# ABSTRAK

Perkembangan media daring telah meningkatkan penyebaran informasi secara cepat, namun juga mendorong munculnya praktik clickbait yang menggunakan judul sensasional untuk menarik perhatian pembaca. Praktik ini dapat menurunkan kualitas informasi dan memengaruhi kredibilitas media. Oleh karena itu, diperlukan metode otomatis untuk mengklasifikasikan judul berita ke dalam kategori clickbait dan non-clickbait. Penelitian ini bertujuan untuk mengevaluasi pengaruh kombinasi N-Gram dan TF-IDF terhadap kinerja varian algoritma Naïve Bayes dalam klasifikasi judul berita clickbait pada dataset *imbalance*.

Penelitian menggunakan dataset berjumlah 8.613 judul berita berbahasa Indonesia yang terdiri dari 3.316 data clickbait dan 5.297 data non-clickbait. Tahapan penelitian meliputi preprocessing teks yang terdiri dari case folding, cleaning, tokenization, stopword removal, dan stemming. Selanjutnya dilakukan ekstraksi fitur menggunakan TF-IDF dengan dua variasi N-Gram, yaitu unigram dan bigram. Model klasifikasi yang digunakan adalah Multinomial Naïve Bayes, Bernoulli Naïve Bayes, dan Complement Naïve Bayes. Evaluasi model dilakukan menggunakan *accuracy*, precision, *recall*, dan *F1-score*.

Hasil penelitian menunjukkan bahwa kombinasi unigram dan Bernoulli Naïve Bayes menghasilkan performa terbaik dengan *accuracy* sebesar 83,52%, precision sebesar 83,55%, *recall* sebesar 83,52%, dan *F1-score* sebesar 83,20%. Sementara itu, seluruh model berbasis unigram menghasilkan performa yang lebih baik dibandingkan model berbasis bigram. Hasil ini menunjukkan bahwa representasi fitur unigram lebih efektif dalam menangkap karakteristik judul berita clickbait pada dataset yang digunakan. Penelitian ini membuktikan bahwa pemilihan representasi fitur memiliki pengaruh terhadap performa klasifikasi, serta menunjukkan bahwa Bernoulli Naïve Bayes merupakan varian yang paling optimal pada skenario penelitian ini.



**Kata Kunci**: *Clickbait*, *Naïve Bayes, Bernoulli Naïve Bayes*, N-Gram, TF-IDF, Klasifikasi Teks

## *ABSTRACT*

*The rapid development of information technology and the internet has significantly transformed the way people consume news. Online news portals have become one of the primary sources of information due to their real-time nature and ease of access. However, intense competition among online media platforms has encouraged the widespread use of clickbait, a practice of creating sensational or misleading headlines to attract readers' attention and increase website traffic. Therefore, an automated method is needed to classify news headlines into clickbait and non-clickbait categories.*

*This study aims to evaluate the effect of combining N-Gram and TF-IDF feature extraction techniques on the performance of Naïve Bayes algorithm variants in clickbait news headline classification using an imbalanced dataset. The dataset consists of 8,613 Indonesian news headlines, including 3,316 clickbait headlines and 5,297 non-clickbait headlines. The research process includes text preprocessing stages, namely case folding, cleaning, tokenization, stopword removal, and stemming. Feature extraction is performed using TF-IDF with two N-Gram representations, namely unigram and bigram. Three variants of the Naïve Bayes algorithm are evaluated: Multinomial Naïve Bayes, Bernoulli Naïve Bayes, and Complement Naïve Bayes. Model performance is assessed using accuracy, precision, recall, and F1-score metrics.*

*The experimental results show that the combination of unigram and Bernoulli Naïve Bayes achieved the best performance, with an accuracy of 83.52%, precision of 83.55%, recall of 83.52%, and F1-score of 83.20%. Furthermore, all unigram-based models outperformed their bigram-based counterparts. These findings indicate that unigram representation is more effective in capturing the characteristics of clickbait headlines in the dataset used. The study demonstrates that feature representation has a significant impact on classification performance and identifies Bernoulli Naïve Bayes as the most effective variant for the clickbait headline classification task under the experimental settings employed.*



*Keywords: Clickbait, Naïve Bayes, Bernoulli Naïve Bayes, N-Gram, TF-IDF, Text Classification*