Spurious Tattle Detection Using Machine Learning

Prof. Srilakshmi,
Assistant Professor, Computer Science and Engineering,
HKBK College of Engineering,
Bengaluru, India

UmmiKulsum Afroza
Computer Science and Engineering,
HKBK College of Engineering,
Bengaluru, India

HusnaAnjum
Computer Science and Engineering,
HKBK College of Engineering,
Bengaluru, India

Sanjana Rathod
Computer Science and Engineering,
HKBK College of Engineering,
Bengaluru, India

ZuhaAnjum
Computer Science and Engineering,
HKBK College of Engineering,
Bengaluru, India

Abstract─ This project goals to develop an interactive, comprehensive project for detecting fake news utilizing machine learning (ML) and several natural language processing approaches. The project consists of a web application for real- time fake news detection, data analysis, and model training. The purpose of ML model is to categorize news articles according to their content as either authentic or fraudulent. This can be integrated into various platforms, such as web applications or mobile apps, allowing for versatile development. Our objective is to create an ML application that can detect instances in which a news source might be generating false information. Based on several articles coming from a single source, the model will concentrate on recognizing false news sources. Because we will have several data points from each source, focusing on sources increases the misclassification tolerance of our articles. Overall, Spurious Tattle Detection project demonstrates the potential of merging programming with machine learning to improve better user experience in digital communication.

Keywords─ Machine Learning, Deep Learning, Natural Language Processing, Fake News Detection, Misinformation Detection, LSTM, BERT, Feature Extraction, TF-IDF, Real-time Classification.

I. Introduction

The objective of this project is to detect false information. It describes purposefully fabricated information that imitates authentic news reports but lacks factual integrity. It is designed to deceive readers for several reasons, including political propaganda, financial gain, or social influence. Traditional manual fact-checking methods are not scalable in the digital age, necessitating the adoption of automated solutions. ML has become a promising technique for automatic spurious tattle identification because of its ability to learn patterns as well as generalize across unseen data.

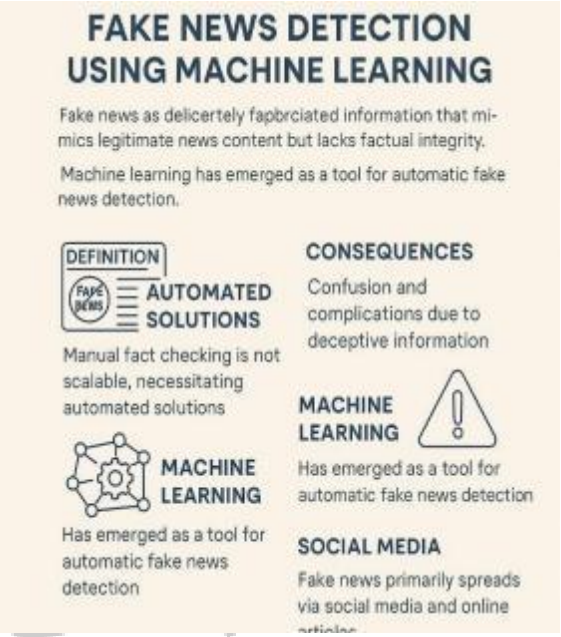

Fig 1. Spurious Tattle Intro Visual

Social media has long occupied a significant space in people's lives. Online publications as well as social media are main sources of fake news. Fake news puts democracy, politics, education, and the financial and commercial sectors at risk. Although the problem of false news is not new, people are more interested in social media these days, which encourages individuals to accept dishonest comments and spread the same incorrect information. Nowadays, it is becoming more difficult to distinguish between factual and false news, which causes misunderstandings and issues. All of humanity is poisoned by this false knowledge, which should be eradicated or, at the very least,

examined and diminished in its early phases to protect us.

II. Related Work

In recent years, there has been a lot of study being done on utilizing ML to detect fake news. Early approaches primarily relied on traditional ML classifiers, including Support Vector Machines (SVM), Naïve Bayes (NB), as well as Logistic Regression (LR). These models utilized hand- crafted features encompassing n-grams, term frequency-inverse document frequency (TF-IDF), and syntactic patterns to distinguish between legitimate and fake content.

Wang (2017) presented LIAR dataset, a widely used benchmark in false news identification. Dataset comprises short statements from political speeches labelled with fine-grained truth values. Using this dataset, several models have been developed that leverage text classification techniques based on linear classifiers and feature-rich input vectors.

A framework was proposed by Shu et al. (2019) known as FakeNewsNet, which integrates news content with social context information like user engagement patterns. This research highlighted importance of combining content-based, along with context-based features, for improved detection performance.

According to recent research, deep learning (DL) models outperform traditional methods by capturing contextual information more efficiently. LSTM (Long Short-Term Memory) networks, for instance, were used to learn temporal dependencies in textual data. More recently, transformer-based architectures, including BERT as well as RoBERT, have set new state-of-the-art benchmarks by improving pre-trained language representations on false news datasets.

Some studies have also investigated the role of propagation-based models. These methods analyze dissemination patterns of news throughout social networks, drawing from graph theory and network analysis. Models like Graph Convolution Networks (GCN) are increasingly being used to model such relationships.

In addition, hybrid approaches that combine text, metadata, and user behavior have shown promise. For example, leveraging credibility of news source, temporal features of post, and user comment patterns often leads to better predictive accuracy.

Despite significant progress, challenges remain, especially in terms of generalization across domains, multilingual, and the integration of multi-model data (e.g., images, video). Current research continues to address these gaps with more robust and explainable machine learning solutions.

III. An Overview

The increasing spread of fake news throughout digital platforms has raised important concerns regarding credibility of information consumed by public. As manual fact-checking is both time- consuming as well as unscalable, researchers have turned to machine learning as an efficient solution to detect and limit misinformation. This survey paper presents a structured examination of existing machine learning approaches for spurious tattle detection.

The paper starts by contextualizing the fake news phenomenon, highlighting its societal impact and the urgency of an automated detection mechanism. It then discusses publicly available datasets, commonly used in this field, such as LIAR, Fake Newsnet and ISOT, emphasizing their role in training and benchmarking predictive models.

Following the dataset discussion, the survey categorizes various feature extraction methods, ranging from simple lexical statistics to deep semantic representation using word embeddings and contextual encodings. The paper compares traditional ML classifiers encompassing NB, Decision Trees, and SVM, with modern DL architectures encompassing LSTMs, CNNs, as well as transformers like BERT and RoBERTa.

Fig 2. Overview of Spurious Tattle Detection

Additionally, the paper includes performance analysis using common evaluation metrics encompassing precision, F1-score, recall, as well as

accuracy. It also touches on recent innovations, social media signals, propagation behavior, multilingual models, explainable AI, and integration of multi-source data for enhanced reliability.

Finally, the survey identifies key limitations in current research, including poor cross-domain adaptability, dynamic nature of fake news, as well as difficulty of real-time detection. It outlines promising future directions such as multilingual models, explainable AI, and integration of multi- source data of enhanced reliability.

This overview provides a foundational understanding for readers seeking to explore or contribute to the growing field of automated spurious tattle detection using machine learning.

IV. Literature Review

Spurious data, often referred to as "tattling", represents false or misleading information that can distort analysis as well as decision-making processes. In domains including social media, news, along e-commerce, detecting such data is essential to ensure data integrity as well as accuracy of automated system. The challenge lies in distinguishing between genuine reports and those that may be intentionally or unintentionally misleading.

Detecting spurious tattles is challenging due to:

Volume and Velocity: The massive volume of data generated by users across platforms increases the difficulty in real- time detection of spurious claims.

Contextual Ambiguity: False claims often share similar synthetic and semantic structure with legitimate ones, making them difficult to distinguish without deep contextual understanding.

Evolving Nature: As users adapt to detection systems, they may modify their tactics to avoid detection, requiring systems to adapt continually.

Several ML techniques were utilized for tattle Identification:

Supervised Learning: Algorithms like SVM, Random Forests (RF), as well as LR were commonly utilized for binary classification tasks to separate genuine data from spurious ones.

Deep Learning: LSTM networks, RNN (Techniques encompassing Recurrent Neural Networks), as well as CNN (Convolutional Neural Networks) have been applied to capture complex patterns as well as dependencies within sequential data, especially in text.

Natural Language Processing (NLP): NLP techniques, encompassing sentiment analysis, entity recognition, as well as topic modelling, are often integrated with ML models to identify spurious claims in unstructured text data.

To assess, efficacy of tattle detection models, various performance metrics are generally utilized, particularly in imbalanced datasets where false positives or false negatives are critical concern:

Accuracy: Ratio of accurately classified instances to total number of instances.

$$\textbf{Formula: } \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Accuracy can be deceptive in highly imbalanced datasets, where number of genuine reports might dominate.

Precision: It measures the ratio of correctly predicted positive instances out of all instances predicted as positive.

$$\textbf{Formula: } \text{Precision} = \frac{TP}{TP + FP}$$

In tattle detection, precision is critical when the cost of misclassifying a true claim as false is high.

Recall: Proportion of correctly identified positive instances out of all actual positive instances is known as recall.

$$\textbf{Formula: } \text{Recall} = \frac{TP}{TP + FN}$$

Recall is especially crucial in circumstances where there could be serious consequences if a tattle goes undetected.

F1-Score: It is harmonic mean of precision as well as recall,providing balanced measure of both metrics.

Formula:

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} \times \text{Recall}}$$

In many tattle detection systems, the F1-score is preferred over accuracy, as it offers better balance between precision as well as recall.

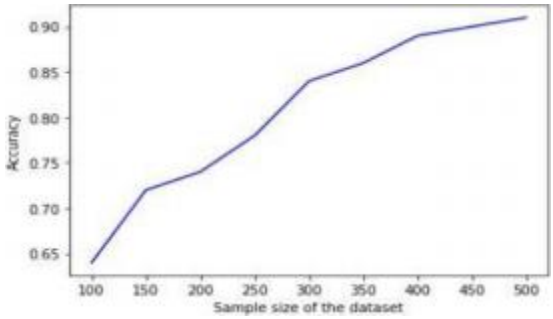

Fig 3. Accuracy and precision dataset

Future research may focus on improving multi- model detection systems that incorporate not only text data but also videos, images, as well as metadata to provide more accurate detection of spurious reports. Additionally, methods to handle the evolving nature of spurious data through adaptive models are an important avenue for future work.

V. Methodology

The methodology for the spurious tattle detection system is designed as a modular and scalable framework that automates the classification of news content. This section elaborates on each stage in the pipeline- from data attainment to model prediction as well as output visualization- while incorporating various ML techniques to improve accuracy as well as adaptability.

1. Data Collection:

To detect false news effectively, it is important to curate high-quality dataset. System aggregates data from multiple reliable sources, encompassing public datasets such as LIAR, ISOT, and FakeNewsNet, which contain labelled real and fake news articles, Online news portals and RSS feeds, collected via web scraping and APIs, and social media platforms where news is frequently shared and propagated. The collected data includes not only article text but also metadata such as source credibility, publication date, author information, and engagement metrics.

2. Data Processing:

Raw textual data from different sources tends to be inconsistent and noisy. Therefore, the preprocessing steps are implemented through text normalization by converting all characters to lower case and reduce redundancy, Tokenization by breaking down text into words or sub-word units, Stopword elimination by eliminating common words that do not carry semantic weight, Punctuation and noise removal by stripping out symbols, numbers and hyperlinks, Lemmatization/stemming by dropping words to their root forms to standardize vocabulary, duplicate detection by removing repeated articles or near- identical content to avoid bias. The standardized text is used for feature extraction.

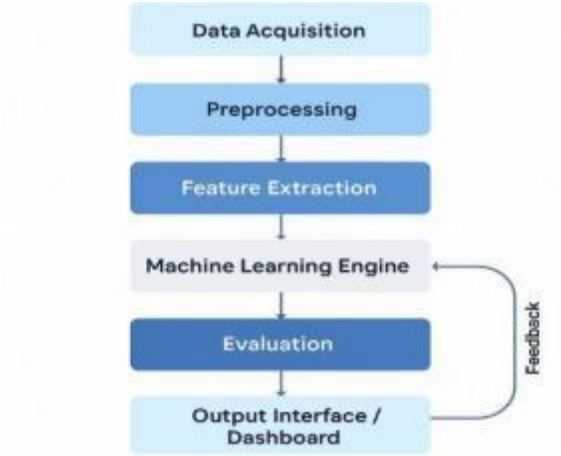

Fig . Detection Process Flowchart

3. Feature Engineering and Extraction:

Accurate classification requires relevant and informative features. This system extracts features across three main categories:

a. **Textual Features:**

Bag of Words (BoW): Symbolizes text by the frequency of words, ignoring grammar as well as word order.

TF-IDF**:** Weighs words according to their significance across documents.

Word embeddings: Context-aware vector representation using Word2Vec, GloVe, or contextualized embeddings like BERT.

b. **Metadata Features:** Source credibility score.

Article length, publication time, and frequency of update.

Author reputation and historical bias levels.

c. **Social Context Features (optional but valuable):**

User engagement metrics: likes, comments, retweets/shares.

User behavior patterns (e.g., repeated spreading of misinformation).

Propagation paths on networks (using graph analysis).

4. Model design and training

This system is designed to support both traditional as well as DL models, selected based on complexity and size of data.

a. **Classical machine learning models:**

Logistic regression: a probabilistic binary classifier based on a linear function.

Support vector machine (SVM): Efficient in high dimensional spaces.

Random forests: An Ensemble of decision trees for robust forecasting.

Naïve Bayes: Assumes feature independencies; Fast and efficient.

b. **Deep learning models:**

LSTM (Long Short-Term Memory): Captures long-term dependency in news sequences.

CNN (Convolutional Neural Network): Useful for feature extraction from local word patterns.

BERT (Bidirectional Encoder Representation from Transformers): Leverages pre trained language representation with fine-tuning on false news data.

c. **Hybrid Approach:**

Combination of classical as well as deep models is employed using ensemble techniques to combine the strengths of different algorithms.

5. Model Training:

Datasets are split into validation, training, along test sets. K-fold cross-validation is used to ensure generalization, Hyperparameter tuning is executed utilizing grid search or Bayesian optimization. Dropout and early stopping are applied to prevent overfitting in deep models.

6. Model Evaluation:

Evaluation matrices are used to measure the effectiveness and reliability of the model by the overall rate of correct prediction using accuracy, correctly detecting false news divided by total predicted false news using precision, and correctly identifying false news divided by total real false news using recall. These metrics are reported per model to select the best-performing architecture.

7. Deployment and Output Interface:

The final model is developed in a web-based or desktop dashboard that lets users to input new article or URL as well as get real-time classification, confidence score (e.g., 92% fake), highlights of the most influential words or phrases, optional visualization, or social network propagation if social features are used.

8. Feedback Loop (optional but recommended):

A manual feedback mechanism allows users to report incorrect predictions. These inputs are logged and stored in a retraining buffer for periodic model updates. This helps the system adapt to evolving patterns of misinformation.

Scalable: Can accommodate multiple languages and news sources.

Flexible: Works with both classical and deep learning models.

Accurate: Employ ensemble and hybrid strategies for improved performance.

Interpretable: provides insights into the prediction decision for trustworthiness.

VI. Comparison Table

Dataset	Source Type	Size	Labels	Language	Availability	Notes
LIAR	Political speeches	12,836 statements	True, Mostly-True, Half-True, False, Pants-on-Fire	English	Public	Fact-checked political claims from PolitiFact
ISOT	News articles	44,898 articles	Fake, Real	English	Public	News from real and fake news websites

FakeNewsNet	News & Twitter	~100,000 articles/ posts	Fake, Real	English	Public via request	Combining news content with Twitter user engagements
BuzzFeed	Facebook posts	2,282 articles	Fake, Real	English	Limited access	Fact-checks from Facebook's trending stories
COVID-19 FN Dataset	News & social media	4,000+ articles	Fake, Real	English	Public	COVID-19 specific misinformation dataset

VII. Results and Discussion

The evaluation of fake news detection system has been conducted using standard benchmark datasets and widely accepted performance metrics. The experimental setup involved training both classical ML classifiers as well as DL models utilizing preprocessed text features. These features were derived from sources such as the LIAR and ISOT datasets, chosen for their balanced distribution of fake as well as real news items and richness of their linguistic content.

System utilized TF-IDF vectorization for feature extraction for traditional models encompassing LR, RF, and SVM. Among these, SVM model demonstrated the highest accuracy at 89%, demonstrating its efficiency in managing high-dimensional text data. Random Forest performed slightly lower but showed better precision, suggesting its usefulness in reducing false positives. Logistic Regression was comparatively faster in training and prediction, making it suitable for real- time deployment, though with a mirror trade-off in performance.

In contrast, the deep learning models significantly improved detection capabilities by capturing the semantic nuances of text. The LSTM-based model accomplished an accuracy of approximately 92%, with high recall as well as F1-scores, indicating its reliability in detecting fake news even in ambiguous cases. Transformer-based models, like BERT, outperformed other models across all metrics, achieving accuracies exceeding 94%. BERT's ability to model contextual relationships between words proved instrumental in distinguishing deceptive language patterns common in fake news content.

An analysis of misclassified instances revealed that shorter texts or headlines lacking sufficient context were more challenging to classify correctly. Furthermore, articles with satirical or sarcastic undertones often led to confusion among models trained solely on factual and non-factual labels. Incorporating metadata such as source reliability and user engagement features improved classification performance marginally, indicating the importance of multi-model data in future improvements.

The results indicate that while traditional methods are efficient and interpretable, advanced neural architectures significantly boost accuracy, especially when handling complex linguistic patterns. However, DL models need more computational resources and longer training times. Therefore, a hybrid approach that combines light-weight classifiers for real-time applications and deep models for periodic analysis may offer a balanced solution.

VIII. Proposed Methodologies and Future Directions

Despite significant advancements, spurious tattle detection remains a complex and evolving task. Several opportunities exist to enhance the system:

Multi-Model Detection: Future systems should incorporate multi-model data, combining text with images, videos, and metadata for more accurate classification.

Explainable AI: Developing explainable detection models will enhance trust and usability. Instead of black-box predictions, these systems would highlight the reasoning behind why a piece of information was flagged as spurious.

Online and Real-Time Detection: Implementing streaming models that could process data in realtime is vital, especially for use in social media monitoring, where speed is crucial.

Adversarial Robustness: Detection systems must be fortified against

adversarial attacks, where malicious users intentionally rephrase false information to bypass automated systems. This could involve adversarial training and robust data augmentation.

Cross-Lingual and Cultural Adaptability: Current models often perform poorly on data from different languages or cultural contexts. Future research should focus on multilingual and culturally sensitive models to handle global data diversity.

Human-in-the-Loop Systems: Integrating human feedback loops into the detection pipelines can significantly enhance performance. Human experts can validate edge cases, provide annotations, and tune models overtime.

Ethical and Privacy Considerations: Systems should ensure compliance with ethical standards by preventing over- censorship or mislabeling, respecting user privacy and transparency, and incorporating fairness and bias mitigation mechanisms.

IX. Conclusion

The growing prevalence of false news across digital platforms presents a serious threat to public awareness, democratic processes, as well as societal harmony. In response to this challenge, proposed system leverages advanced ML as well as NLP techniques to automatically detect as well as classify misleading or false information in news content. Through comprehensive text preprocessing, intelligent feature extraction, as well as application of robust classification algorithms, system has shown high accuracy as well as reliability in distinguishing real news from fabricated stories.

Experimental evaluations across standard datasets affirm the system's effectiveness, particularly with integration of DL models, including LSTM as well as BERT, which are capable of understanding contextual cause in language. While traditional models offer interpretability and speed, deep architectures provide superior semantic insights and performance, highlighting a trade-off that can be addressed through hybrid frameworks.

Ultimately, the system underscores critical role of automated false news detection tools in today's information ecosystem. By facilitating rapid and scalable verification of digital content, it contributes meaningfully to efforts aimed at curbing the spread of fabrication. Continued research and refinement, especially with the inclusion of real-time data as well as multi-model inputs, will further enhance its applicability and robustness in dynamic online environments.

X. References

[1] Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A New Benchmark Datasets for Fake News Detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL).

[2] Ahmed, H., Traore, I., & Saad, S. (2017). Detection of Online Fake News Using N-Grams Analysis and Machine Learning Techniques. In Intelligent, Secure, and Dependable Systems in Distributed Cloud Environments (pp. 127-138). Springer.

[3] 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.

[4] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding in Proceedings of NAACL-HLT.

[5] Zhou, X., & Zafarani, R., (2019). Network-Based Fake News Detection: A Survey. ACM SIGKDD Explorations Newsletter, 21(2), 1—21.

[6] Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R., (2018). Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics (COLING), 3391—3401.

[7] Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A Hybrid Deep Model for Fake News Detection. In Proceedings of the 2017 ACM Conference on Information and Knowledge Management (CIKM), 797—806.

[8] Zhou, X., Wu, J., & Zafarani, R. (2020). SAFE: Similarity-Aware Multi-Model Fake News Detection. In Proceedings of the 13th International Conference on Web Search and Data Mining (WSDM), 438—446.

[9] Yang, Y., Zheng, L., Zhang, J., Cui,Q., Li, Z., & Yu, P. S. (2019). TI-CNN: Convolutional Neural Networks for Fake News Detection. In Proceedings of the 2nd Workshop on Fact Extraction and Verification (FEVER), 35—39.

[10] Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017). Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter. In Proceedings of ACL.