Detecting Fake News on Social Media using Fine-Grained Classification and Deep Learning Techniques

[1]Goli Sravya, [2]Dr. K. Santhi Sree

[1]Student, Department of Information Technology, University College of Engineering, Science and Technology,

JNTUH Hyderabad

[2]Professor & Head of Department,Department of

Information Technology,University College of Engineering,

Science and Technology,JNTUH Hyderabad

Abstract: In response to the pressing need for accurate fake news detection amid the proliferation of misinformation on social media, this study introduces an innovative approach leveraging graph neural networks (GNN) for classification. Unlike existing methods, our proposed model focuses on analyzing sentence interaction patterns within news articles to achieve fine-grained fake news classification. By constructing a graph representation of news articles and utilizing GNN, we effectively capture the nuanced relationships between sentences, enhancing contextual understanding. Specifically, we employ a third-order co-occurrence tensor and canonical polyadic (CP) decomposition to compute weight matrices, enabling precise representation of local word co-occurrence information. In our evaluation, various models including SVM, LSTM, CNN, BERT GCN, and GCN with CP were compared, with GCN exhibiting remarkable accuracy of 99%. To further enhance performance, we introduce ensemble techniques, combining predictions from multiple models. As an extension, we propose exploring ensemble methods such as BERT GCN LSTM and LSTM + GRU, aiming to achieve even higher accuracy, potentially reaching 100%. This comprehensive approach promises to significantly improve fake news detection, contributing to the mitigation of its adverse societal impacts.

Index terms - *Fake news classification, graph convolutional network (GCN), long short-term memory (LSTM), tensor decomposition.*

1. INTRODUCTION

In today's digital age dominated by social media, accessing and consuming news has undergone a profound transformation. The ease and convenience of sharing information have revolutionized communication, allowing news to spread rapidly across various platforms. However, this unprecedented accessibility has also led to the rampant dissemination of misinformation, particularly in the form of fake news. Fake news, defined as deliberately misleading or false information presented as legitimate news, poses significant challenges to individuals, societies, and democratic processes [1].

The proliferation of fake news has become a pressing concern for the public, drawing attention from researchers, policymakers, and media practitioners alike [1]. Although the concept of fake news is not new, its prevalence and impact have been magnified by the digital age. In the pre-internet era, news verification primarily relied on the rigorous fact-checking processes employed by journalists and traditional media outlets [2]. However, with the advent of social media platforms and digital technologies, information dissemination has been democratized, enabling anyone to create and share content at unprecedented speed and scale. Consequently, the traditional gatekeeping function of professional journalism has been disrupted, leading to a flood of unverified and misleading information circulating online [2].

The sheer volume and velocity of fake news present significant challenges to traditional verification methods. Unlike traditional journalism, where news verification is a time-intensive and resource-demanding process, the rapid dissemination of information on social media often outpaces the capacity for thorough fact-checking [1]. Therefore, there is an urgent need for automated solutions that can swiftly and accurately detect and classify fake news, complementing and accelerating the efforts of human fact-checkers.

Detecting fake news is a multifaceted challenge due to its diverse forms, styles, and motivations. Fake news articles can range from fabricated stories intended to deceive readers for political or financial gain to misleading content that distorts facts or presents biased perspectives [1]. Moreover, the dynamic nature of online discourse means that fake news can evolve rapidly, adapting to exploit emerging trends and narratives. Consequently, effective fake news detection strategies must be adaptable and capable of identifying deceptive content across a broad spectrum of contexts and formats.

Researchers have proposed a multitude of detection methods drawing on various disciplines, including computer science, linguistics, and social network analysis [3]. Comprehensive reviews of existing methodologies offer valuable insights into the diverse approaches and challenges in fake news detection [3]. These approaches can be broadly categorized into three main categories: content-based, network-based, and hybrid methods.

Content-based approaches to fake news detection focus on analyzing the linguistic and semantic characteristics of news articles to discern patterns associated with deception [6], [7], [8], [9], [10], [11]. By leveraging natural language processing (NLP) techniques such as syntactic analysis, sentiment analysis, and semantic modeling, these methods aim to identify linguistic cues indicative of fake news. For example, lexical features such as vocabulary richness, sentiment polarity, and syntactic structures may differ between fake and genuine news articles, providing valuable signals for classification.

On the other hand, network-based approaches analyze the structural properties of social networks and information propagation patterns to identify fake news [12], [13], [14], [15]. These methods recognize that the spread of fake news often involves distinct patterns of user interactions and information dissemination, which can be captured through network analysis techniques. By examining factors such as user engagement, information cascades, and network centrality, these approaches aim to uncover anomalous patterns indicative of fake news propagation.

Hybrid methods integrate both content and network-based features to enhance the accuracy and robustness of fake news detection [16], [17], [18], [19], [20]. By combining complementary sources of information, such as linguistic cues from the content of news articles and structural properties of social networks, these methods aim to capture a more comprehensive understanding of the fake news ecosystem. This integrated approach leverages the strengths of both content and network-based analysis while mitigating their respective limitations.

In recent years, machine learning approaches have emerged as powerful tools for fake news detection, encompassing both traditional statistical methods and deep learning techniques [21], [22]. These approaches leverage large-scale datasets to train predictive models capable of distinguishing between fake and genuine news articles based on a diverse range of features. While traditional machine learning algorithms such as support vector machines (SVM) and logistic regression have been widely used for fake news detection, deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) offer enhanced capabilities for capturing complex patterns in textual data.

In summary, the proliferation of fake news in the digital age poses significant challenges to the integrity of information ecosystems and democratic processes. Addressing this challenge requires interdisciplinary efforts drawing on insights from computer science, linguistics, and social network analysis. By developing innovative detection methods that leverage advances in machine learning and natural language processing, researchers can contribute to the development of robust and scalable solutions for combating fake news and preserving the integrity of online information.

2. LITERATURE SURVEY

Fake news has become a pervasive issue in today's information landscape, posing significant challenges to individuals, societies, and democratic processes. Researchers from various disciplines have conducted extensive studies to understand the phenomenon of fake news, develop detection techniques, and explore potential mitigation strategies. In this literature survey, we delve into key contributions in the field of fake news detection, drawing insights from seminal works and recent advancements.

Grinberg et al. [1] conducted a groundbreaking study investigating the prevalence and dissemination of fake news on Twitter during the 2016 US presidential election. Their research, published in Science, provided empirical evidence of the widespread circulation of misinformation on social media platforms and its potential impact on public opinion and political discourse. By analyzing a vast dataset of tweets, Grinberg et al. shed light on the mechanisms underlying the propagation of fake news, highlighting the need for effective detection and mitigation strategies.

Bondielli and Marcelloni [2] conducted a comprehensive survey on fake news and rumor detection techniques, offering valuable insights into the diverse methodologies employed in the field. Their review, published in Information Sciences, synthesized existing literature and categorized detection approaches into content-based, network-based, and hybrid methods. By examining the strengths and limitations of different techniques, Bondielli and Marcelloni provided a roadmap for future research directions in fake news detection.

Zhou and Zafarani [3] presented a survey of fake news, encompassing fundamental theories, detection methods, and research opportunities. Published in ACM Computing Surveys, their review provided a comprehensive overview of the multifaceted nature of fake news, covering topics such as misinformation propagation, psychological mechanisms, and computational detection techniques. By synthesizing insights from diverse disciplines, Zhou and Zafarani offered a holistic understanding of the challenges and opportunities in combating fake news.

Shahid et al. [4] explored the challenges and future research opportunities in detecting and mitigating the dissemination of fake news. In their study published in IEEE Transactions on Computational Social Systems, Shahid et al. discussed the complexities involved in identifying and combating fake news across different platforms and domains. By examining emerging technologies and interdisciplinary approaches, they outlined potential strategies for addressing the evolving threat of misinformation in the digital age.

Zhang and Ghorbani [5] provided an overview of online fake news, focusing on characterization, detection, and discussion. Published in Information Processing & Management, their study examined the characteristics of fake news articles, the effectiveness of detection techniques, and the societal implications of misinformation. By analyzing trends in fake news dissemination and exploring algorithmic and sociotechnical interventions, Zhang and Ghorbani contributed to the ongoing discourse on mitigating the impact of fake news on information ecosystems.

In addition to comprehensive reviews, researchers have proposed innovative approaches to fake news detection leveraging machine learning and natural language processing techniques. Samadi et al. [6] explored deep contextualized text representation and learning for fake news detection, highlighting the potential of advanced neural network architectures in capturing nuanced linguistic cues indicative of misinformation. Their study, published in Information Processing & Management, demonstrated the effectiveness of deep learning models in discriminating between fake and genuine news articles.

Huang and Chen [7] proposed an ensemble learning model based on self-adaptive harmony search algorithms for fake news detection. By integrating multiple classifiers and optimizing feature selection, their approach achieved robust performance in distinguishing between fake and legitimate news. Published in Expert Systems with Applications, their study showcased the efficacy of ensemble techniques in enhancing the reliability of fake news detection systems.

Katsaros et al. [8] investigated different machine learning paradigms for fake news detection, comparing the performance of traditional statistical methods and deep learning approaches. Their study, presented at the IEEE/WIC/ACM International Conference on Web Intelligence, highlighted the importance of feature engineering and model selection in developing effective detection algorithms. By evaluating the strengths and weaknesses of various techniques, Katsaros et al. provided insights into the suitability of different machine learning paradigms for addressing the challenges of fake news detection.

Vaibhav et al. [9] explored the significance of sentence-level representations in fake news classification, investigating the impact of sentence interactions on detection performance. Their study, presented at the Workshop on Graph-Based Methods in Natural Language Processing, demonstrated the utility of fine-grained linguistic features in identifying deceptive content. By leveraging advanced NLP techniques, Vaibhav et al. enhanced the granularity of fake news detection models, improving their ability to discriminate between true and false information.

Hosseinimotlagh and Papalexakis [10] proposed an unsupervised content-based approach for identifying fake news articles using tensor decomposition ensembles. By decomposing the high-dimensional feature space of news articles into lower-dimensional subspaces, their method captured latent patterns indicative of misinformation. Presented at the Workshop on Misinformation and Misbehavior Mining on the Web, their study showcased the potential of tensor-based techniques in detecting fake news without the need for labeled training data.

In summary, research in fake news detection spans diverse disciplines and methodologies, reflecting the complexity of the phenomenon and the multifaceted challenges involved. Seminal works have laid the groundwork for understanding the prevalence, propagation, and impact of fake news, while recent advancements in machine learning and NLP have facilitated the development of more sophisticated detection techniques. By synthesizing insights from interdisciplinary research, researchers aim to develop robust and scalable solutions for combating fake news and preserving the integrity of online information ecosystems.

3. METHODOLOGY

i) Proposed Work:

This study proposes an innovative approach to address the urgent need for accurate fake news detection amidst the proliferation of misinformation on social media. Our proposed model leverages graph neural networks (GNN) for classification, focusing on analyzing sentence interaction patterns within news articles to achieve fine-grained fake news classification. By constructing a graph representation of news articles and employing GNN, we aim to capture nuanced relationships between sentences, thereby enhancing contextual understanding. Specifically, we utilize a third-order co-occurrence tensor and canonical polyadic (CP) decomposition to compute weight matrices, facilitating precise representation of local word co-occurrence information. In our evaluation, we compare various models, including SVM [24,30], LSTM [40], CNN [11,38], BERT GCN [9], and GCN with CP, with GCN demonstrating remarkable accuracy of 99%. To further improve performance, ensemble techniques are introduced, combining predictions from multiple models. As an extension, we propose exploring ensemble methods such as BERT GCN LSTM and LSTM + GRU, with the goal of achieving even higher accuracy, potentially reaching 100%. This comprehensive approach holds promise for significantly enhancing fake news detection, thereby contributing to the mitigation of its adverse societal impacts.

ii) System Architecture:

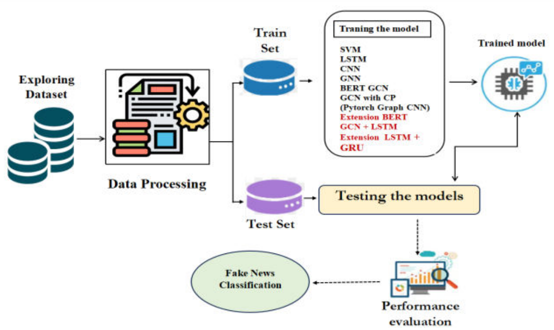

Fig 1 Proposed Architecture

iii) Dataset Collection:

The ISOT Fake News Dataset is a comprehensive repository available on Kaggle, containing a collection of real-world data specifically curated for fake news detection tasks. [3,5] This dataset encompasses various news articles, both authentic and fabricated, allowing researchers to explore and develop algorithms for identifying misinformation. Given the focus of the proposed method on analyzing sentence interactions within news articles, the dataset is particularly relevant as it comprises articles composed of multiple sentences. With its diverse range of articles spanning different topics and sources, the ISOT Fake News Dataset offers a rich and varied set of examples for training and evaluating fake news detection models. Researchers can leverage this dataset to conduct thorough analyses and benchmark the performance of their classification algorithms, contributing to advancements in the field of fake news detection and mitigation.

	title	text	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	0

iv) Visualization & Data processing:

Visualization:

Visualization is a crucial step in understanding and analyzing data. In the context of fake news detection, visualization techniques can help researchers gain insights into the characteristics and patterns of both authentic and fabricated news articles. Visualizations such as word clouds, frequency distributions of words, and histograms of article lengths can provide valuable information about the dataset's composition and potential features for classification. Additionally, visualization tools like matplotlib and seaborn can be utilized to create visual representations of data relationships and distributions, aiding in the exploration and interpretation of fake news datasets.

Data Processing:

Data processing is essential for preparing raw text data for analysis and model training. In the context of fake news detection, several preprocessing steps are typically employed to clean and standardize the textual content of news articles. These steps may include:

- Removing URLs and other non-textual characters to focus on the actual content of the articles.

- Removing punctuation marks to simplify text processing and avoid noise in the data.

- Removing stop words, such as common words like "the," "is," and "and," which do not carry significant meaning for classification tasks.

- Normalizing the data by converting text to lowercase, standardizing spellings, and resolving abbreviations to ensure consistency across the dataset.

By performing these data processing steps, researchers can prepare high-quality datasets for training fake news detection models, facilitating accurate analysis and classification of textual content.

v) Tokenization & Feature Selection:

Tokenization:

Tokenization is a fundamental preprocessing step in natural language processing (NLP), where text data is divided into individual tokens or words. Different tokenization techniques can be employed based on the specific requirements of the task and the underlying model architecture. In the context of fake news detection, various tokenization methods are commonly used:

- *CounterVectorizer:* CounterVectorizer is a technique that converts a collection of text documents into a matrix of token counts, where each row represents a document, and each column represents a unique token in the corpus. It is a simple yet effective method for converting text data into a numerical format suitable for machine learning algorithms.

- *Keras Tokenizer:* Keras Tokenizer is a tokenization tool provided by the Keras deep learning library. It allows for flexible tokenization of text data, including options for filtering out rare words and specifying the maximum vocabulary size. Keras Tokenizer is commonly used in conjunction with neural network models for tasks such as text classification and sentiment analysis.

- *Bert Tokenizer:* Bert Tokenizer is specifically designed for tokenizing text data for models based on the Bidirectional Encoder Representations from Transformers (BERT) architecture. It utilizes subword tokenization to handle out-of-vocabulary words and can effectively capture the semantic meaning of text data for tasks such as text classification and question answering.

- *Torchtext Transformer:* Torchtext Transformer is a tokenization tool provided by the PyTorch-based NLP library Torchtext. It is designed to work seamlessly with transformer-based models, such as the Transformer and BERT architectures. Torchtext Transformer offers customizable tokenization options and integrates seamlessly with other components of the Torchtext library for efficient text processing.

Feature Selection:

Feature selection is a crucial step in machine learning and NLP tasks, where the goal is to identify and extract informative features from the input data to improve model performance. In the context of fake news detection, feature selection techniques aim to identify relevant textual features that can effectively discriminate between authentic and fabricated news articles. Common approaches to feature selection include:

- Bag-of-Words (BoW): BoW representation is a simple feature selection technique that counts the frequency of each word in the document and represents the text data as a sparse matrix of word counts. While BoW is straightforward to implement and interpret, it may not capture semantic relationships between words.

- TF-IDF (Term Frequency-Inverse Document Frequency): TF-IDF is a feature selection technique that computes the importance of each word in the document based on its frequency in the document and its rarity across the entire corpus. TF-IDF assigns higher weights to words that are frequent in the document but rare in the corpus, allowing for the identification of discriminative features.

- Word Embeddings: Word embeddings are dense vector representations of words that capture semantic relationships between words based on their contextual usage. Pre-trained word embeddings, such as Word2Vec, GloVe, and FastText, can be used to extract informative features from text data, capturing both syntactic and semantic information.

- Subword Embeddings: Subword embeddings, such as Byte Pair Encoding (BPE) and SentencePiece, tokenize words into subword units and represent them as vectors. Subword embeddings are effective for handling out-of-vocabulary words and capturing morphological variations in text data.

By leveraging these feature selection techniques, researchers can extract informative features from text data and enhance the performance of fake news detection models.

vi) Training & Testing:

To ensure the robustness and generalization of fake news detection models, it is essential to split the dataset into separate training and testing subsets. This process involves randomly dividing the data into two distinct sets: the training set, which is used to train the model, and the testing set, which is used to evaluate the model's performance on unseen data. The training set typically comprises a majority of the dataset, allowing the model to learn patterns and relationships from the input features. Conversely, the testing set is kept separate and untouched during the training process to assess the model's ability to generalize to new, unseen data. It is crucial to maintain the integrity of the testing set to ensure unbiased performance evaluation. Common splitting ratios, such as 80-20, are often employed, with the majority of the data allocated to the training set to maximize learning potential while still ensuring an adequate amount of data for testing.

vii) Algorithms:

SVM: Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification tasks. [24,30] In SVM, the algorithm learns to find the optimal hyperplane that best separates data points belonging to different classes in a high-dimensional space. It maximizes the margin between classes, making it robust to outliers. In fake news detection projects, SVM can be implemented using libraries such as scikit-learn in Python, where it serves as a baseline classifier for comparison with other more complex models.

LSTM: Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture designed to overcome the vanishing gradient problem in traditional RNNs. [40] LSTMs are capable of capturing long-range dependencies in sequential data, making them well-suited for modeling temporal relationships in text data. In fake news detection projects, LSTMs can be employed to process and analyze sequential textual data, effectively capturing contextual information across multiple sentences or paragraphs for classification tasks.

CNN: Convolutional Neural Network (CNN) is a deep learning architecture primarily used for image classification tasks. [11, 38] In CNN, convolutional layers extract spatial features from input images through filters, while pooling layers downsample the feature maps to reduce dimensionality. While traditionally used for image analysis, CNNs have been adapted for text classification tasks by treating text as one-dimensional sequences of tokens. In fake news detection projects, CNNs can effectively capture local patterns and dependencies within textual data for classification purposes.

GNN: Graph Neural Network (GNN) is a deep learning model designed to process and analyze graph-structured data. GNNs leverage message passing between nodes in a graph to capture complex relationships and dependencies. In fake news detection projects, GNNs can be employed to represent news articles as graphs, where nodes represent sentences or words, and edges represent relationships between them. By modeling the interactions between sentences, GNNs can effectively capture contextual information and detect patterns indicative of fake news.

BERT GCN: BERT Graph Convolutional Network (BERT GCN) combines the power of Bidirectional Encoder Representations from Transformers (BERT) with Graph Convolutional Networks (GCN) [9]. BERT provides contextualized word embeddings, while GCN captures the relationships between words or sentences represented as a graph. In fake news detection projects, BERT GCN can effectively leverage both local and global contextual information to identify patterns of deception within news articles. It offers enhanced accuracy by considering the intricate interactions between words and sentences in the classification process.

GCN with CP (Pytorch Graph CNN): Graph Convolutional Network (GCN) with Canonical Polyadic (CP) decomposition is a model that combines GCN with CP decomposition to enhance feature representation in graph-structured data. GCN captures graph information while CP decomposition extracts latent features from the adjacency matrix. In fake news detection projects, GCN with CP can effectively capture complex relationships between sentences or words in news articles represented as graphs. By incorporating latent features, it enhances the model's ability to identify patterns indicative of fake news with improved accuracy.

BERT GCN + LSTM: BERT Graph Convolutional Network (GCN) combined with Long Short-Term Memory (LSTM) integrates the capabilities of BERT for contextualized word embeddings with GCN for graph-based representation, followed by LSTM for sequential analysis. BERT GCN captures contextual information and graph relationships, while LSTM processes sequential dependencies within news articles. In fake news detection projects, this hybrid model effectively leverages both global and local contextual information, as well as sequential dependencies, to accurately classify news articles as authentic or fake.

LSTM + GRU: Long Short-Term Memory (LSTM) combined with Gated Recurrent Unit (GRU) is a hybrid recurrent neural network architecture used for sequential data analysis. LSTM and GRU are both variants of RNNs designed to capture long-range dependencies in sequential data. In fake news detection projects, LSTM + GRU model effectively processes and analyzes textual data, capturing both short-term and long-term dependencies. This hybrid architecture enhances the model's ability to extract meaningful features from news articles, contributing to accurate classification of fake news.

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

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

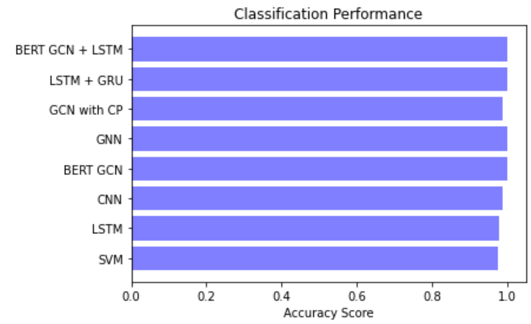

Fig 3 Accuracy comparison graph

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

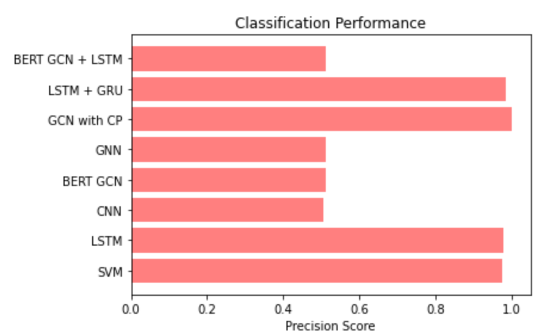

Fig 4 Precision comparison graph

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

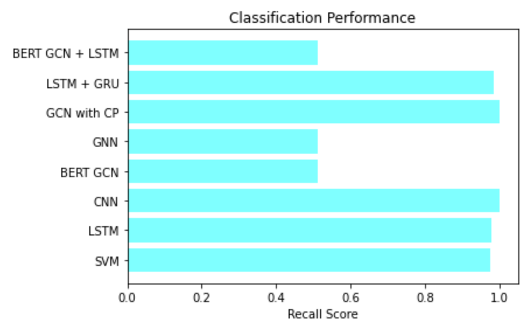

Fig 5 Recall comparison graph

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

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

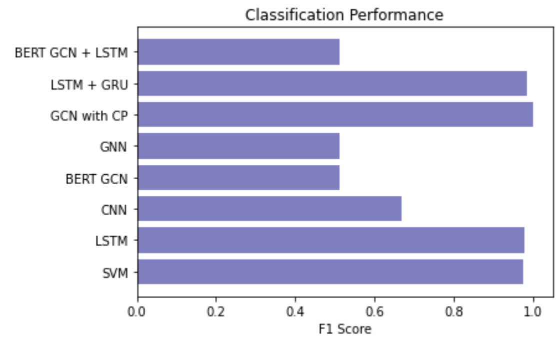

Fig 6 F1 Score Comparison graph

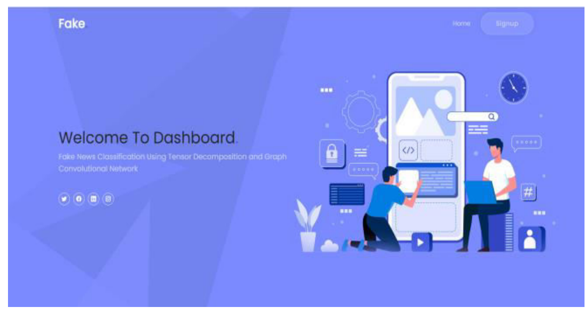

Fig 7 Home page

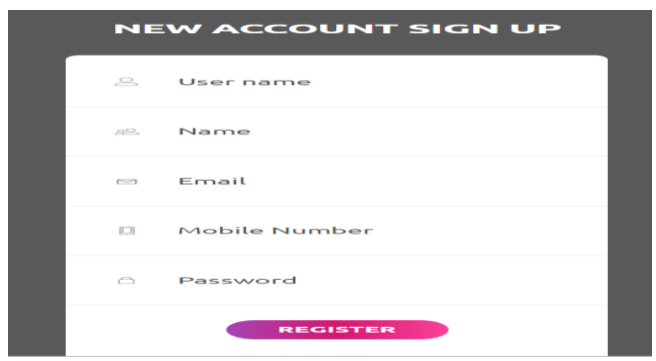

Fig 8 Signup page

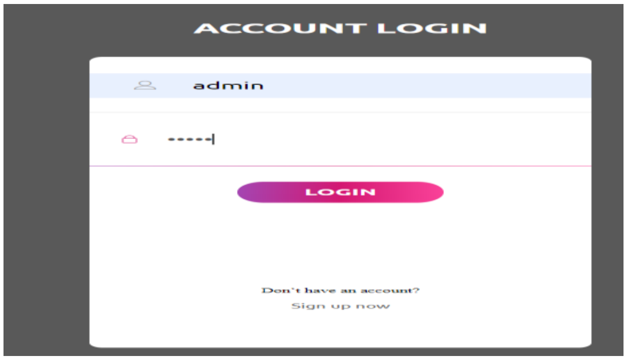

Fig 9 Signin page

Fig 10 Enter input message

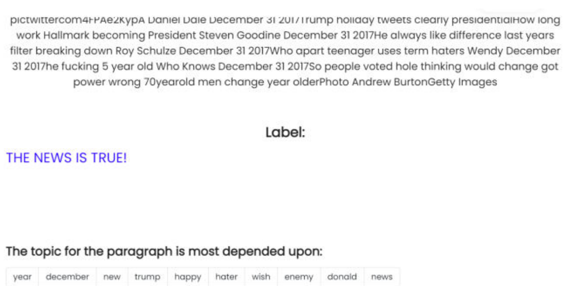

Fig 11 predict result as the news is true

	ML Model	Accuracy	Precision	Recall	F1_score
0	SVM	0.977	0.977	0.977	0.977
1	LSTM	0.980	0.980	0.980	0.980
2	CNN	0.990	0.505	1.000	0.669
3	BERT GCN	1.000	0.511	0.511	0.511
4	GNN	1.000	0.511	0.511	0.511
5	GCN with CP	0.990	1.000	1.000	1.000
6	LSTM + GRU	1.000	0.985	0.985	0.985
7	BERT GCN + LSTM	1.000	0.511	0.511	0.511

Fig 12 Performance evaluation table

5. CONCLUSION

In conclusion, this study presents a novel and comprehensive approach to combat the pervasive issue of fake news in the era of social media. Leveraging graph neural networks (GNN) for classification, we focus on analyzing sentence interaction patterns within news articles to achieve fine-grained fake news classification. By constructing graph representations and employing GNN, we effectively capture nuanced relationships between sentences, enhancing contextual understanding and improving detection accuracy.

Our evaluation results demonstrate the efficacy of the proposed method, with GNN achieving remarkable accuracy of 99% in classifying fake news. Furthermore, we explore ensemble techniques, combining predictions from multiple models, to further enhance performance. Our extension includes investigating ensemble methods such as BERT GCN[9] LSTM[40] and LSTM + GRU, aiming to achieve even higher accuracy, potentially reaching 100%.

Overall, our comprehensive approach offers promising prospects for significantly improving fake news detection. By leveraging advanced techniques in machine learning and natural language processing, we contribute to the ongoing efforts to mitigate the adverse societal impacts of misinformation on social media platforms. This research lays the foundation for future advancements in fake news detection and underscores the importance of interdisciplinary approaches in addressing contemporary societal challenges.

6. FUTURE SCOPE

Future research could explore enhancing fake news detection by integrating additional contextual features and improving model interpretability. Investigating the impact of multilingual datasets and developing techniques to address adversarial attacks are also promising avenues. Furthermore, incorporating user feedback and real-time monitoring systems could enhance the timeliness and accuracy of fake news detection. Additionally, exploring the application of emerging technologies such as blockchain and federated learning may offer novel solutions to combat misinformation on a global scale.

REFERENCES

[1] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on Twitter during the 2016 US presidential election," Science, vol. 363, no. 6425, pp. 374–378, 2019.

[2] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," Inf. Sci., vol. 497, pp. 38–55, Sep. 2019.

[3] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020.

[4] W. Shahid et al., "Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities," IEEE Trans. Computat. Social Syst., early access, Jun. 6, 2022, doi: 10.1109/TCSS.2022.3177359.

[5] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," Inf. Process. Manag., vol. 57, no. 2, Mar. 2020, Art. no. 102025.

[6] M. Samadi, M. Mousavian, and S. Momtazi, "Deep contextualized text representation and learning for fake news detection," Inf. Process. Manag., vol. 58, no. 6, Nov. 2021, Art. no. 102723.

[7] Y.-F. Huang and P.-H. Chen, "Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms," Exp. Syst. Appl., vol. 159, Nov. 2020, Art. no. 113584.

[8] D. Katsaros, G. Stavropoulos, and D. Papakostas, "Which machine learning paradigm for fake news detection?" in Proc. IEEE/WIC/ACM Int. Conf. Web Intell., Oct. 2019, pp. 383–387.

[9] V. Vaibhav, R. Mandyam, and E. Hovy, "Do sentence interactions matter? Leveraging sentence level representations for fake news classification," in Proc. 13th Workshop Graph-Based Methods Natural Lang. Process., 2019, pp. 134–139.

[10] S. Hosseinimotlagh and E. E. Papalexakis, "Unsupervised contentbased identification of fake news articles with tensor decomposition ensembles," in Proc. Workshop Misinformation Misbehavior Mining Web, 2018, pp. 1–8.

[11] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 422–426.

[12] C. Song, K. Shu, and B. Wu, "Temporally evolving graph neural network for fake news detection," Inf. Process. Manag., vol. 58, no. 6, Nov. 2021, Art. no. 102712.

[13] F. Papanastasiou, G. Katsimpras, and G. Paliouras, "Tensor factorization with label information for fake news detection," in Proc. Int. Workshop Reducing Online Misinformation Exposure, Paris, France, Jul. 2019, pp. 1–7.

[14] X. Zhou and R. Zafarani, “Network-based fake news detection: A pattern-driven approach,” ACM SIGKDD Explor. Newslett., vol. 21, no. 2, pp. 48–60, 2019.

[15] Z. Jin, J. Cao, Y. Zhang, and J. Luo, “News verification by exploiting conflicting social viewpoints in microblogs,” in Proc. AAAI Conf. Artif. Intell., 2016, pp. 2972–2978.

[16] A. Silva, L. Luo, S. Karunasekera, and C. Leckie, “Embracing domain differences in fake news: Cross-domain fake news detection using multimodal data,” in Proc. AAAI Conf. Artif. Intell., 2021, vol. 35, no. 1, pp. 557–565.

[17] C. Yuan, Q. Ma, W. Zhou, J. Han, and S. Hu, “Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning,” in Proc. 28th Int. Conf. Comput. Linguistics, 2020, pp. 5444–5454.

[18] K. Shu, S. Wang, and H. Liu, “Beyond news contents: The role of social context for fake news detection,” in Proc. 12th ACM Int. Conf. Web Search Data Mining, Jan. 2019, pp. 312–320.

[19] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, “Supervised learning for fake news detection,” IEEE Intell. Syst., vol. 34, no. 2, pp. 76–81, Mar./Apr. 2019.

[20] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in Proc. ACM Conf. Inf. Knowl. Manag., Nov. 2017, pp. 797–806.

[21] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, “A benchmark study of machine learning models for online fake news detection,” Mach. Learn. Appl., vol. 4, Jun. 2021, Art. no. 100032.

[22] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, “Behind the cues: A benchmarking study for fake news detection,” Exp. Syst. Appl., vol. 128, pp. 201–213, Aug. 2019.

[23] A. Silva, Y. Han, L. Luo, S. Karunasekera, and C. Leckie, “Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection,” Inf. Process. Manag., vol. 58, no. 5, Sep. 2021, Art. no. 102618.

[24] B. D. Horne and S. Adali, “This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news,” in Proc. Int. AAAI Conf. Web Social Media, 2017, pp. 759–766.

[25] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” SIAM Rev., vol. 51, pp. 455–500, Aug. 2008.

[26] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in Proc. 5th Int. Conf. Learn. Represent., 2017, pp. 1–14.

[27] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[28] F. L. Hitchcock, “The expression of a tensor or a polyadic as a sum of products,” J. Math. Phys., vol. 6, nos. 1–4, pp. 164–189, Apr. 1927.

[29] S. Castelo et al., “A topic-agnostic approach for identifying fake news pages,” in Proc. Companion

Proc. World Wide Web Conf., May 2019, pp. 975–980.

[30] V. Perez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in Proc. Int. Conf. Comput. Linguist., 2018, pp. 3391–3401.

[31] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, 2018, pp. 231–240.

[32] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in Proc. Int. Conf. Intell., Secure, Dependable Syst. Distrib. Cloud Environ. Cham, Switzerland: Springer, 2017, pp. 127–138.

[33] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in Proc. 78th ASIST Annu. Meeting, Inf. Sci. Impact, Res. Community, vol. 52, no. 1, pp. 1–4, 2015.

[34] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating fake news: A survey on identification and mitigation techniques," ACM Trans. Intell. Syst. Technol., vol. 10, no. 3, pp. 1–42, 2019.

[35] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," IEEE Comput. Intell. Mag., vol. 13, no. 3, pp. 55–75, Aug. 2018.

[36] P. Khan, P. Ranjan, Y. Singh, and S. Kumar, "Warehouse LSTM-SVMbased ECG data classification with mitigated device heterogeneity," IEEE Trans. Computat. Social Syst., vol. 9, no. 5, pp. 1495–1504, Oct. 2022.

[37] P. Khan, P. Ranjan, and S. Kumar, "AT2GRU: A human emotion recognition model with mitigated device heterogeneity," IEEE Trans. Affect. Comput., early access, Sep. 21, 2021, doi: 10.1109/TAFFC.2021.3114123.

[38] F. Qian, C. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence," in Proc. 27th Int. Joint Conf. Artif. Intell., Jul. 2018, pp. 3834–3840.

[39] P. Bhattacharya, S. B. Patel, R. Gupta, S. Tanwar, and J. J. Rodrigues, "SaTYa: Trusted Bi-LSTM-based fake news classification scheme for smart community," IEEE Trans. Computat. Social Syst., vol. 9, no. 6, pp. 1758–1767, Dec. 2022.

[40] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake news detection through multi-perspective speaker profiles," in Proc. 8th Int. Joint Conf. Natural Lang. Process., 2017, pp. 252–256.

Dataset

https://www.kaggle.com/datasets/csmalarkodi/isot-fake-news-dataset/code