

Sarcasm detection in social media using deep contextual models

Nayana M R¹, Kumar Siddamallappa U¹, Neelamma G² and Sowmya P^{3,*}

¹ Department of Computer Applications (MCA) Davangere University, Davangere

² Department of central library, Davangere University, Davangere

³ Department of studies in Computer Science Davangere University, Karnataka, India.

Kumar Siddamallappa U; Orcid :-0000-0002-1978-3868

Neelamma G; Orcid :-0000-0003-1824-4858

Sowmya P; Orcid :-0009-0007-9059-9262

World Journal of Engineering and Technology Research, 2026, 04(01), 001–007

Publication history: Received on 14 June 2026; revised on 20 July 2026; accepted on 22 July 2026

Article DOI: <https://doi.org/10.53346/wjetr.2026.4.1.0023>

Abstract

Social media platforms are filled with millions of posts created by users that express opinions, emotions, and experiences. Many of these posts have sarcasm in them where the intended meaning is different from the literal meaning, which makes sentiment analysis difficult. Conventional machine learning approaches often fail to capture the contextual and semantic information required for accurate sarcasm detection. Here we propose a deep contextual model based on RoBERTa and LSTM to identify sarcastic and non-sarcastic comments from social media text. The system performs text preprocessing, contextual embedding generation and classification to enhance the accuracy of sarcasm detection. Experimental results show that deep contextual models can outperform traditional machine learning methods and improve sentiment analysis by effective identification of hidden emotions

Keywords: Sarcasm Detection; Sentiment Analysis; Social Media Analysis; Hidden Emotion Recognition

1. Introduction

Social media sites like Twitter, Reddit and Facebook have become major sources of public opinion. People often use sarcasm to express emotions where the literal meaning is opposite to the intended one. Sarcasm detection is an important task to improve sentiment analysis, opinion mining, customer feedback analysis and cyberbullying detection. However, sarcasm detection is still a challenging Natural Language Processing (NLP) task as it requires context, semantics and linguistic patterns. Large context-aware language models such as BERT and RoBERTa have dramatically improved NLP performance by capturing the context of words. The work presented here proposes a combination of RoBERTa and LSTM to increase the accuracy of sarcasm detection in social media text.

Sarcasm detection has therefore become an important research area in NLP. It plays a significant role in improving sentiment analysis, opinion mining, customer feedback analysis, social media monitoring, cyberbullying detection, political opinion analysis, and mental health assessment. Detecting sarcasm enables organizations to better understand customer opinions, monitor public sentiment toward products and services, and make more informed decisions based on user-generated content.

1.1. Problem Statement

Current sentiment analysis systems typically classify text by the literal sense and do not perform sarcasm detection. Consequently, positive statements with negative intent and negative statements with positive intent are often

* Corresponding author: Sowmya P

misclassified. Conventional machine learning techniques rely on manually engineered features, and cannot efficiently extract the contextual associations among words. Hence, there is a need for an intelligent sarcasm detection system that can accurately understand the contextual meaning and identify the hidden emotions.

1.2. Research Objectives

- Detect sarcastic and non-sarcastic comments accurately.
- Identify emotions that are hidden in social media text.
- Deep contextual models for better sentiment analysis.

2. Literature Review

KALaivani and Thenmozhi [1] Proposed a BERT-based sarcasm detection model that uses conversational context (previous messages along with the current message). They fine-tuned BERT to classify comments as sarcastic or non-sarcastic. Burauah et al. [2] Developed a context-aware BERT model that combines the target sentence with surrounding conversation to better understand sarcasm. Parameswaran et al [3] introduced BERT based sarcasm target detection. Achieved about 77–80% accuracy Khatri and Pranav [4] used BERT and GloVe embeddings for detecting sarcasm in Twitter. BERT achieved around 86–89% accuracy, while GloVe-based models achieved about 78–82% accuracy. Yaghoobian et. al [5] Compared multiple deep learning models such as CNN, LSTM, BiLSTM, and BERT for sarcasm detection. Found that BERT produced the best results, with accuracy around 89–91%, outperforming CNN and LSTM models. Aditya Joshi, Pushpak Bhattacharyya [6] Mark James Carman – Automatic Sarcasm Detection: A Survey. A foundational survey covering datasets, methods, and challenges in sarcasm detection. Hamed Yaghoobian, Hamid R.[7] Arabnia, Khaled Rasheed (2021) – Sarcasm Detection: A Comparative Study. Reviews modern sarcasm detection techniques and compares different approaches.

2.1. Research Gap

Most of the previous researches are on contextual embeddings or sequential learning The fusion of RoBERTa and LSTM is able to improve sarcasm detection by capturing contextual semantics and sequential dependencies.

3. Proposed Methodology

We suggest a system that classifies social media comments into sarcastic and non-sarcastic using RoBERTa and LSTM. The methodology includes data collection, preprocessing, contextual embedding generation, model training, testing, and evaluation.

3.1. Dataset Description

The dataset consists of sarcastic and non-sarcastic comments collected from publicly available social media datasets such as Twitter and Kaggle. Each comment is labeled as sarcastic (1) or non-sarcastic (0). The dataset is divided into training, validation, and testing sets.

3.2. Text Preprocessing

- Remove URLs
- Remove hash tags
- Remove special characters
- Convert text to lowercase
- Remove stop words
- Tokenization
- Padding
- Text normalization

3.3. Text Feature Representation

RoBERTa generates contextual embeddings that encode semantic relations between words instead of handcrafted features. These embeddings are then fed to an LSTM network for sequential learning and classification.

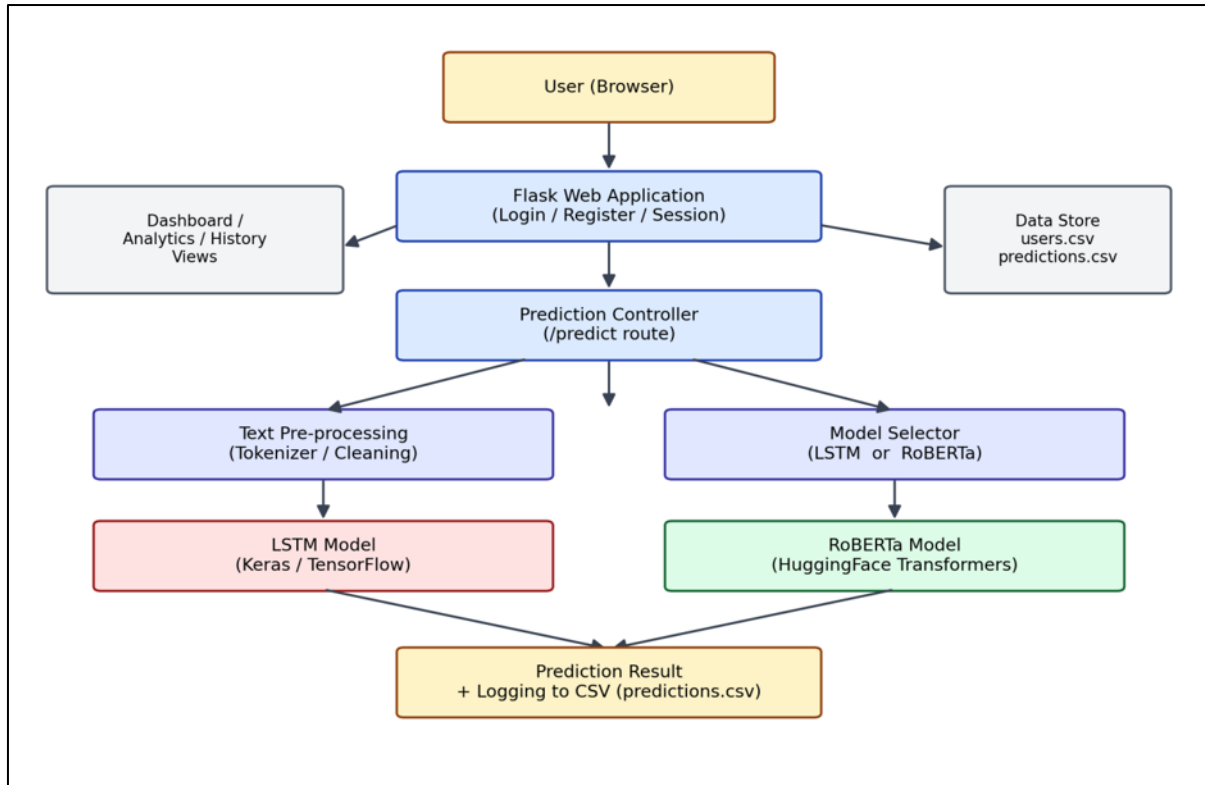


Figure 1 The Data flow diagram of Text Feature representation

3.4. Detection and Classification

- **Input text** : The system receives a text comment or sentence from the user for analysis. Example: *"Great! Another Monday morning."*
- **Text preprocessing**: The text is cleaned, tokenized, and converted into numerical IDs that the model can understand. Example: *"Great! Another Monday morning."*
- **RoBERTa contextual embedding**: RoBERTa converts each word into context-aware embeddings by considering the meaning of surrounding words. Example: In *"Great! Another Monday morning."*, RoBERTa understands that **"Great"** is likely being used sarcastically.
- **LSTM feature learning**: The LSTM learns the sequence and relationships between words to identify sarcastic patterns. Example: It recognizes that **"Great"** followed by **"Another Monday morning"** suggests sarcasm rather than praise.
- **Dense layer**: The Dense layer combines the important features extracted by the LSTM to make the final decision. Example: It gives higher importance to words like **"Great"** and **"Monday morning"** together.
- **Softmax/Sigmoid classifie**: The classifier predicts whether the comment is sarcastic or non-sarcastic based on the learned features. Example: **Output**: *Sarcastic* (Probability = **0.95**) or *Non-Sarcastic* (Probability = **0.05**).
- **Output**: Sarcastic, non-sarcastic

3.5. Experimental Setup

3.5.1. Hardware

- Intel Core i5/i7 Processor
- 8 GB RAM
- 256 GB SSD

3.5.2. Software

- Python, TensorFlow, PyTorch
- Flask
- Jupyter Notebook

- VS Code/Spyder

3.6. Performance Evaluation Metrics

Accuracy: Accuracy measures the overall percentage of correctly classified sarcastic and non-sarcastic comments. Example: If 90 out of 100 comments are predicted correctly, the accuracy is **90%**.

$$\text{Formula: Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: Precision measures how many comments predicted as **sarcastic** are actually sarcastic. Example: If the model predicts 50 comments as sarcastic and 45 are correct, the precision is **90%**.

$$\text{Formula: Precision} = \frac{TP}{TP+FP}$$

Recall: Recall measures how many of the actual sarcastic comments are correctly identified by the model. Example: If there are 60 sarcastic comments and the model detects 54, the recall is **90%**.

$$\text{Formula: Recall} = \frac{TP}{TP+FN}$$

F1-Score: F1-Score is the harmonic mean of precision and recall, providing a balanced performance measure. Example: If both precision and recall are high (e.g., 90%), the F1-score will also be close to **90%**.

$$\text{Formula: F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{precision} + \text{Recall}}$$

Confusion Matrix: A confusion matrix shows the number of correct and incorrect predictions for each class. Example: It displays True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values.

Table 1 The Confusion Matrix

Actual / Predicted	Sarcastic	Non-Sarcastic
Sarcastic	TP (True Positive)	FN (False Negative)
Non-Sarcastic	FP (False Positive)	TN (True Negative)

ROC Curve: The ROC (Receiver Operating Characteristic) curve shows the model's ability to distinguish between sarcastic and non-sarcastic comments. Example: A curve closer to the top-left corner indicates better performance, while an AUC value close to 1.0 represents an excellent classifier.

$$\text{Formula: True Positive Rate (TPR)} = \frac{TP}{TP+FN}$$

$$\text{Formula: True Positive Rate (FPR)} = \frac{FP}{FP+TN}$$

4. Results and Discussion

The ROBERTa-LSTM model is better compared to standard classifiers like SVM, Naïve Bayes, and Logistic Regression in classifying sarcastic texts. Using the deep contextual embeddings helps the model identify sarcastic texts with a higher accuracy than the other machine learning algorithms. The ROBERTa-LSTM model shows better performance with lower error rates in comparison with other methods.

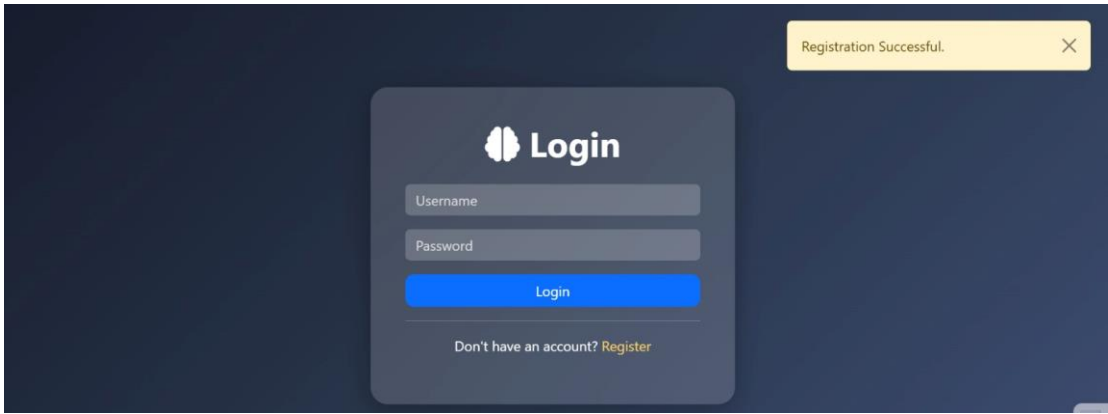


Figure 1 Login Page

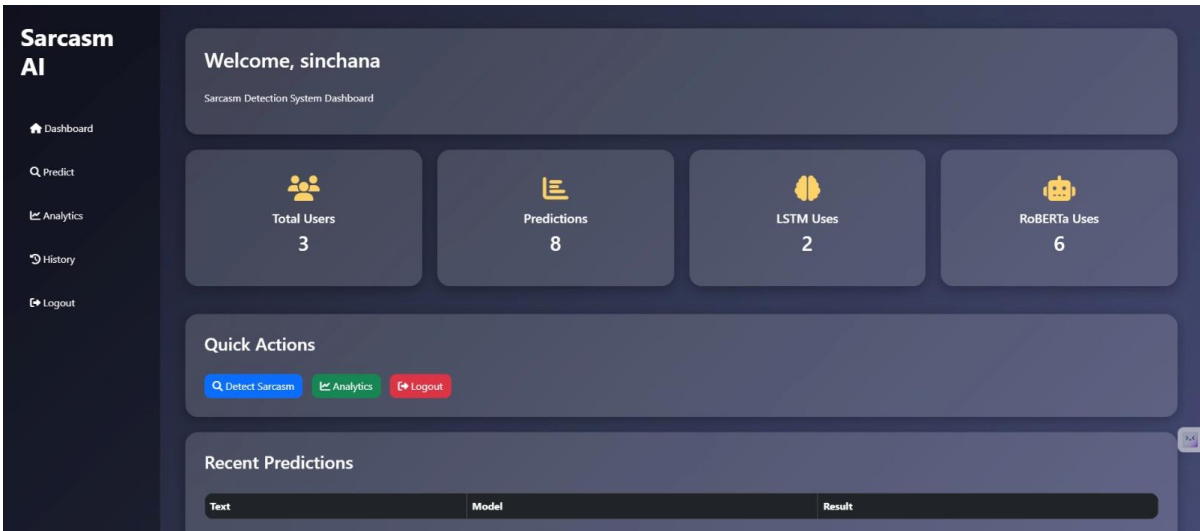


Figure 2 Dashboard Page

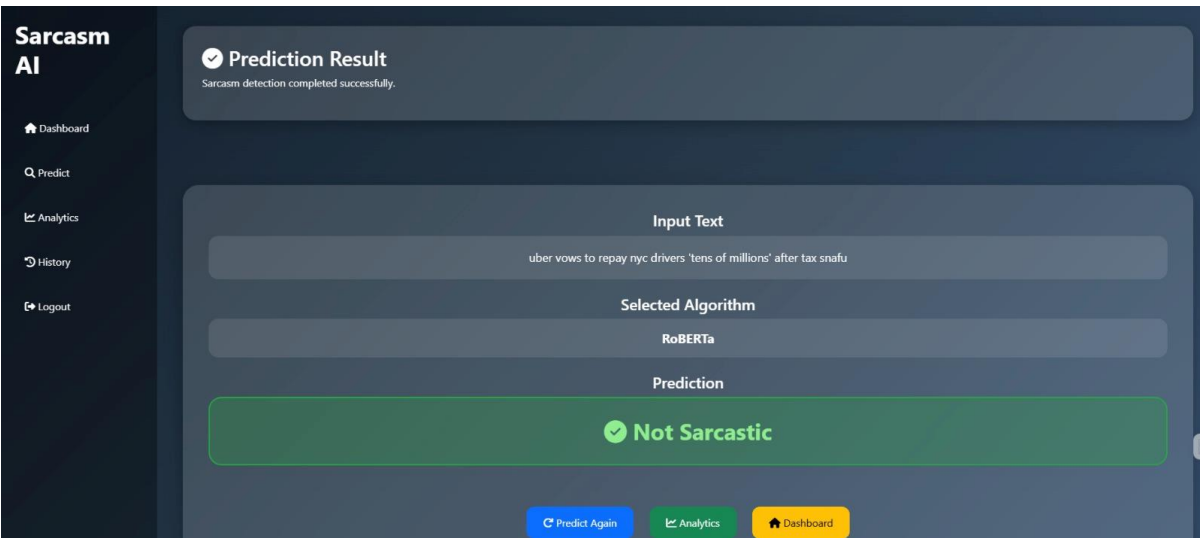


Figure 3 Prediction Result Page

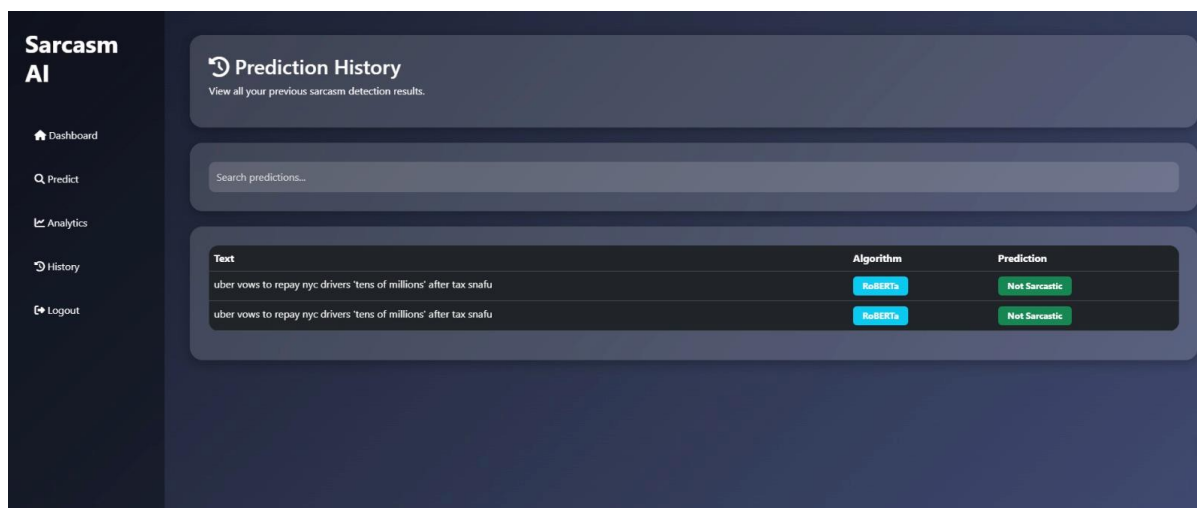


Figure 4 Prediction History Page

5. Conclusion

This research presents a deep contextual sarcasm detection model based on RoBERTa and LSTM that successfully captures the context semantics and sequence information of the input text, thereby improving the accuracy of sarcasm recognition. The proposed method improves the performance of sentiment analysis by mining the hidden emotions contained in the text and reducing the error rate of classification, which can be widely used in customer sentiment analysis, brand monitoring, mental health analysis, and cyberbullying detection.

Future Enhancement

- Support multilingual sarcasm detection.
- Detect sarcasm in images, memes, and videos.
- Optimize the model for real-time social media monitoring.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Kalaivani A., Thenmozhi D. (2020), Sarcasm Identification and Detection in Conversation Context using BERT. This paper compares BERT with traditional machine learning and LSTM approaches for sarcasm detection.
- [2] Arup Baruah, Kaushik Das, Ferdous Barbhuiya, Kuntal Dey (2020) – Context-Aware Sarcasm Detection Using BERT. It explains how conversational context improves sarcasm detection performance.
- [3] Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra, David Eysers (2021). BERT's The Word: Sarcasm Target Detection using BERT. This paper presents transformer-based improvements for sarcasm target detection.
- [4] Elsa Scola, Isabel Segura-Bedmar (2021), Sarcasm Detection with BERT. It compares BiLSTM and BERT models and shows BERT achieving better results.
- [5] Akshay Khatri, Pranav P (2020), Sarcasm Detection in Tweets with BERT and GloVe Embeddings. Focuses on Twitter sarcasm detection using contextual embeddings.
- [6] Aditya Joshi, Pushpak Bhattacharyya, Mark James Carman (2016). Automatic Sarcasm Detection: A Survey. A foundational survey covering datasets, methods, and challenges in sarcasm detection.

- [7] Hamed Yaghoobian, Hamid R. Arabnia, Khaled Rasheed (2021). Sarcasm Detection: A Comparative Study. Reviews modern sarcasm detection techniques and compares different approaches.
- [8] Rafael Jiménez Castro et al. (2026), A Comparative Study of BERT-Based Models for Sarcasm Detection in Social Media Texts. Compares different BERT-based architectures for social media sarcasm detection.