

INTISARI

Penggunaan Twitter selain menjadi media untuk berkomunikasi dengan sesama, dimanfaatkan juga oleh peneliti dan pihak tertentu untuk mengetahui sentimen yang dituliskan oleh publik dalam *tweet*nya. Hal ini menyebabkan Twitter banyak digunakan sebagai objek dalam melakukan analisis sentimen. Namun, di antara banyaknya *tweet* yang disebar di Twitter, terdapat pula spam yang disebar yang dapat mempengaruhi kinerja maupun hasil dari analisis sentimen yang dilakukan. Banyak penelitian tentang analisis sentimen tidak memperhatikan adanya spam tersebut, sehingga data yang seharusnya berlabel spam dan tidak mempengaruhi sentimen terikut ke dalam proses analisis sentimen. Penelitian ini bertujuan untuk mengembangkan sebuah model yang dapat mendeteksi pengguna Twitter sebagai penyebar spam atau tidak dan untuk mencari tahu apakah ada pengaruh yang signifikan dari spam yang ada pada Twitter terhadap hasil analisis sentimen. Penelitian ini dilakukan dalam beberapa tahapan yaitu, 1) Pengembangan Model Deteksi Spam, 2) Pengambilan Data, 3) Filtering Data Spam, 4) Klasifikasi Sentimen, 5) Visualisasi Hasil.

Penelitian ini menggunakan model pengklasifikasi spam yang dapat mengklasifikasikan spam berdasarkan data *tweet* dan metadata yang dimiliki oleh pengguna Twitter. Untuk pengembangan modelnya menggunakan dataset yang sudah tersedia dan diteliti oleh peneliti sebelumnya yang berisi data *tweet* dan metadata yang sudah dilabeli sebagai spam dan tidak spam. Penelitian ini menggunakan 5000 data yang akan diteliti untuk masing-masing topik yang digunakan yaitu data topik Samsung dan Iphone. Selanjutnya dilakukan klasifikasi sentimen pada data yang sudah diolah, pengklasifikasian sentimen menggunakan *tool* VADER karena VADER cocok digunakan pada teks yang terdapat pada media sosial.

Hasil penelitian menunjukkan bahwa terdapat spam dalam jumlah yang cukup besar dalam data *tweet*, namun pengaruh spam tersebut tidak terlalu signifikan terhadap klasifikasi sentimen pada penelitian dengan menggunakan produk Iphone dan Samsung ini. Karena kelas sentimen keseluruhan data tidak berubah menjadi polaritas sentimen yang berbeda. Namun dari penelitian ini ditemukan bahwa spam yang dituliskan pada Twitter tentang produk Samsung dan Iphone kebanyakan merupakan promosi penjualan produk tertentu.

Kata Kunci: Analisis Sentimen, VADER, Twitter, Deteksi Spam, Visualisasi Kata

ABSTRACT

The use of Twitter, apart from being a media for communicating with others, is also used by researchers and certain parties to find out the sentiments written by public in their tweets. This causes Twitter to be widely used as an object in conducting sentiment analysis. However, among the many tweets that are distributed on Twitter, there is also spam that is spread which can affect the performance and results of the sentiment analysis that is done. Many studies on sentiment analysis do not take account to the existence of the spam, resulting the data which is actually labelled as spam being used in the sentiment analysis process. This study aims to develop a model that can classify a Twitter user as spammers or not and to find out the influence of spam on Twitter on the results of sentiment analysis, to find out how significant the spam to the whole sentiment analysis. This research was conducted in several stages, namely, 1) Spam Detection Model Development, 2) Data Collection, 3) Spam Data Filtering, 4) Sentiment Analysis, 5) Result Visualization.

This study uses spam classifier model that can classify spam based on tweet data and metadata owned by Twitter users. For the development of the model, this study uses datasets that are already available and researched by previous researchers which contain tweets and metadata that have been labeled as spam and not spam. This study uses 5000 data to be studied for each topic used, namely Samsung and Iphone topic data. Furthermore, sentiment classification is carried out on the processed data, the sentiment classification stage uses VADER because VADER is suitable for social media texts.

The result shows that there is a fairly large amount of spam in the tweet data, but the effect of spam is not very significant on sentiment analysis in this case study using Samsung and Iphone data. It is due to the overall sentiment class of the data does not change into different polarity of sentiment. However, from this research, it was found that spam that was written on Twitter about Samsung and Iphone was mostly a sales promotion for certain products.

Keywords: Sentiment Analysis, VADER, Twitter, Spam Detection, Word Visualization