

ABSTRACT

Sentiment analysis of products in e-commerce has become a primary concern within the industry. Sentiment analysis provides insights into consumer preferences and behavior during purchases. The aim of this project is to develop a machine learning model using the k-NN algorithm to analyze sentiment in product reviews on Lazada Indonesia. The methodology includes data collection, data preprocessing, analysis, modeling, and evaluation. Review data is sourced from Kaggle.com and processed using Natural Language Processing (NLP) to identify positive, negative, or neutral sentiments. The k-NN model is constructed and assessed. Results indicate that the NLP and k-NN approach can be applied to analyze sentiment in e-commerce product reviews. However, the data used tends to be imbalanced. The initial model achieved 87% accuracy but addressing data imbalance with Random UnderSampler led to a decrease to 57% accuracy while improving the recognition of minority classes. The Random OverSampler technique achieved 96% accuracy, and Synthetic Minority Over-sampling Technique (SMOTE) reached 94%. In a further development, it is crucial to consider the impact of imbalanced data on model predictions and tackle it effectively. By implementing more sophisticated data handling techniques, the model is expected to provide more accurate predictions for e-commerce product reviews.

Keywords: sentiment analysis, e-commerce, NLP, machine learning, k-NN, Imbalanced data