

ABSTRAK

Pemindahan ibu kota adalah keputusan strategis yang memiliki dampak besar pada pembangunan dan pertumbuhan suatu negara. Di Indonesia, rencana pemindahan ibu kota dari Jakarta ke Kalimantan telah menjadi perhatian utama. Di dalam proses ini, penting untuk memperhatikan sentimen masyarakat terkait pemindahan ibu kota guna memastikan keberhasilannya. Tempat masyarakat untuk menyuarakan suaranya biasanya adalah di media sosial, salah satunya adalah twitter. Tujuan penelitian ini adalah untuk mengetahui persepsi masyarakat mengenai pemindahan ibu kota Indonesia yang ada dalam media sosial twitter menggunakan pendekatan *machine learning* analisis sentimen. Metode yang dipilih untuk melakukan analisis sentimen ini adalah menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* dan menggunakan dua macam fitur ekstraksi yaitu *Count Vectorization* dan TF-IDF. *XGBoost* merupakan salah satu algoritma yang menerapkan metode *Ensemble learning* di mana metode ini digunakan untuk meningkatkan kinerja model dengan menggabungkan pembelajaran dari beberapa model terpisah, sehingga meningkatkan efisiensi dan akurasi model tersebut. Dalam penelitian ini juga menerapkan metode *balancing Random Undersampler*, *Stratified K-fold Cross validation*, serta *Hyperparameter tuning* menggunakan *GridsearchCV*. Kombinasi *Hyperparameter* yang digunakan antara lain *Learning rate*, *subsample*, *Max_leaves*, *Max_depth*, *gamma*.

Dari hasil penelitian ini didapat akurasi optimal model *XGBoost* adalah 76,18 % dengan menggunakan data *unbalance*, fitur ekstraksi *Count Vectorization*, dan *Hyperparameter tuning* pada nilai *k-fold* 7. Di sisi lain untuk pengujian model dengan data *testing*, akurasi tertinggi yang didapat adalah 78,67% dengan menggunakan data *unbalance*, fitur ekstraksi *Count Vectorization*, dan *Hyperparameter tuning* pada nilai *k-fold* 9. Hasil ini menunjukkan bahwa model *Xgboost* berhasil digunakan untuk menganalisa sentimen pada data *tweet* pemindahan ibu kota Indonesia. Hasil juga menunjukkan bahwa penggunaan fitur ekstraksi *Count vectorization* lebih baik daripada TF-IDF pada analisa sentimen data *tweet* pemindahan ibu kota Indonesia menggunakan algoritma *XGBoost* dengan perbandingan akurasi tertinggi dari kedua fitur 76,18% : 75,86%. Hasil klasifikasi sentimen menunjukkan bahwa masyarakat mayoritas bersikap netral terhadap pemindahan ibukota negara Indonesia.

Kata Kunci : Analisis Sentimen, Xgboost, *count vectorization*, TF-IDF, *random undersampler*, *hyperparameter tuning*, *gridsearchCV*

ABSTRACT

The relocation of the capital city is a strategic decision with a significant impact on a country's development and growth. In Indonesia, the plan to move the capital from Jakarta to Kalimantan has garnered major attention. In this process, it is crucial to consider public sentiment regarding the capital relocation to ensure its success. Social media platforms, particularly Twitter, are common venues for the public to voice their opinions. This study aims to understand public perception regarding the relocation of Indonesia's capital as expressed on Twitter using a machine learning sentiment analysis approach. The method chosen for this sentiment analysis is the Extreme Gradient Boosting (XGBoost) algorithm, employing two feature extraction techniques: Count Vectorization and TF-IDF. XGBoost is an ensemble learning algorithm designed to enhance model performance by combining learning from several separate models, thereby increasing the model's efficiency and accuracy. This research also incorporates balancing methods such as Random Undersampler, Stratified K-fold Cross-validation, and hyperparameter tuning using GridSearchCV. The hyperparameters used include learning rate, subsample, max_leaves, max_depth, and gamma.

The study found that the optimal accuracy of the XGBoost model was 76.18% using unbalanced data, Count Vectorization feature extraction, and hyperparameter tuning at a k-fold value of 7. In contrast, for the testing data, the highest accuracy achieved was 78.67% using unbalanced data, Count Vectorization feature extraction, and hyperparameter tuning at a k-fold value of 9. These results indicate that the XGBoost model successfully analyzed sentiment in tweets regarding the relocation of Indonesia's capital city. Furthermore, the results show that the use of Count Vectorization feature extraction is more effective than TF-IDF in sentiment analysis of tweets about the capital relocation using the XGBoost algorithm, with the highest accuracy comparison of the two features being 76.18% for Count Vectorization versus 75.86% for TF-IDF. The results of the sentiment classification show that the majority of people have a neutral attitude towards moving the Indonesian capital.

Key word : Sentiment Analysis, Xgboost, count vectorization, TF-IDF, random undersampler, hyperparameter tuning, gridsearchCV