



PROCEEDINGS OF THE

3RD INTERNATIONAL CONFERENCE IN DATA SCIENCE

Autonomous AI Agents: Redefining Decision-Making in a Data-Driven World

25th and 26th of November 2025 at the University of Colombo, Sri Lanka

ICDS 2025



Organised By:



Department of Statistics
University of Colombo

Universiti Malaya
Malaysia

Silver Partner

Knowledge Partner





**Proceedings
of the
International Conference
in Data Science 2025
(ICDS 2025)**

25th - 26th November 2025
at the
**Department of Statistics, University of Colombo,
Colombo, Sri Lanka**

**ICDS 2025 is organised by the Center for Data Science jointly with
the
Department of Statistics, University of Colombo
and
Universiti Malaya, Malaysia**

Publisher: Center for Data Science, University of Colombo
ISSN: 3030-7163

Fake News Detection Using Text Mining and Web Scraping

M M M T Kavinda¹, R R L U I Rajapaksha^{1*}

¹ Faculty of Computing and Technology, University of Kelaniya, Kelaniya, Sri Lanka
*rasikar@kln.ac.lk

Dissemination of fake news has become a vital threat to the lifestyle of people. This paper presents a comprehensive method for detecting fake news in digital information using a combination of deep learning and machine learning techniques. It includes deep learning models such as Bi-directional Encoder Representations for Transformers (BERT) and a recurrent neural network (RNN) with Global Vectors for Word Representation (GloVe) embeddings, as well as machine learning models including Extreme Gradient Boosting (XGBoost), Naive Bayes classifier, support vector machines, and random forests. The study encompasses the usage of various semantic features of text, such as Term frequency (TF), Term-Frequency-Inverse Document Frequency (TF-IDF), N-gram models, and GloVe embeddings. Unlike prior studies that depend completely on semantic embeddings, this study combines explicit linguistic and psychological features with semantic representations to enhance fake news detection accuracy by capturing text explicit features. The study was conducted on a balanced Kaggle dataset of 6,335 articles and achieved 86% accuracy using the RNN with GloVe embeddings, outperforming baseline models, and contributing a robust hybrid framework applicable to real-time fake news detection.

Keywords: Text mining, Web scraping, Ensemble methods, Semantic features, Latent features, Natural Language Processing.