World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Check for updates

A Systematic Review of Text Mining Techniques for Fake News Detection

Faisal Abdullah Althobaiti *

Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia.

World Journal of Advanced Engineering Technology and Sciences, 2025, 17(01), 478-489

Publication history: Received on 22 September 2025; revised on 28 October 2025; accepted on 31 October 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.17.1.1455

Abstract

In today's digital environment, fake news spreads rapidly, undermining trust and confidence. This study investigates using text-mining techniques to detect and combat fake news. Through a detailed analysis of fifteen journals, we examine the effectiveness of textual analysis in distinguishing truth from fakery. By examining mechanisms and outcomes, we aim to inform the development of more reliable detection methods, empower stakeholders to fight misinformation in the digital age, and support accurate information. This study examines text mining for fake news detection. The aim is to use text mining to detect fake news to prevent its spread and to detect the source of fake news to curb its veracity. The study adopts literature review techniques, in which fifteen relevant journals were reviewed. This helps to collect the needed data to analyze fake news detection. Many of the reviewed journals used real and fake news datasets, whereas in the same articles, some percentages are labeled as fake and the rest as real. Also, some studies use 80% of their dataset for training, while the remaining 20% is for testing. Also, some journals used 49.9% of their dataset for training and the remaining 50.1% for testing. Different authors deployed many variables; however, the central aim was to define fake news models' accuracy, precision, and F1 scores. Many of the models are highly accurate to the tune of greater than 80% for fake news detection. Thus, using text mining analytics outperforms the traditional use of journalists, linguists, or media experts to evaluate information credibility. It was found that the strength of text mining analytics lies in its scalability, multimodal analysis, real-time detection, and interpretability. Also, some identified inherent limitations, including semantic complexity, data quality, feature selection, and algorithm biases, were found, and they are the reasons this study calls for multimodal fusion and the development of interpretable models for transparent explanation in future studies.

Keywords: Fake News Detection; Text Mining; Information Credibility; Misinformation; Digital Media

1. Introduction

In the past two decades, the influence of the Internet has transformed the world into a single global village by facilitating real-time access to events through the World Wide Web (WWW) [1]. Concurrently, the rapid increase in the use of social media platforms has assisted with information dissemination and content sharing with friends and followers, which has reshaped how people interact with news and information online [2]. However, this increased accessibility and availability have ushered in concerning trends in the proliferation of fake news, designed to deceive individuals into accepting false narratives [3].

Fake news originates from people and automated bots manipulating information for immoral purposes, such as political agendas and monetary gains, by receiving online traffic or followers [3]. Shockingly, over 70% of the global population has been impacted by the spread of fake news [4]. For example, in the United States of America (USA) presidential elections in 2016, fake news or verified news attracted greater engagement from social media users than published news from conventional news outlets [2].

* Corresponding author: Faisal Abdullah Althobaiti

The consequences of fake news extend far beyond mere misinformation. It poses a threat to democracy, society, economies, and the overall health and well-being of an individual. For instance, during the COVID-19 pandemic, several fake news articles circulated on social media, and one of the major ones was information about unconventional treatments [6]. Fake news mostly does not cause harm immediately; however, it could have a greater negative influence in the long run. For example, the speculation of former President Donald Trump on the efficacy of artificial ultraviolet-C light or drugs containing chloroquine as a cure for COVID-19 was entertained, as many shared the fake news for fun until an individual ingested a cleaning agent containing chloroquine and died [6].

The main cause of the propagation of fake news is the increasing amount of time individuals spend online, especially on social media platforms, which gradually eradicates the traditional media as the primary news source [7]. The shift in consumer behavior could be attributed to several factors, such as the immediacy and accessibility of the news on social media and the ease of interacting and sharing content with other people [8]. For instance, social media platforms were the source of news for 62% of American adults in 2016, which is a 13% increase compared to 2012 [8, 9]. Despite the quality of news produced by traditional outlets, the rapid, inexpensive production and dissemination of news on social media platforms have promoted the spread of fake news [10].

The detection of fake news has become a problem, and large social media giants like Facebook and X are putting measures in place to mitigate the propagation of misinformation across their platforms [11]. For instance, more than 25% of the 30 million tweets from then Twitter (now X) that contained links to news sources during the pre-election periods in the US were highly skewed, biased, and untrue [5]. The scope of fake news goes beyond significant global political events. For example, it was reported by an individual that a golden asteroid on target would hit the earth with a cost of $10 quadrillion worth of precious metals, which was fake news in an attempt to increase the value of Bitcoin [12]. Aside from that, fake news has also infiltrated different facets of public information, making many apprehensive.

Furthermore, according to a survey by the Pew Research Center, misinformation and fake news significantly affect citizens' confidence in their government. For instance, fake news has influenced 68% of Americans' confidence in the government, while 54% of Americans' confidence in each other has also been impacted [14]. 51% of Americans' confidence in their political leaders to get the work done and lead well has been affected by fake news [13]. Some residents in Macedonia used Google AdSense to spread fake news and run politically manipulated pages on Facebook and other websites to make a living [14]. On this note, it is safe to say that fake news is becoming a global pandemic because of the vulnerability of the massive readers and widespread malicious influence. It is, therefore, imperative to study how to use text mining to detect fake news to prevent the spread and detect the source of fake news before it goes viral.

With the growing concern of spreading fake news, exploring effective strategies to identify fake news becomes essential. In this regard, text mining emerges as a promising method. Therefore, this study aims to conduct a systematic review on performing text mining, also known as text analytics or natural language processing (NLP), to detect fake news and mitigate its harmful impacts before it spreads virally [15]. The motivation for this study is that several journals talk about the utilization of text-mining concepts for fake news detection. However, none of the journals provide an overall summary and comparison of what was introduced and what was not in the literature.

The rest of the report consists of the related works conducted on text mining for fake news detection, a discussion of the findings, and a conclusion.

2. Related works

This study selected 15 journals to conduct the systematic review. The inclusion and exclusion criteria used for this study can be seen in Table 1.

Table 1 Selection Criteria

Inclusion Criteria	Exclusion Criteria
Journals are in English.	Journals are not in English.
Journals that are published on or after 2017.	Journals that were published before 2017.
Journals that have open access.	Journals that have no free access.
Journals that have the keywords specified in Table 2.	Journals that do not contain the keywords specified in Table 2.

Additionally, the keywords used and the responses received from the search databases can be seen in Table 2.

Table 2 keywords and database search results

No	Keywords	Database				
		Google Scholar	**Elsevier**	**IEEE**	**DOAJ**	**Scope**
1	Fake news detection	35	21	12	15	17
2	Fake news detection by text mining	75	43	68	78	92
3	Fake news detection by Natural Language Processing (NLP)	57	25	12	37	31
4	Fake news detection by machine learning algorithms	89	53	91	68	73
5	Fake news detection by sentiment analysis	41	25	21	13	15

[16] focused on fake news detection in online articles using semantic features and various machine-learning algorithms. The best performance from the model has 95.66% accuracy, with bigram features performing the best with the random forest classifier, while bigrams outperform unigrams, trigrams, and quadgrams, which shows that word pairs are way more effective than singles. While longer phrases or sentences are an indication of the authenticity of the news. The performance is promising, as bigrams and random forests achieved an accuracy of 95.66%. This implies that semantic features are useful for detecting fake news. As a next step, semantic features may be combined with other linguistic cues and metadata to improve the detection performance.

[17] focused on improving the state-of-the-art techniques to identify fake news on social media by using stylometric (linguistic) features and word vector representations of the textual content. The feature set is exhaustive for the stylometric detection of fake news. The ensemble methods, such as random forest, stochastic gradient descent, and extra trees classifier, perform excellently with stylometric features. Also, TF-IDF vectors and skip-gram Word2Vec features perform well with lower runtime. Logistic regression performs well for every type of word vector feature. Thus, boosting in conjunction with CBOW (Word2Vec) and stylometric features is 95.49% accurate. The use of online training alongside real-time data, considering digital images alongside textual features, and utilizing novel datasets to cover more domains for enhanced organization are recommended. Furthermore, exploration of the use of pre-trained models for enhanced performance can also be considered.

[18] stated that the most related task to fake news detection is how rumors go viral and the veracity classification. Veracity depends on other subtasks, which need opinions that can be extracted from relevant posts. Posts are considered significant sensors for determining rumor veracity. Different forms of rumors include long-term ones like conspiracy theories and short-term ones. Since social media is a tool for the spread of fake news, with its attendant negative impact on users and the general society, fake news comes in phases, such as characterization and detection. In the former, the study introduced basic concepts and principles of fake news in traditional and social media. In the latter, the approach is examined from the perspective of data mining, in which data-oriented, feature-oriented, model-oriented, and application-oriented fake news must be diffused to reduce the negative impact.

[19] proposed a model that has Tf-Idf on unigrams and bigrams with cosine similarity fed into dense neural networks and has 94.31% accuracy. BoW without unigrams and bigrams with cosine similarity fed into dense neural networks has 89.23% accuracy. Pre-trained embeddings (Word2Vec) fed into a dense neural network. In a nutshell, the model performs better than the existing models by 2.5%, even though it struggles with the 'disagree' stance. However, it performs well with 'unrelated,' 'agreed,' and 'discussed.' The strategy of computing Tf-Idf vectors on unigrams and bigrams was effective, as hand-crafted input, like cosine similarity between news articles and headlines, is valuable to the model input. A consistent and smooth learning process can be achieved through experimentation with different hyperparameters and deployment of regularization techniques like Dropout, L2 regularization, cross-validation, and early stopping. Thus, it is possible to build automated fake news detection platforms by performing similar analyses on Twitter and Facebook.

[20] stated that dual emotion signals statistically depend on how viral the news has gone. Also, fake news shows distinct emotional resonance and dissonance, unlike real news, with no clear signal between emotional resonance and dissonance. Additionally, dual emotion features perform better than the existing emotional features and can enhance the performance of different fake news detectors. However, the features work effectively in a simulated real-world case where temporal data splits. Dual emotion features expose the distinctive emotional signals for the detection of fake

news. Mining dual emotion can be the needed solution to the incompetence of using only the semantics for fake news. Also, there is a need to explore multi-modal information, such as emotion in visual content, for precise emotion capturing and the use of a more sophisticated model for dual emotion in future work.

[21] stated that a deep learning model is effective for addressing the complex challenges posed by fake news detection. Important methodologies adopted in the domain were identified, showing the significant role of deep learning models. This shows that the deep learning method has the potential for the detection of fake news using different features like linguistic patterns, social network structures, and content-based characteristics. There is a need to explore and develop deep learning models to detect fake news because of the developing nature of fake news veracity and the challenges of early detection. There is a need to integrate multi-modal information and exploration of new features for the enhancement of accurate and comprehensive detection.

[22] presented that support vector machines and random forests perform better than classification algorithms in terms of accuracy and F1 score measures. About the feature set, the bag-of-words approach achieved the best results. It is common, when dealing with text, that matrices are generated. Such an approach ends up being large. In this sense, DC Distance showed a useful 370 algorithm to reduce dimensionality without losing too much performance. The study recommends that datasets needed to cover a wide range of platforms, languages, and types of fake news must be improved upon for the model's generalizability. Also, there is a need to explore more modalities, such as images and videos and contextual information, to strengthen fake news detection. Additionally, integration of the proposed model into a real-time fact-checking and content moderation system should be investigated to provide support to the efforts at mitigating the spread of misinformation.

[23] came up with a highly efficient prediction model to classify sentiment in COVID-19 pandemic fake news. Also, a comparison between the existing model and the proposed model shows that the latter has a high level of precision. Measures like the confusion matrix, classification rate, and true positives used in evaluating the model come with specific outcomes showing the high performance of the metrics. There is a need to choose an efficient sentiment analysis model to analyze opinions expressed on social media platforms and other microblogging platforms. Especially sentiment analysis of Twitter, now X, data.

[24] The results encompass a comparison of different models based on their accuracy, precision, and F1 score. For instance, LSTM has 93%, which is the highest accuracy model, while logistic regression has 87% accuracy for classifying real-time tweets. Part of the recommendations includes the use of n-gram models in combination with classifiers to enhance accuracy, as well as increase the size of the Twitter dataset, and for verification. Also, the recommendation includes the use of the n-gram model in future work to enhance datasets to improve their accuracy.

[25] includes that the proposed deep learning-based approach performs significantly better than the traditional machine learning techniques deployed to identify sarcasm and irony detection. The CNN model is highly accurate, by 80%, in classifying ironic and sarcastic texts. This shows that the model effectively identifies sarcastic and ironic content. Furthermore, the analysis of the learned CNN model's features and attention weight gives a detailed insight into the linguistic patterns and shows the pathways for the model to learn sarcasm and irony identification. Part of the recommendations includes further research that needs to be done to understand how to incorporate additional contextual information, like the user profile and conversation history, to enhance the detection model. performance in terms of sarcasm and irony. There is a suggestion to explore the applicability of the proposed approach in other languages and domains beyond social media textual content; this will help to evaluate the generalizability. Also, there is a need to examine the potential of a deep learning-based approach for other natural language processing tasks, like sentiment analysis and emotion recognition, where identification of sarcasm and irony could be critical for a thorough and precise understanding of textual content.

[26] has shown how effectively the proposed deep learning-based framework works in text style transfer. The model can successfully transform textual content so it can come with different styles while the semantic content remains intact. For instance, from positive to negative sentiment, or from formal to informal language. The qualitative evaluation done using automatic metrics and human evaluation shows that the generated textual content shows the needed style features while still maintaining a high semantic similarity to the original input. Another finding is that the model's ability to capture and transfer different stylistic features is enhanced. Part of the recommendations in this study is setting a path for future work on text style transfer. It is suggested that there is a need to further explore the use of more sophisticated language models, such as pretrained transformers in the encoder-decoder architectural relationship, which will further enhance the quality of the text generated. Also, there is a recommendation for further investigation of the method to control certain parts of the target style, like the level of formality and sentiment strength, to enhance

a more fine-grained style transfer. Also, there is a need to examine the potential of the application framework in areas like personalized text generation, text simplification, and creative writing assistance.

[27] stated that the fake news usually exhibits catchy title structures, which can help in rapid detection. Also, the use of proper nouns is crucial to differentiating rumors. Furthermore, fake news depends significantly on heuristics, such as catchy titles, which are a form of clickbait rather than strong logical arguments. The study, therefore, is insightful about the inherent challenges and achievements of developing rumor classification systems. It is recommended that media literacy education go a long way in helping users thoroughly evaluate information before they consume and share it. Also, continuing research in social media mining is recommended to detect and resolve rumors promptly before they escalate.

[28] stated that people are prone to believe and share as fast as possible real and fake news that panders to their existing beliefs and sentiments. Also, sensational headlines and emotion-induced language increase the spread and appeal of fake news. Social media algorithms create echo chambers where users are basically exposed to content aligning with their preconceived views. Furthermore, clickbait, shocking images, and persuasive language are parts of the driving forces of the spread of fake news, as people with lower levels of media education are highly prone to believing and sharing fake news. The authors recommend that there is a need to develop a series of interventions that will enhance media literacy, especially among people with low levels of exposure to media literacy. Also, it is recommended that there be a need to have a comprehensive understanding of the crucial role played by social media platforms in enhancing the veracity and spread of fake news. There is a need to disincentivize the psychological factors that make people susceptible to the consumption of fake news and vulnerable to the quick sharing of fake news on social media. Also, aside from the psychological factor, there is a need to understand the digital mechanisms and motivations for spreading fake news; this will help mitigate or combat the spread.

[29] stated that the new dataset, that is, LIAR, has a larger magnitude than the already existing resources and allows machine learning to assess the detection of fake news. Furthermore, the findings show that significant improvement can be attained for fine-grained fake news detection if metadata is integrated with text. It was also found that researchers can use the new dataset to train and test machine learning algorithms. The major recommendation is that LIAR has the potential to be deployed for stance classification, mining arguments, topic modeling, rumor detection, and political natural language processing research.

[30] specified that the most prominent and well-established datasets for the detection of fake news have been identified. The one for analyzing the strengths and weaknesses of the dataset has also been identified. The study recognizes the need for more diverse and representative datasets that address the landscape conducive to fake news. A series of recommendations is proposed by the authors, including the need to develop more diverse and representative datasets that will lead to the early detection of fake news. Also, there is a need to enhance the quality and dependability of the ground truth labels in the dataset to make sure there is accuracy in model training and evaluation. Exploration of the use of multi-modal data, such as text, images, and videos, for the enhancement of fake news detection. A need for collaboration between social media platforms and fact-checking organizations for ease of access and leverage of data for research on fake news. Datasets should be continuously updated and maintained to keep abreast of the ever-evolving fake news tactics and platforms.

The overall summary of the selected 15 journals for this study is provided in Table 3. The brief discussion of the results and recommendations of the selected journals from Table 3 is provided below.

Table 3 Summary of the selected journals

No	Objective of this study	The methodology of text mining utilized	Dataset used	Variables	Ref
1	Fake news detection in online articles using semantic features and various machine-learning algorithms.	Recurrent Neural Networks (RNN) with Naïve Bayes and random forest classifiers.	Real and fake news dataset from Kaggle.com containing 6256 articles, of which 50% are labeled as fake and 50% as real. Also, 80% of the dataset is for training, while the remaining 20% is for testing.	Material dataset, text preprocessing, semantic features including term frequency (TF), term frequency-inverse document frequency (TFIDF), bigrams, trigrams, quadgrams, and vectorized word representations.	[16]
2	To improve the state-of-the-art techniques to identify fake news on social media by using stylometric (linguistic) features and word vector representations of the textual content.	Not Specified.	Fakenews.net and the McIntire Dataset consist of 5405 news articles in the training set, which is 49.9% real and 50.1% fake. Also, 1352 news articles are in the test set.	Stylometric features, which include three different feature sets that are based on authorship attribution and the detection of deception. word vector features, including the count vectors, TF-IDF vectors, CBOW, and skip-gram.	[17]
3	To present an overview of fake news detection and discuss promising research directions	News content models and social context models	Buzzfeed News, LIAR, BS detector, CREDBANK	News content features and social context features	[18]
4	To propose a model for detecting fake news by accurately predicting the stance between headlines and the news articles.	Classifiers such as recurrent neural networks (RNN), Naïve Bayes, and Random Forest use various text representations like bag-of-words, TF-IDF, and pre-trained word embeddings (GloVe, Word2Vec).	Fake news challenges (FNC-1) from the Emergent Dataset created by Craig Silverman, containing 6256 articles, of which 50% are labeled as fake and 50% as real. Also, 80% of the dataset is for training, while the remaining 20% is for testing.	The stances of the news articles and their headlines were grouped into four categories: 'agreed,' 'disagreed,' 'discussed,' and 'unrelated.'	[19]
5	To verify if dual emotion, that is, publisher emotion and social emotion, is different between fake and real news, and to propose dual emotion features for the representation of dual emotion and the connection with regard to the detection of fake news.	Emotion classifiers, emotion lexicons, and neural network architectures like BiGRU and BERT.	RumourEval-19 in the English language, and Weibo-16 and Weibo-20 in Chinese.	Social emotion, publisher emotion, emotion gap, and the veracity of the news (fake or real).	[20]

6	To provide a detailed review of techniques of fake news detection, focusing on deep learning methods and exploring different approaches and methodologies for fake news detection.	Existing Literature.	Fake News, Twitter 15, and Liar.	Linguistic features, social network analysis, graph-based features, and content-based features. Covering commonly used datasets for training and evaluation of fake news detection models.	[21]
7	To propose a framework that can work across different platforms and languages where fake news can be detected.	No specific model is adapted.	Datasets containing news articles from different online sources in various languages, including English, Portuguese, and Spanish.	From the dataset, word embeddings and n-grams are extracted.	[22]
8	Developing a precise approach for sentiment analysis of fake news in relation to COVID-19. The study further aims at analyzing opinions expressed in fake news concerning COVID-19 with the aid of machine learning and deep learning algorithms.	Naive Bayesian, Adaboost, K-nearest neighbors, random forest, logistic regression, decision tree, neural networks, support vector machine, and deep learning models, including CNN, LSTM, RNN, and GRU.	Fake news related to COVID-19.	Content of news articles, sentiment labels, and other relevant features.	[23]
9	To develop a model whereby fake news is detected using machine learning techniques and deep learning algorithms.	Naive Bayes, a support vector machine (SVM), logistic regression, long-short-term memory (LSTM), and a natural network using Keras.	News articles labeled as fake and real.	Tweet features like text, retweet count, favorite count, and source length. User features like user ID, username, and user created at. User description, follower count, user status, user friend counts, user verification, and user profile.	[24]
10	Presenting a new deep learning-based approach to doing sentiment analysis of text to figure out sarcasm and irony. Also, the study intends to develop a model that accurately identifies sarcastic and ironic expressions within textual content, which cannot be figured out by traditional sentiment analysis techniques.	Deep learning-based approach, especially a convolutional neural network (CNN) architecture. Pre-trained word embedding, also known as GloVe, is adopted as a representation of input text that feeds into the CNN model, a trained model.	Several datasets, including the SARC dataset, the Internet Argument Corpus, and the Sarcasm in the Twitter dataset, are used to train and evaluate.	The variables include input text, which is represented as word embeddings and associated labels, including sarcastic, ironic, or non-sarcastic/non-ironic.	[25]
11	Develop a deep learning-based framework for text style transfer that encompasses transforming certain text from a different style while preserving	A deep learning framework consisting of an encoder-decoder architecture with an attention mechanism is deployed.	the Yelp dataset (for sentiment style transfer), the GYAFC dataset (for formality style transfer), and the Civil Comments dataset (for toxicity style	Input text, the target style, and the generated output text.	[26]

	the semantic content of the text. To also have a model that effectively performs the task through numerous applications in natural language and text manipulation.		transfer). The dataset has text samples that are labeled with different styles and are used for the training of the style transfer model.		
12	Addressing the challenges of rumors flying on social media platforms by providing insights into the way Rumors emerge, circulate, and can automatically be assessed for their veracity with the aid of natural language processing (NLP) and data mining algorithms.	NLP data mining algorithms.	Social media posts, tweets, and articles containing rumors.	Text contents, including the titles and body text. Metadata, including the source and user engagement. Linguistic patterns include the use of proper nouns and sentiments. Network features, including retweet patterns and user influence.	[27]
13	Understand the rationale for consuming fake news. Specifically, the study explores the reasons behind the proliferation of sharing fake news on social media, why some people obviously believe it, and the persuasive factors contributing to the veracity of fake news.	Existing Literature.	Fake News.	Title structure, language through the text, image chosen, individual characteristics, news characteristics, social media factors, and psychological factors.	[28]
14	Address the scarcity of labeled benchmark datasets for the detection of fake news and introduce a new benchmark for the automatic detection of fake news. The study also aims to provide a publicly available dataset that will enhance research on this subject matter.	No specific model was stated explicitly.	LIAR, POLITIFACT.com	From the LIAR dataset, the variables are statement content, context, and label. From the POLITIFACT dataset, variables such as source and speaker information were used.	[29]
15	Comprehensive survey and evaluation of already existing datasets for the detection of fake news. The aim is to analyze the characteristics, strengths, and limitations of these datasets so that researchers and other practitioners can be guided in choosing suitable datasets for their research and application of the detection of fake news.	Existing Literature.	Different types of fake news datasets that cover political, financial, and health-related misinformation were used. Also, news articles, social media posts, and fact-checking websites.	Text content, metadata, user information, and ground truth labels.	[30]

3. Discussion

Social media has diminished the influence traditionally held by mass media. The dynamic development of social media platforms, internet media platforms, microblogging, and instant messaging has mitigated the influence of radio, television, and print media.

Modern societies embrace technology with enthusiasm; new offerings by the Information and Communication Technology (ICT) sector are recognized. However, this transformation is accompanied by a new challenge: fake news. This phenomenon has the potential to erode societal trust and credibility, affecting companies, governments, and institutions alike. It stands as one of the negative consequences of the ICT revolution and warrants intervention.

Combatting fake news solely through government policies or state security measures has proven ineffective. Therefore, alternative methods, particularly those that facilitate early detection of fake news before it spreads widely, are imperative. Text mining analytics emerges as a crucial tool in this endeavor.

In the fight against fake news, machine learning methods such as text mining analytics have demonstrated significant success. They enable the development of accurate and efficient models for detecting misinformation in real time. A systematic review conducted in this study underscores the effectiveness of text mining analytics in processing vast amounts of textual data, particularly from social media content and news articles where fake news proliferates. Additionally, integrating text mining with video and image content analysis offers a more comprehensive understanding of how fake news propagates across various media platforms.

Moreover, text mining approaches are not language-biased, making them capable of detecting fake news across multiple languages. This is particularly crucial given the interconnected nature of the global digital landscape. Real-time application of text mining techniques aids in swiftly identifying and capturing fake news, mitigating its impact. The interpretability of text mining techniques allows users to understand the rationale behind classification decisions, fostering trust in automated fake news detection systems.

However, fake news propagators continually evolve their strategies, employing subtle linguistic cues and context-specific content that pose challenges to accurate detection by text mining algorithms, especially in cases of satirical or opinion-based content. Ensuring the availability of high-quality training data is paramount for effective data mining analytics, as biased or flawed datasets can compromise model performance. Biased training data may lead to skewed classifications, particularly concerning sensitive topics such as religion, race, or politics. Additionally, feature selection and representation techniques directly influence the model's ability to capture relevant information from text data.

Choosing the right machine learning algorithm tailored to the characteristics of fake news is essential, as different algorithms exhibit varying capabilities and weaknesses in handling textual data. Furthermore, the choice of evaluation metrics and strategies for handling imbalanced datasets is critical for accurately assessing the performance of fake news detection models.

Text mining is an iterative process, requiring experimentation, tweaking, and thoughtful preprocessing to yield positive outcomes. The journey to fake news detection through text mining underscores the critical role of algorithms and the significance of high-quality training data and preprocessing techniques. Future research should explore multi-modal fusion, combining text mining with images, videos, and metadata to gain a comprehensive understanding of fake news propagation patterns. Additionally, there is a need to develop interpretable models and techniques that provide transparent explanations of classification decisions, enhancing trust and accountability in fake news detection systems.

Moreover, while text mining analytics holds great promise in combating fake news, it is important to acknowledge the potential limitations and challenges associated with this approach. One such challenge often lies with the microlinguistic cues of fake news reporters who use it, namely, the difficulties in text mining algorithms that generate it. In addition to can, reliance on biased or flawed data sets poses another important obstacle. Biased training data can lead to skewed classifications, especially on sensitive topics such as religion, race, and politics, undermining effective research models. Furthermore, the adequacy of cross-sectional representation is needed for a more accurate classification, as insufficient representation may hinder the discriminative power of the model. Choosing an appropriate machine learning algorithm to suit the nature of fake news is also key, as different algorithms exhibit different strengths and weaknesses in processing textual information. Besides, analytical metrics and ways to consume data handling imbalances play an important role in accurately evaluating the false reporting performance of detection models. Finally, although text mining techniques in general. However, it can be explained that there may be some cases where there are no obvious classification decisions. Developing interpretable models and methods that provide clear explanations for classification decisions can enhance the reliability and accountability of false positive detection systems.

Due to these challenges, a nuanced approach is needed to apply text mining analytics in the fight against fake news. While the potential benefits are great, acknowledging and addressing these limitations is important for robust and reliable detection methods. Future research efforts should focus on issues reducing detection algorithms, improving the quality of data sets, overcoming these challenges, and improving false news detection algorithms.

4. Conclusion

In conclusion, this study has highlighted the challenges posed by fake news in the digital age and the potential of text mining analytics in solving this contemporary problem. Through a systematic search of the relevant literature, we investigated various methods and techniques for fake news detection, highlighting the effectiveness of machine learning techniques on large amounts of textual content from social media and news in the reporting of the processing.

Findings from selected journals highlight the importance of data mining research in early detection and mitigation of misinformation, thereby preserving social trust and credibility. When information analytical approaches offer promising methods for dealing with fake news, it is important to acknowledge and address the challenges of biased microlinguistic datasets, indicators, and unweighted use of data.

Going forward, future research efforts should focus on refining detection algorithms, improving data set quality, and addressing the emerging challenges of fake news detection using text mining analytics. In addition to this:

- Creating more diverse and representative datasets: To increase the generalizability of research paradigms, datasets spanning multiple platforms, languages, and pseudotypes are essential.
- Search for advanced machine learning algorithms: Considering the nuances of language and context, researchers should continue to search for and develop advanced machine learning algorithms that match pseudo-text characteristics.
- Integration of different disciplines: Future studies should explore the integration of mining images with image and video content analysis to gain a more comprehensive understanding of false information distribution processes in different media formats. • Developing interpretable models: Translational models are essential to provide clear explanations for classification decisions while promoting reliability and accountability in automated accreditation processes.
- Collaboration between stakeholders: Cooperation between social media platforms, fact-finding organizations, and researchers is essential to improve access to data and enhance investigations in fake news detection.

References

[1] H, Reddy., N, R., M, Gala and A, Bassava. 'Text-mining-based fake news detection using ensemble method" International Journal of Automation and Computing. Vol 17, No, 2. 210-221, 2020. Doi: 10.1007/s11633-019-1216-5.

[2] B. Chang, T. Xu, Q. Liu, E. H. Chen. Study on information diffusion analysis in social networks and its applications. International Journal of Automation and Computing, vol.15, no.4, pp.377–401, 2018. DOI: 10.1007/s11633- 018-1124-0.

[3] K. Shu, A. Sliva, S. H. Wang, J. L. Tang, H. Liu. Fake news detection on social media: A data mining perspective. SIGKDD Explorations Newsletter, vol.19, no.1, pp.22–36, 2017. DOI: 10.1145/3137597.3137600.

[4] Handley, Lucy. "Nearly 70 Percent of People Are Worried about Fake News as a 'Weapon,' Survey Says." CNBC, CNBC, 22 Jan. 2018, www.cnbc.com/2018/01/22/nearly-70-percent-ofpeople-are-worried-about-fake-news-as-a-weapon-survey-says.html.

[5] Bovet, H. A. Makse. Influence of fake news in Twitter during the 2016 US presidential election. Nature Communications, vol.10, no.1, Article number 7, 2019. DOI: 10.1038/s41467-018-07761-2.

[6] K, Englmeier,. "The Role of Text Mining in Mitigating the Threats from Fake News and Misinformation in Times of Corona." Procedia Computer Science, vol. 181, 2020, pp. 149-156, https://doi.org/10.1016/j.procs.2021.01.115. Accessed 29 Apr. 2024.

[7] R, Kozik, et al. "Technical Solution to Counter Potential Crime: Text Analysis to Detect Fake News and Disinformation." Journal of Computational Science, vol. 60, 2022, p. 101576, https://doi.org/10.1016/j.jocs.2022.101576. Accessed 1 May 2024.

[8] S, Afroz, M, Brennan, and R, Greenstadt. Detecting hoaxes, frauds, and deception in writing style online. In ISSP'12.

[9] H, Allcott and M, Gentzkow. "Social media and fake news in the 2016 election". Technical report, National Bureau of Economic Research, 2017.

[10] M Balmas. "When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism". Communication Research, 41(3):430–454, 2014.

[11] P, Bharadwaj and Z, Shao. "Fake news detection with semantic features and text mining". International Journal on Natural Language Computing (IJNLC); Vol, 8, No. 3, pp.17-22, 2019

[12] Babayan, Davit. "Bitcoin Bulls Spreading Fake News About Golden Asteroid: Peter Schiff." NewsBTC, NewsBTC, 29 June 2019, www.newsbtc.com/2019/06/29/bitcoin-bulls-spreading- fake news-about-golden-asteroid-peter-schiff/.

[13] Mitchell, Amy, et al. "Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed." Pew Research Center's Journalism Project, Pew Research Center, 28 June 2019, www.journalism.org/2019/06/05/many-Americans-say-made-up-news-is-a-critical-problem-that needs-to-be-fixed/.

[14] Soares, Isa, and Florence Davey-Atlee. "The Fake News Machine: Inside a Town Gearing up for 2020." CNNMoney, Cable News Network, money.cnn.com/interactive/media/the-macedonia- story/.

[15] A. Mitchell, "Many Americans say made-up news is a critical problem that needs to be fixed," Pew Research Center, https://www.pewresearch.org/journalism/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/ (accessed May 4, 2024).

[16] P. Bharadwaj and Z. Shao, "Fake news detection with semantic features and text mining," International Journal on Natural Language Computing (IJNLC), vol. 8, no. 3, pp. 17-22, 2019.

[17] H. Reddy, N. R., M. Gala, and A. Bassava, "Text-mining-based fake news detection using ensemble method," International Journal of Automation and Computing, vol. 17, no. 2, pp. 210-221, 2020. DOI: 10.1007/s11633-019-1216-5.

[18] K. Shu, A. Sliva, S. H. Wang, J. L. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017. DOI: 10.1145/3137597.3137600.

[19] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake News Detection: A Deep Learning Approach," SMU Data Science Review, vol. 1, no. 3, Article 10, 2018. Available at: https://scholar.smu.edu/datasciencereview/vol1/iss3/10.

[20] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining Dual Emotion for Fake News Detection," in Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia. ACM, New York, NY, USA, pp. 12, 2021. DOI: 10.1145/3442381.3450004.

[21] M. Mridha, A. Hamid, M. Monowar, and S. Rahman, "A comprehensive review on fake news detection with deep learning," IEEE Access, 2021. DOI: 10.1109/ACCESS.2021.3129329.

[22] P. H. Faustini and T. F. Covões, "Fake News Detection in Multiple Platforms and Languages," Expert Systems with Applications, vol. 158, p. 113503, 2020. DOI: 10.1016/j.eswa.2020.113503.

[23] W. Bangyal, R. Qasim, N. ur Rehman, Z. Ahmad, H. Dar, L. Rukhsar, et al., "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 5514220, pp. 14, 2021. DOI: 10.1155/2021/5514220.

[24] M. Thomas, "Comparative analysis of fake news detection using machine learning and deep learning techniques," JETIR, vol. 7, no. 4, pp. 1379-1385, 2020.

[25] B. 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 Proceedings of the 25th International Conference on World Wide Web (WWW), 2017, pp. 905-914.

[26] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, et al., "A Stylometric Inquiry into Hyperpartisan and Fake News," in Proceedings of the 2017 ACM on Web Science Conference, 2017, pp. 157-161.

[27] A. Zubiaga, M. Liakata, R. Procter, and G. Hoi, "Detection and Resolution of Rumours in Social Media: A Survey," ACM Computing Surveys (CSUR), vol. 51, no. 2, pp. 32, 2018.

[28] J. P. Baptista and A. Gradim, "Understanding Fake News Consumption: A Review," Social Sciences, vol. 9, no. 10, p. 185, 2020. DOI: 10.3390/socsci9100185.

[29] W. Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), 2017, pp. 422-426.

[30] M. Caschera et al., "Fake News Detection: A Survey of Evaluation Datasets," PeerJ Computer Science, vol. 7, 2021. DOI: 10.7717/peerj-cs.518.