I.J. Intelligent Systems and Applications, **2021, 3, 1-10**
Published Online June 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2021.03.01

Credibility Detection on Twitter News Using Machine Learning Approach

Marina Azer
Modern Academy for Computer Science and Management Technology, Computer Science Department, Cairo, 11434, Egypt
E-mail: marina.essam991@gmail.com

Mohamed Taha
Benha University, Faculty of Computers and Artificial intelligence, Computer Science Department, Benha,13518, Egypt
E-mail: mohamed.taha@fci.bu.edu.eg

Hala H. Zayed
Benha University, Faculty of Computers and Artificial intelligence, Computer Science Department, Benha, 13518, Egypt
E-mail: Hala.zayed@fci.bu.edu.eg

Mahmoud Gadallah
Modern Academy for Computer Science and Management Technology, Computer Science Department, Cairo, 11434, Egypt
E-mail: mgadallah1956@gmail.com

Received: 24 January 2021; Revised: 11 February 2021; Accepted: 16 March 2021; Published: 08 June 2021

Abstract: Social media presence is a crucial portion of our life. It is considered one of the most important sources of information than traditional sources. Twitter has become one of the prevalent social sites for exchanging viewpoints and feelings. This work proposes a supervised machine learning system for discovering false news. One of the credibility detection problems is finding new features that are most predictive to better performance classifiers. Both features depending on new content, and features based on the user are used. The features' importance is examined, and their impact on the performance. The reasons for choosing the final feature set using the k-best method are explained. Seven supervised machine learning classifiers are used. They are Naïve Bayes (NB), Support vector machine (SVM), K-nearest neighbors (KNN), Logistic Regression (LR), Random Forest (RF), Maximum entropy (ME), and conditional random forest (CRF). Training and testing models were conducted using the Pheme dataset. The feature's analysis is introduced and compared to the features depending on the content, as the decisive factors in determining the validity. Random forest shows the highest performance while using user-based features only and using a mixture of both types of features; features depending on content and the features based on the user, accuracy (82.2 %) in using user-based features only. We achieved the highest results by using both types of features, utilizing random forest classifier accuracy(83.4%). In contrast, logistic regression was the best as to using features that are based on contents. Performance is measured by different measurements accuracy, precision, recall, and F1_score. We compared our feature set with other studies' features and the impact of our new features. We found that our conclusions exhibit high enhancement concerning discovering and verifying the false news regarding the discovery and verification of false news, comparing it to the current results of how it is developed.

Index Terms: Twitter, Credibility Detection, Machine Learning, Content-Based Features, User-Based Features.

1. Introduction

The Social Networks' platforms are used to exchange viewpoints and news that are now considered indispensable origins of information that mostly outweigh the regular sites. Anyone can make an account regardless of age, education, and many other factors. Also, he/she can post what he\she wants, which gives a chance to create fake accounts and share fake information, which significantly impacts decision-making. Organizations, mainly the political, are exceedingly curious about studying and examining Social Networks' substances to estimate the open conclusion and the individual's fulfillment regarding specific topics in the commerce field. Those sites are rapidly developing, particularly

amidst youthful individuals, upon whom the data from unknown sources have a significant influence. Twitter News depends on distinctive sources mainly based on the public. Twitter becomes an environment suitable for the propagation of hearsays due to the absence of superintendence and control. This issue gets to be an issue because a more significant number of individuals rely on Social Networks for getting information, particularly during crises in Ref. [1].

A recent study Ref. [2] showed that rumors disseminated via Facebook within the final 2017 French presidential election left a significant impact on the electors. Different researches expressed that much of the content on Twitter is not true in Ref. [3-5]. Research by El Ballouli et al. Ref. [4] argued that roughly 40% of Facebook's daily posts are untrue. Moreover, Gupta et al.Ref. [5] shows a considerable propagation of false information amid Hurricane Laura. It manifested that 90% of the hearsays were reposted "Shared." They concluded that individuals are mutually exchanging and disseminating information during an emergency, regardless of its untrusted origin. At present, detecting the trueness and authenticity of the information details on Facebook is a critical issue, particularly concerning events.

The problem here is that the information seeker can't distinguish reliable information from false information, so it is vital to building computerized credibility detection systems. But there are different challenges in the news credibility detection task; the problems are a few available datasets that include features about the tweet to be analyzed. Without using technology, it is hard to determine the true and reasonable posts or the information. It takes time and effort. Some studies use Twitter API Ref. [4, 6- 9, 10-12]. Another challenge is the Limited size of tweet length, using informal language. Using non-English words, noise data, and the most crucial challenge in the rumor detection task is the feature extraction phase and choosing the appropriate function or property determined for developing the classification devices' execution. Different functions are features depending on content features, features depending on users' user, and features focusing on the topic. In this work, we attempt to introduce a new function or property that enhances classifiers' performance. The reasons for choosing these features are explained. They affect the performance more than others. We used the dataset, which is the most used dataset in Rumor Detection.

This document tackles a model for converting "rumor information" to be automatically discovered via Facebook. Various artificial intelligence learning methods are used with miscellaneous categories of features. These features focus on content as (the post length, number of sharing, and likes). Suppose the posts include emotions and user\source based features as (if the user verified or not, has a description on his/ her page …) and a combination set of them. Our model is based on particularly 39 features (22 content features, 17 user features). The performance of seven different supervised classifiers: Random Forests (RF), Support Vector Machines (SVM), Logistic Regression (LR), Naïve Bayes (NB), and K-Nearest Neighbor (KNN), max entropy (ME), and conditional random forest (CRF) was contrasted. The suggested pattern or example accomplished an accuracy percent of 83.4% in foreseeing the validity of tweet messages using the Random Forest classifier while combining the content-based and source-based features. The proposed model achieves the following two contributions: 1) introducing new valuable features. 2) Contrasting various artificial intelligence learning methods and discuss the results of each algorithm with each feature sets. The rest of the paper is organized as follows; Section 2 provides some studies focusing on Evacuating Reasonability through applying various methods. Next, Section 3 presents the phases of the suggested pattern. Section 4 demonstrates the results and discusses them. Finally, Section 5 concludes the paper.

2. Related Work

A significant part of the study indicating at deciding the validity of Facebook messages is based on classification. These methods classify posts as true and false utilizing supervised machine learning methods Ref. [13-19]. An essential fact that includes several clarified posts with the information concerning them is utilized to construct mechanical classification devices that can precisely decide a particular post's validity. The trueness of the notes is a vital element influencing the effectiveness of the expectation. The consistency of the discovered information is an additional vital element. Certain studies focus on the post's substance Ref. [16]. However, other research focuses on the original composer of the post Ref. [20]. In this regard, several studies mostly related to this issue are reviewed.

Castillo et al. were considered the primary to make studies on Facebook validity checking Ref. [13, 14]. They use Tweets concerning the most shared themes and proposed a controlled automatic learning pattern to demonstrate to anticipate the validity. They employ different sorts of highlights, a part of which concerns the post's substance, but another part is related to the composer of the post or collected from the relevant topic. The phase of identifying included two phases: the initial phase collects posts that exhibit contents on news details (entitled as Stories or News) through person conclusions (Entitled as Messenger). Another phase centers on the posts entitled NEWS, classifying it as true/false. It uses multiple classifiers such as SVM, decision trees, decision rules, and Bayesian networks on the noted information. However, better execution was performed by J48, decision tree. Another work in this category is the one that appeared on Ref. [21]; the authors suggested a hearsay discovery pattern depends on a consecutive classification device, in which the post is classified a true or false, according to the trained information.

Many of the existing studies consider hearsay's detection also to endure from another topic: they expect that hearsays are persistently wrong, aiming at foreseeing alludes those fake hearsays Ref. [22]. This is often illustrated through the plan of their tests, in which they prepare their discovering models on networks of permanent hearsays, aiming at identifying wrong hearsays. This suggestion is untrue and inapplicable since hearsays are probably true. The

word 'rumor' means unsubstantiated data, which may be considered a while and subsequently becomes true or wrong. They deduct hearsays, with no attention to whether it is true or not. The objective is to mark the very tiny and small posts as hearsays., i.e., very tiny and small posts that include unconfirmed data the quick dissemination, and hence limiting their dangerous effects to come about.

Using a neural network (Lstm) and traditional classification algorithms as SVM and their prediction results to achieve two tasks. Task A is classified posts into (supporting, denying, querying, or commenting) and task B, which determines the veracity of rumors if it is true, wrong, or unverified. In task A f -measure was 0.6, and in task B unverified category has low precision and low f value, which drag down the average f value of the three categories Ref. [23].

Here tweet is collected by tweet crawler Ref. [24]. They focus on identifying fake news and fake accounts on Twitter by using their algorithms which after receiving a Tweet, that user wants to measure its credibility through three phases which take the tweet link, then NER (named entity recognition) get nouns, the topics, social tags, overall sentiment and compare with similar tweets from trust sources in their database to calculate tweet score. They are still developing and trying to improve the quality of the performance.

In Ref. [25] used a model for recognizing fake news; the model was based on using word embedding methods which are glove, fast text, then used training deep learning models which are (RNN), (GRU), and (LSTM) methods with the flair library, they achieved accuracy up to 99.8%.

In Ref. [26] they proposed to show that utilizing entropy-based include a determination on the dataset, employing a pile-gathering type of triple techniques to extend the discovery exactness, they created the suggested rating demonstrate features a way improved discovery percent decreases the fake percent of information occasions and hence identifies the false information precisely.

The work in Ref. [27] focuses on building a new artificial intelligence learning pattern, depending on (NLP) characteristics for discovering false information by utilizing features centered on content and those features of social networking depending on the information. The suggested pattern demonstrates considerable conclusions, giving an average accuracy of 85.20% has appeared

Reference [28] proposed a false page identifier confirming the pages' personality to detect the false pages, mainly using standard terms and defining limited robots. They used three datasets Whatsapp, Instagram, and Facebook. The conclusions discovered increased percentages of accuracy, sharing, and decreased positive percentages of discovering false pages in the previous social networks.

The authors used Twitter data to analyze Abusive Tweets in Ref. [29] based on tweets' contextual and lexical features. The validity of posts was suggested by allotting a number or percent to the content on Facebook for demonstrating its dependability. A comparative think about different rating strategies to bolster adaptability was executed, and a new outlet to the confinements display of now applied methods was discovered.

This work Ref.[30] examined and analyzed customer's online opinions for credibility detection using different machine learning techniques to choose the suitable functions in functions choice set specific rules. Using normal terms was provided. Tests on hotel review databases exhibit the adequacy of the suggested method.

Authors of Ref. [31] applied several mixed features focusing on both the content and the user with supervised machine learning classifiers and found that the combination of features with a random forest classifier achieved the highest accuracy.

Reference[32] The authors applied the features centered on the content and social networking features, with machine learning classifiers: SVM, Random forest, naïve Bayes, MaxEnt, CRF, and found CRF achieved the highest precision and F1-score.

3. Methodology

The proposed model of fake news detection is shown in Figure 1. The model consists of four modules 1.Data collection, 2.Preprocessing, 3.Feature Extraction, 4.Training, Assessing, and testing the model. The execution of each strategy is assessed by measuring accuracy, precision, recall, and F1-Score. Further details on the steps are provided in the subsections below:

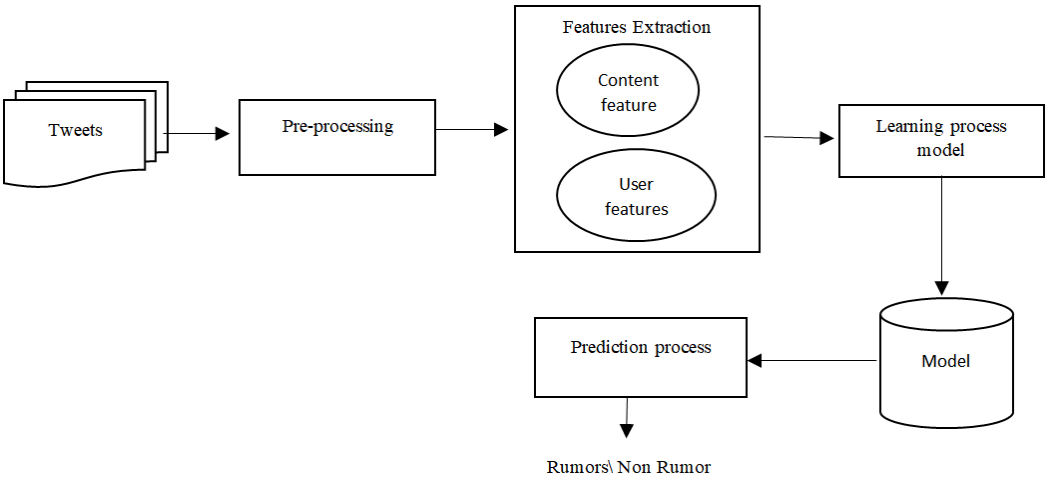


Fig.1. The proposed Model

3.1. Data Collection

There are a few available datasets which is one of the challenges in credibility detection topic. Experiments were conducted using the Pheme dataset is the most commonly used dataset in the credibility detection task. The dataset was collected using Twitter streaming API during five breaking news. Charlie Hebdo :458 rumors , 1,621 non-rumors, Ferguson :284 rumors , 859 non-rumors, German wings crash : 238 rumors , 231 non-rumors ,Ottawa shooting :470 rumors , 420 non-rumors, Sydney siege : 522 rumors,699 non-rumors. The total number of tweets is 5.802 was manually annotated to 3830 (66%) credible and 1792 (34%) non-credible tweets [21].

Regarding data splitting in this step, Regarding the division of information in this step, the dataset is divided into 80% for training, 10% for validation, and 10% for testing the dataset, applying a classified technique. The divided part of training is inserted into ML/DL Models. For making the models, we have benefited from this information and applying the invisible test for assessment.

3.2. Data preprocessing

Data preprocessing is a crucial phase, particularly for social media substance. Twitter data is the well-known unstructured datasets collected of information from individuals entered his/her sentiments, opinion, attitudes, products review, emotions, etc. These datasets need to be subjected to certain refinements by performing preprocessing strategies to the following stages. The essential cleaning operations within preprocessing strategies used in this work are removing unimportant characters, stop-word removal, tokenization, a lower casing, remove repeated letters, auto-correct spelling, and stemming. They will offer us assistance to decrease the size of actual data by evacuating insignificant data. After that, to attain better execution, the preprocessing includes the arrangement of procedures which are listed in the following steps:

- Evacuating insignificant: the punctuation marks as commas, apostrophes, quotes, question marks, and more, which don't include much esteem to the show, are erased.
- Stop Words Removal: a stop word ordinarily refers to the most common words in a language that does not include much meaning to a sentence. These words are expelled from each tweet with the datasets
- Removing non-English words: sometimes, people in social media use non-English words which are deleted from the sentence.
- Remove Repeated Letters and AutoCorrect spelling: repeated letters change the meaning of the word as sooo, haaaaapy should be edited to so, happy.
- Lowercasing: simply it is one of the basic cleaning operations to convert a word to lower cases such as
- Tokenization: It is the key perspective of working with substance data to confine a bit of substance into litter units called tokens. The tokens are tallying segments and sentences, which can be assist broken into words.
- Stemming: stemming is ousting the expansion from and alter it to its root word to decrease the number of word types or classes within the information. For illustration, the words "Making," "Made," and "Maker" will be decreased to the word "make.

We didn't use pos-tagging because we found that pos-tagging doesn't significantly affect the model generation process's accuracy.

3.3. Feature Extraction

Different features (non-lexical, semantic, and stylistic) are tested and examined for their effect on classification

accuracy. Some of the features are calculated as followers to friends rate, the number of statuses to account age rate, the number of friends and followers to account age rate, and more, demonstrating the source's ubiquity. Other information is extracted from the author's past tweet posts, such as the average number of URLs and retweet fraction. The feature set in Table 3 is chosen using the k-best method, which achieved good results, where k is the number of features, and there is an inbuilt class (feature importance and matrix correlation).

We used 39 features; 16 features are new features (12 new content-based features and 4 new user-based features) were combined with other prominent features. Table 3 shows the overall used feature set. New features are represented in table 3 with *, which are not considered in other works Ref. [31, 32] and enhance the classifiers' performance.

Some features didn't enhance the performance that was tested while choosing the appropriate features as (number of replies and the source of the tweet if it is from mobile or web and pos tagging,) in content-based features, and user-based features (user location if the user is near to event or not and also if the user followed by credible users didn't enhance the results because credible users tend to follow few people.

A. Content-based features

Content-based features are the features that focus on the substance of the tweet. The description of the new content-based features is explained in Table 1.

Table 1. The content-based features description

Content-based features	Description
1-Number of positive words.	It means the number of words that represent the positive feeling.
2-Number of negative words.	It means the number of words that represent the negative feeling.
3-Has positive emotions.	It means if the tweet has a positive emotion or not.
4-Has negative emotions.	It means if the tweet has a negative emotion or not.
5-The overall tweet sentiment.	The tweet's overall sentiment, if it is positive, negative, or neutral, to find if it matches the topic's sentiment or not.
6-The overall sentiment of replies	The overall sentiment of replies on the tweet if it matches the sentiment of the tweet or not.
7-The number of questioned comments.	It means the number of questioned comments in the replies of the tweet.
8-Time span	It means the interval time between account creation and the time of posting the tweet.
9-Trusted URLs.	The URLs that appear in non- rumors tweets are classified as trusted URLs to be a factor of the new tweets' credibility process.
10- Have a word 'pray'?	It means if the Tweet has the word 'pray' or not.
11-Have a period?	It means if the Tweet includes a period.
12-Ratio of punctuation marks to words	It is calculated as the number of punctuation marks/number of words per tweet.

In "the number of positive and negative words, the number of positive and negative emotions" features, we found that some works ignored the type of words and just counted the number of words in the Tweet. In contrast, The type of words is an important feature to consider. We notice that the dataset includes five breaking real news that has more negative tweets, sad and angry emotions than in rumors tweets, so the number of positive words and negative words as features is considered. The same in the emotions if the tweet has a sad or happy emotion is not just the number of emotions regardless of the type .and calculated "The overall sentiment of the tweet "to find if it matches the topic's sentiment. The tweet replies are essential to consider; it gives feedback about the tweet. We calculated "The sentiment of the replies" and checked if it matches the sentiment of the tweet or not. Check "the number of questioned replies" on the tweet, reflecting increases in the probability of rumor content. The interval time between account creation and the time of posting the tweet. About' time span' this feature is important because a lot of people create fake accounts, especially for posting some rumored news during emergencies.' trusted URLs a lot of tweets in the dataset have URLs, here we extract URLs used in non-rumors tweets in each event to make them trusted URLs when they appear in test data. It enhances the credibility detection performance. 'word 'pray' Because of the nature of the events included in the dataset. There are dead and injured people in each event, found that the word 'pray' is repeated in non-rumor Tweets, so it also used as a good content-based feature.' Have a period?' If the tweet includes period or not to enhance the credibility and 'Ratio of punctuation marks! To the number of words in the tweet as a sign of non-credibility.

B. User_ based features

User-based features focus on the characteristics of the author (the source) of the tweet, the description of the new user-based features is shown in Table 2.

Table 2. The description of the new user-based features

Content-based features	User-based features
1- **Tweet overall sentiment***	
2-Number of negative words*	
3-Number of positive words*	
4-Tweet has positive emotions*	**1- Number of followers to account age rate***
5-Tweet has negative emotions*	**2- Number of friends to account age rate***
6-Time span*	**3-Rate of statuses to account age***
7- Has word 'pray'*	**4-favorites count***
8-Number of questioned comments*	5-followers\friends ratio.
9-Trusted URLs*	6-listed count.
10-Ratio of punctuation marks to the number of words*	7-Has a description?
	8-Length of description.
11-Replies sentiment*	9-Length of the screen name.
12-Include period*	10-User has URL?
13-number of words.	11- Is verified account?
14-The tweet has a URL?	12-Has a default profile picture?
15-The tweet has hashtags?	13-Average number of hashtags
16- Hashtags count.	14- Average number of URLs
17-has question marks.	15-Average number of mentions.
18-question mark count.	16-Average tweet length.
19-has exclamation marks.	17-Retweet fraction.
20-exclamation mark count.	
21-is retweeted?	
22-retweet count	

Here we calculated the account age and checked the number of followers, friends, and statuses to the account age to check if it a fake account or not. And the favorites count to measure the activity of the account, and which may be a sign to consider it is a fake account or not.

Table 3. The selected Feature set

User-based features	Description
1-Ratio number of the followers to the account age.	calculated as the number of followers/account age
2-Ratio number of the friends to the account age.	calculated as the number of friends/account age
3-Ratio number of the statuses to the account age.	calculated as the number of statuses/account age
4-favorites count.	The number of tweets this user has liked in the account's lifetime.

3.4. Applying Learning Models

We applied seven traditional machine learning classifiers and compared the performance of the models. The models are support vector machine (SVM), random forest (RF), logistic regression (LR), Naïve Bayes (NB), Maximum Entropy (ME), Conditional Random Forest (CRF), and k nearest neighbors (KNN) to find the best one. The algorithms tested with different feature sets, the user-based features, the content-based features, and the combination of them in separated experiments as shown in table 4, table 5, and table 6.

4. Results and Discussion

Our experiments show that choosing the features is a significant factor in credibility detection, and not all features are essential. Some of the features do not affect the performance of the classifiers. After extracting new content and user features. We found an improvement in results than other studies. In figure 2, we compare the accuracy rates between machine learning classifiers in the three experiments and found that user-based features are more capable than content-based features. It is proved to be true after found that the most important features are user-based features as followers count, listed count, verified account and using default profile image, then in the content- features category Retweet count, the overall tweet sentiment, trusted URLs have more effect on the performance of classifiers than others. In the user features category, the Ratio number of the followers to account age, ratio number of the statuses to account age, retweet fraction, and tweet-length have more effect than others.

After using seven different supervised machine learning algorithms with content-based features, user-based features, and a combination of them, we evaluate the results to find the best performance. In Table 4, the algorithms performance while using our content-based features only versus the content features of Ref. [31], Table 5 the performance while using our user-based features only versus user features of Ref. [31] and Table 6 shows the performance of algorithms while using our combination set of content and user features versus the set of Ref. [31]. Random Forests accomplishes higher accuracy rates while using both user-based features and while using combined feature sets, while Logistic Regression is the best classifier while using content-based features. The performance while using the combination of content and user features was better than using content features only, or user features only.

And better than the performance than Ref. [31] feature set. Figure 2 shows the accuracy rates between machine learning classifiers with different feature sets, content-based features only, user-based features only, and a combination of them.

The performance of the algorithms is measured by different measurements, which are:

- **Accuracy:** a degree of completely accurately distinguished tests out of all the tests [33] .accuracy calculated as appeared in Equation (1).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} * 100 \tag{1}$$

- **Precision and Recall:** a degree of the ability of the model to precisely distinguish the existence of a positive class instance is decided by recall [34, 35].precision is shown in Equation (2), and Recall is calculated as appeared in Equation (3).

$$\text{precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

- F1-Score: The consonant cruel of Precision and Recall [36]. F1_score is appeared in Equation (4).

$$F1\_Score = \frac{2 * precision * \text{Re}call}{precision + \text{Re}call} \tag{4}$$

Table 4. The performance using our content-based features versus Ref [31] content features

classifier	Proposed work				Ref [31]			
	Accuracy	Precision	Recall	F1_score	Accuracy	Precision	Recall	F1_score
Random forest	67.7	71	82.3	79.6	61.6	69	76.2	72.4
KNN	69.1	70.2	85.2	80.3	62.1	68.4	79.2	73.4
SVM	70.8	71.1	96.8	83	66.5	67.9	93.6	78.7
Logistic regression	**73.2**	**70.8**	**96.3**	**82.4**	67.1	67.9	93.1	78.5
Naïve Bayes	71.2	70.2	99.7	83.4	66	66.1	99	79.2

Table 5. The performance using user-based features versus Ref [31] user features

classifier	Proposed work				Ref [31]			
	Accuracy	Precision	Recall	F1_score	Accuracy	Precision	Recall	F1_score
Random forest	**82.2**	**82 .6**	**92.4**	**85.2**	77.8	79.5	88.7	83.8
KNN	73.8	79.5	86.7	77.63	70.9	75.7	82.6	78.9
SVM	71.4	69.7	99.6	80.5	66	66.9	96.4	78.9
Logistic regression	69.1	69.4	97.3	80.6	66.1	67.2	95.1	78.7
Naïve Bayes	71.6	69.5	99.8	81.9	66.4	66.6	99	79.6

Table 6. The performance while using our overall feature set versus Ref [31] feature set

classifier	Proposed work				Ref [31]			
	Accuracy	Precision	Recall	F1_score	Accuracy	Precision	Recall	F1_score
Random forest	**83.4**	**83.1**	**95.6**	**88.6**	78.4	79.6	91.6	85.2
KNN	71.3	75.3	84.5	82.3	66.2	72.5	78.9	75.5
SVM	73.5	71.2	93.4	81.3	67	68.7	91.9	78.6
Logistic regression	72.9	70.3	93	80	66.9	68.8	91.2	78.4
Naïve Bayes	72.8	71	96	82	66.7	67.7	94.8	79

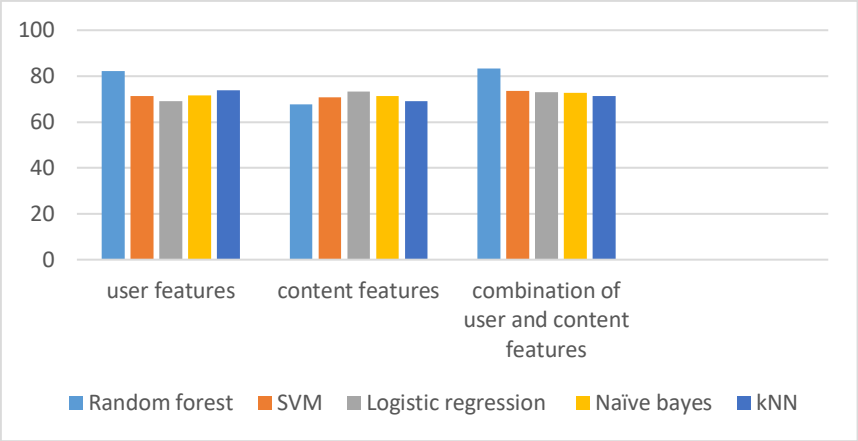


Fig.2. A comparison between accuracy rates of selected classifiers with different feature sets

Table 7. Shows the results of our work comparing with the results of using Ref. [32] features. Found that we get better results while using our feature set with random forest classifier than using Ref. [32] features with the same classifier and another discussed classifiers.

Table 7. The results of using our feature set versus using features of Ref [32]

classifier	Proposed work			Ref [31]		
	Precision	Recall	F1_score	Precision	Recall	F1_score
SVM	.712	.934	.813	0.337	0.483	0.397
Random forest	**.831**	**.956**	**.886**	0.275	0.099	0.145
Naïve Bayes	.71	.96	.82	0.310	0.723	0.434
Max ENT	.62	.69	.60	0.338	0.442	0.383
CRF	.80	.67	.70	0.667	0.556	0.607

5. Conclusion and Future Work

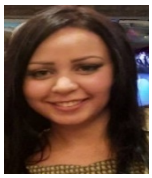
A supervised machine learning framework for untrue news confirmation based on utilizing new content and user-based features is proposed. The preprocessing stage incorporates detailed sentence analysis beginning from evacuating insignificant characters till tokenization and stemming. Pheme dataset was utilized for training and testing models by part the information 80% training, 10% validation, and 10% testing. The feature extraction stage includes extracting not just new features but also features that have an impact on the execution of the classifiers. We examined the importance of features and which have more impact than others and which haven't impact. We chose our feature set using the k-best method, which has an inbuilt class (feature importance and matrix correlation). We analyze using substance-based features only, source-based highlights only, and while using a combination of content-based and user-based features. And we found that user-based features have an impact on the performance more than content-based features. After comparing each feature set, we found that using a combination of content-based features and user-based with applying the Random Forests classifier achieved the best results. We outperformed the Ref. [31, 32] approach in terms of accuracy, precision, recall, and F1- score. For future work, in this paper, the content features are utilized within the binary classification. We expected to use a combination of substance, temporal, and context features to be used in multi-class classification in the future.

References

[1] Sitaula, Niraj, et al. "Credibility-based fake news detection." Disinformation, Misinformation, and Fake News in Social Media. Springer, Cham, 2020. 163-182.
[2] H. Allcott, and M. Gentzkow. "Social media and fake news in the 2016 election." Journal of economic perspectives Vol. 31, No. 2, pp. 211- 36, 2017.
[3] Z. Ashktorab, C. Brown, M. Nandi, and A. Culotta. "Tweedr: Mining twitter to inform disaster response." In ISCRAM, 2014.
[4] R. El Ballouli, W. El-Hajj, A. Ghandour, S. Elbassuoni, H. Hajj, and K. Shaban. "CAT: Credibility Analysis of Arabic Content on Twitter." In Proceedings of the Third Arabic Natural Language Processing Workshop, pp. 62-71, 2017. Nb
[5] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi. "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy." In Proceedings of the 22nd international conference on World Wide Web, pp. 729-736. ACM, 2013.
[6] C. Castillo, M. Mendoza, and B. Poblete. "Information credibility on twitter." In Proceedings of the 20th international conference on World wide web, pp. 675-684. ACM, 2011.
[7] C. Castillo, M. Mendoza, and B. Poblete. "Predicting information credibility in timesensitive social media." Internet Research

Vol. 23, No. 5, pp. 560-588, 2013.
[8] A. Gupta, and P. Kumaraguru. "Credibility ranking of tweets during high impact events." In Proceedings of the 1st workshop on privacy and security in online social media, p. 2, ACM, 2012.
[9] K. Lorek, J. Suehiro-Wiciński, M. JankowskiLorek, and A. Gupta. "Automated credibility assessment on Twitter." Computer Science Vol. 16, No. 2, pp. 157-168, 2015.
[10] A. Zubiaga, M. Liakata, and R. Procter. "Exploiting context for rumour detection in social media." In International Conference on Social Informatics, pp. 109-123. Springer, Cham, 2017.
[11] N. Hassan, W. Gomaa, G. Khoriba, and M. Haggag. "Supervised Learning Approach for Twitter Credibility Detection." In 2018 13th International Conference on Computer Engineering and Systems (ICCES), pp. 196-201. IEEE, 2018.
[12] S. Sabbeh, and S. Baatwah. "Arabic news credibility on twitter: an enhanced model using hybrid features.", journal of theoretical & applied information technology Vol. 96, No. 8, 2018.
[13] Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter." Proceedings of the 20th international conference on World wide web. ACM, 2011.
[14] Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Predicting information credibility in time-sensitive social media." Internet Research 23.5 (2013): 560-588.
[15] Gupta, Aditi, and Ponnurangam Kumaraguru. "Credibility ranking of tweets during high impact events." Proceedings of the 1st workshop on privacy and security in online social media. ACM, 2012.
[16] Zubiaga, Arkaitz, Maria Liakata, and Rob Procter. "Learning reporting dynamics during breaking news for rumour detection in social media." arXiv preprint arXiv:1610.07363 (2016).
[17] Gupta, Aditi, and Ponnurangam Kumaraguru. "Credibility ranking of tweets during high impact events." Proceedings of the 1st workshop on privacy and security in online social media. ACM, 2012.
[18] Lorek, Krzysztof, et al. "Automated credibility assessment on Twitter." Computer Science 16.2) (2015): 157-168.
[19] El Ballouli, Rim, et al. "CAT: Credibility Analysis of Arabic Content on Twitter." WANLP 2017 (co-located with EACL 2017) (2017): 62.
[20] Alrubaian, Majed, et al. "Reputation'based credibility analysis of Twitter social network users." Concurrency and Computation: Practice and Experience 29.7 (2017): e3873.
[21] Zubiaga, Arkaitz, et al. "Pheme dataset of rumours and non-rumours." Figshare. Dataset (2016).
[22] Zubiaga, Arkaitz, et al. "Detection and resolution of rumours in social media: A survey." ACM Computing Surveys (CSUR) 51.2 (2018): 1-36.
[23] Sedhai, Surendra, and Aixin Sun. "Semi-supervised spam detection in Twitter stream." IEEE Transactions on Computational Social Systems 5.1 (2017): 169-175.
[24] Sato, Koichi, Junbo Wang, and Zixue Cheng. "Credibility Evaluation of Twitter-Based Event Detection by a Mixing Analysis of Heterogeneous Data." IEEE Access 7 (2018): 1095-1106.
[25] Kula, Sebastian, et al. "Sentiment analysis for fake news detection by means of neural networks." International Conference on Computational Science. Springer, Cham, 2020.
[26] Akinyemi, Bodunde, Oluwakemi Adewusi, and Adedoyin Oyebade. "An Improved Classification Model for Fake News Detection in Social Media.", international journal of Information Technology and Computer Science, Vol.12, No.1, PP.34-43, 2020.
[27] Shubham Bauskar, Vijay Badole, Prajal Jain, Meenu Chawla "Natural language processing based hybrid model for detecting fake news using content-based features and social features." International Journal of information Engineering and Electronic Business, Vol.11, No.1, PP.1-10, 2019.
[28] Ali M. Meligy, Hani M. Ibrahim, Mohamed F. Torky" Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks." International Journal of Computer Network and Information Security, Vol.7, No.1, PP.31-39, 2014.
[29] Priya Gupta, Aditi Kamra, Richa Thakral, Mayank Aggarwal, Sohail Bhatti, Vishal Jain "A Proposed Framework to Analyze Abusive Tweets on the Social Networks." International Journal of Modern Education & Computer Science, Vol. 10, No. 1, PP.46-56, 2018.
[30] Naznin Sultana, Sellappan Palaniappan, "Deceptive Opinion Detection Using Machine Learning Techniques", International Journal of Information Engineering and Electronic Business, Vol.12, No.1, pp. 1-7, 2020.
[31] Hassan, Noha Y., et al. "Supervised learning approach for twitter credibility detection." 2018 13th International Conference on Computer Engineering and Systems (ICCES). IEEE, 2018.
[32] Zubiaga, Arkaitz, Maria Liakata, and Rob Procter. "Exploiting context for rumour detection in social media." International Conference on Social Informatics. Springer, Cham, 2017.
[33] Shu, Kai, et al. "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media." Big Data 8.3 (2020): 171-188.
[34] Shahi, Gautam Kishore, and Durgesh Nandini. "FakeCovid--A Multilingual Cross-domain Fact Check News Dataset for COVID-19." arXiv preprint arXiv:2006.11343 (2020).
[35] Zhou, Xinyi, et al. "Recovery: A multimodal repository for covid-19 news credibility research." Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.
[36] Memon, Shahan Ali, and Kathleen M. Carley. "Characterizing covid-19 misinformation communities using a novel twitter dataset." arXiv preprint arXiv:2008.00791 (2020).

Authors' Profiles



Marina Azer is currently a Lecturer Assistant in the computer science department, Modern Academy for Computer Science and Management Technology, Egypt, since 2012. She achieved her master's degree from Helwan University. She has worked on several research topics.

Mohamed Taha is an Assistant Professor at Benha University, Faculty of Computers and Artificial intelligence, Computer Science Department, Egypt. He received his M.Sc. degree and his Ph.D. degree in computer science at Ain Shams University, Egypt, in February 2009 and July 2015. His research interest's concern: Computer Vision (Object Tracking-Video Surveillance Systems), Digital Forensics (Image Forgery Detection – Document Forgery Detection - Fake Currency Detection), Image Processing (OCR), Computer Network (Routing Protocols - Security), Augmented Reality, Cloud Computing, and Data Mining (Association Rules Mining-Knowledge Discovery). Taha has contributed more than 20+ technical papers to international journals and conferences.

Hala H. Zayed received the B.Sc. in electrical engineering (with honor degree) in 1985, the M.Sc. in 1989, and Ph.D. in 1995 from Benha University in electronics engineering. She is now a professor at the faculty of Computers and Artificial intelligence, Benha University. Her areas of research are computer vision, biometrics, machine learning, and image processing.

Mahmoud Gadallah received the B.Sc. in electrical engineering in 1979, the M.Sc. in 1984 Faculty of Engineering, Cairo University, and Ph.D. in 1991 from Cranfield Institute of Technology (Cranfield University now), United Kingdom. He is now a professor at the modern academy for computer science and Management Technology, Cairo, Egypt. He has worked on several research topics as image processing, pattern recognition, computer vision, and natural language processing.

How to cite this paper: Marina Azer, Mohamed Taha, Hala H. Zayed, Mahmoud Gadallah, "Credibility Detection on Twitter News Using Machine Learning Approach", International Journal of Intelligent Systems and Applications(IJISA), Vol.13, No.3, pp.1-10, 2021. DOI: 10.5815/ijisa.2021.03.01