

# A Comparative Study on Handling Imbalanced Data in Indonesian Hate Speech Detection Using FastText and BiLSTM

Akmal Muhamad Faza<sup>1\*</sup>, Yuliant Sibaroni<sup>2</sup>, Sri Suryani Prasetyowati<sup>3</sup>

<sup>1,2,3</sup>*School of Computing, Telkom University  
Bandung, Indonesia*

[akmalfaza@student.telkomuniversity.ac.id](mailto:akmalfaza@student.telkomuniversity.ac.id)

## Abstract

Online hate speech poses a significant threat to social harmony in Indonesia, necessitating effective automated detection systems. This study addresses the challenge of data imbalance, a common issue in hate speech datasets, by developing a Bidirectional Long Short-Term Memory (BiLSTM) model with FastText word embeddings. We systematically compare three oversampling techniques—Random Oversampler, SMOTE, and ADASYN—across varying degrees of imbalance in the Indonesian Hate Speech Superset dataset (14,306 comments), a gap in existing literature. Evaluated using Stratified K-fold Cross-Validation with Accuracy, Precision, Recall, and F1-score, our results indicate that oversampling generally enhances model performance, particularly for the minority class. The optimal oversampling strategy depends on imbalance severity: SMOTE achieved the best balance trade-off within Recall (78.9%) and F1-score (75.3%) on the original dataset, while Random Oversampling was superior for severely imbalanced scenarios, reaching F1-scores of 60.6% (30% minority) and 38.6% (10% minority). These findings offer vital insights for building more adaptive hate speech classification systems in the Indonesian context with imbalanced data distribution.

**Keywords:** BiLSTM, Deep Learning, FastText, Hate Speech, Imbalanced, Oversampling