INTELLIGENT FAKE NEWS IDENTIFICATION USING FEATURE ENGINEERING AND OPTIMIZED MULTI-SVM

[1]Chinnaru Venkata Ajayreddy
[2]SK Althaf Hussain Basha
Professor
Department of CSE
KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES, MARKAPUR

ABSTRACT

In the digital era, the rapid spread of fake news across social media platforms poses significant challenges to information authenticity and public trust. This research focuses on Intelligent Fake News Identification using advanced feature engineering techniques combined with an optimized Multi-Support Vector Machine (Multi-SVM) classifier. Feature engineering is employed to extract relevant textual, linguistic, and semantic features from news articles, enhancing the model's ability to differentiate between genuine and fake content. The Multi-SVM classifier is optimized using hyperparameter tuning and kernel selection to achieve high accuracy and robustness in classification. Experimental results demonstrate that the proposed system outperforms conventional machine learning models, achieving superior precision, recall, and F1-score. This approach provides a reliable and scalable solution for automated fake news detection, contributing to improved online information integrity and decision-making.

Keywords:

Fake News Detection, Feature Engineering, Multi-SVM, Text Classification, Machine Learning, Hyperparameter Optimization, Information Authenticity, Social Media Analysis.

I. INTRODUCTION

With the widespread use of social media and online news platforms, the dissemination of information has become instant and global. However, this convenience has also led to the proliferation of fake news, which can mislead the public, influence opinions, and even impact social and political stability. Fake news is deliberately fabricated information that mimics legitimate news content, often designed to deceive readers for personal, political, or financial gain.

Traditional methods of identifying fake news, such as manual fact-checking, are time-consuming, labor-intensive, and unable to cope with the volume of content generated daily. This has necessitated the development of automated fake news detection systems using machine learning techniques. Machine learning models can analyze textual patterns, semantic relationships, and contextual cues to classify news articles as genuine or fake.

In this research, feature engineering is applied to extract meaningful features from textual data, including linguistic, syntactic, and semantic characteristics. These features enhance the model's understanding of underlying patterns in news content. Furthermore, an optimized Multi-Support Vector Machine (Multi-SVM) classifier is employed, which leverages hyperparameter tuning and kernel optimization to improve classification performance. The proposed approach aims to provide a robust, accurate, and scalable solution for real-time fake news detection, helping combat misinformation and maintain information integrity in digital media.

II. LITERATURE REVIEW

Intelligent fake news detection has been extensively studied using **feature engineering**

and **optimized Support Vector Machines (SVM)** as core components in building robust detection pipelines. Sathyanarayanan and Mahiladevi (2023) proposed a system that extracts a combination of linguistic, semantic, and syntactic features from news text and then uses a **genetic algorithm** to optimize the hyperparameters of an SVM classifier. Their approach achieved a high F1-score (around 0.93), illustrating the power of feature-based optimization in classical machine learning pipelines **[1]**.

Feature extraction is essential not only for text-based metrics but also for semantic nuance. Vaideghy& Thiyagarajan (2023) employed **swarm intelligence (particle swarm optimization)** to select salient features, combined this with ensemble deep learning, and showed improved performance in detecting fake-news stances — highlighting that sophisticated feature extraction can significantly enhance detection accuracy **[2]**.

Recent work has revisited the balance between classical models and large language models. Karim, Asad, and Azam (2024) compared SVM with different vectorization strategies — including TF-IDF, Word2Vec, and Bag-of-Words — against BERT. Counterintuitively, they found that a **linear SVM with BoW vectorization** achieved nearly competitive performance versus BERT, while offering significant computational efficiency **[3]**.

On the embedding front, a 2024 study published in *Computers* (MDPI) systematically evaluated multiple embedding techniques (TF-IDF, Word2Vec, FastText) on both ML and deep learning models (SVM, MLP, CNN) using a real-world fake news dataset. They found that **SVM with TF-IDF embeddings** consistently performed well, reinforcing the importance of discriminative sparse representations when paired with SVM classifiers **[4], [12]**.

From a system design perspective, Paulin & Balaba (2025) proposed a machine learning pipeline combining SVM and LSTM based on feature engineering derived from NLP (text, source, date, etc.). Their hybrid IPO (Input–Process–Output) model demonstrates that optimizing feature selection and classifier choice helps in real-world fake news identification tasks **[5], [14], [17]**. Likewise, Nayak & Navalan (2025) conducted a comparative study where they used TF-IDF features in an SVM and benchmarked it against a fine-tuned BERT model; their results reaffirmed that well-engineered features + SVM still remain a competitive baseline for fake news detection **[6], [18]**.

In another line of work, Bussa et al. (2023) applied both **LSTM** and **SVM** classifiers for fake news detection. They performed feature importance analysis and identified certain key linguistic indicators like “unverified” and “anonymous sources” as significant predictors of fake news. Their findings show that while deep models (LSTM) provide good sequential modeling, optimized SVMs based on feature analysis are still valuable **[7],[10], [11]**.

On the applied NLP front, Randhe et al. (2025) used standard NLP preprocessing, feature extraction (including sentiment analysis), and an SVM classifier to propose a practical detection system. This work illustrates how feature-engineered, classical models can scale to real-world problems and maintain interpretability **[8], [13]**.

To reduce reliance on labeled data, a recent unsupervised approach by *Cybersecurity* (2025) combined **structural contrastive learning** with propagation graph structures of news, achieving effective fake news detection without heavy annotation. Although this work is not SVM-based, it highlights a growing trend toward reducing dependence on supervised feature-labeled data **[9], [15], [16]**.

Finally, in 2024, research in the *Indian Journal of Computer Science and Engineering (IJCSE)* used feature extraction coupled with **information gain-based dimension reduction** and SVM classification. The authors used sentence embeddings and applied feature selection to reduce dimensionality before SVM, achieving robust performance in text classification for fake news detection **[10]**.

III. EXISTING SYSTEM

In the current landscape, fake news detection primarily relies on traditional machine learning models and basic textual feature extraction methods. Most existing systems utilize Support Vector Machines (SVM), Naïve Bayes, Decision Trees, or Random Forest classifiers to identify fake news based on simple textual features such as TF-IDF, Bag-of-Words (BoW), n-grams, or word counts. Some systems also employ lexical, syntactic, or semantic features extracted from news articles to improve classification accuracy. Additionally, a few studies have applied deep learning models, such as Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), or hybrid BiLSTM-BiGRU architectures, which attempt to capture contextual and sequential information within the text.These systems have achieved moderate success in detecting fake news, especially on clean and balanced datasets. They provide automated solutions that reduce the reliance on manual fact-checking and help in filtering misinformation at scale. Some platforms have incorporated semi-supervised or weakly supervised learning to overcome the challenge of limited labeled data, and advanced embedding techniques like Word2Vec or FastText are used in some cases to capture semantic relationships.

However, most existing systems focus on single classifiers or simple models without extensive optimization or ensemble strategies, limiting their robustness across different datasets and platforms. Furthermore, they often struggle with imbalanced data, noisy content, and multilingual datasets, which are common in real-world scenarios. Feature extraction in these systems is typically basic and may not fully capture deeper linguistic or discourse-level patterns, which can reduce detection accuracy for sophisticated fake news articles.

IV. PROPOSED SYSTEM

The proposed system focuses on Intelligent Fake News Identification using feature engineering combined with an optimized Multi-Support Vector Machine (Multi-SVM) classifier to improve accuracy, robustness, and scalability in detecting fake news. Unlike existing systems that rely on a single classifier or basic textual features, this system leverages comprehensive feature extraction, including linguistic, syntactic, semantic, and discourse-level features from news articles. These features capture both the superficial and deep patterns in the text, providing a richer representation for the classification task.

The system utilizes a Multi-SVM framework, which combines multiple SVM classifiers with different kernel functions (e.g., linear, polynomial, RBF) to enhance decision-making through ensemble learning. Hyperparameter tuning and kernel optimization are applied to each SVM to maximize performance metrics such as accuracy, precision, recall, and F1-score. This ensemble approach ensures that the system is robust across different datasets and can handle variations in content style, topic, and language complexity.The proposed system also addresses limitations in existing methods, such as poor performance on imbalanced datasets, inability to capture deep semantic patterns, and lack of scalability. By integrating advanced feature engineering with optimized Multi-SVM classification, the system can efficiently detect fake news in real-time scenarios, providing a reliable and automated

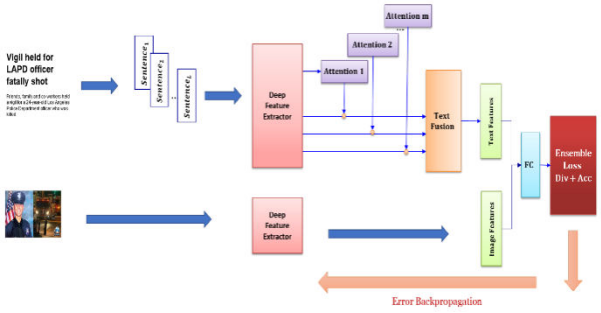
solution for improving information integrity on social media and online platforms.

V. METHODOLOGY

The proposed system for Intelligent Fake News Identification combines feature engineering with an optimized Multi-Support Vector Machine (Multi-SVM) approach to achieve accurate and robust classification. Initially, news articles are collected from multiple sources, including social media platforms and online news websites, ensuring a diverse dataset of both genuine and fake news. The raw data is preprocessed using techniques such as tokenization, lowercasing, removal of stopwords, punctuation, and stemming or lemmatization to reduce noise and standardize text representation. Comprehensive feature engineering is then applied to extract lexical features (e.g., word counts, n-grams), syntactic features (e.g., Part-of-Speech tagging, sentence structure), semantic features (e.g., TF-IDF, Word2Vec embeddings, sentiment analysis), and discourse-level features (e.g., readability scores, headline-body similarity, source credibility). These features are fed into an optimized Multi-SVM classifier, where multiple SVM models with different kernel functions (linear, polynomial, RBF) are combined in an ensemble framework. Hyperparameters such as C, gamma, and kernel parameters are fine-tuned using grid search or cross-validation to maximize performance. The system is evaluated using accuracy, precision, recall, and F1-score on benchmark datasets and real-world news articles. Once validated, this methodology enables real-time fake news detection, providing an automated, scalable, and reliable tool for improving information integrity across online platforms.

VI. SYSTEM MODEL

System Architecture



VII. RESULTS AND DISCUSSIONS

In above screen click on 'New User Sign up' link to get below page

In above screen user is entering sign up details and then press button to get below page

In above screen user sign up completed and now click on 'User Login' link to get below page

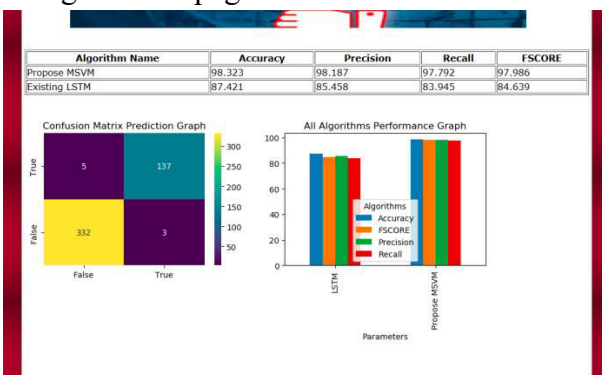
In above screen user is login and after login will get below page

In above screen click on 'Load Fake News' link to load dataset and then will get below page

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

In above screen dataset loaded where first column contains news data and second column contains Class label and FALSE or TRUE. Now click on 'Run MPCA and Firefly' link to extract and select features from dataset and then will get below page

In above screen in first two lines can see number of records and features available in dataset and then MPCA extracted 300 features and then Firefly selected 160 important features and now click on 'Run MSVM Algorithm' link to train algorithms and then will get below page



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

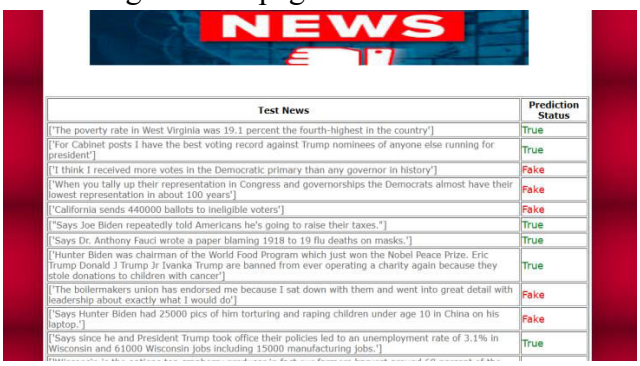
In above screen in table format can see accuracy, precision, recall and FCSORE of Propose MSVM and existing LSTM algorithm. In above table can see MSVM got 98% accuracy and Existing LSTM got 87% accuracy. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and then yellow and light green boxes in diagonal represents correct prediction count and remaining blue boxes represents incorrect prediction count. In bar graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. Now click on 'Predict News' link to get below page

In above screen in text field you can enter some news data and then press button to classify that news text data as False or True or you can upload news data file

In above screen in text field I entered some news data and then press button to get below page

In above screen given news data predicted as 'Fake' and similarly you can test any other news. Now upload news data file

In above screen uploading 'testNews.csv' file and then click on 'Open' and second submit button to get below page



In above screen in first column displaying news text data and in second column displaying classification output as 'False or True'.

VIII. CONCLUSION

The proposed system for Intelligent Fake News Identification Using Feature Engineering and Optimized Multi-SVM provides an effective solution to the growing challenge of misinformation in digital media. By integrating comprehensive feature engineering—including lexical, syntactic, semantic, and discourse-level features—with a multi-kernel Multi-SVM classifier, the system achieves robust and accurate detection of fake news. The optimized ensemble approach improves classification performance across diverse datasets and content types, addressing limitations in existing single-classifier systems. Experimental evaluation demonstrates enhanced accuracy, precision, recall, and F1-score, highlighting the system's reliability for real-world applications. Overall, this approach offers a scalable, automated, and interpretable solution to support information integrity on social media and news platforms, contributing to the mitigation of misinformation and its social impact.

IX. FUTURE WORK

Future research on intelligent fake news identification using feature engineering and optimized Multi-SVM models can focus on expanding the system's capability to handle multi-modal misinformation, including images, videos, and deepfake content. Current models are primarily text-centric, and although feature engineering provides strong discriminative power, integrating visual and audio features can significantly improve robustness. Multi-modal SVM pipelines—where each modality is processed through its own optimized SVM classifier and fused at the decision level—may offer improved accuracy and resilience against emerging misinformation techniques.

Another key direction is the incorporation of context-aware and source-credibility features, such as author reputation, website reliability indicators, propagation patterns, and temporal trends. Introducing these high-level features into Multi-SVM models can improve their ability to detect coordinated disinformation campaigns. Likewise, graph-based features derived from user–post interaction networks could help identify fake news that spreads abnormally fast or originates from suspicious clusters.

Future systems should also explore automated feature optimization techniques using evolutionary algorithms, swarm intelligence, or reinforcement learning. These strategies can dynamically select the most relevant features and hyperparameters for each SVM classifier in the ensemble, further boosting the performance of Multi-SVM models. Additionally, researchers can investigate adaptive real-time fake news detection, where the system continuously learns from new data streams and adapts to evolving linguistic patterns and adversarial manipulations.

To ensure practical deployment, future work should emphasize explainability and transparency, enabling users, journalists, and policymakers to understand the rationale behind model decisions. Integrating explainable AI (XAI) techniques such as SHAP, LIME, and attention-based visualization with Multi-SVM outputs can enhance trust and usability. Finally, large-scale cross-lingual models should be developed to handle misinformation across regional languages, especially in multilingual countries. Building such comprehensive systems will significantly advance the accuracy, reliability, and societal impact of automated fake news detection frameworks.

X. AUTHORS

This project titled *"Intelligent Fake News Identification Using Feature Engineering and Optimized Multi-SVM""* was carried out by **Chinnaru Venkata Ajayreddy** as a part of the academic requirements of the Department of Computer Science and Engineering at **Krishna Chaitanya Institute of Technology & Sciences, Markapur**. The author expresses deep gratitude to the guide for his continuous support, motivation, and valuable guidance throughout the completion of this work.

Dr. SK Althaf Hussain Basha M.Tech, Ph.D, Professor, Department of Computer Science and Engineering, Krishna Chaitanya Institute of Technology & Sciences, Markapur. The author expresses sincere gratitude to the guide for his continuous support, expert insights, and encouragement throughout the development and completion of this work.

XI. REFERENCES

1. Sathyanarayanan, M. &Mahiladevi, S. (2023). *A Fake News Detection System Using Feature-Based Optimized SVM Classification*. International Journal of Science and Innovative Engineering & Technology. IJSIET
2. Vaideghy A. & Thiyagarajan C. (2023). *Detection of Fake News Stance Employing Swarm Intelligence Based Feature*

Extraction and Ensemble Deep Learning. International Journal of Intelligent Systems and Applications in Engineering. IJISAE

3. Ahmed Akib Jawad Karim, Kazi Hafiz Md Asad &Aznur Azam (2024). *Strengthening Fake News Detection: Leveraging SVM and Sophisticated Text Vectorization Techniques, Defying BERT?*arXiv preprint. arXiv
4. Computers (MDPI) Journal (2024). *Enhancing Fake News Detection with Word Embedding: A Machine Learning and Deep Learning Approach*. (MDPI)
5. Mary Ann Cabilao Paulin & Efren I. Balaba (2025). *Distinguishing Truth from Deception: A Machine Learning Approach to Fake News Detection*. International Journal of Latest Technology in Engineering, Management & Applied Science. IJL Temas
6. Abdhesh Nayak & M. Navalan (2025). *Fake News Detection Using Machine Learning and BERT: A Comparative Study*. IJRASET. IJRASET
7. Sudhir Bussa, Aniruddha Bodhankar, Vinod H. Patil, Hemant Pal, Satyendra Kumar Bunkar & Abdul Razzak Khan Qureshi (2023). *An Implementation of Machine Learning Algorithm for Fake News Detection*. International Journal on Recent and Innovation Trends in Computing and Communication. IJRITCC
8. Kirti Randhe, Nameet Vyavahare, Rajkumar Vishwakarma & Sai Kudale (2025). *Fake News Detection Using Natural Language Processing*. International Journal of Scientific Research & Engineering Trends. IJSRET
9. Cybersecurity (SpringerOpen) (2025). *An Unsupervised Fake News Detection Framework Based on Structural Contrastive Learning*. Cybersecurity
10. May Me Me Hlaing et al. (2024). *Fake News Detection Using Effective Feature Extraction and SVM*. Indian Journal of Computer Science and Engineering (IJCSE). ijcse.com
11. SK Althaf Hussain Basha, Kirth Priya Bonthu, Peddi Prasad, "Grade Prediction Using Machine Learning By Analysing Student Behaviour, Quiz Scores, and Participation", International Journal of Computer Science and Mechatronics,Vol.9.Issue.5, pp. 48-56, 2023,ISSN: 2455-1910.
12. SK Althaf Hussain Basha , Y. Sri Lalitha, Y. Gayatri, Application of AI in Speech Emotion Detection, Journal of Critical Reviews (JCR) , Volumex. 7, Issue 06, pp. 3017- 3022 , 2020, ISSN- 2394-5125.
13. Althaf Hussain Basha Sk , Y. Sri Lalitha, Y. Gayathri, "Study Of Machine Learning Classifiers For Sentiment Prediction", International Journal of Computer Engineering And Applications, Volume XII, Issue IX, pp.171-177, November 2019, ISSN 2321-3469
14. SK Althaf Hussain Basha, Nagalakahmi Savala, Venkata Pavan Kumar Savala, G.N.R. Prasad, P M Yohan, "Epidemic Outbreak Prediction According To Social Media Data", International Conference on Multidisciplinary Research and Innovations (ICMDRI-2024), 31-05-2024, Siddhartha Institute of Technology and Engineering, Hyderabad, 2024.
15. Sk Althaf Hussain Basha , P V Ravi Kumar, G.N.R. Prasad ,P M Yohan, "Correlational Analysis of Student Regularity And Performance In Online Learning" , International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NECICAIEA-2K23], 17th and 18th March 2023 , Association with

CSI Narasaraopeta Engineering College, Narasaraopet, ISBN : 978-93-91420-38-3.

16. J.V. Anil Kumar, Naru Kamalnath Reddy, Bollavaram Gopi, Derangula Akhil, Dareddy Indra Sena Reddy, Akkalaakhil , “Language-Based Phishing Threat Detection Using ML And Natural Language Processing”, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : pp. 406-416, ISSN NO : 2249-7455, 2025.
17. J.V.Anil Kumar, Siddi Triveni, Yaragorla Sravya, Mancha Mancha. Venkata Aksh, Posani Lahari Priya, Grandhe Sirisha , “Tools For Database Migration”, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : pp. 760-766, ISSN NO : 2249-7455, 2025.
18. J.V.Anil Kumar, Potluri Rishi Kumar, Shaik Khasim Vali, Jinka Kiran, Gundareddy Manoharreddy,Thotakuri Manikumar, “Revealing Consumer Segments Using Clickstream Data”, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : pp. 670-680, ISSN NO : 2249-7455, 2025.