



## INTISARI

### DETEKSI KEASLIAN PENGGUNA MENGGUNAKAN TEKS DI TWITTER

Oleh

Rahmita Iftar Rizma  
12/334616/PA/14849

Meningkatnya jumlah pengguna pada media sosial, menimbulkan beberapa masalah atau isu baru. Salah satunya berkaitan dengan maraknya akun-akun palsu. Orang-orang dari dunia media, hiburan, dan politik biasanya menjadi korban dari akun-akun palsu tersebut.

Maka dalam penelitian kali ini akan dilakukan identifikasi terhadap teks dari media sosial berupa *tweet* untuk menentukan keasliannya. Dari *tweet* tersebut, kemudian akan diambil beberapa fitur seperti jumlah *hashtag*, jumlah *mention*, jumlah alamat *web* yang dicantumkan, jumlah masing-masing kelas kata, jumlah entitas bernama, jumlah abreviasi, dan jumlah akronim untuk selanjutnya diklasifikasi. Proses klasifikasi dilakukan dengan menggunakan metode *Support Vector Machine (SVM)* dengan kernel linear, RBF dan polinomial. Kemudian untuk pengujiannya digunakan metode *K-Fold Cross Validation*, dengan mengambil nilai *k* sebesar 10. Percobaan klasifikasi dilakukan pada dua jenis *dataset*, yaitu *dataset* 2 kelas dan *dataset* 12 kelas.

Hasil dari penelitian ini adalah model klasifikasi *dataset* 2 kelas dapat menentukan yang mana akun tokoh publik asli dan akun yang hanya meniru akun asli berdasarkan *tweet* yang diproses, dimana tingkat akurasi rata-rata tertinggi sebesar 86% dengan menggunakan kernel RBF. Sedangkan untuk hasil klasifikasi *dataset* 12 kelas, model klasifikasi dapat membedakan masing-masing akun tokoh publik dan bukan tokoh publik dengan tingkat akurasi tertinggi sebesar 46% menggunakan kernel RBF.

Kata kunci : *text mining*, klasifikasi, *author identification*, *Support Vector Machine*, *tweet*, Twitter



## **ABSTRACT**

### **AUTHENTICATION DETECTION OF USER USING TEXT IN TWITTER**

By

Rahmita Iftar Rizma  
12/334616/PA/14849

The increase in the number of users on social media has raised several new problems or issues. One of them is related to the increase in the number of fake accounts. Celebrities in the media, entertainment, and political domains are regular victims of these fake accounts. In this research, the identification of text from social media will be carried out in the form of tweets to determine its authenticity.

Several features will be taken from each tweet, such as the number of hashtags, number of mentions, number of web addresses attached, number of each part-of-speech, number of named entities, number of abbreviations, and number of acronyms. The classification process is carried out using the Support Vector Machine (SVM) method. For the kernel function, this research tried linear, RBF, and polynomial kernel. Then for the evaluation used the K-Fold Cross Validation method. The value of k is 10. This research did experiment using two types of datasets, which is 2-class dataset and 12-class dataset.

As a result, the classification model of 2-class dataset can distinguish between real account and fake account of public figure from the tweet with the highest average of the accuracy is 86% using RBF kernel. And the 12-class dataset experiment can distinguish between user account from the tweet posted with highest accuracy of 46% using RBF kernel.

**Keywords** : text mining, classification, author identification, Support Vector Machine, tweet, Twitter