IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 11, Issue 11, November 2024

DOI: 10.17148/IARJSET.2024.111114

Analysis of Deceptive News Recognition in Online Platforms

Prathap[1], Preetham N[2], Mahesh S[3], Sagar S[4], Manoj Naik[5]

Department of Artificial Intelligence and Machine Learning, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka[1-5]

Abstract: The challenge posed by misinformation is critical because it confuses public perception and undermines trust in the traditional news ecosystem, which has accuracy and truth as cores. To combat this rapid spreading fake news, we propose a tool that detects and classifies information posted on social media as false information. This system analyzes user-submitted text by cross-referencing it with verified data from trusted repositories. Based on this analysis, the model categorizes the content as either authentic or fabricated with clear labeling in its output. This solution not only enhances the detection of misleading content but also bolsters public trust and reduces the damaging effects of false information.

Keywords: False information detection, semantic analysis, sentence-level features, disinformation, text categorization

I. INTRODUCTION

In recent years, social media has gained immense popularity as a platform for accessing news due to its convenience, rapid dissemination, and affordability. However, it also serves as a catalyst for the swift propagation of "fake news," which refers to intentionally misleading or false information. This phenomenon has emerged as a significant societal concern, as fake news shared on social platforms often triggers widespread misinformation with far-reaching consequences. Consequently, detecting fake news on social media has become a critical area of research, drawing significant attention from academia, industry, and policymakers. Effective solutions are needed to address this growing challenge.

The prevalence of fake news on social platforms and messaging applications has risen dramatically. During emergencies, unverified information spreads at an alarming rate, causing confusion, panic, and public distrust. Individuals often accept forwarded content without verification, further amplifying the issue. Social media has evolved into a dominant channel for distributing false information, magnifying the adverse effects of misinformation. The unchecked spread of fake news undermines the social good, skews public perceptions, and manipulates audiences into accepting biased or entirely false narratives as truth. This widespread manipulation poses serious risks to society and the news ecosystem.

This false information is crafted intentionally to sow confusion and foster skepticism, leaving individuals increasingly unable to distinguish fact from fiction. The growing unreliability of social media feeds, blogs, and online news sites underscores the pressing need for computational tools to evaluate the credibility of online content. In this context, this paper discusses effective methods to combat the proliferation of fake content in digital news.

Initially, data from various online platforms, such as timesofindia.com and news.google.co.in, are integrated into the system's database for detecting fake news, with a focus on the political domain. If no direct matches are found in the database, the system performs a cross-referencing operation using a search engine to fetch the top five to seven results for validation. This step ensures higher reliability in the classification process.

Subsequently, machine learning models are developed through rigorous experimentation, utilizing advanced NLP techniques to preprocess and standardize text. These methods convert raw content into vector representations, allowing a quantitative comparison with existing vectors in the database. A similarity score is computed, based on which the content is classified as either ‘real’ or ‘fake.’ The system then displays the corresponding label to the user, providing a robust and reliable mechanism for identifying fake news.

II. LITERATURE SURVEY

This section reviews research on detecting fraudulent messages, highlighting methodologies, strengths, and limitations. Common approaches include machine learning, natural language processing (NLP), rule-based models, and hybrid

DOI: 10.17148/IARJSET.2024.111114

systems. Machine learning identifies patterns in large datasets, while NLP analyzes linguistic features for context. Rule-based models offer transparency, and hybrid methods combine strengths for better accuracy. Challenges include reliance on labeled data, high computational costs, and difficulty adapting to new fraud tactics. Techniques like feature engineering and anomaly detection are used but often struggle to keep up with evolving fraud strategies, pointing to the need for more adaptive and scalable solutions..

SR. No.	TITLE	PUBLISHER	DESCRIPTION	BENEFITS	LIMITATIONS
01	Design Study of False News: A Multidisciplinary Methodology for Content Sharing and Trust in Digital Media [1]	Jaigris Hodson, Royal Roads University; Brian Traynor, Mount Royal University (2018 IEEE International Professional Communication Conference)	Researchers propose a multidisciplinary approach to examining fake news, merging algorithmic techniques, psychometric data, and qualitative analyses of user interactions.	1) Offered creative strategies for fake news identification via algorithms and behavioral studies. 2) Highlighted ways to combine various tools and methodologies in a multidisciplinary framework to analyze user behavior and improve trust, news sharing, and the handling of false content.	The study depends on labeled datasets and domain experts to train systems for detection and classification. More efforts are needed to explore how humans form judgments about fake news.
02	False and Spam Content: Identifying Deceptive Messages During Disasters on Social Media [2]	Meet Rajdev and Kyumin Lee, Department of Computer Science, Utah State University (2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology)	A case study on the 2013 Moore Tornado and Hurricane Sandy, analyzing methods to identify false and spam content.	1) Conducted analysis of events like the 2013 Moore Tornado. 2) Early results revealed 96.43% accuracy with an F-measure of 0.961 for spam and fake message detection.	Overlooked prompt design to predict spam tweets during data streams with high volume.
03	Detecting Deceptive News Using Natural Language Processing via Attribution-Based Models [3]	Terry Traylor, U.S. Marine Corps, Fargo, ND; Jeremy Straub, Gurmeet, Nicholas Snell (Dept. of Computer Science, North Dakota State University), Fargo, ND (2019 IEEE 13th International Conference on Semantic Computing)	Explored classifiers, semantics, and an evolving framework for news verification.	1) Proposed influence-mining techniques to detect false articles and advertising. 2) Developed an initial detection tool with promising results for distinguishing deceptive news.	Attribution-based classifiers struggled with news lacking direct quotes or containing misleading citations, limiting system performance.
04	Assessing Credibility of Textual Claims on the Internet [4]	Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum (Max Planck Institute for Informatics, Saarbrücken, Germany)	A method that evaluates claims using communication between article language and web source credibility. Experiments with Snopes and Wikipedia hoaxes demonstrated its effectiveness.	1) Aimed at automating claim verification by identifying sources in media, processed via supervised classifiers. 2) Focused on evaluating claims without assumptions about their structure or source.	Could not consider nuances like denial, speaker intent, or the broader context of claims.

DOI: 10.17148/IARJSET.2024.111114

05	Message Verification System for Social and Mobile Applications [5]	Ankur Gupta, Purnendu Prabhat, Rishi Gupta, Sumant Pangotra, Suave Bajaj (Model Institute of Engineering and Technology, Jammu, India) (2017 International Conference on Next Generation Computing and Information Systems)	A system enabling message authenticity checks for mobile applications.	1) Developed a hierarchical database using trusted data sources to verify messages. 2) Allowed users to share validated content with authenticity scores, reducing misinformation spread.	Depended on extensive data sharing and forwarding, which raised scalability challenges.

III. PERFORMANCE ANALYSIS

The study on user experience—particularly the way users interact with and identify fake news—has been explored in only a few studies; [2] presents a novel approach to analyzing user behavior, including factors such as trust, loyalty, visual design, and usability, to support the detection of fake news. Furthermore, this paper presents findings from an analysis carried out on news platforms that show how user behavioral patterns affect the accuracy of fake news identification.

Natural Language Processing is introduced as a novel approach in fake news detection in the paper [3]. In 2019, Terry Taylor et al. proposed NLP-based methodologies for finding fake news and provided a technical assessment that can be used by future researchers. The paper employs tools such as TextBlob and the SciPy Toolkit to create a specialized classifier capable of identifying misleading information. It also reports the performance of the classifier along with the precision of the system while giving this the ability to process large datasets.

In [4], message credibility is identified as a crucial issue. A 2016 study by Kashyap Popat et al. explores the task of assessing the credibility of information being disseminated. The work introduces techniques that apply a supervised classifier to assess the reliability of sources of news. The authors demonstrate the practical validity of their work with real-life experiments on Snopes.com and Wikipedia, demonstrating how automation can help in achieving more precise credibility assessments.

In [5], Rajdev et al. (2015) came up with a system aimed at detecting spam messaging during natural disasters by using techniques of classification and feature detection of the information shared on Twitter. The system utilizes both flat and hierarchical classification methods for information about legitimate content and fake content, while feature detection identifies messages based on distinct characteristics. It reached an accuracy rate of 96.43%, and hence, the proposed system is a good foundation that will benefit real-time spam detection during crises.

IV. METHODOLOGY

Various techniques have been proposed to verify the authenticity of messages. However, it has been observed that understanding the semantics of a message is pivotal to grasping its true context. The context provides insight into the intent behind the message, a crucial aspect in determining its truthfulness. While comparing sentences in their original text form to detect semantic differences can be challenging, we employ a method that identifies relationships between adjacent words, called word embedding. This technique converts textual data into numerical vectors, making comparison simpler and more efficient. Thus, our methodology utilizes word embedding techniques to effectively capture the semantics embedded in the text.

A. **Word Embeddings**

Word embeddings are a transformative technique in natural language processing (NLP) that bridge the gap between human language and machine understanding. By mapping words into a vector space in n-dimensional space, embeddings ensure that semantically similar words or those used in similar contexts are located closer to each other. Unlike traditional approaches like bag-of-words or term frequency-inverse document frequency (TF-IDF), which treat words as independent entities, embeddings model the relationships and contextual usage of words effectively. Based on Zellig Harris's "distributional hypothesis" [10], this approach leverages the idea that words appearing in similar contexts often share similar meanings, thus offering a deeper linguistic understanding.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 11, Issue 11, November 2024
DOI: 10.17148/IARJSET.2024.111114

B. **Word2Vec**

Word2Vec, developed by Tomas Mikolov et al. [7] at Google in 2013, has become a foundational model for generating word embeddings. It trains on large-scale datasets to capture contextual and semantic relationships between words. Word2Vec uses two primary architectures: Continuous Bag of Words (CBOW), which predicts a target word from surrounding context words, and Skip-gram, which predicts surrounding words from a target word. This approach creates dense, fixed-size vector representations of words, which are computationally efficient and capable of encoding complex linguistic patterns. Its ability to model analogies, such as "king - man + woman = queen," showcases its power in capturing relationships between concepts.

C. **Vector Representation and Similarity Measures**

Vectorized word representations enable quantitative comparisons of semantic similarities between words, phrases, or texts. Techniques like cosine similarity measure the angular closeness of vectors, with a higher cosine value indicating greater similarity. These measures play a pivotal role in applications like sentiment analysis, topic modeling, and fake news detection, where contextual nuances must be accurately assessed. The ability to distinguish subtle differences or similarities between word meanings enhances model precision.

D. **Integration with Fake News Detection**

Our fake news detection system leverages word embeddings to capture semantic nuances and contextual patterns in text. By transforming input text into vectors and comparing them to preprocessed datasets, the system identifies inconsistencies that often signal fabricated content. Embedding-based approaches allow the model to analyze subtle distinctions in word usage, tone, and context. These representations, combined with similarity measures, enable the detection of fake news with improved accuracy while maintaining computational efficiency, even when processing extensive text datasets. The system effectively filters misinformation by combining linguistic insights with scalable technology.

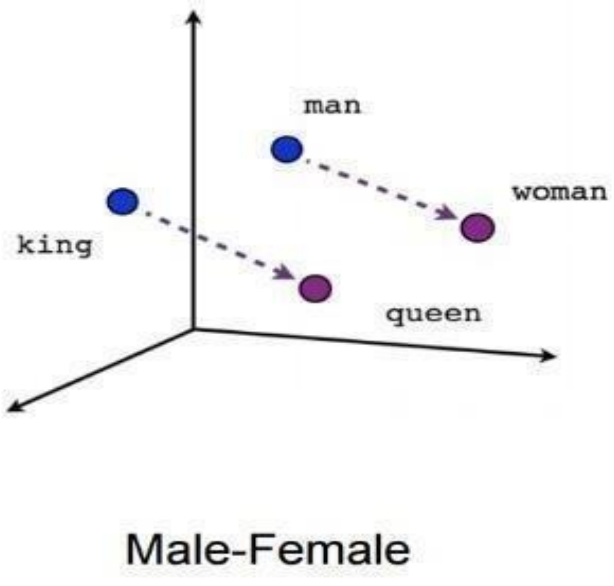


Fig 3. Word Embedding

Source: "Fake News Detection" by Moin Khan, Rishi Chouhan, Amisha Jain, and Sakeeb H. Sheikh.

An illustrative example would be: woman = man – king + queen. Word2Vec creates word embeddings using two primary methods: Skip-gram [8] and Continuous Bag of Words (CBOW).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

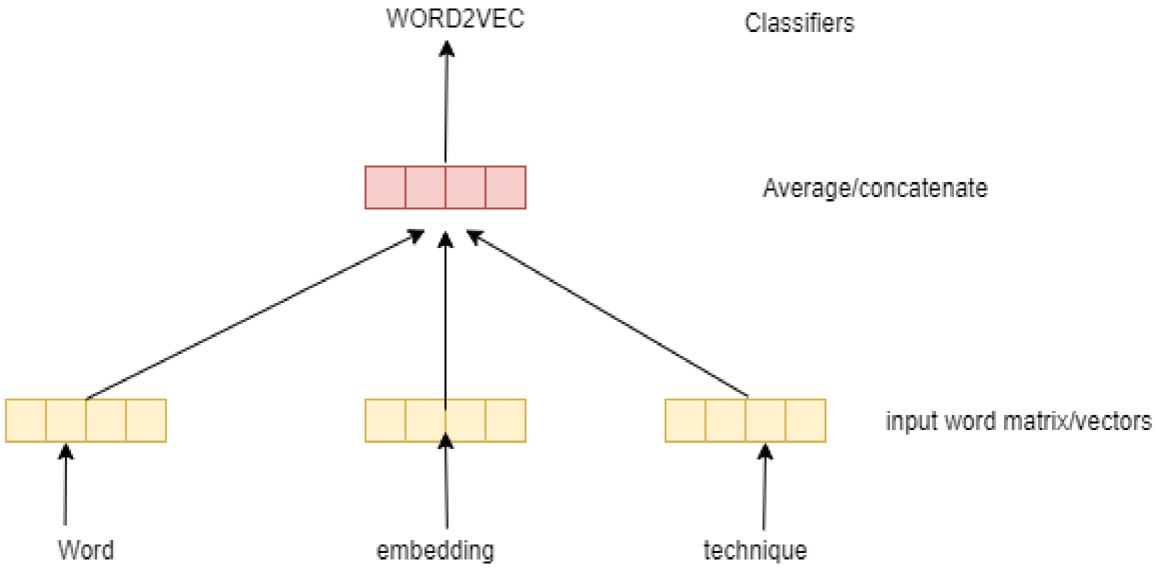
Impact Factor 8.066 ⵋ Peer-reviewed / Refereed journal ⵋ Vol. 11, Issue 11, November 2024

DOI: 10.17148/IARJSET.2024.111114

- **Continuous Bag of Words (CBOW)**

The CBOW creates word embeddings by predicting a target word from the surrounding context words within a specified sliding window. For example, for the sentence "Word2Vec is a word embedding technique," the model predicts the target word "Word2Vec" using context words like {word, embedding, technique}. The input word is represented with one-hot encoding initially; the neural network compares the predicted word with the actual target word and computes the error in the output. By repeated readjustments of its internal weights, the network learns a reasonable vector representation of the target word.

.



Source: "Fake News Detection" by Moin Khan, Rishi

Fig 4. CBOW Algorithm Sketch

This is in contrast to CBOW, which predicts the vector of the context words surrounding the current word and does not predict the current word by its context. Here, the idea is to predict the word vectors for context words such as {word, embedding, technique}, where the input word {Word2Vec} is used as the reference.

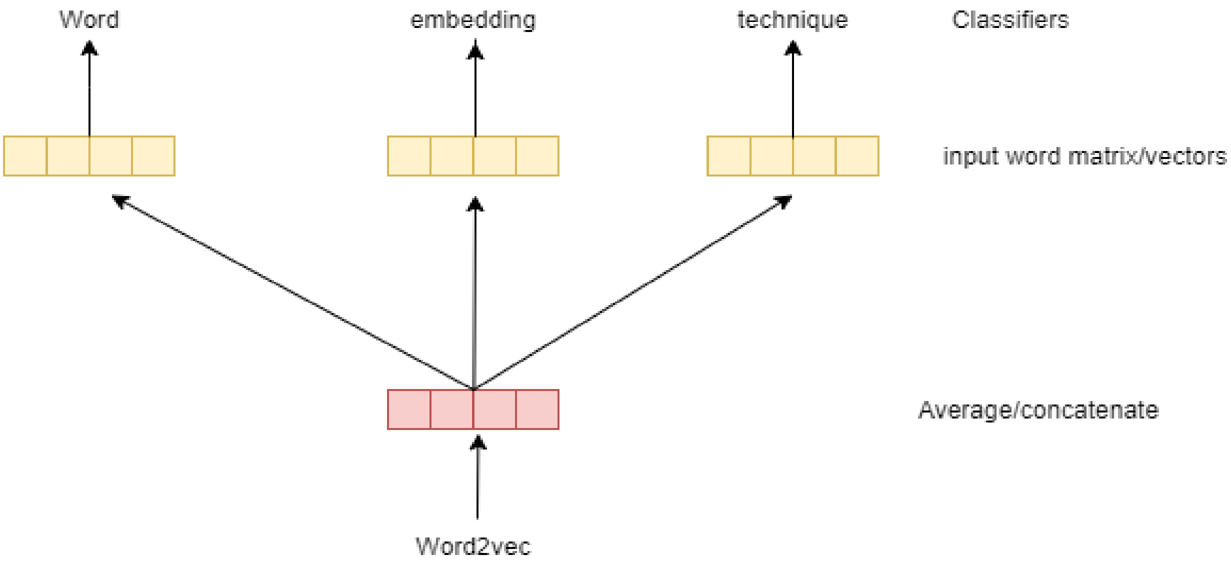


Fig 5.Skip Gram

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 11, Issue 11, November 2024

DOI: 10.17148/IARJSET.2024.111114

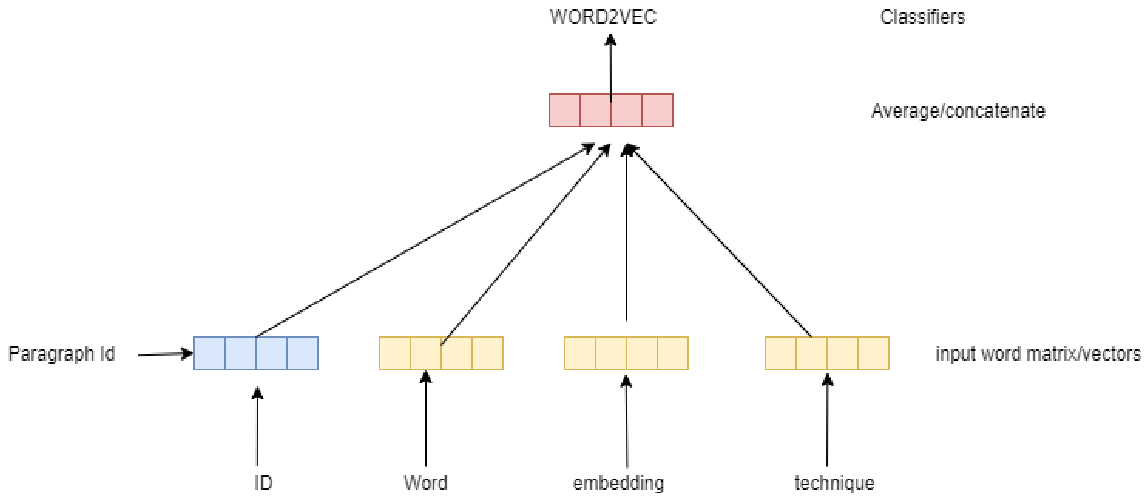
Source: "Fake News Identification on Social Media" by Moin Khan, Rishi Chouhan, Amisha Jain, Sakeeb H. Sheikh.

Each technique has distinct advantages and limitations. According to Mikolov, the Skip-gram model works particularly well with smaller datasets and is effective at capturing the meaning of rare words. In contrast, the CBOW model tends to be faster and produces better representations for words that appear frequently.

A. Doc2Vec

Doc2Vec [9] is an extension of Word2Vec, where it generates vector representations for larger text units like documents, paragraphs, or sentences. Representing documents numerically is more challenging than words due to their unstructured nature, requiring a different methodology. Mikolov and Le solved this by introducing an additional vector, the Paragraph ID, which is integrated into the Word2Vec model. Like Word2Vec, Doc2Vec offers two versions of its model:

Distributed Memory Version of Paragraph Vector, or PV- DM: This method is analogous to the CBOW model but introduces a document-specific vector, in addition to the words surrounding it, for the prediction. The document vector D would be learned together with the word vectors W during training, thus creating a numeric and overall document representation.



Source: "Fake News Detection" by Moin Khan, Rishi

Fig 6. .PV-DM

Distributed Bag of Words version of ParagraphVector (PV-DBOW)

While it is very similar to the Skip-gram model in Word2Vec, the Doc2Vec model is far more efficient than the Skip-gram model as it doesn't store individual word vectors, only requiring a smaller amount of computational memory; what it does instead is train on sets of documents and, during training, produce a word vector W for each word and a document vector D for each document.

DOI: 10.17148/IARJSET.2024.111114

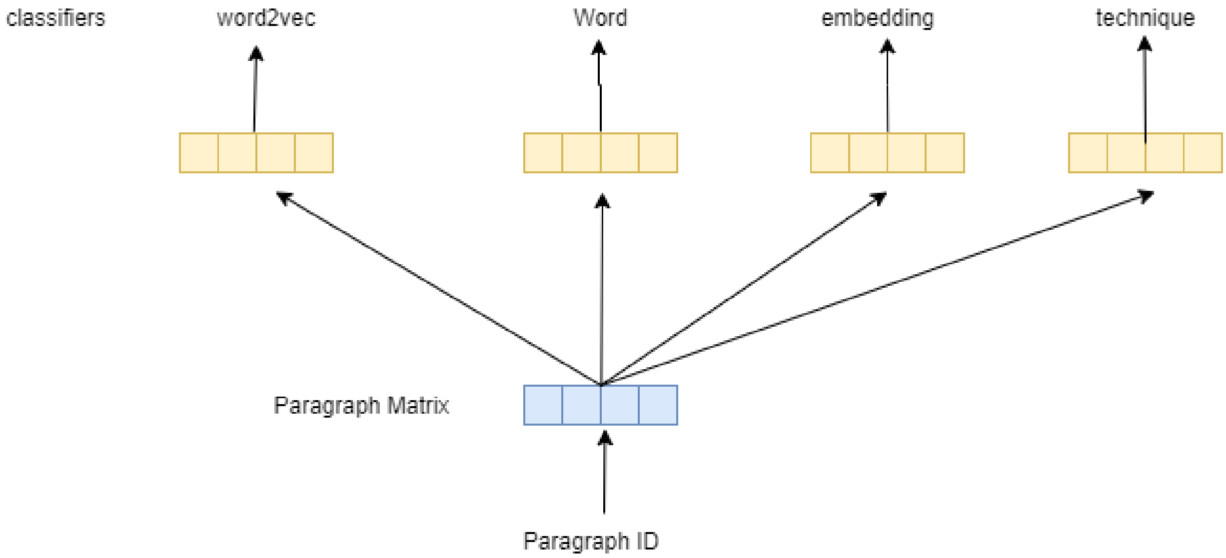


Fig 7. PVDBOW

Source: "Fake News Identification on Social Media" by Moin Khan, Rishi Chouhan, Amisha Jain, Sakeeb. H. Sheikh.

V. EXPECTED RESULTS

The literature survey on fake news detection systems yields thefollowing insights:

1. **Diverse Approaches**: Various methodologies, including natural language processing and machine learning classifiers, have been explored. For example, Traylor et al. emphasize in-article attributions for effective classification.

2. **User Engagement**: Studies highlight that user experience factors, such as trust and loyalty, significantly influence therecognition of fake news, as noted by Hodson and Traynor.

3. **Credibility Assessment**: Popat et al. demonstrate the effectiveness of using supervised classifiers to assess the credibility of news sources, validating claims through platforms like Snopes.com.

4. **Context-Specific Detection**: Rajdev et al. focus on misinformation during natural disasters, achieving 96.43% accuracy, which underscores the importance of targeteddetection strategies.

5. **Advancements in Word Embeddings**: The use of Word2Vec and Doc2Vec models facilitates semantic analysis of messages, enhancing authenticity verification through numerical representation.

6. **Framework for Future Research**: Identified gaps lead to a proposed conceptual framework for fake news detection, addressing previous shortcomings and emphasizing semantic understanding.

7. **Application Development Implications**: Insights from the survey inform the design of a proposed fake news identification application, integrating machine learning models to enhance user awareness.

VI. CONCLUSION

We reviewed existing fake news detection systems and identified the gaps in the contemporary approaches. This insight is what informed and built into the revised system architecture that we explain below. We propose a new framework designed to address such gaps, focusing on detecting fake news for the public and evaluating existing methods to identify fake political and other types of false news on social media.

This then relates to our suggestions of a fake news detection application and explores more possibilities for practical applications.

 This work is licensed under a Creative Commons Attribution 4.0 International License

International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⋇ Peer-reviewed / Refereed journal ⋇ Vol. 11, Issue 11, November 2024
DOI: 10.17148/IARJSET.2024.111114

VII. FUTURE SCOPE

The future of fake news detection systems presents several promising avenues for research and development:

1. **Enhanced Machine Learning Techniques:** Further exploration of advanced machine learning models, including deep learning and ensemble methods, can improve the accuracy and efficiency of fake news detection.
2. **Integration of Multi-Modal Data**: Incorporating multi- modal data, such as images, videos, and audio, alongside text analysis can provide a more holistic understanding of content authenticity and combat misinformation.
3. **Real-Time Detection Systems:** Developing real-time detection systems that can analyze and classify news content as it is shared on social media platforms will enhance the ability to curb the spread of fake news.
4. **User-Centric Design:** Focusing on user experience and engagement strategies, such as personalized feedback and educational tools, can empower users to better identify and question the authenticity of the information they encounter.
5. **Cross-Lingual and Cross-Cultural Studies:** Investigating fake news detection across different languages and cultural contexts can reveal unique challenges and strategies, leading to more universally applicable solutions.
6. **Collaboration with Fact-Checking Organizations:** Establishing partnerships with fact-checking organizationscan improve the reliability of detection systems and provide users with verified information.
7. **Policy and Ethical Considerations:** Future research should also address the ethical implications of automated fake news detection, including privacy concerns and the impact of algorithmic bias on information dissemination.

REFERENCES

[1]. Hodson, J., & Traynor, B. (2018). Design Exploration of Fake News: A Transdisciplinary Methodological Approach to Understanding Content Sharing and Trust on Social Media. *2018 IEEE International Professional Communication Conference*.
[2]. Traylor, T., Straub, J., Gurmeet, & Snell, N. (2019). Classifying Fake News Articles Using Natural Language Processing to Identify In- Article Attribution as a Supervised Learning Estimator. *IEEE International Conference on Semantic Computing (ICSC)*.
[3]. Popat, K., Mukherjee, S., Strotgen, J., & Weikum, G. (2016). Credibility Assessment of Textual Claims on the Web. *ACM*.
[4]. Rajdev, M., & Lee, K. (2015). Fake and Spam Messages: Detecting Misinformation during Natural Disasters on Social Media.
[5]. Zhang, Y., Rahman, M. M., Braylan, A., Dang, B., Chang, H.-L., Kim, H., McNamara, Q., Angert, A., Banner, E., Khetan, V., McDonnell, T., Nguyen, A., Xu, D., Wallace, B., & Lease, M. (2016). Neural Information Retrieval: A Literature Review.
[6]. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *Proceedings of the Workshop at ICLR*.
[7]. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Proceedings of the Neural Information Processing Systems (NIPS)*.
[8]. Le, Q. V., & Mikolov, T. (2014). Distributed Representations of Sentences and Documents. *Proceedings of the Neural Information Processing Systems (NIPS)*.
[9]. Harris, Z. (1954). Distributional Structure. *Word*.
[10]. Mikolov, T. (2012). Statistical Language Models Based on Neural Networks. PhD thesis, Brno University of Technology.
[11]. Mikolov, T., Le, Q. V., & Sutskever, I. (2013). Exploiting Similarities Among Languages for Machine Translation. *CoRR*, abs/1309.4168.
[12]. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Phrases and their Compositionality. *Advances in Neural Information Processing Systems*.
[13]. Pennington, J., Socher, R., & Manning, C. D. (2014). *GloVe: Global Vectors for Word Representation*. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543.
[14]. Zhang, Z., Robinson, D., & Tepper, J. (2019). *Detecting Fake News for Reducing Misinformation Risks Using a Hierarchical Attention Model with Semantic Textual Similarity*. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD).
[15]. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).