

ABSTRAK

Analisis sentimen merupakan riset komputasi untuk mencari atau mengidentifikasi sikap, pendapat, emosi, evaluasi atau pemikiran dalam sesuatu bacaan. Analisis sentiment ialah proses text analytics memperoleh beragam data internet dan platform dari media sosial. UMKM (Usaha Mikro, Kecil, dan Menengah) ialah sektor usaha dengan potensial dan bertanggung jawab untuk meningkatkan kesejahteraan Masyarakat. *K-Nearest Neighbor* (KNN) Algoritma KNN bekerja mencari K titik data terdekat dari titik data yang baru, kemudian mengklasifikasi kelas atau nilai dari tetangga-tetangganya. Namun, KNN memiliki kelemahan karena sangat dipengaruhi oleh nilai K dan jarak yang menghasilkan akurasi yang kurang optimal. Sehingga perlu penambahan bobot. Algoritma *Weighted K-Nearest Neighbours* (WKNN) meningkatkan akurasi klasifikasi dengan menambah bobot ke setiap titik data tetangga berdasarkan jaraknya dari item yang dikategorikan, sehingga mengoptimalkan proses klasifikasi. Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode yang digunakan mengonversi data teks ke data numerik dengan tujuan untuk memberikan bobot pada setiap kata atau fitur. Algoritma feature selection untuk memilih fitur penting di dokumen teks antara lain yaitu *information gain*. *Information Gain* adalah metode untuk mengukur perubahan nilai entropi suatu dataset disegmentasi dari fitur. Augmentasi data menghitung informasi relevan yang diberikan suatu fungsi tentang kelas atau label suatu data.

Kata kunci : Analisis Sentimen, UMKM, Rumah Makan Bebek Rizki, *K-Nearest Neighbor* (KNN), *Weighted K-Nearest Neighbours* (WKNN), *Term Frequency-Inverse Document Frequency* (TF-IDF), *Information Gain*

ABSTRACT

Sentiment Analysis is a computational research method aimed at identifying attitudes, opinions, emotions, evaluations, or thoughts within a text. It involves the process of text analytics to extract various data from the internet and social media platforms. Micro, Small, and Medium Enterprises (MSMEs) are a business sector with great potential and responsibility for improving community welfare. K-Nearest Neighbor (KNN) is an algorithm that identifies the K nearest data points to a new data point and classifies the class or value based on its neighbors. However, KNN has weaknesses as it is highly influenced by the value of K and the distance metric, which may lead to suboptimal accuracy. Therefore, weighting is necessary. The Weighted K-Nearest Neighbors (WKNN) algorithm enhances classification accuracy by assigning weights to each neighboring data point based on its distance from the classified item, thus optimizing the classification process. Term Frequency-Inverse Document Frequency (TF-IDF) is a method used to convert text data into numerical data, aiming to assign weights to each word or feature. Feature selection algorithms, such as Information Gain, are used to select important features in text documents. Information Gain measures the change in entropy values of a dataset segmented by features. Data augmentation calculates the relevant information provided by a function about the class or label of the data.

Keywords: Sentiment Analysis, MSMEs, Bebek Rizki Restaurant, K-Nearest Neighbor (KNN), Weighted K-Nearest Neighbors (WKNN), Term Frequency-Inverse Document Frequency (TF-IDF), Information Gain