

ABSTRACT

Purpose: This research aims to develop a system for detecting the credibility of information on social media X by classifying tweets as credible or non-credible. Additionally, it seeks to improve the accuracy of classification and prediction of information credibility using feature extraction methods, semantic features, feature expansion, and optimization.

Methods: The system is built using a deep learning approach with Long Short-Term Memory (LSTM), Term Frequency-Inverse Document Frequency (TF-IDF), Robustly optimized BERT Approach (RoBERTa), Global Vector (GloVe), and Particle Swarm Optimization (PSO). The dataset consists of 54,766 Indonesian tweets from social media X, focusing on the 2024 General Election and using several keywords such as 'Pemilu 2024', 'Pilpres 2024', 'anies baswedan', 'Prabowo', '#GanjarPranowo', and '#debatCapres'.

Result: The results of this study show that the highest accuracy achieved is 89.09% using LSTM with an 80:20 data split, baseline unigram, RoBERTa, Top1 corpus IndoNews, and PSO of the LSTM model's hyperparameters, resulting in a highly significant statistical improvement of 0.96% over the baseline model.

Novelty: This research contributes to information credibility classification research using RoBERTa to add semantic features and GloVe to expand features by utilizing a built corpus and finding similar words to connect with these expanded features. Additionally, PSO is applied to find the optimal hyperparameters, thereby improving the performance and accuracy of the LSTM classification model.

Keywords: Information Credibility, Social Media X, GloVe, LSTM, PSO