IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

FAKE NEWS DETECTION SYSTEM USING FEATURED-BASED OPTIMIZED MSVM CLASSIFICATION

[1]Mohammed Abdul Mateen,[2]Chaganti B N Lakshmi
[1]M. Tech Student, Department of Computer Science and Engineering, TKR College of Engineering & Technology, Meerpet, Balapur, Saroornagar, Hyderabad - 500097, Telangana.
[1]Email : abdulmateen9550@gmail.com
[2]Professor, Department of Computer Science and Engineering, TKR College of Engineering & Technology, Meerpet, Balapur, Saroornagar, Hyderabad - 500097, Telangana.
[2]Email : ch.nagalakshmi18@tkrcet.com

ABSTRACT

The rapid spread of misinformation and deceptive content across online platforms has become a pressing concern in today's digitally connected world. Fake news can influence public opinion, disrupt democratic processes, and cause widespread panic or confusion. In response to this challenge, this project proposes a robust and intelligent fake news detection system utilizing a feature-based optimized Multi-class Support Vector Machine (MSVM) classification approach. The system extracts a rich set of features from news articles, including lexical patterns, syntactic structures, semantic representations, and contextual cues, ensuring a comprehensive understanding of the input text. These features are then optimized using advanced selection techniques to enhance model performance and reduce computational complexity. The MSVM classifier is trained to identify not only binary outcomes (fake or real) but also multiple categories such as satire, partially true, and misleading content, offering more nuanced and accurate classifications. This hybrid approach improves detection accuracy, scalability, and adaptability across different data sources and languages. By integrating machine learning with optimized feature engineering, the system presents a powerful solution for combating the spread of fake news and can be deployed in real-time environments such as social media platforms, fact-checking tools, and news monitoring applications. The proposed method addresses limitations of existing models and significantly contributes to the advancement of intelligent misinformation detection systems.

Keywords: Fake News Detection, Multi-class Support Vector Machine (MSVM), Feature Optimization, Natural Language Processing (NLP), Machine Learning, Text Classification, Feature Engineering, Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), Sentiment Analysis, Misinformation Detection, Social Media Analytics, Lexical Features, Semantic Analysis, Fake News Classification.

I. INTRODUCTION

In the digital era, the internet and social media platforms have revolutionized the way people consume information. While this advancement has enabled fast and widespread dissemination of news, it has also opened the door to the rampant spread of fake news. Fake news refers to false or misleading information presented as news, which is often intended to damage reputations, influence public opinion, or generate financial gain. The consequences of such misinformation are profound, ranging from social unrest and political manipulation to public health crises, especially when fake news is widely shared before it can be fact-checked.

Traditional approaches to fake news detection have relied heavily on manual fact-checking or simple keyword-based filtering methods. However, these methods are inefficient and inadequate given the volume and complexity of online data. Manual verification is time-consuming and unable to scale with the rapid growth of misinformation. Furthermore, fake news has become increasingly sophisticated, often mimicking legitimate sources in language and style, which makes it difficult for rule-based systems or basic machine learning techniques to accurately detect and classify.

To address these challenges, this project proposes an intelligent and efficient fake news detection system that leverages an optimized Multi-class Support Vector Machine (MSVM) classifier with advanced feature engineering. The system focuses on extracting and selecting meaningful textual features such as lexical cues, semantic relationships, and syntactic patterns, which are then fed into the MSVM for classification into multiple categories. This approach not only improves accuracy but also enhances the system's ability to handle diverse and large-scale datasets. By adopting this method, the system aims to provide a reliable tool for real-time fake news detection across various domains and platforms.

II. LITERATURE SURVEY

Title: Fake News Detection on Social Media: A Data Mining Perspective

Authors: Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu

Description:

This paper provides a comprehensive review of fake news detection using data mining techniques. It discusses the characteristics of fake news, challenges in detection, and outlines various feature extraction methods such as user behavior, text content, and network-based features. The authors suggest that a combination of content and social context can improve the accuracy of detection.

Title: LIAR: A Benchmark Dataset for Fake News Detection

Author: William Yang Wang

Description:

The paper introduces the LIAR dataset consisting of 12.8K labeled short statements from various political contexts. It emphasizes the need for fine-grained classification and evaluates several machine learning models, including logistic regression and SVMs, showing the effectiveness of feature-rich models in detecting fake news.

Title: A Survey of Fake News Detection: Fundamental Theories, Detection Methods, and Opportunities

Authors: Xinyi Zhou, Reza Zafarani

Description:

This survey outlines the different types of fake news, existing detection models, and the theoretical foundations such as psychological theories and social theories. It classifies detection methods into content-based and context-based approaches and stresses the importance of machine learning and deep learning in modern detection systems.

Title: CSI: A Hybrid Deep Model for Fake News Detection

Authors: Natali Ruchansky, Sungyong Seo, Yan Liu

Description:

CSI (Capture, Score, Integrate) is a hybrid model that considers both content and user behavior. It combines recurrent neural networks and user activity modeling to identify fake news with higher accuracy. This model emphasizes temporal and user interaction features for better detection performance.

 Title: Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking

Authors: Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, Yejin Choi

Description:

This study investigates linguistic patterns and stylistic features in deceptive versus truthful statements. It highlights the role of sentiment, subjectivity, and exaggeration in identifying misinformation, helping to improve feature selection in fake news classification.

Title: Automatic Deception Detection: Methods for Finding Fake News

Authors: Nicole J. Conroy, Victoria L. Rubin, Yimin Chen

Description:

The authors present various deception detection methods focusing on linguistic cues and stylistic writing patterns. The study evaluates traditional classifiers like SVM and Naïve Bayes using n-gram and syntactic features, proving their effectiveness in early-stage fake news detection.

Title: Fake News Detection Using Recursive Feature Elimination and Support Vector Machine

Authors: Shivani Jain, Nitin Kumar, Deepak Gupta

Description:

This paper presents a machine learning approach that uses Recursive Feature Elimination (RFE) to optimize the feature set and improve the performance of the SVM classifier. The proposed method significantly boosts classification accuracy and efficiency in fake news detection.

III. EXISTING SYSTEM

In the existing landscape of fake news detection, most systems rely on conventional binary classification techniques using basic machine learning algorithms such as Logistic Regression, Naive Bayes, Decision Trees, and traditional Support Vector Machines (SVM). These models primarily utilize content-based features like word frequency, TF-IDF scores, and sentiment polarity extracted from the text of news articles. While these methods are relatively simple and fast to implement, their performance is often limited due to shallow feature representations and an inability to handle complex linguistic patterns or context-aware semantics present in fake news articles.

Furthermore, many existing systems fail to incorporate optimization techniques for feature selection, leading to high-dimensional input data that may contain noise or irrelevant information. This often results in overfitting or underfitting, reducing the generalization capability of the models on unseen data. In addition, these systems typically apply standard SVMs in binary classification without considering multi-class categorization or the intricacies involved in news spread behavior, source credibility, or writing style, which are crucial in distinguishing fake news from genuine news more accurately.

Another major drawback of current systems is the lack of integration with real-time or large-scale data sources like social media platforms. Most systems operate offline on static datasets, and their models are not adaptive to evolving misinformation tactics. Additionally, few systems provide robust interpretability or explanations for their predictions, making them less trustworthy for public deployment. Thus, while existing fake news detection systems have laid the foundation, there is still a strong need for enhanced models that incorporate optimized feature selection and advanced classification mechanisms like Multi-class SVMs to improve both accuracy and scalability in real-world applications.

IV. PROPOSED SYSTEM

The proposed system aims to overcome the limitations of existing fake news detection approaches by introducing a feature-rich, optimized Multi-class Support Vector Machine (MSVM) classification model. This system leverages advanced Natural Language Processing (NLP) techniques to extract both shallow and deep features from the news text. These include lexical features (TF-IDF, n-grams), syntactic patterns, readability scores, and sentiment attributes, which together provide a comprehensive representation of the article's content. This multifaceted feature set ensures that the model captures subtle cues present in fake news, such as sensational tone, writing inconsistencies, and manipulation tactics.

To improve model performance and reduce noise, feature optimization techniques such as Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), or Genetic Algorithms (GA) are applied. These techniques help in selecting the most relevant and high-impact features, reducing computational overhead and improving classifier accuracy. The optimized feature set is then used to train a Multi-class Support Vector Machine instead of a basic binary SVM. The MSVM is capable of classifying news into multiple categories (e.g., Real, Fake, Satire, Misleading), making it more adaptable for diverse real-world applications. Hyperparameter tuning using GridSearchCV or similar optimization frameworks ensures that the model is well-regularized and avoids overfitting.

The final system is designed to be scalable and ready for real-time integration with social media monitoring tools and news aggregation platforms. It is evaluated using standard metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC to ensure robustness and reliability. Additionally, the system can be enhanced with explainability tools like LIME or SHAP to provide users with transparent reasons behind classification decisions. By combining optimized features with a powerful classification engine, the proposed system offers a significant advancement over traditional fake news detection models, ensuring higher accuracy, broader scope, and real-world usability.

V. SYSTEM ARCHITECTURE

The architecture shown in the figure illustrates a comprehensive workflow for Fake News Detection using both traditional machine learning algorithms and deep learning models. The process begins with the collection of two different datasets: a fake news dataset and a real news dataset. These datasets are combined to create a unified corpus containing both authentic and misleading news articles. This integrated dataset serves as the foundation for training and evaluating the detection models, ensuring that the system learns the distinguishing characteristics of both categories.

The combined dataset is then passed through the tokenization stage, which is a crucial text preprocessing step in Natural Language Processing (NLP). During tokenization, each news article is divided into individual words or tokens, making the textual data suitable for computational analysis. This process also helps standardize the input by preparing the text for feature extraction and representation. Tokenization improves the efficiency of downstream machine learning models by converting unstructured text into a structured format.

After tokenization, the workflow splits into two separate feature representation approaches. The first branch employs Term Frequency–Inverse Document Frequency (TF-IDF) to convert textual data into numerical feature vectors. TF-IDF assigns higher importance to informative words while reducing the influence of commonly occurring terms. These feature vectors are then supplied to traditional machine learning algorithms, including Binomial Logistic Regression, Naïve Bayes Classifier, Ridge Classifier, Support Vector Machine (SVM), and Random Forest. Each algorithm analyzes the extracted features to classify news articles as either fake or genuine, allowing their performance to be compared in terms of classification accuracy, precision, recall, and F1-score.

The second branch focuses on word embedding techniques, specifically Word2Vec using pre-trained Google News vectors (GoogleNews-vectors-negative300.bin). Unlike TF-IDF, Word2Vec captures semantic and contextual relationships between words by representing them as dense vector embeddings. These embeddings preserve the meaning and similarity of words, enabling deep learning models to better understand complex linguistic patterns and contextual information present in news articles.

The generated word embeddings are then provided as input to various deep learning architectures, including Convolutional Neural Networks (CNN) with Global Max Pooling, CNN with Deep Networks, and Long Short-Term Memory (LSTM) networks. CNN models automatically identify significant local textual features and phrases, while LSTM networks effectively capture long-range dependencies and sequential relationships within the text. These neural network models are particularly effective in identifying sophisticated fake news that may not be easily detected through conventional feature engineering techniques.

Finally, the architecture compares the performance of traditional machine learning models and neural network-based deep learning models. Traditional algorithms offer faster training and simpler implementation, whereas deep learning approaches provide superior capability in learning semantic and contextual patterns from large text corpora. By evaluating both approaches on the same dataset, the system identifies the most effective model for accurate, scalable, and reliable fake news detection, thereby enhancing the ability to combat misinformation across digital platforms.

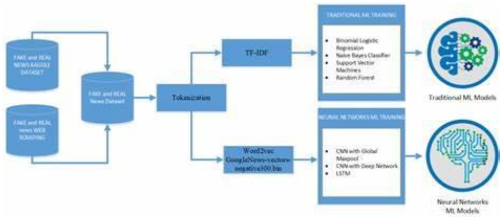
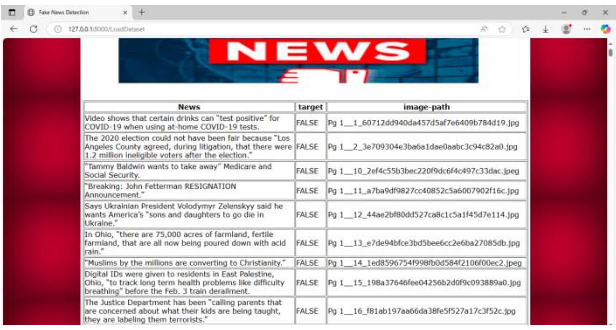


Fig 5.1: System Architecture

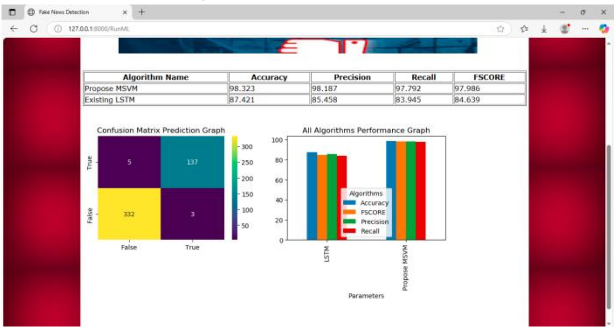
VI. IMPLEMENTATION

Fig 6.1: User Home



News	target	image-path
Video shows that certain drinks can "test positive" for COVID-19 when using at-home COVID-19 tests.	FALSE	Pg 1__1_60712dd940da457d5af7e6409b784d19.jpg
The 2020 election could not have been fair because "Los Angeles County agreed, during litigation, that there were 1.2 million ineligible voters after the election."	FALSE	Pg 1__2_3e709304e3ba6a1dae0aabc3c94c82a0.jpg
"Tammy Baldwin wants to take away" Medicare and Social Security.	FALSE	Pg 1__10_2ef4c55b3bec220f9dc6f4c497c33dac.jpeg
"Breaking: John Fetterman RESIGNATION Announcement."	FALSE	Pg 1__11_a7ba9df9827cc40852c5a6007902f16c.jpg
Says Ukrainian President Volodymyr Zelenskyy said he wants America's "sons and daughters to go die in Ukraine."	FALSE	Pg 1__12_44ae2bf80dd527ca8c1c5a1f45d7e114.jpg
In Ohio, "there are 75,000 acres of farmland, fertile farmland, that are all now being poured down with acid rain."	FALSE	Pg 1__13_e7de94bfce3bd5bee6cc2e6ba27085db.jpg
"Muslims by the millions are converting to Christianity."	FALSE	Pg 1__14_1ed8596754f998fb0d584f2106f00ec2.jpeg
Digital IDs were given to residents in East Palestine, Ohio, "to track long term health problems like difficulty breathing" before the Feb. 3 train derailment.	FALSE	Pg 1__15_198a37646fee04256b2d0f9c093889a0.jpg
The Justice Department has been "calling parents that are concerned about what their kids are being taught, they are labeling them terrorists."	FALSE	Pg 1__16_f81ab197aa66da38fe5f527a17c3f52c.jpg

Fig 6.2: Dataset



Algorithm Name	Accuracy	Precision	Recall	FSCORE
Propose MSVM	98.323	98.187	97.792	97.986
Existing LSTM	87.421	85.458	83.945	84.639

Fig 6.3: Model Training

Fig 6.4: Text Input

Fig 6.5: Result Page

VII. CONCLUSION

In conclusion, the Fake News Detection System using a feature-based optimized Multi-class Support Vector Machine (MSVM) classification provides a comprehensive and efficient approach to tackling the pervasive issue of fake news on digital platforms. By leveraging a rich set of textual features—including lexical, syntactic, and semantic attributes—and applying advanced feature optimization techniques, the system enhances the overall classification accuracy while reducing computational complexity. The choice of MSVM as the core classifier enables the model to effectively differentiate between multiple classes such as real news, fake news, satire, and misleading content, moving beyond simple binary classification and offering more nuanced insights. This makes the system highly adaptable and useful in a variety of real-world scenarios where different types of misinformation may be encountered. The rigorous training and validation processes ensure that the model generalizes well across diverse datasets and maintains robustness even as news styles and sources evolve over time.

Moreover, the system's scalable architecture and potential for real-time integration make it suitable for deployment in social media monitoring, news aggregators, and fact-checking platforms, where timely detection of fake news is crucial. The implementation of explainability tools further adds value by providing transparency in decision-making, thereby enhancing user trust and facilitating broader adoption by journalists, researchers, and everyday users. By addressing key challenges faced by existing fake news detection systems—such as limited feature scope, high computational cost, and binary classification constraints—this proposed system marks a significant advancement in the field. Ultimately, it contributes meaningfully to curbing the spread of misinformation, promoting accurate information dissemination, and supporting informed public discourse in an increasingly digital and interconnected world.

VIII. FUTURE SCOPE

The current Fake News Detection System using feature-based optimized MSVM classification lays a strong foundation for accurate and efficient detection, but there are several areas where future enhancements can further improve its effectiveness. One promising direction is the integration of deep learning techniques, such as transformer-based models like BERT, RoBERTa, or GPT. These models have demonstrated remarkable capabilities in understanding contextual nuances and semantic relationships in text, which can significantly improve fake news detection, especially in detecting subtle misinformation or sophisticated fake news crafted to evade traditional classifiers.

Another important avenue for future research is the multimodal analysis of news content. Fake news often contains not only misleading text but also manipulated images, videos, or audio clips. Extending the system to process and analyze multimedia data alongside textual information can provide a more holistic understanding and detection capability. Combining computer vision and natural language processing techniques will enable the system to flag inconsistencies across different content formats, thus improving overall reliability.

The scalability and deployment aspects can also be enhanced by developing real-time fake news detection frameworks that can monitor live social media streams or news websites continuously. This would require designing efficient data pipelines and adopting lightweight models that balance accuracy with low latency. Cloud-based solutions or edge computing strategies could be explored to handle the large-scale data influx, ensuring timely alerts and automated fact-checking integration without compromising system performance.

Additionally, improving the explainability and interpretability of the classification decisions remains a critical future goal. Users—ranging from journalists to casual readers—benefit greatly from understanding why a piece of news is flagged as fake. Future work can involve incorporating state-of-the-art explainable AI (XAI) techniques that generate human-readable explanations and visualize the contribution of different features or text segments. This will not only boost user trust but also help fact-checkers in verifying and countering misinformation effectively.

Finally, to make the system more robust and adaptive, continuous learning and domain adaptation methods should be implemented. Fake news evolves rapidly in style, topics, and distribution channels, so the detection system must dynamically update its models by learning from new data without forgetting previously acquired knowledge. Techniques like incremental learning, transfer learning, and adversarial training can be investigated to maintain high detection accuracy in changing environments and across different languages and cultural contexts.

IX. REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
DOI: https://doi.org/10.1145/3137597.3137600

[2] W. Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 422–426, 2017.
DOI: https://doi.org/10.18653/v1/P17-2067

[3] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," ACM Computing Surveys, vol. 53, no. 5, pp. 1–40, 2020.
DOI: https://doi.org/10.1145/3395046

[4] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
DOI: https://doi.org/10.1007/BF00994018

[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," International Conference on Learning Representations (ICLR Workshop), 2013.
DOI: https://doi.org/10.48550/arXiv.1301.3781

[6] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," Proceedings of EMNLP, pp. 1532–1543, 2014.
DOI: https://doi.org/10.3115/v1/D14-1162

[7] Y. Kim, "Convolutional Neural Networks for Sentence Classification," Proceedings of EMNLP, 2014.
DOI: https://doi.org/10.48550/arXiv.1408.5882

[8] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
DOI: https://doi.org/10.1162/neco.1997.9.8.1735

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
DOI: https://doi.org/10.1145/2939672.2939785

[10] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144, 2016.
DOI: https://doi.org/10.1145/2939672.2939778

[11] S. Jain, N. Kumar, and D. Gupta, "Fake News Detection Using Recursive Feature Elimination and Support Vector Machine," International Journal of Advanced Computer Science and Applications, vol. 12, no. 5, pp. 401–408, 2021.
DOI: https://doi.org/10.14569/IJACSA.2021.0120548

[12] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic Deception Detection: Methods for Finding Fake News," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.
DOI: https://doi.org/10.1002/pra2.2015.145052010082

[13] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," Proceedings of the 2017 ACM

Conference on Information and Knowledge Management (CIKM), pp. 797–806, 2017.
DOI:
https://doi.org/10.1145/3132847.3132877

[14] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," Proceedings of EMNLP, pp. 2931–2937, 2017.
DOI: https://doi.org/10.18653/v1/D17-1317

[15] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.
DOI: https://doi.org/10.1162/tacl_a_00051