

ABSTRAK

Olahraga sudah menjadi bagian dari kehidupan manusia dan kehidupan suatu negara. Salah satu ajang olahraga yang ditunggu-tunggu khususnya warga anggota ASEAN (Association of Southeast Asian Nations) adalah SEA (Southeast Asian Games) yang diadakan pada tanggal 5 Mei 2023 hingga 17 Mei 2023 di Kamboja. Kamboja sebagai tuan rumah SEA Games mendapatkan banyak kritikan. Kritikan ini banyak sekali disampaikan pada platform sosial media khususnya twitter. Dari sebuah tweet pada media sosial twitter ini dapat dilakukan penelitian untuk mengetahui sentiment terhadap SEA Games 2023. Data tweet yang akan diolah sebanyak 2443 dengan berbagai kata kunci. Data kemudian akan diklasifikasikan ke dalam tiga jenis yaitu negatif, positif, dan netral dengan menggunakan metode K-Nearest Neighbor (K-NN). Pada penelitian ini juga menerapkan metode data *balancing* yaitu SMOTE, Tomek Link, dan SMOTE-Tomek Link. Hasil dari penelitian ini didapatkan data yang tidak diseimbangkan dengan K=5 memiliki akurasi 71,75% dan data yang diseimbangkan menggunakan SMOTE dengan K=1 memiliki akurasi 81,17%, Tomek Link dengan K=9 memiliki akurasi 70,68%, dan SMOTE-Tomek Link dengan K=1 memiliki akurasi 80,57%. Hasil dari penelitian ini menunjukkan bahwa metode data *balancing* menggunakan SMOTE memiliki akurasi yang lebih tinggi dalam melakukan analisis sentimen ini.

Kata kunci : SEA Games, Analisis Sentimen, *K-Nearest Neighbor*, Data *Balancing*, SMOTE, Tomek Link, SMOTE-Tomek Link

ABSTRACT

Sports have become a part of human life and the life of a country. One of the sporting events that is eagerly awaited by members of ASEAN (Association of Southeast Asian Nations) is the SEA (Southeast Asian Games) which will be held from 5 May 2023 to 17 May 2023 in Cambodia. Cambodia as the host of the SEA Games has received a lot of criticism. This criticism has been expressed a lot on social media platforms, especially Twitter. From a tweet on Twitter social media, research can be carried out to determine sentiment towards the 2023 SEA Games. The tweet data that will be processed is 2443 with various keywords. The data will then be classified into three types, namely negative, positive and neutral using the K-Nearest Neighbor (K-NN) method. This research also applies data balancing methods, namely SMOTE, Tomek Link, and SMOTE-Tomek Link. The results of this research showed that unbalanced data with $K=5$ had an accuracy of 71.75% and data balanced using SMOTE with $K=1$ had an accuracy of 81.17%, Tomek Link with $K=9$ had an accuracy of 70.68%, and SMOTE-Tomek Link with $K=1$ has an accuracy of 80.57%. The results of this research show that the data balancing method using SMOTE has higher accuracy in carrying out this sentiment analysis.

Keywords: SEA Games, Sentiment Analysis, K-Nearest Neighbor, Data Balancing, SMOTE, Tomek Link, SMOTE-Tomek Link.