Sentiment Analysis

|  |  |
| --- | --- |
| 000  00  00  00  00  00  00  00  00  00  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 | 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |

|  |
| --- |
|  |
| Andi Mohammad Adnan Kofi  **Muhammad Husain Toding B.** |
|  |
|  |
|  |
|  |
|  |

Abstract

Sentiment analysis is the process of determining whether a piece of writing is positive, negative, or neutral. There are many methods that use to solve this problem. The methods that used to solve this problem are Naïve Bayes, SVM, and Regression. And by using those methods, the comparison between them will resulting which method that gives the best result by depending on the accuracy for each method.

Introduction

In this digital era, there are many features inside technology that exist. Like the readers can give feedback or comments by online in some sites. So, the sites also know how to react for reader that gives comments or feedback. In order to know what the comment refers to, whether it is positive or negative comments, there is a method that called sentiment analysis. **Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It’s also known as opinion mining, deriving the opinion or attitude of a speaker.** A common use case for this technology is to discover how people feel about a particular topic. So, it will give results which comment is positive or negative from given data test using 3 different methods which are Naïve Bayes, SVM, and Regression. By comparing those 3 methods, it will resulting which one is the best method that use to solve this problem by considering the accuracy of each method.

Methods

For the library to processing textual data is TextBlob. It provides a simple API for diving into common NLP tasks such as POS-Tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. In here it only used for sentiment analysis. There are 3 methods that will be use. The first method that will use is Naïve Bayes classifier. Naive Bayes classifiers are a collection of classification algorithms based on **Bayes’ Theorem**. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class. In here the Naïve Bayes works to classify the result whether the comments positive or negative. The second method is SVM. This method is also use to classify the comment result. It will find the best hyperplane that separates the data points of two different classes.in here it used linear kernel. And the last method that will use is regression. It will find the vector pattern to classify the comment results.

|  |  |
| --- | --- |
|  |  |

Experiment

In this experiment, the method is by using Naïve Bayes approach to classify the result by using the data set from IMDb. IMDb itself is an online database of information related to world films, television programs, home videos and video games, and internet streams, including cast, production crew, personnel and fictional character biographies, plot summaries, trivia, and fan reviews and ratings. So, from the given input, it will give an output whether the input/comment is positive or negative. There are some variation of input example that will conduct. Like if the input is only positive comment, the input only negative comment, the input is contain both, and the last is the input is not listed in most informative features (i.e word “Like”). So, for example for the first variation if the input is only positive like “it was a funny movie”, it will result positive. Because the word “funny” is exist in most informative features and its define as positive word. And for the second variation if the input is only negative comment like “it was a bad movie to watch”, it will result negative. Because the word “bad” is exist in most informative features and its define as negative word. For the third variation which is the comment contain both word (positive and negative) like “the actor was great but the movie was boring”, it will result negative. And for the last input if the comment contains the word that not in the list of most informative features like “good” word. So, it will result negative comment even though the word “good” is a positive.

|  |  |
| --- | --- |
| 000  00  00  00  00  00  00  00  00  00  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 | 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |

# Analysis

From the experiment above, by using Naïve Bayes classifier, the result will heavily depend on the most informative features. The more most in-formative features shown, most likely the result will be the same as that feature’s class. And the accuracy from the data test is 74.4 %. And for the other method which is SVM, the accuracy from the data test is just 44.2 % and for Regression, the accuracy from the data test is 74%.

# Conclusion

|  |  |
| --- | --- |
|  |  |

|  |  |
| --- | --- |
|  |  |

|  |  |
| --- | --- |
|  |  |

The conclusion is the best method to use for solving the sentiment analysis problem is by using Naïve Bayes approach that has highest accuracy between other methods.

.

|  |  |
| --- | --- |
|  |  |