ABSTRACT

OPTIMIZING NAÏVE BAYES-DRIVEN BUKALAPAK SENTIMENT ANALYSIS ON TWITTER USING DATA ANALYSIS AND MODEL REFINEMENT TECHNIQUES

By

N.A. Irsan

19/440456/PA/19086

Sentiment analysis for the Indonesian stock market using Twitter data presents challenges related to data availability, language barriers, and cultural differences. This research aims to comprehensively analyze sentiment in the context of Bukalapak, focusing on optimizing the sentiment analysis results. The study utilizes Tweepy to access the Twitter API for data retrieval. Two machine learning base models, Multinomial Naïve Bayes (MNB) and Bernoulli Naïve Bayes (BNB), are employed to create an optimized, hyperparameter tuned model.

Based on a gathered dataset with the keyword "Bukalapak," the study found that the Bukalapak elicits predominantly positive sentiments, followed by neutral and negative sentiments. When tested using the same dataset, the Bernoulli Naïve Bayes model achieved better performance accuracy (81.8%) compared to the Multinomial Naïve Bayes model (71%). Through hyperparameter tuning, the optimized version of the Bernoulli Naïve Bayes model achieved the best performance with an accuracy score of 95.3%.

This study contributes to understanding sentiment on Twitter within the context of Bukalapak. It provides insights into the sentiment distribution towards the Bukalapak and highlights the superior performance of the hyper tuned Bernoulli Naïve Bayes model for sentiment analysis. These findings offer valuable guidance for data-driven decision-making and effective sentiment analysis in similar contexts.

Keywords: Sentiment Analysis, Naïve Bayes Classifiers, Twitter