

ABSTRAK

Analisis sentimen memiliki tujuan yaitu untuk mengetahui apakah polaritas sebuah data yang berbentuk teks mengarah pada sifat positif atau negatif. Salah satu topik yang sedang ramai diperbincangkan pada media sosial *Twitter* adalah tentang peralihan dari mobil konvensional menuju mobil listrik di Indonesia. Salah satu hal yang menjadi tolak ukur keberhasilannya adalah bagaimana sentimen masyarakat mengenai mobil listrik di Indonesia. Penelitian ini memiliki tujuan utama untuk melihat efektivitas dan mendapatkan kinerja metode *Random Forest* dalam melakukan analisis sentimen terkait dengan topik peralihan mobil listrik serta mengetahui reaksi masyarakat terhadap kebijakan peralihan mobil listrik di Indonesia. Diperoleh data sejumlah 7774 *tweets*. Setelah melalui proses filtrasi, data yang dihasilkan adalah 2012 yang digunakan untuk melatih model. Tahapan yang dilakukan pada penelitian ini menggunakan model penelitian *data science* CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Pada penelitian ini digunakan teknik *Feature extraction* TF-IDF (*Term Frequency-Inverse Document Frequency*). Digunakan juga teknik SMOTE (*Synthetic Minority Over-sampling Technique*) yang merupakan bagian dari skenario pengujian. Pada skenario pertama, dengan menggunakan 1675 data yang dihasilkan setelah proses filtrasi, didapatkan akurasi sebesar 77.49%. Pada skenario kedua, menggunakan 1788 data yang didapatkan setelah penerapan teknik SMOTE didapatkan akurasi sebesar 77.74%. Pada skenario ketiga, menggunakan 1048 data yang didapatkan setelah proses filtrasi dan *K-means clustering* dengan nilai $k=4$ didapatkan hasil akurasi sebesar 98.47%. Pada skenario keempat, dengan menggunakan 1544 data yang didapatkan setelah proses filtrasi, *K-means clustering* dengan nilai $k=4$ dan SMOTE, didapatkan hasil akurasi sebesar 99.55%. pada skenario kelima, dilakukan evaluasi tanpa menerapkan *cross-validation*, dengan menggunakan 1675 data, hasil akurasi yang didapatkan adalah 77.34%. Pada skenario pengujian keenam, hasil menunjukkan suatu prediksi terhadap opini masyarakat mengenai peralihan mobil listrik di Indonesia cenderung bersifat negatif.

Kata Kunci: Mobil Listrik, Analisis Sentimen, *Random Forest*, klasifikasi, *Twitter*

ABSTRACT

Sentiment analysis aims to determine whether the polarity of textual data leans towards a positive or negative nature. A trending topic on the Twitter social media platform is the transition from conventional to electric cars in Indonesia. A key metric for success is gauging public sentiment regarding electric cars in the country. This research aims to assess the effectiveness and performance of the Random Forest method in sentiment analysis related to the electric car transition, while also understanding public reactions to the policy shift in Indonesia. A dataset comprising 7774 tweets was obtained, and after filtration, 2012 tweets were used for model training. The research follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) data science model. The TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction technique and SMOTE (Synthetic Minority Over-sampling Technique) were employed in testing scenarios. In the first scenario, with 1675 filtered data, an accuracy of 77.49% was achieved. The second scenario, using 1788 data after applying SMOTE, resulted in an accuracy of 77.74%. In the third scenario, using 1048 filtered data and K-means clustering with $k=4$, an accuracy of 98.47% was obtained. The fourth scenario, incorporating 1544 filtered data, K-means clustering with $k=4$, and SMOTE, yielded an accuracy of 99.55%. In the fifth scenario, without cross-validation and using 1675 data, the accuracy was 77.34%. The sixth testing scenario revealed a negative trend in predicting public opinion on the transition to electric cars in Indonesia.

Keywords: *Electric Cars, Sentiment Analysis, Random Forest, Classification, Twitter*