Applied Soft Computing Journal 100 (2021) 106983

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Multiple features based approach for automatic fake news detection on social networks using deep learning

Somya Ranjan Sahoo [a], B.B. Gupta [a,b,c,*]

[a] Department of Computer Engineering, National Institute of Technology Kurukshetra, India
[b] Department of Computer Science and Information Engineering, Asia University, Taiwan
[c] Macquarie University, Australia

ARTICLE INFO

Article history:
Received 7 August 2020
Received in revised form 8 November 2020
Accepted 30 November 2020
Available online 9 December 2020

Keywords:
Online Social Network
Fake news
Deep learning
Hybrid approach

ABSTRACT

In recent years, the rise of Online Social Networks has led to proliferation of social news such as product advertisement, political news, celebrity's information, etc. Some of the social networks such as Facebook, Instagram and Twitter affected by their user through fake news. Unfortunately, some users use unethical means to grow their links and reputation by spreading fake news in the form of texts, images, and videos. However, the recent information appearing on an online social network is doubtful, and in many cases, it misleads other users in the network. Fake news is spread intentionally to mislead readers to believe false news, which makes it difficult for detection mechanism to detect fake news on the basis of shared content. Therefore, we need to add some new information related to user's profile, such as user's involvement with others for finding a particular decision. The disseminated information and their diffusion process create a big problem for detecting these contents promptly and thus highlighting the need for automatic fake news detection. In this paper, we are going to introduce automatic fake news detection approach in chrome environment on which it can detect fake news on Facebook. Specifically, we use multiple features associated with Facebook account with some news content features to analyze the behavior of the account through deep learning. The experimental analysis of real-world information demonstrates that our intended fake news detection approach has achieved higher accuracy than the existing state of art techniques.

1. Introduction

The pattern of communication with each other has changed after mid-1990. Multiple online social networks like Facebook, Twitter, Instagram, and WhatsApp ease the distribution of user's real-time information between multiple users over the same and different networks. Due to multiple characteristics of online social networks like, ease of use, faster transformation and less expensive, it becomes the significant way of communication and information sharing. Nowadays, almost all the social network users access the news through online channels. However, due to the increasing popularity of OSNs, the use of the Internet becomes an ideal way of communication and spreading of fake news. The spreading of fake news in the form of misleading content, fake reviews, fake rumors, advertisements, fake speech regarding politics, satires and many more. Currently, fake news is spread faster in social media rather than mainstream media [1]. Manipulated information's by multiple propagandist to convey political and other influential messages over the network. For examples, in the 2019 general election time in India, multiple users created fake accounts to spread fake news on Twitter and Facebook to attract people. Furthermore, a considerable amount of misleading and incredible information is generated by the multiple users and display through the social network platform [2–4]. It also created a potential threat to multiple communities and had a profoundly negative impact over the users through multiple advertisement, online shopping and social messaging in terms of 3 V depicted in Fig. 1.

Some information spreads are confused social network users by triggering them and distrust. Detection and identification of Fake news on a social platform is a challenging task. The quick spreading nature of fake news affects millions of users and their true environment. Creation of fake news is not a new problem in the social network platform. Multiple companies and the reputed person use multiple social media network for advertising their product and built reputation. All these operation influences many users to shared and like that news. By this process, the fake news also spread over the network. In term of a topic, the content of the fake news, style and media platform changes time to time and fake news attempt to distort linguistic form. Also, this content

* Corresponding author at: Department of Computer Engineering, National Institute of Technology Kurukshetra, India.
E-mail address: gupta.brij@gmail.com (B.B. Gupta).

https://doi.org/10.1016/j.asoc.2020.106983
1568-4946/

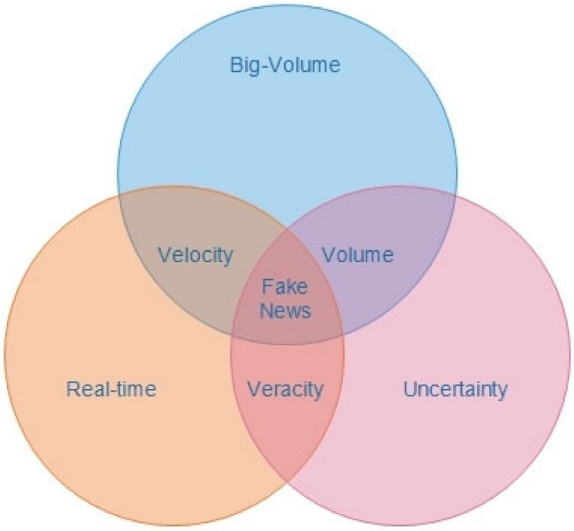

Fig. 1. Fake news in terms of 3 V.

was mocking the actual news. Sometimes, fake news holds true evidence inside the fake context to support a nonfactual claim [5]. Especially, after the US presidential election in 2016, fake news has attained more aid by academician and researchers [6]. People use multiple manual approaches for fact-checking systems such as FaceChek.org and PolitiFact.com. These websites play a vital role in detecting fake news over the network. But all these software's need expert study for a timely response. Also, all these fact-checking systems are mainly focusing on political issues. Besides, a large amount of information created, shared, liked and commented in an online social network platform by the user. Many fake profiles are also spread fake news over the social network platform through multiple posts [7]. All these shared contents make it difficult to detect in the social platform due to bulk information. The fake news spreads over the network cannot be exposed through a blind eye. To distinguish fake news from genuine one, some characteristics of the fake news must be studied. As per awareness, numerous article and blogs are written. These articles provide multiple tricks and tips to handle misleading information. While each author built a different framework and model by analyzing multiple characteristics to identify fake news in social platform related to the text. As per Fig. 1, the content in the social platform increases without any verification and everyone easily write and dispersed the fake material over the internet [8].

Many WebPages published fake news such as deversuardian.com, ABCnews.com and so on. On the other hand, multiple types of fake news such as false statements, fake advertisements, conspiracy theories, satire news, and fake news. These varieties affect human lives in every aspect. All these news dominate the publics' opinion, interest, and decision. There are several characteristics agreed upon by multiple author and researchers to detect fake news relating to text, the response in the form of share, live and the source from where it generated. The traditional approach for detecting fake news by a human editor and expert journalist does not analyze the volume of news created through the social network platform. For this purpose, new computational techniques are required to detect it timely. But some manual expert verification was needed, including a computational approach to identify the fake news. These processes required some third party application to check the fact-finding regarding particular topics like Snopes and Fact checker websites. All computational approaches used for fact-finding and fake detection are based on some predictive models to classify whether the news is fake or not based on previous study. One of the main challenges in training such a framework and models is a suitable feature study subject to multiple categories of news. Further, the designing of a proper model based on the estimated source is challenging. To eradicate all these difficulties, an adequate frame work is required to detect fake news on the Facebook platform by analyzing the characteristics of the news and user activities.

To achieve this goal, we have followed some principle framework. Based on the multiple features study using experiments and literature survey, a fake news detection framework is proposed for detecting fake news in the Facebook platform on user's home page. The proposed framework helps the user to organize and characterize multiple features related to the user account and shared content in the form of fake news. In order to evaluate the efficiency of our proposed framework, we collected fake and legitimate news from different profiles those are participating in spreading news. The collected information passes through multiple machine learning and deep learning-based analyzer for better decision making.

The remaining of the paper are organized as follows. Section 2 describes the multiple literature survey and state of arts of fake news detection on multiple social network platforms. In Section 3, we elaborate the proposed framework based on the multiple features study for detecting fake news on the Facebook platform. The experimental study related to machine learning and deep learning describes in Section 4. In Section 5, we elaborate multiple experimental evaluations based on multiple features and classifications. Chrome extension based detection system with multiple classifiers describes in Section 6. In Section 7, we describe the comparative study of our framework with another state of art techniques. Finally, in Section 8 we conclude our paper with some research direction.

2. Related work

The processes of detecting fake news in the social network platform have been studied since people use the virtual environment for communication. Fake news are defined as misleading information propagated through multiple social network users using different services for certain financial gains. The advancement of social network platform invites users to send and communicate personal information to their belongings. It also attracts many researchers to protect user information from multiple threats. The detection of malicious content in the form of fake news is a challenging task for the researchers. Some researcher uses multiple graph-based and machine learning approaches to detect malicious contents in the form of fake news. The process of flowing fake news was thrust into the spotlight during the presidential election in the United States in the year 2016 through the social network platform. The effect of that news was not clear but, it propagates some rumor, confusion, and deception among users [9]. Due to the emerging media environment that streamlines partition polarization, fragmentation with misleading political information spreading even more widely [10]. The content in the social network as fake news has spurred more doubt about the user's honesty, dignity, openness, nationalism, and sometimes stability, resulting in fears and anger among the users. It also unprecedentedly changed country [11]. Mainly, the detection of fake news is based on news content and social contexts. Based on the textual and visual aspect, news

content features are extracted. Some contexts are also extracted based on sensational emotions. Also, latent textual representation is framed using deep neural network [12] and tensor factorization [13]. Recently, multiple researches focus on detecting fake news based on adversarial learning [14], user response generating [15], semi-supervised detection [16], supervised and unsupervised detection [17] and other methods also. The authors in [18] describe a complete review of the fake news detection process in social media platform, including fake news characterization on psychological theories. All the prospective was related to fake news detection solved through multiple machine learning approaches. In [19], the authors proposed a model based on two machine learning approaches to detect hoaxes or not-hoaxes news spread over the social network platform like Facebook. But the detection of the content analyzes based on users like or shared contents. In [20], the authors describe two different approaches for detecting fake news. Both of these approaches applied simultaneously for faster and secure detection of Fake news. The major categories applied by the authors are linguistic cue approaches using machine learning and network study approach. Multiple approaches applied to detect three types of fake news like serious reporting, based on their pros and cons and predictive modeling study in multiple social network platforms. The study identified multiple posts, shared contents and news content in the form of audio, video, and texts. By analyzing multiple posts of the user authors in [21] describes the detection of the fake news in social platform based on the serious reporting, their pros and cons, text analytics and multiple predictive modeling. In [22] authors also describe the detection of fake news based on a novel algorithm called DETECTIVE that operates using Bayesian inference. It also employs posterior sampling to trade off exploitation activity. Based on the propagation of content over social network platform, the authors in [23] designed a novel approach called 'Traceminor' to identify fake news spreads over a different network using deep learning classifiers. It also analyzes the behavior of the content if certain information is absent. In [24], the authors analyzes the content of the social network spreads through multiple bots. Study of bots prevents spreading of malicious news over the social platform and also detect the account easily those are involved in spreading. The resultant study says curbing of the social bot is an effective strategy for mitigating the spreads of misinformation over social platform. Based on the corresponding article bodies, a fake news detection system was proposed by the authors in [25]. The model can be applied to detect fake news in the clickbait detection scenario. The accuracy of the detection system scores 89.59. By measuring users trust level authors in [26] proposed a fake news detection system. The author's analyzes representative of both experienced user who can identify fake or malicious information in the form of news. Also, the authors do a comparative study of multiple profiles who disclose their potential to distinguish fake news. Authors in [27], uses news content and social context information to detect fake news. In this study, the author analyzes multiple datasets from a different perspective and differentiate the activities also. Based on the multiple study authors develop one detection system called FakeNewsNET to identify and analyze multiple fake news spread over social platform. Convolution neural network-based study of fake news detection proposed by authors in [28]. By using explicit and latent features into a unified feature space, authors proposed the model called TI-CNN to analyze image and text for fake news study. In [29], the authors proposed a machine learning approach for detecting fake news by analyzing multiple features extracted from different stories including source from where it generates. In this study, authors measure the prediction performance of proposed approaches and design one auto-detection system for fake news detection. An image-based fake news detection system describes by the author in [30]. The authors proposed a scheme and study the performance of multiple images to detect forgery against the image to image translation using compression and ideal condition of the image. A fake news detection model was developed using the n-gram and machine learning approach by the authors in [31]. In this article the author uses multiple features extracted using two different methods and tested using six different machine learning environments for study. The Term Frequency Inverted Document Frequency (TF–IDF) as feature extraction and Support vector machine (SVM) as machine learning analyzer gives better accuracy as compared to others. An automatic fake news credibility inference model called Fake Detector was developed by the authors in [32] to detect fake news in social network platform. The author analyzes multiple attributes like user profile features, the connection between users and creator of the fake news using deep diffusive neural model to learn the representative of news articles. The author in [33] proposed a graph neural network with continual learning-based approach for detection of fake news on social media platform. In this approach the author uses graph neural network dealing with non-Euclidean data for the analysis. They usually used unseen data for the implementation by avoiding some text content. A fake news detection approach proposed by the author in [34] by analyzing supervised artificial intelligence algorithms in social media account. The author used twenty-three intelligent classification approaches using public data available. The author in [35] used SentiWordNet to consider the cognitive cues of the text to facilitate opinion mining. It further incorporated GRNN (Bidirectional Gated Recurrent Neural Network) objective factors such as sentiment and credibility score to provide fake news detection platform. Furthermore, a novel multi-level voting ensemble model is proposed by the author in [36]. The proposed system has been tested on three datasets using twelve classifiers. These ML classifiers are combined based on their false prediction ratio. It has been observed that the Passive Aggressive, Logistic Regression and Linear Support Vector Classifier (LinearSVC) individually perform best using TF–IDF, CV and HV feature extraction approaches, respectively, based on their performance metrics. Various other approaches are also proposed in the literature to protect social networks [37] and others [38,39] from different attacks [40–42]. Although all the above studies mentioned the study and detection of fake news in the social network environment by using different approaches are effective but certain limitations were found.

- The selected features are not sufficient for fake news study.
- Some of the machine learning approaches was used to detect malicious information in the form of fake news. But, the study in the form of accuracy is very low i.e. detection rate is less.
- The processing time for detection is more and no online detection system available for the user to implement on their homepage.

In summary, the detection mechanism for fake news in the social network platform is a challenging task. However, our proposed model overcomes the issues related to previous studies and implemented the same at user's homepage to detect fake news. Also, our framework uses both machine learning and deep learning to detect malicious information in the form of fake news on Facebook.

3. Proposed framework

Virtual communication platform users affected through fake content shared by various users. In this section, we present a chrome extension based fake news detection approach. Our proposed approach uses machine learning and deep learning-based analysis including various features associated with the user's profile to identify fake news. Our proposed approach identifies the fake news in a reactive manner. Moreover, our approach collects different news disseminated by the Facebook users and various features associated with different profiles to analyze fake activity. Additionally, the resultant output is shown in the form of pop up box in chrome environment at user's homepage. Keeping the above scenario in mind, we argue that the proposed approach is more suitable for detecting fake news on Facebook platform.

3.1. Architectural details

Today, a lot of fake content is spreading on Facebook through its users and it reaches to different legitimate users in the same network or in the different network. Our proposed approach can detect fake news spread by different users like an advertisement, messages, images, and other contents. Also, our approach identifies the content as fake or benign using machine learning and deep learning classifiers. Our proposed approach consists of three different phases as shown in Fig. 2. In phase I, all the information is collected using our crawler and Facebook API through user analyzed features. The preprocessing and refining of the collected dataset are also done in phase 1. In phase II, all the collected information passes through machine learning and deep learning-based classifier for classification. Finally, in phase III, the best suitable classifier is selected for detecting fake news through chrome extension.

3.2. Characteristics analysis based on Facebook users and news content information

We have collected various profiles and their shared content features. Then, we have analyzed features using our chrome extension to classify fake news on Facebook. Based on the user behavior on social network platform, we have collected multiple features that can help and share information to predict malicious content in the form of fake news.

All the collected information from user profile and shared content selected features are elaborated in Tables 1 and 2. Also, various characteristics of fake news based on multiple features are shown in Fig. 3. This figure clearly shows the scope and variety of fake information spread on OSNs platform. The four major components of the image are, creator/spreader, social content, news content and target victim. Again, all major components are subdivided into number of groups on the basis of their use and spreading information.

3.3. Crawler for feature extraction and data collection

In order to compute different features and to extract information from Facebook profile of users, our crawler uses chrome environment. The details of the extraction process through multiple features analysis is shown in Algorithm 1. Our crawler runs over a specific user's account and collects public information available on user's homepage like followers, a number of friends, likes, and other information. Also, our crawler uses Facebook API to extract the user's private information on request.

Algorithm1: Collected dataset based on the extracted feature (UCF features and NCF features)

Input: A list of Facebook profile user

Output: A label data set generation

for each Profile (FB^P) in the crawler list **do**

Pull out FB^P's features using crawler;

Store the extracted feature sets in an EXCEL file;

Based on the available content, Extract new features by using Facebook API and crawler;

for each UCF by multiple profiles i.e. Post shared P_{shared} **do**

Extract set of the crude dataset from multiple features like $\#^{Messages}$, $\#^{Sharedmsg}$ $\#^{Comment}$, $\#^{likes}$, $\#^{Multimedia\ content\ shared}$, $\#^{Follower}$, $\#^{Followings}$ from FB^P as an input parameter to the crawler.

Extracted feature store in EXCEL file.

End

for each NCF by multiple profiles i.e. News Post shared NP_{shared} **do**

Extract set of the crude dataset from multiple features like $\#^{Source}$, $\#^{Bodytext}$, $\#^{Headlines}$, $\#^{Texts}$, $\#^{Images}$, $\#^{Wordcount}$, $\#^{Sentense}$ from FB^P as an input parameter to the crawler

Extracted feature store in EXCEL file.

End

Parse EXCEL files and calculate the total and average values for each user post

Analyze the performance based on the stored result in the file.

End

By analyzing multiple features, our crawler extracts the user information and stores it in XLS file. The XLS file gets automatically updated when the user visits any profile. We have extracted the profile information from multiple profiles including those which are spreading fake news. We have collected more than fifteen thousand news from 5000 different profiles including fake and real news. All the stored information is pre-processed for further operation. We have selected some new features including existing features collected by various researchers. Basically, our new features are profile based features. Based on our analysis, we can say that fake users are spreading malicious information because fake news is used to highlight their profiles and attract other users toward their profile.

3.4. Construction of raw dataset

For detecting fake news on Facebook, we have collected multiple user posts and generate a raw dataset. Based on the selected features our crawler extracts both relevant and irrelevant information from multiple profiles. Therefore, to eradicate irrelevant information, we have applied some filtering techniques. Our crawler extracts news that hampers user reputation, content liked by the users, pornographic content, and recent news in the form of text, images or videos. Also, the crawler extracts some political blogs and recent trends that are trending in society.

3.5. Construction of baseline dataset from the raw dataset

For the detecting the fake news on Facebook, baseline dataset is required for preprocessing. We have constructed baseline dataset from the raw dataset by applying some filtering techniques. Every user and his contents were labeled as a fake or legitimate. Fake news is dispersed by fake users over Facebook through spam content like images with the link, hashtag, URLs, etc. Based on the guideline provided by the service provider and current research we have categorized each profile and their posts into different groups. The details of the extracted dataset depicted in Table 3.

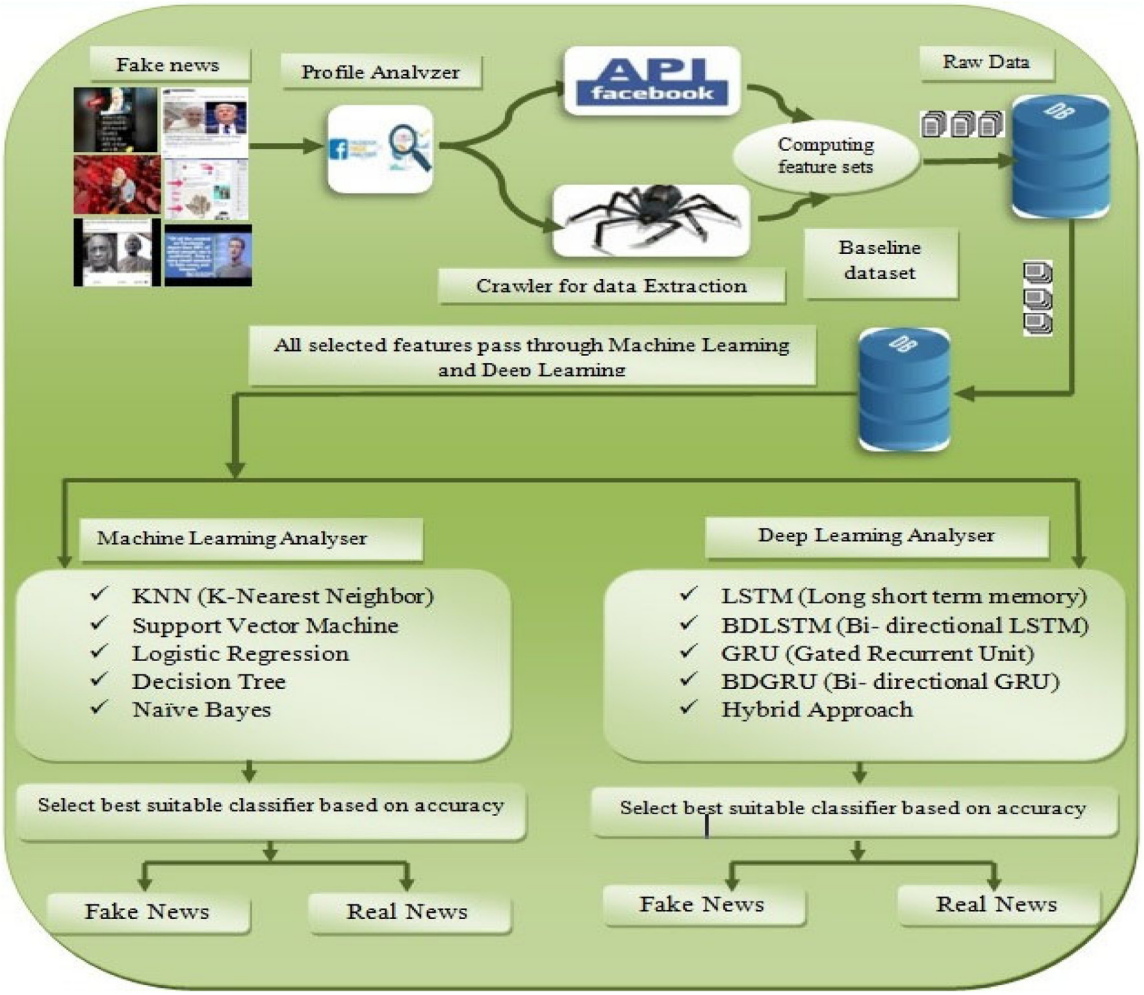

Fig. 2. Framework for Fake news detection.

4. Classification approaches based on machine learning and deep learning

In this section, we discuss the details regarding multiple classifiers which are used for our proposed chrome extension based on fake content detection approach. All of these classifiers process our dataset collected from different Facebook profiles. Descriptions of the various classifiers are listed below.

4.1. Classification techniques based on machine learning

To classify the contents as fake and genuine based on their features and behavior, we have used various machine learning based classification approaches described below:

KNN (K Nearest Neighbors) — It is a simple classification algorithm that stores all available conditions and classifies a new condition based on the similarity measure i.e. called distance function. In this classification technique, the output is a class membership. An object is classified based on the popularity index of its neighbor including the assigned object to the class common among its neighbor. Basically, the distance measured through Euclidean, Manhattan, and Minkowski functions. Also, all the above distance measures are only for continuous variables but for categorical variables hamming distance used. It basically brings up the issues related to standardization of numerical values between zero and one when the dataset contains both numerical and categorical variables.

SVM (Support Vector Machine) is a discriminative classifier describes with a separating hyper plane. By providing labeled data, the algorithm produced an optimal hyper plane which distinguishes the input content in a better manner. Basically, this hyper plane is dividing the plane into two different parts where each class justifies a certain class of output depicted in Fig. 4. The multiple tuning parameters are used in SVM called regularization parameter and gamma. Both of these parameters are used from nonlinear classification line with more accuracy within a specified time.

Logistic Regression: If the dependent variable is binary (Dichotomous) logistic regression study is appropriate due to the predictive nature of the study. It represented based on an equation with certain input parameters. The multiple input parameter (X) are combined using multiple weights linearly of using some coefficient values to predict an output called Y. The basic objective of this classifier is to model the output as a binary value zero (0) or one (1) based on the depicted equation.

$$Y = e^{(b0+b1*X)}/(1 + e^{(b0+b1*X)}) \quad (1)$$

where, Y = predictive output

b_0 = intercept term or bias

b_1 = Coefficient for single input values (X)

Decision tree: Decision tree classifier built the classification and regression model like a tree structure. All the data set divided into a number of smaller subsets and smaller subsets are associated to form a decision tree in an incremental way. The final

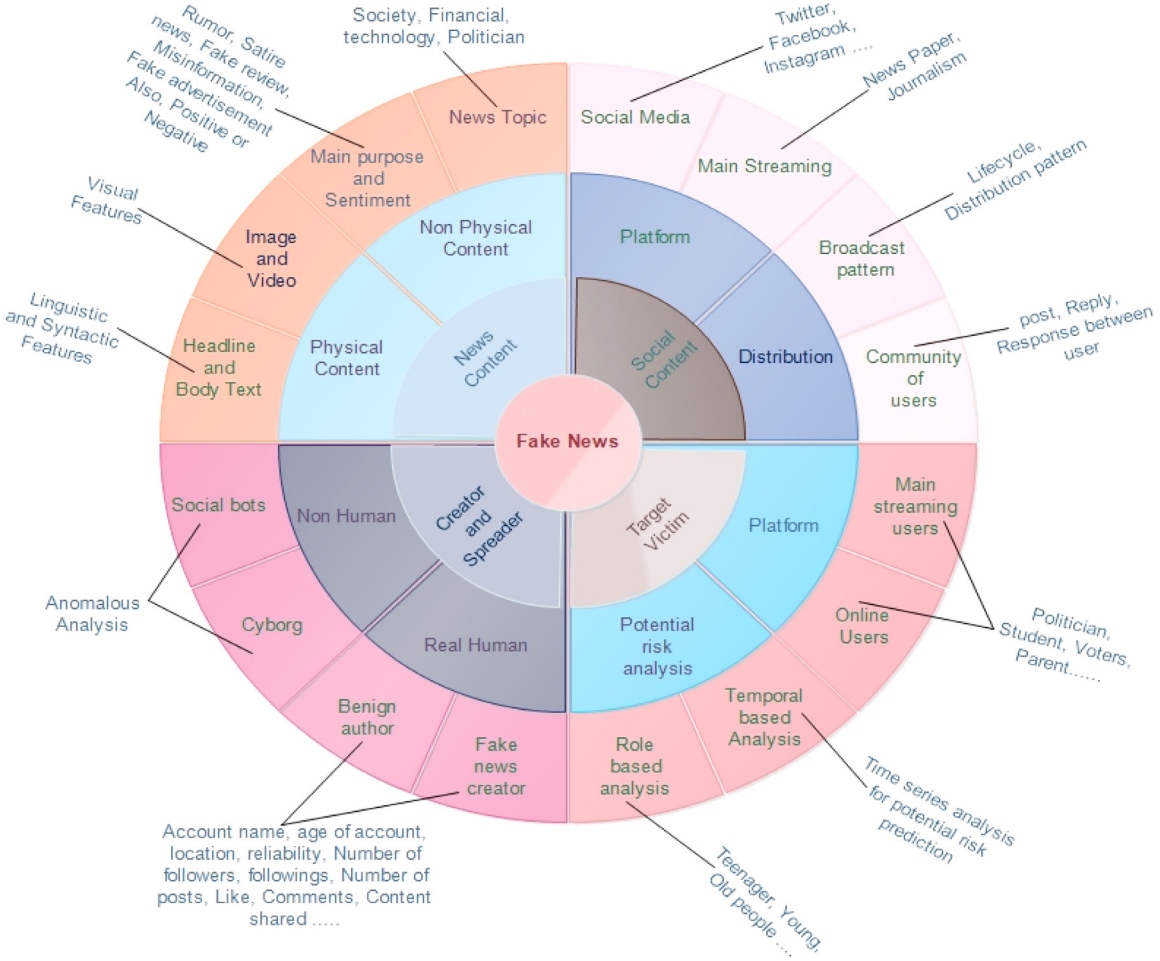

Fig. 3. Multiple characteristics of Fake news.

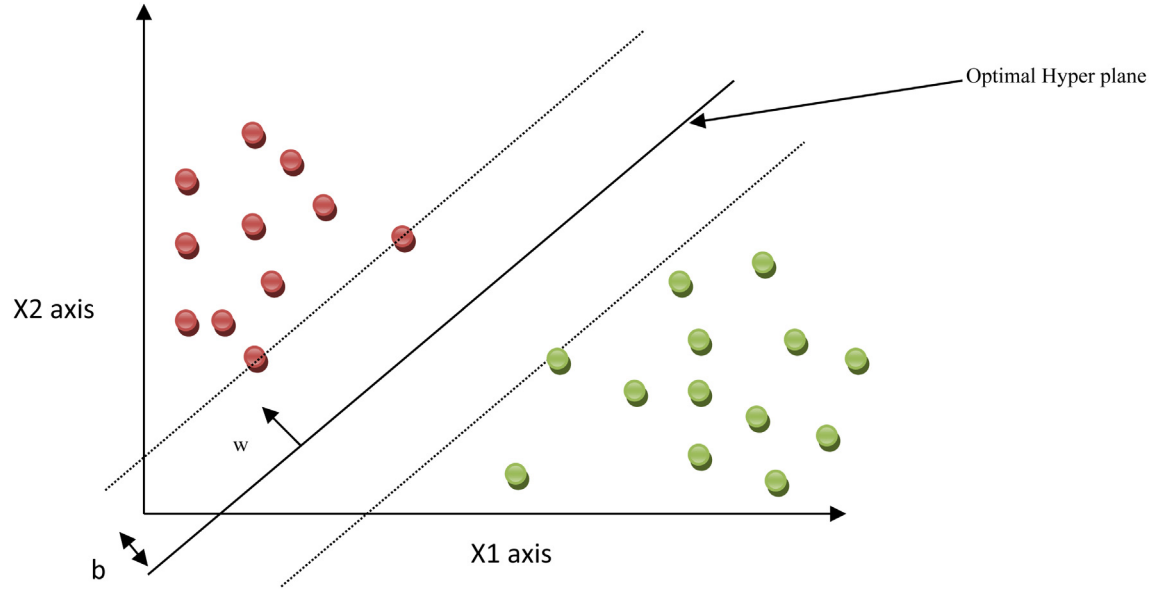

Fig. 4. SVM (Support Vector Machine).

result is the combination with decision node and tree. The decision node has two outputs in the form of yes (1) or no (0). The leaf node shows multiple decisions in the form of classification. The core algorithm related to decision tree is ID3. It uses the principle of Entropy and Information Gain to build the decision tree. The entropy is used to calculate the homogeneity of a sample. If the sample matches properly it shows the output as zero (0) and if the sample has equally divided it has the output as one (1). The

Table 1
User content features for study.

Feature No.	Feature	Description	Feature No.	Feature	Description
UCF^{1}	Profile ID	Every profile has one unique profile ID.	UCF^{10}	Number of following	Follow the people by your liking.
UCF^{2}	Profile name	Profile name of the user varies based on users choice	UCF^{11}	Number of events	It shows the number of events user participated.
UCF^{3}	Date of join	It describes how old the profile is.	UCF^{12}	Number of post shared (image, text, video)	It shows the number of content shared by the user over social network platform.
UCF^{4}	All friends	Total number of friends of the user.	UCF^{13}	Number of URL shared	URLs are the hyperlinks shared by the user with different posts.
UCF^{5}	Profile picture	It shows pictorial identity of the user.	UCF^{14}	Number of Tags	Tagging shows the pointing to a specific person through multiple posts.
UCF^{6}	Number of group join	Number of different group's user participated.	UCF^{15}	Number of Hashtag	It is a way of sharing clickable links with different users.
UCF^{7}	Number of page like	This feature shows association of the user in different contents.	UCF^{16}	Number of newly added friends	It identifies how many new users are added to the profile recently.
UCF^{8}	News post	It identifies the interest of the user and the multiple events user have participated.	UCF^{17}	Recent post like or shared	It shows the recent user interest on multiple posts.
UCF^{9}	Profile with photo guard	To protect the image from unauthorized user, user uses safety principle as photo guard.	UCF^{18}	Present location	User frequently update their location and shared in social network to show their presence
UCF^{19}	Number of stories shared	User shared the multiple events as story in their profile to view by multiple users.	UCF^{20}	Messages with spam words	User share their content with spam words to affect other user.

Table 2
News content features for study.

Feature No.	Feature	Description	Feature No.	Feature	Description
NCF^{1}	Source	It describes the author or publisher of the news article.	NCF^{7}	Linguistics based(chapter, word, sentence, document, quoted word, external link, etc.)	These features capture the different writing style and sensational headlines to identify fake news.
NCF^{2}	Headline	It describes main highlight of the topic and catch the reader's attention.	NCF^{8}		
NCF^{3}	Body text	Text that elaborate the detail story of the topic. It also highlights the angle of publisher.	NCF^{9}		
NCF^{4}	Text	It highlights the story as textual representation i.e. in readable format	NCF^{10}	Statistical features (Count, image ratio, Multi image ratio, Hot image ratio, Short image ratio)	It shows the statistical study of the content that shared over the network including image, video and other information.
NCF^{5}	Images (Image with text, image with hyperlink)	Content of the shared news article that provides visual description about that activities or events. It also posted by some user including caption as text and links.	NCF^{11}	Images (Clarity source, Coherence, similarity distribution, Diversity source, Clustering Score)	It describes the overall statistical study of fake news spread through images.
NCF^{6}	Videos		NCF^{12}	Date of post	It describes the posted date of the content.

below-mentioned equation shows the entropy with two different attributes for our data set.

$$\mathrm{E}\,(\mathrm{T}, \mathrm{X}) = \sum_{c \epsilon X} P\,(c)\, E(c) \tag{2}$$

Also, the decision tree algorithm depends on Information Gain. Information Gain varies with the study of entropy. To construct the decision tree, finding the attribute that produces the highest information gain is required. The final result of Information Gain based on the entropy depicted in equation……

$$\mathrm{GAIN}\,(\mathrm{T}, \mathrm{X}) = \mathrm{Entropy}\,(\mathrm{T}) - \mathrm{Entropy}\,(\mathrm{T}, \mathrm{X}) \tag{3}$$

Naïve Bayes: Naïve Bayes classifier is a machine learning classification model based on probabilistic principle. The principle of

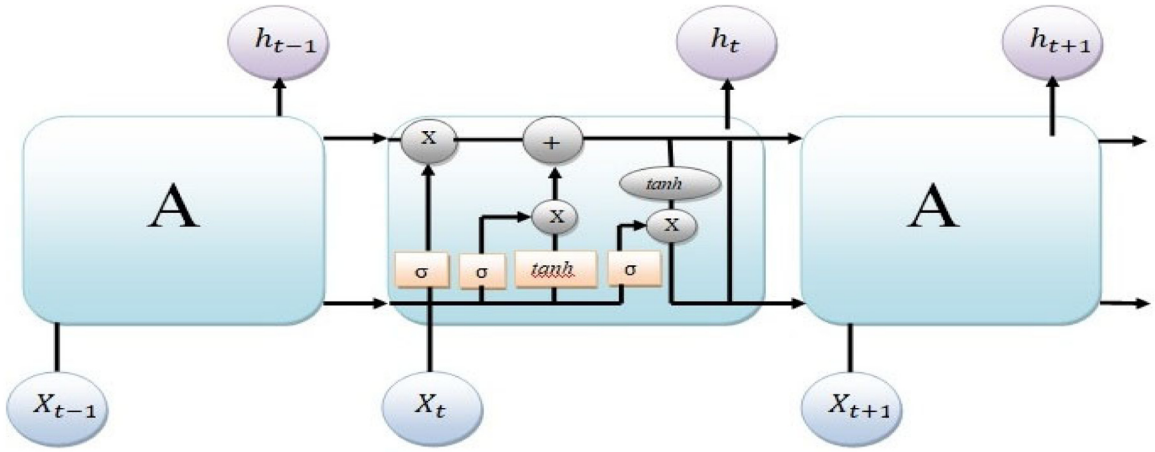

Fig. 5. Working procedure of LSTM.

Table 3
Details of generated dataset for fake news detection approach by our crawler.

Profile parameters (Fake and legitimate)	Total number of information related to multiple profiles (Facebook)
Total profiles	5026
Total news collected	15 328
Total number of post	42 256 893
Total number followers	3 685 643
Total number of followings	2 354 782
Total number of likes by the user	16 236 669
Total umber of listed count	67 976
Total URL's shared	2609

this classifier is based on Bayes theorem depicted in equation ….. . Our prediction of the result based on our dataset depends on multinomial class based classifier for classifying fake news from real one in a social network platform.

$$P(A|B) = (P(B|A) * P(A))/P(B) \tag{4}$$

4.2. Deep learning-based classification techniques

As compared with machine learning classifiers, deep learning classifiers use its own data processing for learning. The basic definition of deep learning says, it is a kind of machine learning that achieves more power and flexibility by learning to represent the world as a nested hierarchy of concepts with each concept related to the simpler one with abstract representation. It uses the principle of neural networks with a large number of parameters called layers in different architecture called unsupervised pre-trained network, convolution network, recurrent network and recursive neural network. For our proposed approach, we have used LSTM (Long Short term memory). The details of the LSTM are depicted below.

LSTM (Long Short Term Memory): LSTM is a special kind of recurrence neural network handle long term dependencies. The main objective of LSTM is to remembering information for a longer period of time. The basic flow control of information in LSTM describes in Fig. 5.

The main flow of information in LSTM based on cell state the horizontal line in the top of the figure. It behaves like a conveyor belt because of the flow of nature in cell state is straight down. The operation process of LSTM does not have the property to remove and add information to the cell state. For the flow of information in-between LSTM, it uses multiple gates. Gates are composed of sigmoid function and pointwise multiplication operators. What the actual information we are going to throw is the first step of LSTM. All these decisions are made by the sigmoid layer. It is also called forget gate. The information through forget gate is in the form of h_{t-1} and x_t. The output of the forget gate is either zero (0) or one (1) for each number in the cell state. The operation 1 and 0 says either completely keep it or completely get rid of this. The complete operation related to the flow of information based on the multiple sigmoid value describes with the equations are as follows.

$$ft = \sigma\ (Wf.[ht - 1, xt] + bf) \tag{5}$$

$$i_t = \sigma\ (W_i.[h_{t-1}, x_t] + b_i) \tag{6}$$

$$C'_t = \tanh\ (W_C.[h_{t-1}, x_t] + b_C) \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * C'_t \tag{8}$$

$$o_t = \sigma\ (W_o.[h_{t-1}, x_t] + b_o) \tag{9}$$

where, ft = Forget state, i_t = Input state , C'_t = Intermediate state, C_t = Cell state and o_t = Output state.

The complete output of the LSTM operation is based on the cell state with filtered output. First, the sigmoid layer decides what part of the cell state be the output. Then the overall process passes through tanh and multiplying with the output of the sigmoid gate to get the decided part output.

5. Experimental evaluation based on collected features

Processing of multiple features to detect fake news in the machine learning environment may not predict a better accuracy on Facebook platform. Hence, we have processed the multiple features in machine learning algorithms as well as in deep learning algorithms for better analysis of fake news. The selection of features on the basis of user's profile and shared content provides a better input to our approach to predict the decision. A set of new features are extracted by our crawler and Facebook API. Also, we have evaluated the performance of multiple posts extracted from different Facebook profiles on the basis of selected and some existing features for processing. We have evaluated news content and user content features separately for detecting fake news. But, the detection rate for individual features is less when compared to a group of features i.e. combination of both news content and user content features. To train and test the dataset, we have uses 10 fold cross-validation techniques for better identification and analysis. The objective of our approach is how accurately our multiple features-based approach distinguishes the fake news from real news in the online environment on the user's home

Table 4
Confusion matrix for fake news detection framework.

Accuracy		Classification in the form of actual output	
		Fake News	Real News
Classification in the form of predictive output	Fake news	True positive	False negative
	Real news	False positive	True negative

Table 5
Analysis of features based on its category and classification techniques.

	Features	Different measures	Multiple classifiers				
			KNN	SVM	Logistic regression	Decision tree	Naïve Bayes
Study of features based on its category and combining approach	Users profile content features	TP Rate	0.941	0.962	0.928	0.943	0.948
		FP Rate	0.005	0.003	0.007	0.006	0.005
		Precision	0.938	0.961	0.921	0.940	0.943
		Recall	0.940	0.958	0.926	0.938	0.947
		F measure	0.940	0.951	0.926	0.938	0.945
		MCC	0.928	0.960	0.920	0.931	0.941
		ROC Area	0.938	0.959	0.926	0.934	0.941
		PRC Area	0.935	0.957	0.925	0.937	0.948
	News content features	TP Rate	0.923	0.939	0.928	0.930	0.936
		FP Rate	0.007	0.006	0.007	0.006	0.006
		Precision	0.930	0.935	0.921	0.929	0.934
		Recall	0.929	0.932	0.929	0.927	0.932
		F measure	0.926	0.939	0.926	0.927	0.932
		MCC	0.926	0.937	0.928	0.928	0.931
		ROC Area	0.928	0.939	0.927	0.921	0.937
		PRC Area	0.929	0.931	0.928	0.924	0.931
	(Users profile features) + (news content features)	TP Rate	0.993	0.993	0.990	0.991	0.986
		FP Rate	0.004	0.003	0.007	0.006	0.06
		Precision	0.994	0.994	0.991	0.991	0.978
		Recall	0.992	0.993	0.991	0.990	0.986
		F measure	0.991	0.990	0.993	0.992	0.982
		MCC	0.990	0.989	0.989	0.983	0.960
		ROC Area	0.990	1.000	0.993	0.990	0.998
		PRC Area	0.993	1.000	0.992	0.991	0.996

Table 6
Accuracy of different machine learning and deep learning classifiers.

Features	Measure in %	Machine learning classifiers					Deep learning classifiers
		KNN	SVM	Logistic Regression	Decision tree	Naïve Bayes	LSTM
Users profile content features	Accuracy	94.1	96.2	92.8	94.3	94.8	96.3
News content features	Accuracy	92.3	93.9	92.8	93.0	93.6	91.1
Users profile features + News content features	Accuracy	99.3	99.3	99.0	99.1	98.6	99.4

page. After a successful operation in multiple machine learning and deep learning algorithms which are using our dataset, the result in the form of confusion matrix depicted in Table 4.

The predicted results through confusion matrix also measure various parameters like recall (Sensitivity), true positive rate (TPR), true negative rate (Specificity), ROC (Receiver operating curve), etc. The performance results of different parameter are depicted in Table 5. We have analyzed the performance of different categories of user profiles, news content features and combination of both. We have also compared results and pick the best suitable results for our chrome extension that can run on user's home page. As compared to machine learning, deep learning-based analysis gives better accuracy and detection rate. In machine learning environment, SVM (Support vector machine) gives better accuracy compared to KNN, Logistic Regression, Decision Tree, and Naïve Bayes classifiers.

We have observed that, the combination of features based on users' content and news content gives better accuracy as compared to other features in machine learning and in deep learning environment. From our analysis we can see that the number of messages with hashtags, URL, and picture with multiple captions contain more fake news as compared to other posts. The entire classification tasks with parameters like accuracy are shown in Fig. 6. Also, the analysis of multiple features based on their performance is shown in Figs. 6a to 6b. The TP rate in KNN and SVM are more as compared to other classification approaches i.e. 99.3%. Also, we have found the accuracy of multiple classifiers on the basis of confusion matrix and compared with that of deep learning-based analysis for better decision making. The analysis of accuracy in different classification techniques and in deep learning-based analysis is shown in Table 6. We have observed better accuracy in deep learning-based analysis i.e. in LSTM due to long time storage of previous results. As compared to SVM in machine learning, LSTM gives better accuracy of 99.42% when we have considered user profile and news content features.

Our approach predicts better results in deep learning environment as compared to machine learning environment using both user profile and fake news features. The graphical analysis of the results is shown in Fig. 7.

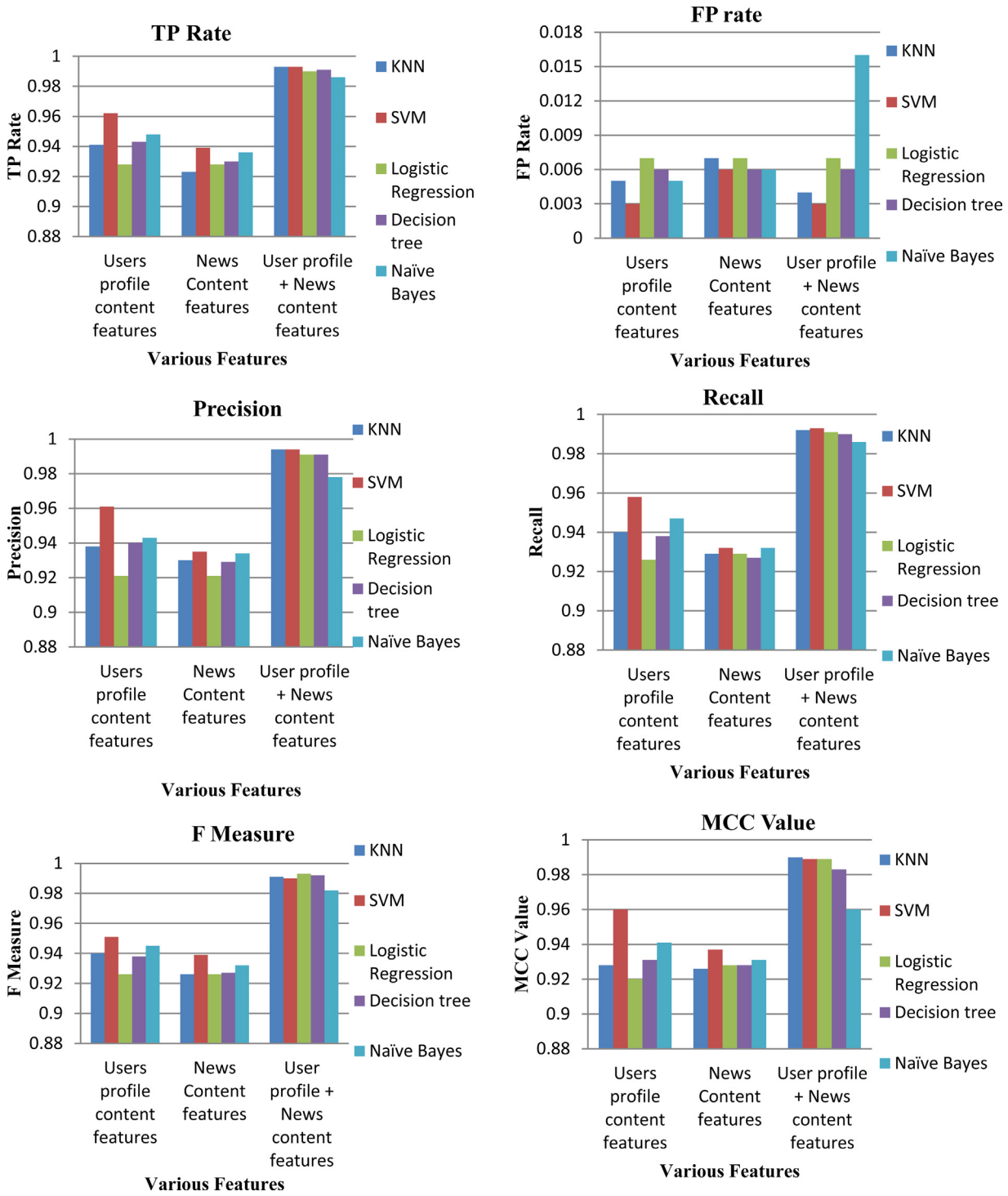

Fig. 6a. Study of multiple features in different classifiers.

6. Chrome extension-based auto detection system

After analyzing the performance of multiple features in machine learning and deep learning based analyzer, we processed the selected content for detecting fake news. This approach works as a chrome extension which gets triggered by the user from its home page and it extracts multiple features from different profiles for analysis and self-learning. This extension uses HTML and CSS as a front-end and JavaScript as the back-end. It combines certain files and display the results as a pop-up to user using popup.html. To defile the extension name, version, files and the script manifest.json file is used. When the user clicks on the extension, two files are loaded: first, popup.html which displays the pop-up window for result; second, popup.js which contains JavaScript code that generates .csv file. User initiates the fake content detection after selecting the text which needs to be analyzed. Fig. 8 demonstrates the working of extension for fake content detection. Firstly, user starts the detection process by clicking on the extension and it starts extracting the content from user's profile for analysis. The selected features are used to

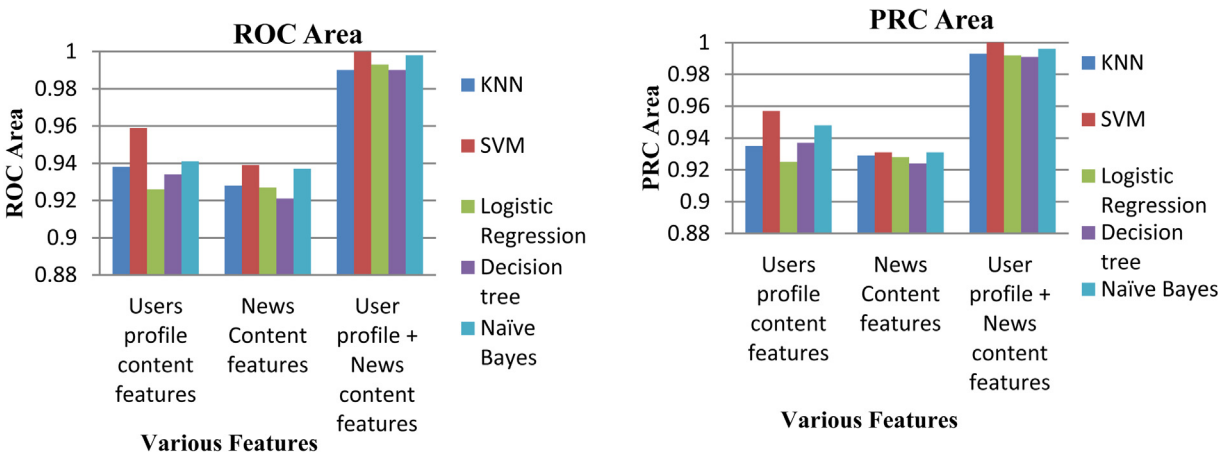

Fig. 6b. Study of multiple features in different classifiers.

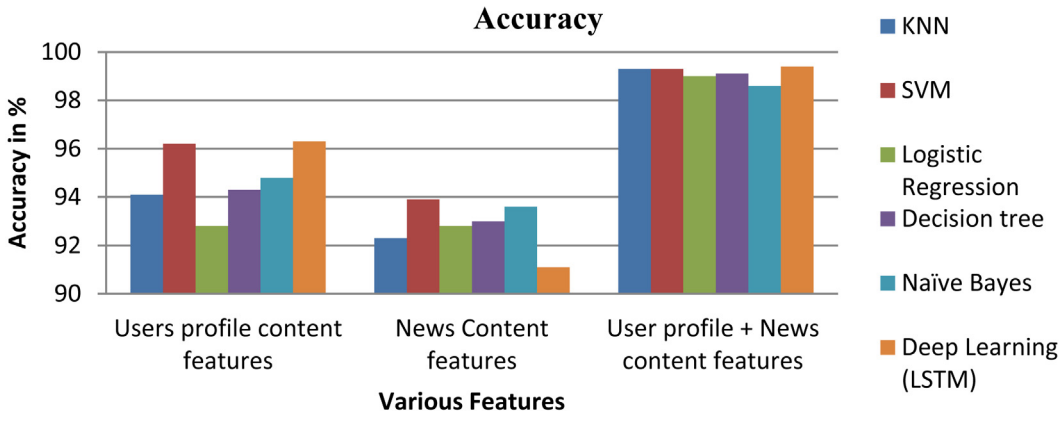

Fig. 7. Comparison of accuracy in different classification techniques.

train the extension. At server side, our approach train different algorithms with the extracted user content and news content features. The proposed approach automatically updates the content of the user's home page after every click. The Predicted Output and Decision Making in Chrome Extension on the basis of feature analysis depicted in Algorithm 2 and Algorithm 3. After analyzing the content, it displays the results as depicted in Figs. 9a and 9b.

Algorithm 2: Predict using classifiers
Input: A list of User profile feature and news content features from the dataset (UCF^1, UCF^2…….UCF^N And NCF^1, NCF^2, ……NCF^N)
Output: Predicted output in the form of Fake news or real
Dataset← Read_Dataset (File_Path)
Model←Multiple classifier( )
N← Fatures_List [0, 1……N-2]
Profile-Type ←Features_List [1……N-1]
Model.Fit (Features, Profile_Type)
Predict_Output = Model.Predict(Features_List)

Algorithm 3: Predictor for decision maker as Fake news or real news
Input: List of Features provided to model
Output: Decision support system in the form of value
Predict ← Predict_using_classifier (Features_List)
then
If(Predict ==1)
Value ← "The content related to Fake news"
else
Value ← "The content related to real news"

7. Comparative analysis of our approach with other approaches

Our proposed approach analyzes various Facebook posts shared by the users on social network platform to detect fake news. Our approach reveals fast rate of detection as compared to other approaches as depicted in Table 7.

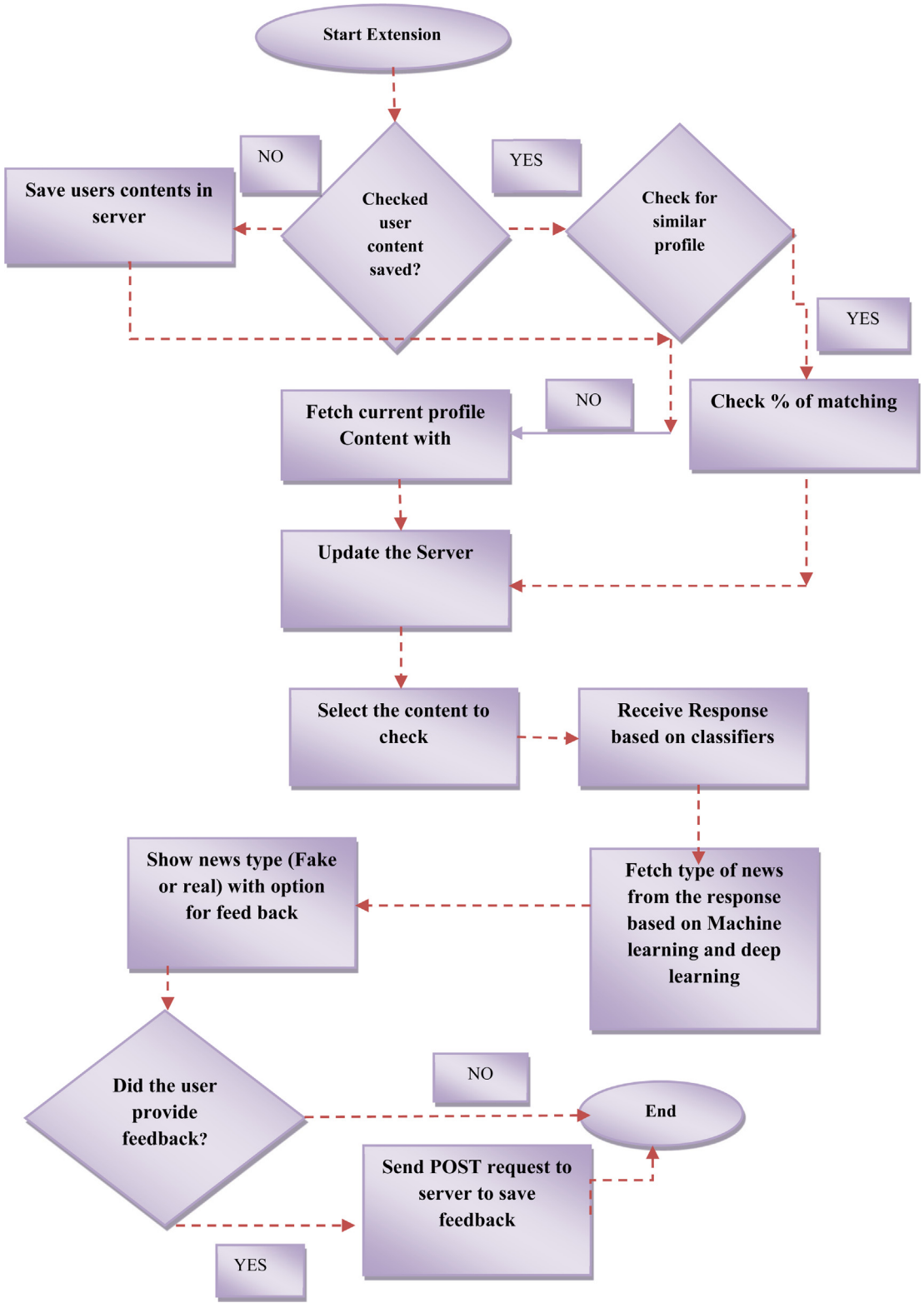

Fig. 8. Flow control of fake news detection system.

8. Conclusion and future work

In this article, we have proposed a fake news detection approach for Facebook users using machine learning and deep learning classifiers in chrome environment. Our approach analyzes both user profile and news content features. In our proposed work, we have developed a chrome extension that uses crawled data extracted by our crawler. Also, to boost up the performance of chrome extension, we have used deep learning algorithm

Fig. 9a. Study of fake news based on KNN, SVM and Logistic regression.

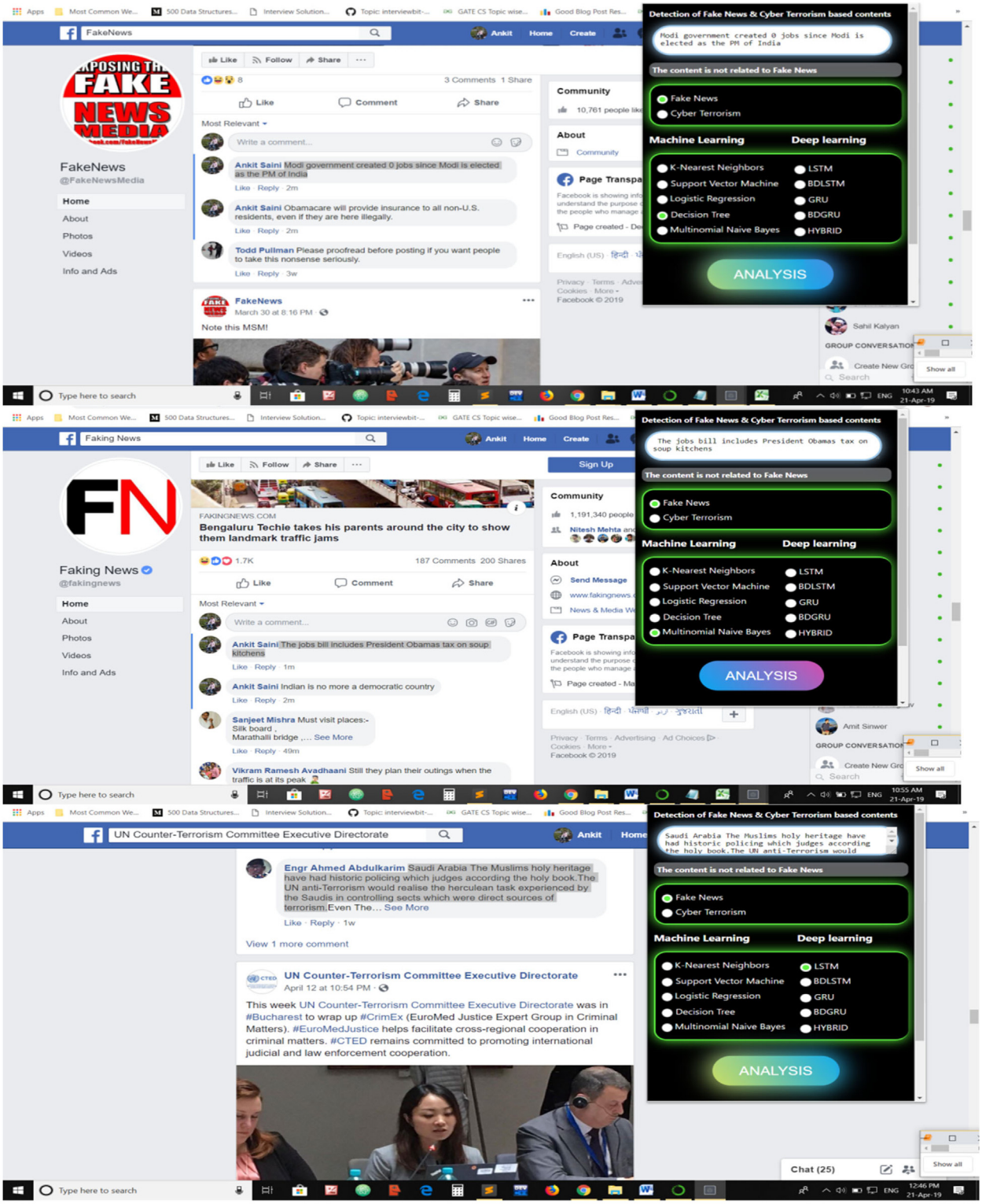

Fig. 9b. Study of fake news based on Decision tree, Naïve Bayes and Deep learning (LSTM).

called Long Short-Term Memory. Our deep learning algorithm achieved excellent performance with 99.4% accuracy as compared to machine learning algorithms. The experimental results revealed that the approach successfully deployed at user side on chrome environment to detect the fake news in real time by analyzing user information and shared posts. As future work, it can be enhanced by analyzing the features using other deep learning algorithms like bidirectional LSTM, bidirectional GRU and some hybrid approach based deep learning classifier for

Table 7
Comparative analysis of our approach with other approaches.

Approach	Description	Dataset used	Results	Benefits	Limitations
Ahmad et al. 2017 [31]	Fake news detection approach was developed using n-gram and machine learning approaches. Features are extracted using two different methods and tested using six different machine learning algorithm for analysis.	12 600 fake news article and 12 600 real news articles.	Accuracy in SVM = 85%, LSTM = 92% and in KNN = 77%	This approach used n-gram model to detect automatically fake news by analyzing fake reviews.	Only user reviews are used for fake news analysis. Accuracy is low due to limited number of features are used for analysis.
Reis et al. 2019 [29]	Machine learning approach for detecting fake news by analyzing multiple features extracted from different stories including source from where it generates. Also measure the prediction performance of proposed approaches to design auto detection system.	2282 BuzzFeed news related to US election.	Accuracy in RF = 85% KNN = 80% SVM = 79%	Detection of fake news in social platform by using media related content and user profile contents. Various classification techniques are used including KNN and SVM.	Accuracy for detecting fake account is very due to small dataset.
Yang et al. 2018 [28]	using explicit and latent features into a unified feature space authors proposed the approach called TI-CNN to analyze image and text for fake news analysis	using explicit and latent features into a unified feature space authors proposed the approach called TI-CNN to analyze image and text for fake news analysis	using explicit and latent features into a unified feature space authors proposed the approach called TI-CNN to analyze image and text for fake news analysis	using explicit and latent features into a unified feature space authors proposed the approach called TI-CNN to analyze image and text for fake news analysis	using explicit and latent features into a unified feature space authors proposed the approach called TI-CNN to analyze image and text for fake news analysis
Shu et al. 2018 [27]	Based on the multiple analysis authors develop one detection system called FakeNewsNET using Convolution neural network to analyze fake news.	15 257 Buzzfeed news content and 23 865 politifact news content including fake and real.	Accuracy is 92% for Buzzfeed news and 93.6 for politifact news	Detection of fake news based on the news content, social context and dynamic information leads to better analyzes fake news.	CNN classification better analyzes the image content as compared to text advertisement. Accuracy is less due to a smaller number of features.
Our proposed approach	Our proposed approach incorporates the behavior of multiple features associated with Facebook account and analyzes the behavior of those accounts through machine learning and deep learning-based classifiers.	More than 15 000 news contents from different Facebook users including both fake and real news.	Accuracy KNN = 99.3 SVM = 99.3 Logistic regression = 99.0 Decision tree = 99.1 LSTM = 99.4	Proposed approach analyzes both user content and news content features for analyzing fake news on Facebook.	Deep learning algorithm takes more time to train and test the approach as compared to machine learning algorithm.

better decision making with more datasets. Also, we can use boost up the performance.

CRediT authorship contribution statement

Somya Ranjan Sahoo: Conceptualization, Methodology, Evaluation, Writing - original draft. **B.B. Gupta:** Visualization, Investigation, Software, Validation, Writing - review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research work is being supported by sponsored project grant (i) YFRF, under the project Visvesvaraya PhD Scheme of Ministry of Electronics & Information Technology, Government of India and being implemented by Digital India Corporation and (ii) project grant (SB/FTP/ETA-131/2014) from SERB, DST, Government of India.

References

[1] M. Balmas, When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism, Commun. Res. 41 (3) (2014) 430–454.
[2] S. Kaushik, C. Gandhi, Ensure hierarchal identity based data security in cloud environment, Int. J. Cloud Appl. Comput. (IJCAC) 9 (4) (2019) 21–36.
[3] S.R. Sahoo, B.B. Gupta, Classification of multiple attacks and their defence mechanism in online social networks: a survey, Enterprise Inf. Syst. 13 (6) (2019) 832–864.
[4] C. Li, Z. Zhang, L. Zhang, A novel authorization scheme for multimedia social networks under cloud storage method by using MA-CP-ABE, Int. J. Cloud Appl. Comput. (IJCAC) 8 (3) (2018) 32–47.
[5] Song Feng, Ritwik Banerjee, Yejin Choi, Syntactic stylometry for deception detection, in: ACL'12.
[6] B.D. Horne, S. Adali, This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news, 2017, ArXiv eprints.
[7] S.R. Sahoo, B.B. Gupta, Hybrid approach for detection of malicious profiles in twitter, Comput. Electr. Eng. 76 (2019) 65–81.
[8] H. Ahmed, Detecting Opinion Spam and Fake News using N-Gram Study and Semantic Similarity (Ph.D. thesis), 2017.
[9] M. Barthel, A. Mitchell, J. Holcomb, Many Americans believe fake news is sowing confusion, Pew Res. Center 15 (12) (2016).
[10] E. Bakshy, S. Messing, L.A. Adamic, Exposure to ideologically diverse news and opinion on Facebook, Science 348 (6239) (2015) 1130–1132.

[11] C.R. Sunstein, On Rumors: How Falsehoods Spread, Why We Believe Them, and What Can Be Done, Princeton University Press, 2014.
[12] Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, Jiliang Tang, Multi- Source Multi-Class Fake News Detection. in: COLING, 2018.
[13] Seyedmehdi Hosseinimotlagh, Evangelos E. Papalexakis, Unsupervised content-based identification of fake news articles with tensor decomposition ensembles, 2018.
[14] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, J. 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, 2018, pp. 849–857.
[15] F. Qian, C. Gong, K. Sharma, Y. Liu, Neural user response generator: fake news detection with collective user intelligence. in: IJCAI, Vol. 18, 2018, pp. 3834–3840.
[16] Gisel Bastidas Guacho, Sara Abdali, Neil Shah, Evangelos E. Papalexakis, Semi-supervised content-based detection of misinformation via tensor embeddings. in: ASONAM, 2018.
[17] Seyedmehdi Hosseinimotlagh, Evangelos E. Papalexakis, Unsupervised content-based identification of fake news articles with tensor decomposition ensembles, 2018.
[18] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD Explor. Newsl. 19 (1) (2017) 22–36.
[19] E. Tacchini, G. Ballarin, M.L. Della Vedova, S. Moret, L. de Alfaro, Some like it hoax: Automated fake news detection in social networks, 2017, arXiv preprint arXiv:1704.07506.
[20] N.J. Conroy, V.L. Rubin, Y. Chen, Automatic deception detection: Methods for finding fake news, Proc. Assoc. Inf. Sci. Technol. 52 (1) (2015) 1–4.
[21] V.L. Rubin, Y. Chen, N.J. Conroy, Deception detection for news: three types of fakes, in: Proceedings of the 78th ASIS & T Annual Meeting: Information Science with Impact: Research in and for the Community, American Society for Information Science, 2015, p. 83.
[22] S. Tschiatschek, A. Singla, M. Gomez Rodriguez, A. Merchant, A. 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, 2018, pp. 517–524.
[23] L. Wu, H. Liu, Tracing fake-news footprints: Characterizing social media messages by how they propagate, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, ACM, 2018, pp. 637–645.
[24] C. Shao, G.L. Ciampaglia, O. Varol, A. Flammini, F. Menczer, The spread of fake news by social bots, 2017, pp. 96–104, arXiv preprint arXiv:1707.07592.
[25] P. Bourgonje, J.M. Schneider, G. Rehm, From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles, in: Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, 2017, pp. 84–89.
[26] K. Shu, S. Wang, H. Liu, Understanding user profiles on social media for fake news detection, in: 2018 IEEE Conference on Multimedia Information Processing and Retrieval, MIPR, IEEE, 2018, pp. 430–435.
[27] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media, 2018, arXiv preprint arXiv:1809.01286.
[28] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, P.S. Yu, TI-CNN: Convolutional neural networks for fake news detection, 2018, arXiv preprint arXiv:1806.00749.
[29] J.C. Reis, A. Correia, F. Murai, A. Veloso, F. Benevenuto, E. Cambria, Supervised learning for fake news detection, IEEE Intell. Syst. 34 (2) (2019) 76–81.
[30] F. Marra, D. Gragnaniello, D. Cozzolino, L. Verdoliva, Detection of GAN-generated fake images over social networks, in: 2018 IEEE Conference on Multimedia Information Processing and Retrieval, MIPR, IEEE, 2018, pp. 384–389.
[31] H. Ahmed, I. Traore, S. Saad, Detection of online fake news using n-gram study and machine learning techniques, in: International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, Springer, Cham, 2017, pp. 127–138.
[32] J. Zhang, L. Cui, Y. Fu, F.B. Gouza, Fake news detection with deep diffusive network model, 2018, arXiv preprint arXiv:1805.08751.
[33] Y. Han, S. Karunasekera, C. Leckie, Graph neural networks with continual learning for fake news detection from social media, 2020, arXiv preprint arXiv:2007.03316.
[34] F.A. Ozbay, B. Alatas, Fake news detection within online social media using supervised artificial intelligence algorithms, Physica A 540 (2020) 123174.
[35] V. Sabeeh, M. Zohdy, A. Mollah, R. Al Bashaireh, Fake news detection on social media using deep learning and semantic knowledge sources, Int. J. Comput. Sci. Inf. Secur. (IJCSIS) 18 (2) (2020).
[36] S. Kaur, P. Kumar, P. Kumaraguru, Automating fake news detection system using multi-level voting model, Soft Comput. 24 (12) (2020) 9049–9069.
[37] S. Gupta, N. Gugulothu, Secure NoSQL for the social networking and e-commerce based bigdata applications deployed in cloud, Int. J. Cloud Appl. Comput. (IJCAC) 8 (2) (2018) 113–129.
[38] A. Tewari, B.B. Gupta, Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework, Future Gener. Comput. Syst. 108 (2020) 909–920.
[39] Y. Huang, B. Li, Z. Liu, J. Li, S.M. Yiu, T. Baker, et al., ThinORAM: Towards practical oblivious data access in fog computing environment, IEEE Trans. Serv. Comput. (2019).
[40] M.A. Alsmirat, F. Al-Alem, M. Al-Ayyoub, Y. Jararweh, et al., Impact of digital fingerprint image quality on the fingerprint recognition accuracy, Multimedia Tools Appl. 78 (3) (2019) 3649–3688.
[41] Z.A. Al-Sharif, M.I. Al-Saleh, L.M. Alawneh, Y.I. Jararweh, et al., Live forensics of software attacks on cyber–physical systems, Future Gener. Comput. Syst. 108 (2020) 1217–1229.
[42] D. Li, L. Deng, et al., A novel CNN based security guaranteed image watermarking generation scenario for smart city applications, Inform. Sci. 479 (2019) 432–447.