# Mul-GAD vs. Shallow Learning Methods on CORA and Citeseer Benchmarks

Assignee Research

May 31, 2026

**Abstract**

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the performance gap between Mul-GAD and traditional shallow learning methods (e.g., SVM, Random Forest) on benchmark datasets like CORA or Citeseer when measured by F1-score and AUC-ROC. Sentiment analysis of product reviews on e-commerce platforms plays a critical role in automatically understanding customer satisfaction and providing actionable insights for sellers seeking to improve product quality. This paper presents a comprehensive benchmarking study. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Benchmarking Logistic Regression, SVM, and LightGBM Against BiLSTM with Attention for Sentiment Analysis on Indonesian Product Reviews. Research question: What is the performance gap between Mul-GAD and traditional shallow learning methods (e.g., SVM, Random Forest) on benchmark datasets like CORA or Citeseer when measured by F1-score and AUC-ROC?.

## 2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

## 3 Results

10 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 3.8/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## 5 Extracted Claims

| Claim | Verified | Confidence |
|---|---|---|
| PyCaret setup() automatically handled missing value imputation and MinMax feature normalization prior to 10-fold cross-v | × | 0.06 |
| Hyperparameter optimization for the Deep Learning model was conducted over multiple trials; the best configuration (Tria | × | 0.03 |
| The Deep Learning model was trained for up to 10 epochs with early stopping based on validation loss, reaching an optima | × | 0.04 |
| All experiments were executed using Google Colab with GPU acceleration. | × | 0.03 |
| Both ML and DL models are evaluated using Accuracy, Precision, Recall, and F1-Score (macro-averaged). | × | 0.09 |
| For the ML models, metrics are averaged across 10 cross-validation folds. | × | 0.05 |
| For the DL model, metrics are computed on the held-out test set of 3,946 samples. | × | 0.08 |
| Logistic Regression emerged as the best-performing model overall in the 10-fold cross-validation results for all three M | × | 0.10 |
| The PyCaret classification module was used to orchestrate the ML experimental pipeline with train_size=0.8 and automatic | × | 0.03 |
| Three algorithms were benchmarked: Logistic Regression (LR), Support Vector Machine (SVM) with a linear kernel, and Ligh | ✓ | 0.29 |
| All ML models were evaluated using 10-fold Stratified Cross-Validation on the training set, ensuring each fold maintains | × | 0.09 |
| The Deep Learning architecture was developed within the PyTorch framework, trained from scratch without pre-trained embe | × | 0.03 |
| The Deep Learning architecture consists of an Embedding Layer with a vocabulary size of 13,600 and dense vectors of dime | × | 0.04 |

## References


- http://arxiv.org/abs/2604.25452v1

- http://arxiv.org/abs/2106.16020v1
- http://arxiv.org/abs/2212.05478v1