

## **ABSTRACT**

*The COVID-19 vaccine has become a trending topic among Twitter social media users in Indonesia. This is a public response to the COVID-19 vaccination program carried out by the Indonesian government. This program aims to suppress the spread of the COVID-19 virus in Indonesia. Various opinions about the COVID-19 vaccine have been written by the public on Twitter social media. These opinions can be used for research in the field of sentiment analysis. Many researchers have conducted research on sentiment classification with COVID-19 tweet data from Twitter using various algorithms such as Logistic Regression, Random Forest, and Support Vector Machine (SVM). However, from the variety of uses of these algorithms, in general, accuracy problems are still a concern for today, even accuracy problems related to multi-class classification still require further research. This study proposes a stacking ensemble classifier method to produce better accuracy by combining Logistic Regression, Random Forest, and Support Vector Machine (SVM) algorithms as first-level learners and using Logistic Regression as a meta-learner for the multi-class classification of COVID-19 vaccine topics on Twitter. Based on the evaluation, the proposed Stacking Ensemble Classifier model shows 86% accuracy, 85% precision, 86% recall, and 85% f1-score. But, second Ensemble Classifier model shows 51% accuracy, 52% precision, 51% recall, and 52% f1-score on balanced dataset with small amount.*

*Keywords: COVID-19 Vaccines, Ensemble Method, Multi-class Classification, Sentiment Analysis, Stacking Ensemble Classifier*

## ABSTRAK

Vaksin COVID-19 menjadi *trending topic* di kalangan pengguna media sosial *Twitter* di Indonesia. Hal tersebut merupakan respon masyarakat terhadap program vaksinasi COVID-19 yang dilakukan oleh pemerintah Indonesia. Program ini bertujuan untuk menekan penyebaran virus COVID-19 di Indonesia. Berbagai opini tentang vaksin COVID-19 telah ditulis oleh masyarakat di media sosial *Twitter*. Opini - opini tersebut dapat digunakan untuk penelitian di bidang analisis sentimen. Banyak peneliti telah melakukan penelitian mengenai klasifikasi sentimen dengan data *tweet* COVID-19 dari *Twitter* menggunakan berbagai algoritma seperti *Logistic Regression*, *Random Forest*, dan *Support Vector Machine* (SVM). Namun, dari ragam penggunaan algoritma tersebut pada umumnya permasalahan akurasi masih menjadi perhatian para peneliti saat ini, bahkan permasalahan akurasi terkait klasifikasi multi-kelas masih membutuhkan penelitian lebih lanjut. Penelitian ini mengusulkan metode *stacking ensemble classifier* untuk menghasilkan akurasi yang lebih baik dengan menggabungkan algoritma *Logistic Regression*, *Random Forest*, dan *Support Vector Machine* (SVM) sebagai *first-level learners* serta menggunakan *Logistic Regression* sebagai *meta-learner* untuk klasifikasi multi-kelas topik vaksin COVID-19 pada *Twitter*. Berdasarkan hasil evaluasi, model *Stacking Ensemble Classifier* yang diusulkan menunjukkan nilai *accuracy* sebesar 86%, *precision* 85%, *recall* 86%, dan *f1-score* 85%. Namun dengan model *Stacking Ensemble Classifier* kedua menunjukkan nilai *accuracy* sebesar 51%, *precision* 52%, *recall* 51%, dan *f1-score* 52% dengan menggunakan dataset yang seimbang namun jumlah data yang lebih sedikit.

**Kata Kunci:** Analisis Sentimen, *Ensemble Method*, Klasifikasi Multi-kelas,  
*Stacking Ensemble Classifier*, Vaksin COVID-19