

ABSTRAK

Instagram merupakan media sosial yang memiliki jumlah pengguna aktif tiap bulannya mencapai dua juta pengguna. Fitur yang ditawarkan oleh Instagram tidak hanya dapat memberikan dampak positif namun dapat dimanfaatkan oleh segelintir orang untuk kepentingan pribadi yang biasanya dilakukan oleh akun palsu. Banyaknya pengguna Instagram berbanding lurus dengan jumlah akun palsu yang beredar dimana akun palsu ini memiliki dampak negatif terhadap pengguna lainnya diantaranya penipuan, *cyberbullying*, *spamming*, serta penyebaran berita dan informasi palsu. Untuk menjaga kepentingan pengguna lain serta mengurangi dampak negatif yang diberikan dari banyaknya akun palsu, penting untuk dilakukan klasifikasi akun palsu menggunakan sistem kecerdasan mesin.

Terdapat beberapa penelitian sebelumnya mengenai topik klasifikasi akun palsu menggunakan berbagai algoritma diantaranya K-NN, *neural network*, *naive bayes*, *support vector machine* (SVM), dan model linear. Dari hasil penelitian sebelumnya SVM memiliki performa yang baik diantara algoritma lainnya, sehingga pada penelitian ini menggunakan algoritma SVM. Metode SVM memiliki kekurangan dimana kinerjanya sangat bergantung pada pemilihan fitur. Untuk mengatasi hal tersebut dilakukan optimasi dengan menggunakan seleksi fitur *principal component analysis* (PCA) untuk menghilangkan fitur yang tidak atau kurang penting dengan tujuan meningkatkan kinerja generalisasi, efisiensi, dan akurasi dari model klasifikasi. Dataset yang digunakan bersumber dari penelitian sebelumnya dengan topik yang sama mengenai klasifikasi akun palsu yang dibagikan secara publik melalui GitHub dan digunakan oleh beberapa peneliti lainnya. Dataset tersebut memiliki jumlah data sebanyak 1194 data.

Dari hasil penelitian dan percobaan yang dilakukan, seleksi fitur menggunakan PCA didapatkan dari fitur original yang berjumlah sembilan fitur, menjadi enam fitur penting yakni *follower_following_ratio*, *user_is_private*, *username_length*, *user_has_profil_pic*, *user_biography_length*, dan *username_digit_count*. Pengujian menggunakan metode *10-fold cross validation* terhadap model SVM dengan kernel RBL, C 10, dan *gamma* 0.1 yang dioptimasi dengan seleksi fitur *principal component analysis* (PCA) dihasilkan akurasi meningkat sebanyak 2% dengan akurasi sebesar 94.7%.

Kata Kunci : Instagram, akun palsu, *support vector machine*, *principal component analysis*, klasifikasi.

ABSTRACT

Instagram is a social media platform with a monthly active user base reaching two million users. The features offered by Instagram not only have a positive impact but can also be exploited by some individuals for personal gain, typically carried out by fake accounts. The increasing number of Instagram users is directly proportional to the proliferation of fake accounts, which have negative effects on other users, including fraud, cyberbullying, spamming, and the dissemination of fake news and information. To safeguard the interests of other users and reduce the negative impact caused by the abundance of fake accounts, it is important to classify fake accounts using machine learning systems.

Several previous studies have explored the topic of fake account classification using various algorithms, including K-NN, neural network, naive bayes, support vector machine (SVM), and linear models. From the results of previous research, SVM has shown good performance among other algorithms, leading to its selection for this study. The SVM method has a drawback where its performance heavily relies on feature selection. To address this issue, optimization was carried out using principal component analysis (PCA) for feature selection, aiming to eliminate unimportant features and enhance the overall performance, efficiency, and accuracy of the classification model. The dataset used in this study is sourced from previous research on the same topic of fake account classification, shared publicly through GitHub and utilized by several other researchers. The dataset comprises a total of 1194 data points.

From the results of the research and experiments conducted, feature selection using PCA was obtained from the original features, which amounted to nine features, reduced to six important features, namely follower_following_ratio, user_is_private, username_length, user_has_profil_pic, user_biography_length, and username_digit_count. Testing using the 10-fold cross-validation method on the SVM model with RBF kernel, C 10, and gamma 0.1 optimized with principal component analysis (PCA) feature selection resulted in an accuracy increase of 2%, with an accuracy of 94.7%..

Keyword : *Instagram, fake account, support vector machine, principal component analysis, classification.*