

ABSTRAK

Penggunaan email sebagai media komunikasi digital terus meningkat pada individu maupun organisasi karena bersifat cepat, efisien, dan terdokumentasi. Peningkatan intensitas pertukaran informasi melalui email turut mendorong tingginya arus pesan masuk, termasuk pesan yang tidak relevan atau tidak diinginkan. Perkembangan tersebut diikuti oleh bertambahnya email spam yang tidak hanya mengganggu produktivitas, tetapi juga berpotensi menjadi media *phishing*, penipuan, maupun penyebaran *malware* sehingga meningkatkan risiko terhadap keamanan informasi.

Untuk mengatasi hal tersebut, diperlukan sistem klasifikasi yang mampu membedakan email spam dan ham secara akurat. Penelitian ini menganalisis dan membandingkan kinerja dua metode representasi teks, yaitu TF-IDF dan Word2Vec, pada model *Support Vector Machine* (SVM) kernel RBF untuk klasifikasi spam email pada dataset berbahasa Indonesia dan Inggris. Tahapan penelitian meliputi pengumpulan data, *pre-processing* (*cleaning*, *case folding*, *tokenization*, *stopword removal*, *lemmatization*, dan *stemming*), ekstraksi fitur menggunakan TF-IDF dan Word2Vec, serta evaluasi menggunakan *confusion matrix*. Pemilihan parameter terbaik dilakukan melalui optimasi parameter pada metode representasi teks menggunakan *Grid Search* dan skema *K-Fold Cross Validation*. Selain itu, penelitian ini juga mengembangkan sistem klasifikasi spam email yang diuji menggunakan *black-box testing*.

Hasil pengujian menunjukkan bahwa optimasi parameter meningkatkan performa kedua metode, dengan peningkatan Word2Vec yang relatif lebih besar dibanding *baseline*. Pada dataset Bahasa Indonesia, TF-IDF + SVM memperoleh *Accuracy* 0.9498 dan *F1-Score* 0.9416, meningkat menjadi *Accuracy* 0.9594 dan *F1-Score* 0.9540 setelah optimasi. Pada dataset Bahasa Inggris, TF-IDF + SVM memperoleh *Accuracy* 0.9521 dan *F1-Score* 0.9443, meningkat menjadi *Accuracy* 0.9558 dan *F1-Score* 0.9490. Sementara itu, Word2Vec + SVM meningkat dari *Accuracy* 0.9063 menjadi 0.9491 pada dataset Indonesia, dan dari *Accuracy* 0.9100 menjadi 0.9476 pada dataset Inggris setelah optimasi. Secara keseluruhan, TF-IDF menunjukkan performa yang lebih optimal dan konsisten pada kedua dataset, sehingga efektif digunakan sebagai pendekatan klasifikasi spam email berbasis *machine learning*.

Kata kunci: Spam Email, TF-IDF, Word2Vec, *Support Vector Machine*, Klasifikasi Teks

ABSTRACT

Email usage as a digital communication medium continues to increase among individuals and organizations because it is fast, efficient, and well-documented. The growing intensity of information exchange via email also leads to a higher volume of incoming messages, including irrelevant or unsolicited content. This development is accompanied by an increase in spam emails, which not only disrupt productivity but also have the potential to facilitate phishing, fraud, and malware distribution, thereby raising information security risks.

Due to the diverse and evolving textual characteristics of spam, manual filtering becomes inefficient. Therefore, an accurate classification system is required to automatically distinguish spam from ham emails. This study analyzes and compares the performance of two text representation methods, namely TF-IDF and Word2Vec, using a Support Vector Machine (SVM) with an RBF kernel for spam email classification on Indonesian- and English-language datasets. The research workflow includes data collection, text pre-processing (cleaning, case folding, tokenization, stopword removal, lemmatization, and stemming), feature extraction using TF-IDF and Word2Vec, and model evaluation using a confusion-matrix-based approach. Parameter selection is optimized through Grid Search with K-Fold Cross Validation. In addition, this study develops a spam email classification system, which is validated using black-box testing.

The results show that parameter optimization improves the performance of both representations, with Word2Vec exhibiting a relatively larger gain from its baseline. For the Indonesian dataset, TF-IDF + SVM achieves an Accuracy of 0.9498 and F1-score of 0.9416, increasing to an Accuracy of 0.9594 and F1-score of 0.9540 after optimization. For the English dataset, TF-IDF + SVM obtains an Accuracy of 0.9521 and F1-score of 0.9443, improving to an Accuracy of 0.9558 and F1-score of 0.9490 after optimization. Meanwhile, Word2Vec + SVM improves from an Accuracy of 0.9063 to 0.9491 on the Indonesian dataset and from 0.9100 to 0.9476 on the English dataset after optimization. Overall, TF-IDF demonstrates more consistent and slightly superior performance across both datasets, indicating its effectiveness for spam email classification using machine learning.

Keywords: Spam Email, TF-IDF, Word2Vec, Support Vector Machine, Text Classification