

KLASIFIKASI DATA KOMENTAR COVID 19 PADA TWITTER MENGUNAKAN METODE KLASIFIKASI NAÏVE BAYES

NAUFALDY PURWANTO

ABSTRAK

Banyak sekali pengguna *Twitter* yang mencurahkan pendapat melalui *tweet* yang mereka kirimkan pada media social *Twitter*. Salah satu *Tweet* yang menjadi perbincangan pada tahun 2019 silam yaitu mengenai Pandemi *Covid19*. Banyak sekali *tweet* yang membahas tentang Pandemi *Covid19* tersebut. Mereka selalu membahas tentang bagaimana dampak positif dan negatif dari Pandemi *Covid19*. Dari permasalahan tersebut, dilakukanlah penelitian mengenai analisis sentiment pada data *Twitter* yang berkaitan dengan Pandemi *Covid19* dan menggunakan metode *Naïve Bayes*. Data yang diambil yaitu data dari *tweet* menggunakan library *tweet-harvest* yang sudah terintegrasi dengan *Twitter API*. Sebelum melakukan proses klasifikasi data, data tersebut akan dilakukan pemberian label yaitu label positif dan label negative secara manual. Setelah itu, dilakukan praproses dengan tahapan *data cleaning*, *tokenization*, *case folding*, *stopword removal*, *stemming*. Kemudian melakukan pembobotan kata menggunakan *CountVectorizer*. Tahap selanjutnya dilakukan perbandingan hasil pembagian data 80% data latih dan 20% data uji dengan 70% data latih dan 30% data uji. Dengan mengklasifikasikan menggunakan metode *Naïve Bayes*. Hasilnya pembagian data dengan 70% data latih dan 30% data uji menghasilkan tingkat akurasi tertinggi yaitu 91%.

Kata Kunci : Analisis Sentimen, Klasifikasi, Pandemi *Covid19*, *Naïve Bayes Classification*, *Twitter*

CLASSIFICATION OF COVID 19 COMMENT DATA ON TWITTER USING THE NAÏVE BAYES CLASSIFICATION METHOD

NAUFALDY PURWANTO

ABSTRACT

A lot of Twitter users express their opinions through the tweets they send on Twitter social media. One of the tweets that became a conversation in 2019 was about the Covid19 Pandemic. Lots of tweets discussing the Covid19 Pandemic. They always discuss how the positive and negative impacts of the Covid19 Pandemic. From these problems, research was conducted on sentiment analysis on Twitter data related to the Covid19 Pandemic and using the Naïve Bayes method. The data taken is data from tweets using the tweet-harvest library that has been integrated with the Twitter API. Before carrying out the data classification process, the data will be labeled, namely positive labels and negative labels manually. After that, preprocessing is carried out with the stages of data cleaning, tokenization, case folding, stopword removal, stemming. Then do word weighting using CountVectorizer. The next stage is to compare the results of data division of 80% training data and 20% test data with 70% training data and 30% test data. By classifying using the Naïve Bayes method. The results of data sharing with 70% training data and 30% test data resulted in the highest accuracy rate 91%.

Keywords : *Sentiment Analysis, Classification, Covid19 Pandemic, Naïve Bayes Classification, Twitter*