ABSTRAK

Shafira, Wildan Ayu. 2026. Perbandingan Kinerja *XGBoost* dan *LightGBM* dengan *Hyperparameter Tuning* pada Klasifikasi Berita *Hoax* Berbahasa Indonesia. **Pembimbing**: (1) Lutfi Ali Muharom, S.Si., M.Si.; (2) Daryanto, M. Kom.

Penyebaran berita hoax berbahasa Indonesia melalui internet dan media sosial semakin meningkat dan berpotensi menimbulkan kesalahpahaman, kepanikan, hingga perpecahan di masyarakat, sehingga dibutuhkan sistem klasifikasi otomatis yang akurat dan efisien. Penelitian ini bertujuan membandingkan kinerja algoritma XGBoost dan LightGBM dalam mengklasifikasikan berita hoax berbahasa Indonesia, baik pada kondisi parameter default (baseline) maupun setelah dilakukan hyperparameter tuning menggunakan Grid Search. Dataset yang digunakan terdiri atas 31.353 berita berbahasa Indonesia yang diperoleh dari platform Kaggle, dengan proporsi 10.381 berita hoax dan 20.972 berita non-hoax. Metodologi penelitian mengacu pada kerangka kerja CRISP-DM, meliputi tahap cleaning, case folding, tokenizing, stopword removal, dan stemming pada tahap preprocessing, dilanjutkan dengan ekstraksi fitur menggunakan TF-IDF. Validasi model dilakukan menggunakan skema 5-Fold Cross Validation, dengan evaluasi berdasarkan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa hyperparameter tuning mampu meningkatkan performa kedua algoritma, dengan model XGBoost Tuned dan LightGBM Tuned sama-sama mencapai accuracy sebesar 97,67% dan F1-score sebesar 95,08%. Namun, LightGBM Tuned membutuhkan waktu pelatihan yang jauh lebih efisien (1.385 detik) dibandingkan XGBoost Tuned (5.425 detik), sehingga LightGBM dengan hyperparameter tuning ditetapkan sebagai model yang lebih optimal untuk klasifikasi berita hoax berbahasa Indonesia pada penelitian ini.

Kata kunci: berita *hoax*; *hyperparameter tuning*; klasifikasi teks; *LightGBM; XGBoost*

ABSTRACT

Shafira, Wildan Ayu. 2026. *Performance Comparison of XGBoost and LightGBM with Hyperparameter Tuning for Indonesian Hoax News Classification.*
Advisors: (1) Lutfi Ali Muharom, S.Si., M.Si.; (2) Daryanto, M.Kom.

The spread of Indonesian-language hoax news through the internet and social media continues to increase and has the potential to cause misunderstanding, panic, and social division, so an accurate and efficient automatic classification system is required. This study aims to compare the performance of the XGBoost and LightGBM algorithms in classifying Indonesian-language hoax news, both under default parameters (baseline) and after hyperparameter tuning using Grid Search. The dataset used consists of 31,353 Indonesian-language news articles obtained from the Kaggle platform, comprising 10,381 hoax and 20,972 non-hoax articles. The research methodology follows the CRISP-DM framework, including cleaning, case folding, tokenizing, stopword removal, and stemming in the preprocessing stage, followed by feature extraction using TF-IDF. Model validation was performed using a 5-Fold Cross Validation scheme, with evaluation based on accuracy, precision, recall, and F1-score. The results show that hyperparameter tuning improved the performance of both algorithms, with the tuned XGBoost and tuned LightGBM models both achieving an accuracy of 97.67% and an F1-score of 95.08%. However, the tuned LightGBM model required substantially less training time (1,385 seconds) compared to the tuned XGBoost model (5,425 seconds), establishing tuned LightGBM as the more optimal model for Indonesian-language hoax news classification in this study.

Keywords: *hoax news; hyperparameter tuning; text classification; LightGBM; XGBoost*