

ABSTRAK

Nama : Arnesha Febra Syavera
Program Studi : Teknik Informatika
Judul : Klasifikasi Spam Tweet Covid-19 Pada Twitter
Menggunakan Metode Naïve Bayes
Dosen Pembimbing : Muhamad Soleh, M.Kom

Penelitian ini membahas bagaimana penerapan Naïve Bayes bisa mengklasifikasi tweet ke dalam dua kategori yaitu *spam* dan *non spam* berdasarkan *hashtag* dan *tweet* yang bisa diklasifikasi berbahasa Indonesia dan bagaimana agar dapat meningkatkan perhitungan akurasi dalam sebuah sistem klasifikasi *spam* dan *non spam* pada *twitter* menggunakan fitur *trending topics* dan *hashtag*. Dalam menghadapi permasalahan tersebut, maka dari itu pada tugas akhir telah membangun suatu sistem untuk mengklasifikasi *spam* dan *non spam* dengan metode Naïve Bayes. Dalam mengimplementasikan sistem ini, tahapan yang dilakukan meliputi pengumpulan data menggunakan *twitter API*, pembuatan *dataset*, pemberian label secara manual, pengolahan data, pemilihan model, pengujian model, perhitungan akurasi dan validasi. Menggunakan metode Naïve Bayes yang memiliki kinerja baik untuk pengklasifikasian data dokumen yang mengandung angka maupun teks dan kemampuannya mengklasifikasikan dokumen dengan sederhana dan kecepatan komputasinya yang tinggi. Dalam penelitian ini *dataset* akan melewati proses *training* dan *testing*. Hasil dari pengujian (*testing*) sistem klasifikasi ini menunjukkan bahwa metode Naïve Bayes tingkat akurasi yang didapat sebesar 98.84% untuk data latih (*training*) dan 75.14% untuk data uji (*testing*). Tingkat akurasi ini didapat dari hasil pengujian terhadap 615 data *tweet*. Dan Algoritma Naïve Bayes dapat memprediksi tweet ke dalam kelas *spam* dan *non spam* di dalam proses validasi menggunakan 433 data *tweet*.

Kata Kunci: Covid19, Twitter, tweet, spam, Naïve Bayes.

ABSTRACT

This study discusses how the application of Naïve Bayes can classify tweets into two categories, namely spam and non-spam based on hashtags and tweets that can be classified in Indonesian and how to improve accuracy calculations in a spam and non-spam classification system on Twitter using the trending topics and hashtags. In dealing with these problems, therefore, in this final project, we have built a system to classify spam and non-spam with the Naïve Bayes method. In implementing this system, the steps involved include data collection using the Twitter API, dataset creation, manual labeling, data processing, model selection, model testing, accuracy calculations and validation. Using the Naïve Bayes method which has good performance for classifying document data containing numbers and text and its ability to classify documents with simplicity and high computational speed. In this research, the dataset will go through a training and testing process. The results of testing this classification system show that the Naïve Bayes method has an accuracy rate of 98.84% for training data and 75.14% for testing data. This level of accuracy is obtained from the test results of 615 tweet data. And the Naïve Bayes algorithm can predict tweets into spam and non-spam classes in the validation process using 433 tweet data.

Keywords: Covid19, Twitter, tweets, spam, Naïve Bayes.