

ABSTRAK

Perkembangan teknologi informasi di Indonesia memicu masifnya penyebaran disinformasi atau hoaks melalui portal berita digital, sehingga membutuhkan sistem deteksi otomatis yang presisi. Penelitian ini bertujuan mengomparasi performa algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) berkernel *linear* untuk mengklasifikasikan berita nasional menjadi kategori Hoaks dan Fakta. *Dataset* penelitian berjumlah 2.910 data berita periode 2025, yang terdiri dari 1.510 berita Hoaks dan 1.400 Fakta. Teks diproses melalui tahapan *preprocessing* (*case folding*, *cleansing*, *stopword removal*, *stemming* Sastrawi) dan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) berdimensi 8.000 kosa kata. Pembagian data latih dan uji menggunakan rasio 80:20. Hasil evaluasi model pada 582 data uji menunjukkan bahwa *Naive Bayes* menghasilkan akurasi 93,81%, presisi 93,84%, *recall* 93,81%, dan *F1-score* 93,82%. Di sisi lain, SVM memperoleh metrik performa merata sebesar 93,64%. Meskipun *Naive Bayes* unggul tipis secara statis dan lebih tangguh menekan *False Negative* (15 data), pengujian *5-Fold Cross Validation* membuktikan SVM memiliki stabilitas generalisasi yang lebih tinggi dengan rata-rata *F1-score* 93,94% dibanding *Naive Bayes* (93,39%). SVM juga lebih optimal meminimalisir *False Positive* (17 data). Analisis kesalahan menemukan adanya keterbatasan bias tematik pada pembobotan leksikal saat memproses entitas bernama seperti media sosial. Kesimpulannya, *Naive Bayes* direkomendasikan untuk sistem yang memprioritaskan pemblokiran hoaks secara agresif, sedangkan SVM disarankan untuk sistem moderasi yang mengutamakan keadilan jurnalistik guna mencegah alarm palsu.

Kata Kunci: *Deteksi Hoaks, Naive Bayes, Portal Berita, Support Vector Machine, TF-IDF.*

ABSTRACT

The rapid development of information technology in Indonesia has triggered the massive spread of disinformation or hoaxes through digital news portals, thus requiring a precise automated detection system. This study aims to compare the performance of the Naive Bayes and Support Vector Machine (SVM) algorithms in classifying national news into Hoax and Fact categories. The research dataset consists of 2,910 news items from the 2025 period, comprising 1,510 Hoax and 1,400 Fact news. The texts were processed through preprocessing stages (case folding, cleansing, stopword removal, Sastrawi stemming) and Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction with a dimension of 8,000 vocabularies. The data was split into training and testing sets using an 80:20 ratio. Evaluation results on 582 test data showed that Naive Bayes achieved an accuracy of 93.81%, precision of 93.84%, recall of 93.81%, and F1-score of 93.82%. Meanwhile, SVM obtained an even performance metric of 93.64%. Although Naive Bayes had a slight static advantage and was more robust in suppressing False Negatives (15 data), 5-Fold Cross-Validation testing proved that SVM possessed higher generalization stability with an average F1-score of 93.94% compared to Naive Bayes (93.39%). SVM was also more optimal in minimizing False Positives (17 data). Error analysis revealed a thematic bias limitation in lexical weighting when processing named entities like social media platforms. In conclusion, Naive Bayes is recommended for systems prioritizing aggressive hoax blocking, while SVM is suggested for content moderation systems emphasizing journalistic fairness to prevent false alarms.

Keywords: Hoax Detection, News Portal, Naive Bayes, Support Vector Machine, TF-IDF.