ABSTRACT

The 2024 Aceh-North Sumatra National Sports Week (PON) XXI has become a major event that has sparked a variety of public reactions on social media, especailly on the X platform. This research seeks to examine audience sentiment towards the organisation of PON XXI to gain a deeper understanding of public opinion trends. Data was collected through a crawling process using Tweet Harvest with relevant keywords, resulting in 2,503 posts. The collected data was then classified into three sentiment categories: positive, negative, and neutral. This study utilised the Support vector Machine (SVM) algorithm as the primary classification method, applying four types of kernels: linear, polynomial, Radial Basis Function (RBF), and sigmoid. The classification process was conducted in two approaches: with and without the application of the Synthetic Minority Over-sampling Technique (SMOTE) as a method for addressing data imbalance. Additionally, experiments were conducted with three training and testing data ratio combinations: 90:10, 80:20, and 70:30. The testing results showed that the SVM classification model combined with the SMOTE technique yielded more optimal results compared to the model without SMOTE. Of all the combinations tested, the best performance was achieved by the SVM model with RBF kernel and an 80:20 ratio, which produced an accuracy of 87%, precision of 86%, recall of 85%, and an F1-Score of 85%. As the final step, the best model was implemented into a Streamlit-based dashboard that presents analysis results in the form of visualisations, such as wordclouds, n-gram distributions, and the distribution of frequently appearing hashtags related to the event. Additionally, the dashboard is equipped with a new sentiment prediction feature, developed from an SVM-based sentiment classification model.

Keywords: *Sentiment Analysis, PON XXI Aceh-North Sumatra 2024, SMOTE, SVM, Text Classification*