

ABSTRAK

Wendy Ariyani, Putu (2021), *Analisis Sentimen Masyarakat terhadap Virus Corona Berdasarkan Opini dari Twitter menggunakan Metode Naïve Bayes dan K-Nearest Neighbor*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I : Dr. I Made Gede Sunarya, S.Kom., M.Cs. dan Pembimbing II : Dr. I Gede Aris Gunadi, S.Si. M.Kom.

Kata Kunci : Virus Corona, COVID-19, Twitter, Analisis Sentimen, *Naïve Bayes*, K-Nearest Neighbor

Sejak penyebaran Virus Corona, banyak masyarakat yang mengutarakan pendapatnya melalui media social Twitter dalam menanggapi penyebaran Virus Corona. Berbagai opini yang diutarakan masyarakat dapat menjadi sebuah acuan untuk mengetahui Sentimen Masyarakat terhadap Virus Corona. Diperlukan analisis sentiment untuk mengetahui sentiment opini yang muncul pada social media. Tujuan penelitian ini adalah untuk mengetahui implementasi dan tingkat akurasi dari metode *Naïve Bayes* dan K-Nearest Neighbor. Data yang digunakan diambil dari twitter mengenai Virus Corona. Jumlah data yang digunakan sebanyak 2000 data tweet. Dokumen dengan sentiment positif yaitu 1320 dan dokumen dengan sentiment negative yaitu 689. Data yang terkumpul akan dibagi untuk digunakan sebagai Data Latih dan Testing untuk proses klasifikasi. Implementasi *Naïve Bayes* dan KNN dalam analisis sentimen masyarakat terhadap virus corona, dimulai dari tahap preprocessing data yang terdiri Normalisasi data (menghilangkan URL dan username), Case Folding (mengubah semua huruf menjadi huruf kecil), Tokenizing (memilih menjadi beberapa kata), Stopword Removal (menghilang kata yang sering muncul), serta stemming (mengubah sebuah kata menjadi bentuk umumnya). Setelah melalui tahap preprocessing, kemudian dilakukan tahap pembobotan TF-IDF. Hasil dari pembobotan TF-IDF akan diklasifikasi menggunakan metode *Naïve Bayes* dan KNN, sehingga nanti akan mendapatkan hasil perbandingan klasifikasi dari kedua metode tersebut. Hasil klasifikasi menggunakan metode *Naïve Bayes* diperoleh akurasi sebesar 0.83 dan error rate sebesar 0.17. Sedangkan untuk hasil klasifikasi menggunakan metode KNN diperoleh akurasi sebesar 0.78 dan error rate sebesar 0.21. Perbandingan performansi metode *Naïve Bayes* dan KNN menunjukkan bahwa *Naïve Bayes* lebih baik dalam mengklasifikasikan data Covid-19.

ABSTRACT

Wendy Ariyani, Putu (2021), *Analysis of Public Sentiment against Corona Virus Based on Opinion from Twitter Using the Naïve Bayes and K-Nearest Neighbor Methods*. Thesis, Computer Science, Post Graduate Study Program, Ganesha University of Education.

This thesis has been supervised and approved by Supervisor I : Dr. I Made Gede Sunarya, S.Kom., M.Cs. and Supervisor II : Dr. I Gede Aris Gunadi, S.Si. M.Kom.

Keywords: Corona Virus, COVID-19, Twitter, Sentiment Analysis, *Naïve Bayes* , K-Nearest Neighbor

Since the spread of the coronavirus, many people have expressed their opinions through social media Twitter in response to the spreading of the Corona Virus. The various opinions expressed by the public can be a reference to knowing the public's feelings about the coronavirus. A sentiment analysis is needed to find out the sentimental opinions that emerge on social media. The aim of this study was to know the implementation and level of accuracy of the Naïve Bayes and K-Nearest Neighbor methods. The data used was taken from Twitter about the coronavirus. The amount of data used is as much as 2000 data tweets. A document with positive sentiment is 1320, and a document with negative sentiment is 689. The data collected will be divided for use as training and testing data for the classification process. The implementation of Naïve Bayes and KNN in the analysis of public sentiment against the coronavirus, starting from the stage of data preprocessing consisting of Data Normalization (removing URLs and usernames), Case Folding (transforming all letters into small letters), Tokenizing (choosing into a few words), Stopword Removal (deleting words that appear frequently), and stemming (mengubah sebuah kata menjadi bentuk umumnya). After the pre-processing stage, then the TF-IDF weighing stage is carried out. The results of TF-IDF weighing will be classified using the Naïve Bayes and KNN methods, so that the results of the classification comparison of the two methods will be obtained later. The results of the classification using Naïve Bayes method obtained an accuracy of 0.83 and an error rate of 0.17. For classification results using the KNN method, an accuracy of 0.78 and an error rate of 0.21 is obtained. A comparison of the performance of Naïve Bayes and KNN methods suggests that Naiv Bayes is better at classifying Covid-19 data.