# Fake News Detection on Social Media using Deep learning and Semantic Knowledge Sources

Journal of Computer Science IJCSIS, Rasha Al Bashaireh

*IJCSIS Vol 18 No. 2 February 2020 Issue*

# Fake News Detection on Social Media using Deep learning and Semantic Knowledge Sources

Vian Sabeeh
Computer Science and Engineering Department
Oakland University
Rochester,MI,USA
e-mail: viansabeeh@oakland.edu

Mohammed Zohdy
Electrical and Computer Engineering Department
Oakland University
Rochester,MI,USA
e-mail: zohdyma@oakland.edu

Atiqul Mollah
Computer Science and Engineering Department
Oakland University
Rochester,MI,USA
e-mail: mullah@oakland.edu

Rasha Al Bashaireh
Computer Science and Engineering Department
Oakland University
Rochester,MI,USA
e-mail: ralbashaireh@oakland.edu

***Abstract* - The rapid diffusion of false information on social networks emphasizes the need for automatic fake news detection systems. The majority of current fake news detection approaches are either based on news content analysis or social context features. To improve the identification of fake news, it is necessary to explore supplemental information that includes the interaction between the user and the news. Also, some of the existing works adopt knowledge-based approaches for fake news detection, which exploit external sources for facilitating fact-checking. However, the entire content on social networks is not all trustworthy for cross-referencing using external sources that include malicious accounts or news for diffusing fake news. Thus, credibility analysis is essential to verify the trustworthiness of news to improve the detection accuracy. The comments of the users on social networks are the most reliable signals of the user intent, and at the same time, the fake reviews are written by the illegitimate users who seek to mislead the people to make erroneous decisions. To highlight these constraints, this work presents a model referred to as detecting fake news on Social media Platforms through Opinion mining and Trustworthiness of user and event (SPOT) that incorporates opinion mining on user comment, and credibility analysis of Twitter metadata. The proposed method exploits the knowledge source during feature extraction to gain knowledge of a certain event for improving the performance. The proposed method employs SentiWordNet to consider the cognitive cues of the text to facilitate opinion mining. It further employs a bidirectional Gated Recurrent Neural Network (GRNN) incorporating objective factors, such as sentiment and a credibility score to provide efficient decisions for fake news detection. The experimental results reveal that the proposed SPOT approach significantly outperforms the existing DUAL approach with 14.15% higher accuracy.**

***Keywords: social network; fake news; deep learning; credibility; semantic knowledge source; sentiment analysis; and decision making.***

## I. INTRODUCTION

Nowadays, social media have served as significant news sources for people over the world due to its low cost, fast dissemination, and easy access. It has massively disseminated news information at anytime and anywhere in the world through the development of virtual networks [1]. Owing to the explosive growth of social networks over the past decade, an unprecedented scale of information is now available online. As a consequence, social media networks have become a significant alternative online information source for traditional media during critical events. The identification of trustworthy news sources becomes complex owing to the growth of misleading information. The convenient features and the openness of social media also foster fake news, which deliberately conveys misinformation to the public for different purposes, including political or financial manipulation [2]. Fake news is extensively disseminated to the users without the confirmation that causes severe threats to society and also makes it nontrivial to discover [3,4]. The fake news on social media has the competence to impel people to distrust society. Notably, the presence of fake news in social media platforms leads to break the authenticity and balance of the news system, social trust, and political polarization [5]. On the other hand, with the steadily increasing diverse content, automatic fake news detection has become a crucial challenge [6]. To cope with this deceptive information, the users progressively demand efficient fake news detection technology [7].

A significant amount of scientific methods has been proposed to suppress the rumors automatically online by applying machine learning techniques [8]. The fake news detection mechanism uses various features for combating fake news in the social media that includes news content, user information, and propagation of all social posts [9]. The machine learning-based conventional methods for fake news detection leverages handcrafted features and excludes the various roles of real post and retweets in the production of information and diffusing process, which makes the existing detection inefficient. Also, as the scale of data grows, the machine learning model seems to fall behind the deep learning model. The deep learning is an endeavor to learn high-level feature representations through the hidden layers from large-scale data for providing accurate results. This representation is a distinctive part of the deep learning method and a significant step ahead of machine learning methods. Recently, the surge of analyzing the fake news on the social networks has targeted to resolve the critical problems using the deep learning together with the web services and emerging data sets to discover the deceiving content [10]. Therefore, most of the earlier work heavily relies on the deep learning model. However, even with the deep learning algorithm, it is incredibly challenging to successfully discover whether the information of the news is real or fake while getting in with the fake news detection. For example, when people adopt a decision, they cross-refer these user comments, in which the dependence of the comments helps to identify the fake news. Furthermore, the analysis of user comment helps people to procure the correct information.

Even with the characteristics of sophisticated feature extraction, deep learning based fake news detection is still a challenge, because the news content is designed to resemble the fact to deceive readers; and without the auxiliary information or fact-checking, it is often difficult to identify the veracity. To cope with this constraint, a sentiment analysis method for capturing the sentiment orientations from short text is essential. However, it is a challenging task to be done and induces the confirmation biases during detection by sentiment analysis alone. Therefore, it is essential to extract the subjectivity from the natural language news context based on the corpus and knowledge source for precise decision making. Despite the rich opinionated information available on Twitter regarding the specific news available on Twitter, not every source is trustworthy or helpful in delivering reliable information regarding the event. Assessing the trustworthiness of news is an arduous and cumbersome task when the fake news is intentionally written to mislead the reader. The credibility of the user and tweet often has a vital role in real-world events, as rumors and fake news also spread with real news. Thus, the proposed SPOT method presents a deep


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

learning based fake news detection model that employs both the sentiment and credibility analysis for effectively identifying the fake news on Twitter. Also, it utilizes a semantic knowledge source for identifying the most relevant and potential features of Twitter and getting aware of events for detecting fake news.

*A. Scope & Motivation*

The unprecedented existence of rich information on Online Social Networks (OSNs) behaves like a double-edged sword that imparts an opportunity for malicious accounts to manipulate the consequences of public events and alter the characteristics of the people. For instance, in 2016, the US election was severely affected by the circulation of false news largely through social media [11]. The majority of people appear to believe fake claims, which swirled around the social sites before the 2016 presidential election and broadly supported Donald Trump over Hillary Clinton [12]. Even though such fake claims did not modify the election's results, its prevalence is yet the main concern. Fake information promotes misperceptions over the voters and brings about distrust of legitimate news. In this context, it induced severe risk to American democracy. Also, these kinds of activities haven't just declined user experience, but also violate the terms of conditions for social media sites. Identification of such deceptive information about the event has been a more important, yet quite challenging task that is still difficult to understand. Even for a human, it is hard to distinguish truthful information from misinformation without detail. To address these concerns, the development of automated fake news detection is significant.

The main intention of the fake news detection system is to leverage various machine learning and natural language processing techniques for combating fake news on social media platforms. It brings up public awareness to discern between falsehood and legitimate news. Moreover, the fake news detection model helps to ensure the people receive authentic information and retain trustworthiness in the news ecosystem. The proposed fake news detection method identifies the implausible content and circumvents the dissemination of malicious information in the early stages on Twitter. Moreover, it detects the fake news on newly arrived events on social media networks.

*B. Contributions*

The main contributions of the proposed fake news detection model are as follows:

- The proposed SPOT method ensures effective fake news detection on Twitter by utilizing the Bi-directional Gated Recurrent Neural Network (GRNN) deep learning model.
- The proposed method exploits the semantic knowledge sources such as Wordnet, and Wikipedia during the feature extraction for retrieving the vocabulary enriched potential features and acquiring knowledge of the event that facilitates fake news detection.
- The SPOT approach performs the opinion mining on Twitter comments for effectively capturing the opinion of the users from their natural language users' content in which the proposed system exploits the Sentiwordnet corpus to discover the different sentiment polarity from the natural language text.
- Moreover, the proposed SPOT method computes the credibility score of the user, tweet, and event through the retrieved Twitter metadata for ensuring the efficiency of the fake news detection system.
- The experimental framework shows the superior performance of the proposed approach than the existing approach.


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The research work is structured as follows: Section II reviews the previous works related to fake news detection in social media. Section III explains the problem statement and its formulation. Section IV describes the working procedure of the proposed method. Section V explains the experimental evaluation to examine the performance of the proposed method. Finally, Section VI presents the conclusion of the SPOT approach.

## II. RELATED WORK

At the beginning of the decade, the scholarly community began to make efforts for the issue of "misinformation" [13], then "rumor" identification [14], and currently fake news detection and prevention [15]. The automatic fake news detection model explores the heterogeneous data, including social network structure, user demography and activity, content, and news propagation for preventing the propagation of fake news. It took the advantages of deep geometric learning to learn the propagation pattern of fake news [16]. The Multi-source Multi-class Fake news Detection framework (MMFD) [17] exploits the combination of LSTM, CNN, and fully connected network for fusing the data from the multiple sources and it establishes the veracity of genuine/fake statements to detect the fake news automatically. However, it lacks to consider information about user interaction, social networks, and temporal information. The Discourse level Hierarchical Structure of Fake news detection model (DHSF) builds the discourse level structures for each news article in a hierarchical nature for identifying the fake news in an automated manner [18]. The user profile based fake news detection model [19] exploits different users' features concerning the user engagement on news articles for discovering the fake news. Moreover, it examines the relationship between the user profile and the trustworthiness of news articles to identify the fake news. However, the lack of consideration of user credibility and the social context does not contribute the promising results. To provide promising detection results, the work [20] analyzes the content along with the features of the social context of the threshold rule. Even though it employs the heuristics for effective fake news detection, it lacks in sophisticated natural language processing. The rumors on social networks have an impact on the public's lives, which spoils the social cohesion, personal reputation, and national stability. Thus, it is essential to detect and block the propagation of rumors. To filter out the potential fake news sources, the work [21] designs the tool that effectively detects the fake news. The detection algorithm seeks to perform Bayesian inference for identifying fake news to avert the propagation of misinformation. It leverages the crowd signals to discover the fake news and also learns about the flagging behavior of users over some time. Even though it leverages the flagging behavior of users without the knowledge and learning about the user declines the accuracy of detection [22]. The fake news detection based on unsupervised learning algorithms in [23] captures the opinions of the users from the news article engagement with social media and gathers their opinions in an unsupervised manner whereas the user's credibility and veracity of news are considered as the latent random variables. Moreover, it attacks the inference issue by exploiting the collapsed Gibbs sampling approach. However, it lacks to consider the user profiles and news contents while detecting the fake news. An automatic mechanism [24] attempts to categorize true and fake news stories in the popular Twitter threads. It involves four main processes; feature extraction for prediction accuracy, dataset alignment, pre-set feature selection, and evaluation of model transfer. The rumor identification model [25] relies on the GRU-based RNN deep neural network model for detecting the rumor propagation. It utilizes the believability concept for automatically identifying the users spreading misinformation. The rumor identification model exploits the GRU-based RNN deep neural network model without the inclusion of users liking and replying, thus creating adverse effects on the results. The hybrid LSTM based fake news detection model [26] includes the speaker profiles for confirming the credibility of news articles. It incorporates the credit history of users for improving the detection accuracy. However, it disregards what kind of topics the speaker lies about. In order to address the issue in the fake news detection, the work [27] exploits the deep multi-label classifier, which is capable of classifying the given news statement into six fine-grained fake news categories. However, the classifications of the short news statement into fine-grained fake news classes are the most complicated task. To detect the suspicious or verified news post on Twitter, the work [28] exploits the linguistic neural networks along with the linguistic features that are useful for the fine-grained classification whereas the grammar and syntax have little impact. To avert the social disruption, the Credible Early Detection (CED) model detects the rumors as soon as possible before it propagates. However, the lack of attention to the users' profile for rumor detection may mislead the detection. The hybrid deep fake news detection model [29] classifies the tweets not only based on text, but also using the images, reactions, likes, tags, and shares. Despite that, it does not extract the temporal evolution of user engagement on news articles. To deal with the newly arrived events, the Event Adversarial Neural Network framework (EANN) learns the transferable feature representation for detecting the recent unseen record of the event [30]. With the support of the tri-relationship between the news items, publishers, and users, the work [31] seeks to detect the fake news in the social media platform which encourages the trustworthiness of the news system. However, the lack of effective opinion mining on user behavior makes it challenging to automate fake news detection. The hybrid DUAL attention model [32] leverages attention-based bi-directional Gated Recurrent Units (GRU) for capturing the latent features from the content of the news. Additionally, it exploits a deep neural network for learning side information such as the speaker's profile information and reviews. However, without the credibility assessment, the attention mechanism to the latent representation of text can be ineffective.

## III. PROBLEM STATEMENT AND FORMULATION

### *A. Problem Statement*

In social media, the propagation of fake news poses a negative impact on the trustworthiness of the information. The detection of deceptive information among the vast quantities of data is a challenging task. To cope up with this constraint, the earlier researchers employ various learning algorithms to automate the fake news detection to some extent. However, the accuracy of such automated methods is confined and quite often demands an external force for identifying the fake news. They lack to emphasize awareness of political or social context and the engagement of the user with the news while the interpretation of the news, besides novel natural language processing methods, is still missing that decline the general trust and confidence towards the news ecosystem. Because of the lack of attention on contextual and semantic information, there are severe obstacles in the identification of fake news. Thus, it necessitates the semantic knowledge source for capturing the enriched vocabulary words from the ambiguous natural language text. However, most traditional attention models for fake news detection did not take advantage of opinion mining, mining of short text of users brings rich sentiment information and plays a critical role in the identification of fake news. On the other hand, most existing methods of sentiment analysis still face challenges


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

when finding proper senses of a text across domains in which the words are unpredictable. Therefore, it is essential to have the knowledge source for procuring the domain knowledge that provides linguistic and pragmatic knowledge in a specific domain for precise detection. However, the lack of information verifiability and accountability in the existing attention models afford the social users an excellent opportunity to diffuse deceptive information through the network. Another challenge is that the credibility of any user is affected through their relationships with the other users, and also temporary social alignment. The additional challenge is that the identification of an ill-intentioned user can create fake news much like real news. In this sense, it is challenging to ascertain the trustworthiness of users and their content. Addressing the issues above in user credibility evaluation in social networks entails developing robust techniques for estimating user and content credibility.

*B. Problem Formulation*

Let, $U = \{u_i\}=\{u_1,u_2, ...,u_n\}$ be a set of 'n' users on Twitter in which 'i' varies from '1' to 'n', $T = \{t_1,t_2, ...,t_m\}$ is a set of tweets posted by each user. An event 'e' on Twitter is considered as a set of more than one tweet. It is worth mentioning that this work emphasizes the emotion of the user towards news for accurate detection. Given an event 'e,' the proposed work evaluates sentiment score ($SS_c$) for each comment (c) for a tweet(t). The score ($SS_c$) returns the degree of emotion in a given tweet (t) by looking up the SentiWordNet. In particular, in problem formulation, the SPOT focuses on the problem of determining which of the tweet, events, and users on Twitter are deemed as credible. In each event 'e' in Twitter, the SPOT approach aims at assessing the scores '$SS_c$,' and 'CS' to detect the opinion and credibility label of tweet 't' in a set T, users in a set 'U' and news event of tweets. In credibility score (CS) evaluation, various kinds of heterogeneous information in model effectively incorporated, including user credibility '$CS_u$' event credibility '$CS_e$' and tweet credibility '$CS_t$' in which the larger value of CS [0, 1] represents that the user u has the highest credibility. High credibility users are closer to the latent features of true news. Further, it focuses on the identification of the most intelligent malicious user who can exactly mimic legitimate news. Table 1 illustrates the notations used in the proposed algorithm and its explanation.

## IV. METHODOLOGY

The proposed work focuses on developing the fake news detection model for the social network. Initially, the SPOT approach applies the text preprocessing of the social media text messages such as tweets, involving Natural Language Processing (NLP). After preprocessing the social media messages or posts, the proposed approach extracts the critical features of the input, text data in terms of cognitive factors. In subsequence, the proposed approach provides the extracted features to the GRNN model to classify the tweets as rumors or non-rumors that is fake news or not with the analysis of the surrounding context in the tweet sentence.

To effectively detect whether the posted information is fake news or real news, the proposed approach utilizes the contextual features such as the total number of posts at a particular time interval, the length of the posts, the sentiment score from its comments, and the credibility score. By looking up the SentiWordNet, the SPOT effectively determines the opinion about the corresponding post with the help of the semantic knowledge sources. Furthermore, the proposed method evaluates the credibility score for the user, and event regarding the tweet for improving the decision making the process. Thus, the proposed SPOT approach effectively detects fake news with improved response time.


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE 1: NOTATIONS AND ITS DEFINITIONS

| Variables | Meaning |
|---|---|
| E | Social event in Twitter, e={1,2,…,m} |
| M | Total number of social events in Twitter |
| U | Generalized social user in Twitter, i.e. $\{u_i\}=\{u_1,u_2, ...,u_n\}$ |
| T | Tweets in a Twitter, t={1,2,...,k} |
| K | Total number of tweets in a particular event 'e.' |
| t(e) | Tweet in an event 'e.' |
| $SS_c^{t(e)}$ | Sentiment score of $c^{th}$ comment for a tweet 't' in event 'e.' |
| Nc | Total number of comments for a tweet 't.' |
| C | Comments for a tweet 't.' |
| $SS_e$ | Sentiment score of event 'e.' |
| $GCS_{(u_i,e)}$ | Generalized Credibility Score of a particular user ($ui$) for $e^{th}$ event |
| $W_1, W_2, W_3$ | Weighting parameters |
| $(RT_u^e)$ | Number of retweets generated from the tweets posted by the user (u) on event(e) |
| $(RT_{u_i}^e)$ | Number of retweets generated for all the tweets posted by the user ($u_i$) in an event (e) |
| $FW_{u_i}$ | Number of followers to a particular user '$u_i$.' |
| $FD_{u_i}$ | Number of friends of a particular user '$u_i$.' |
| $T_{tp(u_i)}$ | Tweet posted time of a particular user '$u_i$' in Twitter, (i.e. time = year) |
| $T_{pc(u_i)}$ | A profile created time of a particular user '$u_i$' in Twitter, (i.e., time = year) |
| $(UR_u^e)$ | Number of user responses for the tweets posted by the user (u) on event 'e' |
| $(UR_{u_i}^e)$ | Number of user responses for the tweets posted by a particular user '$u_i$' on event 'e' |
| $CS_t$ | Credibility score of the tweet 't.' |
| $N_u(t)$ | Number of users who posted a particular tweet 't' |
| $CS_e$ | Credibility score of an event 'e.' |
| $N_t(e)$ | Total number of tweets in the event(e), |
| $CS_{u_i}$ | Credibility score of a particular user '$u_i$.' |
| $N_u$ | Total number of social users on Twitter in an event 'e.' |

The SPOT approach incorporates four main phases, including data transformation, opinion mining, credibility analysis, and fake news detection. Fig. 1 shows the fake news detection using the deep learning model based on emotion analysis on user comments.

### *A. Data transformation*

The ever-growing utilization of text messages in a social network becomes a significant cornerstone of crowd wisdom extraction specifically, in terms of emotions; hence, sentiment analysis and classification are a potential task. The major challenge in this domain is to handle the data in terms of noise, emoticons, relevance, slangs, and folksonomies. Thus, the proposed method applies various preprocessing on Twitter data for the fortification of emotion classification. In the context of data transformation, the proposed SPOT method heavily relies on the tweet preprocessing and feature engineering for transforming the unstructured tweet into potential relevant information.

#### *1) Tweet preprocessing:*

Ordinarily, Twitter users are more exposed to typographical and spelling errors and the usage of slang and abbreviations. Also, to emphasize their feelings, they utilize punctuation and exclamation marks. Thus, Twitter includes a considerable amount of noise while it retrieves raw information from the tweet. The reduction of noise in the tweet renders the smooth, effective and easy, efficient classification process, whereas the quality of the data is considered as the critical factor in the successful performance of the machine learning classifier. Therefore, it is necessary to normalize and cleanse the raw tweet data, which is referred to as preprocessing.


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

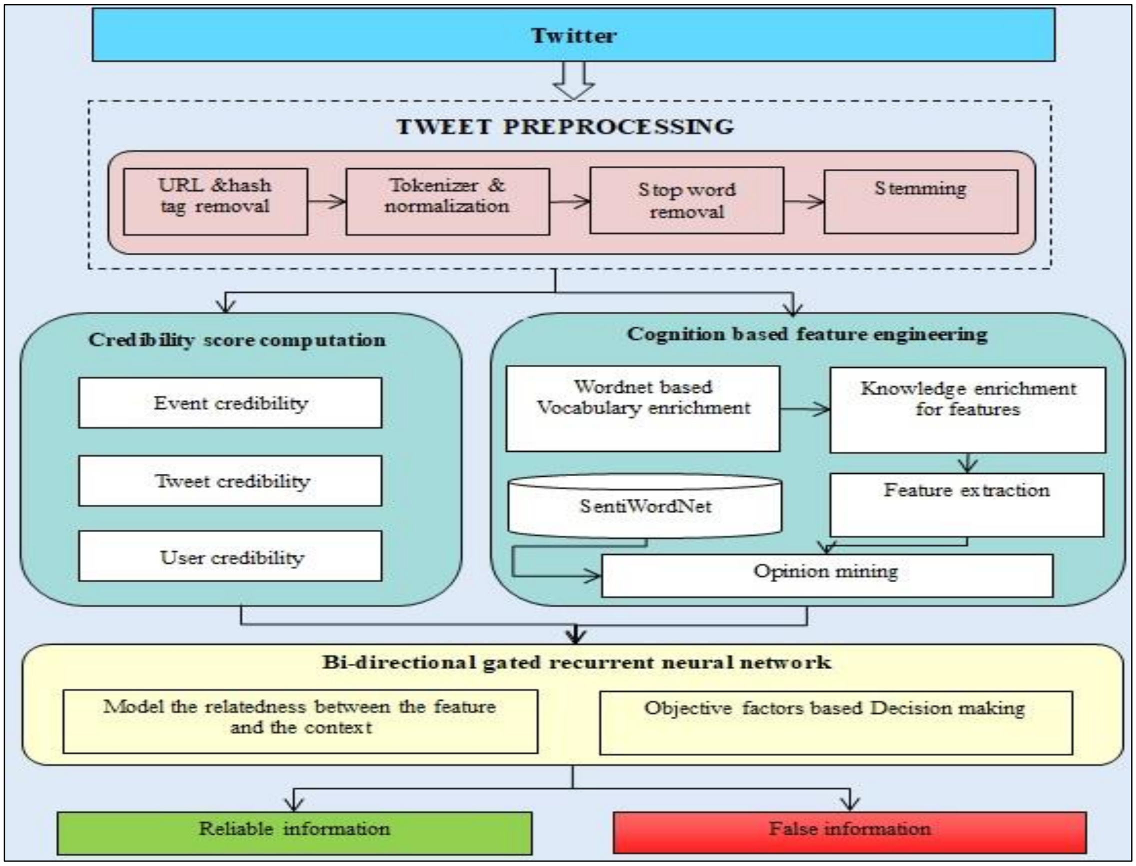


Figure.1 Block Diagram of the SPOT approach

In the context of preprocessing, the tweets are collected from the source; then it has undergone various preprocessing techniques for procuring valid segments from the tweet in which the word does not have potential information for emotion analysis is discarded. Accordingly, the tweet preprocessing helps to retain the semantic information presented in the tweet, which brings beneficial results [33].

*2) Cognition-based feature engineering:*

In Twitter, a user posted natural language format tweets are more prone to grammatical errors, word shortening, inappropriate punctuations, and exaggeration. Furthermore, the word limit restraint drives a user to restrict their content and their opinion. Such inconsistencies make it harder to resolve natural language processing tasks like sentiment analysis. In the context of sentiment analysis, the user expressed opinionated short text is nuanced by subtleties at syntactic, semantic, lexical and pragmatic levels. Also, the complexity and the ambiguity of natural language text make it harder to filter out non-opinionated terms in the text. On the other hand, the correct inference of semantics in the text is non-trivial and entails an analysis of concept, content, and context. The right selection of features is a key step by because of irrelevant and redundant features negatively impact the predictive power and performance of the classifier. Besides, the identification latent features present in the users' text help to contribute important cues to determine if the given news article is real or not. In this work, the proposed method addresses the problem of fake news detection by identifying the useful feature to boost the detection accuracy of the classifier.


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

To accomplish this goal, the proposed method explores various kinds of features that are extracted from the users' text and news sources from the knowledge base. This makes it possible to discover the inherent stylistic variance between the false and legitimate news. In particular, the proposed approach exploits Wikipedia information on retrieving the event-related and vocabulary enriched features for acquiring knowledge of the event. Among the massive extracted feature set, the SPOT approach recognizes the most related co-occurring features through the continuous skip-gram model. To further enrich the feature set, it applies the genetic algorithm for optimally selecting the subset of relevant features. The genetic algorithm based feature selection process evaluates the features regarding the mutual information-based fitness function. Eventually, it emphasizes cognitive features [34] along with semantic and syntactic features for improving the quality. To this end, the SPOT approach builds a negation handling framework by taking into consideration cognitive cues in the text such as polarity. The identification negation in the text helps to reflect the exact knowledge behind users' text which provides absolute information, and thus boosts the detection performance.

*B. Opinion mining on user comments*

Most often, people share their opinions in the form of reviews on Twitter to express their feelings or ideas. Several individuals and organizations seek to utilize this natural language information for better decision making. Therefore, opinion mining is necessary to recognize the opinion of the people from their unstructured natural language reviews. Accordingly, the SPOT approach attempts to analyze the users' replies for identifying the users' conception regarding the specific tweet that helps to reveal to the trustworthiness of the user's tweet. In this context, the opinion mining on user comments helps to discriminate against the satire news from the real news towards their content. For instance, credible users' comments, such as " this is false information," is clear evidence for identifying fake news. The proposed methodology attempts to capture the sentiments from the natural language users' content which is useful for identifying the users' perspective. It exploits the SentiWordNet corpus to discover the different sentiment polarity from the natural language text. By exploiting the SentiWordNet, the SPOT effectively identifies the processed user review to uncover the sentiment of the text, whereas the proposed model, generate the sentiment score ($SS_e$) based on the user comment or reactions regarding the specific tweet.

$$SS_e = \frac{1}{k}\sum_{t(e)=1}^{k}\left(\sum_{c=1}^{nc} SS_c^{t(e)}\right)/nc \quad (1)$$

According to the user comment sentiment score ($SS_c$), the SPOT method evaluates the opinions of the users for each event that is the event level sentiment score ($SS_e$) by using Eq. (1). In this equation, the term t(e) and $SS_c^{t(e)}$ represents the tweet of a specific event (e) and the sentiment score of the comment (c) for a tweet(t) under the particular event (e), the term 'k' and term '$n_c$' denotes total number of user tweets in a event 'e' and the number of the comments in a particular tweet 't(e)' respectively in which k ⊆m.

*C. Credibility score computation*

The recognition of fake news in social media not only relies on sentiment analysis, but also relies on user comments, and the truthfulness of user and tweet. Since the fake news content on social media mimics the real news


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

in order to make misinformations look like legitimate news. In addition, it becomes worse when handling natural language text. Besides the content of the news, there is a necessity to analyze the social context, including social behavior, and the user engagements for evaluating credibility. During crisis events, the public is more exposed to go under the false information or rumors on Twitter owing to emotional vulnerability, which drastically declines the potential of information enclosed on its tweets. To resolve such conflicts, the SPOT approach attempts to evaluate the credibility or trustworthiness of a user and tweet. In the process of assessing tweet credibility, the proposed approach assesses the credibility in three levels, such as user level, tweeter level, and the event level. Also, it is essential to know the interaction-based features of users, as this is a crucial factor while identifying socially reliable users. Therefore, the proposed method incorporates many vital features of tweets and the user's social behavior for estimating the credibility of the users on Twitter. Accordingly, by using Eq. (2) the proposed method computes a generalized credibility score of a user ($u_i$) based on the features of retweet rate (RT), relationships between friends (FD) and followers (FW), age of the user in Twitter, and the number of users' reactions (UR) regarding the specific tweet (t).

$$\mathrm{GCS}_{(u_i,\mathrm{e})}=\mathrm{W}_1\left(\frac{\log\left(\mathrm{RT}^{\mathrm{e}}_{u_i}\right)}{\sum_{\mathrm{u}}\sum_{\mathrm{e}}\log(\mathrm{RT}^{\mathrm{e}}_{\mathrm{u}})}\right)+\mathrm{W}_2\left(\frac{\log\left(\left(\mathrm{FW}_{u_i}\right)\text{-}\left(\mathrm{FD}_{u_i}\right)\right)}{\left(\mathrm{T}_{\mathrm{tp}(u_i)}\text{-}\mathrm{T}_{\mathrm{pc}(u_i)}\right)}\right)+\mathrm{W}_3\left(\frac{\left(\mathrm{UR}^{\mathrm{e}}_{u_i}\right)}{\sum_{\mathrm{u}}\sum_{\mathrm{e}}\log(\mathrm{UR}^{\mathrm{e}}_{\mathrm{u}})}\right) \quad (2)$$

In Eq. (2), the first term denotes the reputation of the user in social media in which numerator term refers the total number of retweets ($RT^{e}_{ui}$) for the tweets posted by a particular user ($u_i$) in a particular event (e) and the denominator term indicates the total number of retweets for the tweets posted by all the users (u) in a event (e), $u_i \in u$. The second term reflects the popularity and social influence of users by their followers–friends relationship with Twitter. Whereas, in the numerator, the terms FW and FD refers the number of followers $\left(\mathrm{FW}_{u_i}\right)$ and number of friends $\left(\mathrm{FD}_{u_i}\right)$ for a particular user ($u_i$) in Twitter and denominator term estimates the age of a particular user ($u_i$) in Twitter. In the denominator, $\mathrm{T}_{\mathrm{tp}(u_i)}$ represents the period at which the tweet posted on Twitter and $\mathrm{T}_{\mathrm{pc}(u_i)}$ indicates the period at which the user profile is created. Further, the third term computes the user responses from the number of replies, the term $\left(\mathrm{UR}^{\mathrm{e}}_{u_i}\right)$ denotes the number of user replies for the tweets posted by a particular user ($u_i$) in a specific event (e), and the term ($UR^{e}_{u}$) implies the total number of replies for the tweets posted by all the users (u) in a event (e). In the Eq. (2), the value of weighting parameters is $W_1+W_2+W_3=1$, $0< W_1,W_2,W_3<1$, $W_1=W_2=0.4$, $W_3=0.2$, $W_1,W_2 > W_3$. In the weighting parameter, the user comments have lower weights. It is because the user comment does not reflect the credibility of a user directly. The reputation and the influence of users on social media easily reveal the credibility of a user for ensuring valuable information propagation. Thus, the proposed SPOT method computes the credibility score of tweets ($CS_t$) using Eq. (3). By exploiting (3), the proposed approach significantly reflects the degree of reliability and confidence of the tweet, and therefore reveals the acceptable information regarding a certain user's credibility.

 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\mathrm{CS_t}=\frac{\sum_{\mathrm{u}} \mathrm{GCS}_{(u_i,\mathrm{e})}}{\mathrm{N_u(t)}} \quad (3)$$

In (3), $N_u(t)$ represents the total number of users posted a particular tweet (t), and $GCS_{(ui,e)}$ denotes the generalized credibility score of the user. The numerator section in (3) gives the summation result of the GCS of the users who post for the corresponding tweet't'.

$$\mathrm{CS_e}=\frac{\sum_{\mathrm{t=1}}^{\mathrm{k}} \mathrm{CS_t}}{\left(\mathrm{N_t(e)}\right)} \quad (4)$$

To assess the presence of credible information, the focus on tweets along with the specific event is essential, whereas the event credibility relies on the knowledge regarding that event. In Twitter, the tweets of the credible users contribute the trustworthiness to a particular event in which the event acquires the credibility from the relevant tweets. Hence, the proposed method assesses the credibility score of the event on Twitter by using Eq. (4) for enhancing the quality of the system. In Eq. (4), the term $N_t(e)$ represents the total number of tweets in the event(e), $CS_t$ denotes the credibility score for all the tweets in the event(e) whereas t={1,2, ...k}, 'k' represents the number of tweets that belong to a particular event (e), k ⊆m .

$$\mathrm{CS}_{u_i}=\frac{\mathrm{CS_t}}{\mathrm{N_u}} \quad (5)$$

There is a possibility of propagation of fake news with proper context through credible users, which has no purpose of conveying false information or mislead the people. In this context, even the credible user may propagate the false information owing to their high trustworthiness of friends and followers. Accordingly, the evaluation of user credibility based on the social relationship alone lacks to identify the fake news. Thus, the proposed SPOT method attempts to reevaluate the credibility of users based on their tweet information by using Eq. (5). In Eq. (5), the term $N_u$ denotes the total number of users.

*D. Bi-directional gated recurrent neural networks based decision making*

To capture the interrelatedness between the words in tweets, the proposed SPOT model employs a gated recurrent neural network that effectively models the tweet-level semantic and syntactic relation in the tweets. Apart from the gated recurrent neural network, the remaining deep learning model lacks to explicitly capture the intrinsic semantic information of tweets, such as co-references, dependency relations, and negation scopes. Moreover, in order to effectively extract the discourse relation, the proposed approach applies the bi-directional GRNN model over sentence representations of the tweet in right-to-left and left-to-rights direction, respectively. Fig. 2 demonstrates the perceptron diagram for the bi-directional GRNN model in which the word vector $w_1,w_2\ldots w_w$ are given as input.

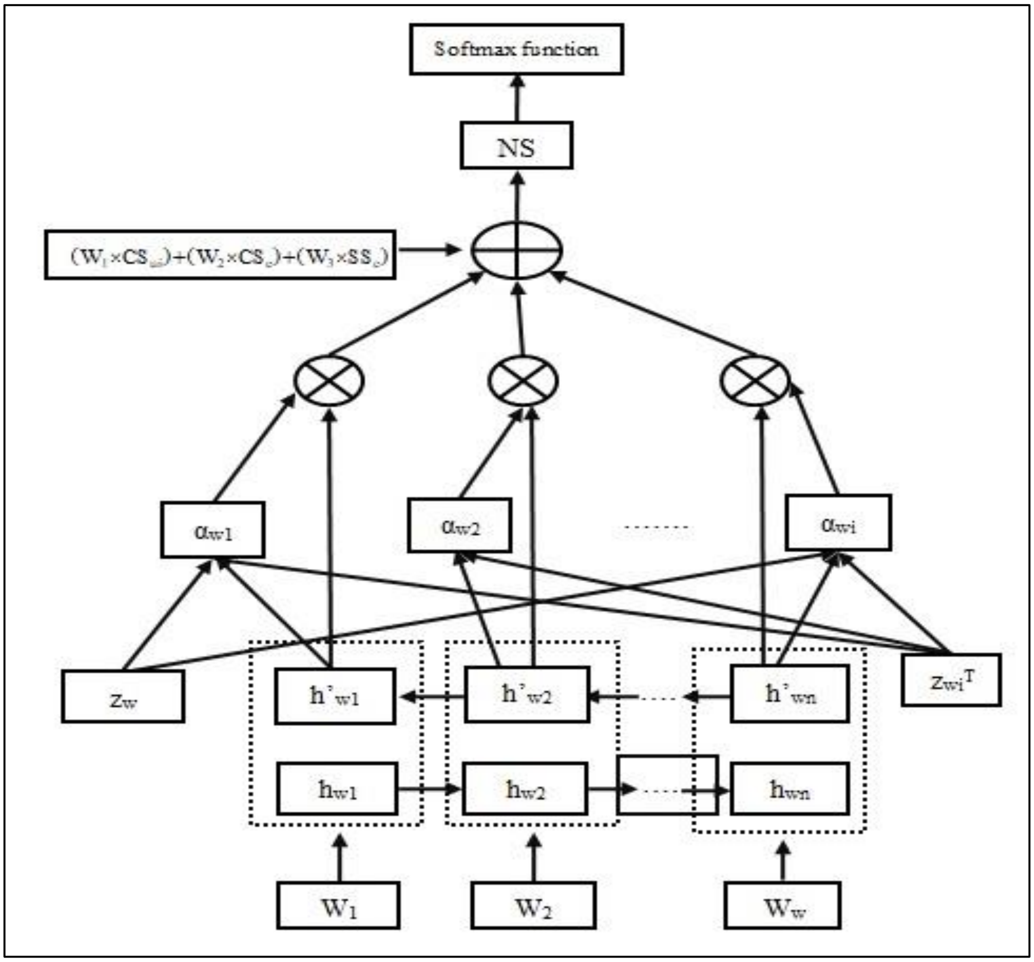


Figure 2. Perceptron diagram for the bidirectional GRNN

It results in a sequence of forwarding state $h_1$, $h_2$, .., $h_n$ and backward state $h_n$', $h_{n-1}$', ..., $h_1$'. In the initial, the proposed method embeds each word present in the tweets into the vector using the embedding matrix. With the help of bidirectional GRNN, the proposed SPOT method acquires the word annotation using summarizing information in both the direction of words, whereas the concatenation of forwarding and backward hidden state absolutely provides the annotations of words in the tweet. In the final layer, bi-directional GRNN uses the softmax function to map the non-normalized vector into a probability that was computed by using the Eq. (6), whereas $Y_i$ represents each element in the vector Y.

$$\text{Softmax function}=\frac{\exp^{(Y_i)}}{\sum_{j=1}^{k}\exp^{(Y_j)}} \qquad (6)$$

In the fake news detection, the entire words in the tweets do not equally contribute to the meaning of the sentence. Thus, the proposed method adopts the attention mechanism for capturing the informative words. In particular, the spatial information is convincing evidence in false information detection in which the potential users who are considered as more informative than others while their spatial location is similar to the event location. Hence, the proposed method involves attention mechanisms for considering the significance of various word vectors.

$$z_i=\tanh\left(W_w\left(h_{wi} \oplus h'_{wi}\right)+b_w\right) \quad (7)$$

In initial, for each word 'w' is one tweet, the proposed method integrates the weights ($W_w$) into a bi-directional state vector by using (7), whereas the $z_i$ and $b_w$ represent the hidden representation of $h_i$ and bias of the word.

$$\propto_{wi}=\frac{\exp\left(z_{wi}^{T} z_w\right)}{\sum_i \exp(z_{wi}^{T} z_w)} \quad (8)$$

Subsequently, the proposed method attempts to compute the normalized importance ($\alpha_{wi}$) with the help of softmax function by utilizing (8) in which $w_i$ refer to the word in a sentence. After that, it is estimated the news score 'NS' based on the sentence vector by utilizing the Eq. (9) in the manner of the weighted sum of annotation of words regarding the weights.

$$NS=\sum_{wi} \propto_{Wi}\left(h_{wi} \oplus h'_{wi}\right)+\left(W_1 \times CS_{u_i}\right)+(W_2 \times CS_e)+(W_3 \times SS_e) \quad (9)$$

In Eq. (9), the proposed SPOT method additionally considers the credibility score of a user ($CS_{ui}$) and event ($CS_e$) with the sentiment score ($SS_e$) for detecting the fake news, whereas $W_1$, $W_2$, and $W_3$ are the weighting parameters for revealing the importance of the term, $W_1+W_2+ W_3=1$, $W_1=W_2=0.4$, $W_3=0.2$, $W_1,W_2 > W3$. In the weighting parameters, the $SS_e$ receives lesser importance than the remaining terms. This is because the user comment based sentiment score may vary owing to the variation of users' intention regarding the event. On the other hand, the credibility analysis based on social behavior, user profile, and the event information contributes more informative and effective information. Considering the additional measures in the Eq. (8), helps to acquire the precise results in the fake news detection. Algorithm 1 explains the procedure of the proposed SPOT approach.

## V. EXPERIMENTAL EVALUATION

To demonstrate the performance of the proposed SPOT approach, the experimental framework compares the proposed approach with the existing DUAL approach using the test data set.

### *A. Dataset*

The experimental setup evaluates the proposed approach by utilizing the PHEME dataset [35]. It incorporates Twitter rumors and non-rumor conversations threads posted during the period of breaking news.


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

**Input:** Twitter data including tweet (t), user comment (c), users' information
**Output:** Fake news score, NS
Let, user (u), event(e), tweet (t)
**// Sentiment score calculation**
    **for** each user comment (c) **do**
        Captures the sentiment score ($SS_e$) using the equation (1)
    **endfor**
**// Credibility score computation**
**Input: tweets(t), UR**
    **for** each tweet t **do**
        **for** each user reaction (UR) **do**
            Compute the generalized credibility score ($GCS_{ui,e}$) using equation (2)
            Computes the credibility score of tweets ($CS_t$) using the equation (3)
            Credibility score of event in Twitter by using the equation (4)
            Computes the credibility score of user by using the equation (5)
        **endfor**
**endfor**
**// fake news detection**
**Input: features of tweets, $CS_e$. $CSu_i$, $SS_e$**
    **for** each tweet t **do**
        Compute hidden representation $u_i$ word in tweet using the equation(7)
        Compute the normalized importance weight ($\alpha_{wi}$) using equation (8)
        compute 'NS' using equation (9)
    **endfor**

Algorithm 1: Fake News Detection Algorithm

The tweets are about the nine newsworthy events including 'Ottawa shooting', 'Sydney siege', 'Charlie Hebdo shooting', 'Germanwings plane crash', 'Putin missing', 'prince to play in Toronto', 'Ferguson unrest', 'Gurlitt collection', 'Michael Essien contracted Ebola' whereas each event is updated with veracity information. The event 'Charlie Hebdo shooting' comprises of 1,621 non-rumors and 458 rumors, the 'Ferguson unrest' event incorporates 859 numbers of non-rumors tweets and 284 numbers of rumors, the event 'Germanwings Crash' includes 238 rumours and 231 non-rumours, 'Sydney siege' event comprises of 522 rumors and 699 non-rumours, and the event 'Ottawa shooting' incorporates 470 rumours and 420 non-rumours. Furthermore, in the PHEME dataset each event has tweet ID, source tweet, and the set of reactions towards the source tweet.

*B. Experimental setup*

The experimental model implements the proposed SPOT and the existing DUAL algorithm using the python programming language. The experiments conducted on Ubuntu 16.04 64 bit machine. The software settings, including python version 3.6.8 run on Spyder IDE 3.3.6. It exploits the machine learning libraries, including pandas, numpy, nltk processing, sklearn preprocessing, sklearn metrics, keras preprocessing, and gensim.

*C. Evaluation metrics*

By formulating fake news detection as a classification problem, the proposed method defines the metrics include accuracy, precision, recall, and f-measure that allows evaluating the performance of a classifier under various perspectives.

1) *Precision:*

$$\text{Precision}=\frac{TP}{TP+FP} \qquad (10)$$


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Precision in Eq. (10) is the ratio between the total number of correctly detected fake news and the total number of fake news articles detected by the system.

*2) Recall:*

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \qquad (11)$$

A recall as in Eq. (11) is a ratio between the total number of correctly detected fake news and the total number of news articles that belong to fake news.

*3) Accuracy:*

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad (12)$$

Accuracy in Eq. (12) is the percentage of the total number of correctly detected fake and real news from the overall news articles.

*4) F1-score:*

$$\text{F1-score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \qquad (13)$$

F1-score in (13) combines the precision and recall results in fake news detection that provides overall prediction performance.

*5) Error rate:*

$$\text{Error rate}=\frac{\text{FP+FN}}{\text{TP+FP+TN+FN}} \qquad (14)$$

The error rate metric in (14) helps to estimate the misclassification error of the classifier. In this equation, FP and FN denote the false positive and false negative, in which false positive is the number of real news that is incorrectly classified as fake news and false negative is the number of fake news that is incorrectly labeled as the real news. The term, TP and TN represent the true positive and true negative, whereas true positive is the number of fake news is correctly classified as fake news, and true negative is the number of real news that is correctly classified as real news.

*6) Log loss:*

$$\text{log loss}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log p_{ij} \qquad (15)$$

The log loss metric in Eq. (15) helps to evaluate the performance of the classification model. In this equation, N and M represent the number of data instances and possible labels, yij and pij indicate whether or not class label j is the correct classification for ith data instance and the probability of assigning j label to ith data instance. An ideal classifier has a Log Loss value of precisely zero.

*7) Geometric mean (G-mean):*

$$\text{G-mean}=\sqrt{\text{TP}\times\text{TN}} \qquad (16)$$


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

G-mean in Eq. (16) is the square root of the product of TP and TN. It helps to improve the accuracy of each class when keeping accuracies are balanced.

*8) Receiver Operating Characteristics (ROC) curve:*

It measures the overall performance of the classifier. It shows the relation between the false detection rate and the correct detection rate at different threshold settings. This curve shows the performance of the classification model over various classification thresholds.

*D. Experiment results*

*1) Training data size Vs. Precision:*

The precision value of both the SPOT and the existing DUAL approaches is illustrated in Fig 3. This figure demonstrates that the performance of proposed SPOT and existing DUAL approaches while the variation of training data size from 3000 to 6000. The precision value increases with the increase of the dimension of training data size. Accordingly, the SPOT approach linearly increases its precision value by 6.96%, but the existing DUAL approach increases only 1%, while increasing the training data size from 3000 to 6000, which indicates that it precisely computes the fake news in the social media. It is because the proposed SPOT method utilizes the semantic knowledge sources, including Wordnet, and Wikipedia knowledge source to procure the rich vocabulary enriched potential features and gain the knowledge about the Twitter event for ensuring the efficient classification results and the right amount of valuable information. The proposed approach also concerned with the cognitive features for further improving the classification result. Therefore, the SPOT approach achieves 97.26% of precision value even when dealing with the 5628 training data sample. On the other hand, the DUAL approach only attains 87% of precision value owing to the lack of absolute extraction of conceptual and domain knowledge during feature extraction.

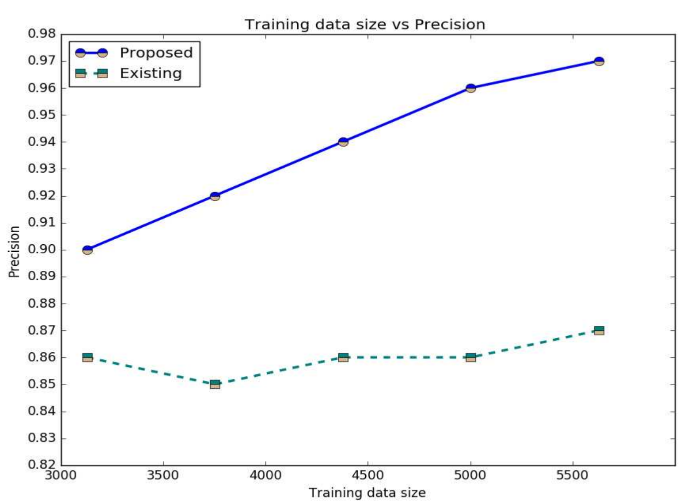


Figure 3. Training data size Vs. Precision


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

*2) Training data size Vs. Recall:*

Fig. 4 compares the results regarding the recall rate of the SPOT approach and the existing DUAL approach with the variation of the total amount of training data size from 3000 to 6000. The SPOT approach considers that the opinion mining on user comments is essential to find out the exact intention of the users. The SPOT approach attains a higher recall rate of 6.28% than the existing DUAL approach when the training data size is 3127 samples. It is because the SPOT approach captures the opinions of the user during the opinion mining phase, which assists to inherently identify the truth level of the news content by analyzing the opinions of the users that helps to reflect the truth of the news content. It captures the inherent emotions of the users and recognizes their opinion through sentiment extraction rather than considering the extracted features alone for decision making in fake news detection that helps to improve the recall rate. Hence, the extraction of sentiment in the users' text deals with the extraction of opinions in the unstructured text associated with the topic of the document being analyzed. It helps to recognize the attitude of the speaker regarding the topic. The proposed approach attains the highest recall rate of 97.27%, even when increasing the size of training data as maximum. However, the DUAL approach performs poorly in ensuring the accurate detection of fake news when compared to the SPOT approach. Accordingly, the existing DUAL approach accomplishes only 85% of recall when handling the maximum number of input samples.

*3) Fake News Ratio (FNR) Vs. Accuracy:*

Fig. 5 shows the accuracy for both the proposed and existing approaches when varying the fake news ratio from 0.1 to 0.5. Fake News Ratio (FNR) is the ratio between the number of fake news and the total number of news. When FNR= 0.1, the SPOT approach attains a higher value of the accuracy value by 99.15%, but the DUAL approach only obtains 85% of accuracy value owing to the lack of focus on the sentiment of the users. Moreover, the SPOT approach achieves and maintains better accuracy than the existing DUAL approach even with the higher value of FNR.

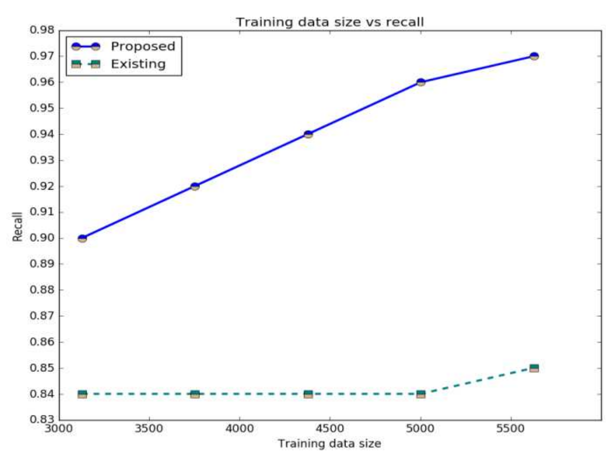


Figure 4. Training data size Vs. Recall

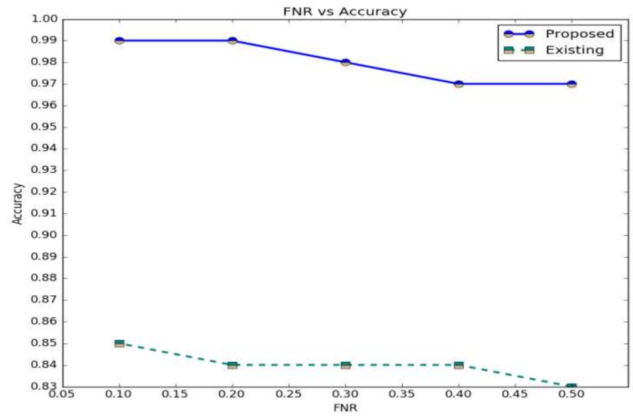


Figure 5. Fake News Ratio (FNR) Vs. Accuracy

The accuracy results show that the proposed method retains the higher rate of accuracy rather than existing methods while trying to discover the non-credible sources among a large number of fake news content present in the system. Accordingly, the proposed method accomplishes 99% of accuracy when FNR is 0.2, and it maintains the accuracy value up to FNR=0.5. However, the existing DUAL approach suddenly drops its accuracy rate from 85% to 83% due to the lack of attention mechanism for credibility analysis. Furthermore, the proposed SPOT approach performs the credibility analysis to discover the implausible news content that's posted on Twitter and prevent the propagation of malicious information. In this context, the credibility analysis helps to find out the user who is more vulnerable to spread malicious information and to compute the erroneous news content that helps to improve the fake news detection on Twitter.

*4) Fake News Ratio (FNR) Vs. F-Score:*

The F-score of both the SPOT and the DUAL approaches are depicted in Fig. 6 with the variation of the FNR. When varying the FNR from 0.1 to 0.5, the proposed SPOT approach initially attains the higher f-score and attempts to maintain the optimal f-score value for various FNR value. When the FNR value is 0.3, the SPOT approach accomplishes 14% greater f-score value compared to the existing DUAL approach. It is because the proposed SPOT approach additionally considers the spatial information of the users while computing the fake news score based on the bidirectional GRNN model. In essence, the credible users who are in a location nearer to the event place provided potential information that has a higher possibility of real information compared to the farthest credible users. Therefore, the user’s spatial information based fake news detection helps to provide convincing evidence for detecting the fake news and improve the performance.

5) *Training data size Vs. Error rate:*

Fig. 7 shows the error rate for both the proposed SPOT and existing DUAL approaches when varying the number of training data sizes from 3000 to 6000. Error rate provides the misclassification rate of the classifier. In essence, the higher level of error rate indicates that the classifier is lacking to discover the right class for the data samples that leads to misclassification. When the training data size is 4377, the DUAL approach attains the higher value of error rate as 16% owing to the incorrect classification of input data instances. On the other hand, the proposed SPOT approach acquires the lowest error rate of 6%. It is because the SPOT approach captures more informative words from the tweets and social information of the users through the semantic knowledge source and opinion mining rather than the existing DUAL approach. This potentially extracted informative word assists in improving the detection accuracy of the fake news detection model. From the figure, it is observed that even when increasing the dimension of the data, the proposed method optimally declines the error rate, but the existing method retains the error rate.

*6) Training data size Vs. Log loss:*

Fig. 8 compares the results of the log loss value of the SPOT approach and the existing DUAL approach with the variation of the number of training data from 3000 to 6000. According to the probabilities, the log loss metric penalizes the classification model for each false classification to quantify the accuracy.

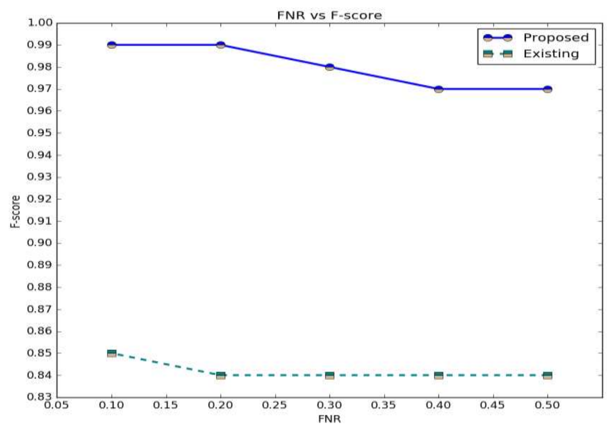


Figure 6. Fake News Ratio (FNR) Vs. F-score

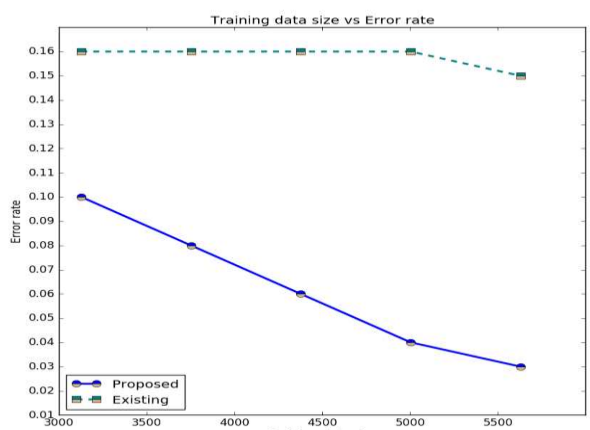


Figure 7. Training data size Vs. Error rate

The minimal value of log loss represents superior predictions. From Fig. 8, it is observed that the proposed SPOT approach linearly decreases the log loss value and returns the minimal rate of 5%, even while increasing the number of training data size as 5628. In contrast, the existing DUAL approach attains 38% of log loss when escalating the size of the data. Even though the existing method considers the social engagement of users towards news on Twitter provides auxiliary information to detect the fake news, but utilizing social engagements of the user is rather limited. The proposed SPOT method considers the correlation between the users, user-news stance engagement relation, latent user features, and user credibility to detect malicious news from the real news. Therefore, the proposed SPOT method attains a lower rate of log loss than the existing DUAL method.

*7) Length of tweets Vs. G-means*

Fig. 9 demonstrates that the classification performance of both the proposed and existing approaches while the variation of length of tweets from 5 to 25. To evaluate the proposed and existing classifier model on an imbalanced dataset G-means metric is used. This metric represents the balance between the classification performance of the minority and majority classes. When the length of tweets varies from 5 to 25, the proposed SPOT approach obtains the optimum value of the g-mean, and then, it attempts to maintain the optimal value. When the length of the user's tweet is 5, the DUAL approach acquires only 80.9% of the g-mean value and maintains it until only the number of lengths of tweets is 15. At the same time, the SPOT approach attains 98% g-mean value and maintains it even when there is the tweet length is 20, which seems the proposed SPOT approach perfectly classifies the actual fake news as fake news and actual real news as real news. It is because the SPOT approach effectively captures the content of the news even when the user posted their tweet in a shorter length by performing the opinion mining on a tweet, which helps to improve the performance. Even though the existing methods extract the relevant features through the deep learning model; it lacks to capture the news content for the shorter length text owing to the absence of opinion mining on the users' text in the Twitter.

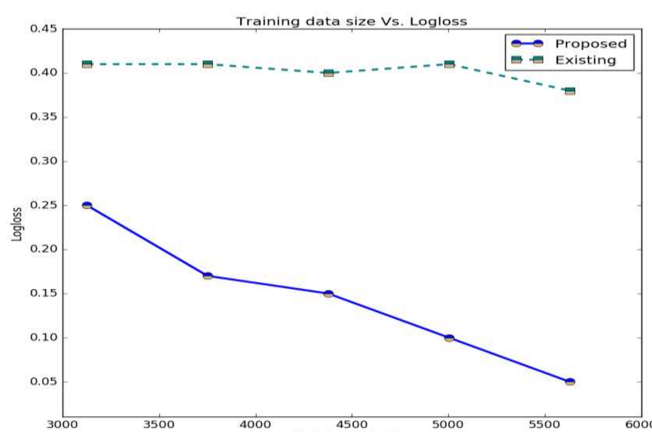


Figure 8. Training data size Vs. Log loss


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

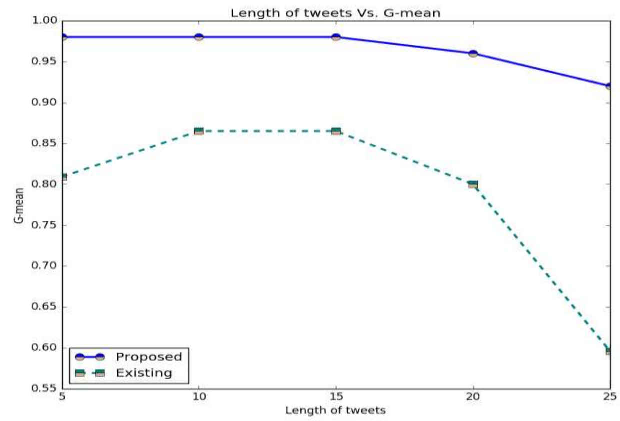


Figure 9. Length of tweets Vs. G-means

*8) ROC curve:*

The visualization of the ROC curve and True Positive Rate (TPR) and False Positive Rate (FPR) of the classifier model shown in Fig 10. The ROC curve enables a comparison of the performance of classifiers by considering the trade-off in the TPR and FPR which evaluate the performance of the classifiers when they have the unbalanced classes.

From the ROC curve, it is noted that the proposed model performs well, rather than the existing model. Accordingly, the SPOT approach accomplishes the ROC curve value of 99% for both the classes, at the same time; the existing DUAL approach attains only 89% of the ROC curve value for both the classes. It reveals that the proposed SPOT approach significantly separates the fake news from the real news rather than the existing DUAL approach.

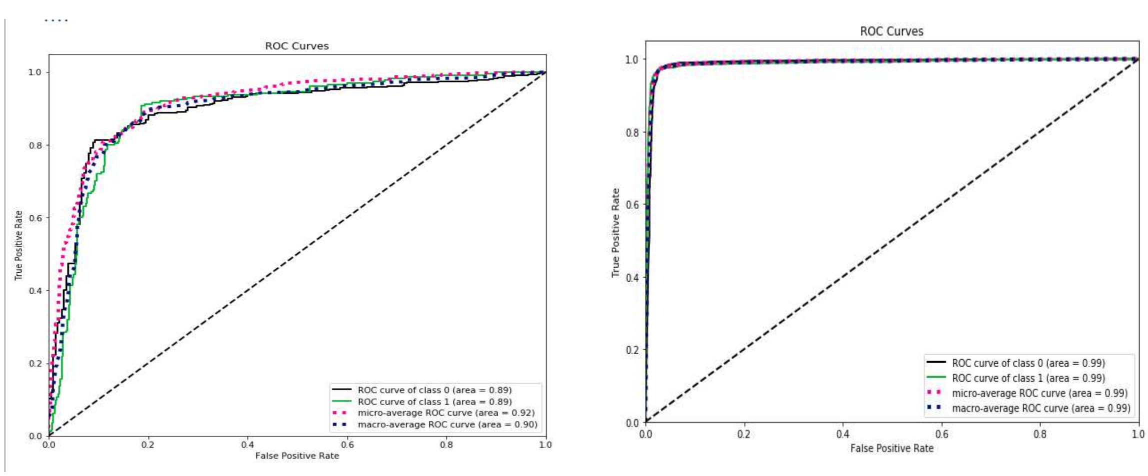


Figure 10. ROC curve for the existing and proposed model

## VI. CONCLUSION

This paper presented the SPOT approach, a fake news detection method based on semantic knowledge source and deep learning. The proposed SPOT method exploits the semantic knowledge source during the feature extraction for improving the feature selection and also it pays much attention for the retrieval of potential features that eases the opinion mining. To detect the false news stories in Twitter effectively, the SPOT method performs the opinion mining on user comments for mining the opinion of the user, which helps to improve the fake news identification. Furthermore, the proposed method assesses the credibility of user and event in the Twitter by using the social behavior and user profile and interaction between the news and user for acquiring high-level efficiency in the detection. Evaluation results on real-world Twitter fake news datasets show the effectiveness of the proposed SPOT approach and the significance of opinion mining and credibility assessment in fake news detection. It is worth mentioning that the SPOT approach attains excellent fake news detection performance in the earlier stage. Accordingly, the proposed approach achieves a better accuracy of the recall rate of 97% compared to the existing approach.

Future direction:

- As the misinformation evolves rapidly in Twitter, it is worth to extend the SPOT approach by exploring and selecting the useful features over the massive features using the feature selection methods.
- The proposed approach will be further extended by focusing on media content to identify malicious information accurately.
- The future work will consider the specific writing style of the manipulators in Twitter while detecting the fake news for improving the detection performance.

## REFERENCES

[1] Park, Chang Sup, “Do social media facilitate political learning? Social media use for news, reasoning and political knowledge”, The Journal of Social Media in Society, Vol.6, No.2, pp.206-238, 2017.

[2] Pierri, Francesco, and Stefano Ceri, “False News On Social Media: A Data-Driven Survey”, arXiv preprint arXiv:1902.07539, 2019.

[3] Allcott, Hunt, Matthew Gentzkow, and Chuan Yu, “Trends in the diffusion of misinformation on social media”, No.w25500, National Bureau of Economic Research, 2019.

[4] Azzimonti, Marina, and Marcos Fernandes, “Social media networks, fake news, and polarization”, No.w24462, National Bureau of Economic Research, 2018.

[5] Zhou, Xinyi, Reza Zafarani, Kai Shu, and Huan Liu, “Fake News: Fundamental Theories, Detection Strategies and Challenges”, In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp.836-837, 2019.

[6] Campan, Alina, Alfredo Cuzzocrea, and Traian Marius Truta, “Fighting fake news spread in online social networks: Actual trends and future research directions”, In IEEE International Conference on Big Data (Big Data), pp.4453-4457, 2017.

[7] Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, “Fake news detection on social media: A data mining perspective”, ACM SIGKDD Explorations Newsletter, Vol.19, No.1, pp.22-36, 2017.

[8] Aphiwongsophon, Supanya, and Prabhas Chongstitvatana, “Detecting Fake News with Machine Learning Method”, In IEEE 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp.528-531, 2018.

[9] Zhou, Xinyi, and Reza Zafarani, “Fake News: A Survey of Research, Detection Methods, and Opportunities”, arXiv preprint

arXiv:1812.00315, 2018.

[10] Pham, Trung Tin, “A Study on Deep Learning for Fake News Detection”, 2018.

[11] Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer, “Fake news on Twitter during the 2016 US Presidential election”, Science, Vol.363, No.6425, pp.374-378, 2019.

[12] Allcott, Hunt, and Matthew Gentzkow, “Social media and fake news in the 2016 election”, Journal of economic perspectives, Vol.31, No.2, pp. 211-36, 2017.

[13] Amoruso, Marco, Daniele Anello, Vincenzo Auletta, and Diodato Ferraioli, “Contrasting the spread of misinformation in online social networks”, In Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, International Foundation for Autonomous Agents and Multiagent Systems, pp.1323-1331, 2017.

[14] Zubiaga, Arkaitz, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter, “Detection and resolution of rumours in social media: A survey”, ACM Computing Surveys (CSUR), Vol.51, No.2, pp.32, 2018.

[15] Parikh, Shivam B., and Pradeep K. Atrey, “Media-Rich Fake News Detection: A Survey”, In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp.436-441, 2018.

[16] Monti, Federico, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein, “Fake News Detection on Social Media using Geometric Deep Learning”, arXiv preprint arXiv:1902.06673, 2019.

[17] Karimi, Hamid, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang, “Multi-source multi-class fake news detection”, In Proceedings of the 27th International Conference on Computational Linguistics, pp.1546-1557, 2018.

[18] Karimi, Hamid, and Jiliang Tang, “Learning Hierarchical Discourse-level Structure for Fake News Detection”, arXiv preprint arXiv:1903.07389, 2019.

[19] Shu, Kai, Suhang Wang, and Huan Liu, "Understanding user profiles on social media for fake news detection”, In IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 430-435, 2018.

[20] Della Vedova, Marco L., Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, and Luca de Alfaro, “Automatic Online Fake News Detection Combining Content and Social Signals”, In IEEE 22nd Conference of Open Innovations Association (FRUCT), pp.272-279, 2018.

[21] Aldwairi, Monther, and Ali Alwahedi, “Detecting Fake News in Social Media Networks”, Procedia Computer Science, Vol.141, pp.215-222, 2018.

[22] Tschiatschek, Sebastian, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause, “Fake news detection in social networks via crowd signals”, In Companion of the The Web Conference 2018 on The Web Conference 2018, International World Wide Web Conferences Steering Committee, pp.517-524, 2018.

[23] Yang, Shuo, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu, “Unsupervised Fake News Detection on Social Media: A Generative Approach”, 2019.

[24] Buntain, Cody, and Jennifer Golbeck, “Automatically identifying fake news in popular Twitter threads”, In IEEE International Conference on Smart Cloud (SmartCloud), pp.208-215, 2017.

[25] Rath, Bhavtosh, Wei Gao, Jing Ma, and Jaideep Srivastava, “From retweet to believability: Utilizing trust to identify rumor spreaders on Twitter”, In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp.179-186, 2017.

[26] Long, Yunfei, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang, “Fake news detection through multi-perspective speaker profiles”, In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Vol.2, pp.252-256, 2017.

[27] Roy, Arjun, Kingshuk Basak, Asif Ekbal, and Pushpak Bhattacharyya, “A Deep Ensemble Framework for Fake News Detection and Classification”, arXiv preprint arXiv:1811.04670, 2018.

[28] Volkova, Svitlana, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas, “Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter”, In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vol.2, pp.647-653, 2017.

[29] Ruchansky, Natali, Sungyong Seo, and Yan Liu, “Csi: A hybrid deep model for fake news detection”, In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp.797-806, 2017.

[30] Wang, Yaqing, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao, "EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection”, In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp.849-857, 2018.


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[31] Shu, Kai, Suhang Wang, and Huan Liu, "Beyond news contents: The role of social context for fake news detection", In ACM Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp.312-320, 2019.

[32] Dong, Manqing, Lina Yao, Xianzhi Wang, Boualem Benatallah, Quan Z. Sheng, and Hao Huang, "DUAL: A Deep Unified Attention Model with Latent Relation Representations for Fake News Detection", In International Conference on Web Information Systems Engineering, pp.199-209, 2018.

[33] Symeonidis, Symeon, Dimitrios Effrosynidis, and Avi Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis",Expert Systems with Applications, Vol.110, pp.298-310, 2018.

[34] Roets, Arne. "'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions", Intelligence, Vol.65 pp.107-110, 2017.

[35] Available online at: https://figshare.com/articles/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078, assessed on April 2019. , assessed on September 2019.