

INTISARI

Meningkatnya angka tindak kejahatan di Indonesia dan tingginya intensitas pemberitaan kriminal di media daring mendorong kebutuhan Humas Kepolisian Republik Indonesia untuk melakukan pemantauan berita pada *dashboard* Berita Indonesia oleh xPlore Pustakadata.id secara lebih terstruktur berdasarkan kategori tindak kejahatan. Sementara itu, pemantauan berita sesuai kategori kejahatan yang diberitakan pada *dashboard* saat ini hanya bisa dilakukan dengan metode pencarian kata kunci yang terbatas untuk analisis lebih dalam. Untuk mendukung kebutuhan tersebut, penelitian ini bertujuan mengembangkan sistem klasifikasi otomatis berbasis *Artificial Intelligence* (AI) dengan pendekatan *Natural Language Processing* (NLP) yang mampu mengelompokkan artikel berita kriminal sesuai dengan kategori tindak kejahatan dalam Kitab Undang-Undang Hukum Pidana (KUHP). Tujuh model klasifikasi teks diuji dalam penelitian ini, yaitu *Multinomial Naive Bayes* (MNB), *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Decision Tree* (DT), *Random Forest* (RF), *K-Nearest Neighbor* (KNN), dan IndoBERT, dengan pendekatan representasi teks *CountVectorizer* dan TF-IDF. Evaluasi dilakukan berdasarkan akurasi, presisi, *recall*, *f1-score*, serta waktu pelatihan dan pengujian. Hasil penelitian menunjukkan bahwa keseluruhan model menunjukkan performa baik yang didukung oleh kualitas dataset yang seimbang dan teknik representasi teks yang informatif. Model IndoBERT memberikan performa tertinggi dengan akurasi 98%, namun membutuhkan waktu pelatihan sekitar 18 menit. Sementara itu, model SVM dengan TF-IDF terbukti menjadi pilihan paling optimal untuk implementasi pada sistem nyata seperti *dashboard* xPlore Pustakadata.id karena mampu mencapai akurasi 97% dengan waktu pelatihan hanya 0,14 detik sehingga model SVM cocok untuk klasifikasi cepat dan efisien, sedangkan IndoBERT tetap relevan digunakan untuk proses klasifikasi skala besar berbasis *batch-processing*.

Kata kunci : Klasifikasi teks, Tindak Kejahatan, *Natural Language Processing*, *Machine Learning*, *Deep Learning*

ABSTRACT

The rising crime rate in Indonesia, coupled with the increasing volume of criminal news coverage in online media, highlights the need for the Public Relations Division of the Indonesian National Police to monitor news more systematically through the Berita Indonesia dashboard by xPlore Pustakadata.id. However, the current keyword-based approach used to filter crime-related news on the dashboard is limited in its ability to support deeper analysis. To address this challenge, this study aims to develop an automated classification system powered by Artificial Intelligence (AI) and Natural Language Processing (NLP) that can accurately categorize criminal news articles based on the types of crime defined in the Indonesian Penal Code (KUHP). Seven text classification models were evaluated in this research: Multinomial Naive Bayes (MNB), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), and IndoBERT. Two text representation techniques, CountVectorizer and TF-IDF, were employed. Each model was assessed using accuracy, precision, recall, F1-score, and training/testing time. Findings show that all models performed well, which can be attributed to the balanced dataset and informative text representations. Among them, IndoBERT achieved the highest performance with 98% accuracy, although it required a longer training time of around 18 minutes. On the other hand, SVM with TF-IDF stood out as the most practical model for real-world implementation, achieving 97% accuracy with a training time of just 0.14 seconds. While SVM is ideal for fast and efficient classification tasks, IndoBERT remains a valuable choice for large-scale or batch-processing systems where deep contextual understanding and accuracy are paramount.

Keywords : *Text classification, Criminal Acts, Natural Language Processing, Machine Learning, Deep Learning*