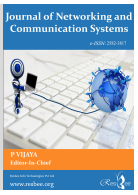
Vol.4 No.4 Oct 2021

Journal of Networking and Communication Systems

Received 17 August, Revised 15 September, Accepted 29 October



Fractional order Fish Migration Optimization Algorithm for Spam Detection in Twitter Data Stream

Shireen Meher

254 Clarendon Park, Leicester, United Kingdom

Abstract: One of the dangerous death-causing diseases called cancer can be cured by exploiting a renowned technique called chemotherapy. Social network sites continue to increase their popularity because of the high practice of the internet. By exploiting social media such as Twitter as well as Facebook the peoples become linked with each other. Due to this, impetuous communication has increased, such as spam and it is used in gathering information in marketing or individual to find out crime over people. On Twitter, spam detection is considered one of the important problems due to the small text as well as the numerous language inconsistencies in social media. Therefore, it is very important to formulate a spam detection technique that pretenses the capability to detect spam messages by exploiting Twitter data. This work formulates a new spam detection approach by exploiting a Twitter data stream. Therefore, the transformation of data is performed on the input data by exploiting the YJ transformation to make the data appropriate for processing. By exploiting the Renyi entropy as well as the DBN feature fusion is carried out. In addition, the detection of spam is carried out by exploiting the Generative Adversarial Network (GAN) that is trained by exploiting the adopted Fractional-order Fish Migration Optimization (Fo-FMO) algorithm.

Keywords: Communication, Data stream, Internet, Social media, Twitter.

Nomenclature

Abbreviations	Descriptions
SVM	Support Vector Machine
I2FELM	Incremental Fuzzy-kernel-regularized Extreme Learning Machine
LSTM	Long Short Term Memory
DT	Decision Tree
YJ	Yeo-Jhonson
RF	Random Forest
XGBoost	eXtreme Gradient Boosting
S3D	Semi-Supervised Spam Detection
ML	Machine Learning
KNN	K-nearest Neighbor
FC	Fractional Calculus
RF	Random Forest
DBN	Deep Belief Network
SOM	Self-Organizing Maps
NN	Neural Network
CNN	Convolutional Neural Networks
RBF	Radial Basis Function
NB	Naive Bayesian
BP	Back Propagation Neural Network
ELM	Extreme Learning Machine

1. Introduction

Daily, over 500 million tweets are posted on Twitter which is not possible to control and filter manually. Therefore, to make the advancement as well as usage of Machine learning recognition as well as the filtering models an imperative requirement. By the exclusively pronounced employ of automation, the task is aggravated that is at present an important module of the mistreatment scene on Twitter. For

https://doi.org/10.46253/jnacs.v4i4.a5

spammers, the maximum click rate, as well as propagation of effectual messages, creates Twitter as a striking platform. Raising the activities of spamming has unfavorably affected the user experience and numerous tasks namely analysis of user behavior as well as recommendations. Numerous conventional researches on Twitter spam concentrate on blocking accounts that are to recognize as well as block spammers as well as spam users. The social graph, as well as user tweets, use, and spammer detection, were devised as an optimization issue [2]. Likewise, from users' tweets, and demographics the information is extracted, which shares the social connection as well as URLs are used as features in general ML approaches to recognize the spam users. Nevertheless, the account blocking technique is considered minimum efficient for spammers who might act as legitimate users by posting non-spam content frequently. Legitimate users might even be hurt by blocking spammers that occur to endowment permissions to a 3rd part application that posts spam tweets in his user name. Even though blocking as well as spammer accounts is considered an important task, tweet level detection of spam is important to struggle over the spamming on a fine-grained level as well as aids to recognizing the spam tweets rather than wait for the users to recognize the spammers. The training data set must be incessantly updated to tackle altering sharing of features in the tweet stream [1].

The procedure via the spammers alters their behavior as well as characteristics to remove the recognition systems. The machine learning recall system generally asymmetrically impacts the population drifts. Other spammers do not radically alter their characteristics, as well as they, stay dependably recognized. To motivate an alteration in the perception of the supervised system, the asymmetrical deterioration is exploited. To discover the spammers rather than the detection, the tools are exploited in the wild [4]. Currently, the challenges in Twitter spam research are feature selection and detection approach selection. Here, the characterization is presented as follows. Predecessor study frequently chooses the indistinguishable type of characteristics in feature selection for instance content-based and user profile-based characteristics for identification. Generally, numerous kinds of social networks' abnormal user characteristics are diverse from those of general users, and it is not sufficient to precisely articulate the data state. Researchers chiefly exploit supervised machine learning approaches to contract with spam detection in social networks in the selection of algorithms. On the basis of the classification scheme, the investigators have modeled the numerical form ideas to recognize the spam users. Here, the supervised Machine Learning approach has been divided into a single classification approach as well as a combined classification approach such as SVM, RBF, meta-classifiers (Decorate, Logit Boost), NB, KNN, BP-NN BP, ELM, DT, RF, and XGBoost [15]. The real dataset of social networks exerts a long tail effect that is it is an unbalanced dataset with several non-spams far beyond the spam [7] [5].

The major objective of this work is to propose a technique to carry out spam detection with Twitter data. By taking into consideration of three stages such as transformation, feature integration as well as spam detection process is performed. At first, Twitter is fed to carry out the data transformation stage in which the Yeo-Jhonson transformation is exploited to transform the data to create it appropriate for additional processing. If the data transformation process is finished, the Twitter data is subjected to the feature integration stage, in which the integration of the feature is accomplished by exploiting Renyi entropy and DBN. Subsequent to the completion of the feature fusion process, the process of spam detection is accomplished by exploiting the developed Fo-FMO-based GAN. Moreover, spam detection is accomplished by exploiting GAN that is trained by exploiting adopted Fo-FMO.

2. Literature Review

In 2020, Nour El-Mawass et al [1] investigated the probability of exploiting the outcome of supervised classification systems for the detection of spammers. Here, the hypothesis systems were exploited to detect the spammers while their recall was away from ideal. Subsequently, the classifier's outcome was proposed in a probabilistic graphical approach. Then the alternative definition based on a bipartite users-content interaction graph was adopted. Also, Markov Random Field was modeled on a graph of similar users and calculates prior beliefs by exploiting a collection of conventional classifiers.

In 2020, Yuliya Kontsewaya et al [2], worked on the spam detection process to minimize the number of spam by exploiting a classifier to recognize it. Here, mainly the machine learning approach was exploited for the most precise classification of spam. To verify an email text a natural language processing method was exploited to recognize the spam. In order to verify the results, the adopted model was compared with the conventional models such as KNN, NB, SVM, DT, Logistic regression, and RF.

In 2021, Ashraf Neisari et al [3], developed a new method to recognize fake reviews from real ones by exploiting the linguistic features. To carry out the classification of the reviews unsupervised learning by means of SOM in combination with CNN was exploited. By arranging semantically-similar words around an image pixel or consistently a SOM grid cell the reviews into images were transformed. The ensuing

review images were therefore subjected to CNN for supervised training as well as subsequent classification.

In 2020, ZHIJIE ZHANG et al [4], worked on spam detection on Twitter, where it was analyzed as user attribute, activity, and content. In this work, a new approach named spam detection technique was modeled on the basis of the regularized ELM, named Improved I2FELM. It was exploited to identify the Twitter spam precisely. The experimentation analysis stated that the adopted approach can proficiently recognize the unbalanced and balanced dataset. In addition, with some characteristics were exploited, the I2FELM efficiently recognized the spam.

In 2020, Surendra Sedhai and Aixin Sun [5], developed an S3D model for the detection of spam at the tweet level. The developed model comprises the two most important frameworks. In the real-time model, the detection of spam framework operates and the framework update module operates in batch mode. The detection of spam framework comprises four lightweight detectors a blacklisted domain detector to label tweets comprising blacklisted URLs; to label tweets, a near-duplicate detector which was near-duplicates of self-assuredly pre-labeled tweets; consistent ham detector to label tweets that were posted by trusted users as well as which do not comprises spammy words; multi-classifier based detector labels residual tweets.

3. Proposed Renyi-based DBN and Fo-FMO-based GAN for Detection of Twitter Spam

Twitter is represented as a well-liked micro-blogging social networking site that permits users to post the message and is considered the tweets. Clicking maximum rates as well as effective propagation of messages creates Twitter as an effective platform for spammers. The user experiences as well as numerous tasks of recommendation have effects the raises in spam activities. The classification of spam is considered the trending research, particularly in the social network sites that have increased the spammer's spam activity. Nowadays, spam attackers are raising and various social media users are revealed to the spam attacks. Therefore, the main objective is to recognize the spam by exploiting the detection of the spam model by taking into consideration of the Twitter data stream.

For spam detection, the new approach is formulated in this paper by exploiting Twitter data. Moreover, the input twitter data is fed to the data transformation stage whereas Yeo-Jhonson (YJ) transformation [6] is exploited to transform the Twitter data to construct it appropriate for further processing. Subsequent to the transformation, by exploiting the DBN [8] as well as Renyi entropy [7] the feature fusion is carried out. The features are allocated on the basis of the utmost values of Renyi entropy by taking into consideration of features in feature fusion. Subsequently, the feature grouping is carried out by exploiting the number of chosen features. For grouping features, the FC [9] is used. The attained features are subjected to GAN [10] [11] whereas the spam detection is carried out by training GAN exploiting the adopted Fo-FMO approach. Fig.1 demonstrates the Architectural model of the adopted approach for spam detection on Twitter.

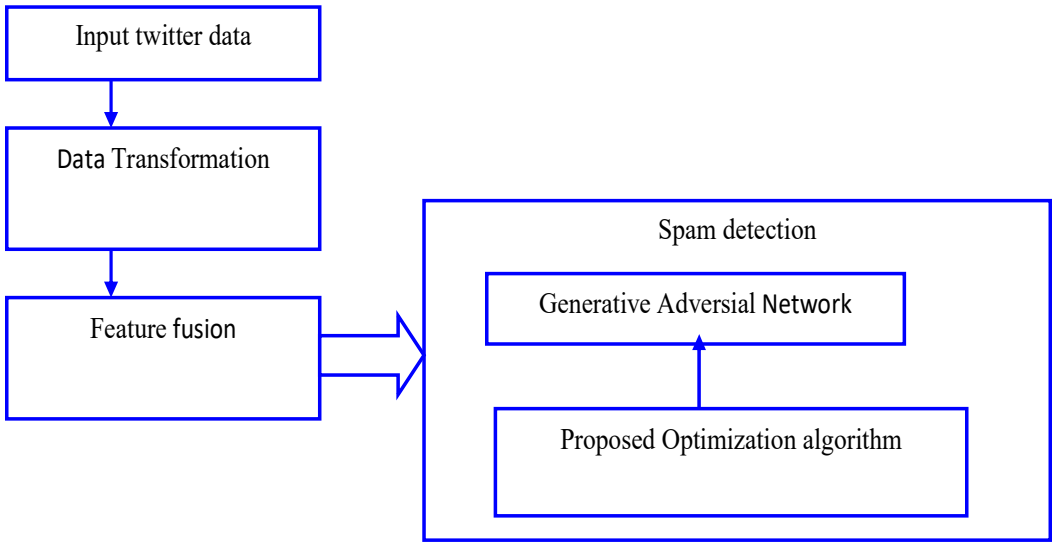


Fig.1. *Architectural model of adopted approach for spam in detection Twitter*

Let us consider input data given as A by exploiting diverse attributes, and it is stated in eq. (1), wherein A_{cd} indicates the Twitter data comprised in the database A with c^{th} attribute in d^{th} data. Moreover, a indicates total data points, and b count of attributes is used. Each data is subjected to the data transformation stage, in which the data transformation is performed to create it appropriate for further processing.

$$A = \{A_{c,d}\}; (1 \le c \le b)(1 \le d \le a) \tag{1}$$

3.1. Yeo-Jhonson Transformation for Data Transformation

Using a mathematical model, the data transformation is the process it is carried out and is developed on the data. Here this technique, a widespread transformation adapted to transform data is YJ transformation. This process is used to compress enormous data. For instance, if the processing of enormous data is performed, subsequently the minute values get inundated by superior values. Hence, YJ transformation is considered by certifying visualization to be clearer $A_{c,d}$ is considered as the input Twitter data, which is fed to the data transformation stage, wherein input twitter data is pre-processed by exploiting the YJ transformation [6]. The YJ transformation has been experimented with on the basis of the Box-Cox transformation property and YJ transformation is indicated as follows:

$$P_d = H(q,\delta) = \begin{cases} \dfrac{\left((q+1)^{\delta}\right)-1}{\delta} & ; q \ge 0, \delta \\ \log(q+1) & ; q \ge 0, \delta = 0 \\ -\dfrac{\left((-q+1)^{2-\delta}\right)-1}{2-\delta} & ; q < 0, \delta \ne 2 \\ -\log(-q+1) & ; q < o, \delta = 2 \end{cases} \tag{2}$$

wherein δ denotes parameter, q indicates log volatility that is a positive number, and P_d q indicates log volatility that is a positive number, the transformed data. The most important advantage of YJ transformation it can be used to generate the probability space.

3.2. Renyi Entropy and DBN for Features Integration

Features fusion is an integration of diverse data sources to generate the precise reliable and positive information presented by the data source. The advantage of fusion with the feature is that several feature vectors are mined on the basis of the same patterns and emulate each featured pattern. The integration as well as optimization of these features not only obtain effective discriminant information but although evades additional data to a particular degree. Moreover, the approaches on the basis of feature fusion can increase the recognition rate. In this technique, the integration of features is carried out with 2 steps in that feature re-ordering and feature grouping are carried out.

3.2.1. Features Re-ordering

By exploiting the maximum entropy values, the features are reordered in feature reordering. To investigate the data uncertainty, the Renvi entropy [7] metric is exploited and also exploited to increase the general information with diverse functions. The maximum accessibility of entropy has aimed the suitability for a particular operation. Therefore, the input data entropy is used to aim for the difference. Eq. (3) indicates the Renyi entropy of P_d, wherein, j signifies a real number, η signifies constant, a signifies total images; $1 \le d \le a$, $\Pr ob(P_d)$ signifies probability distribution of transferred Twitter data P_d.

$$R(P_d) = \frac{1}{1-\eta} \log\left(\sum_{d=1}^{a} \Pr ob(P_d)^j\right) \tag{3}$$

3.2.2. Features Grouping

In order to integrate the features, fractional calculus is used to group the features in that the fractional notion is exploited. To evaluate the optimal solution the FC [9] is exploited by exploiting the preceding iterations and it is ascertained to obtain the series features. In addition, the effectuality is improved by exploiting the adjacent features. To fuse the features, the FC is used and the performance is improved with the integration process by exploiting the individual features. Therefore, on the basis of the FC, the feature integration is carried out whereas the groups are created by considering the selected features

and the formulation is derived using the FC which is devised as eq. (4). Here, α signifies fractional coefficient, D signifies count of features to be selected, and $n = \frac{B}{D}$, B signifies total count of features.

$$K_m^{new} = \sum_{\substack{s=1 \\ s=s+B/n \\ t=1 to n}}^{D} \frac{\alpha}{t} K_s \tag{4}$$

Each feature with FC derivation is depicted as follows. Therefore, the attained formulation can be given on the basis of the fractional idea following FC [9], and the features are represented as,

$$K_1^{new} = \alpha K_1 + \frac{\alpha}{2} K_6 \tag{5}$$

$$K_2^{new} = \alpha K_2 + \frac{\alpha}{2} K_7 \tag{6}$$

$$K_3^{new} = \alpha K_3 + \frac{\alpha}{2} K_8 \tag{7}$$

$$K_4^{new} = \alpha K_4 + \frac{\alpha}{2} K_9 \tag{8}$$

$$K_5^{new} = \alpha K_5 + \frac{\alpha}{2} K_{10} \tag{9}$$

Therefore, the feature vector comprising the chosen features is indicated as,

$$M = \left\{K_1^{new}, K_2^{new}, \ldots, K_3^{new}, \ldots, K_D^{new}\right\} \tag{10}$$

From FC [27] the fraction coefficient helps to resolve derivative and integral formulation. The Laplace transform is exploited by the fractional order to resolve the derivative as well as integral formulation. Additionally, the fractional-order exploited the intrinsic memory capacity that aids to resolve the enormous data. To ascertain the fraction coefficient α input the features K_1 to K_5 DBN and produce α that in general ranges among $[0 to 1]$. The training data is taken into consideration that represents the features attained and be apt to be of size 100×10 in DBN. Moreover, all features are fed to DBN to obtain the objective value. The training data is divided into class 0 as well as 1 which class 0 indicates 60×10 and class 1 indicates 40×10. The class 0 data and class 1 data mean are calculated and are modeled into 1×10. Subsequently, the cosine similarity is calculated among the class 0 data as well as 1 data to obtain the objective value.

3.3 Architecture of DBN Classifier

The DBN [8] classifier is modeled by taking into consideration of 2 RBM layers as well as a single MLP layer. No association is comprised between visible neurons and hidden neurons in the DBN classifier. Additionally, the connection is positioned among hidden as well as visible hidden neurons. The input fed to the DBN classifier is a feature vector y that is subjected to RBM layer-1. The RBM layer-1 output presents the input to RBM layer-2 so that outcome generated from RBM layer-2 is extra presented as an input to the MLP layer. The input subjected to RBM layer-1 is devised as,

$$x^1 = \left\{x_1^1, x_2^1, \ldots, x_k^1, \ldots, x_l^1\right\}; 1 \le k \le l \tag{11}$$

$$y^1 = \left\{y_1^1, y_2^1, \ldots, y_o^1, \ldots, y_p^1\right\}; 1 \le o \le p \tag{12}$$

wherein, l denotes the total count of visible neurons, x_k^1 denotes k^{th} visible neurons of RBM layer-1, y_o^1 denotes o^{th} hidden neurons as well as p denotes the total number of neurons present in the hidden layer. Each neuron of the hidden and visible layer comprises a bias. Therefore, two biases connected with neurons of both layer in RBM layer-2 is stated as follows,

$$g^1 = \left\{g_1^1, g_2^1, \ldots, g_k^1, \ldots, g_l^1\right\} \tag{13}$$

$$h^1 = \left\{h_1^1, h_2^1, \ldots, h_o^1, \ldots, h_p^1\right\} \tag{14}$$

wherein, g_k^1 signifies bias of k^{th} visible neuron and h_o^1 signifies bias of o^{th} hidden neuron. Thus, the weights of RBM layer-1 are formulated as,

$$w^1 = \left\{w_{ko}^1\right\}; 1 \le k \le l; 1 \le o \le p \tag{15}$$

wherein, w_{ko}^1 signifies weight among k^{th} visible as well as o^{th} hidden neurons so that their weight vector size is presented as $[\kappa \times \ell]$. Hence, the outcome produced from the hidden layer of RBM layer-1 is computed on basis of the bias as well as weight connected with neurons as well as it is devised as,

$$y_o^1 = \lambda\left[h_o^1 + \sum_k x_k^1 w_{ko}^1\right] \tag{16}$$

wherein, λ signifies activation function. Therefore, the RBM layer -1 output is devised as,

$$y^1 = \left\{y_o^1\right\}; 1 \le o \le p \tag{17}$$

The RBM layer-2 is developed based on the outcome produced from the hidden layer of RBM layer-1 as well as outcome is similar to in eq. (17). The ultimate outcomes of RBM layer-s are represented in Eq. (16) and are fed as RBMlayer-2 visible layer. Thus, the count of neurons in the visible layer is similar to the hidden layer neurons of RBM layer-1 and it is stated in eq. (18). $\left\{y_o^1\right\}$ signifies the RBM layer-1 output. Therefore, RBM layer-2 hidden layer indication is represented as in eq. (19). The RBM layer-2 visible and hidden layer biases pretense similar indications as stated in Eq. (13) and Eq. (14), however, they are stated as g^2 and h^2. Hence, the RBM layer-2 weight vector is indicated as eq. (20), wherein, w_{ko}^2 states weight amid k^{th} visible neuron as well as o^{th} hidden neuron so weight size is indicated as $[v \times v]$. h_o^2 represents bias connected with o^{th} hidden neuron. Therefore, the hidden layer output is indicated as eq. (22).

$$x^2 = \left\{x_1^2, x_2^2, \ldots, x_p^2\right\} = \left\{y_o^1\right\}; 1 \le o \le p \tag{18}$$

$$y^2 = \left\{y_1^2, y_2^2, \ldots, y_o^2, \ldots, y_p^2\right\}; 1 \le o \le p \tag{19}$$

$$w^2 = \left\{w_{ko}^2\right\}; 1 \le k \le l \text{ and } 1 \le o \le p \tag{20}$$

$$y_o^2 = \lambda\left[h_o^2 + \sum_k x_k^2 w_{ko}^2\right] \forall x_k^2 = y_o^1 \tag{21}$$

$$y^2 = \left\{y_o^2\right\}; 1 \le o \le p \tag{22}$$

The aforesaid formulation is RBM layer-1 output that is fed as an input to the MLP layer so that number of neurons lies in the input layer is p. Therefore, MLP layer input is stated as eq. (23), wherein, s denotes the count of neurons comprised in the input layer that is presented by RB layer-2 output $\left\{y_o^2\right\}$. The MLP hidden layer is indicated as eq. (23) and (24), X states the total count of neurons in the hidden layer.

$$e = \left\{e_1, e_2, \ldots, e_o, \ldots, e_p\right\} = \left\{y_o^2\right\}; 1 \le o \le p \tag{23}$$

$$r = \left\{r_1, r_2, \ldots, r_X, \ldots, r_Y\right\}; 1 \le X \le Y \tag{24}$$

Suppose that F_X represents the X^{th} hidden neuron bias, wherein $X = 1,2,\ldots,Y$. Hence, the MLP layer is indicated as eq. (25), T represents the total number of neurons comprised in the output layer.

$$L = \left\{L_1, L_2, \ldots, L_S, \ldots, L_T\right\}; 1 \le S \le T \tag{25}$$

The MLP layer consists of 2 weight vectors in that one is among input as well as a hidden layer, wherein the other is among hidden as well as output layer. Let us w^A state the weight among input and hidden layers which is indicated as eq. (26). w_{oX}^A denotes weight among o^{th} input neurons as well as X^{th} hidden neurons so that size of w^A is $p \times Y$.

$$w^I = \left\{w_{oX}^A\right\}; 1 \le o \le p; 1 \le X \le Y \tag{26}$$

The hidden layer output is stated as eq. (27), V_X denotes hidden neuron bias as well as $e_o = y_o^2$, as input to MLP is outcome produced from RBM layer-2.

$$r_X = \left[\sum_{o=1}^{p} w_{oX}^A * e_o\right] V_X \forall e_o = y_o^2 \tag{27}$$

Therefore, weights among hidden as well as output layers are stated as w^B and are devised as,

$$w^B = \left\{w_{XS}^B\right\}; 1 \le X \le Y; 1 \le S \le T \tag{28}$$

Hence, the output vector is computed by exploiting weight w^B as well as hidden layer outcome is indicated as in eq. (29). r_X indicate hidden layer output as well as w_{XS}^B signifies weight among X^{th} hidden neuron and S^{th} output neurons.

$$L_S = \sum_{X=1}^{Y} w_{XS}^B * r_X \tag{29}$$

4. Proposed Fo-FMO for Spam Detection

In this section, spam detection is performed by exploiting the adopted model, and recognition is performed by exploiting the feature vector. Subsequent to the transformation the features are attained the extracted features are subjected to detection of spam module that is done by exploiting the GAN and the training is carried out using the adopted optimization model. The classifier's internal approach parameters are optimally tuned by exploiting the adopted optimization technique to increase the effectuality of the detection of spam. From the input of Twitter data, the objective of the adopted model is to identify the spam on the basis of the attained feature vector.

4.1 GAN architecture

In order to obtain the detection of spam, the feature vector T is fed to the GAN. GAN [10] [11] indicates the deep learning classifier that obtained a precise level of access in the spam detection field. To generate the precise detection, the GAN is mainly exploited that outcomes in complicated scenarios. The GAN consists of 2 diverse components a discriminator as well a generator. Moreover, the generator assures to confound discriminator via the conceivable data generation. Here, to determine fake data from the real group of data the discriminator is used. Concurrently, both the discriminator as well as the generator are trained to obtain global convergence. By exploiting the fully connected Neural Network, the generator is modeled in GAN. By exploiting the data samples of other distributions, the GAN maps the samples. In addition, to detect the spam precisely the mapping process is exploited.

Let us assume N as data points that indicate the feature vector and subjected as the input to GAN. E_{λ} as the generative model distribution, f indicate the high dimensional arbitrary variable, E_{υ} indicates the arbitrary variable, as well as E_{υ} indicates the random variable, indicates real data distribution correspondingly. Moreover, $G(N).$ indicates the generator maps the feature vector. In addition, $I(.)$ represents the generator and $J(.)$ represents the discriminator functions. Therefore, the value function $C(J,I)$ is indicated as,

$$C(J,I) = U_{O\sim V_{data}}\left[\log J(W)\right] + Q_{O\sim V_g}\left[\log(1 - J(W))\right] \tag{30}$$

$$C(J,I) = Q_{O\sim V_{data}}\left[\log J(W)\right] + Q_{O\sim V_s}\left[\log(1 - J(I(w)))\right] \tag{31}$$

wherein, $J(W)$ indicate the sigmoid function that is used to compute the probability outcome to detect renewable energy by exploiting the discriminator, and $I(w)$ indicate synthetic information by taking into consideration the distribution. Therefore, $Q_{I\sim K}$ represents the random variable expectation of data I samples from the distribution K. In addition, the discriminator is used to detect renewable energy. The loss function is modeled by taking into consideration of binary classification and cross-entropy so that the discriminator loss function is stated as,

$$\hbar_J = -\frac{1}{\vartheta}\sum_{\rho=1}^{\vartheta} G_{\rho}\log\left(J\left(W_{\rho}\right)\right) - \frac{1}{\vartheta}\sum_{\rho=1}^{\vartheta}(1 - G_{\vartheta})\log(1 - J(W_{\vartheta})) \tag{32}$$

wherein, ϑ denotes the number of samples. The generator attempt to lessen the increase of discriminator and is modeled as loss function of generator $\hbar_L$ stated as,

$$\hbar_L = \max_P . C(J,I) \tag{33}$$

4.2 Training of GAN

The GAN is trained by exploiting the adopted optimization model named Fo-FMO approach. Nevertheless, the classifier weight is trained by exploiting the Fo-FMO algorithm [12], to choose the best weights to attain the update procedures. In this work, this optimization method is also exploited to detect spam.

A new FOFMO approach [12] integrates the FMO approach with the idea of fractional order. Here, the update scheme of fractional order velocity is exploited, and also the new offspring location for the proposed model is generated on the basis of the global optimal particle.

$$d_{offset} = \frac{E_r . U_s^t}{a + b.\left(U_s^t\right)^{x'}} \tag{34}$$

$$U_s^t = P^t - P^{t-1} \tag{35}$$

To enhance the capability of exploitation, the FC is exploited to update velocity. From (34), because of a=

2:25; b = 36:2; x = 2:23, it can be identified that for the complete fraction, evaluated with the numerator, the denominator is much higher.

The grayling's migration is back to reproduce the new off-springs while they have classified to maturity. The graylings must breed in a location that is more conducive to survival. Hence, the locations of the new off-springs must be near to the global optimal particle and it is stated in eq. (36).

$$p_i^{new} = p_{gbest} + rand.\left(p_i^{old} - p_{gbest}\right) \quad (36)$$

5. Results and Analysis

In this section, experimentation analysis of the adopted model with the traditional models by exploiting precision, F-measure, and recall was demonstrated. Here, the experimentation of each approach was performed by varying training data. Moreover, the performance was analyzed by exploiting the 4 datasets. Moreover, the database comprises 4 datasets, such as a 5k-continuous dataset, a 5k-random dataset, a 95k-continuous dataset, and a 95k-random dataset [13]. Each dataset comprises 12 features and these features were used for the analysis.

Figs 2 and 3 demonstrate the analysis of techniques by exploiting the precision, recall, and F-measure parameter by taking into consideration of 4 datasets. Here, the proposed method is compared with the conventional models such as CNN, LSTM, SVM [14], and NN. By means of the analysis, it is exhibited that the adopted model attains better performance as evaluated with the conventional techniques.

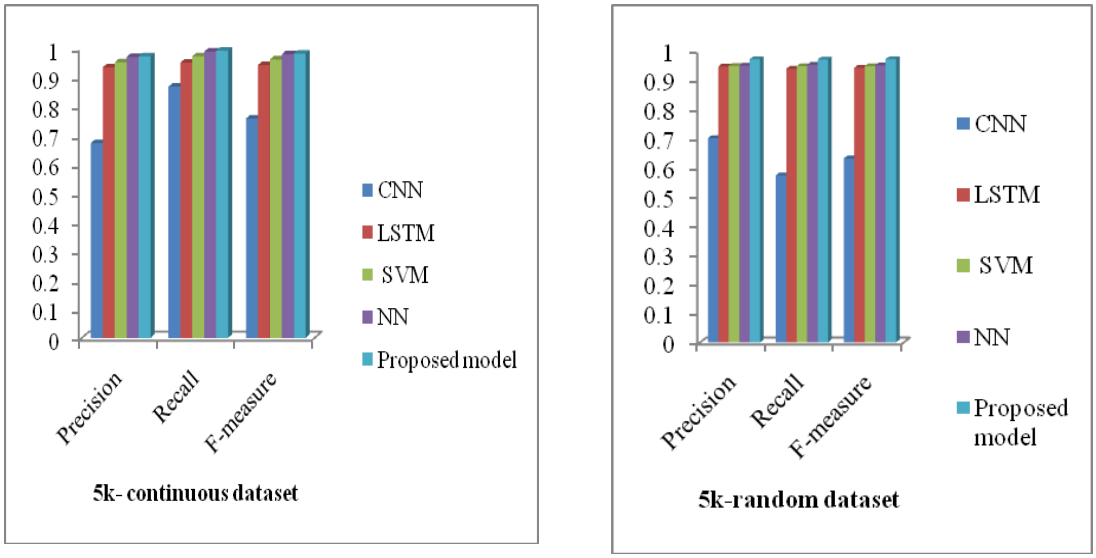


Fig.2. *Analysis of the proposed and conventional models for (a) 5k-continuous dataset (b) 5k-random dataset*

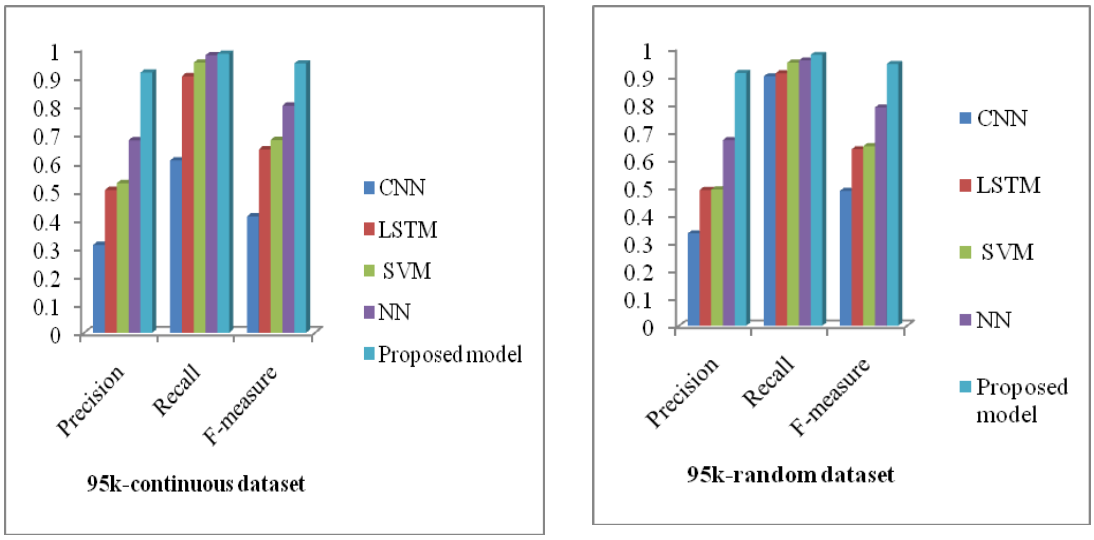


Fig.3. *Analysis of the proposed and conventional models for (a) 95k-continuous dataset (b) 95k-random dataset*

6. Conclusion

In this work, a new method was developed for the detection of spam by exploiting Twitter data. Moreover, the Twitter data experiences the transformation of data to make the data appropriate for additional processing. By exploiting the Yeo-Jhonson transformation, data transformation was performed. For feature integration, the transformed data was used in which the Renyi, as well as DBN, was exploited. Here, the feature integration was done by means of two steps in that feature grouping as well as feature order was carried out. The feature reordering was performed by exploiting the Renyi entropy as well as the features grouping was performed by exploiting the FC to obtain the sequential features. To discover the fractional coefficient, the features were subjected to DBN. The detection of spam was done by exploiting GAN in which the adopted FoFMO technique was used to train the GAN. The developed FoFMO presents higher performance as evaluated with the traditional techniques with maximum precision, maximum recall, and maximal F-measure.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

Reference

[1] Nour El-MawassPaul HoneineLaurent Vercouter,"SimilCatch: Enhanced social spammers detection on Twitter using Markov Random Fields", Information Processing & Management, 29 June 2020.

[2] Yuliya KontsewayaEvgeniy AntonovAlexey Artamonov,"Evaluating the Effectiveness of Machine Learning Methods for Spam Detection", Procedia Computer Science, 22 July 2021.

[3] Ashraf NeisariLuis RuedaSherif Saad,"Spam review detection using self-organizing maps and convolutional neural networks", Computers & Security, 17 April 2021.

[4] Z. Zhang, R. Hou and J. Yang, "Detection of Social Network Spam Based on Improved Extreme Learning Machine," IEEE Access, vol. 8, pp. 112003-112014, 2020.

[5] S. Sedhai and A. Sun, "Semi-Supervised Spam Detection in Twitter Stream," in IEEE Transactions on Computational Social Systems, vol. 5, no. 1, pp. 169-175, March 2018.

[6] Tsiotas, G., "On the use of non-linear transformations in Stochastic Volatility models", Statistical Methods and Applications, vol. 18, no. 4, pp.555, 2009.

[7] Zheng Y, Qin Z, Shao L, Hou X, "A novel objective image quality metric for image fusion based on Renyi entropy," Inf. Technol. J, vol.7, no.6, pp.930-5, 2008.

[8] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," Neural Comput., vol. 18, pp. 1527–1554, 2006.

[9] Bhaladhare, P.R. and Jinwala, D.C., "A clustering approach for the-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm," Advances in Computer Engineering, 2014.

[10] Gao, Y., Kong, B. and Mosalam, K.M., "Deep leaf bootstrapping generative adversarial network for structural image data augmentation", Computer Aided Civil and Infrastructure Engineering, vol. 34, no. 9, pp.755-773, 2019.

[11] Pascual, S., Bonafonte, A. and Serra, J., "SEGAN: Speech enhancement generative adversarial network", arXiv preprint arXiv:1703.09452, 2017.

[12] B. Guo, Z. Zhuang, J. -S. Pan and S. -C. Chu, "Optimal Design and Simulation for PID Controller Using Fractional-Order Fish Migration Optimization Algorithm," in IEEE Access, vol. 9, pp. 8808-8819, 2021.

[13] Venkateswarlu B and Viswanathan V,"Optimized generative adversarial network with fractional calculus based feature fusion using Twitter stream for spam detection",Information Security Journal: A Global Perspective, August 2021.

[14] Bhagyalakshmi V,Dr.Ramchandra and Dr.Geeta D,"Arrhythmia Classification Using Cat Swarm Optimization Based Support Vector Neural Network",Journal of Networking and Communication Systems, vol. 1, no. 1, October 2018.

[15] Meher Pratham Achampetkar,"Risk assessment and Health monitoring in WBSN using an optimization-based Deep Learning Model",Journal of Networking and Communication Systems,vol. 4, no. 2, April 2021.