

ABSTRACT

Rio Tamyiz Maulana, 11120017

COMPARATIVE PERFORMANCE OF NAÏVE BAYES CLASSIFIER AND SUPPORT VECTOR MACHINE ON SENTIMENT ANALYSIS OF THE 2024 ELECTION USING SEMMA METHOD

Undergraduate Thesis, Department of Information Systems, Faculty of Computer Science and Information Technology, Gunadarma University, 2024.

Keywords: Sentiment Analysis, Machine Learning, Naïve Bayes, SVM.

(xii – 74 – attachment)

This research collects and documents tweets regarding the 2024 election using a Python scraper with 24 selected keywords, implementing a lexicon-based labeling method to classify tweets into positive and negative sentiment categories. The SEMMA (Sample, Explore, Modify, Model, Assess) methodology is employed to analyze public sentiment on Twitter. The initial dataset of 24,305 tweets is reduced to 20,480 after the cleaning process, which will be used in the analysis phase. The labeling results indicate that 78% (15,993 tweets) contain negative sentiment, while 21.9% (4,487 tweets) express positive sentiment. The Support Vector Machine (SVM) exhibits higher performance accuracy compared to Naïve Bayes across all three data partitioning scenarios, with the highest accuracy reaching 0.918 in the 85:15 data split. Conversely, Naïve Bayes consistently shows lower accuracy, achieving a maximum of 0.801 in the 55:45 data split and a minimum of 0.797 in the 85:15 split. Overall, SVM is recommended as a more effective algorithm for sentiment analysis of tweets related to the 2024 election, providing significant contributions to understanding public sentiment and data-driven decision-making.

References (1900-2024)