

ABSTRACT

SENTIMENT ANALYSIS IN TWITTER FOR E-KTP SUBJECT

By

Astri Charolina
13/356480/PPA/04433

Currently Twitter has been widely used as a source of data for research. Research themes are quite popular and often raised, among others: the analysis of sentiment, topical cluster of tweets. Issues raised in various studies also varied, such as issues related to politics, economics, social, culture, corruption, sports, market analysis for certain products and so forth. One of the hottest topics that is currently a trend Twitter at this time is about the case of mega corruption e-KTPd. This study aims to develop a system that can cluster tweets in positive, negative and neutral sentiment classes generated by the e-KTP tweet-themed sentiment analysis process.

The algorithm used in the sentiment analysis process is Naive Bayes Classifier. As for clustering used Lingo algorithm and Suffix Tree Clustering (STC). In the application development process, used the library API Twitter4J for tweet. Other than that, also used Carrot² Framework in which there is a library for clustering, the algorithm Lingo and STC.

The resulting system is then tested. The testing process consists of two parts, namely the test of sentiment analysis accuracy and clustering testing. In the process of testing accuracy with test data as much as 4 groups of data, with each numbered 50, 100, 150 and 200 tweet. The test results show the average system accuracy level in analyzing sentiment is 89.67%. With the results of this accuracy it can be said that Naive Bayes Classifier feasible to be implemented. While the clustering test obtained results that in terms of processing time, STC algorithm is faster when compared with Lingo algorithm with the average processing time of 66.2 milli second. While in terms of number of clusters, Lingo algorithm produces more cluster number when compared to STC algorithm.

Keywords: Sentiment Analysis, Clustering, Naive Bayes Classifier, STC.

INTISARI

ANALISIS SENTIMEN TWEET BERTEMA KARTU TANDA PENDUDUK ELEKTRONIK (E-KTP)

Oleh

Astri Charolina
13/356480/PPA/04433

Saat ini Twitter telah banyak digunakan sebagai sumber data bagi penelitian. Tema penelitian yang cukup populer dan sering diangkat antara lain: analisis sentimen, topical cluster of tweets. Isu-isu yang diangkat dalam berbagai penelitian pun beragam, misalnya isu terkait dengan politik, ekonomi, sosial, budaya, korupsi, olahraga, analisis pasar untuk produk tertentu dan lain sebagainya. Salah satu topik terhangat yang saat ini menjadi tren perbincangan pengguna Twitter saat ini adalah mengenai kasus mega korupsi e-KTP. Penelitian ini bertujuan untuk mengembangkan sistem yang dapat mengkluster kumpulan tweet bertema E-KTP dan dapat mengklasifikasikan tweet-tweet pada setiap topik atau cluster ke dalam kelas sentimen positif, negatif dan netral.

Algoritma yang digunakan dalam proses analisis sentimen adalah *Naive Bayes Classifier*. Sedangkan untuk *clustering* digunakan *Suffix Tree Clustering* (STC). Dalam proses pengembangan aplikasi, digunakan *library* API Twitter4J untuk pengambilan tweet. Selain itu juga digunakan *Carrot² Framework* yang di dalamnya terdapat *library* untuk clustering, yaitu algoritma Lingo dan STC.

Sistem yang dihasilkan selanjutnya diuji. Proses pengujian terdiri dari dua bagian, yaitu pengujian akurasi analisis sentimen dan pengujian clustering. Pada proses pengujian akurasi dengan data uji sebanyak 4 kelompok data, dengan masing-masing berjumlah 50, 100, 150 dan 200 tweet. Hasil pengujian menunjukkan tingkat akurasi rata-rata sistem dalam menganalisis sentimen adalah sebesar 89.67%. Dengan hasil akurasi ini maka dapat dikatakan bahwa *Naive Bayes Classifier* layak untuk diimplementasikan. Sedangkan pada pengujian *clustering* didapatkan hasil bahwa dari segi waktu pemrosesan, algoritma STC adalah lebih cepat jika dibandingkan dengan algoritma Lingo dengan waktu pemrosesan rata-rata 66.2 *mili second*.

Kata Kunci: Analisis Sentimen, Clustering, Naive Bayes Classifier, STC.