

ABSTRACT

My Tel-U is a private social media application used by the Telkom University academic community to access academic information and express opinions through its timeline feature. Posts shared on the timeline often reflect users' sentiments toward campus services or the environment. However, previous sentiment analysis approaches have relied on small datasets and traditional methods that struggle to handle informal and mixed-language Indonesian texts. This study aims to address these limitations by evaluating the effectiveness of Large Language Models (LLMs), including BERT, RoBERTa, and IndoBERT, in comparison to traditional models like Support Vector Machine (SVM), for sentiment analysis on imbalanced and noisy data from My Tel-U. The research was conducted through three experimental scenarios, which include examining the effect of different oversampling techniques (SMOTE, random oversampling, and no oversampling) on IndoBERT, the performance comparison among LLMs, and comparison between LLMs and SVM using TF-IDF and Word2Vec representations. This study is among the first to systematically explore the use of Indonesian LLMs for sentiment classification in private educational social media platforms. Results show that IndoBERT outperforms all other models with an F1-Macro score of 0.8423, achieving strong performance without requiring oversampling. This indicates that pretraining on an Indonesian corpus significantly enhances contextual understanding. In contrast, English-based LLMs and traditional SVM models demonstrate lower performance due to their limitations in handling Indonesian linguistic structures. These findings demonstrate the practical potential of integrating language-specific LLMs into institutional feedback and monitoring systems, enabling universities to better understand and respond to user sentiments.

Keywords: sentiment analysis, large language models, indobert, imbalanced data, my tel-u