Spam Detection using Recurrent Neural Networks

[1]S. Karishma, [2]V. Akila, [3]V. Govindasamy

[1]M.Tech Student, [2]Assistant Professor, [3]Associate Professor, [1,2]Department of Computer Science and Engineering, [3]Department of Information Technology, [1,2,3]Pondicherry Engineering College, Puducherry, India, [1]16karish@pec.edu, [2]akila@pec.edu, [3]vgopu@pec.edu

Abstract - Spam is well defined as the unsolicited bulk messages or junk mail will send to email address or phone number that are generally marketable in nature and also carry malicious documents. The main issue of spam is that it can download malicious files which can attack the computers, smart phones and networks, utilize network bandwidth and storage space, degrades email servers and can cause attacks in our devices like spyware, phishing and ransomware. In the existing approach, an exploratory analysis of supervised machine learning algorithms has done and the performance has been evaluated. The drawback of existing approach is that the performance of supervised machine learning algorithms decreases as we increase the size of the dataset. In order to overcome such drawbacks, an efficient spam detection using recurrent neural networks using the BiGRU model has been proposed. By implementing this, it has been achieved with better accuracy of 99.07%. From this, it is concluded that BiGRU model has better performance than existing approaches.

Keywords — Spam, Machine Learning, Deep Learning, Recurrent Neural Networks, Bidirectional Long Term Short Memory, Bidirectional Gated Recurrent Unit.

I. INTRODUCTION

Spamming is the usage of sending an unwanted message particularly marketing as well as sending messages frequently to the same email address. Spamming remains economically feasible since promoters haven't any operational costs afar their mailing lists management, servers, organizations, ranges of IP and domain names and it's hard to grip senders responsible for their bulk mailings. The most widely used spam is Email spam are automated unwanted messages send to your email address which are mostly profitable in nature and also carry malwares. The word 'spam' is used in associated abuses in other broadcastings like instant messaging, newsgroup, web search engine, blogs, Wikipedia, online classified ads, mobile phone messaging, internet forum, junk fax transmissions, social media, mobile apps, television advertising and file sharing [13]. Spam circulates 14.5 billion messages worldwide for each day. Spam occupies 45% of inbox messages. The United States ranks first in generating spam email, followed by Korea as the next leading supplier of unwanted email. The greatest prevailing type of spam is related to marketing that receives around 36% of all spam messages. The subsequent prevailing type of spam is related to adult content that receives around 31.7% of all spam messages. Since spam has flooded both the personal and corporate world, it has affected the individuals and firms worried about spam. It has been discovered in a survey organized in 2005 in which 53% of people had lost confidence in communicating through email due to spam [14].

The main issues of spam are [15]:

- Spam usually carries file attachments that contains malware in it. If the user downloads the malware supplement, the malware program is activated and installed. The intruder on the other side can make use of malware to spy and collect the unsuspicious user's information.
- Spam are the main cause of phishing attacks that can act like a genuine website and prompt an unsuspicious user to enter user credentials like Aadhar card number, ATM card details and further personal details.
- Spam cause network traffic that results in communication overloaded and hence consumes lot of bandwidth.

A. *Reducing or Preventing Spam*

Do not open spam

Never open and response to any spam messages. Even if it is opened, never download or click on any attachments that result in downloading malwares into the device [15].

DOI : 10.35291/2454-9150.2020.0305

Never stop spam using unsubscribe link

It is suggested that never unsubscribe the mail. By clicking on unsubscribe list approving that the email address is valid and the mail is received. Instead of unsubscribing the mail, the better option is never open the spam mail by adding the email to block list [15].

Never reveal address visibly

Never submit the email address publicly to any websites, comment sections, blogs etc. For e.g., some users type their email ID in comment box so that the spammer able to see your email ID and contact the owner. Spammers have specific software and flatterers created for the purpose of continuously crawling websites and gathering new email addresses [15].

Use anti-spam tools

Spamihilator, Spam Assassin, MailWasher Free, CleanMail HomePaid etc. are the anti-spam tools are available for filering spam. Of which, some are free and some are commercial tools. [15].

The summary of this paper are as follows:

1. A novel recurrent neural network of BiLSTM and BiGRU architecture model has been proposed.

2. An experimental study of hyperparameters settings for both the models has been experimented.

3. We have improved the spam detection performance by BiGRU model and the results were compared with other machine learning and deep learning techniques.

The plot of the paper is discussed here. Section 2 tells about the literature survey. Section 3 gives the detail about the deep learning, recurrent neural network, BiLSTM and BiGRU. Section 4 explains about the detailed performance evaluation, hyperparameter settings and evaluation metrics of different machine learning and deep learning algorithms. Section 5 winds up the paper.

II. LITERATURE SURVEY

In [2], spam cluster creation has done for each category using neural classification and SVM technique is used to identify the spam and the idea of neural network's multiclass classifier to categorize the data received online. The value_generator algorithm produces the each token text's ASCII value provided in the category data and produces a combined representable value that will be developed as the training data to both SVM and Neural Network. The train_neural algorithm trains the Neural Network after the neuron's successful training in the Hidden layer. The proposed work has improved performance of the kernel function of SVM may lead to more accurate results. In [3], ALO is used as an optimal feature selection and boosting classifier is used to forecast based on the features. This technique is applied with several methods of classification such as KNN, SVM and Bagging were applied on CSDMC2010 and SpamAssassin dataset. The experimental outcomes presented that the proposed method gets minimal selected features and maximal classification precision. In [4], describes the wordnet ontology and to attain least set of features optimality, feature dimensionality drop was included during techniques of feature selection like principal component analysis and correlation feature selection has been applied for classification algorithms like Logistic Regression, SVM, Random Forest and NB. These methods were applied EnronSpam dataset. The proposed work has greatly major performance with improved accuracy and less time. In [5], describes three significant contributions. Initially, frequency difference and category ratio based feature selection function TFDCR to select the most selective features regardless of the number of samples in each classification has applied. Then, an incremental learning approach is applied that permits the classifier to upgrade the discriminant task successfully. Finally, an empirical function known as selectionRankWeight is presented to improve the previous feature set which governs new structures to carry robust perceptive ability from received set of emails. These methods were implemented on EnronSpam, ECML and PU dataset. The proposed work has improved classification accurateness and false positive error reduction. In [6], describes the review's text feature extraction using Natural Language Processing (NLP) to filter the POS tagging, n-gram term frequencies, stemming, stop word and filtration of punctuation marks to reduce the number of words in the datasets has been applied for classification algorithms like Logistic Regression, SVM, RF, DT, GBT and NB. Spearman's connection coefficient is a genuine size of the unique relationship worth between coordinated evidence. These methods were applied on TripAdvisor dataset. In [7], proposed KNN classification method by conjoining Spearman's correlation coefficient as distance measure. These methods were applied on Spambase dataset. Experimental outcomes presented a noteworthy progress in accuracy with higher F-measure compared with existing algorithms. In [8], SVM is used to detect the spam but the computational complexity decreases due to high dimensional datasets. In [9], the combined PSO and ANN for selecting features and SVM to classify for self-organizing map and K-means based on conditions AUC. These methods were applied on UCI repository dataset. The outcomes show that the proposed technique has better correctness for spam detection and email classification. In [10], NB-SVM is a hybrid spam filtering algorithm which requires the advantages of both NB and SVM. NB algorithm is having fast classification and also requires small dataset but it has low accuracy performance because NB has major limitation of assumption of independence of the features extracted from training

DOI : 10.35291/2454-9150.2020.0305

samples and SVM algorithm have highest accuracy performance due to their high precision rate and recall rate. The result has more accuracy than both separately implemented NB and SVM. In [11], the preprocessing has done using TF-IDF frequency. These methods were applied on SMS Spam Collection V.1 and Spam SMS Dataset 2011-12 dataset. It is concluded that CNN has high accuracy while compared with other machine learning algorithms to detect spam SMS.

III. PROPOSED WORK

A. Deep Learning

Deep learning is a subset of a classified family of machine learning techniques depends on artificial neural networks. Learning can be categorized into supervised, semi-supervised and unsupervised. The word "deep" in "deep learning" denotes the quantity of layers over which the data is transmitted. More specifically, deep learning structures have an important credit assignment path (CAP) depth. The CAP is the order of transmissions from input to output by describing potentially connective links between input and output. Deep learning architectures can be built using greedy layer-by-layer technique. Deep learning supports to discrete these concepts and to select the best features to increase performance. Deep learning techniques eliminate feature engineering by converting the data into compressed intermediate illustrations similar to principal components and develop layered structures that eliminate redundancy for supervised learning tasks.

B. Recurrent Neural Network

Recurrent Neural Network (RNN) is an ideal result to overcome the progressive learning problem in the traditional neural network. It has a distinctive characteristic of storing the information while reading the input sequence with each time step hence they are ‘‘stateful’’. Constraint allotment is a vital property of RNN that generalizing the services to the model to the input arrangement of different extents. The fundamental RNN structure is similar to feed forward neural network with the variance in the contacts between the neurons. As a replacement for unidirectional links where data flows into neurons from one layer to another, the neurons are permitted to have directed rounds in the network. They have self-loop or links.

C. Long Short Term Memory

Long Short Term Memory (LSTM) is a kind of recurrent neural network design applied on the field of deep learning. LSTM has feedback connection that is unrelated to standard feed forward neural networks. It cannot only process sole data points but also entire sequences of information. LSTM unit contains input gate, output gate, forget gate and a cell. The cell recollects data over arbitrary time breaks and therefore the three gates control the data flow into and out of the cell. LSTM networks are compatible for categorizing, handling and producing guesses supported statistic data, subsequently there are often delays of unidentified period between main events during time sequences. LSTMs were established to overcome the discharging and disappearing gradient problems which will be come across when training traditional RNNs.

The LSTM unit contains a recollection cell that has three gates defined below:

Input gate (i): The input gate computes the sum of input that is allowed to pass through it and is calculated by:

$$i = \sigma\,(x_t\,U^i + s_{t-1}\,W^i) \qquad (1)$$

The sigmoid function plots the input value between [0, 1] and is multiplied by the weight vector (Ui). This helps the gate to bring about the quantity of input that is transferred through the input gate.

Forget gate (f): It helps the network by choosing the data from the earlier level to transfer to the succeeding level. The sigmoid function maps the value of this function between 0 and 1. It is calculated by:

$$f = \sigma\,(x_t\,U^f + s_{t-1}\,W^f) \qquad (2)$$

If no input wants to be transferred to the succeeding level, the preceding memory is multiplied with the zero vector, that creates the input value zero. Likewise, if the memory at s_{t-1} needs to pass to next level it is multiplied by 1 vector. If only certain part of the input is to be passed, then the resultant vector is multiplied with the input vector.

Output gate (o): It describes the output delivered at each step of the network and is calculated by:

$$o = \sigma\,(x_t\,U^o + s_{t-1}\,W^o) \qquad (3)$$

BiLSTM means bidirectional LSTM, which means the signal transmits backward as well as forward in time.

D. Gated Recurrent Unit

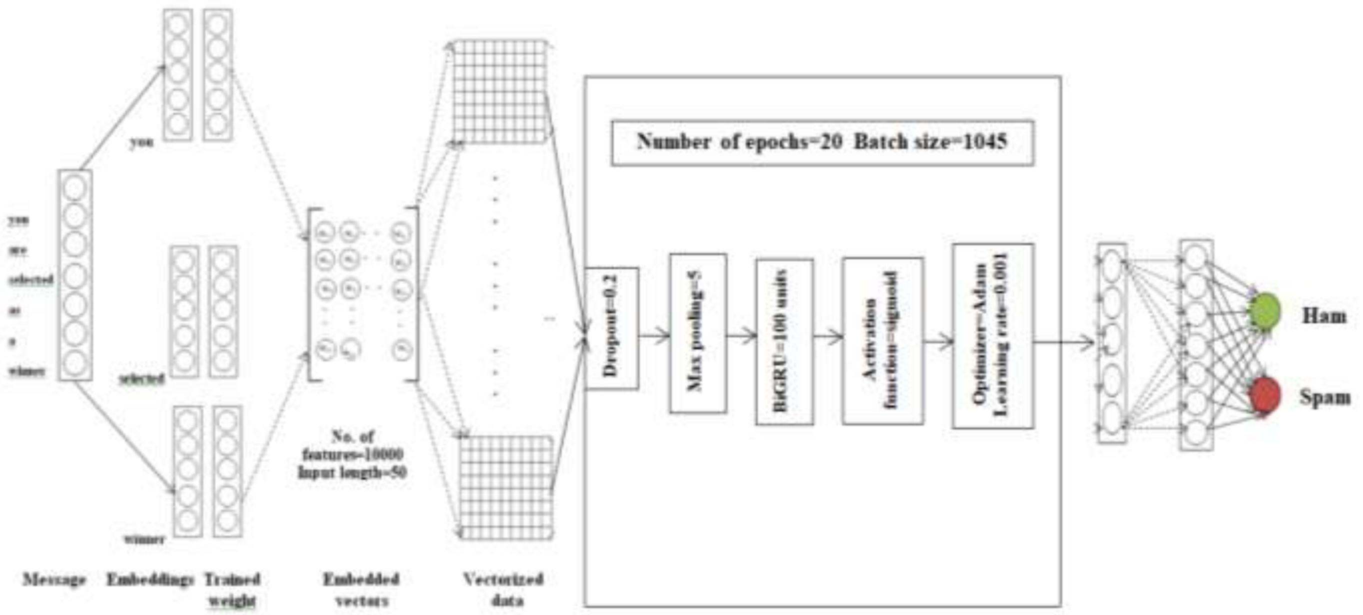
GRU is a type of deep learning algorithm that is enhanced from the LSTM algorithm to minimize the complication of the algorithm by using update gate and reset gate. The update gate is used to regulate hidden state volume to be forwarded to the next state. The reset gate is used to define the consequence of the previous hidden state information.

Update Gate (z): It determines how much of the past information needs to be passed along into the future. It is similar to the Output Gate in an LSTM recurrent unit.

$$z = \sigma\,(x_t\,U^z + s_{t-1}\,W^z) \qquad (4)$$

Reset Gate (r): It defines how much of the past information to forget. It is similar to the combination of the Input Gate and the Forget Gate in an LSTM recurrent unit.

DOI : 10.35291/2454-9150.2020.0305



$$r = \sigma (x_t U^i + s_{t-1} W^r) \quad (5)$$

BiGRU means Bidirectional GRU's are a kind of bidirectional recurrent neural networks with only the input and forget gates. It allows for the use of information from both previous time steps and later time steps to make predictions about the current state.

IV. PERFORMANCE EVALUATION

A. Dataset description

The SMS Spam Collection v.1 is a bundle of SMS tagged messages are gathered for SMS Spam research collected by Tiago Agostinho de Almeida and José María Gómez Hidalgo. It contains 5,574 SMS out of which 4,827 (86.6%) are ham messages and 747 (13.4%) are spam messages.

B. Hyperparameters

The constraints which support in boosting the LSTM network are described as follows:

Initialization of Word Vector

The manuscript should be translated into vector system, so that they can be applied in the learning model as input [12].

Activation Function

On the convolved features from the word vectors, the activation function returns the top feature that is implemented when the same window's features with many filters are concatenated. The activation function is a nonlinear function such as tanh, ReLU or sigmoid function which compresses the value of the vectors between the particular ranges. For example: sigmoid function maps the value of the vector ranged between [- 1, 1] [12].

Optimization

Optimizer is needed for minimizing the inaccuracy function that is computed when the faults are back transmitted to the preceding layer in a network. Adam algorithm is used in the proposed work to optimize the error that familiarizes learning rate according to the criteria. The learning frequency is familiarized according to the constraints with greater inform for often arising constraints and lesser inform for rare constraints [12].

Dropout

Maximum of the neural networks affect from the over fitting issue especially when the data volume is less. To minimalize the amount of over fitting is to reduce the network size rather than increase the amount of data can be overcome by one of the regularizing procedure called dropout. In the dropout level indicated network part, the hidden nodes is deleted or dropped. Due to this, the co-adaptation of the nodes is avoided and thus decreases the network over fitting [12].

Epochs

Using the same training data, the training method is repeated number of times is known as the epoch. The number of epochs depends on the training data. The complex and noisy data requires more epochs. High number of epochs might result in data over fitting and that would outcome in better training data accurateness but poorer test accurateness [12].

Features

The most important words in the corpus is considered as features. The number of features restricts the model usage to the maximum repeating words rather than holding entire features. This supports in decreasing over fitting effect due to rare words that are nullified and cannot take part in the detection of spam [12].

BiLSTM and BiGRU Units

BiLSTM and BiGRU units are the amount of memory cells in the BiLSTM and BiGRU network. It denotes the capable of remembering the facts and matches it with the preceding evidence. The information in the memory units

DOI : 10.35291/2454-9150.2020.0305

are moved further in the following time period for the advance training.

C. Hyperparameter Settings

Table- I: Hyperpararameter Settings

Parameter	Value
Number of features	10000
Dropout	0.2
Activation function	Sigmoid
Optimization	Adam
Learning rate	0.001
Max pooling	5
BiGRU units	100
Epoch	20

D. Experiments

The execution of the BiLSTM and BiGRU model was took place with Tensorflow backend on the Keras 2.0 API via Python 3.6.5 in Windows 10 64-bit operating system. The data is preprocessed by removing include stemming, noise removal and stop words and embedded. Each embedded message is trained and weight is allocated. When the suitable features are selected, pick an amount to allocate the features to generate features vectors before classification. Then, it is converted into matrix form and data is vectorized. Then apply max pooling to flattern the vectorized data. Weighted features are generally displayed as bag of words or vector space model. The trained dataset is forwarded into DNN classifier. Finally, the dataset is separated into 75% of training set and 25% of testing set and classify the given message is spam or ham using BiGRU model. Adam optimizer yields better performance by providing the minimal loss during the training of the model. The model was tested with different sizes of featuring maps and the pooling window. The best outcomes were generated with the feature map size is 64 and pooling window size is 5. The model executed well with 0.2 of dropout rate. The best results were obtained of using 5 grams. The model was additional verified with the batch size 1024 with 20 epochs. By implementing this model, it has been achieved with better accuracy of 99.07% while comparing with existing approaches.

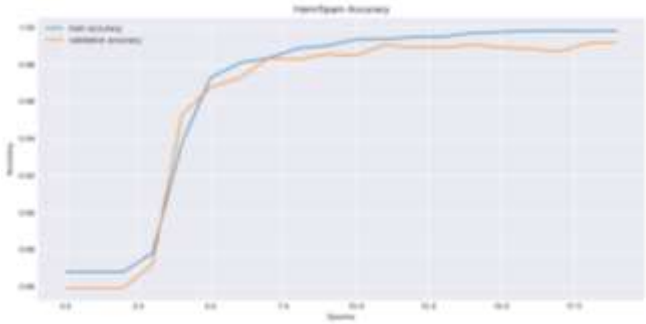


Fig. 2. Accuracy graph evaluated using BiGRU model related with number of epochs

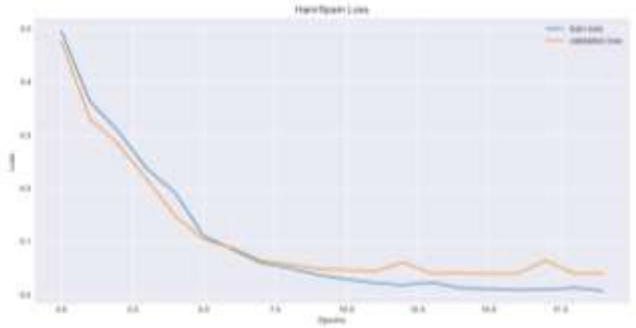


Fig. 3. Loss graph evaluated using BiGRU model related with number of epochs

E. Performance Metrics

Accuracy

To find the percentage of correctly classified instances.

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (6)$$

Precision

Precision refers to the true positive to the total predicted positive.

$$\text{Precision} = \frac{tp}{tp + fp} \quad (7)$$

Recall

Recall refers to the true positive to the total actual positive.

$$\text{Recall} = \frac{tp}{tp + fn} \quad (8)$$

F_1-score

F_1 - score is the harmonic average of precision and recall.

$$F_1 - \text{score} = 2 \, . \, \frac{\text{precision . recall}}{\text{precision + recall}} \quad (9)$$

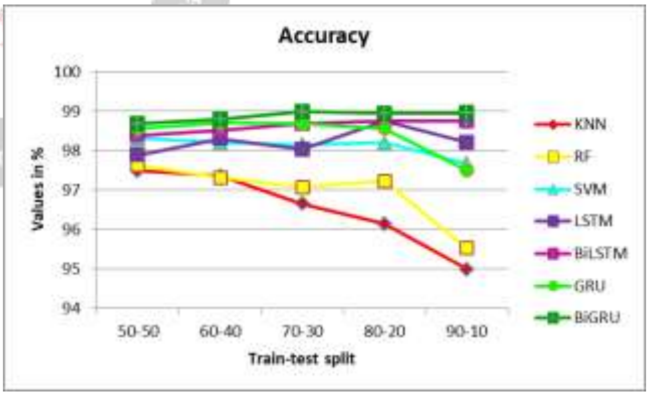


Fig. 4. Accuracy graph

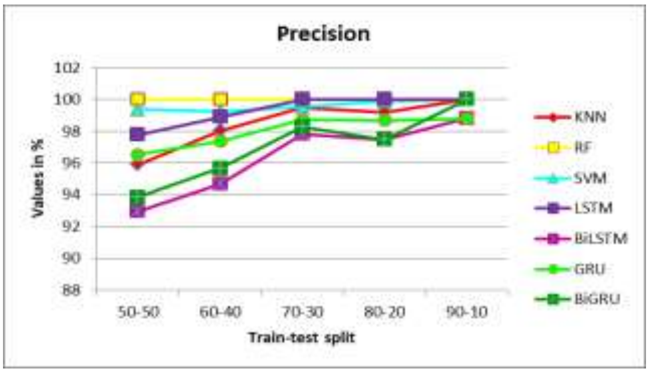


Fig. 5. Precision graph

DOI : 10.35291/2454-9150.2020.0305

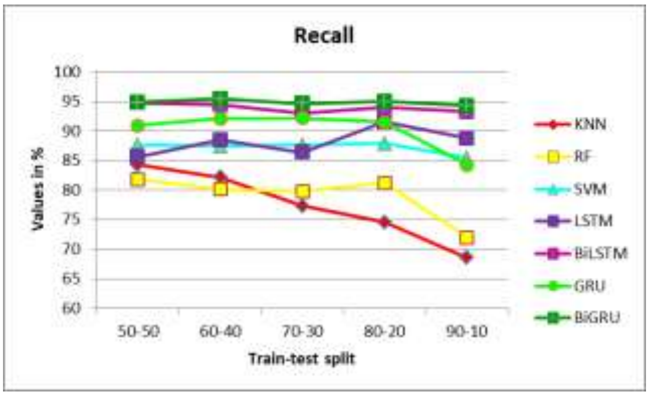


Fig. 6. Recall graph

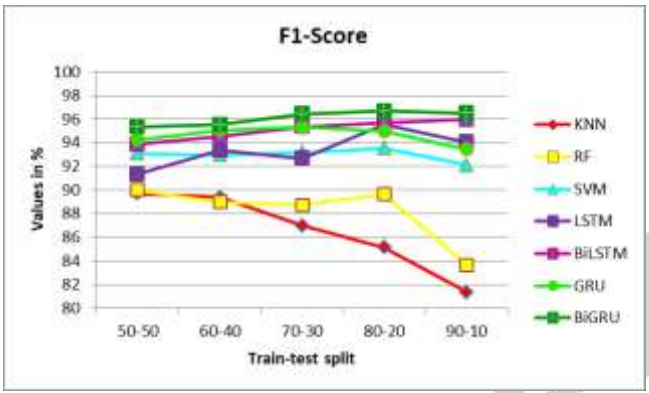


Fig. 7. F1-Score graph

V. CONCLUSION

This paper explained about the implementation of various machine learning classification algorithms such as KNN, Random Forest and SVM and deep learning algorithms such as LSTM, GRU, BiLSTM and BiGRU for SMS spam detection. Out of which, BiGRU model has better accuracy of 99.07% while compared with other machine and deep learning algorithms. This can be further applied on datasets of different languages and different datasets like twitter, online social reviews.

REFERENCES

[1] Emmanuel Gbanga Dada, Joseph Stephen Bassi, Haruna Chiroma, Shafi Muhammad Abdulhamid, Adebayo Olusola Adetunmbi and Opeyemi Emmanuel Ajibuwa, "Machine Learning for email spam filtering: reviews, approaches and open research problems", *Elsevier, Helion 5,* May 2019.

[2] Sanjeev Dhawan and Simran, An Enhanced Mechanism of spam and category detection using Neuro-SVM", *International Conference on Computational Intelligence and Data Science (ICCIDS 2018),* 429-436, 2018.

[3] Amany A. Naem, Neveen I. Ghali and Afaf A. Saleh, "Antlion optimization and boosting classifier for spam email detection", *Future Computing and Informatics Journal 3*, 436-442, 2018.

[4] Eman M. Bahgat, Sherine Rady, Walaa Gad and Ibrahim F. Mohawad, "Efficient email classification approach based on semantic methods", *Ain Shams Engineering Journal 9*, 3259-3269, 2018.

[5] Gopi Sanghani and Ketan Kotecha, "Incremental personalized E-mail spam filter using novel TFCDR feature selection with dynamic feature update", *Expert Systems with Applications 115, Elsevier*, 287-299, July 2018.

[6] Wael Etaiwi and Ghazi Naymat, "The Impact of applying Different Preprocessing Steps on Review Spam Detection", *The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017), Procedia Computer Science 113, Elsevier,* 273-279, 2017.

[7] Ajay Sharma and Anil Suryawanshi, "A Novel Method for Detecting Spam Email using KNN Classification with Spearman Correlation as Distance Measure", *International Journal of Computer Applications (0975 – 8887)* Volume 136, Issue 6, February 2016.

[8] Zahra S. Torabi, Mohammad H. Nadimi-Shahraki and Akbar Nabiollahi, "Efficient Support Vector Machines for Spam Detection: A Survey", *(IJCSIS) International Journal of Computer Science and Information Security*, Volume 13, Issue 1, January 2015.

[9] Mohammad Zavvar, Meysam Rezaei and Shole Garavand, "Email Spam Detection Using Combination of Particle Swarm Optimization and Artificial Neural Network and Support Vector Machine", *I.J. Modern Education and Computer Science*, Volume 7, 68-74, 2016.

[10] Diksha S. Jawale, Ashwini G.Mahajan, Kalyani R. Shinkar and Vaishnavi V. Katdare, "Hybrid spam detection using machine learning", *International Journal of Advanced Research, Ideas and Innovations in Technology*, Volume 4, Issue 2, 2828-2832, 2018.

[11] Mehul Gupta, Aditya Bakliwal, Shubhangi Agarwal and Pulkit Mehndiratta, "A Comparative Study of Spam SMS Detection using Machine Learning Classifiers", *Eleventh International Conference on Contemporary Computing (IC3)*, 2-4 August, 2018, Noida, India.

[12] Gauri Jain, Manisha Sharma and Basant Agarwal, "Spam detection in social media using convolutional and long short term memory neural network*", Annals of Mathematics and Artificial Intelligence, Springer*, January 2019.

[13] https://en.wikipedia.org/wiki/Spamming

[14] https://www.spamlaws.com/spam-stats.html

[15] https://atechjourney.com/how-dangerous-are-spam-mails.html

DOI : 10.35291/2454-9150.2020.0305