

ABSTRAK

Twitter merupakan salah satu sosial media yang banyak digunakan terutama di Indonesia. Twitter dapat diakses melalui website maupun aplikasi yang tersedia di *app store* dan *play store*. Untuk dapat menggunakan fasilitas kita harus memiliki akun dan login ke akun tersebut sehingga bisa menfollow akun lain, mengirimkan tweet, meretweet tweet orang, membalas tweet, mengirimkan pesan ke akun lain. Twitter juga menampilkan topik yang sedang banyak di perbincangkan yang disebut Trending Topik yang tersedia untuk berbagai wilayah di dunia. Dengan fasilitas yang beragam banyak orang bermain twitter untuk sekedar dijadikan hiburan, mencari informasi berita terbaru bahkan menyampaikan pendapat mengenai berbagai hal.

Penelitian ini dilakukan untuk mengklasifikasikan tiap tweet yang mengandung kata 'corona' sehingga diketahui apakah tweet tersebut bersifat positif, negatif atau netral. Tweet yang digunakan sejumlah 982 tweet yang diambil secara random dengan kata kunci corona menggunakan *twitter API*.

Metode yang digunakan adalah algoritme *naïve bayes classfier*, tweet yang berbentuk teks akan di *preprossesing* sehingga masuk ke perhitungan *TF-IDF*, dari bobot tiap kata akan dihitung probabilitasnya. Sebelum menghitung probabilitas data dibagi menggunakan *Kfold Cross Validation*. Setelah itu diklasifikasikan berdasarkan probabilitas tertinggi.

Penggunaan metode *naïve bayes classfier* pada penelitian ini menghasilkan akurasi tertinggi pada 9-fold yaitu sebesar 90,825%

Kata kunci : Algoritme *naïve bayes classfier*, tweet, twitter, *TF-IDF*, *twitter API*, *kfold cross validation*, corona.

ABSTRACT

Twitter is one of the most widely used social media, especially in Indonesia. Twitter can be accessed through the website or application available in the app store and play store. To be able to use the facility we must have an account and log into that account so that we can follow other accounts, send tweets, retweet people's tweets, reply to tweets, send messages to other accounts. Twitter also features a topic that is being discussed a lot called Trending Topics which are available for various regions of the world. With various facilities, many people play Twitter for entertainment, looking for the latest news information and even expressing opinions on various things.

This research was conducted to classify each tweet containing the word 'corona' so that it is known whether the tweet is positive, negative or neutral. The number of tweets used was 982 tweets that were taken randomly with the keyword corona using the twitter API.

The method used is the naïve Bayes classifier algorithm, tweets in the form of text will be preprocessed so that they enter the TF-IDF calculation, from the weight of each word the probability is calculated. Before calculating the probability the data is divided using Kfold Cross Validation. After that it is classified based on the highest probability.

The use of the naïve Bayes classifier method in this study resulted in the highest accuracy at 9-fold, amounting to 90.825%

Keywords: Naïve Bayes classifier algorithm, tweet, twitter, TF-IDF, twitter API, kfold cross validation, corona.