

ABSTRAK

Media sosial, khususnya platform X, menjadi ruang utama bagi masyarakat untuk menyampaikan opini terhadap isu publik yang berkembang, termasuk isu terkait Pertamina yang kerap memunculkan respons beragam. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap isu Pertamina di platform X ke dalam dua kelas, yaitu positif dan negatif, menggunakan pendekatan analisis sentimen berbasis teks. Data penelitian dibatasi pada tweet berbahasa Indonesia pada periode 1 September 2025-25 Desember 2025 dengan jumlah 1.819 tweet. Tahapan penelitian dilakukan melalui pengumpulan data, preprocessing teks, pembentukan fitur menggunakan TF-IDF dengan n-gram (1,2), serta proses klasifikasi menggunakan algoritma *Naive Bayes*, khususnya *Complement Naive Bayes*, dengan pemilihan parameter terbaik menggunakan *GridSearchCV* sehingga diperoleh nilai *alpha* terbaik sebesar 0,05. Hasil pengujian menunjukkan nilai akurasi sebesar 0,8407 dengan distribusi label sebanyak 938 tweet negatif dan 881 tweet positif. Berdasarkan *confusion matrix*, prediksi benar diperoleh pada 157 data negatif dan 149 data positif, sedangkan kesalahan klasifikasi terjadi pada 31 data negatif terbaca positif dan 27 data positif terbaca negatif. Hasil tersebut menunjukkan bahwa kombinasi *preprocessing*, TF-IDF, dan *Complement Naive Bayes* mampu memberikan kinerja yang baik dalam memetakan sentimen publik terhadap isu Pertamina di platform X pada periode penelitian.

Kata kunci: analisis sentimen, platform X, TF-IDF, *Complement Naive Bayes*, *Naive Bayes*

ABSTRACT

Social media, particularly the X platform, has become a primary space where people express opinions about emerging public issues, including topics related to Pertamina that often trigger diverse public reactions. This study aims to classify public sentiment toward Pertamina-related issues on X into two classes, positive and negative, using a text-based sentiment analysis approach. The dataset was limited to Indonesian-language tweets collected from 1 September 2025 to 25 December 2025, totaling 1,819 tweets. The research procedure consisted of data collection, text preprocessing, feature extraction using TF-IDF with n-gram (1,2), and sentiment classification using the Naive Bayes algorithm, specifically Complement Naive Bayes, with optimal parameter selection through GridSearchCV, resulting in the best alpha value of 0.05. The evaluation results show an accuracy of 0.8407, with a label distribution of 938 negative tweets and 881 positive tweets. Based on the confusion matrix, correct predictions were obtained for 157 negative instances and 149 positive instances, while misclassifications occurred in 31 negative instances predicted as positive and 27 positive instances predicted as negative. These findings indicate that the combination of preprocessing, TF-IDF, and Complement Naive Bayes provides good performance for mapping public sentiment toward Pertamina-related issues on the X platform within the study period.

Keywords sentiment analysis, X platform, TF-IDF, Complement Naive Bayes, Naive Bayes