ABSTRAK

Olimpiade Paris 2024 menjadi salah satu perhelatan olahraga terbesar yang menarik perhatian publik di seluruh dunia, memicu beragam respons yang terekam dalam banyak unggahan digital di media sosial, termasuk platform X (Twitter). Hal ini menimbulkan tantangan dalam mengidentifikasi berbagai jenis sentimen yang berkembang di masyarakat. Respons publik yang beragam dapat mempengaruhi persepsi terhadap pelaksanaan acara dan dampaknya terhadap citra Olimpiade secara keseluruhan. Oleh karena itu, pemahaman yang tepat mengenai sentimen publik diperlukan suntuk mendukung pengambilan keputusan yang responsif dan efektif terhadap respons masyarakat di masa depan.

Pada penelitian kali ini, dilakukan analisis sentimen terhadap Olimpiade Paris 2024 menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma *Random Forest*. Metodologi penelitian ini mencakup pengumpulan data dari platform X (Twitter), pelabelan data secara manual dan berbasis *lexicon*, *preprocessing* teks, pembobotan menggunakan TF-IDF, pelatihan model menggunakan Random Forest, serta evaluasi performa model berdasarkan akurasi, *precision, recall*, F1-score. Model Random Forest dilatih dengan skenario pelabelan dan *negation handling* untuk memperoleh performa terbaik.

Hasil penelitian menunjukkan bahwa model terbaik diperoleh pada model dengan pelabelan *lexicon* dengan *negation handling*, mencapai akurasi tertinggi sebesar 82%, *precision* 82%, *recall* 91%, dan F1-score 86%. Penggunaan *negation handling* meningkatkan akurasi model sebesar 2%, sementara model dengan pelabelan *lexicon* didapatkan nilai akurasi 2% lebih tinggi dibandingkan model dengan pelabelan manual. Temuan ini menunjukkan bahwa pendekatan yang menggabungkan *negation handling* dan pelabelan *lexicon* mampu meningkatkan performa analisis sentimen pada model dengan algoritma *Random Forest* dan TF-IDF.

Kata Kunci: Analisis Sentimen, Random Forest, Decision Tree, TF-IDF, Olimpiade Paris 2024.

ABSTRACT

The Paris 2024 Olympics has become one of the largest sporting events, capturing global attention and sparking a variety of responses recorded in numerous digital posts on social media, including the X platform (formerly Twitter). This creates challenges in identifying the different types of sentiments emerging within society. The diverse public responses can influence perceptions of the event's execution and its impact on the overall image of the Olympics. Therefore, a proper understanding of public sentiment is necessary to support responsive and effective decision-making in future public interactions.

This study conducts sentiment analysis of the Paris 2024 Olympics using the Term Frequency-Inverse Document Frequency (TF-IDF) method and the Random Forest algorithm. The research methodology includes data collection from the X platform (Twitter), manual and lexicon-based data labeling, text preprocessing, TF-IDF weighting, model training using Random Forest, and model performance evaluation based on accuracy, precision, recall, and F1-score. The Random Forest model was trained with labeling scenarios and negation handling to achieve the best performance.

The results show that the best model was achieved with lexicon labeling and negation handling, yielding the highest accuracy of 82%, precision of 82%, recall of 91%, and F1-score of 86%. The use of negation handling improved the model's accuracy by 2%, while the lexicon labeling model achieved an accuracy 2% higher than the manual labeling model. These findings suggest that the approach combining negation handling and lexicon labeling enhances the performance of sentiment analysis using Random Forest and TF-IDF algorithms.

Keywords: Sentiment Analysis, Random Forest, Decision Tree, TF-IDF, Paris 2024 Olympics.