

KLASIFIKASI TEKS PADA TWITTER MENGGUNAKAN METODE *K-NEAREST NEIGHBORD* DAN SELEKSI FITUR *INFORMATION GAIN*

(Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua
DPR Ibu Puan Maharani di *Twitter*)

Alfa Yunitra¹, Agung Nilogiri², Deni Arifianto³

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email: ¹Alfayunitra12@gmail.com

ABSTRAK

Indonesia adalah salah satu negara dengan pemakai internet yang paling cepat berkembang dengan sekitar 196 juta penduduk atau 73,3% dari populasi terjangkau internet. Perkembangan teknologi pada tahun 2017 mencapai 3,4 miliar atau 45% dari total populasi di dunia, dan pada tahun 2022 perkembangan teknologi juga diproyeksikan akan menyentuh 4,8 miliar atau sekitar 60% dari populasi di dunia. Oleh karena itu dalam penelitian ini akan menentukan perhitungan presisi dan akurasi tingkat terbaik untuk melakukan sebuah klasifikasi komentar pengguna di twitter ke dalam kelas sentimen positif juga, sentiment negatif. Metode yang digunakan yaitu *K-Nearest Neighbord* dan seleksi fitur *Information Gain*. Tujuannya untuk menganalisis tingkat perbandingan akurasi dan performa presisi *Information Gain* dan *K-Nearest Neighbor* dalam mengelompokkan komentar pengguna twitter ke dalam kelas sentimen positif dan sentiment negatif. Setelah melakukan proses perhitungan dengan metode *K-Nearest Neighbor* dan seleksi fitur *Information Gain* dengan uji coba beberapa fold antara lain 3,5, dan 9 dimana nilai k pada knn 3,5,9. Dan telah didapatkan hasil akurasi terbaik tanpa fitur 61.6% pada fold 5 dengan nilai $k=3$. Sedangkan hasil akurasi terbaik dengan menggunakan fitur *selection* sebesar 71.3% pada fold 9 dengan nilai $k=9$.

Kata kunci: *Klasifikasi, K-nearest neighbord, Information Gain.*

**TEXT CLASSIFICATION ON TWITTER USING THE K-
NEAREST NEIGHBORD METHOD AND INFORMATION GAIN
FEATURE SELECTION**

**(Case Study: Public Opinion Response to the Chair of the DPR,
Mrs. Puan Maharani, on Twitter)**

Alfa Yunitra¹, Agung Nilogiri², Deni Arifianto³

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email: 1Alfayunitra12@gmail.com

ABSTRACT

Indonesia is one of the countries with the fastest growing internet users with around 196 million residents or 73.3% of the population with internet access. Technological developments in 2017 reached 3.4 billion or 45% of the world's total population, and in 2022 technological developments are also projected to reach 4.8 billion or around 60% of the world's population. Therefore, this research will determine the best level of precision and accuracy calculations for classifying user comments on Twitter into positive sentiment classes as well as negative sentiment. The method used is K-Nearest Neighbor and Information Gain feature selection. The aim is to analyze the comparative level of accuracy and precision performance of Information Gain and K-Nearest Neighbor in grouping Twitter user comments into positive sentiment and negative sentiment classes. After carrying out the calculation process using the K-Nearest Neighbor method and selecting the Information Gain feature by testing several folds, including 3, 5, and 9, where the k value in knn is 3, 5, 9. And the best accuracy results without features were obtained, 61.6% on fold 5 with a value of k=3. Meanwhile, the best accuracy results using the selection feature were 71.3% on fold 9 with a value of k=9.

Keywords: *Classification, K-nearest neighbor, Information Gain.*