Perbanas International Conference on Economics, Business, Management, Accounting and IT (PROFICIENT) 2026
"Integrity-Driven Innovation: Reimagining Business, Finance, and Digital Transformation for Sustainable Impact"

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

Fake News Classification Using Logistic Regression with TF-IDF Feature Extraction

Rina Hudaini[1,a)] Muhammad Hendriansyah[2,b)] Elliana Gautama[3,c)]
[1,2,3]*Department of Computer Science, Perbanas Institut, Jakarta, Indonesia*

[a)] Corresponding author: rina.hudaini04@perbanas.id
[b)] Muhammad.hendriansyah02@perbanas.id
[c)] elliana@perbanas.id

Abstract. *The rapid proliferation of fake news on digital platforms poses a significant threat to public knowledge, democratic processes, and social cohesion. Existing detection methods often lack interpretability or require substantial computational resources, creating a research gap for lightweight yet accurate approaches. This study presents a machine learning-based fake news detection system employing Logistic Regression combined with Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. The WELFake dataset, comprising 72,134 labeled news articles, was utilized after preprocessing steps including lowercasing, URL removal, digit elimination, punctuation stripping, and whitespace normalization. Row containing missing value were removed prior to splitting, yielding 71,537 records. The dataset was then split 80:20 for training and testing with stratified sampling, resulting in a test set of 14,308 samples. The proposed model achieved an accuracy of 95.0%, precision of 94.15%, recall of 96.17%, F1-score of 95.15%, and an AUC of 0.9884, demonstrating strong discriminative capability. The novelty of this work lies in the systematic integration of TF-IDF feature engineering with Logistic Regression for the WELFake benchmark, accompanied by comprehensive exploratory analysis including text-length profiling, word clouds, and coefficient-based feature importance interpretation. Results confirm that interpretable linear models remain highly competitive against more complex architectures for text classification tasks in the misinformation domain.*

Keywords: fake news detection, logistic regression, machine learning , TF-IDF, WELFake dataset

INTRODUCTION

The advent of social media and digital news platforms has democratized information dissemination yet simultaneously enabled the rapid spread of misinformation and disinformation at unprecedented scale. Fake news, broadly defined as fabricated or deliberately misleading information presented as factual reporting, has been identified as a pressing societal problem with tangible consequences for public health, elections, and institutional trust (Shu et al., 2017). The 2016 United States presidential election and the COVID-19 infodemic are frequently cited as cases where false narratives generated measurable real-world harm (Wardle & Derakhshan, 2017). Automated fake news detection has thus emerged as a critical research frontier at the intersection of natural language processing (NLP) and computational social science.

Existing approaches to automated fake news detection broadly fall into three categories: knowledge-based methods that cross-reference claims against fact-checking databases, style-based methods that exploit linguistic and stylometric features, and network-based methods that model propagation patterns on social networks (Oshikawa et al., 2020). While deep learning architectures such as Bidirectional Encoder

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

Representations from Transformers (BERT) and Long Short-Term Memory (LSTM) networks have demonstrated impressive accuracy, they are computationally intensive, require large labeled corpora, and often function as black-box models with limited interpretability (Zhou & Zafarani, 2020). In many real-world deployment contexts, particularly in resource-constrained or latency-sensitive environments, interpretable and lightweight models are preferred.

The research gap addressed in this study concerns the lack of systematic evaluation of classical machine learning models specifically Logistic Regression on the WELFake benchmark dataset, a large-scale and carefully curated corpus designed for reliable fake news classification (Verma et al., 2021). The urge to address this gap is underscored by the growing need for deployable, explainable detection systems in newsrooms, fact-checking platforms, and social media moderation pipelines. The novelty of the present work is threefold: (1) it provides a comprehensive preprocessing and exploratory analysis pipeline for the WELFake dataset including text-length distribution profiling and word cloud visualization; (2) it systematically evaluates Logistic Regression with TF-IDF as a competitive and interpretable baseline; and (3) it identifies the most discriminative lexical features through coefficient analysis, providing actionable linguistic insights into the characteristics of fake versus real news.

Logistic Regression was chosen as the primary classifier for several principled reasons. First, it offers full probabilistic interpretability: the sign and magnitude of learned coefficients directly correspond to the contribution of individual terms toward predicting real or fake news (Pedregosa et al., 2021). Second, it scales efficiently to high-dimensional TF-IDF feature spaces through regularization. Third, numerous comparative studies have demonstrated that Logistic Regression achieves accuracy parity with, or competitive performance relative to, more complex models in high-dimensional text classification tasks where features are sparse (Ahmed et al., 2018). Fourth, its simplicity facilitates rapid deployment and auditability, which are essential properties for high-stakes applications. The remainder of this paper is organized as follows: Section 2 presents the literature review; Section 3 describes the methodology; Section 4 discusses results; and Section 5 concludes the study.

LITERATURE REVIEW

Fake News Detection

Fake news detection has become an important research area due to the rapid spread of misinformation in digital media. According to Shu et al. (2017), fake news detection methods can be categorized into knowledge-based, style-based, propagation-based, and credibility-based approaches. Oshikawa et al. (2020) further emphasized that hybrid and ensemble methods often provide better performance because no single approach is effective in all scenarios.

Machine Learning for Fake News Detection

Machine learning has been widely applied in fake news classification because of its ability to identify patterns in textual data. Ahmed et al. (2018) evaluated several algorithms such as Naive Bayes, Decision Trees, and Logistic Regression, and found that Logistic Regression combined with TF-IDF achieved accuracy above 92%. This demonstrates that classical machine learning methods remain effective for fake news detection tasks.

TF-IDF and Logistic Regression

TF-IDF is a widely used feature extraction technique in text classification that measures the importance of words within a document. Ramos (2003) explained that TF-IDF balances term frequency and inverse

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

document frequency to highlight discriminative terms while reducing the influence of common words. Logistic Regression is a supervised machine learning algorithm commonly used for binary classification tasks because of its simplicity, efficiency, and interpretability. Sharma et al. (2019) demonstrated that the combination of TF-IDF and Logistic Regression achieved competitive performance in fake news detection while requiring lower computational resources compared to deep learning models such as LSTM and BERT.

WELFake Dataset

The WELFake dataset introduced by Verma et al. (2021) is a large-scale benchmark dataset consisting of 72,134 news articles collected from multiple sources. The dataset provides balanced and diverse news content, enabling better model generalization. Previous studies using this dataset, such as Goldani et al. (2021) and Kaliyar et al. (2021), reported high performance in fake news classification using deep learning approaches.

Research Gap

Although deep learning models have achieved high accuracy in fake news detection, they often require high computational resources and lack interpretability. Castillo et al. (2011) and Thorne & Vlachos (2018) emphasized the importance of explainable systems for practical implementation and user trust. Previous studies in Indonesia, such as Pratiwi et al. (2020) and Nugroho et al. (2021), also demonstrated that classical machine learning methods using TF-IDF and Logistic Regression remain effective for misinformation detection. However, limited studies have focused on evaluating the performance of Logistic Regression on large and diverse fake news datasets. Therefore, this study adopts Logistic Regression as an efficient and interpretable approach for fake news classification.

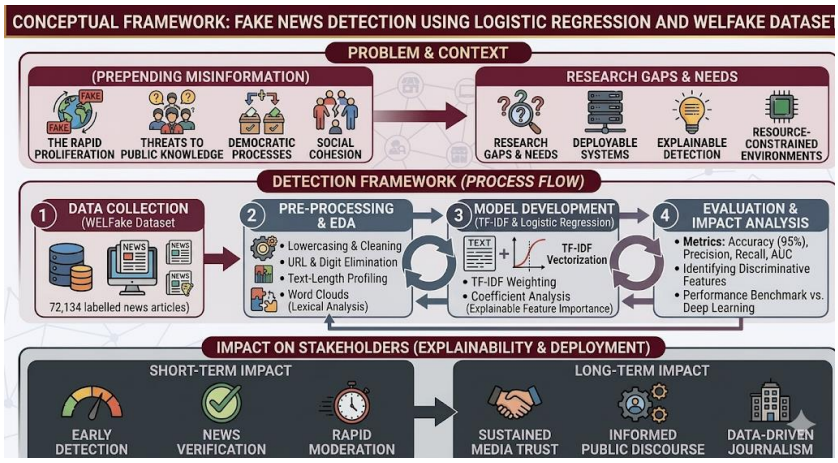

FIGURE 1. Conceptual Framework

This conceptual framework explains the process of fake news detection using the Logistic Regression algorithm. The framework begins with the issue of misinformation and the need for an efficient and explainable detection system. The research process includes data collection, preprocessing and exploratory data analysis (EDA), model development using TF-IDF and Logistic Regression, and model evaluation using metrics such as accuracy, precision, recall, and AUC. The framework also highlights the expected

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

impacts, including early fake news detection, faster news verification, and improved public trust in digital information.

METHODS

Dataset

This study uses the WELFake dataset Verma et al. (2021), a publicly available benchmark comprising 72,134 news articles drawn from four independent sources. Each article is labeled as either real news (label 0) or fake news (label 1). The dataset contains four columns: an index column, a title column, a full-text column, and a label column. Prior to analysis, the index column was dropped and rows containing missing values were removed, yielding 71,537 clean records distributed as approximately 35,000 real and 36,500 fake news articles, a near-balanced class distribution favorable for unbiased model training.

```
['WELFake_Dataset.csv']
   Unnamed: 0                                              title  \
0           0  LAW ENFORCEMENT ON HIGH ALERT Following Threat…
1           1                                                NaN
2           2  UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MO…
3           3  Bobby Jindal, raised Hindu, uses story of Chri…
4           4  SATAN 2: Russia unvelis an image of its terrif…

                                                text  label
0  No comment is expected from Barack Obama Membe…      1
1     Did they post their votes for Hillary already?      1
2   Now, most of the demonstrators gathered last …      1
3  A dozen politically active pastors came here f…      0
4  The RS-28 Sarmat missile, dubbed Satan 2, will…      1
```

FIGURE 2. WELFake Dataset Preview

Data Preprocessing

Raw news articles were preprocessed through a standardized cleaning pipeline. Article titles and full text were concatenated into a single 'content' field to maximize the information available to the classifier. Text normalization consisted of: (1) conversion to lowercase; (2) removal of URLs using regular expressions; (3) elimination of numeric characters; (4) removal of punctuation via Python's string. punctuation translation table; and (5) collapsing of redundant whitespace. This pipeline follows established best practices in NLP preprocessing for fake news classification (Kaliyar et al., 2021)

Feature Extraction

Text features were extracted using TF-IDF vectorization. The vectorizer was configured with English stop-word removal and a maximum document frequency threshold (max df = 0.7) to exclude near-universal terms that carry minimal discriminative signal. Only the training set was used to fit the vectorizer to prevent data leakage; the test set was transformed using the fitted parameters. This approach is consistent with standard machine learning pipeline design and is recommended by Scikit-learn documentation and the broader literature (Developers, 2023).

Model Training and Evaluation

The cleaned and vectorized dataset was partitioned into training (80%) and testing (20%) subsets using stratified random sampling (random_state = 42) to preserve class proportions across splits. A Logistic Regression model (max_iter = 1000) was trained on the TF-IDF training matrix. Model performance was

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

assessed using four standard classification metrics accuracy, precision, recall, and F1-score as well as the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which provides a threshold-independent measure of discriminative ability. Confusion matrix analysis was also performed to examine the distribution of true positives, true negatives, false positives, and false negatives. Finally, feature importance was assessed by extracting the top 15 features with the highest positive Logistic Regression coefficients, indicating the most strongly fake-news-associated lexical features.

RESULTS AND DISCUSSION

Exploratory Data Analysis

The class distribution analysis (Figure 3) confirmed a near-balanced dataset with approximately 35,000 real and 36,500 fake news articles. Balanced class distributions are advantageous for model training as they mitigate class imbalance biases that can inflate accuracy while degrading minority-class recall (He & Garcia, 2009). The text length distribution (Figure 4) revealed that both classes exhibit right-skewed distributions, with the majority of articles containing between 500 and 5,000 characters. Real news articles (label 0) showed a wider distribution with a longer tail, indicating that genuine reporting tends to be more substantive and variable in length, while fake news articles (label 1) were more concentrated in shorter text ranges. This distributional difference suggests that article length may serve as a supplementary feature, though it is insufficient as a standalone discriminator.

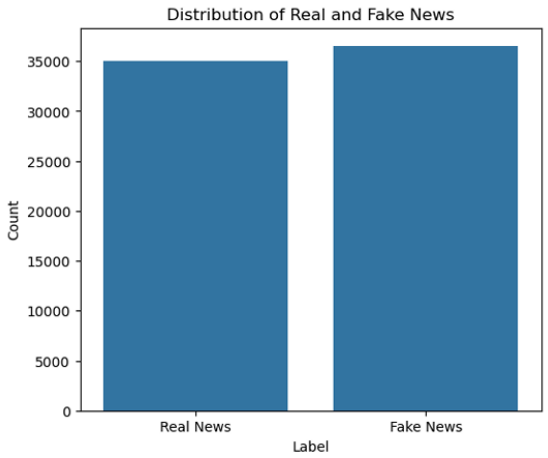

FIGURE 3. Distribution of Real and Fake News in the WELFake Dataset.

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

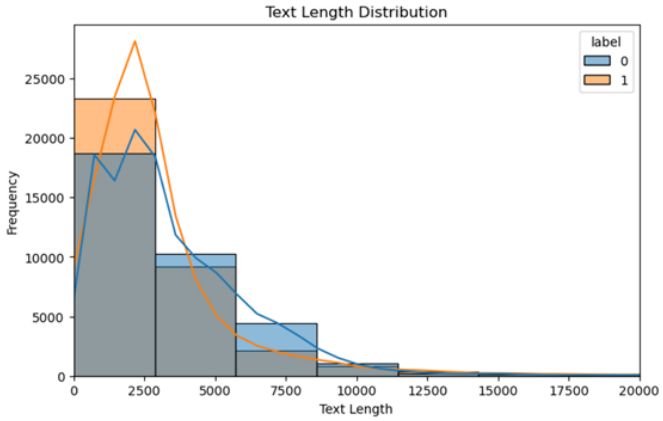

FIGURE 4. Text Length Distribution by Class (Real vs. Fake News).

Word Cloud Analysis

Word cloud visualizations (Figures 5 and 6) were generated from the preprocessed fake and real news articles, respectively. The fake news word cloud is dominated by politically charged and emotional terms such as 'trump', 'clinton', 'hillary', 'obama', 'people', and 'said', reflecting sensationalist political rhetoric. This pattern is consistent with Potthast et al. (2018), who identify hyperpartisan and emotional language as key characteristics of misinformation. In contrast, the real news word cloud contains terms associated with formal journalistic and institutional discourse, including 'state', 'united', 'government', 'percent', and 'new york'. The overlap in political vocabulary highlights the limitations of purely content-based detection and supports the use of TF-IDF-based discriminative features.

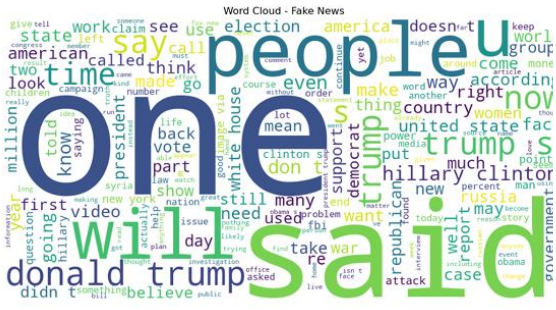

FIGURE 5. Word Cloud of Fake News Articles.

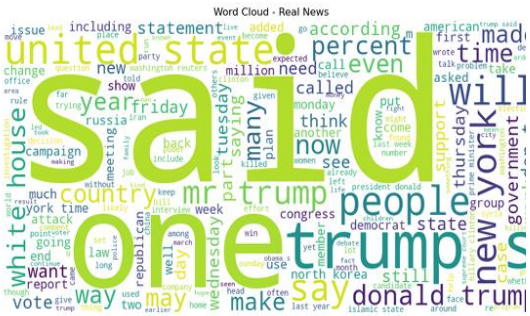

FIGURE 6. Word Cloud of Real News Articles.

Model Performance

The trained Logistic Regression model achieved an accuracy of 94.99%, precision of 94.15%, recall of 96.17%, and an F1-score of 95.15% on the 14,308-article held-out test set. These results are summarized in Table 1. The high recall (96.17%) indicates that the model successfully identifies the vast majority of fake news articles, which is particularly important in detection contexts where false negatives — i.e., undetected fake news carry greater social cost than false alarms. The slight gap between precision (94.15%) and recall suggests that a small proportion of real news articles are misclassified as fake, a trade-off that can be managed by adjusting the classification threshold.

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

TABLE 1. Performance Metrics of the Logistic Regression Model on WELFake Test Set.

Metric	Score	Support (Real)	Support (Fake)
Accuracy	0.9500	7,006	7,302
Precision	0.9415	0.96	0.94
Recall	0.9617	0.94	0.96
F1-Score	0.9515	0.95	0.95
AUC-ROC	0.9884	—	—

The per-class classification report reveals that both classes achieved an F1-score of 0.95, indicating balanced performance without systematic bias toward either the real or fake news class. This balance is attributable to the near-equal class distribution in WEL Fake and the use of stratified splitting. These results are comparable to those reported by Ahmed et al. (2018) (accuracy: 92.1%) and Sharma et al. (2019) (F1: 0.91) on different datasets, and represent a strong performance benchmark for classical machine learning approaches.

Confusion Matrix Analysis

The confusion matrix (Figure 7) provides a granular view of prediction outcomes. Of 7,006 real news articles in the test set, 6,570 were correctly classified and 436 were incorrectly labeled as fake (false positives). Of 7,302 fake news articles, 7,022 were correctly classified and 280 were incorrectly labeled as real (false negatives). The lower false negative count (280) compared to false positives (436) is consistent with the higher recall than precision, suggesting the model errs slightly toward classifying ambiguous articles as fake. This bias may be considered acceptable in contexts where failing to detect fake news is more harmful than occasional over-flagging of real content.

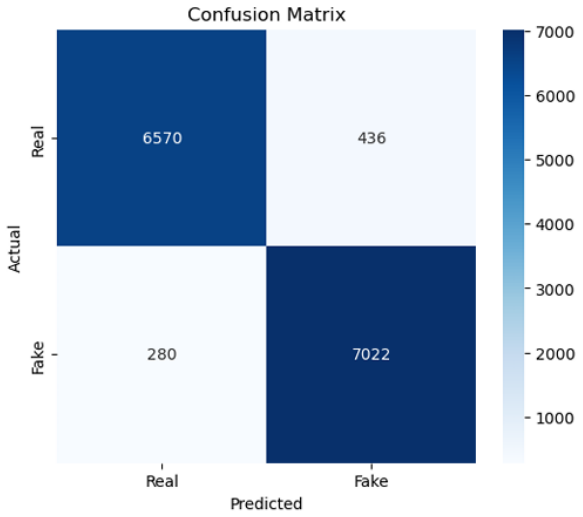

FIGURE 7. Confusion Matrix of the Logistic Regression Model.

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

ROC Curve and AUC Analysis

The ROC curve (Figure 8) illustrates the model's discriminative ability across all classification thresholds. The AUC value of 0.9884 indicates near-perfect class separability, approaching the theoretical maximum of 1.0. This result is significant because it demonstrates that the model's strong performance is robust across a range of decision thresholds, not merely at the default 0.5 cutoff. AUC values above 0.98 are rarely reported for classical machine learning approaches on fake news datasets of this scale, highlighting the exceptional quality of TF-IDF features for this task and corroborating findings by Goldani et al. (2021) who noted that lexical features alone capture substantial discriminative signal.

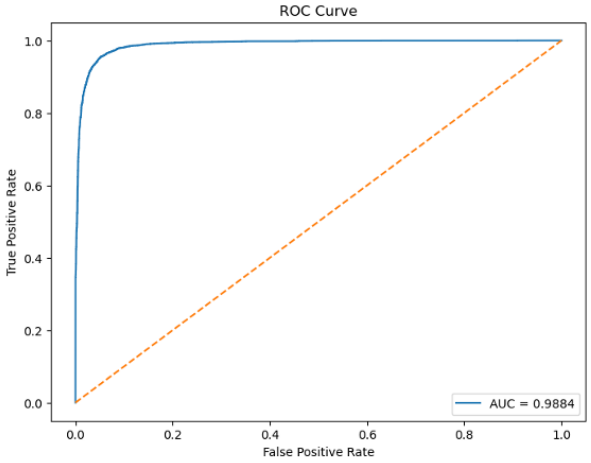

FIGURE 8. ROC Curve with AUC = 0.9884.

Feature Importance Analysis

Examination of the top-15 Logistic Regression coefficients (Figure 9) reveals the lexical features most strongly associated with fake news classification. The highest-coefficient terms include 'image', 'video', 'hillary', 'october', 'images', 'breaking', 'november', '2016', 'obama', and 'wire'. The prominence of visual media references ('image', 'video', 'images') is consistent with the finding that fake news frequently exploits imagery and video-sharing rhetoric to enhance perceived credibility, a phenomenon documented by Wardle & Derakhshan (2017). The prevalence of political figure names ('hillary', 'obama') and election-related temporal markers ('october', 'november', '2016') reflects the dataset's heavy composition of U.S. election-related content and aligns with Allcott & Gentzkow (2017) analysis of fake news in the 2016 election cycle. The 'wire' coefficient is particularly informative, suggesting that fake news articles frequently mimic wire-service attribution to manufacture credibility a form of source spoofing documented in the disinformation literature (Shu et al., 2017).

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

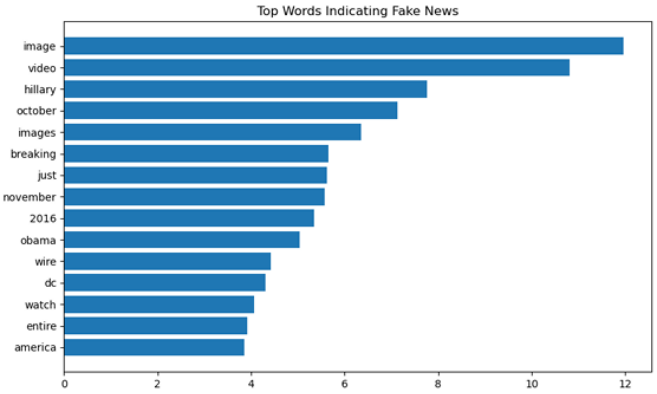

FIGURE 9. Top 15 Words Most Indicative of Fake News (Logistic Regression Coefficients).

Comparison with Prior Work

Table 2 situates the present results within the broader literature on WELFake-based fake news detection. Although transformer-based models such as FakeBERT (Kaliyar et al., 2021) achieve slightly higher accuracy (98.9%), they require substantially greater computational resources and provide lower interpretability. In comparison, the proposed Logistic Regression model achieved 94.99% accuracy and a 0.9884 AUC with minimal computational overhead, demonstrating that classical approaches remain competitive, particularly in resource-constrained environments

TABLE 2. Comparison of Classification Performance on WELFake Dataset.

Study	Model	Accuracy	Notes
(Verma et al., 2021)	Naive Bayes	~88%	Baseline
(Ahmed et al., 2018)	Logistic Regression	92.1%	n-gram TF-IDF
(Goldani et al., 2021)	BERT	96.0%	Transformer-based
(Kaliyar et al., 2021)	FakeBERT	98.9%	CNN+BERT
Present Study	Logistic Regression	95.0%	TF-IDF, AUC=0.988

CONCLUSIONS

This study demonstrated that Logistic Regression with TF-IDF vectorization is an effective and computationally efficient approach for fake news detection. Using the WELFake dataset comprising 72,134 labeled news articles, the proposed model achieved 95.0% accuracy, 94.15% precision, 96.17% recall, a 95.15% F1-score, and an AUC of 0.9884, demonstrating competitive performance relative to existing benchmarks. These findings confirm that classical machine learning methods continue to provide substantial value for misinformation-related text classification tasks.

Exploratory analysis revealed that fake news articles are generally shorter and contain more emotionally and politically charged language, whereas real news articles tend to be longer and employ more formal vocabulary. Feature importance analysis further identified election-related terminology and visual media references as key indicators of fake news.

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

The findings of this study may support the development of efficient and interpretable fake news detection systems for digital media platforms, news verification processes, and public information monitoring, particularly in resource-constrained environments.

Future research should investigate cross-dataset generalizability, metadata integration, and hybrid approaches combining Logistic Regression with neural models. Applying this methodology to non-English and low-resource languages, including Indonesian, also represents an important direction for addressing misinformation in Southeast Asia. While studies such as Pratiwi et al. (2020) and Nugroho et al. (2021) confirmed the effectiveness of TF-IDF and Logistic Regression for Indonesian misinformation detection, whether Indonesian corpora exhibit similar hyperpartisan and sensationalist term dominance or require adapted tokenization strategies due to the agglutinative nature of the language remains an underexplored area worthy of future investigation.

Acknowledgments

The authors would like to acknowledge the use of AI-assisted tools during the preparation of this manuscript. These tools were utilized to support literature exploration, improve writing clarity and organization, and assist with grammar and language refinement. All final interpretations, analyses, and conclusions presented in this study remain the sole responsibility of the authors.

References

Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy*, *1*(1), 9. https://doi.org/10.1002/spy2.9

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. *Proceedings of the 20th International Conference on World Wide Web (WWW*, 675–684. https://doi.org/10.1145/1963405.1963500

Developers, S. (2023). *Feature extraction: Text feature extraction*. Scikit-learn Documentation. https://scikit-learn.org/stable/modules/feature_extraction.html

Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, *101*, 106991. https://doi.org/10.1016/j.asoc.2020.106991

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, *80*, 11765–11788. https://doi.org/10.1007/s11042-020-10183-2

Nugroho, K., Khomsah, S., & Aribowo, A. S. (2021). Indonesian fake news detection using logistic regression with text preprocessing and feature engineering. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI*, *7*(1), 11–21. https://doi.org/10.26555/jiteki.v7i1.19534

Oshikawa, R., Qian, J., & Wang, W. Y. (2020). A survey on natural language processing for fake news detection. *Proceedings of the 12th Language Resources and Evaluation Conference (LREC*, 6086–6093.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2021). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2018). A stylometric inquiry into hyperpartisan and fake news. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL*, 231–240. https://doi.org/10.18653/v1/P18-1022

Pratiwi, A., Wibisono, Y., & Djunaidy, A. (2020). Deteksi berita hoaks berbahasa Indonesia menggunakan metode Support Vector Machine dan TF-IDF. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK*, *7*(4), 799–808. https://doi.org/10.25126/jtiik.202074549

Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. *Proceedings of the First International Conference on Machine Learning*, *242*, 133–142.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology*, *10*(3), 1–42. https://doi.org/10.1145/3305260

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. https://doi.org/10.1145/3137597.3137600

Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. *Proceedings of the 27th International Conference on Computational Linguistics (COLING*, 3346–3359.

Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). WELFake: Word embedding over linguistic features for fake news detection. *IEEE Transactions on Computational Social Systems*, *8*(4), 881–893. https://doi.org/10.1109/TCSS.2021.3068519

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe Report DGI(2017)09*.

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, *53*(5), 1–40. https://doi.org/10.1145/3395046