

ABSTRAK

Pengelompokan konflik sosial menggunakan teknik *text mining* dilakukan untuk mengetahui jenis-jenis konflik. Pada penelitian ini pengolahan data dilakukan menggunakan bahasa *R* dengan aplikasi *RStudio*. Data didapatkan dari penambangan komentar masyarakat tentang konflik sosial di media sosial *twitter* dan menghasilkan 3952 data. Setelah data didapatkan lalu diolah menggunakan metode *text cleaning* untuk menghapus angka, simbol dan alamat *url*. Berikutnya melakukan *text processing* dengan langkah-langkah *stemming*, *tokenizing*, *stopword removal* dan *TF-IDF*. Metode *TF-IDF* mendapatkan hasil “konflik agraria” menjadi topik konflik sosial tertinggi dari penambangan data *twitter*. Penelitian ini menggunakan metode *unsupervise* dengan algoritma *K-Means*. Pertama yaitu menentukan jumlah *cluster* dengan menggunakan metode *elbow plot* dan menghasilkan sebanyak 6 *cluster*. Kedua, menghitung jarak tiap *cluster* dan mendapatkan nilai 0,566. Terakhir yaitu memasukkan seluruh anggota terpilih kesetiap *cluster* yang tersedia. Evaluasi *K-Means* dilakukan menggunakan metode *Sum of Square Error* dan mendapatkan nilai akurasi sebesar 97,2%. Jarak antar *cluster* dievaluasi menggunakan metode *sillhouette coefficient* dan mendapatkan akurasi sebesar 96,7%. Kesimpulan pada penelitian ini adalah algoritma *K-Means* berhasil membuat pengelompokan topik berdasarkan jarak keterkaitan antar kata-kata.

Kata Kunci : *Clustering, K-Means, Konflik Sosial, Text Mining, Twitter.*

ABSTRACT

Grouping of social conflicts using text mining techniques is done to find out the types of conflicts. In this study, data processing was carried out using the R language with the RStudio application. The data was obtained from mining community comments about social conflicts on social media twitter and produced 3952 data. After the data is obtained, it is processed using the text cleaning method to remove numbers, symbols and url addresses. Next, do text processing with the steps of stemming, tokenizing, stopwords removal and TF-IDF. The TF-IDF method results in “agrarian conflict” being the highest social conflict topic from twitter data mining. This study uses the unsupervised method with the K-Means algorithm. The first is to determine the number of clusters using the elbow plot method and produce as many as 6 clusters. Second, calculate the distance of each cluster and get a value of 0.566. The last is to enter all selected members into each available cluster. K-Means evaluation was carried out using the Sum of Square Error method and obtained an accuracy value of 97.2%. The distance between clusters was evaluated using the silhouette coefficient method and obtained an accuracy of 96.7%. The conclusion of this study is that the K-Means algorithm succeeded in making topic groupings based on the distance between words.

Keyword : *Clustering, K-Means, Social Conflict, Text Mining, Twitter.*