

ABSTRAK

IMPLEMENTASI *NATURAL LANGUAGE PROCESSING* (NLP) DASAR PADA ANALISIS SENTIMEN TAGAR #KABURAJADULU

Oleh

ZIKWAN ISMAIL

Fenomena penggunaan tagar #KaburAjaDulu di media sosial mencerminkan respons publik terhadap kondisi sosial, ekonomi dan politik di Indonesia, khususnya dari masyarakat yang mempertimbangkan untuk menetap di luar negeri. Penelitian ini bertujuan untuk menganalisis opini publik yang terungkap melalui unggahan bertagar tersebut menggunakan pendekatan *Natural Language Processing* (NLP) dasar. Peneliti mengumpulkan 5.198 data unggahan dari platform X (*Twitter*) dan *TikTok* melalui proses *scraping*, kemudian menerapkan tahap preprocessing teks seperti *cleaning*, *tokenization*, *stopword removal*, *case folding* dan *stemming* menggunakan *library Python* seperti *nlk* dan Sastrawi. Data yang telah diproses dilabeli secara otomatis oleh *volunteer* dan digunakan untuk melatih dua model klasifikasi sentimen: *Naïve Bayes* dan *Logistic Regression*. Evaluasi model dilakukan dengan metrik *accuracy*, *precision*, *recall* dan *F1-score* melalui pendekatan *cross-validation*.

Hasil penelitian menunjukkan bahwa *Logistic Regression* memberikan performa klasifikasi yang lebih unggul dibandingkan *Naïve Bayes* pada seluruh *dataset*, dengan akurasi mencapai 89% pada data *TikTok*, dibandingkan 65% oleh *Naïve Bayes*. Selain itu, model *Logistic Regression* menunjukkan ketahanan yang lebih baik dalam menangani distribusi data yang kompleks dan tidak seimbang. Secara keseluruhan, penelitian ini menunjukkan efektivitas metode NLP dasar dalam memetakan opini publik secara digital dan menegaskan bahwa *Logistic Regression* lebih adaptif dalam memahami konteks sentimen dari teks sosial media. Temuan ini memberikan kontribusi penting bagi studi lanjutan mengenai analisis opini publik dan pengambilan keputusan berbasis data sosial.

Kata Kunci: Analisis Sentimen, Pemrosesan Bahasa Alami (NLP), *Twitter* (X), *TikTok*, #kaburajadulu, *Naïve Bayes*, Regresi Logistik.

ABSTRACT

IMPLEMENTATION OF BASIC NATURAL LANGUAGE PROCESSING (NLP) IN THE ANALYSIS OF SENTIMENT HASHTAG #KABURAJADULU

By

ZIKWAN ISMAIL

The phenomenon of using the hashtag #KaburAjaDulu on social media reflects the public's response to social, economic, and political conditions in Indonesia, particularly among those considering settling abroad. This study aims to analyze public opinion expressed through posts using this hashtag using a basic Natural Language Processing (NLP) approach. The researchers collected 5,198 posts from the X (Twitter) and TikTok platforms through scraping, then applied text preprocessing steps, including cleaning, tokenization, stopword removal, case folding, and stemming, using Python libraries such as nltk and Sastrawi. The processed data was manually labeled by volunteers and used to train two sentiment classification models: Naïve Bayes and Logistic Regression. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics through a cross-validation approach.

The results showed that Logistic Regression outperformed Naïve Bayes across all datasets, achieving 89% accuracy on TikTok data compared to 65% by Naïve Bayes. Additionally, the Logistic Regression model demonstrated better resilience in handling complex and imbalanced data distributions. Overall, this study highlights the effectiveness of basic NLP methods in mapping public opinion digitally and confirms that Logistic Regression is more adaptive.

Keywords: *Sentiment Analysis, NLP, Twitter(X), TikTok, #kaburajadulu, Naïve Bayes, Logistic*