INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Transformative User Credibility Assessment on Twitter: A RNN based Heuristic Approach

Vinita Nair[1]*, **Jyoti Pareek**[2]

1 Research Scholar, Department of Computer Science, Gujarat University, Ahmedabad, 380009, Gujarat, India
2 Professor and Head, Department of Computer Science, Gujarat University, Ahmedabad, 380009, Gujarat, India

OPEN ACCESS

Received: 08-12-2023

Accepted: 22-04-2024

Published: 14-05-2024

Citation: Nair V, Pareek J (2024) Transformative User Credibility Assessment on Twitter: A RNN based Heuristic Approach. Indian Journal of Science and Technology 17(20): 2056-2063. https://doi.org/10.17485/IJST/v17i20.3108

* **Corresponding author**.

vinitanair@gujaratuniversity.ac.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To construct a comprehensive weighted multi-dimensional model to assess the impact of influence score of Twitter users, considering the credibility based on user profile, their tweets and social interactions aiming to empower users in distinguishing fake news or misinformation. **Methods:** The credibility evaluation is formulated based on text analysis, user account attributes, and user social engagement. We've gathered around 100,000 tweets from 100 users using Tweepy API over a six-month duration for the purpose of evaluating credibility. The collected tweets spanned diverse professions namely politics, entertainment, business, science, sports, and trending topics. We chose to utilize a self-devised deep active learning model to classify and label the unlabelled data instead of engaging in time-consuming human annotation for the tweets we gathered. **Findings:** The obtained accuracy for influence score evaluation for Recurrent Neural Network, Random Forest, Naïve Bayes, Decision Tree, and Support Vector Machine are 89.03%, 79.10%, 81.59%, 73.06% and 79.45% respectively. Upon reviewing and analysing the outcomes, RNN surpassed all other models achieving an exceptional accuracy of 89.03%. **Novelty:** Employing a weighted multi-dimensional framework, it systematically evaluates the influence score by considering the credibility of both users and tweets within the context of Twitter. Weighted features are instrumental in capturing the relative importance of different elements, leading to a more refined and context-aware decision-making process. In contrast to earlier research, which predominantly centred on the credibility of individual tweets, our research work shifts the focus to a broader perspective, encompassing the credibility of users, their tweets and their overall social influence. By incorporating user influence score, the framework not only empower users in discerning fake news or mis-information but also elevates their ability to gauge the reliability of information, offering a nuanced approach to news credibility analysis.

Keywords: Active Learning; Credibility Score; User Influence; Twitter; Machine Learning; Recurrent Neural Network

1 Introduction

What sets our proposed framework apart from the solutions found in existing literature (Table 3) is its unique ability to effectively evaluate and determine User Influence score based on weighted multi-dimensional aspects namely Text Credibility, User Credibility and Social Credibility. The text credibility involves assessing the credibility through text analysis of the tweet, user credibility relies on user account attributes, and social credibility is determined by attributes indicating user engagement and influence on Twitter. Furthermore, the experimental results also represent an evaluation of the bot score utilizing Botometer[1]. Botometer is a public tool, built using supervised machine learning classifier to distinguish bot-like and human-like accounts on Twitter based on user features. The multifaceted uses of reputation and influence on Twitter extend across various domains, encompassing political engagement, the propagation of rumours, analysis of human mobility, transportation studies, and investigations into epidemiological trends. It is imperative to identify and analyse a specific cohort of highly active users, as they wield significant influence in shaping and disseminating trends, ideas, information, and rules within the Twitter community[2]. Despite their novelty, the approaches studied in the literature predominantly employ supervised techniques, relying on labelled data to differentiate between fake and non-fake labels. In contrast, the paper describes the application of Recurrent Neural Network on self-tailored Twitter dataset, and the credibility assessment involves labelling through a deep active learning technique. Many fake news detection models primarily focus on textual content and user profile features. There is a notable research gap in the development of robust models capable of effectively analysing and integrating information from various modalities. This involves incorporating weighted features to enhance the precision of credibility assessment in detecting fake news, and our research stands out in this regard. The inclusion of weighted features serves to assign varying degrees of importance or significance to different aspects of the data. This allows the model to prioritize and emphasize specific features during the analysis, contributing to a more nuanced and accurate evaluation, especially in the context of credibility assessment in fake news detection.

2 Methodology

To study the user credibility, the tweet and user entities are to be studied, followed by social presence of user actions such as retweeting, retweeting with comment, reply to the tweet and liking the tweet. We represent it as user X is a follower of user Y when X carries out the action by following Y.[2]

2.1 Data Collection

Using the Tweepy API, we collected 100,000 tweets derived from 100 individual users over the duration of six months. The extraction involved real-time tweets based on various query parameters like, user handle, time duration, keyword, latest trends, etc. We gathered tweets from different professional backgrounds such as entertainment, business, politics, science, sports, etc. The tweets were selected for the users who are non-bots, having public tweets, non-zero friends and followers, active on twitter. The users' identity is concealed to ensure confidentiality and have annotated them as User A, Topic A and so for the reader's purpose. Tweets have been chosen specifically from non-bot users, non-zero friends and followers, active on twitter. Our main hypothesis is that to analyze and evaluate credibility level of user considering user tweets, user timeline and user social connection. For our experimentation, we took into account the features outlined in Table 1. The dataset is further pre-processed, lemmatized, balanced prior the training process. The features used for training the model were 'text', 'favorite_count', 'retweet_count', 'followers_count', 'friends_count', 'status_count'.

2.2 Data annotation scheme

Data annotation for the scrapped tweets were achieved by applying a self-devised deep active learning approach to annotate our unlabeled dataset, with sentiment polarity serving as the information criterion. Active learning is machine learning approach that involves selecting the most informative samples for labeling to improve the model's performance with fewer labelled samples. The approach is applied to small hand-labelled dataset to query unlabeled data points. The goal is to reduce the amount of labelled data needed for training while improving the model's accuracy.

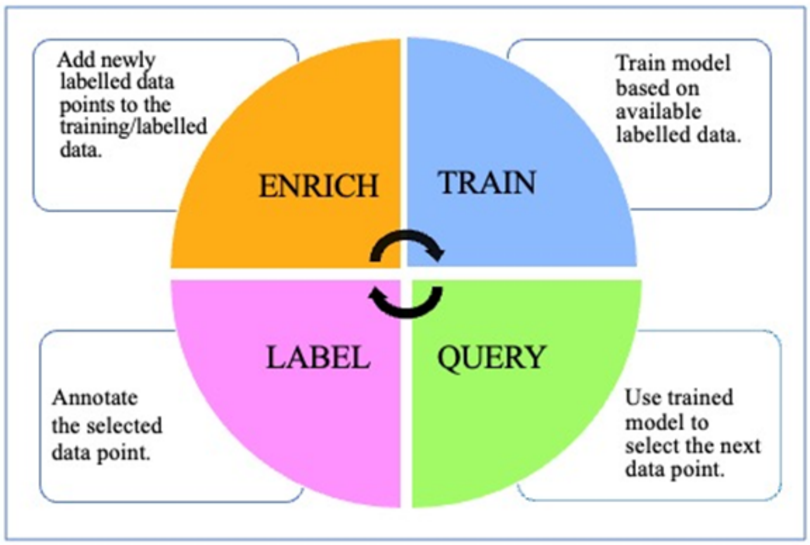

Fig 1. The Iterative Pathway of Active Learning Workflow

Figure 1 elaborates the process of iteratively selecting and labeling the unlabeled instances based on active learning model, progressively improves the model's performance by minimizing resources utilized for manual labeling. Customization of the process of query pattern, data efficiency, varied data selection strategies, real-time, human loop-in collaboration are other characteristics summing up to the excellence of active learning model. To compute credibility scores for tweets, we needed training data, which was scrapped using Tweepy API. The normalized dataset is labelled as credible and non-credible using deep active learning technique. Sentiment polarity is considered as a parameter for human annotation for unlabeled dataset. The human annotated dataset was prepared with class labels as credible and non-credible based on characteristics that affects the tweet veracity included: lack of evidence or context, emotional manipulation, sensational and conspiracy theories, in-complete or poor grammar, non-formal language, etc. These characteristics alone are not definitive for evaluating tweet credibility as more fact-checking and cross-reference techniques are required for complete tweet veracity.

2.3 Feature Engineering

To assess our experiment, we've categorized features into three distinct groups based on their scope, that will assist to measure the User Influence Score. Table 1 outlines the selected attributes for assessing user, tweet, and social engagement credibility.

Table 1. Influence Score Determinants for Twitter Users - A holistic examination of Tweet Content, User Profile, and Social Engagement Features

FEATURES FOR TEXT CREDIBILITY

Continued on next page

Table 1 continued

Tweet Features*	Description
Retweet Count (Sc_{TwRT})	Number of times a given tweet was retweeted.
Favourite Count ($ScTw_F$)	Number of times a given tweet was marked as favourite.
Relevant Score Ratio (Sc_{TwRW})	Ratio of real values other than special characters or stop words to the total words.
Sentiment Score [Sent-Tweet + Sent-Emoji + Sent-Image] (Sc_{TwS})	Sentiment polarity of the entire tweet including the sentiment of emojis used in each tweet. Sentiment of text in the image is also evaluated (if available)
Sentiment of Tweet Replies (Sc_{TwR})	Sentiment of replies for each tweet.
Positive Words Ratio (Sc_{TwP})	Ratio of positive words to the total words.
Negative Words Ratio (Sc_{TwN})	Ratio of negative words to the total words.
Neutral Words Ratio(Sc_{TwNe})	Ratio of neutral words to the total words.
Emoji Score Ratio (Sc_{TwEmo})	Ratio of emojis to the total words.
Punctuation Score Ratio (Sc_{TwPun})	Ratio of punctuation to the total words.
Hashtag Score Ratio (Sc_{TwH})	Ratio of hashtags to the total words.
User Mention Score Ratio (Sc_{TwM})	Ratio of user mentions to the total words.
URL Score Ratio (Sc_{TwURL})	Ratio of URLs to the total words.
URL Domain Count (Sc_{TwDM})	Ratio of domains embedded in URLs to the total extracted domains.
FEATURES FOR USER PROFILE CREDIBILITY	
Tweet Features*	**Description**
User Location (Sc_{UL})	True (1), if user location is mentioned in user profile, otherwise False (0)
User URL (Sc_{UU})	True (1), if user URL is mentioned in user profile, otherwise False (0)
User Verified (Sc_{UV})	True (1), if user is verified, otherwise False (0)
User Description (Sc_{UD})	True (1), if user has provided his/her description mentioned in user profile, otherwise False (0)
User Geo-Location (Sc_{UG})	True (1), if user has shared geo-location mentioned in user profile, otherwise False (0)
Account Age of User – Creation / Active Age (Sc_{UAge})	Active age of a user on Twitter
Average of last 50 tweets, cred score (Sc_{UAvg50})	Average tweet credibility score, for last 50 tweets for a user.
FEATURES FOR USER SOCIAL ENGAGEMENT CREDIBILITY	
Tweet Features*	**Description**
FollowersCountRatio to Friend_Count Age ($UScl_{FlCR}$)	Ratio of friend_count to followers_count of a user.
FriendsCountRatio to Account Age ($UScl_{FrCR}$)	Ratio of friend_count to active age of a user.
StatusCountRatio to FollowersCount ($UScl_{StFl}$)	Ratio of statuscount to followers_count of a user.

* The user influence determinants features with feature name and corresponding notation are subsequently utilized in Equations (1), (2), (3) and (4)

For our experiment, participants were selected from diverse fields such as politics (Narendra Modi, Donald Trump, Barack Obama, etc.), business (Ratan Tata, Gautam Adani, Elon Musk, Tim Cook, Bill Gates, etc.), companies (Google, Microsoft, Reliance Group, Twitter, etc.), actors (Shahrukh Khan, Akshay Kumar, Kapil Sharma, Mahendra Singh Dhoni, etc.), news & television (The New York Times, ESPN, CNN, Reuters, NatGeo, etc.), music (A. R. Rehman BTS, Shakira, BrunoMars, Jennifer Lopez, etc.), Sports (Sachin Tendulkar, Virat Kohli, MS Dhoni, Cristiano Ronaldo, etc.), trending acts/movies (The Kerala Story, Pathan, Adipurush, MeTooMovement etc.). For example, the tweet with ID 1672747830463586308 : #usermention# Population collapse is a severe danger to the future of civilization, tweeted by User A, the following information is collected. We additionally collected 100 tweets posted by User A. Thereafter we calculated Text Credibility, User Profile Credibility score and User Social Influence Score based on Algorithm 1.0 as described using Equations (1), (2), (3) and (4).

Table 2. Comprehensive Weighted Multi-dimensional Assessment of User A - User Influence Score and Additional metrics

Credibility Metric	Metric Value
Tweet ID	1672747830463586308
Text	#usermention# Population collapse is a severe danger to the future of civilization
Sentiment Score	0.5
Retweet Count	2017
Favourite Count	13676

Continued on next page

Table 2 continued

Words Count	13
Relevant Word Count	8
Follower Count	144857623
Friend Count	336
Follower Count Ratio	2.31
Friend Count Ratio	23.87
Average Tweet Credibility for last 100 tweets	0.78
Tweet Credibility Score (A)	0.43
User Credibility Score (B)	7.68
User Influence Score (C=A+B)	8.10 / 10
Botometer Score	0.9 / 5

2.4 Algorithm to enhance User Credibility Assessment

This self-devised weighted algorithm sets us apart from existing solutions and is customized to effectively address the challenges of credibility analysis. The algorithm describes an overall analysis and assessment of user profile, tweet analysis and user social interaction patterns for cumulative credibility score evaluation.

ALGORITHM 1.0: Enhancing User Credibility Assessment: A Weighted Multi-dimensional approach for Evaluating User Influence Scores in the Twitter verse.

INPUT: User profile data, including account details, tweet history, social interaction counts for Twitter data.

OUTPUT: Cumulate credibility score, indicating credibility level of User on Twitter based on user profile and social interactions along with extensive tweet analysis.

Step 1. Data Retrieval: Retrieve twitter user profile, tweet history and social interaction data using Tweepy API.

Step 2. Labelling Data: Label unlabelled data using deep active learning model.

Step 3. Data cleaning and normalization: Tweet content basic pre-processing.

Step 4. Loading of Dataset: Load CSV file for collected and labelled data with user profile, tweet, and follower details with social interaction counts.

Step 5. Feature identification: may include user account age, geolocation, verification status, follower count, content analysis and user engagement and many more.

Step 6. Weighing and Scaling: For user credibility influence score, assign weights to each feature. Weights are assigned based on ratio between degree to frequency of feature for the given dataset.

Step 7. Credibility Score Calculation: Blended overall credibility score evaluation combining of weighted and scaled features. (Equations (1), (2), (3) and (4))

Step 8. Performance metrics: Comparative analysis of performance metrics for each classification machine learning algorithms.

Step 9. Retraining model: Train model regularly for accommodating any updates in feature weights or for inclusion of new insight to improve model accuracy.

Equations (1), (2), (3) and (4) were derived by selecting a distinct set of user and tweet features and subsequently adjusting the weights through iterative trial-and-error experiments.

Equation (1). **Tweet content-based credibility score**

$$TweetCredibilityScore\ (TwCr_wc_u) = (0.08 * Sc_{TwRT}) + (0.075 * ScTw_F) + (0.08 * Sc_{TwRW}) + (0.1 * Sc_{TwS}) + (0.1 * Sc_{TwR}) + (0.055 * Sc_{TwP}) + (0.055 * Sc_{TwN}) + (0.055 * Sc_{TwNe}) + (0.06 * Sc_{TwEmo}) + (0.06 * Sc_{TwPun}) + (0.08 * Sc_{TwH}) + (0.08 * Sc_{TwM}) + (0.06 * Sc_{TwURL}) + (0.06 * Sc_{TwDM}) \quad (1)$$

Equation (2). **User profile-based credibility score**

$$UserProfileCredibilityScore\ (UsrPrCrdSc_u) = (WL * Sc_{UL}) + (WU * Sc_{UU}) + (WD\ (*Sc_{UD}) + (WV * Sc_{UV}) + (WG * Sc_{UG}) + (WC * Sc_{UAge}) + (WA100 * Sc_{UAvg100}) \quad (2)$$

Equation (3). **User social engagement credibility score**

$$Social\_engagement\_score\ (UsrSCLSc_u) = (UScl_{FlCR} * 0.45) + (UScl_{FrCR} * 0.35) + (UScl_{StCR} * 0.20) \quad (3)$$

Equation (4). **User Influence Score - cumulative user credibility score**

$$UserInfluenceScore\ (UsrCrdInflSc_u) = (0.45 * TwCrdSc_u + 0.2UsrPrCrdSc_u + 0.35 * UsrSCLSc_u) \quad (4)$$

3 Result and Discussion

Within this section, a comparative overview of different studies is presented. Table 3 outlines the research contributions made by various authors.

Table 3. A Comparative Exploration of Research Findings

Sr. No.	Authors	Classifiers	Features	Dataset	Accuracy (%)
1.	Iftene A et. al[3]	Neural Network	retweets count, favourite count, word count, relevant word count ratio and user profile features.	Real time tweets	85.2
2.	Abu-Salih et. al[4]	CredSaT, IDF, NDGC	Time-aware, domain-based and user features.	Real time tweets	85.6
3.	Garcia et. al[5]	NOFACE framework + LDA + Clustering + Apriori	followers, favourites, retweets and mentions.	a. COVID-19 December 2019 b. US elections November 2020	silhouette coefficient- 0.0095 and 0.12 respectively for mentioned datasets.
4.	Sitaula et. al[6]	SVM (RBF Kernel), Linear SVM, Logistic Regression, Random Forest, AdaBoost, Naive Bayes, Gradient Boosting, Decision Tree	Source and content features	a. BuzzFeed b. PolitiFact	Linear SVM = 77 (BuzzFeed) Linear SVM = 82 (PolitiFact)
5.	Cardinale et. al[7]	T-CREo	Content, account details and followers.	Real time tweets	86
6.	Azer et. al[8]	Naïve Bayes, SVM, KNN, Logistic Regression, Random Forest	User and tweet based	Pheme dataset	83.4
7.	Khan et. al[9]	Random Forest, MLP	social profiles, tweets credibility, sentiment score, and h-indexing score	Real time tweets	90
8.	Ahmad et. al[10]	Random forest, SVM, Naive Bayes, KNN, J48 Decision tree, and MLP.	user-based content-based and hybrid features	Real time tweets	Random forest = 96.59 J48 decision tree = 96.54 SVM = 95.65 Naïve Bayes = 95.39 KNN= 95.11 MLP = 95.03
9.	Proposed framework	RNN	Extensive features accommodating User profile, User tweets, user engagement and user influence score as described in Table 1	Real time tweets	RNN = 89.03, Random forest = 79.10, Naïve Bayes = 81.59, Decision Tree = 73.06, SVM = 79.45

The proposed model assessed the cumulative weighted influence score, considering ranking models such as user popularity, metadata analysing, and community-powered methods[11]. During the course of research latest 100 tweets and retweets are taken in account as the essential features, acknowledging their significance in determining profile credibility, as highlighted[12]. Supervised classifiers Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine are evaluated, and comparative results are presented in Figure 3. Among the evaluated supervised methods for credibility evaluation Naïve Bayes demonstrated impressive accuracy of 81.59% for user influence score. In our quest to enhance the model's accuracy, we incorporated Recurrent Neural Network (RNN), renowned for effectively tackling challenges in Natural Language Processing, and leveraging them enabled notable improvements in our results as shown in Figure 2.

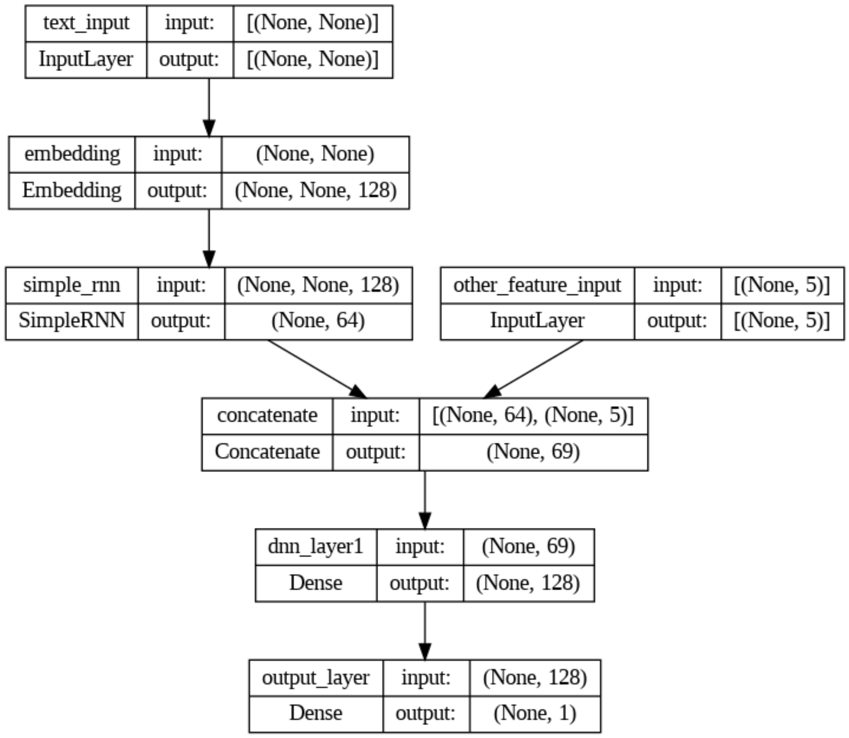

Fig 2. RNN model summary for User Influence evaluation for Twitter Users

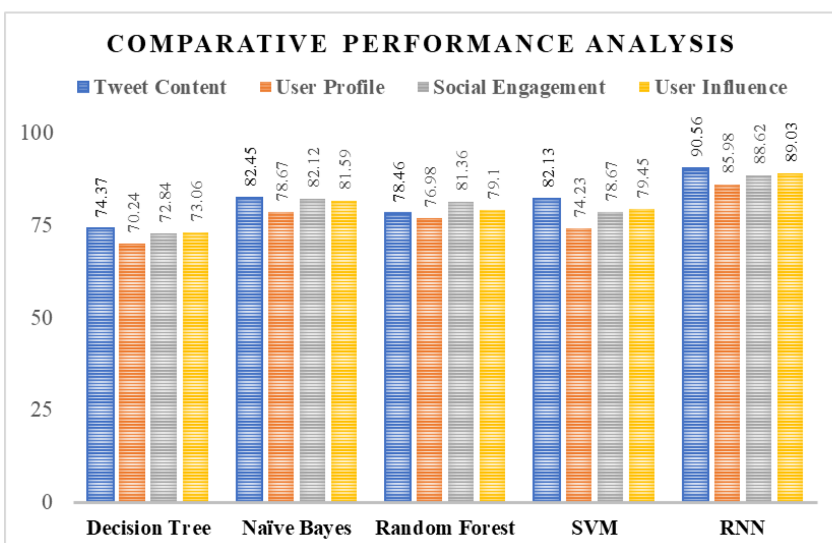

Fig 3. Comparative Analysis of Tweet content, User Profile and User Social engagement features impacting Influence Score for Twitter Users

4 Conclusion

This study delineates a comprehensive overview of the weighted multi-dimensional model for evaluating user influence score for Twitter users considering user content, user profile and user social engagement score as crucial factors in the evaluation process. The obtained outcomes for influence score evaluation for Recurrent Neural Network, Random Forest, Naïve Bayes, Decision Tree, and Support Vector Machine were 89.03%, 79.10%, 81.59%, 73.06% and 79.45% accuracy respectively. As a part of future work, we aim to develop a real-time system capable of capturing and analyzing user tweets instantaneously. The system may offer real-time credibility score for the evaluated content, enabling prompt and reliable assessment of user credibility. Additionally, we plan to delve into exploring further features, particularly user network and behavioral aspects that impact the overall credibility of both Twitter users and their tweets. By incorporating these additional dimensions, we aspire to gain deeper insights and enhance the comprehensive evaluation of user credibility in our future research.

References

1) Yang KC, Ferrara E, Menczer F. Botometer 101: social bot practicum for computational social scientists. *Journal of Computational Social Science*. 2022;5(2):1511–1528. Available from: https://dx.doi.org/10.1007/s42001-022-00177-5.
2) Al-Yazidi S, Berri J, Al-Qurishi M, Al-Alrubaian M. Measuring Reputation and Influence in Online Social Networks: A Systematic Literature Review. *IEEE Access*. 2020;8:105824–105851. Available from: https://dx.doi.org/10.1109/access.2020.2999033.
3) Iftene A, Gîfu D, Miron AR, Dudu MS. A real-time system for credibility on twitter. In: Proceedings of the Twelfth Language Resources and Evaluation Conference. European Language Resources Association. 2020;p. 6166–6173. Available from: https://aclanthology.org/2020.lrec-1.757.
4) Abu-Salih B, Wongthongtham P, Chan KY, Zhu D. CredSaT: Credibility ranking of users in big social data incorporating semantic analysis and temporal factor. *Journal of Information Science*. 2018;45(2):259–280. Available from: https://dx.doi.org/10.1177/0165551518790424.
5) Diaz-Garcia JA, Ruiz MD, Martin-Bautista MJ. NOFACE: A new framework for irrelevant content filtering in social media according to credibility and expertise. *Expert Systems with Applications*. 2022;208:1–15. Available from: https://dx.doi.org/10.1016/j.eswa.2022.118063.
6) Sitaula N, Mohan CK, Grygiel J, Zhou X, Zafarani R. Credibility-Based Fake News Detection. In: Disinformation, Misinformation, and Fake News in Social Media. Lecture Notes in Social Networks. 2020;p. 163–182. Available from: https://doi.org/10.1007/978-3-030-42699-6_9.
7) Cardinale Y, Dongo I, Robayo G, Cabeza D, Aguilera A, Medina S. T-CREo: A Twitter Credibility Analysis Framework. *IEEE Access*. 2021;9:32498–32516. Available from: https://dx.doi.org/10.1109/access.2021.3060623.
8) Azer M, Taha M, Zayed HH, Gadallah M. Credibility Detection on Twitter News Using Machine Learning Approach. *International Journal of Intelligent Systems and Applications*. 2021;13(3):1–10. Available from: https://dx.doi.org/10.5815/ijisa.2021.03.01.
9) Khan T, Michalas A. Seeing and Believing: Evaluating the Trustworthiness of Twitter Users. *IEEE Access*. 2021;9:110505–110516. Available from: https://dx.doi.org/10.1109/access.2021.3098470.
10) Ahmad F, Rizvi SAM. Identification of user's credibility on twitter social networks. *Indonesian Journal of Electrical Engineering and Computer Science*. 2021;24(1):554–563. Available from: https://dx.doi.org/10.11591/ijeecs.v24.i1.pp554-563.
11) De Silva THM. A Network Analysis Based Credibility Ranking Model to Combat Misinformation on Twitter. 2021. Available from: https://dl.ucsc.cmb.ac.lk/jspui/handle/123456789/4528.
12) Li J, Paananen V, Suryanarayana SA, Huusko E, Kuutila M, Mäntylä M, et al. It is an online platform and not the real world, I don't care much: Investigating Twitter Profile Credibility With an Online Machine Learning-Based Tool. In: Proceedings of the 2023 Conference on Human Information Interaction and Retrieval. ACM. 2023;p. 117–127. Available from: https://doi.org/10.1145/3576840.3578308.