

ABSTRAK

Keberadaan twitter telah banyak digunakan masyarakat luas untuk menulis suatu opini terhadap semua hal yang ada, salah satunya membicarakan tentang tokoh-tokoh publik tertentu. Tokoh yang dinilai merupakan tokoh yang dianggap layak untuk dipilih sebagai pemimpin suatu negara terutama di Indonesia. Oleh karena itu, penelitian ini mencoba untuk menganalisis twitter berbahasa Indonesia yang memiliki opini tentang tokoh-tokoh publik di Indonesia. Pada penelitian ini terdiri dari beberapa tahap. Tahap pertama adalah proses *pre-processing* yang terdiri dari proses *case folding*, tokenisasi, *stemming*, normalisasi kata, dan *stopword removal*. Selanjutnya pada tahap kedua adalah dengan melakukan pembobotan kata menggunakan metode TF-IDF. Pada tahap terakhir yaitu melakukan klasifikasi dengan metode klasifikasi *tweet* yang berisi tentang pihak tertentu. Metode klasifikasi yang dilakukan dalam penelitian ini adalah *K-Nearest Neighbor*. Hasil yang diperoleh dari analisis sentimen terhadap tokoh publik menggunakan algoritma *K-Nearest Neighbor* memperoleh akurasi tertinggi ketika menggunakan *20-fold* dengan jumlah tetangga nilai $k=1$ sebesar 68,3%.

Kata kunci : klasifikasi, tf-idf, *K-Nearest Neighbor*.

ABSTRACT

The existence of Twitter has been widely used by the public to write an opinion on all things that exist, one of which is talking about certain public figures. Those who are being talked are public figures who are considered worthy to be elected as the leader of a country especially in Indonesia. Therefore, this research attempts to analyze Indonesian-language Twitter which has opinions on public figures in Indonesia. This research consists of several stages. The first stage is pre-processing process consisting of case folding, *tokenisasi*, stemming, *normalisasi kata*, and stopword removal. The second stage is word weighting using TF-IDF method. Finally, the last step is doing the classification using *Tweet* classification that contains certain parties. The classification method used in this research is K-Nearest Neighbor. The sentiment analysis toward public figures using K-Nearest Neighbor algorithm results the highest accuracy when using 20-fold with the number of neighboring $k=1$ at 68,3%.

Key words: classification, tf-idf, *K-Nearest Neighbor*.