

## INTISARI

### **IMPLEMENTASI *ACTIVE LEARNING* DENGAN SVM PADA KLASIFIKASI EMOSI APLIKASI DASBOR ANALISIS TWITTER**

Ragajiwa Asa

19/447319/SV/17013

Dasbor Analisis Twitter merupakan sebuah aplikasi berbasis web yang berfungsi untuk menganalisis *tweet* berbahasa Indonesia dari 151 akun pilihan. Salah satu fitur yang terdapat dalam aplikasi tersebut adalah klasifikasi emosi dengan menggunakan *Support Vector Machines* (SVM) dan TF-IDF untuk ekstraksi fitur. Terbatasnya *dataset* klasifikasi emosi *tweet* berbahasa Indonesia membuat metode *active learning* menjadi pilihan untuk dapat meningkatkan akurasi model supaya dapat *reliable* ketika memprediksi data baru. Proyek akhir ini menjelaskan bagaimana cara mengimplementasikan metode *active learning* ke dalam Dasbor Analisis Twitter dengan memanfaatkan data tak berlabel yang dihasilkan aplikasi. *Active learning* pada proyek akhir ini menggunakan skenario *stream-based selective sampling* dan memilih data dengan *least confidence uncertainty sampling*.

Pada proyek akhir ini dilakukan pengujian membandingkan performa *active learning* dengan pemilihan data secara acak pada penambahan data 20, 10, dan 50 di setiap iterasi. Menurut hasil yang didapat, model dengan *active learning* memiliki akurasi yang lebih baik dibandingkan dengan model acak. Model *active learning* dapat meningkatkan akurasi model 0,23% hingga 1,09% dari akurasi model dasar pada 66,85%. Peningkatan akurasi ini membuktikan bahwa *active learning* memungkinkan untuk dapat meningkatkan reliabilitas model klasifikasi emosi *tweet* berbahasa Indonesia.

Kata kunci: Twitter, *Machine learning*, Klasifikasi emosi, SVM, *Active learning*

## **ABSTRACT**

### **IMPLEMENTATION OF ACTIVE LEARNING WITH SVM IN EMOTION CLASSIFICATION OF TWITTER ANALYTICS DASHBOARD APPLICATION**

Ragajiwa Asa

19/447319/SV/17013

*The Twitter Analysis Dashboard is a web-based application that analyzes Indonesian-language tweets from 151 selected accounts. One of the features in the application is emotion classification using Support Vector Machines (SVM) and TF-IDF for feature extraction. The limited dataset of Indonesian-language tweet emotion classification makes active learning a choice to increase the model's accuracy to become reliable when predicting new data. This final project explains how to implement active learning into the Twitter Analysis Dashboard by utilizing unlabeled data generated by the application. Active learning in this project uses a stream-based selective sampling scenario and selects data with least confidence uncertainty sampling.*

*A test was conducted to compare the performance of active learning with random data selection in adding 20, 10, and 50 data on each iteration. According to the results obtained, the active learning model had better accuracy than the random model. The active learning model can improve the model's accuracy from 0,23% to 1,09% from the baseline model accuracy of 66,85%. This improvement in accuracy proves that active learning can improve the reliability of emotion classification models for Indonesian language tweets.*

*Keywords: Twitter, Machine learning, Emotion classification, SVM, Active learning*