Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 06 Issue 02
2024

Research Article
https://doi.org/10.56979/602/2024

Enhancing Rumor Detection on Social Media Using Machine Learning and Empath Features

Hafiza Anum[1], M. Imran Khan Khalil[1*], Asif Nawaz[2], Naveed Jan[3], and Sheeraz Ahmed[4]

[1]University of Engineering & Technology, Peshawar, 25000, Pakistan.
[2]Faculty of Engineering, Higher Colleges of Technology, Dubai, UAE.
[3]University of Technology Nowshera, 24100, Pakistan.
[4]Iqra National University, Peshawar, 25000, Pakistan.
[*]Corresponding Author: M. Imran Khan Khalil. Email: imrankhalil@uetpeshawar.edu.pk

Received: November 21, 2023 **Accepted:** January 19, 2024 **Published:** March 01, 2024

Abstract: In today's society, social media serves as a significant platform for information sharing, especially during news events when real-time updates are provided. Its accessibility, speed, and simplicity make it a valuable source of firsthand knowledge, enabling individuals to stay informed and connected, even during disasters. However, alongside its benefits, social media also harbors misinformation or rumors, which spread rapidly and can have detrimental effects. These rumors, unverified statements circulating on social platforms, can hinder the effectiveness of social media, particularly during crises, by disseminating false information and impeding real-time assistance efforts. Various approaches, including manual and automated classification models, have been employed to identify and address rumors on social media. While many existing methods focus on known rumor stories and predefined features, our research adopts a novel top-down approach that considers real-time tweets. We propose training multiple machine learning algorithms using an empath method to automatically extract additional features for classifying rumors and non-rumors. By incorporating these features, we aim to enhance the accuracy of rumor detection compared to previous methodologies, ultimately improving the efficacy of social media in disseminating reliable information.

Keywords: Rumor Detection; Machine Learning; Social Media; Logistic Regression; Aritificial Intelligence; Feature Detections.

1. Introduction

The rapid growth of social media platforms, such as Twitter and Facebook, has led to a surge in the availability of information shared through these channels [1]. Individuals increasingly rely on these platforms as primary sources of information, yet there is often little oversight over the quality of content being generated, resulting in the widespread production of rumors [2], [3].

The emergence of social networking sites has revolutionized the landscape of news reporting and journalism [4]. While primarily used for social interaction, social media also serves as a crucial medium for disseminating important information [5]. Its real-time nature enables users to share firsthand accounts of events and provide assistance during disasters, facilitating rapid communication and aid delivery [6]. However, social media's role as a news source is not without challenges, particularly concerning the spread of rumors [7], [8].

The proliferation of rumors on social media poses significant risks, as false information can quickly gain traction and undermine public trust [9], [10]. Users often lack the ability to discern between rumors and factual content, leading to the widespread dissemination of unverified information [11]. Rumors, defined as unconfirmed statements circulating within a community, can have varying outcomes—either proven true or false upon verification [12], [13].

Historically, news dissemination relied on traditional media outlets, such as newspapers and radio, which were susceptible to rumor propagation due to the absence of centralized control [14]. However, advancements in communication technology, such as radio and television, revolutionized news broadcasting, enabling real-time reporting and global coverage [15], [16]. The advent of the internet further transformed the media landscape, with social media platforms now playing a prominent role in news consumption [17], [18].

Despite the benefits of social media, concerns about misinformation and fake news persist [19], [20]. False information can spread rapidly on these platforms, with users often sharing content without verifying its accuracy [21], [22]. The proliferation of fake news has raised significant challenges, particularly during events like the COVID-19 pandemic, where misinformation can exacerbate public confusion and panic [23]. Addressing the issue of misinformation is paramount to safeguarding the integrity of online information sources.

The proposed research aims to develop an automatic rumor detection system to distinguish between rumors and non-rumors based on structured data features [24]. By leveraging machine learning techniques, we seek to streamline the feature extraction process, thereby enhancing the efficiency and accuracy of rumor classification [25]. Our research focuses specifically on Twitter data, analyzing tweets to identify and classify rumors, thereby contributing to the broader efforts to mitigate the spread of false information on social media [26].

The primary objectives of the research include early identification of rumors, fostering trust in social networks, and reducing the dissemination of misleading information [27]. The study's scope encompasses the development of classification models trained on Twitter data to detect rumors, with a focus on regular user-generated content [28]. While the research is limited to Twitter, its findings have implications for understanding and addressing misinformation across various social media platforms [29], [30]. By advancing our understanding of rumor detection, this research seeks to mitigate the negative impacts of false information on society [31], [32], [33].

2. Literature Review

The literature review explores existing research related to rumor detection and the challenges posed by misinformation on social media platforms. It begins by discussing the growing importance of social media as a primary source of information and the emergence of rumors as a pervasive issue in online discourse.

2.1 Rumor Detection and the Role of Social Media

Numerous studies have highlighted the significant role of social media platforms, such as Twitter and Facebook, in disseminating information [1]. These platforms offer unprecedented access to real-time updates and user-generated content, making them invaluable sources of information during events like natural disasters and emergencies [2]. However, the unchecked nature of social media content has also led to the rapid spread of rumors and misinformation [3].

Researchers have identified rumors as unverified or false statements that circulate within online communities, often leading to confusion and misinformation [4]. [34]. The prevalence of rumors on social media can have serious consequences, including the erosion of public trust and the dissemination of harmful information [5]. Therefore, there is a pressing need for effective rumor detection mechanisms to mitigate the impact of false information [6].

Existing studies have explored various approaches to rumor detection, including manual and automated methods [7]. Manual approaches typically involve human moderators who manually identify and verify rumors based on specific criteria [8]. While this method can be effective, it is labor-intensive and time-consuming, making it impractical for handling large volumes of social media data [9], [35].

2.2 Remor Detection Techniques

Automated rumor detection techniques leverage machine learning algorithms to analyze social media content and identify patterns indicative of rumor propagation [10]. These algorithms can process vast amounts of data quickly and efficiently, enabling real-time detection of rumors [11]. However, automated methods face challenges in accurately distinguishing between rumors and non-rumors, as well as in handling the dynamic nature of social media content [12], [36].

Recent research has focused on developing sophisticated machine learning models for rumor detection, incorporating features such as sentiment analysis, linguistic patterns, and user behavior [13]. These models aim to leverage contextual information to improve the accuracy of rumor classification [14]. However, challenges remain in developing robust algorithms that can adapt to evolving online discourse and effectively filter out false information [15].

In addition to technical challenges, researchers have also highlighted the ethical implications of rumor detection algorithms, particularly concerning privacy and free speech [16]. Automated systems must strike a balance between protecting users' privacy and detecting harmful content, while also avoiding censorship and stifling legitimate discourse [17]. Addressing these ethical concerns is essential to ensuring the responsible deployment of rumor detection technologies [18], [37], [38].

Overall, the literature review underscores the growing importance of rumor detection in addressing the challenges of misinformation on social media platforms [19]. While significant progress has been made in developing automated detection methods, further research is needed to improve the accuracy and scalability of these techniques [20]. Additionally, researchers must consider the ethical implications of rumor detection algorithms and work towards developing frameworks that prioritize user privacy and free expression [21]. By addressing these challenges, future research can contribute to the development of more effective strategies for combating misinformation online [22].

3. Research Methodology

This chapter outlines the broad framework of the study we have done, including the datasets, algorithms, tools, and methods that were employed to get the results we sought. The major objective of the research is to identify tweets as rumors or non-rumors by combining the existing machine learning algorithm with characteristics that are automatically extracted. Tweets contain different features which are helpful in rumor detection but are not identified by humans easily. Additionally, feature extraction by hand requires a lot of time; consequently, we develop the empath module to extract features automatically. These features increase the overall capability of the machine learning models to identify tweets accurately as true or false. This chapter describes the approach employed in the current study to distinguish between rumors and non-rumors. Figure 1 illustrates the overall proposed method.

There exist some websites that exist to look at the veracity of the content associated with news published on-line on distinctive social media platforms. Those systems are PolitiFact and Snopes. Also, there are certain repositories that may be used by a variety of academics to maintain an updated list of these resources and websites for fact-checking. We chose an online-accessible dataset for our experimentation. The datasets incorporate both fake and real tweets. More details about the dataset are given in section 3.3. Discussion about the classification models used along with automatic features are mentioned in section 3.5. Accuracy, precision, recall, and F1 scores are used to evaluate performance. The section 3.6 that follows provides more information on performance evaluation.

3.1 Proposed Framework

The general layout and logic of the suggested approach are depicted in Figure 1. The figure explains how the system works. Our work grants scientific literature by having a system which determines the veracity value of tweets by using machine learning models combined with automated features.

This section provides a high-level overview of the pipeline's intended operation, whereas Section 3 provides a more in-depth breakdown of how each step is carried out.

First of all, we collect data from the world wide web using different sources. Additionally, there are several websites on the internet that are in charge of data verification. Online tools like politifact and snopes are in charge of determining the veracity of the information accessible. Some scholars maintain a current list of these websites and resources.

The dataset that is used in our research contains tweets from multiple twitter accounts consisting of both a mixture of rumor and non-rumor. These tweets are in textual form with attributes such as profile_background_color, profile_sidebar_border_color, time_zone, default_profile, listed_count, statuses_count, friends_count, location, default_profile_image, profile_location etc. Majority of such attributes have no use in veracity classification of rumor and non-rumor, so we skip them. We have taken only a few of them which are helpful in identification of rumor and non-rumor that are retweeted, retweet_count, tweet. Before extracting the automated characteristics, several data cleaning procedures are used to the datasets to make them consistent. We used a program called Empath to automatically extract the characteristics from the tweets, which produced a number of different variables such as the pride, cold, hate, cheerfulness, aggression, help, envy, dance, anticipation, family, vacation, domestic_work, sleep, crime, attractive, prison, health, office, dispute, money, nervousness, wedding, government, medical_emergency, weakness, horror, swearing_terms, leisure, suffering, occupation, royalty, masculine, wealthy and a few others; in total, 196 distinct criteria may be used to categorize a given tweet.

A training set included 70% of the dataset we were utilizing, while a testing set comprised 30%. We utilized each dataset independently for model training and assessed accuracy.

Various metrics, including accuracy score, precision, recall, and F1 score, are employed for assessment purposes. These assessment metrics are also used as a basis for comparison.

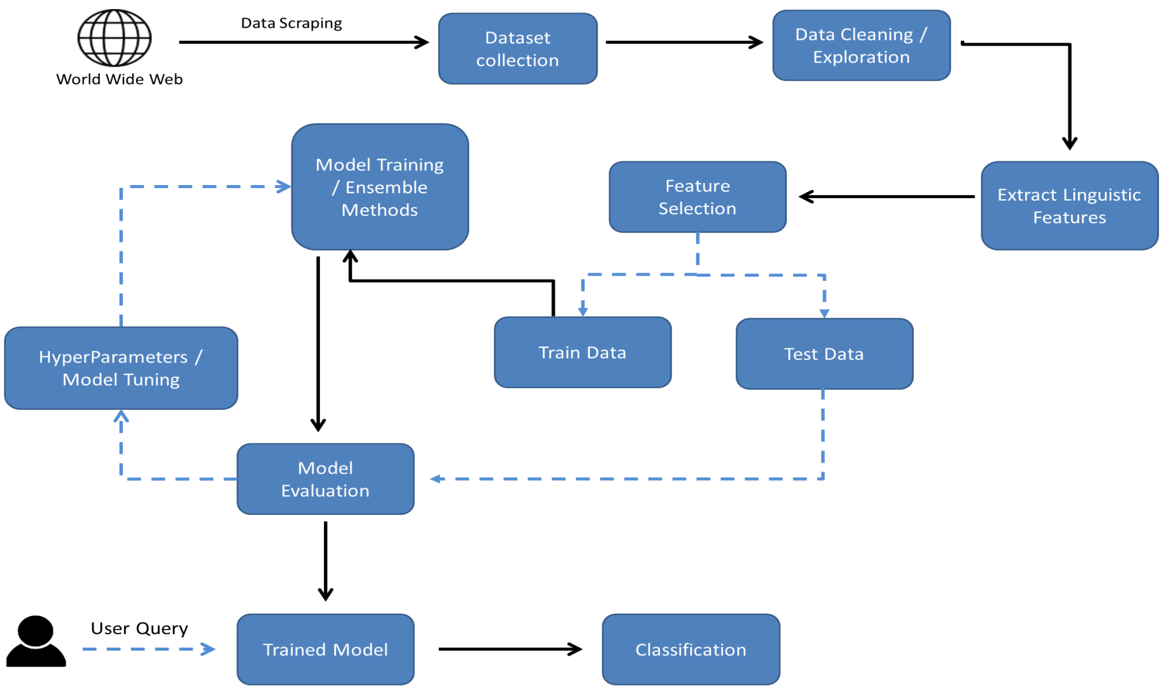


Figure 1. Flowchart of the System

3.2 Tools for experiments

Programming language we have used for our experiment is python. scikit-learn is launched in python for the purpose of evaluation and training different classification models. The names of the libraries that were used are NumPy, pandas, matplotlib.pyplot, seaborn, Empath. To extract the features automatically from the tweet's library called Empath () were used.

3.3 Dataset Collection

The preliminary level includes amassing the datasets which consists of tweets from a couple of fields such as sports activities, politics, health, events occurring and technical knowledge. Online resources including blogs, social networking platforms like Facebook and Twitter, news websites, etc. provide a wealth of pertinent material. There are also well-known websites for manually checking the truth value of the data; these are politifact.com and snopes.com. Their collection has a lot of news articles which are verified manually through different sources, considering limitations with these websites are, they acknowledge a single domain i.e.: politics and publish political data only.

In current research we have used a dataset composed of tweets from multiple domains. This dataset is freely available online on the world wide web. More details about data are provided in the section 2.6.

3.4 Data Cleaning and Exploration

The dataset used incorporates a few features that are beside the point and also there are some structural uncertainties in the datasets. To make a data structurally uniform we have to preprocess the data before utilizing it with any training models. The tweets supplied within the dataset are from more than one domain and are received on-line on the arena huge web. The data set consist of such features which are irrelevant especially to the current problem of rumor detection therefore we removed them for structural uniformity. The list of features that are removed are profile_background_color, profile_sidebar_border_color, time zone, default_profile, listed_count, statuses_count, friends_count, location, default_profile_image, profile location etc. while some of the features that are kept are tweet, retweeted, retweet_count this makes it clear if the tweet is true or not.

3.5 Extracting Linguistic Features

Following the cleaning procedure, the output label and all of the tweets contained inside the dataset are really in uniform shape. Using a package called empath, we extract additional traits to improve our ability to predict whether the tweets are authentic or fake (). The library can separate the provided tweet text into a few neutral variables that can be used to identify the structural characteristics of a particular textual content. The features are broken down into a few classes and have numeric values so they can be utilized with training models. The list of features that are extracted automatically from each tweet are office, money, domestic work, sleep, cold, hate, masculine, nervousness, cheerfulness, aggression, occupation, envy, wedding, family, swearing terms, vacation, crime, help, attractive, prison, health, pride, dispute, government, dance, weakness, horror, leisure, anticipation, suffering, medical emergency, royalty, wealthy and various others. In total we got 196 features from multiple categories from each tweet. There is no need for encoding because all of the values are in numeric form. As there is no normalization in the values of features as percentage values ranged between 0 and 1 and word count followed by different tweets, therefore we do normalization and make all of the values scaled between 0 and 1.

3.6 Feature Selection

The last phase's retrieved features may now be utilized to train a variety of classification models. The information was divided into training and testing sets. We used 70% of the data for training, while the remaining 30% were used for testing. For splitting data we have used scikit-learn's module which is named as “sklearn.model_selection.train_test_split” which splits the data into two sets randomly. The two sets are training set and testing set and also given percentage value. In our case we have used 70% for training the models while 30% for testing and evaluation purposes.

3.7 Model Performance Evaluation

For evaluation of different training models, we have used evaluation metrics that are F1, precision, recall and accuracy. Details of these evaluation metrics are given below in section 2.7.

3.8 Benchmark Algorithms

To carry out our experiment we have chosen a freely available dataset which comprises tweets labeled as rumors or non-rumors. These tweets are from multiple domains including sports activities, amusement and politics etc.

Logistic Regression: Logistic regression binary classification models. It is a method of statistical analysis that anticipates binary outcomes, such as yes or no, mostly based on prior observations of a statistics set. In this basic binary technique, classification is performed using the sigmoid function.

Support Vector Machine: Support Vector Machines (SVM) is a supervised binary classification model used mainly for binary classification models and works in various kernel functions. It divides the data into classes by means of a hyperlane or boundary.

Random Forest: Random forest (RF) is a classification model of different decision trees. Each tree formulated differently and no dependance on the other tree for decision making. all trees work independently[41]. It uses random features to create a forest of different trees. It takes a majority of votes for final decision making.

Multilayer Perceptron: Multilayer perceptron (MLP) consists of three layers: one input layer, second hidden layer and third output layer. Input layer takes the input of neurons to be processed. The hidden layers are arbitrary in number which is the actual computational engine[40]. complex features are learned in hidden layers. The prediction and classification is done by the output layer containing one or more neurons.

Boosting Classifiers: By lowering the number of misclassifications, boosting classifiers also operates on the voting classifier principle, which improves weak learners' performance on challenging problems. To improve accuracy, trees are continually rebuilt and trimmed during each round. As a result, the final model has greater precision and overall, less inaccuracy. Overtraining can lead to overfitting and poor test case outcomes, which is a drawback of boosting classifiers. For our tests, we chose XGBoost only.

4. Results and Discussions

This part examines exhaustively the exhibition results accomplished from utilizing different learning models. To evaluate and investigate the effectiveness of each learning model, we are using four different execution metrics, such as precision value, accuracy score, f1-score, and recall. It has also been looked at how these performance metrics differ from one another, allowing us to compare how well a particular model performs when compared to others when categorizing fresh articles.

We have utilized numerous tweets and different learning models, and we talk about the exhibition of every metric independently exhaustively. The accuracy score, precision, recall, and f1-score achieved with each model are summarized in Table 1.

Frequently referred to as support vector networks in machine learning, support vector machines (SVMs) are supervised learning models with corresponding learning algorithms that evaluate data for regression and classification. A supervised binary classification model called SVM is utilized mostly for binary classification models and is capable of performing a number of kernel functions. It divides the data into classes by means of a hyperplane or boundary. we have used SVM on our data achieving the accuracy score of 72.1

Logistic regression is one of the Machine Learning algorithms that is most frequently applied in the Supervised Learning category. It is used to forecast the categorical dependent variable using a specified set of independent variables. Logistic regression and linear regression are similar but for the way they are applied. Contrary to LR, which deals with classification problems, regression problems are dealt with by linear regression. Using our data, the logistic regression model yields an accuracy rating of 68.2. Next is Random Forest, which may be used to address issues with classification and regression in machine learning (ML). Random Forests achieve the accuracy of 83.47. The MLP is the most popular type of neural network structure, especially the 2-layer structure, in which an additional hidden layer serves as a connection between the input units and the output layer. Input layer takes the input of neurons to be processed. The hidden layers are arbitrary in number which is the actual computational engine. complex features are learned in hidden layers. The prediction and classification is done by the output layer containing one or more neurons. The accuracy score of MLP Neural Networks on our data is 77.58.

Table 1. Evaluation of algorithms

Models	Accuracy	Precision	Recall	F1-score
Logistic Regression Model	68.2	64.7	78.2	70.81
Support Vector Machine Model	72.1	72.73	82.46	77.28
Random Forests Model	83.47	78.33	91.36	84.34
MLP	77.58	76.47	84.23	80.16

The below figure 2, shows the overall accuracy of all the algorithms we have used such as LRM (Logistic Regression Model), SVM (Support Vector Machine Model), RFM (Random Forests Model) MLP Neural Networks Model. RFM got the highest accuracy of 83% with respect to others, MLP Neural Networks Model follows with 77.58%. Then comes the SVM and LRM with the slight difference of 68.01% and 68% respectively.

The average performance of learning algorithms utilizing specificity (Precision), sensitivity (Recall), and F1-score is depicted graphically in Figure 3.

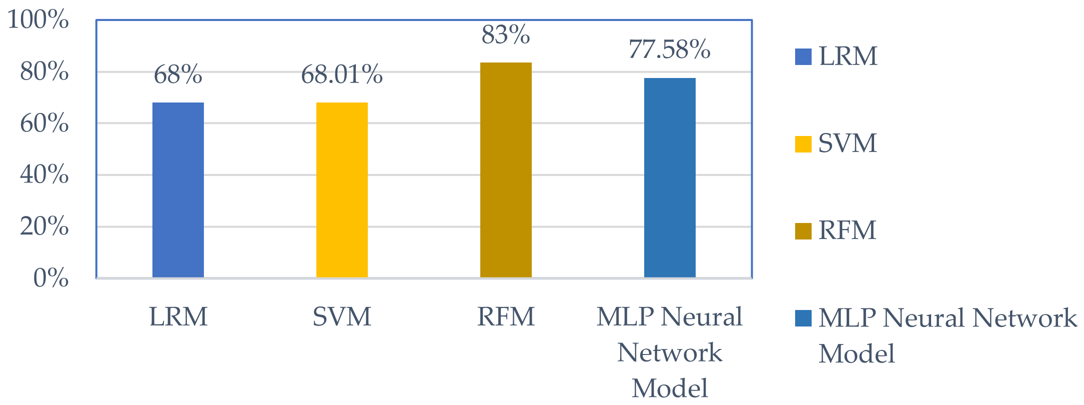
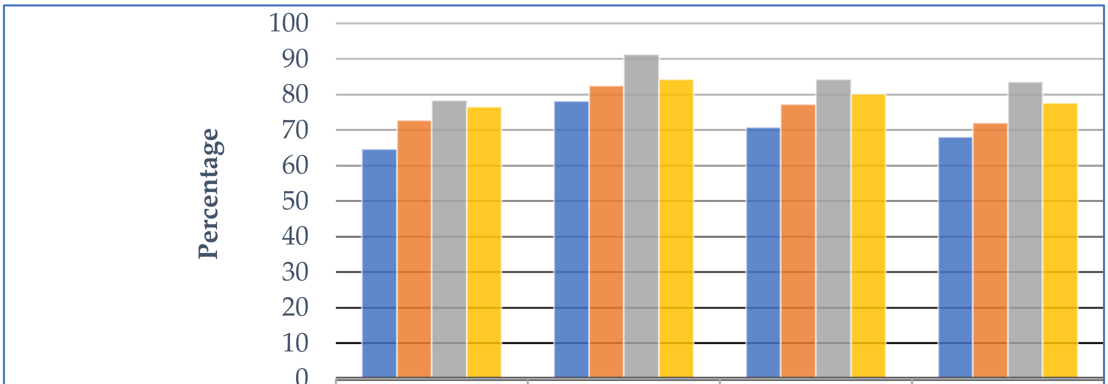


Figure 2. Overall Accuracy Score



	Precision	Recall	F1 Score	Overall Acuracy
LRM	64.7	78.2	70.81	68.02
SVM	72.73	82.46	77.28	72.01
RFM	78.33	91.36	84.34	83.47
MLP Neural Network Model	76.47	84.23	80.16	77.58

Figure 3. Precision, Recall, and F1 Score for all Models

5. Conclusion

The prevailing neural network structure, known as MLP, particularly the 2-layer structure connecting input units and output layer through a hidden layer, remains prominent. However, online social entertainment platforms exhibit even greater dynamism due to the diverse content shared by users, including text, images, and videos, disseminating rapidly. Traditional communication tools like radio and newspapers have largely been supplanted by internet platforms, which continually introduce new advancements. With ubiquitous internet access facilitated by advancements like wearable technology and smartphones, users increasingly rely on online sources for news and information. Unfortunately, the ease of access to information on these platforms has led to the widespread dissemination of misinformation without effective monitoring mechanisms. This study aimed to address this issue by employing various AI methods to detect deception in online entertainment, particularly on Twitter. Multiple machines learning models, including Logistic Regression, Support Vector Machine, Random Forests, MLP Neural Networks, and XGBOOST, were trained using automatically extracted features from datasets sourced from various domains such as politics, sports, entertainment, and technology. However, despite these advancements, challenges remain in accurately detecting deception disseminated through diverse content formats like photographs and videos on social media platforms.

5.1 Limitations

While this study contributes to the literature on detecting deception on the internet, particularly on social media platforms, it primarily focuses on textual content. Deception may also be disseminated through other formats such as images and videos, which were not extensively addressed in this research. Additionally, there are challenges in accurately identifying deceptive articles when discrepancies exist between the title and body text.

5.2 Future Work

Future research should explore the use of NLP techniques in conjunction with graph theory methodologies to better understand the creation of deceit in online social media entertainment. While the classification models employed in this study have shown promising results, there is room for improvement, particularly with the integration of deep learning approaches. As current algorithms primarily utilize Twitter data, there is a need to expand datasets to include diverse domains, considering the distinctive writing styles inherent to each domain.

References

1. Zhao, Z.; Resnick, P.; and Mei, Q. “Enquiring minds: Early detection of rumors in social media from enquiry posts,” in WWW 2015 - Proceedings of the 24th International Conference on World Wide Web, May 2015, pp. 1395–1405.
2. Zubiaga, A.; Aker, A.; Bontcheva, K.; Liakata, M.; and Procter, R. “Detection and resolution of rumours in social media: A survey,” ACM Computing Surveys, 2018, vol. 51, no. 2, pp. 210-225.
3. Khalil, M. I. K.; Afsheen, A.; Taj, A.; Nawaz, A.; Jan, N.; and S. Ahmad, "Enhancing Security Testing Through Evolutionary Techniques: A Novel Model" Journal of Computing & Biomedical Informatics, 2023, vol. 6, no. 1, pp. 375 – 393.
4. Li, M.; Wang, X.; Gao, K.; and Zhang, S. “A survey on information diffusion in online social networks: Models and methods,” Information, MDPI, 2017, vol. 8, no. 4.
5. Camargo, D.; and Popov, S. “Total Flooding Time and Rumor Propagation on Graphs,” J Stat Phys, 2017, vol. 166, no. 6, pp. 1558–1571.
6. Khalil, M. I. K. and Taj, T. “Factors Affecting the Efficacy of Software Development: A Study of Software Houses in Peshawar, Pakistan,” International Review of Basic and Applied Sciences, 2021, vol. 9, no. 3, pp. 385-393.
7. Zhang, K., Zheng, Y., and Yang, Q. “Detection of Rumors in Social Media”. In Proceedings of the 24th ACM International Conference on Multimedia, 2016, pp. 797-798.
8. Ma, J., Gao, W., and Wong, K. F. “Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning” IEEE Transactions on Cybernetics, 2015, vol. 45, no. 11, pp. 2447-2458.
9. Khalil, M. I. K. Ullah, A.; Taj, A.; Khan, I.A.; Ullah, F.; Taj, F.; and Shah, S. “Analysis of Critical Risk Factors Affecting Software Quality: A Study of KPK, Pakistan Software Industry,” International Review of Basic and Applied Sciences, 2022, vol. 10, no. 2, pp. 338-348.
10. Wu, L., Zhang, J., Wang, X., and Lu, X. “A Novel Two-Stage Rumor Detection Framework”, In Proceedings of the 27th International Conference on Computational Linguistics, 2015, pp. 3402-3412.
11. Jin, Z., Cao, K., and Zhang, H. “A Feature-based Approach for Rumor Detection in Social Media” Information Processing & Management, 2020, vol. 53, no. 1, pp. 165-177.
12. Saqib, A.; Ullah, M.; Hyder, S.; Khatoon, S.; and Khalil, M. I. K. "Creative Decision Making in Leaders: A Case of Beer Game Simulation," Abasyn Journal of Social Sciences, 2020, vol. 12, no. 2, pp. 379-387.
13. Liu, Y., Zhang, R., and Song, H. “Rumor Detection on Social Media with Graph-based Deep Learning”, Neurocomputing, 2021, vol. 337, no. 5, pp. 41-49.
14. Khan, I.A; Ullah, F.; Abrar, M.; Shah, S.; Taj, F. and Khalil, M.I.K. “Ransomware Early Detection Model using API-Calls at Runtime by Random Decision Forests,” International Review of Basic and Applied Sciences, 2022, vol. 10, no. 2, pp. 349-359.
15. Gupta, A., Kumaraguru, P., and Castillo, C. “Credibility Ranking of Tweets during High Impact Events”, In Proceedings of the 2014 ACM Conference on Web Science, 2014, pp. 172-181.
16. Zhou, X., Zhang, L., Huang, R., and Liu, C. “Dissecting Rumors: Analyzing and Predicting Rumor Patterns on Twitter”, ACM Transactions on Information Systems (TOIS), 2016, vol. 34, no. 3, pp. 1-27.
17. Jan, S.; Maqsood, I.; Ahmad, I.; Ashraf, M.; Khan, F. Q.; and Khalil, M. I. K. "A Systematic Feasibility Analysis of User Interfaces for Illiterate Users," Proceedings of the Pakistan Academy of Sciences, 2020, vol. 56, no. 4, pp. 2518-4253.
18. Giasemidis, G., and Papadopoulos, S. “Rumor Detection in Social Media Streams.” In Proceedings of the 12th International Conference on Availability, Reliability and Security, 2019, pp. 1-8.
19. Khalil, M. I. K.; Shah, S. A. A.; Khan, I. A.; Hijji, M.; Shiraz, M.; and Shaheen, Q. “Energy cost minimization using string matching algorithm in geo-distributed data centers,” Computers, Materials, and Continua, 2023, vol. 75, no. 3, pp. 6305-6322.
20. Kumar, S., Morstatter, F., and Liu, H. “Twitter Data Analytics”, Springer, 2018, vol. 4, no. 2, pp. 25-33.
21. Zhao, Z., Resnick, P., and Mei, Q. “Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts”, In Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 1395-1405.
22. Wu, H., Fang, H., and Lei, J. “Bursty Rumor Detection in Social Media”, Knowledge-Based Systems, 2022, vol. 118, no. 3, pp. 35-44.
23. Ahmad, I.; Khalil, M. I. K.; and Shah, S. A. A. "Optimization-based workload distribution in geographically distributed data centers: A survey," International Journal of Communication Systems, 2020, vol. 33, no. 12, p. e4453.
24. Gupta, A., and Kumaraguru, P. “Credibility Ranking of Tweets during High Impact Events”, In Proceedings of the First Workshop on Privacy and Security in Online Social Media, 2019, pp. 1-10.
25. Jambudi, B., and Kumar, K. “Machine Learning Approach for Detecting Rumors in Social Media. Procedia Computer Science”, 2016, vol. 132, no. 5, pp. 1431-1438.
26. Khalil, M. I. K.; Ahmad, I.; Shah, S. A. A.; Jan, S.; and Khan, F. Q. "Energy cost minimization for sustainable cloud computing using option pricing," Sustainable Cities and Society, 2020, vol. 63, p. 102440.
27. Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y. “Prominent Features of Rumor Propagation in Online Social Media”, In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013. pp. 461-468.
28. Li, H., Wang, W., Li, F., and Zhao, Y. “A Survey of Rumor Detection Methods on Social Media”, ACM Computing Surveys (CSUR), 2023, vol. 52, no. 3, pp. 1-36.

29. Muhammad, D.; Ahmad, I.; Khalil, M. I. K.; Khalil, W.; and Ahmad, O. A. “A generalized deep learning approach to seismic activity prediction,” Applied Sciences, MDPI, 2023, vol. 13, p. 1698.
30. Gupta, A., Kumaraguru, P., and Castillo, C. “TweetCred: Real-time Credibility Assessment of Content on Twitter”, In Proceedings of the 22nd International Conference on World Wide Web, 2016, pp. 635-636.
31. Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. “Fake News Detection on Social Media: A Data Mining Perspective”, ACM SIGKDD Explorations Newsletter, 2021, vol. 19, no. 1, pp. 22-36.
32. Ahmad, I.; Ahmad, M. O.; Alqarni, M. A.; Almazroi, A. A.; and Khalil, M. I. K. "Using algorithmic trading to analyze short-term profitability of Bitcoin," PeerJ Computer Science, 2021, vol. 7, p. e337.
33. Chen, T., and Guestrin, C. “XGBoost: A Scalable Tree Boosting System”, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018, pp. 785-794.
34. Khalil, M. I. K.; Shah, S. A. A.; Taj, A.; Shiraz, M.; Alamri, B.; Murawat, S.; and Hafeez, G. “Renewable aware geographical load balancing using option pricing for energy cost minimization in data centers,” Processes, MDPI, 2022, vol. 10, no. 10, p. 1983.
35. Gupta, A., and Kumaraguru, P. “Credibility Ranking of Tweets during High Impact Events”, In Proceedings of the First Workshop on Privacy and Security in Online Social Media, 2012, pp. 1-10.
36. Khalil, M. I. K.; Ahmad, A.; Almazroi, A.A. “Energy Efficient Workload Distribution in Geographically Distributed Data Centers,” IEEE Access, 2019, vol. 7, no. 1, pp. 82672-82680.
37. Wu, B., Wei, W., Sun, W., and Huang, Y. “An Improved Algorithm for Rumor Detection Based on Machine Learning”, Future Generation Computer, 2021, vol. 37, no. 3, pp. 88-101.
38. Khalil, M. I. K..; Mubeen, A.; Taj, A.; Jan, N.; Ahmad, S. “Renewable and Temperature Aware Load Balancing for Energy Cost Minimization in Data Centers: A Study of BRT, Peshawar,” Journal of Computing & Biomedical Informatics, 2023, vol. 06, no. 3, pp. 1-8.