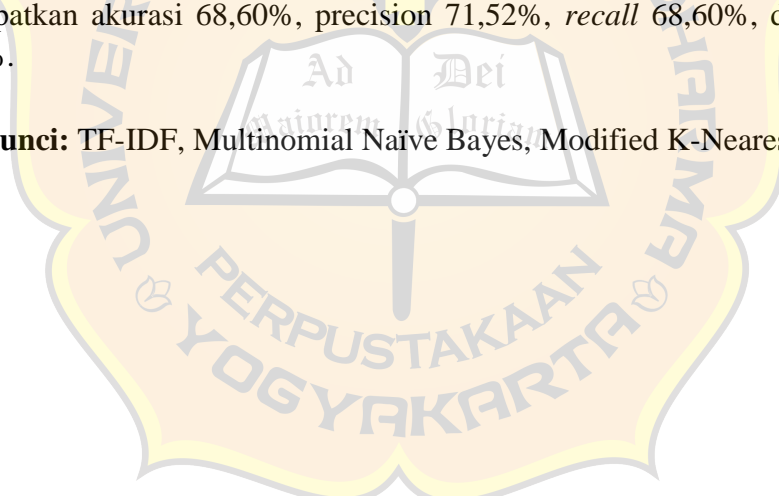


## ABSTRAK

Indonesia merupakan negara dengan populasi besar dan keragaman budaya yang sangat kaya, dengan lebih dari 1.340 kelompok etnik dan bahasa daerah. Namun, meskipun banyak bahasa daerah digunakan oleh jutaan orang, pengembangan dan penelitian di bidang Pemrosesan Bahasa Alami (Natural Language Processing/NLP) terhadap bahasa-bahasa tersebut masih sangat terbatas. Hal ini mengakibatkan minimnya data linguistik yang tersedia untuk eksplorasi dan pengembangan lebih lanjut. Sebagai upaya pelestarian bahasa lokal, komunitas NusaX telah menyediakan kumpulan teks paralel multibahasa berkualitas tinggi untuk 10 bahasa daerah, seperti Aceh, Jawa, dan Sunda, yang dapat digunakan untuk analisis sentimen dan penelitian lainnya. Melanjutkan inisiatif tersebut, penelitian ini berfokus pada analisis sentimen Bahasa Dayak Maayan, salah satu bahasa daerah dari Kalimantan Tengah. Penelitian ini menggunakan algoritma Multinomial Naïve Bayes (MNB) dan Modified K-Nearest Neighbor (MKNN) untuk mengklasifikasikan sentimen dalam teks berbahasa Dayak Maayan. Hasil evaluasi menunjukkan bahwa algoritma MNB memiliki performa yang lebih baik, dengan akurasi sebesar 74,40%, precision 75,10%, recall 74,40%, dan *F1-score* 73,77% dengan parameter *alpha* 0,1 dan *k-fold* = 10. Sedangkan MKNN mendapatkan akurasi 68,60%, precision 71,52%, recall 68,60%, dan *F1-score* 67,86%.

**Kata kunci:** TF-IDF, Multinomial Naïve Bayes, Modified K-Nearest Neighbor.



## ABSTRACT

Indonesia is a country with a large population and rich cultural diversity, comprising more than 1,340 ethnic groups and regional languages. However, despite the widespread use of many regional languages by millions of people, research and development in Natural Language Processing (NLP) for these languages remain very limited. This has led to a lack of available linguistic data for further exploration and development. To address this, the NusaX community has provided high-quality parallel multilingual text datasets for 10 regional languages, such as Acehnese, Javanese, and Sundanese, which can be used for sentiment analysis and other studies. Building on this initiative, this study focuses on sentiment analysis of the Dayak Maayan language, a regional language from Central Kalimantan. The research utilizes Multinomial Naïve Bayes (MNB) and Modified K-Nearest Neighbor (MKNN) algorithms to classify sentiment in Dayak Maayan text. The evaluation results indicate that the MNB algorithm performs better, achieving an accuracy of 74.40%, precision of 75.10%, *recall* of 74.40%, and *F1-score* of 73.77% with an alpha parameter of 0.1 and k-fold = 10. Meanwhile, MKNN achieves an accuracy of 68,60%, precision of 71.52%, *recall* of 68,60%, and *F1-score* of 67.86%.

**Keywords:** TF-IDF, Multinomial Naïve Bayes