

## ABSTRACT

---

*Social media has become a tool for citizens to express their opinions and comments. In this study, we followed a sentiment analysis approach to classify Instagram comments. Our methodology involved constructing an Indonesian-based social media data dictionary, collecting social media data, performing data pre-processing, training and classification, and implementing the model on social media. We categorized the comments into several categories, namely, health, education, administration, economy, infrastructure, citizenship, social and environment. We built an Indonesian-based social media data dictionary taken from news portal, comprising 2,205 records. Additionally, we scraped 800 Instagram comments. Subsequently, we pre-processed the dataset and trained it using two models: multinomial logistic regression and gradient boosting. Our experimental results show that the Gradient Boosting model demonstrated higher accuracy than the Multinomial Logistic Regression, both on the training and testing dataset: 98.07% and 73.11%, respectively. The results indicate that the Gradient Boosting is more reliable than the Multinomial Logistic Regression for our small dataset. However, we identified that our model misclassified certain comments into different categories. This result need for further analysis to improve the accuracy and precision of the models capabilities.*

*Keywords: social media, sentiment analysis, logistic regression, gradient boosting.*