Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 7, pp. 2659–2665
DOI : 10.47974/JIOS-2199

Optimization of fake news detection models using advanced machine learning techniques

Ashish R Savaliya *
Amrita Sharma †
Department of Computer Science and Engineering
Indus University
Ahmedabad 382115
Gujarat
India

Monika Arora §
Department of Mathematics
Miranda House
University of Delhi
Delhi 110007
India

Abstract

The reality that social media, blogs, and websites can be utilized by anybody makes a part of issues. Fake news can be a huge issue that can hurt individuals or indeed make struggle between nations. Spread online, deceiving individuals into tolerating something that isn't genuine. It can be unsafe since it can impact what individuals think and do based on wrong data that spreads all over. That's why it's imperative to oversee and control social media. Computer programs that learn can distinguish wrong data. This article proposes organizing a way to discover fake news by analyzing particular parts of the news. The unused innovation can tell the contrast between fake news and genuine news. Lemmatization is the meth-od of finding the fundamental frame of a word employing a reference direct. We found a few imperative focuses by utilizing two distinctive strategies. Take out the rehashed words and the strategy utilized to discover out how vital they are. Vector calculation implies figuring out the comes about of including, subtracting or increasing vectors. We utilized three bunches of information. Accessible online: Fake-or-Real-News, Media-Eval, and ISOT. We utilized eleven diverse ways to total the assignment. The current framework: exactness,

* *E-mail:* savaliyaashish.21.rs@indusuni.ac.in (Corresponding Author)
† *E-mail:* amritasharma.dcs@indusuni.ac.in
§ *E-mail:* monika.arora@mirandahouse.ac.in

the zone under-neath accuracy, assessment, and F1 score. We arrange to progress precision for the Fake-or-Real dataset.

Subject Classification: *Primary 68T07, Secondary 68T50, 62H30, 68Q32.*

Keywords: *Optimization techniques, Machine learning, Fake news, Artificial intelligence, Social media.*

1. Introduction

In today's world, the spread of fake news is a big problem for people's opinions, democracy, and social order. The fast spread of false information online means we need to create good ways to find and stop it. This study looks at different ways to find fake news and compares how accurate and efficient they are. This research aims to see how well these advanced models work at spotting fake news by checking how accurate they are at classifying it and how efficiently they use computer resources. The study compares different methods to find the best ways to fight misinformation online.

Incorrect information can be shared as written words, audio recordings, and images. People share photos, videos, and music on social media sites like Facebook and Twitter [7], [2].

Recent optimization-oriented machine learning approaches for information systems and misinformation analysis have demonstrated that hybrid and tuned models significantly improve classification accuracy and robustness [5], [6].

Fake news has been a problem for a long time. People used to believe this news, especially if it wasn't true. It can be hard to know if the news is real or fake. Since there’s no one controlling the internet, it's become easier to share harmful content. A humbler parcel of machine learning that requires portion information to make a smart choice [3]. This paper centers on utilizing common machine learning methodologies to recognize fake news. The data we utilized didn't have adequate information.

It is evident that various people neglect the truth roughly what's happening on social media right by and by. We have to be get liberated of the old things; we require some-thing much prevalent; we require something improved; we ought to be more relent-less; we need to make development. A much better way to help us. Select out which information on social media is veritable and which isn't [11]. Several survey and attribution-based approaches have also been explored to understand misinformation propagation and detection strategies [10], [13].

This proposes a methodology to find out on the off chance that something is honest to goodness or fake. Turning reasonable news into less complex words utilizing principal machine learning techniques. Utilizing procedures to form choices by voting. The calculations being utilized are outstandingly basic. Bayes

(NB), straight back vector machine (LSVM), calculated backslide (LR), subjective timberland (RF), and passive-aggressive are specific sorts of machine learning calculations. The Father and SGD are two differing sorts of classifiers [14].

2. Proposed Framework of optimization

This parcel of the article looks at the prescribed framework by comparing it. This information can be utilized by anyone to help prepare the system. To start with, get liberated of any futile information, letters, or numbers. TF-IDF and DOC2VEC are both techniques utilized in common lingo taking care of to talk to and get it the meaning of words and sentences. They offer assistance with tests and make information less requesting to induce it and more correct. Utilized to drag out basic information from the data. Similar optimization-driven ML frameworks have been successfully applied in information systems to enhance decision accuracy and computational efficiency [5]. The overall workflow of the proposed fake news detection framework is illustrated in Figure 1.

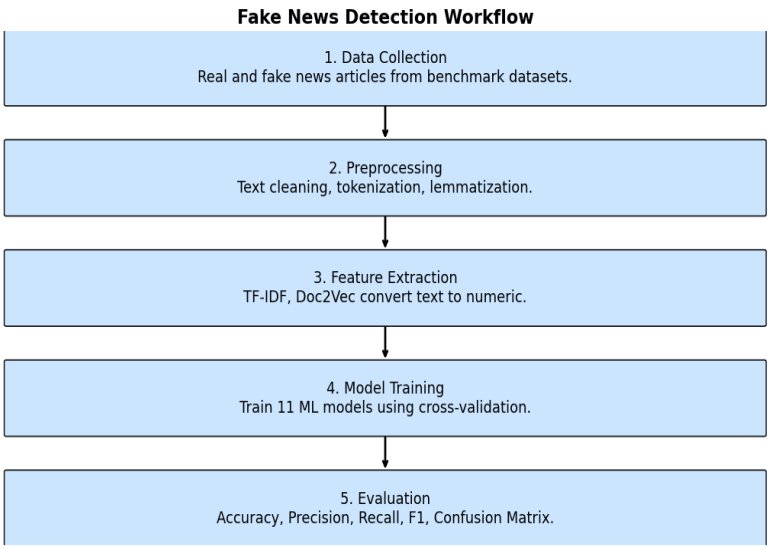

Figure 1
Workflow for Fake News Detection technique of ML Algorithm.

2.1 *Data Preprocessing*

Information should be arranged before taking out the important details. This information could have cool letters and numbers, extra and unnecessary space. Start by removing all special characters also known as strange letters and numbers. The last part of planning the content is called lemmatization. This

reduces words to their basic meaning. For example, there will be one computer left. Lemmatization helps avoid using the same words repeatedly.

2.2 *Dataset Section*

It's truly imperative to part the information into two parts: one for preparing and one for testing to see at the diverse sorts of classification models. In this segment, the. Information is part into two segments: one for learning and one for checking. A portion is utilized to plan for the appear, and it's called a prep set. The other portion is utilized to check how great the classification is. It's a set of apparatuses for testing. We utilize the test set to make strides how things work. The dataset contains the ID, title, creator, and content data. The information was taken from the Kaggle dataset. The way we part the information into parts is called k-fold cross-validation. [7], [9], [15]. Language-specific NLP studies further highlight challenges in adapting fake news detection techniques across linguistic contexts [12].

3. Proposed Method Contribution for optimization

To ensured class distribution analysis and consistent label encoding. Early analysis of class imbalance using histograms and statistics that helps to design balanced models and assess data bias. Hybrid and forensic-based detection methods have been investigated for content manipulation and misinformation analysis in multimedia contexts [2], [3], [8]. Used regex-based cleaning: re.sub(r'[^A-Za-z\s]', '', text) and removed stop words using NLTK to improves semantic quality of input and reduces noise, ensuring each word contributes meaningfully Used

1. $\text{LogLoss} = -\frac{1}{N} * \Sigma [y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)]$

N = Total number of samples
y_i = Actual label (0 or 1)
$\hat{y}_i$ = Predicted probability of being class 1 (fake news)

In Optimization Phase used Logistic Regression(C=1.2) increases sensitivity to mis-classification. XGBoost(max_depth=4, subsample=0.8, colsample_bytree=0.8) ena-bles sparse-aware learning, reducing overfitting and enhancing generalization.

Our approach followed a three-phase evaluation: Baseline → Improved → Hyperparameter Tuned, ensuring a thorough performance comparison. Compared to traditional models, the improved and tuned XGBoost model achieved over 95% accuracy and outperformed all others across precision, recall, and F1-score. The use of custom preprocessing (stopword removal, lemmatization) and bi-gram feature engineering significantly enhanced model

learning. Intermediate improvements like changing model parameters, restricting feature size, and adjusting max depth yielded clear performance gains. The hyperparameter tuning process, although computationally expensive, optimized model settings further and improved stability across metrics [1], [4]. Figure 2 presents the comparative performance trend across baseline, improved, and fine-tuned models.

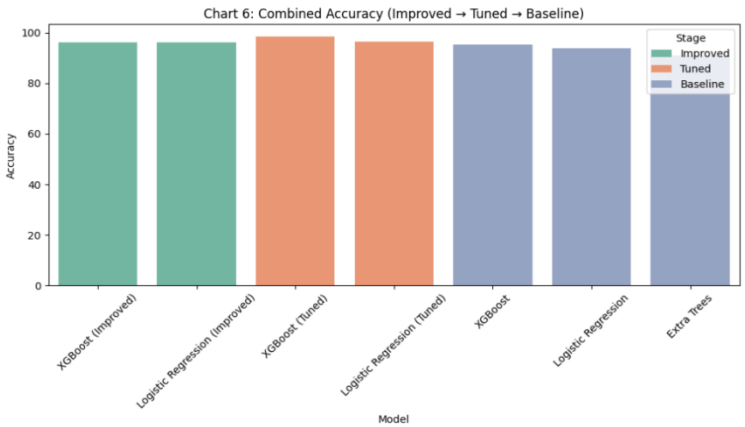

Figure 2

Clearly reflects the overall effectiveness of the proposed methodology Base-line → Improvement → Fine-tuning pipeline helps identify the best version of each model.

The table has the ML Algorithm models with Improved and tuned results for Accuracy, Precession, Recall and F1 Score. By using tuned parameter the accuracy has been achieved and optimized. The optimized hyper parameter settings and corresponding performance metrics for each machine learning model are summarized in Table 1.

Table 1

Best hyper tuning parameter for optimization Fake News Detection ML Algorithm.

Model	Accuracy	Precision	Recall	F1 Score
XGBoost (Improved)	96.15385	92.85171	95.46521	96.14032
Logistic Regression (Improved)	96.09615	93.27182	94.83972	96.04923
XGBoost (Tuned)	98.38462	93.58268	92.92416	98.25225
Logistic Regression (Tuned)	96.57692	93.54971	89.01486	96.22596
XGBoost	95.38462	94.95733	95.69977	95.32710
Logistic Regression	93.80769	93.09946	94.40969	93.75000
Extra Trees	91.03846	95.99824	85.34011	90.35596

4. Conclusion

Using text cleanup, two-word pairs, and TF-IDF changes greatly improved how well the model worked. Baseline models showed us how well things could perform, and the adjustments we made showed that our step-by-step approach was helpful. The accuracy of the model was improved to make sure it can correctly find fake news and is trustworthy. The system works well and is very accurate (over 95%) with strong F1-scores in improved and adjusted models. Our method can grow and be changed for other text classification tasks, and we expect to see similar optimization.

Overall, the comparative assessment of distinctive approaches appears that show optimization through parameter tuning, crossover structures, and deep learning integration significantly boosts both accuracy and efficiency.

References

[1] C. Dong, X. Chen, R. Hu, J. Cao, and X. Li, "Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection," arXiv preprint arXiv:2112.08935 (2021).

[2] N. Kaur, N. Jindal, and K. Singh, "Efficient Hybrid Passive Method for the Detection and Localization of Copy-Move and Spliced Images," *Turkish Journal of Electrical Engineering & Computer Sciences,* vol. 29, no. 2, pp. 561–582 (2021), doi: 10.3906/ELK-2001-138.

[3] P. Meel and D. K. Vishwakarma, "Deep Neural Architecture for Veracity Analysis of Multimodal Online Information," in *Proc. 11th Int. Conf. Cloud Computing, Data Science & Engineering (Confluence),* pp. 7–12 (2021), doi: 10.1109/Confluence51648.2021.9377172.

[4] D. Varshney and D. K. Vishwakarma, "ArtiMarker: A Novel Artificially Inflated Video Marking and Characterization Method on YouTube," in *Proc. 5th Int. Conf. Computer, Communication and Signal Processing (ICCCSP)*, pp. 244–249 (2021), doi: 10.1109/ICCCSP52374.2021.9465393.

[5] S. Desai, N. A. Rahul, V. Latke, T. Deshmukh, P. N. Mahalle, and L. Umate, "Hybrid machine learning models for solving complex optimization problems in information systems," *Journal of Information and Optimization Sciences*, vol. 46, no. 4-B, pp. 1253–1263 (2025).

[6] P. Malhotra and S. K. Malik, "Fake news detection using supervised machine learning techniques," *Journal of Information and Optimization Sciences,* vol. 43, no. 1, pp. 7–15 (2022).

[7] Z. A. Jawad and A. J. Obaid, "A systemic literature overview of Fake News Challenge (FNC-1) dataset and its use in fake news detection schemes," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 26, no. 4, pp. 1197–1206 (2023).

[8] D. Baines and R. J. R. Elliott, "Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic," *Discussion Paper 20-06, Department of Economics,* University of Birmingham, Birmingham, U.K., (Apr. 2020). [Online]. Available: https://repec.cal.bham.ac.uk/pdf/20-06.pdf. [Accessed: Jul. 29, 2026].

[9] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information," *Big Data*, vol. 8, no. 3, pp. 171–188 (2020).

[10] S. Ahmed, K. Hinkelmann, and F. Corradini, "Combining Machine Learning with Knowledge Engineering to Detect Fake News in Social Networks—A Survey," in *Proc. AAAI 2019 Spring Symp. Combining Machine Learning with Knowledge Engineering (AAAI-MAKE 2019)*, Palo Alto, CA, USA, CEUR Workshop Proc., vol. 2350 (Mar. 25–27, 2019).

[11] B. D. Horne and S. Adali, "This Just In: Fake News Packs A Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire Than Real News," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 11, no. 1, pp. 759–766 (May 2017), doi: 10.1609/icwsm.v11i1.14976.

[12] C. Busioc, S. Ruseti, and M. Dascalu, "A Literature Review of NLP Approaches to Fake News Detection and Their Applicability to Romanian-Language News Analysis," *Transilvania*, vol. 50, no. 10, pp. 65–71 (Oct. 2020), doi: 10.51391/trva.2020.10.07.

[13] T. Traylor, J. Straub, and W. Snell, "Classifying fake news articles using NLP," in *Proc. IEEE Int. Conf. on Semantic Computing* (ICSC), pp. 505–509 (2019), doi: 10.1109/ICSC.2019.8665587.

[14] F. A. Abdulghani and N. A. Z. Abdullah, "Hybrid deep learning model for Arabic text classification," *Journal of Information and Optimization Sciences,* vol. 43, no. 8, pp. 1901–1908 (2022).

[15] P. Prajapati and A. Thakkar, "Extreme multi-label learning," *Journal of Information and Optimization Sciences*, vol. 40, no. 4, pp. 983–1001 (2019).

Received April, 2025