

ABSTRAK

Wirya Yasa, I Nyoman Wirya (2026), *Analisis Sentimen Komentar Youtube Terhadap Kebijakan Pajak dan Insentif Kendaraan Listrik di Indonesia Menggunakan Naïve Bayes dan K-Nearest Neighbor*. Tesis, Ilmu Komputer, Program Pasca Sarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I: Dr. I Made Gede Sunarya, S.Kom., M.Cs. dan Pembimbing II: Dr. Putu Hendra Suputra, S.Kom., M.Cs.

Kata-kata kunci: analisis sentimen, youtube, *Naïve Bayes Classifier*, *K-Nearest Neighbor*, kendaraan listrik.

Perubahan kebijakan pemerintah mengenai pajak dan insentif kendaraan listrik di Indonesia menimbulkan beragam tanggapan masyarakat yang banyak disampaikan melalui media sosial, khususnya YouTube. Besarnya jumlah komentar yang dihasilkan membuat proses analisis manual kurang efektif, sehingga diperlukan pendekatan analisis sentimen berbasis *machine learning* untuk mengidentifikasi kecenderungan opini masyarakat secara otomatis. Tujuan utama dari penelitian ini mencakup penerapan metode *Naïve Bayes Classifier* (NBC) dan *K-Nearest Neighbor* (K-NN) dalam analisis sentimen komentar YouTube terkait kebijakan pajak dan insentif kendaraan listrik di Indonesia, serta membandingkan kinerja kedua metode tersebut. Data penelitian diperoleh melalui proses scraping komentar dari kanal media berita di YouTube menggunakan Python dan YouTube Data API. Total data yang diperoleh sebanyak 3.507 komentar. Setelah proses pelabelan dan seleksi data, diperoleh 2.297 komentar terdiri atas 1.833 komentar negatif dan 464 komentar positif. Tahapan pengolahan data meliputi *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, *stemming*, pembobotan kata menggunakan TF-IDF, klasifikasi menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor*, serta evaluasi menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *Receiver Operating Characteristic – Area Under the Curve* (ROC-AUC). Hasil penelitian menunjukkan bahwa pada data yang tidak seimbang, metode *K-Nearest Neighbor* memberikan akurasi terbaik sebesar 90%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 86%. Namun, setelah dilakukan penyeimbangan data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), metode *Naïve Bayes* menunjukkan peningkatan performa dan menjadi model terbaik dengan akurasi 88%, *precision* 62%, *recall* 94%, *F1-score* 75%, serta AUC 96%. Sementara itu, *K-Nearest Neighbor* pada data yang telah diseimbangkan hanya menghasilkan akurasi 39%, sehingga menunjukkan penurunan performa yang signifikan. Hasil tersebut membuktikan bahwa penerapan SMOTE mampu meningkatkan kemampuan *Naïve Bayes* dalam mengklasifikasikan kelas minoritas secara lebih seimbang, sehingga metode ini lebih efektif digunakan untuk analisis sentimen komentar YouTube mengenai kebijakan pajak dan insentif kendaraan listrik di Indonesia.

ABSTRACT

Wirya Yasa, I Nyoman (2026), Sentiment Analysis of YouTube Comments Regarding Electric Vehicle Tax Policies and Incentives in Indonesia Using Naïve Bayes and K-Nearest Neighbor. Thesis, Computer Science, Postgraduate Program, Universitas Pendidikan Ganesha.

This thesis has been approved and reviewed by Supervisor I: Dr. I Made Gede Sunarya, S.Kom., M.Cs., and Supervisor II: Dr. Putu Hendra Suputra, S.Kom., M.Cs.

Keywords: sentiment analysis, YouTube, Naïve Bayes Classifier, K-Nearest Neighbor, electric vehicles.

The Indonesian government's policies regarding electric vehicle (EV) taxation and incentives have generated diverse public responses, particularly on social media platforms such as YouTube. The large volume of user-generated comments makes manual analysis inefficient; therefore, a machine learning-based sentiment analysis approach is required to automatically identify public opinion trends. This study aims to implement the Naïve Bayes Classifier (NBC) and K-Nearest Neighbor (K-NN) algorithms for sentiment analysis of YouTube comments related to Indonesia's electric vehicle tax and incentive policies, as well as to compare the performance of both classification methods. The research data were collected by scraping comments from four YouTube news channels using Python and the YouTube Data API. A total of 3,507 comments were collected. After the labeling and data selection processes, 2,297 comments were retained, consisting of 1,833 negative comments and 464 positive comments. The data processing stages included text preprocessing (cleaning, case folding, normalization, tokenization, stop-word removal, and stemming), TF-IDF feature weighting, sentiment classification using the Naïve Bayes and K-Nearest Neighbor algorithms, and performance evaluation using a confusion matrix with the metrics of accuracy, precision, recall, F1-score, and Receiver Operating Characteristic – Area Under the Curve (ROC-AUC). The experimental results indicate that, when applied to the original imbalanced dataset, the K-Nearest Neighbor algorithm achieved the highest accuracy of 90%, while the Naïve Bayes Classifier achieved an accuracy of 86%. However, after applying the Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset, the Naïve Bayes Classifier demonstrated superior performance, achieving an accuracy of 88%, precision of 62%, recall of 94%, F1-score of 75%, and an AUC of 96%. In contrast, the performance of the K-Nearest Neighbor algorithm significantly decreased, yielding an accuracy of only 39% on the balanced dataset. These findings demonstrate that the implementation of SMOTE effectively improves the Naïve Bayes Classifier's ability to classify minority-class instances, making it a more suitable approach for sentiment analysis of YouTube comments regarding electric vehicle tax and incentive policies in Indonesia.