

## ABSTRAK

Kemajuan teknologi interaksi sosial tidak sebaik perilaku sosial masyarakat dalam bersosial media. Flaming adalah salah satu jenis dari cyberbullying yang sering dijumpai di platform media sosial manapun. Menurut Satalina (2014), flaming merupakan pesan teks yang isinya merupakan kata-kata yang penuh amarah dan frontal (Willard, 2005). Salah satu media sosial yang dengan mudah menemukan umpatan terhadap orang lain adalah Twitter. Pada umumnya orang Indonesia senang mengumpat dengan menggunakan nama jenis hewan sebagai bentuk ekspresinya.

Topik tersebut menjadi sumber data dan dasar dalam penelitian ini. Klasifikasi adalah metode yang dapat mengelompokkan tweet berdasarkan maknanya. Algoritma klasifikasi yang cukup populer salah satunya adalah K-Nearest Neighbor. Ada pula klasifikasi yang mengandalkan kamus sentimen seperti Leksikon. Penelitian ini mencoba mengukur metode mana yang memiliki akurasi yang lebih baik antara KNN dan pendekatan kamus Sentimen Leksikon Indonesia menggunakan Rule Based Method dalam mengklasifikasi antara tweet flaming dan netral yang mengandung kata umpatan.

Penelitian ini dilakukan dengan jumlah data sebanyak 100 tweet, dimana semua tweet-nya mengandung kata umpatan 'anjing'. Kedua metode ini menggunakan masukan yang sama yaitu berupa teks. Pada tahap KNN terdapat beberapa proses seperti pre-processing, pembobotan TF-IDF, K-Fold Cross Validation, dan mengukur jarak tetangga. Akurasi kedua metode ini diukur menggunakan Confusion Matrix dengan berdasarkan klasifikasi manual oleh Bapak A. Danang Satria Nugraha, M.A sebagai triangulator.

Dari hasil pengujian yang telah dilakukan, algoritma K-Nearest Neighbor menghasilkan akurasi tertinggi sebesar 73.0% dengan menggunakan 10 fold saat jumlah tetangga terdekat berjumlah  $K = 1$ . Sedangkan pada pendekatan kamus Sentimen Leksikon Indonesia dengan menggunakan Rule Based Method memperoleh hasil akurasi sebesar 64%.

Kata Kunci : Tweet, Klasifikasi, K-Nearest Neighbor, Leksikon, Rule Based Method

## ABSTRACT

The advancement of social interaction technology is not as good as social behavior of society in social media. Flaming is one of those types of cyber bullying that is often found on any social media platform. According to Satalina (2014), flaming is a text message whose contents are anger-filled and frontal words (Willard, 2005). One of the social media that easily finds a swearing of others is Twitter. In general, Indonesians love to swear by using animal names as their expressions.

The topic is a data source and the basis for this study. Classification is a method that can group tweets by their meaning. The classification algorithm is quite popular one of which is K-Nearest Neighbor. There are also classifications that rely on sentiment dictionaries like lexicon. This study tried to gauge which method has better accuracy between KNN and the dictionary approach Indonesian Lexicon sentiment uses the Rule Based Method of classifying the flaming and neutral tweets containing the word swear.

This research is done with a total of 100 tweets, of which all of the tweets contain the word ' dog '. Both methods use the same input as text. At the KNN stage there are several processes such as pre-processing, TF-IDF-weighted, K-Fold Cross Validation, and distance-measuring neighbors. The accuracy of these two methods is measured using the Confusion Matrix based on manual classification by Mr. A. Danang Satria Nugraha, M.A as the triangulator.

From the test results, the K-Nearest Neighbor algorithm resulted in the highest accuracy of 73.0% by using 10 fold when the nearest number of neighbors was  $K = 1$ . Meanwhile, in the dictionary approach of Indonesian lexicon by using Rule Based Method, the accuracy of 64% is achieved.

Keyword : Tweet, classification, K-Nearest Neighbor, Lexicon, Rule Based Method