

ABSTRACT

Social media has been a part of society since few years ago, especially in Indonesia. One of the most popular social media among Indonesian citizen is Twitter. Social media become popular since it allows people to post their opinion, experience or whatever concern them. After Joko Widodo won the Presidential election in 2014, people tend to post their opinion about his performance as a president.

This research aims to analyze the opinions that posted by Twitter users in Indonesia using Naive Bayes Classifier method and WEKA version 3.6.13. This research conducted on two datasets which origin from same data resources. Each dataset contains 1862 tweets and manually classified. The first dataset classified into four classes: negative, critic, suggestion and positive. The second dataset classified into two classes: negative and positive. This research was conducted to answer three hypotheses.

The result of this research which use the significance level of 1% showed that the dataset which contains four classes has worse accuracy, precision and recall value than the dataset which contains two classes. Furthermore, different tokenization method and stopwords list will also affect the result of the analysis process.

Keywords: sentiment analysis, Naive Bayes Classifier, Twitter

INTISARI

Media sosial sudah menjadi bagian dari kehidupan masyarakat Indonesia dan berbagai manfaat sudah banyak dirasakan. Salah satu media sosial yang banyak digunakan oleh masyarakat adalah Twitter. Berbagai macam opini banyak dilontarkan setiap harinya oleh para penggunanya, dan tidak luput diantaranya adalah opini terhadap pemerintah di bawah kepemimpinan Presiden Joko Widodo.

Penelitian ini bertujuan untuk melakukan analisis terhadap opini yang dikeluarkan oleh pengguna internet yang menggunakan layanan Twitter dengan menggunakan metode klasifikasi bernama Naive Bayes Classifier. Penelitian ini dilakukan dengan memanfaatkan aplikasi bernama WEKA dengan data yang didapat dari Twitter untuk menjawab tiga pertanyaan penelitian yang diajukan. Dataset yang digunakan dalam penelitian ini adalah sebanyak dua buah dengan masing-masing dataset berisi 1862 buah tweet yang berhasil dikumpulkan kemudian diklasifikasikan secara manual. Dataset pertama memiliki empat kelas atribut yaitu negatif, kritik, saran dan positif, sedangkan dataset kedua memiliki dua kelas atribut yaitu negatif dan positif.

Hasil pengujian dengan tingkat signifikansi sebesar 1% pada penelitian ini menunjukkan bahwa *dataset* yang memiliki empat kelas atribut menghasilkan nilai parameter yang berupa *accuracy*, *precision* dan *recall* lebih rendah dibandingkan dengan *dataset* yang memiliki dua kelas atribut. Selain itu, penggunaan metode tokenisasi dan *stopwords list* juga memiliki pengaruh terhadap hasil analisis.

Kata kunci – analisis sentimen, Naive Bayes Classifier, Twitter