

# Review of: "Comparative Analysis of Machine and Deep Learning Techniques for Text Classification with Emphasis on Data Preprocessing"

Bhavna Bajpai<sup>1</sup>

<sup>1</sup> Dr. C.V. Raman University

Potential competing interests: No potential competing interests to declare.

The review of the research paper titled “Comparative Analysis of Machine and Deep Learning Techniques for Text Classification with Emphasis on Data Preprocessing” by Dr. Saikat Gochhait<sup>1</sup>. The abstract provides valuable insights into the study, and I’ll summarize the key points for you.

## Abstract Summary:

- **Objective:** The research aims to evaluate the performance of different deep learning algorithms for text classification, specifically focusing on issues related to long short-term memory (LSTM).
- **Background:** Physician-written discharge medical notes contain essential information about patients’ health. Deep learning algorithms have proven effective in extracting valuable insights from unstructured medical notes.
- **Dataset:** The study utilizes the Titanic Disaster Dataset for preprocessing. Preprocessing is crucial due to the abundance of unnecessary information in textual data.
- **Data Cleanup:** The authors eliminate duplicate rows and fill in missing values to enhance data quality.
- **Algorithms Evaluated:**
  - Traditional machine learning algorithms: Naive Bayes (NB), gradient boosting (GB), and support vector machine (SVM).
  - Deep learning algorithms: Bidirectional LSTM (BiLSTM) using Conditional Random Fields (CRFs).
- **Findings:**
  - BiLSTM outperforms other models and baseline research, achieving a remarkable classification accuracy of 98.5%.
- **Author Affiliations:**
  - Dr. Saikat Gochhait is affiliated with:
    - Symbiosis Institute of Digital and Telecom Management, Symbiosis International Deemed University, Pune, India.
    - Neuroscience Research Institute, Samara State Medical University, Russia.

In summary, this research sheds light on the effectiveness of deep learning techniques, particularly BiLSTM, for text

classification tasks. The study emphasizes the importance of data preprocessing and highlights the superior performance of BiLSTM in comparison to other models<sup>1</sup>. The keywords associated with this work include machine learning, textual data, bi-directional long short-term memory, deep learning, and text classification.