

PERBANDINGAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE PADA KASUS ANALISIS SENTIMEN TERHADAP DATA VAKSIN COVID-19 DI TWITTER

oleh:

Rizki Anom Raharjo
NIM 1929101020

ABSTRAK

Organisasi Kesehatan Dunia (WHO) secara resmi menyebut virus Covid-19 sebagai pandemi global, oleh karena itu semua negara di dunia berusaha meminimalkan dampak yang ditimbulkan oleh virus tersebut, yaitu dengan mengembangkan vaksin. Dalam konteks pandemi Covid-19, pemerintah Indonesia juga meminta dan mendorong masyarakat untuk turut serta mendukung vaksinasi, namun upaya tersebut sebenarnya memiliki kelebihan dan kekurangan, sehingga banyak masyarakat yang mengutarakan pendapatnya di jejaring sosial salah satunya Twitter. Penelitian ini bertujuan untuk mengetahui hasil penerapan analisis sentimen dan mengukur performansi algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) terhadap data vaksin Covid-19 dengan cara mengklasifikasikan data tersebut ke dalam kelas positif dan negatif. Data tweet yang didapat kemudian dilakukan *text preprocessing* untuk mengoptimalkan pengolahan data. Terdapat 4 tahapan *text preprocessing* antara lain *Case Folding*, *Tokenizing*, *Filtering*, dan *Stemming*. Penelitian ini mengkaji kinerja *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) dengan menambahkan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) yang bertujuan untuk memberikan bobot pada hubungan kata (*term*) sebuah dokumen. Kemudian melakukan *splitting* data yaitu membagi data *training* 80% dan data *testing* 20% dengan harapan mendapatkan model dengan performansi terbaik dan yang terakhir melakukan visualisasi data tweet dengan menggunakan *Word Cloud* agar bisa menarik sebuah kesimpulan. Hasil klasifikasi data tweet vaksin Covid-19 menggunakan algoritma *Naïve Bayes Classifier* mendapatkan nilai *accuracy* sebesar 81%, *precision* sebesar 80%, *recall* sebesar 99%, dan *f1-score* sebesar 89%, Sedangkan untuk algoritma *Support Vector Machine* mendapatkan nilai *accuracy* sebesar 87%, *precision* sebesar 88%, *recall* sebesar 96%, dan *f1-score* sebesar 92%.

Kata Kunci: *Perbandingan Metode Klasifikasi, Analisis Sentimen, Vaksin Covid-19, Twitter, Naïve Bayes Classifier, Support Vector Machine.*

COMPARISON OF NAÏVE BAYES CLASSIFIER METHODS AND SUPPORT VECTOR MACHINE IN CASE OF SENTIMENT ANALYSIS OF COVID-19 VACCINE DATA ON TWITTER

Rizki Anom Raharjo
NIM 1929101020

ABSTRACT

The Covid-19 virus has been officially designated as a global pandemic by the World Health Organization (WHO), so that all countries in the world are trying to minimize the impact caused by the virus, namely by developing vaccines. The Indonesian government also seeks and urges the public to participate in supporting vaccination in the context of the Covid-19 pandemic, but these efforts actually lead to pros and cons so that many people give their opinion on social media, one of which is twitter. This study aims to determine the results of applying sentiment analysis and measuring the performance of the Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) algorithms on Covid-19 vaccine data by classifying the data into positive and negative classes. The tweet data obtained is then done by text preprocessing to optimize data processing. There are 4 stages of text preprocessing, including Case Folding, Tokenizing, Filtering, and Stemming. This study examines the performance of Nave Bayes Classifier (NBC) and Support Vector Machine (SVM) by adding the TF-IDF (Term Frequency-Inverse Document Frequency) technique which aims to give weight to the word relationship (term) of a document. Then do data splitting, namely dividing 80% training data and 20% testing data in the hope of getting a model with the best performance and finally visualizing tweet data using Word Cloud in order to draw a conclusion. The results of the Covid-19 vaccine tweet data classification using the Naïve Bayes Classifier algorithm get an accuracy value of 81%, precision 80%, recall 99%, and f1-score 89%, while the Support Vector Machine algorithm gets an accuracy value of 87%, 88% precision, 96% recall, and 92% f1-score.

Keywords: Comparison of Classification Methods, Sentiment Analysis, Covid-19 Vaccine, Twitter, Naïve Bayes Classifier, Support Vector Machine