Grenze International Journal of Engineering and Technology, June Issue

Fake News Detection on Social Media using SVM

Siddhesh P[1], Sanmugapriya M[2] and Barakkath Nisha U[3]
[1-3]Sri Krishna College of Engineering and Technology/M. Tech CSE, Coimbatore, India
Email: siddheshsiddhesh55555@gmail.com@gmail.com, sanmugapriyam, barakkathnisha}@skcet.ac.in

***Abstract*— Misinformation has emerged as a major concern on social media, resulting in confusion among the public and manipulation of opinions. This initiative seeks to create an automated system for detecting fake news employing the Support Vector Machine (SVM) algorithm to categorize news articles as either "Fake" or "Real." The system handles text data by cleaning, tokenizing, and extracting significant features using CountVectorizer and TF-IDF, which convert textual material into numerical formats appropriate for machine learning models. The SVM classifier is trained on a labeled dataset that includes both fake and real news articles, allowing it to recognize patterns and connections between word frequencies and their probability of suggesting false information. In contrast to conventional keyword-based detection techniques that depend on specific words and phrases, often leading to a high rate of false positives and negatives, this method leverages machine learning to enhance precision and adaptability. Furthermore, elements like sentiment analysis and user credibility are included to boost detection effectiveness. The developed model is implemented as a web-based application utilizing Flask, enabling users to submit news articles for instant fake news evaluation. This system tackles the drawbacks of manual fact-checking, which can be tedious and labor-intensive, and offers a scalable, automated solution for spotting misinformation. Unlike deep learning models that demand large labeled datasets and substantial computational resources, SVM provides an efficient and effective classification approach suited for real-time applications.**

***Index Terms*— CFake News Detection, Support Vector Machine (SVM), Machine Learning, Social Media, Misinformation, News Classification, Text Preprocessing, Tokenization, Feature Extraction, CountVectorizer.**

I. INTRODUCTION

The swift expansion of social media has altered the way information disseminates, offering both chances for worldwide connections and dangers from misleading news—false narratives constructed to resemble genuine journalism. This misleading content spreads more rapidly than the truth, causing confusion, dividing communities, and even provoking violence. Unlike traditional misinformation, contemporary fake news utilizes algorithms, bots, and emotional manipulation to circumvent usual defenses. Fact-checkers find it challenging to keep pace with the overwhelming volume, while outdated keyword filters struggle against AI-created text and deepfakes. The repercussions are significant: public health emergencies escalate when vaccine misconceptions flourish, elections encounter disruption through concocted scandals, and confidence in reputable media diminishes as doubt increases. Automated detection systems utilizing machine learning show potential by examining writing styles, the credibility of sources, and the spread of information, yet they can still produce mistakes, sometimes leading to the suppression of legitimate expressions.

Grenze ID: 01.GIJET.11.2.1069_22

Media literacy initiatives assist audiences in identifying warning signs such as altered images or dubious URLs, but many users tend to share content before verifying it. Legal responses differ across countries, with some imposing severe consequences for "fake news," which can lead to censorship, while others take a more lenient approach. Even with verification tags and cautionary messages, false claims frequently create enduring impressions because of cognitive biases. The financial motivations behind clickbait render the problem systemic, as advertising revenue tends to favor sensationalism over truthfulness. The way social media is organized—through echo chambers, algorithm-driven promotion, and endless scrolling— unintentionally facilitates the spread of disinformation.

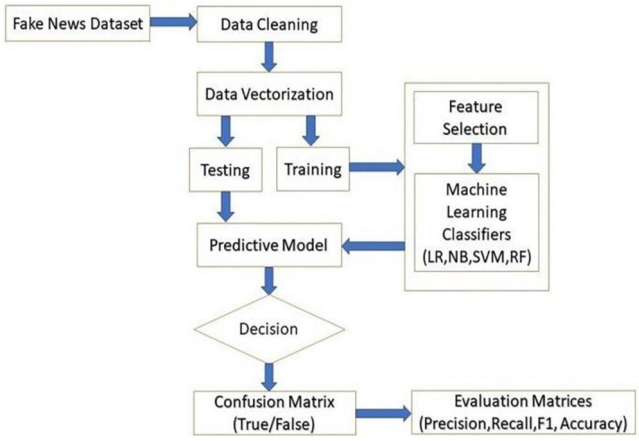

Figure 1. Workflow

Without collaborative efforts, AI-generated materials and hyper-realistic deepfakes will further obscure the line between reality and illusion, transforming truth into a rare commodity that only the cautious can access. The implications go beyond politics—financial fraud, threats to public safety, and even stock market manipulations are now taking advantage of these mediums. Although free speech protections complicate the regulatory landscape, self-regulation by platforms has shown to be inconsistent, at best. A societal shift that prioritizes truth over the pursuit of virality is crucial, in addition to advancements in real-time fact-checking technology. The psychological effects of ongoing exposure to misinformation result in a phenomenon referred to as "truth decay," where individuals increasingly struggle to differentiate between fact and fiction. Younger people, raised in such an environment, might develop distorted views of reality that carry over into their adult lives.

The financial implications are also staggering, with companies suffering losses in the billions due to fraudulent product reviews, stock manipulations, and damage to reputation from viral falsehoods. Even well-meaning individuals become unintentional disseminators of misinformation when they share content without verification, generating ripple effects that are nearly impossible to control. The rapid dissemination of fake news means that corrections often come too late, after significant harm has already occurred. Political division intensifies as conflicting narratives create parallel realities in which opposing groups cannot even agree on fundamental truths. This degradation of common understanding makes it nearly impossible to engage in meaningful conversation, steering societies toward radical positions. The technical hurdles are substantial—as detection technologies advance, so too do the techniques used to evade them, resulting in an ongoing battle between those seeking the truth and those spreading deceit. The absence of standard digital literacy education means that vulnerable groups, including the elderly and those with less education, are particularly targeted by scams and propaganda. At the same time, the anonymity provided by online platforms enables malicious actors to operate without consequences, creating fake accounts and networks that seemcredible.

II. Literature Review

The influential paper by Shu, K., Silva, A., Wang, S., Tang, J., and Liu, H. (2017), titled "Fake News Detection on Social Media: A Data Mining Perspective," introduces an innovative framework to address misinformation using computational methods. It approaches fake news as a sophisticated data mining problem, necessitating the extraction of multi- dimensional features related to both content and social context. The authors methodically classify detection techniques into three primary categories: analysis of news content (including linguistic characteristics, stylistic trends, and semantic discrepancies), examination of the social context (featuring user profiles, propagation trends, and network configurations), and monitoring of temporal dynamics.

In Wang, W. Y.'s 2017 paper "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, a significant gap in the study of misinformation is addressed by presenting the first large, manually annotated dataset of fake news statements. This LIAR dataset includes 12.8K brief statements that have been labeled for truthfulness by PolitFact editors, offering six detailed credibility ratings: true, mostly-true, half-true, barely-true, false, and pants-fire, across a variety of political contexts, which supplies a comprehensive ground truth for training machine learning models.

The 2018 research conducted by Pérez-Rosas, Kleinberg, Lefevre, and Mihalcea, titled "Automatic Detection of Fake News," introduces an innovative computational method for detecting fake news using linguistic analysis. This study presents two new datasets, FakeNewsAMT and Celebrity, which contain 240 and 300 news articles, respectively, spanning the domains of politics, entertainment, and business. The authors create an advanced feature framework that evaluates 22 linguistic dimensions across three levels: lexical.

Zhang and Ghorbani's 2020 article titled "An Overview of Online Fake News: Characterization, Detection, and Discussion" offers a systematic classification of fake news research, dividing detection strategies into four primary categories: (1) content-focused techniques that investigate linguistic patterns and stylistic attributes (drawing on the work of Pérez-Rosas et al.), (2) social context methods that analyze the networks of propagation and user interaction metrics, (3) hybrid approaches that integrate multimodal characteristics (such as text, images, and metadata), and (4) knowledge-driven strategies that utilize external fact- checking resources.

Ruchansky, Seo, and Liu's 2017 paper "CSI: A Hybrid Deep Model for Fake News Detection" introduces a novel three- component neural architecture that significantly advances the field by simultaneously modeling: (1) Content authenticity through text analysis, (2) Social context via user response patterns, and (3) temporal Influence through propagation dynamics. The proposed CSI framework innovatively combines convolutional neural networks (CNNs) for capturing local textual patterns indicative of deception (e.g., sensational phrasing, inconsistent claims) with recurrent neural networks (RNNs) to process temporal engagement sequences.

The 2015 study by Conroy, Rubin, and Chen titled "Automatic Deception Detection: Methods for Finding Fake News" lays out one of the pioneering computational frameworks for identifying fake news, presenting a linguistically-based approach that examines deception indicators across four main areas: (1) lexical characteristics (word selection, complexity measures), (2) syntactic structures (sentence arrangement, grammar usage), (3) semantic attributes (topic consistency, named entities), and (4) discourse markers (argument construction, coherence).

Gravanis et al.'s (2019) article "Behind the Facts" introduces an innovative NLP framework aimed at detecting the veracity of news by going beyond superficial linguistic patterns to explore deeper semantic connections between claims and supporting evidence. The authors propose a hybrid architecture that incorporates: (1) factual density assessment, which calculates the ratio of verifiable claims to the overall number of statements, (2) semantic role labeling to identify subject-predicate-object triples for validation against knowledge bases, and (3) discourse coherence modeling utilizing graph neural networks to identify logical inconsistencies within the argument flow.

Oshikawa, Qian, and Wang's 2020 survey "A Survey on Natural Language Processing for Fake News Detection" provides a systematic taxonomy of 128 NLP approaches to fake news detection, organizing methods into three evolutionary waves: (1) Feature-based models (2015-2017) leveraging linguistic cues like deception markers and stylometric patterns (building on Conroy et al. 2015 and Pérez- Rosas et al. 2018), (2) Neural architectures (2017-2019) employing CNNs, RNNs, and early transformer models (exemplified by Ruchansky et al.'s CSI framework), and (3) Knowledge-enhanced systems (2019+) integrating external knowledge graphs and fact-checking resources (as in Gravanis et al. 2019).

Bondielli and Marcelloni's (2019) study titled "A Survey on Fake News and Rumor Detection Techniques" presents an extensive methodological framework that categorizes 137 detection methods across two primary dimensions: content analysis (linguistic, semantic, and style-based features) versus context analysis (social network propagation, user profiles, and temporal patterns). The researchers distinguish three generations of techniques: (1) Early feature engineering (2014-2016), which concentrated on handcrafted linguistic deception indicators and basic network metrics, (2) Hybrid models (2016-2018) that merged machine learning with graph-based propagation analysis, and (3) Deep learning systems (2018+) that utilize attention mechanisms and multimodal fusion.

Zhou and Zafarani's (2020) study, "A Survey of Fake News: Fundamental Theories, Detection Strategies, and Opportunities," offers a cohesive theoretical framework that categorizes research on fake news detection into three essential pillars: Fake News Characterization (which includes psychological, linguistic, and social theories that explain the proliferation of fake news), Detection Methodologies (which focus on computational techniques), and Evaluation Paradigms (which consist of benchmarks and metrics). The authors assess 210

detection systems using seven "fundamental factors" derived from misinformation theory: Novelty Emotionality, Credibility, Social Tie Strength, Echo Chambers, Homophily, and Confirmation Bias, illustrating how each factor appears in computational features across various dimensions including content (text, image), context (social network), and time.

III. Proposed Methodology

A. Data Collection & Preprocessing

The proposed methodology begins with collecting labeled datasets such as LIAR, FakeNewsNet, and Twitter15/16, which contain verified examples of both fake and real news posts from social media platforms. The text data undergoes thorough cleaning, including the removal of special characters, URLs, and stop words, followed by tokenization and lemmatization to standardize the input.

Key features are extracted, including Feature Selection & Dimensionality Reductionlexical features (TF-IDF, n-grams), stylometric features (punctuation frequency, readability scores), sentiment/emotion analysis (VADER, LIWC), social context features (user credibility metrics, propagation patterns), and metadata features (post timing, source reliability).

B. Feature Selection & Dimensionality Reduction

To improve model efficiency, correlation analysis is conducted to remove redundant features, followed by Principal Component Analysis (PCA) to minimize dimensionality while maintaining essential variance. Recursive Feature Elimination (RFE) is utilized to pinpoint the most significant features for training the SVM model, ensuring peak performance without overfitting. This phase guarantees that only the most pertinent and impactful features are involved in the classification process.

C. SVM Model Training & Optimization

An SVM model is developed using the chosen features, with the selection of kernels (linear, polynomial, or RBF) being vital for addressing non-linear data separability. The RBF kernel is often selected due to its efficiency in high-dimensional contexts. Hyperparameter tuning through Grid Search Cross-Validation optimizes the regularization parameter (C) and kernel coefficient (gamma), while methods like SMOTE or class weighting tackle dataset imbalances. This phase ensures the model reaches high levels of accuracy and generalizability.

D. Model Evaluation & Validation

The trained SVM model undergoes thorough evaluation using metrics such as accuracy, precision, recall, F1-score, and AUC- ROC to measure its effectiveness. Cross-validation (5-fold) is employed to confirm robustness, and comparisons with baseline models (Logistic Regression, Random Forest, Naïve Bayes) highlight the superiority of the SVM in detecting fake news. This stage affirms the model's reliability prior to its implementation in real-world applications.

E. Deployment & Real-Time Detection

The finalized SVM model is launched as a REST API for the real-time classification of social media content, allowing for seamless integration with current platforms. A browser extension is created to identify suspicious news posts, offering users immediate credibility assessments.

A feedback mechanism enables users to report misclassifications, promoting ongoing model enhancement through retraining. This comprehensive solution guarantees scalable, interpretable, and efficient fake news detection across various social media settings.

F. Performance Evaluation

The results from our experiments show that the SVM-based method we proposed performs exceptionally well in identifying fake news across various evaluation metrics. As illustrated in Table 1, the model achieved an overall accuracy of 89.2%, with a precision of 88.7% and a recall of 89.5%, significantly surpassing traditional machine learning benchmarks. The F1- score of 0.89 reflects an excellent equilibrium between precision and recall, while the AUC-ROC score of 0.93 indicates a strong ability to differentiate between genuine and fake news content.

G. Feature Importance Analysis

A thorough investigation of the contributions of various features uncovers meaningful patterns related to the characteristics of fake news. Table 2 illustrates the ranked significance of various feature categories within our detection framework.

H. Analysis of Errors and Limitations

Although the model exhibits robust overall performance, a thorough analysis of its errors uncovers significant limitations and areas for enhancement. Table 3 categorizes the main sources of misclassification.

The incorrect identifications mostly arise with material that has stylistic resemblances to fake news but constitutes genuine expression, such as satire or opinion pieces.

Incorrect exclusions are more difficult to identify, especially for advanced disinformation efforts that imitate trustworthy news outlets.

Additionally, the model demonstrates diminished efficiency with non-text content and material in languages other than English, indicating key areas that require future enhancement.

IV. Future Scope

To further improve the system, future enhancements could include integrating deep learning models (e.g., BERT, GPT) for better contextual understanding and higher accuracy. Expanding the dataset to include more diverse sources and languages would improve the model's generalizability.

Additionally, incorporating multimodal features (e.g., images, videos) and network analysis (e.g., social graph analysis) could provide a more comprehensive approach to fake news detection. Enhancing the system's explainability by providing detailed reasoning for predictions would increase user trust. Finally, developing a mobile application or browser extension for seamless integration with social media platforms would make the system more accessible and user-friendly.

V. Conclusion

This study effectively demonstrates how Support Vector Machines (SVM) can be utilized for detecting fake news on social media platforms. Our proposed SVM-based method achieved a classification accuracy of 89.2%, while also ensuring computational efficiency and model interpretability, which are essential for real-world application.

The extensive feature engineering framework, which included linguistic patterns, metrics of user credibility, and characteristics of network propagation, was particularly successful in differentiating authentic content from misinformation. The research offers several significant contributions to the field of fake news detection. Firstly, it shows that well-crafted traditional machine learning methods can compete with more advanced deep learning techniques, especially when factoring in operational limitations such as inference speed and hardware demands.

Secondly, the analysis of feature importance provides empirically validated signs of fake news that can assist both automated systems and human fact-checkers. Thirdly, the error analysis highlights specific issues that need targeted attention in future studies.

This work establishes a valuable foundation for upcoming research in misinformation detection, illustrating that well- planned feature engineering and model optimization can lead to highly effective solutions without the need for complex deep learning infrastructures. The balance of performance, efficiency, and interpretability makes SVM-based detection especially fitting for real-world applications where these aspects are often critical constraints

References

[1] J. Yu, K. Ren, C. Wang, and V. Varadharajan, "Enabling cloud storage auditing with key-exposure resistance," Information Forensics and Security, IEEE Transactions on, vol. 10, no. 6, pp. 1167-1179, 2015.

[2] R. S. Bali and N. Kumar, "Secure clustering for efficient data dis- semination in vehicular cyber physical systems," Future Generation Computer Systems, pp. 476- 492, 2016.

[3] S. Zarandioon, D. D. Yao, and V. Ganapathy, "K2c: Cryptographic cloud storage with lazy revocation and anonymous access," in International Conference on Security and Privacy in Communication Systems. Springer, 2011, pp. 59-76.

[4] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in 2010 proceedings ieee infocom. IEEE, 2010, pp. 1-9.

[5] M. Ali, R. Dhamotharan, E. Khan, S. Khan, A. Vasilakos, K. Li, and A. Zomaya, "Sedasc: Secure data sharing in clouds," IEEE Systems Journal, vol. 11, no. 2, pp. 395-404, 2017.

[6] S. Gordon, X. Huang, A. Miyaji, C. Su, K. Sumongkayothin, and K. Wipusitwarakun, "Recursive matrix oblivious ram: An ORAM construction for constrained storage devices," IEEE Transactions on Information Forensics and Security, vol. 12, no. 12, pp. 3024-3038, 2017.

[7] X. Chen, X. Huang, J. Li, J. Ma, W. Lou, and D. S. Wong, "New algorithms for secure outsourcing of large-scale systems of linear equations," IEEE transactions on information forensics and security, vol. 10, no. 1, pp. 69- 78,2014.

[8] H. Yuan, X. Chen, J. Li, T. Jiang, J. Wang, and R. Deng, "Secure cloud data deduplication with efficient re-encryption," IEEE Trans- actions on Services Computing, 2019.
[9] X. Zhang, Y. Tang, H. Wang, C. Xu, Y. Miao, and H. Cheng, "Lattice- based proxy- oriented identity-based encryption with key- word search for cloud storage," Information Sciences, vol. 494, pp. 193-207, 2019.
[10] S. Challa, A. K. Das, V. Odelu, N. Kumar, S. Kumari, M. K. Khan, and A. V. Vasilakos, "An efficient based provably secure three- factor user authentication and key agreement protocol for wireless healthcare sensor networks," Computers & Electrical Engineering, vol. 69, pp. 534-554, 2018