ABSTRAK

Studi ini mengkaji reaksi masyarakat terhadap rencana pemerintah menaikkan tarif PPN dari 11% menjadi 12% pada tahun 2025. Kebijakan ini memicu berbagai opini warganet di media sosial X. Data dikumpulkan melalui web crawling dari Oktober hingga Desember 2024, menghasilkan 1.871 data. Setelah praproses teks seperti cleaning, case folding, tokenize, stopword removal, dan stemming, dataset menjadi 1.806 data. Sebanyak 1.000 data dilabeli negatif, netral, dan positif secara manual oleh pakar bahasa untuk memastikan validitas. Sisanya, 806 data tanpa label, digunakan untuk pengujian. Kata-kata dibobot menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Tiga algoritma machine learning dibandingkan: Support Vector Machine (SVM), Random Forest, dan Decision Tree. Hasil menunjukkan SVM mencatat akurasi tertinggi 94%, diikuti Random Forest 93%, dan Decision Tree 91%. Mayoritas sentimen bersifat negatif, menunjukkan ketidakpuasan masyarakat terhadap kebijakan ini. Studi ini membuktikan bahwa machine learning efektif menangkap persepsi publik melalui media sosial dan dapat menjadi masukan bagi pemerintah.

Kata Kunci: PPN 12%, Sentiment Analysis, Machine Learning, Social Media X, SVM, Random Forest, Decision Tree

ABSTRACT

This study examines people's reactions to the Indonesian government's plan to adjust the VAT rate from 11% to 12%, which is scheduled to take effect in 2025. This policy triggered a variety of opinions among netizens, especially on the social networking service X. To explore public opinion, data was collected through web crawling techniques from October to December 2024, resulting in 1,871 records. Then the dataset was preprocessed by text cleaning, case folding, tokenization, stopword removal, and stemming, and the dataset was reduced to 1806. In addition, up to 1000 data will be manually labeled, negative, neutral, positive, by language experts to ensure that each sentence has the appropriate label. These data are used for testing and training, then up to 806 unlabeled data are used as final testing. At the word weighting stage, the Term Frequency-Inverse Document Frequency (TF-IDF) method is used to perform the process. In this study, three machine learning algorithms were used to compare the classification performance, namely Support Vector Machine (SVM), Random Forest, and Decision Tree. Based on the evaluation results, the SVM algorithm recorded the highest accuracy rate of 94%, followed by Random Forest with 93% and Decision Tree with 91%. The results showed a predominance of negative sentiments, indicating public dissatisfaction with the policy. This study proves that machine learning techniques can be effectively used to capture public perceptions through social media, which in turn can be a benchmark for the government to make decisions that will be enforced.

***Keyword**: PPN 12%, Sentiment Analysis, Machine Learning, Social Media X, SVM, Random Forest, Decision Tree.*