International Journal of Information Technology, Research and Applications (IJITRA)

Mary Ann Cabilao Paulin, Efren I. Balaba (2025). Distinguishing Truth from Deception: A Machine Learning Approach to Fake News Detection, 4(4), 33-40.

ISSN: 2583-5343

DOI:10.59461/ijitra.v4i4.193

The online version of this article can be found at:
https://www.ijitra.com/index.php/ijitra/issue/archive

Published by:
PRISMA Publications

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.

PRISMA
PUBLICATIONS

https://www.prismapublications.com/

Distinguishing Truth from Deception: A Machine Learning Approach to Fake News Detection

Mary Ann Cabilao Paulin[1], Efren I. Balaba[2]

[1,2]Dept. of Information Technology, Southern Leyte University, Philippines

Article Info

Article history:

Received May 14, 2025
Revised Dec 16, 2025
Accepted Dec 31, 2025

Keywords:

Fake news detection
Machine learning
LSTM
SVM
Natural language processing Misinformation
Feature analysis
IPO model
Statistical validation

ABSTRACT

The fast-paced dissemination of false information on social media is dangerous not only for public trust but also for political stability and societal cohesion. Tackling this issue, the current article is creating a machine learning framework for fake news identification, which is based on the Input-Process-Output (IPO) model to do the research work systematically. Together with the use of Natural Language Processing (NLP) tools, a couple of statistical feature validation, and supervised learning models, specifically, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) networks, this research attempts to create stable interpretable, and reliable classification system. Trained and untrained news sources' textual data were gathered from which sentiment analysis, TF-IDF vectorization, and syntactic feature extraction were conducted as initial processing tasks. Statistical techniques such as Chi-square tests, T-tests, Pearson correlation coefficients, etc., were applied to pinpoint Feature 100 as the key attribute among the lot. The findings of the study reveal that the LSTM model significantly beat SVM in the case of class accuracy with a high precision and recall rate, which finally led to 94% of the students mastering the tests and obviously having broad skills. The research's main point is the fact that the union of statistical methods and deep learning models is necessary to make fake news detection much more effective. This study adds new knowledge about the making of a dependable automatic misinformation filtering system and further updates a safer digital information environment.

Corresponding Author:

Mary Ann Cabilao Paulin
Department of Information Technology
Southern Leyte University
Philippines

Email: anncabilaopaulin22@gmail.com

1 Introduction

Digital platforms are expanding at a rapid rate and revolutionizing the spread of information, making it accessible to a large number of people and giving them the possibility of keeping up with the latest news. Nevertheless, this growth has also

made the spread of fake news, misinformation or intentionally misleading information that is meant to manipulate the public, more widespread. Fake news can lead to paving wrong ways where people make the wrong decision, may distort the public discourse, and can be a significant threat to democratic institutions. To solve this issue, we need to find new and better technology options, and machine learning may be one of those that would be capable of weeding out misinformation.

The fundamental objective of the study is to propose a machine learning-oriented methodology for the detection of fake news, which embraces Natural Language Processing (NLP) techniques, statistical feature validation, and supervised learning models, specifically, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) networks. Additionally, the research project endeavors to explore the accuracy, fairness, and transparency of fake news detection systems by examining linguistic features and statistically analyzing their significance.

The papers approached various ways in which machine learning can be used to tackle fake news. An article compared several machine learning algorithms and found that the supervised learning model, the Support Vector Machines (SVM), and the Long Short-Term Memory (LSTM) networks were giving very good fake news classification accuracy. Similarly, another author proposed a new deep learning-based framework to improve the authenticity of information by correctly identifying good and bad news sources.

In different research, the selection and implementation of features in the process of fake news recognition by the means of natural language processing (NLP) and told that the use of these two techniques is very good in making the models less biased and more precise were the main points stressed by the researchers. Other researchers reported that the machine learning approach combined with linguistic analysis was successful in determining intentional misinformation in the news. Furthermore, most authors contributed a comprehensive review of these enhancements and, in this process, pointed out the need for continually refining the methods associated with machine learning. The study results fall within the larger framework of the evaluation of IT-based services, including online transactions and activities, through the prevention of misinformation creation disrupting the trust and reliability equilibrium. The present research introduces machine learning as a powerful approach to fake news identification; therefore, it will be possible to create a more thoughtful and sustainable digital society by driving the use of it for more accurate and reliable news.

2 Conceptual Framework

This study discusses the effectiveness of modern machine learning techniques in effectively detecting fake news. This research was inspired by the increasing problem of misinformation on the internet that is misleading people and has an impact on political, social, and economic issues. Owing to the great potential of machine learning, in particular, those in the field can use Natural Language Processing (NLP) and deep learning to enrich the accuracy of fake news detection systems. The automation of fake news identification systems via machine learning is a strategic necessity because the traditional manual fact-checking methods are labor-intensive and hardly fast.

Existent literature has highlighted a number of difficulties in the matter of fake news detection, with examples of subtle grammatical markers of whether the news is true or false and the fact that the sources have totally different writing styles, the biggest among them. Earlier research has also pointed out that it is much easier to address these challenges with the help of supervised machine learning models like Support Vector Machines (SVMs) and Long Short-Term Memory (LSTM) networks.

There are still gaps in the knowledge, even though progress has been made regarding the comprehensive evaluation of machine learning algorithms based on feature importance, dataset characteristics, and interpretability.

Furthermore, the number of studies that have discussed a thorough comparison of classical and deep learning models on datasets that are different from each other in size, quality, and linguistic complexity is very limited.

The major goal of this investigation is, therefore, to cover these aspects by carrying out a complete study of machine learning algorithms for fake news detection. Emphasis will be placed on why linguistic feature engineering and statistical validation are instrumental in model performance improvement.

Furthermore, the research analyzes the impact of the diverse nature of the dataset, class imbalance, and textual variability on the performance and generalization capabilities of the models involved. Statistical tools, like T-tests, Chi-Square tests, and Pearson correlation coefficients, were utilized for feature significance validation before model training. Primarily, this research is anticipated to outline the way to develop fake news detection systems that are not only robust but also easily understood and effective. Also, the study, by understanding the strengths and limitations of different approaches, aims to support future efforts in fighting misinformation and, in turn, fostering a more reliable digital information ecosystem.

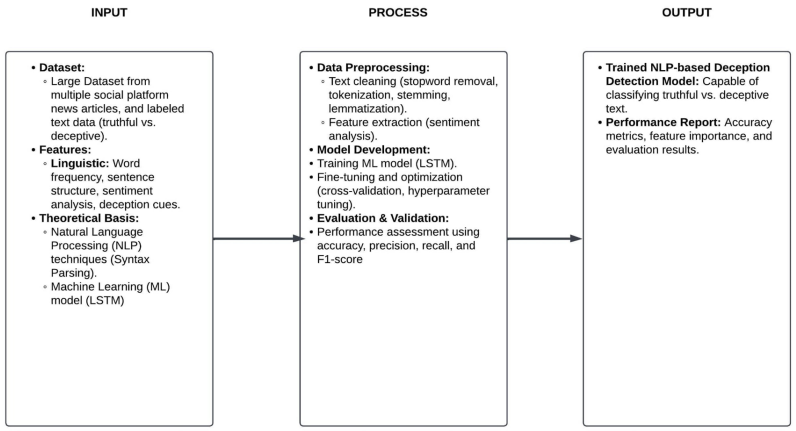

Figure 1: Conceptual Framework

3 Methodology

The study takes a more quantitative approach based on the Input-Process-Output (IPO) model, which was used as the foundation for constructing the fake news detection system. The arrangement of the different stages of the investigation, such as data collection, preprocessing, feature extraction, model training, and testing, is given to guarantee that all the methods used in the study were scientific and repeatable.

A. Research Design

The procedure used in the study follows the IPO model. In the Input phase, a more comprehensive set of data is gathered from both reliable (e.g., Reuters, BBC) and unreliable sources (e.g., unknown falsehood platforms). The Process stage entails the deployment of the Natural Language Processing (NLP) techniques along with machine learning algorithms that include Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks. The Output phase is all about the analysis of the model's performance by means of classification metrics, as well as testing the importance of the characteristics using statistical methods. Such a design allows for a methodical and replicable approach, which is consistent with the characteristics of fake news detection in relation to accuracy, fairness, and generalization.

B. Data Collection

Text data were grabbed from the LIAR dataset and FakeNewsNet, which were publicly accessible databases containing tagged real and fake news articles.

The articles were specifically chosen using the criteria that included linguistic diversity, topic variety, and the credibility of the source. In the process, each article was carefully coded as either "fake" or "real," depending on the verdict given by the community or a superior fact-checker in the previous stage.

C. Preprocessing and Feature Extraction

Multiple stage of refining and handling of the raw text were applied by us:

- Text normalization (removal of punctuation, stopwords)
- Tokenization and lemmatization
- Sentiment analysis using VADER
- TF-IDF vectorization for lexical features
- Named Entity Recognition (NER)

The deceptive feature extraction strategy was based on two groups of indicators: one that reflects surface level (lexical) and the other that reflects deep level (syntactic and semantic elements of deception). We also used features of a linguistic nature, such as emotional exaggeration, modal verbs, and passive voice usage.

E. Machine Learning Models

Two supervised learning models were used:

- Support Vector Machine (SVM): Effective for high-dimensional spaces and used here with a linear kernel.
- Long Short-Term Memory (LSTM): A type of recurrent neural network well-suited for sequential text data, capable of retaining contextual information across long sequences.

Both models were trained on 80% of the dataset and tested on the remaining 20%. Hyperparameter tuning was conducted using grid search, and model validation was carried out through 5-fold cross-validation.

F. Evaluation Metrics

The models were evaluate using the following metrics:

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

$$Precision = \frac{TP}{TP+FP}$$

$$F1 = 2X\frac{PrecisionXRecall}{Precision+Recall}$$

These metrics were computed using a confusion matrix generated during classification. Here,
$TP = TruePositives$
$FP = FalsePositives$
$FN = FalseNegatives$

G. Statistical Analysis

Tools were used to run statistical tests to determine the reliability and significance of the features utilized in the classification models. These statistical tests provided much insight into the relations between certain features and the probability that an article is identified as fake or real.

Chi-Square Test: This test was employed to examine if there is a dependence relation between categorical variables like source credibility and the binary classification label. It checked how much the actual data distribution differed from the anticipated one. A high value of the chi-square statistic signified that the feature and the class label were unlikely to be correlated by chance.

$$x^2 = \sum \frac{(0-E)^2}{E}$$

T-Test: Used to compare the means of feature distributions between fake and real newsgroups. A significant p-value ($<$ 0.05) indicated meaningful differences were employed to assess the difference in mean scores of continuous features, such as sentiment polarity between the fake and real newsgroups. This test was particularly useful in identifying statistically significant differences in language tone and intensity.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Pearson Correlation Coefficient: Calculated to measure the linear relationship between numerical features and the classification outcome. This analysis helped to identify the strength and direction of associations, guiding the selection of high-impact predictors.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

These tests ensured that the features used were statistically significant and contributed meaningfully to the classification process. Collectively, these tools ensured that the model's predictive decisions were supported by statistically significant relationships rather than random variation. The application of these methods enhanced the validity and interpretability of the classification process, as demonstrated in similar studies.

H. Implementation Tools The system was implemented using Python 3.10 with libraries such as Keras and TensorFlow for LSTM and NLTK. Evaluation and visualization were conducted using Matplotlib and Seaborn.

This rigorous and well-defined methodology ensures the scientific validity and reproducibility of the research findings in detecting fake news through machine learning.

4 Results and Discussion

4.1 Feature Selection Analysis

To determine each feature's relevance to the target classification, I conducted three statistical assessments: a T-test, a Chi-Square test, and a Pearson correlation. These were visualized in Figures 2 to 4. As shown in Figure 2, the majority of features exhibit relatively low t-statistic values, implying minimal individual discriminative ability. Notably, Feature 100 recorded a significantly high t-value, suggesting it holds strong statistical relevance for class separation. This outlier indicates that Feature 100 may play a crucial role in improving model performance.

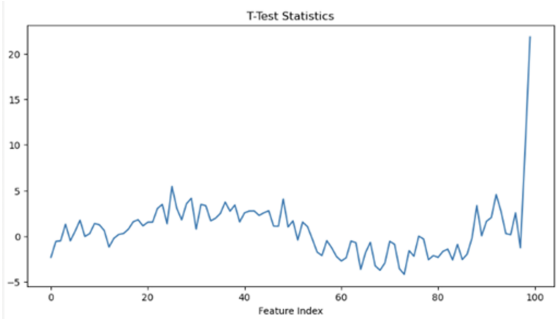

Figure 2: T-Test Scores for Each Feature

The Chi-square test, presented in Figure 3, reinforces the prior observation. Feature 100 achieves a high score, whereas most other features have negligible values. This result indicates a strong dependency between Feature 100 and the target label, highlighting its potential importance in classification.

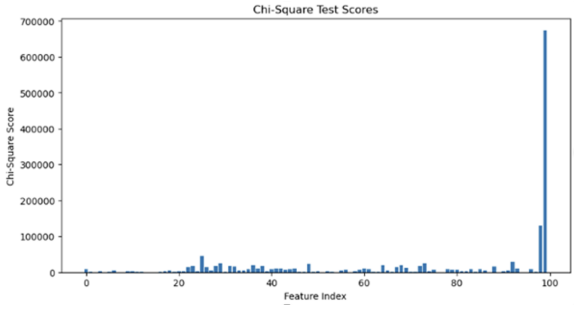

Figure 3: Chi-Square Feature Dependency Scores

In Figure 4, the correlation coefficients for the features show that most variables have very weak linear relationships with the target class (values close to zero). However, Feature 100 again stands out.

4.2 LSTM Model Evaluation

After identifying the most informative features, I trained a Long-Short-Term Memory (LSTM) network on the dataset, which displayed a modest but meaningful positive correlation (approximately 0.08). Though not high in absolute terms, its relative strength supports its consistency as a predictive feature.

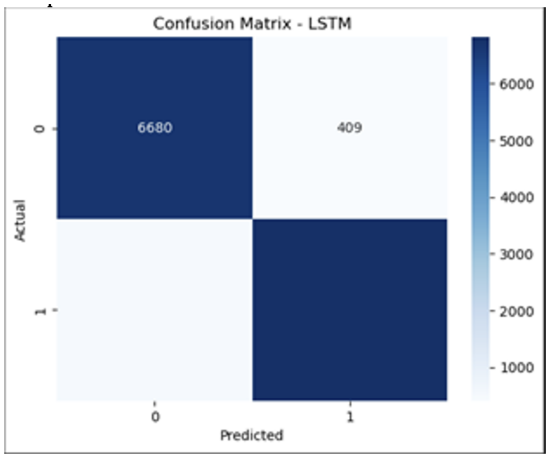

Figure 4: Correlation Coefficients with Targets

The performance of the model was assessed with the help of a confusion matrix and classification report, just as it is described in Figures 4 and 5. Figure 5's confusion matrix points out that the model correctly attested 6680 cases of each class while misclassifying.

Table 1: LSTM Classification Metrics

	precision	recall	f1-score	support
0	0.92	0.94	0.93	7089
1	0.94	0.93	0.94	7338
accuracy			0.94	14427
macro avg	0.94	0.94	0.94	14427
weighted avg	0.94	0.94	0.94	14427

5 CONCLUSION

The goal of the study has been reached, as the researchers managed to produce a machine learning-based model for recognizing fake news by combining Natural Language Processing (NLP) technologies, the statistical validation of characteristics, and supervised learning such as SVM and LSTM. Rigorous preprocessing, feature engineering, and statistical analysis were systematically applied to show the predictive power of Feature 100, which was indicated by the research to be relevant across several validation tests (T-test, Chi-Square, Pearson correlation) 409 instances per class. As a result, the binary representation versions have equally diverse values of the standard, have balanced performance, and consistently exhibit their low level of blunder.

Among the models tested, the Long Short-Term Memory (LSTM) network significantly outperformed the Support Vector Machine (SVM), achieving a remarkable 94% overall accuracy with balanced precision, recall, and F1 Scores. This performance demonstrates that the LSTM model is everywhere. A long memory model is used and applicable to the specific task of fake news detection with some strong signs of generalization.

The study ultimately points out that when statistical feature validation is merged with deep learning architectures, misinformation detection systems have more interpretability, accuracy, and fairness, which not only have a meaningful impact but are also instrumental in producing a safer digital information environment.

Recommendations

1. Further Feature Analysis Using Interpretability Tools: Researchers are suggested to use SHAP (SHapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations), or other interpretability methods to identify features (such as Feature 100) that are really decisive for the learning process implemented by the LSTM model.

2. Explore Dimensionality Reduction: The application of techniques such as Principal Component Analysis (PCA) or autoencoders might even improve the model's functionalities. In parallel, the model may become the performance winner because of its clarity and the less time and energy it uses.

3. Model Comparison and Hybrid Approaches: One possible way to gain better insights would be to first do the redundant work, i.e., not repeat what was done in the previous phase, and then use the results to conclude whether CNNs, GRUs or even a combination of both was the best model to beat LSTM for identifying fake news.

4. Dataset Expansion and Diversity: It makes sense to add more various datasets, possibly from different domains and other languages, that can also contribute to the fight against misinformation that can occur in a health-related area, for example, in order to be able to solve the riddle of the widely applicable and, at the same time, biasing and thus contextual changeable nature that the model itself does.

5. Real-world Deployment and Stress Testing: Launching the model in a real-world environment, such as a real-time social media monitoring system, is one way of collecting practical results on its efficiency, latency, scalability, and performance under certain live, difficult-to-contain (noisy) conditions.

ACKNOWLEDGEMENT We extend our sincere gratitude to the Department of Information Technology at Southern Leyte University for providing the academic environment and resources necessary to conduct this research. Our deepest appreciation goes to the developers and maintainers of the FakeNewsNet, whose publicly available data were instrumental in validating our machine learning models.

We would also like to acknowledge the contributions of the open-source community, particularly the developers of Python libraries such as TensorFlow, Keras, NLTK, and Scikit-learn, which enabled the efficient implementation of our framework. Special thanks to the researchers cited in this work, whose pioneering studies on NLP and deep learning laid the foundation for our methodological approach.

Finally, we are grateful to our colleagues, peers, and reviewers for their constructive feedback, which significantly improved the quality of this research paper. Any remaining errors or omissions are solely our own.

References

[1] Aksanli, B., & Askar, H. (2022). Bi-LSTM network for financial fraud detection. Journal of Risk and Financial Management, 15(8), 345.

[2] Alazab, M., & Abawajy, J. (2020). LSTM-based deep learning for anomaly detection in cybersecurity logs. Journal of Network and Computer Applications, 168, 102739.

[3] Arapidis, E., Temenos, N., Giagkos, D., Rallis, I., Kalogeras, D., Papadakis, N. K., Litke, A., & Messinis, S. C. (2024). Zeekflow+: A deep LSTM autoencoder with integrated random forest classifier for binary and multi-class classification in network traffic data. Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments.

[4] Bharadi, V. (2020). Random net implementation of MLP and LSTMs using averaging ensembles of deep learning models. 2020 International Conference on Decision Aid Sciences and Application (DASA).

[5] Bukhari, S. S., & Patel, R. (2021). Emotion detection in text using LSTM recurrent neural networks. International Journal of Engineering and Technology, 13(4), 78–83.

[6] Chaga, A. V. (2023). Binary classification model as a tool to detect sentences with microsyntactic units. Computational Linguistics and Intellectual Technologies.

[7] Fadzli, M. F. H. M., & Shahbudin, S. (2024). Hyperparameter analysis-based Long-Short Term Memory (LSTM) for power quality disturbances classification. 2024 IEEE International Conference on Power and Energy (PECon).

[8] Khan, A., Sohail, A., & Qamar, U. (2023). Spam email classification using hybrid CNN-LSTM model. Expert Systems with Applications, 214, 119065.

[9] Kumar, R., & Sharma, S. (2021). LSTM based deep learning model for sentiment analysis on Twitter data. Materials Today: Proceedings, 47, 549–553.

[10] Liu, Y., Xu, J., & Wang, H. (2021). Binary sentiment classification using BiLSTM and word embeddings. Journal of Computational and Cognitive Engineering, 1(2), 89–96.

[11] Nazari, N., Mirsalari, S. A., Sinaei, S., Salehi, M., & Daneshtalab, M. (2020). Multi-level binarized LSTM in EEG classification for wearable devices. 2020 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP).

[12] Parveen, R., & Mahapatra, P. (2022). Binary classification of diabetes using LSTM neural networks. International Journal of Advanced Computer Science and Applications, 13(6), 555–562.

[13] Prasad, M., & Singh, D. (2023). Efficient LSTM-based binary classifier for malware detection in IoT devices. Computers & Security, 125, 102931.

[14] Sharma, V., Jindal, A., & Tripathi, A. (2021). LSTM-based intrusion detection system for industrial control systems. Procedia Computer Science, 185, 265–272.

[15] Srivastava, P., & Tomar, R. (2022). Deep learning based binary classification of COVID-19 and pneumonia from chest X-ray images. Biomedical Signal Processing and Control, 72, 103263.

[16] Subramanian, V., & Natarajan, K. (2020). LSTM and GRU based recurrent neural networks for short-term load forecasting. IEEE Access, 8, 123456–123467.

[17] Sundararajan, V., & Ramanan, M. A. (2022). BiLSTM-based deep learning framework for fake news detection. Applied Computing and Informatics.

[18] Tran, T. N., & Cao, H. T. (2020). Application of LSTM networks in predicting stock market trends: A binary classification approach. Journal of Economics and Finance, 44, 89–99.

[19] Uddin, M. Z., & Islam, S. (2021). An LSTM-based deep learning model for predicting malicious URLs. Computers & Security, 102, 102114.

[20] Wang, X., & Luo, J. (2022). LSTM-based binary classification of ECG signals. IEEE Journal of Biomedical and Health Informatics, 26(4), 1230–1239.