

ABSTRACT

Twitter is a popular social networking website where users post and interact with messages known as “tweets”. This serves as a mean for individuals to express their thoughts or feelings about different subjects. Various different parties such as consumers and marketers have done sentiment analysis on such tweets to gather insights into products or to conduct market analysis. Furthermore, our sentiment analysis predictions are able to improve. In this report, we will attempt to conduct sentiment analysis on “tweets” using various different machine learning algorithms. We attempt to classify the polarity of the tweet where it is either positive or negative. If the tweet has both positive and negative elements, the more dominant sentiment should be picked as the final label. The data provided comes with emoticons, usernames and hash tags which are required to be processed and converted into a standard form. It also needs to extract useful features from the text such unigrams and bigrams which is a form of representation of the “tweet”. We use various machine learning algorithms to conduct sentiment analysis using the extracted features. However, just relying on individual models did not give a high accuracy so we pick the top few models to generate a model