

INTISARI

Analisis sentimen adalah salah satu metode *machine learning* yang dapat digunakan untuk mengidentifikasi pendapat/opini dari sebuah data yang berbentuk teks, misalnya apakah dalam sebuah teks mengandung emosi senang, sedih, marah, dan lain-lain. Analisis sentimen banyak diterapkan untuk berbagai kebutuhan, terutama diluar negeri. Contoh penerapannya misalnya dalam bentuk *chatbot* untuk dapat menghasilkan respons yang natural, atau dalam bentuk pemantauan marketing perusahaan untuk mengetahui opini konsumen terhadap perusahaan tersebut apakah opininya positif atau negatif. Namun, metode Analisis sentimen yang sudah diterapkan di luar negeri ini tidak dapat serta-merta diterapkan untuk teks berbahasa Indonesia.

Untuk menghasilkan model analisis sentimen yang akurat dalam bahasa Indonesia, pertama-tama perlu dilakukan teknik *pre-processing* yang berbeda dari teknik *pre-processing* yang dilakukan terhadap data berbahasa Inggris. *Pre-processing* adalah sebuah proses pengolahan data-data sebelum diterapkan di dalam model *machine learning*. Untuk menemukan teknik *pre-processing* yang sesuai untuk teks bahasa Indonesia, perlu dilakukan pengujian dengan menggunakan berbagai macam teknik *pre-processing* yang sudah ada dengan model analisis sentimen sederhana untuk menemukan teknik *pre-processing* yang menghasilkan akurasi yang paling tinggi. Hasil dari percobaan pada tugas akhir ini kemudian akan diterapkan dalam bentuk *tools* yang dapat digunakan untuk tahap *pre-processing* pada data berbentuk teks.

Setelah melakukan rangkaian percobaan menggunakan berbagai macam teknik *pre-processing*, telah terbukti bahwa penggunaan teknik *pre-processing* dapat meningkatkan akurasi dari sebuah model analisis sentimen yang menerapkan algoritma SVM, KNN, dan SGD. Dengan menerapkan penghilangan *noise* pada sebuah data kasar, nilai akurasi dari model analisis sentimen dapat meningkat sebesar 3% sampai 6%, selanjutnya dengan menambahkan teknik *pre-processing* lainnya seperti *stemming* dan penghilangan *stopwords*, pada data yang telah dihilangkan *noise*-nya, akurasi dapat semakin meningkat sebesar 0.5% sampai 2%. *Tools* *pre-processing* yang dibuat berdasarkan hasil percobaan tersebut dapat mengurangi jumlah kata pada sebuah baris data dari 16,6 kata per baris data menjadi 13,9 kata per baris data.

Kata kunci : Analisis sentimen, *Machine Learning*, *Text pre-processing*, *Natural Language Processing*

ABSTRACT

Sentiment analysis is one of many methods in machine learning which is used to identify an opinion from text data, for example whether a specific text data contains specific emotions like happiness, sadness, and anger. Sentiment analysis is already widely especially overseas, this implementation including chatbot to deliver a natural, human-like response or in company marketing department to find out consumer opinions on the company whether they react positively or negatively to company's actions. But sentiment analysis methods which already implemented overseas can't haphazardly implemented on Indonesian based text.

To generate a accurate sentiment analysis model in Indonesian language, firstly a different pre-processing method is needed to be implemented on Indonesian language text. Pre-processing is a process of preparing data before being implemented on a machine learning model. To find which pre-processing method works best in Indonesian language, a test using various pre-processing methods is needed which then is tested on a simple sentiment analysis model to find which method yields the best accuracy. The result from conducting this experiment is then applied into a tools that can pre-process text data.

After conducting various experimentation using different pre-processing techniques, it is proven that the usage of pre-processing techniques can improve the accuracy of a sentiment analysis model which implement SVM, KNN, and SGD algorithm. By using noise removal method on a raw data, the accuracy yielded from a sentiment analysis model seems to improve by 3% to 6%. In addition to that, by applying stemming and stopwords removal methods, the accuracy from the model increased even further by 0.5% to 2%. The pre-processing tools built based on this experiment is able to reduce the number of words per data from 16,6 words per data to 13,9 words per data.

Keywords : Sentiment analysis, Machine learning, Text pre-processing, Natural Language Processing