

## ABSTRACT

### CLASSIFYING PRO-VACCINE AND ANTI-VACCINE POSTS IN SOCIAL MEDIA USING TF-IDF WEIGHTED NAIVE BAYES CLASSIFIER

Nindito Weda Bagaskara

15/376923/PA/16439

A tweet is a post made by a user on the social media called Twitter. It is a form of a message that is composed with a maximum of 280 characters. A tweet can contain anything from information to opinions. Several researches have been done to see whether these tweets containing opinions can be analyzed for the purpose of sentiment classification. Naive Bayes is an algorithm that is often used for the purpose of classification on twitter data. A Bayesian classifier allows a weighting algorithm such as TF-IDF to be implemented within it to improve performance.

In this research, the goal is to implement TF-IDF weighting to a Naive Bayes classifier in order to produce a high performance score when used for the purpose of classifying pro-vaccine, anti-vaccine, and neutral tweets. The results obtained in this research gave the average accuracy of 84% for the system's ability of distinguishing between pro-vaccine, anti-vaccine, and neutral tweets. Based on the experiment results obtained, although the TF-IDF weighted Naive Bayes classifier can indeed be satisfactorily accurate, it is un-optimized to be used on a dataset that contains a large amount of sarcasm and ambiguity.

*Key Words: Naive Bayes, Term Frequency-Inverse Document Frequency, Classification*

## ABSTRAK

### KLASIFIKASI POST PRO-VACCINE DAN ANTI-VACCINE DI MEDIA SOSIAL MENGGUNAKAN KLASIFIKASI NAIVE BAYES DENGAN PEMBOBOTAN TF-IDF

Nindito Weda Bagaskara

15/376923/PA/16439

Sebuah *tweet* adalah sebuah posting yang dibuat oleh seorang user di media sosial bernama *Twitter*. Posting tersebut terdiri dari maksimum 280 karakter. Sebuah *tweet* dapat berisi segala sesuatu mulai dari informasi hingga opini. Beberapa peneliti telah mengevaluasi apakah *tweet* ini berisi pendapat yang dapat dianalisis dengan tujuan klasifikasi *sentiment*. *Naive Bayes* adalah sebuah algoritma yang sering digunakan untuk tujuan klasifikasi data *Twitter*. Sebuah *Bayesian classifier* memungkinkan pembobotan algoritma seperti *TF-IDF* untuk diimplementasikan di dalamnya untuk meningkatkan performa.

Tujuan dari penelitian ini adalah untuk mengimplementasikan pembobotan *TF-IDF* pada suatu *classifier Naive Bayes* dengan tujuan meningkatkan nilai performa ketika digunakan untuk tujuan klasifikasi *tweet* pro-vaccine, anti-vaccine, dan netral. Hasil yang diperoleh dari penelitian ini menunjukkan kemampuan sistem mempunyai akurasi rata-rata 84% untuk membedakan antara *tweet* pro-vaccine, anti-vaccine, dan netral. Berdasarkan hasil penelitian, dapat disimpulkan bahwa meskipun sebuah *classifier Naive Bayes* dengan pembobotan *TF-IDF* dapat mempunyai akurasi yang cukup baik, akan tetapi belum optimal ketika digunakan dalam suatu dataset yang berisi penuh dengan sarkasme dan ambigu.

**Key Words:** *Naive Bayes, Term Frequency-Inverse Document Frequency, Classification*