

ABSTRAK

Fadhilah Az-zahra, 2020, Analisis Sentimen Pada Twitter Shopee Indonesia Menggunakan *K-Nearest Neighbor* dengan Pembobot *Term frequency-inverse document frequency*, Skripsi, Program Studi Statistika, Universitas Muhammadiyah Semarang. Pembimbing: I.Dr Rochdi Wasono, M,Si., II. Tiani Wahyu Utami, M Si

Twitter termasuk media sosial yang diminati oleh banyak masyarakat dan menjelma menjadi penyiar informasi yang amat cepat saat ini. Informasi yang dikeluarkan serta bersirkulasi via media ini amat bebas serta banyak ragam, layaknya berita, pendapat, pertanyaan, kritikan, komentar baik bersifat positif ataupun negatif. Klasifikasi ialah kaidah di teks mining yang menghimpunkan muatan mengacu pada kesamaan skripnya. Lewat kaidah ini membolehkan tweets yang tersedia di Twitter digolongkan jadi satu bersandarkan jenisnya. Misalkan, substansi sepakbola, voli, serta tenis diguguskan pada kategori olahraga. Prosedur pada klasifikasi dimulai memakai preprocessing, selanjutnya dilakukan pembobotan kata, lalu kategorisasi yang terdiri dari Preprocessing dan beberapa tahapan, yaitu pembersihan dokumen, tokenizing, stopword removal, serta stemming. Metode pembobotan kata yang dipergunakan dalam skripsi ini ialah Term Frequency–Inverse Document Frequency (TF-IDF) & memakai K-Nearest Neighbor (K-NN) dengan data latih 80:20. Metode KNN merupakan klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Pengujian akurasi dari klasifikasi tweets pada Twitter dengan memakai metode K-Nearest Neighbor (K-NN) menghasilkan akurasi dimana total data berjumlah 1300, dengan angka k yang dimasukkan ialah 1, 3, 5, serta 7 dan 9 masing-masing hasilnya k = 1, akurasi sebesar 77,31%; k = 3, ketepatan 75,38%; k = 5, ketepatan 75,38%; k = 7, ketepatan 75,38% dan k=9 ketepatan 75,38%.

Kata Kunci : :Preprocessing, K- Nearest Neighbor, Term frequency-inverse document frequency