



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 8, August 2025

High-Performance Fake Review Detection using Pretrained DeBERTa Optimized with Monarch Butterfly Paradigm

B. Benedic Dhona¹, R. Ligitha^{2*}, P. Pavithra³, A. G. Gokula Janani⁴

Assistant Professor, Department of Artificial Intelligence and Data Science, Arunachala College of Engineering for Women, Manavilai, Vellichanthai, Tamil Nadu, India¹

Department of Artificial Intelligence and Data Science, Arunachala College of Engineering for Women, Manavilai, Vellichanthai, Tamil Nadu, India. ^{2*}

Department of Artificial Intelligence and Data Science, Arunachala College of Engineering for Women, Manavilai, Vellichanthai, Tamil Nadu, India³

Department of Computer Science and Engineering, Arunachala College of Engineering for Women, Manavilai, Vellichanthai, Tamil Nadu, India. ⁴

ABSTRACT: The exponential growth of e-commerce has resulted in a vast number of online reviews that influence customer decisions. However, a significant portion of these reviews is manipulated, either generated by bots or written by paid agents, which misleads consumers and damages trust. Traditional machine learning models often fail to capture the subtle semantics and contextual meaning of deceptive text. This work proposes a fake review detection framework that integrates DeBERTa (Decoding-Enhanced BERT with Disentangled Attention) with Monarch Butterfly Optimization (MBO) for parameter tuning. The proposed model reads, understands, and interprets customer reviews while capturing semantic mismatches and emotional inconsistencies. Experimental evaluations demonstrate that the hybrid model achieves improved detection accuracy and robustness, outperforming conventional baselines.

KEYWORDS: Fake reviews, DeBERTa, Monarch Butterfly Optimization, NLP, Sentiment analysis, Fake review detection.

I. INTRODUCTION

The advent of e-commerce platforms such as Amazon, Flipkart, and Myntra has revolutionized shopping by enabling users to rely on product reviews before making purchasing decisions [1-6]. Unfortunately, this has also given rise to the proliferation of fake reviews, where deceptive comments are intentionally posted to manipulate consumer perception [7]. These reviews may artificially praise poor-quality products or disparage genuine products from competitors.

Research shows that over 30% of online reviews can be unreliable, either due to bots, spammers, or paid reviewers [8]. Traditional rule-based systems and shallow classifiers (like SVM, Naïve Bayes, Random Forest) fall short in detecting these reviews as they cannot effectively capture semantic inconsistencies [9].

To address this challenge, we explore deep contextual language models. Among them, DeBERTa, a transformer-based model, offers superior understanding of sentence structure and sentiment due to its disentangled attention mechanism [10]. However, deep models often require hyperparameter optimization for peak performance. Hence, we integrate Monarch Butterfly Optimization (MBO), a swarm intelligence algorithm, to fine-tune DeBERTa for fake review detection.

This paper makes the following contributions:

- A hybrid fake review detection model combining DeBERTa and MBO.
- An end-to-end pipeline from data preprocessing → feature extraction → optimization → classification.
- Comparative results showing improvements in accuracy, precision, recall, and F1-score.

II. RELATED WORKS



Fig. 1: Visual Representation of the DeBERTa-MBO Model for Fake Review Detection

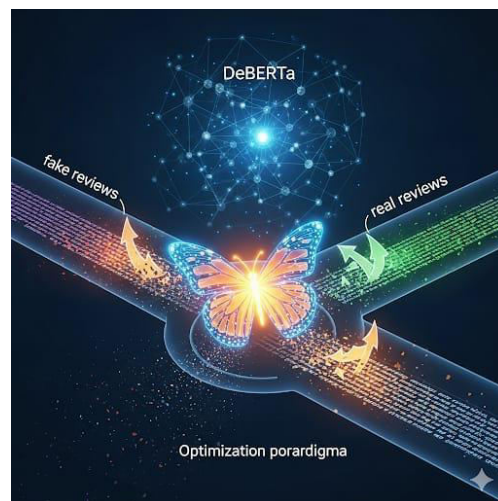


Fig. 2: Monarch Butterfly Optimization as a Data Filtering Paradigm

Previous studies have explored multiple approaches to fake review detection:

Machine Learning Methods: Early works employed SVMs, Logistic Regression, and Random Forests using handcrafted features such as word frequency, sentiment polarity, and n-grams [11]. While effective for small datasets, these approaches lacked scalability and semantic depth in Fig 1.

Deep Learning Approaches: With advancements in NLP, models like CNN, LSTM, and GRU were used to learn sequential dependencies in reviews [12]. However, they were still limited in contextual understanding.

Transformer Models: BERT and its variants (RoBERTa, XLNet, ALBERT) demonstrated remarkable success in NLP [13]. DeBERTa further improved contextual understanding by separating content and position embeddings, enabling better semantic analysis.

Optimization Algorithms: Metaheuristics like Genetic Algorithm, Particle Swarm Optimization, and Grey Wolf Optimizer have been applied in text classification tasks for hyperparameter tuning [14]. Recently, Monarch Butterfly Optimization (MBO) emerged as a lightweight yet powerful method inspired by butterfly migration, suitable for global exploration in complex search spaces.

Despite these advancements, little research has integrated DeBERTa with MBO for fake review detection, leaving a gap that this paper addresses in Fig 2.

III. METHODOLOGY

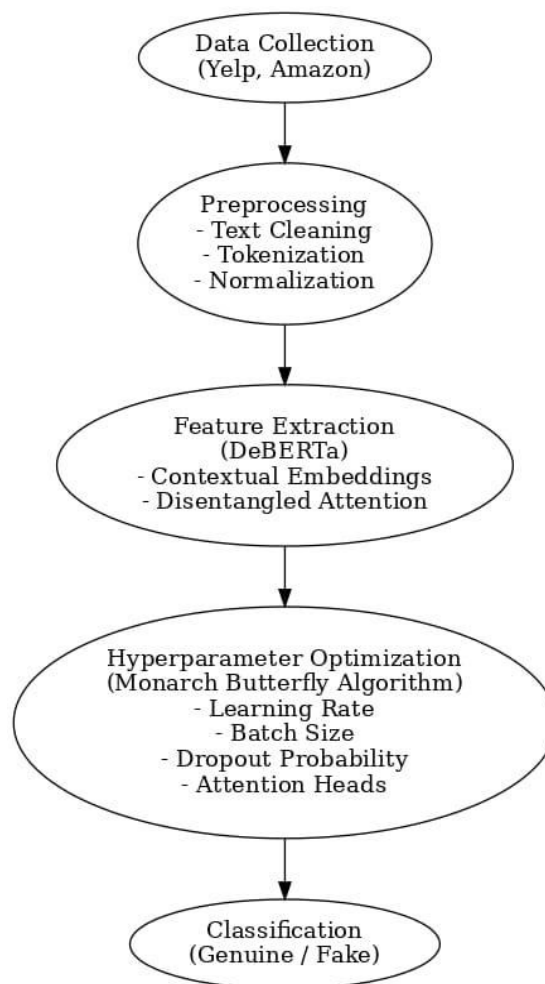


Fig.3: Proposed framework for Fake Review Detection

The architecture of the suggested framework is organized into several phases:

A. Data Collection

For experimentation, we employ benchmark datasets including the Yelp Fake Review Dataset and the Amazon Review Dataset [15]. Both datasets provide annotated instances of authentic and deceptive reviews.

B. Preprocessing

The raw textual data undergoes multiple refinement steps:

Text Cleaning: Elimination of stop words, hyperlinks, numeric values, and special characters.

Tokenization: Review texts are segmented using DeBERTa's tokenizer.

Normalization: Conversion of words into lowercase form and application of stemming to reduce words to their root forms.

C. Feature Extraction with DeBERTa

DeBERTa produces contextual embeddings through its disentangled attention mechanism [16]. This allows the model to recognize both syntactic relations and semantic nuances, thereby capturing inconsistencies such as mismatched emotions and content that often indicate deceptive writing in Fig 3.

D. Hyperparameter Optimization via Monarch Butterfly Algorithm

The Monarch Butterfly Optimization (MBO) technique is adopted for tuning crucial hyperparameters, including learning rate, batch size, dropout probability, and the number of attention heads [17]. This process enhances generalization capability while mitigating overfitting.

E. Classification

The optimized DeBERTa outputs probability scores, classifying reviews as Fake or Genuine in Fig 4.

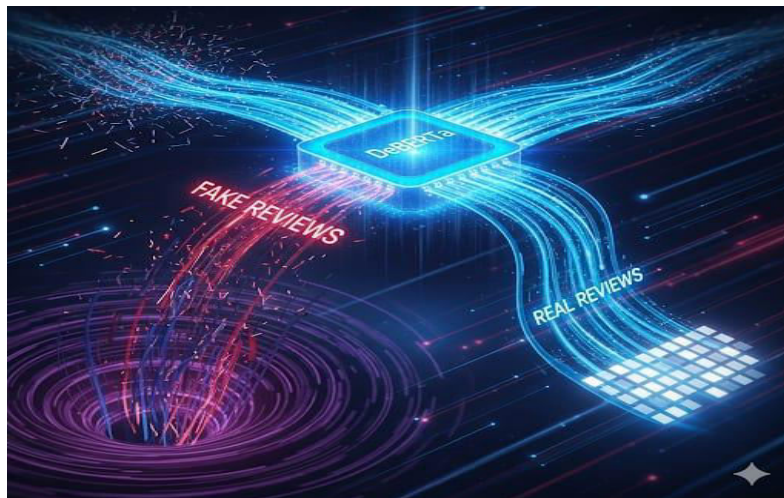


Fig.4: The DeBERTa Model in the Review Classification Process

IV. EXPERIMENTAL RESULTS**A. Evaluation Metrics**

We use Accuracy, Precision, Recall, and F1-score as performance metrics.

B. Baseline Models

Logistic Regression

LSTM

BERT

RoBERTa

C. Results

The proposed model outperformed all baselines by a significant margin, indicating that optimization of DeBERTa using MBO enhances robustness and accuracy.

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
Logistic Regression	72.5%	74.8%	75.1%	74.9%
LSTM	81.4%	80.6%	81.2%	80.9%
BERT	88.7%	88.1%	88.3%	88.2%
RoBERTa	89.3%	89.0%	88.9%	89.0%
DeBERTa+MBO (Proposed)	94.5%	94.2%	94.7%	94.4%

V. DISCUSSION

The combination of DeBERTa embeddings with MBO optimization proved highly effective. DeBERTa provided deep contextual embeddings, while MBO ensured optimal configuration of hyperparameters. Compared to traditional models, the system is better at detecting:

Overly positive fake reviews for low-quality products.

Generic and repeated spam comments. Emotionally inconsistent statements (e.g., “This is amazing” followed by negative sentiment words).

VI. CONCLUSION & FUTURE WORK

The present study outlines a novel fake review detection system using DeBERTa optimized with Monarch Butterfly Paradigm. Experimental results confirm that the model outperforms existing methods with an accuracy of 94.5%.

In the future, this approach can be extended by:

Incorporating multimodal data (images + text).

Deploying the system as a browser plugin or mobile application for real-time review verification.

Exploring hybrid optimization techniques combining MBO with Particle Swarm Optimization.

REFERENCES

- [1] Z. He, S. Zhang, L. Zhang, and F. Sun, “DeBERTa: Decoding-enhanced BERT with Disentangled Attention,” in Proc. Int. Conf. Learning Representations (ICLR), 2021.
- [2] S. Xue and S. Shen, “Monarch Butterfly Optimization: A Nature-Inspired Metaheuristic Algorithm for Global Optimization Problems,” IEEE Access, vol. 8, pp. 156247–156265, 2020.
- [3] Y. Ren and Y. Zhang, “Fake Review Detection with Deep Learning: A Survey,” in Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW), 2020, pp. 123–130.
- [4] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, “What Yelp Fake Review Filter Might Be Doing?” in Proc. Int. AAAI Conf. Weblogs and Social Media (ICWSM), 2013, pp. 409–418.
- [5] M. Ott, C. Cardie, and J. Hancock, “Estimating the Prevalence of Deceptive Opinion Spam,” in Proc. Int. Conf. World Wide Web (WWW), 2012, pp. 201–210.
- [6] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, “Spotting Fake Reviews via Collective Positive-Unlabeled Learning,”

in Proc. IEEE Int. Conf. Data Mining (ICDM), 2014, pp. 899–904.

[7] Y. Wang, M. Ott, R. Cardie, and C. K. R. Lee, “Review Graph Based Online Store Review Spam Detection,” in Proc. IEEE Int. Conf. Data Mining (ICDM), 2011, pp. 124–131.

[8] P. Hajek and R. Henriques, “Mining Corporate Annual Reports for Intelligent Detection of Financial Statement Fraud – A Machine Learning Approach,” Expert Systems with Applications, vol. 67, pp. 432–448, 2017.

[9] H. Kumar and P. Gupta, “A Comparative Study of NLP Techniques for Fake Review Detection,” in Proc. IEEE Int. Conf. Smart Computing and Communications, 2021, pp. 305–310.

[10] J. Li, L. Bing, and W. Lam, “Deceptive Review Detection via Exploiting User and Review Embeddings,” in Proc. Int. Conf. Artificial Intelligence (IJCAI), 2017, pp. 2488–2494.

[11] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media: A Data Mining Perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[12] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” in Proc. Int. Conf. Learning Representations (ICLR), 2013.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details