supervised machine learning probabilistic classifier, our team shows that we are generally able to determine the sentiment of text. We say “generally” because no learning algorithm is 100% effective. That being said, our team trained our Naïve Bayes classifier on datasets comprised of a two class, positive and negative, system. By extending the Naïve Bayesian classifier to look at twitter we hope to find some clarity and determine the general consensus in the overabundance and oversaturation of news to which we are exposed.

**Naïve Bayes**

A Naïve Bayes classifier works by applying Bayes Theorem in a supervised manner. The classifier is trained on a pre-classified dataset to allow the classifier to learn what words comprise a positive block of text and what words comprise a negative block of text. These pre-classified samples allow us to build up a bag of words for each class presented in the training set.

Since the Naïve Bayes classifier calculates the probability that a given statement belongs to each class, the position of the words in the block of text and its immediate context is negligible. Ignoring the word’s position in the text may seem counterintuitive, but this approach actually works quite well[[1]](#footnote-1). Each word, given a class, gets assigned a probability.



Naïve Bayes Formula

Stated above is the general formula that is used to classify new data. In this formula, *v* is the class (positive or negative) and *w* is the word in a given text. The value of the text is thus classified by the maximum, normalized, probability computed by aggregating the product of each word in the text, given each class.

We designed our Naïve Bayes classifier to have a few more features than just the bare-bones classifier. We gave the user the option of choosing whether or not they want the classifier to continue learning, to discriminate between what it learns and what it does not learn, to use an English language stemmer, or to write newly classified data back to its training set (so as to not lose any progress).

If the classifier continues learning, each word in the block of text is added as a key value pair to the given class’s vocabulary. Therefore, if this word appears in a new block of text, the newly learned context will add more weight to the given class. As a safety measure, to minimize the chance of learning a word incorrectly (say a bad word is used in a positive context), we decided to only add the word to the dictionary if the classifier is at least 90% certain of the classification. Our team chose 90% since it was empirically the best choice. Anything less than or greater than 90% had diminishing returns.

To deal with stop words and strange characters found in tweets, the classifier removes the stop words and symbols from the sentence so as to not skew the probability. Lastly, our team implemented an English language stemmer to converge multiple tenses of a given word to one word in the class vocabularies[[2]](#footnote-2). This serves to strip away the unnecessary words, characters and suffixes to allow the vocabulary to only include relevant words.

Porter Stemmer Examples

**Twitter Integration**

The Sentiment Classifier is a standalone classifier that can work in many contexts. However, for this paper we integrated it with a simple PHP twitter aggregator, so that we have an easy way to collect relevant items to classify.

Using an open source PHP Twitter API Wrapper, we host a simple web app on an apache server to aggregate tweets by either a Twitter handle or Twitter hashtag[[3]](#footnote-3). Along with choosing to search by a Twitter handle or hashtag, the user has the option to specify the number of tweets he or she wants sourced. Once Tweets are sourced via the query in a web GUI, we populate a text file with just the contents of the Tweets. This text file is then passed into the Naïve Bayes classifier for classification.

**Results**

After revising our original proposal, we settled on aiming for an accuracy of 65% when classifying statements. At first we were aiming to accomplish at least 80%, but had no support to back up that claim.  After looking at similar projects and seeing how unrealistic our first goal was, we revised our definition of success. [[4]](#footnote-4)

Under ideal circumstances, the classifier has the capability of achieving ~77% accuracy. The testing set is ideal because it does not contain neutral sentences[[5]](#footnote-5). Since the training set does not contain any neutral sentences, our classifier can perform quite accurately. Our Naïve Bayes classifier does not do well in the case where it is exposed to neutral text because we only consider the classifications of positive or negative text. That is not to say that our classifier cannot handle the neutral class given adequate data, but it has not been trained to do so.

It is worth stating, for clarity, that we used two distinct datasets for training and testing. Each of these two sets are pre-classified in a consistent manner. Therefore, our classifier can immediately check the classifier’s efficacy.

Checking the classifier’s efficacy against Twitter data is slightly more challenging. In real life, not every block of text can be neatly categorized as either positive or negative. There are many shades between the two. That being said, our implementation of the Naïve Bayes classifier seems to empirically meet our goal of reaching at least 65% accuracy. Furthermore, we noticed that our classifier is more effective and accurate when it is set to the learning mode.

The base case for testing accuracy was after training the classifier with the original training set, and evaluating the testing set with learning off. This configuration achieved a ~76% accuracy. When this same test was performed after running hundreds of tweets through the classifier with learning was turned on, we obtained an accuracy of ~77%.

**Project Evolution**

Currently, our Naïve Bayes classifier is only being trained on a two-class system - positive and negative. Koppel and Schler (2006) in a research paper presented their findings that a neutral class can improve overall accuracy of a classifier conducting sentiment analysis[[6]](#footnote-6). Therefore, in the future, we would invest more resources into vetting more training data to cover all edge cases and incorporate all the classes necessary to adeptly represent real-life situations.

**Conclusions**

In conclusion, our team was able to successfully create a Naïve Bayes classifier to accurately classify text in an ideal environment and adeptly classify text in a non-ideal environment, such as Twitter. While there is room for improvement, namely adding a neutral class to the training set, our classifier surpassed our expectations of 65% by achieving an accuracy score of ~77% on a dataset from a University of Michigan sentiment analysis competition.

1. [CS 188 Lecture](https://www.youtube.com/watch?v=DNvwfNEiKvw&t=438s) [↑](#footnote-ref-1)
2. [Porter Stemmer](http://www.tartarus.org/~martin/PorterStemmer/index.html) [↑](#footnote-ref-2)
3. [Twitter API PHP](https://github.com/J7mbo/twitter-api-php) [↑](#footnote-ref-3)
4. [Twitter Sentiment Analysis](http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2857/3251) [↑](#footnote-ref-4)
5. Dataset is from [UMichigan](https://inclass.kaggle.com/c/si650winter11/data) [↑](#footnote-ref-5)
6. [Neutral Class - Sentiment Analysis](http://blog.datumbox.com/the-importance-of-neutral-class-in-sentiment-analysis/) [↑](#footnote-ref-6)