ABSTRACT

ChatGPT is an artificial intelligence–based application developed by OpenAI and available on the Google Play Store to assist users in various conversational and information-seeking needs. The large number of user reviews, both positive and negative, makes sentiment analysis essential to understand users' perceptions of the application's performance. In this study, review data were collected using a scraping technique with the Google Play Scraper library, resulting in 10,000 Indonesian language reviews with the application ID "com.openai.chatgpt" during the period of July to August 2025. The collected data then underwent a preprocessing stage to clean irrelevant characters before applying word weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The class distribution was further balanced using the Synthetic Minority Over-sampling Technique (SMOTE) to prevent the model from being biased toward the majority class. The classification process employed the Support Vector Machine (SVM) algorithm, which is known for its effectiveness in handling high-dimensional data. The experimental results show that the combination of TF-IDF and SMOTE significantly improves model performance, achieving an accuracy of 97.81%, precision of 99.12%, recall of 98.50%, F1-score of 98.81%, and AUC of 99.85%. Therefore, this approach is proven to be accurate, stable, and effective for conducting sentiment analysis on ChatGPT user reviews.

Keywords: Sentiment Analysis, ChatGPT, TF-IDF, SMOTE, Support Vector Machine

ABSTRAK

ChatGPT merupakan aplikasi berbasis kecerdasan buatan yang dikembangkan oleh *OpenAI* dan tersedia di *Google Play Store* untuk membantu pengguna dalam berbagai kebutuhan percakapan serta pencarian informasi. Banyaknya ulasan pengguna, baik positif maupun negatif, menjadikan analisis sentimen penting dilakukan untuk mengetahui persepsi pengguna terhadap kinerja aplikasi tersebut. Pada penelitian ini, data ulasan diperoleh melalui teknik *scraping* menggunakan pustaka *Google Play Scraper* sebanyak 10.000 ulasan berbahasa Indonesia dengan *application ID* "com.openai.chatgpt" pada periode Juli hingga Agustus 2025. Data yang terkumpul kemudian diproses melalui tahapan *preprocessing* guna membersihkan teks dari karakter yang tidak relevan sebelum dilakukan pembobotan kata menggunakan metode *Term Frequency Inverse Document Frequency* (TF-IDF). Selanjutnya, dilakukan penyeimbangan distribusi kelas menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) agar model tidak bias terhadap kelas mayoritas. Proses klasifikasi dilakukan dengan algoritma *Support Vector Machine* (SVM) yang dikenal efektif dalam memisahkan data berdimensi tinggi. Hasil pengujian menunjukkan bahwa kombinasi *TF-IDF* dan *SMOTE* mampu meningkatkan kinerja model secara signifikan, ditunjukkan oleh nilai akurasi 97,81%, *precision* 99,12%, *recall* 98,50%, *F1-score* 98,81%, serta *AUC* 99,85%. Dengan demikian, pendekatan ini terbukti akurat, stabil, dan efektif untuk analisis sentimen pengguna aplikasi *ChatGPT*.

Kata kunci: Analisis Sentimen, *ChatGPT, TF-IDF, SMOTE*, *Support Vector Machine.*