

ABSTRAK

ALDIVA WIBOWO, NIM 4183550007 (2022). Perbandingan Algoritma Machine Learning Support Vector Machine (SVM) dan Random Forest untuk Mendeteksi Konten Negatif Pada Website.

Internet saat ini sudah berkembang sebagai tempat untuk mencari informasi yang memberikan dampak positif kepada kehidupan sehari-hari masyarakat. Tetapi, internet saat ini juga menjadi tempat untuk menyebarkan informasi yang memberikan dampak negatif dari penyebaran konten negatif kepada masyarakat. Untuk memberantas website yang memberikan dampak negatif seperti pornografi, perjudian, dan penipuan pemerintah telah membangun sistem laporan dan DNS (Domain Name System) yaitu *Trust Positif*. Namun langkah tersebut masih kurang efektif karena hanya didasarkan pada DNS dan laporan masyarakat. Tujuan dari penelitian ini adalah untuk menggunakan *Machine Learning* untuk membuat model klasifikasi konten negatif website berdasarkan isi teks website. Metode yang digunakan adalah *Support Vector Machine* dan *Random Forest* sebagai perbandingan. Proses penelitian diawali dengan melakukan preprocessing pada data teks dengan pendekatan dari NLP (*Natural Language Processing*) lalu mengaplikasikan metode *Machine Learning*. Hasil dari pengujian menunjukkan bahwa *Support Vector Machine* lebih unggul dibandingkan dengan *Random Forest* dalam penelitian ini. *Accuracy* dari *Support Vector Machine* sebesar 97%, *Precision* sebesar 90% dan *Recall* sebesar 91%, sedangkan untuk *Accuracy* pada *Random Forest* sebesar 92%, *Precision* sebesar 71%, dan *Recall* sebesar 86%. Nilai yang didapatkan adalah hasil pengujian dengan menggunakan 526 url website.

Kata Kunci : Konten Negatif, Klasifikasi, Machine Learning, Support Vector Machine, Random Forest.

ABSTRACT

ALDIVA WIBOWO, NIM 4183550007 (2022). *Comparison of Machine Learning Support Vector Machine (SVM) Algorithms and Random Forests to Detect Negative Content on Websites.*

The internet has now developed as a place to find information that has a positive impact on people's daily lives. However, the internet is now also a place to disseminate information that harms the spread of negative content to society. To eradicate websites that have negative impacts such as pornography, gambling, and fraud, the government has built a reporting system and a DNS (Domain Name System), namely Positive Trust. However, this measure is still ineffective because it is only based on DNS and public reports. The purpose of this study is to use Machine Learning to create a website negative content classification model based on the website text content. The method used is a Support Vector Machine and Random Forest as a comparison. The research process begins with pre-processing text data using the NLP (Natural Language Processing) approach and then applying the Machine Learning method. The results of the tests show that the Support Vector Machine is superior to the Random Forest in this study. The Accuracy of the Support Vector Machine is 97%, Precision is 90% and Recall is 91%, while for Accuracy in Random Forest is 92%, Precision is 71%, and Recall is 86%. The value obtained is the result of testing using 526 website URLs.

Keywords: Negative Content, Classification, Machine Learning, Support Vector Machine, Random Forest.