

Abstract

Twitter is a social networking service and microblogging which allows users to send and read text-based message up to 140 characters, with a very rapid progress Twitter became the object of analysis is excellent for various purposes, one of the study were in demand today with the social media Twitter is analysis sentiment and opinion mining. To carry out opinion mining on Twitter constrain themselves because of data limitations Twitter from twit users are only limited by 140 characters, in addition to the research opinion mining usually only focused on classification or clustering of data but not much to explain stage preprocessing, basically preprocessing good will generate process mining is good as well, it needs a variety of ways to maximize the process of preprocessing on Twitter one of which is the process of stemming by implementing algorithms Soundex where the algorithm is expected to maximize the process of stemming the preprocessing of process mining on Twitter, in addition to this method will be paired with a variety of algorithms weighting Term Frequency (TF), Feature Term Presence (TP), Term Frequency-Inverse DocumentFrequency (TF-IDF) to find a partner the right algorithm to support the classification process is good, the classification is done by the method of Naïve Bayes which can then be analyzed how the soundex algorithm for stemming the influence and effect of the weighting algorithm when applied to the classification process, in addition to the research is expected to produce algorithms that give a good contribution to the process of data stemming Twitter and learn how to influence the weighting algoritma when paired with soundex algorithm. After an examination of stemming compared the results stemming algorithm soundex with porter then showed to the test data as much as 300 twit that soundex slightly superior then classified data of stemming the soundex with some algorithm weighting the results obtained accuracy values are the same, based on the analysis found that the algorithm weighting does not affect the classification results.

Keywords : Twitter, *Algoritma Soundex, Stemming, Preprocessing, Naïve bayes*