International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 1, February 2020, pp. 486~493
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i1.pp486-493

A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing

Abhishek Kumar[1], Jyotir Moy Chatterjee[2], Vicente García Díaz[3]
[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Himachal Pradesh, India
[2]Department of IT, LBEF (APUTI), Kathmandu, Nepal
[3]Department of Computer Science, Universidad de Oviedo, Asturias

Article Info

Article history:

Received Aug 17, 2019
Revised Aug 30, 2019
Accepted Sep 17, 2019

Keywords:

Feature extraction
Fraudulent emails
Hybrid method
Phishing attacks
Phishing detection
Probabilistic neural network (PNN)
Support vector machine (SVM)

ABSTRACT

Phishing attacks are one of the slanting cyber-attacks that apply socially engineered messages that are imparted to individuals from expert hackers going for tricking clients to uncover their delicate data, the most mainstream correspondence channel to those messages is through clients' emails. Phishing has turned into a generous danger for web clients and a noteworthy reason for money related misfortunes. Therefore, different arrangements have been created to handle this issue. Deceitful emails, also called phishing emails, utilize a scope of impact strategies to convince people to react, for example, promising a fiscal reward or summoning a feeling of criticalness. Regardless of far reaching alerts and intends to instruct clients to distinguish phishing sends, these are as yet a pervasive practice and a worthwhile business. The creators accept that influence, as a style of human correspondence intended to impact others, has a focal job in fruitful advanced tricks. Cyber criminals have ceaselessly propelling their techniques for assault. The current strategies to recognize the presence of such malevolent projects and to keep them from executing are static, dynamic and hybrid analysis. In this work we are proposing a hybrid methodology for phishing detection incorporating feature extraction and classification of the mails using SVM. At last, alongside the chose features, the PNN characterizes the spam mails from the genuine mails with more exactness and accuracy.

Corresponding Author:

Jyotir Moy Chatterjee,
Department of Information Technology,
Lord Buddha Education Foundation (APUTI),
Opposite Maiti Devi Temple, Kathmandu, 44600-Nepal.
Email: jyotirchatterjee@gmail.com

1. INTRODUCTION

Phishing is a cybercrime in correlation with others, for example, hacking. The expression of phishing is a minor departure from the word fishing. The thought is that trap is tossed out with the expectations that a client will snatch it and nibble into it simply like the fish. Phishing is equipped for harming e-commerce since it makes client lose their trust on the web. To make clients sophisticated of most recent phishing assaults, some global associations, for example, anti-phishing working group (APWG), have distributed phishing alarms on their sites [1]. As indicated by APWG patterns report first quarter 2014 [2], the quantity of phishing destinations expanded by 10.7 percent over the final quarter of 2013 and furthermore the installment administrations are the most focused on industry part. Messages can be sorted into three [3] Ham, Spam and Phishing. Ham is requested and genuine email while spam is a spontaneous email. Then again, phishing is a spontaneous, beguiling, and possibly unsafe email. Phishing emails are made by false individuals to emulate genuine E-banking emails.

The deceptive phishing which is identified with social engineering methods, rely upon molded email that misrepresentation from an authentic organization or bank. At that point, through a link inside the email, the assailant endeavors to misdirect clients to counterfeit Websites. These forged Web destinations are intended to misleadingly get monetary information (usernames, passwords, debit/credit card numbers, and individual data, and so forth.) from certified clients [4].

Phishing is a cybercrime in correlation with others, for example, hacking. The expression of phishing is a minor departure from the word fishing. The thought is that trap is tossed out with the expectations that a client will snatch it and nibble into it simply like the fish. Phishing is equipped for harming e-commerce since it makes client lose their trust on the web. To make clients sophisticated of most recent phishing assaults, some global associations, for example, anti-phishing working group (APWG), have distributed phishing alarms on their sites [1]. As indicated by APWG patterns report first quarter 2014 [2], the quantity of phishing destinations expanded by 10.7 percent over the final quarter of 2013 and furthermore the installment administrations are the most focused on industry part. Messages can be sorted into three [3] - Ham, Spam and Phishing. Ham is requested and genuine email while spam is a spontaneous email. Then again, phishing is a spontaneous, beguiling, and possibly unsafe email. Phishing emails are made by false individuals to emulate genuine E-banking emails.

The deceptive phishing which is identified with social engineering methods, rely upon molded email that misrepresentation from an authentic organization or bank. At that point, through a link inside the email, the assailant endeavors to misdirect clients to counterfeit Websites. These forged Web destinations are intended to misleadingly get monetary information (usernames, passwords, debit/credit card numbers, and individual data, and so forth.) from certified clients [4].

The ongoing improvements in internet and mobile innovation pulled in maximum business foundations to provide their administrations internet, along banks, stocks and e-commerce. As individuals progressively depend on Internet administrations to complete their exchanges, Internet extortion turns into an extraordinary risk to individuals' security and privacy. Phishing is the primary sorts of web misrepresentation; that depends on tricking clients to distribute or proclaim their personal data (counting passwords & Visa numbers), phishing is characterized as a digital assault which conveys socially-designed messages to people through e-correspondence channels (email, SMS, telephone call) so as to convince users to do activities (like enter accreditations, debit/credit card number, etc.) for the assailants advantage; such activities could be influencing an online business site client to enter his certifications to a forged site (overseen by the aggressor) like the first site and after that the assailant utilizes them to mimic the client. So as to induce the sufferer individual client to login to such a forged site, the socially designed message attracts a dream to the client that he needs to perform such activity, for example, cautioning the client about record suspension or that the site administrator is mentioning him to reset his password [5].

Phishing assaults utilize email messages and sites that are planned in an expert way to be like emails and sites from genuine foundations and associations (more often than not the client is a client for those associations), to influence clients into revealing their own or money related data. The aggressor would then be able to utilize gathered delicate client data for his advantage. Clients can be fooled into uncovering their data either by giving touchy data by means of a web structure, answering to mock emails, or downloading and introducing Trojans, which search clients' PCs or screen clients' online exercises so as to get data [6].

Most recent advances in phishing assault research have expanded stream cognizance of how people choose decisions concerning suspicious mails. In any case, the precise activity of various message-express components, including how and why they sway people's choices and decisions, remains dim. We have utilized a technique for feature extraction, extricated some unique sort of features from the text and pictures then the feature selection done over the separated features and the SVM classifier going about as a twofold classifier and arrange the genuine and phished emails with strategies like Text parsing, word tokenization and Stop words expulsion alongside Bayes classification strategy using probabilistic neural network. Finally, alongside the chose features, the Probabilistic Neural System (PNN) characterizes the spam mails from the genuine mails with more exactness and accuracy. The next sections are arranged as follows section 2 provides the literature survey about the various email phishing detection and analysis methods. Section 3 presents our proposed methodology in details such as methods used and dataset used for the purpose along with our proposed algorithm. The uniqueness of the proposed methodology is presented in section 4. Section 5 discusses about the results we got along with detailed discussion and section 6 concludes the research work.

2. LITERATURE REVIEW

In 2014 [7] examines & reported the utilization of random forest machine learning method in classification of phishing assaults, with the real goal of building up an improved phishing email classifier with better prediction precision & less quantities of features. In 2015 [4] investigated which dimension reduction method is better for characterizing content information like emails. [8] proposed a smart model for recognition of phishing emails which relies upon a preprocessing stage that concentrates a set of features concerning diverse email parts. [9] studied earlier works away at how to avoid end-clients from succumbing to email (spear) phishing assaults.

In 2016 [10] analyzed the impact of three social engineering procedures on clients' decisions of the fact that it is so sheltered to click on a link in an email. One of the constraints of their work was the utilization of a comfort test of college understudies joined up with subjects on business and information systems. Such a sample may not really mirror the capacities of the more extensive populace and, in this way, it restrains the generalizability of their discoveries. In 2016 [6] provided a smart classification model for identifying phishing emails utilizing information revelation, data mining and content handling procedures. They present the idea of phishing terms weighting which assesses the heaviness of phishing terms in each email. In 2017 [11] suggested an idea called TORPEDO to improve phish recognition by giving just-in-time and just-in-place reliable tooltips. These assistance individuals to distinguish phish links inserted in emails. TORPEDO's tooltips contain the genuine URL with the area featured.

In 2018 [12] uncovers a concerning gap amongst the server-side spoofing discovery and the real protection on clients. They exhibited that most email suppliers enable manufactured messages to get to client inbox, while coming up short on the vital cautioning instrument to tell clients (especially on versatile applications). Portrays a utilization of a deep sequence-to-sequen CNN model to the creation check task for email messages [21]. Gave a best in class review over Bitcoin related innovations and total up different difficulties [22]. Tried to give a sensible increasingly significant comprehension about the IoT in BD structure close by its various issues and difficulties and focused on giving possible arrangements by ML methodology [24]. Uncovered about how to recognize phishing emails from genuine emails [25].

In 2019 [13] explored how three of these components, which have not been broadly inspected in past research, impact decisions of email trust and convincingness, explicitly the utilization of misfortune and reward-based impact systems, real structure signs, and referencing a remarkable recent development. Introduced a strategy while in transit to characterize an instrument for computerized distinguishing proof of standards of human influence in social engineering, inside phishing messages [14]. According Ruskanda [19] considered the impacts of pre-processing stages on the presentation of supervised spam classifier algorithms. Tests were led on two generally utilized supervised spam classifier algorithms: NB & SVM. introduced another phishing email recognition model named THEMIS, which is utilized to show emails at the email header, the email body, the character level & the word level at the same time [20]. Focused on the latest advancement over inquires about concerning machine learning for big data processing and various methods with regards to present day processing situations for different social applications [23].

3. PROPOSED SYSTEM

Feature extraction, feature selection and classification are the important steps in the email phishing detection. For the purpose of feature extraction, we extracted some special kind of features from the text and images then the feature selection done over the extracted features and the support vector machine classifier acting as a binary classifier and classify the legitimate and phished emails [15]. Figure 1 show depicts our proposed approach for the email spam filtering method.

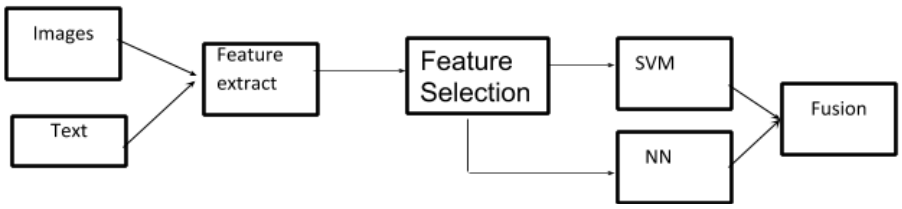

Figure 1. Proposed flow diagram

3.1. Text parsing, word tokenization

The email subject and body content are parsed & tokenized into tokens, if the email body is machine-comprehensible content increase language-organized then the HTML labels unit parsed to remove the content and affirm URLs. Additionally, if the email contains connections, they will even be parsed and tokenized [16].

3.2. Stop words removal

Here, a portion of the regular words may appear to be of almost no significance or trash zone unit a long way from the separated tokens, the normal stop words grasp the tokens "the", "then", "he",… and so on this preprocessing helps to decreasing the likenesses among messages and improves the presentation of the arranged model to distinguish the email phishing [17].

3.3. Dataset

The dataset used includes of 1705 emails out of that 1291 area unit ham and 404 area unit phished. The ham emails area unit collected from an in public accessible dataset and phished emails area unit a combination of emails from numerous sources.

3.4. Proposed algorithm

Start
Step1 -Function narratives
Step 2-Preprocess of text data collected from different sources
Step 3-Feature extraction stage
{X=μ}; where μ is feature vector with anomalies
Step 4-Form a set of Feature vector from documents data set
X=α {Say} where α Preprocessed feature vectors
Step5-Punctuations erased and conversion performed
Step6- List of stop hand removal and normalization has been performed for better preprocessing of document dataset
{∑X =0} now in order to perform normalization i.e. {α1+ α2+ α3+ α4+…… αn}
Step7- Now hybrid approach of SVM along with new probabilistic neural network applied for classifications unlike existing approach
Hybrid Approach = {SVM+PNN + NLP}

The advantage of proposed approach is let's assume if any classified data is missed by one classifier the other classifier in the collaborative approach will correct it and increased accuracy has been shown in the results section. We have done XOR operation to make SVM act as binary classifier and results depicted in Table 1.

Table 1. XOR operation with SVM

SVM +NLP	NN	OUT
0	xor 0 -	0
1	xor 1 -	0
0	xor 1 -	1
1	xor 1 -	1

4. NOVELTY IN IMPLEMENTATION APPROACH

The proposed model is based on dynamic approach unlike existing approach although the proposed work also approached with inbuilt dataset collected from different internet source to provide dynamic dimension to the work and improve the accuracy as compare to existing work.
The novelties in proposed approach are as follows

1. Features from different number and source from documents are collected first in order to provide dynamic dimension and better accuracy to the results which has been shown in the result section, Hybrid methodology provided better accuracy comparatively.
2. The second parameter that has been taken is setting the threshold value like constraints to remove all the features that do not appear more than two times, that means removal of infrequent words will be performed unlike the existing work discussed in related work section.
3. The proposed work has not blindly extracted features from document but narratives are processed to better phishing of email tackling methods apart from that further it has been tokenized which has proved to increase the accuracy than existing methods discussed.

4. The other dimension which has been approached to get better accuracy than all comparative methods is to remove empty documents as our datasets has not considered empty documents enter many times and increase the complexity and reduce the accuracy of results.
5. Ultimately the part where proposed work has overcome the accuracies shown by other conventional and proposed approach is classification part .The training of dataset is one of the most crucial stage which can increase or decrease the accuracy The proposed approach focuses on collaborative approach of SVM along with the probabilistic approach of neural network over the large number of datasets.

5. RESULT AND DISCUSSION

The dataset consisting of extracted options is divided then fed into 5 classifiers and results noted. 10-fold cross validation technique has been used for partitioning the initial knowledge sample into coaching set and take a look at set. K-fold cross validation [17]. In k fold cross validation, the data set is willy-nilly split into k reciprocally exclusive subsets of around equal sizes [17]. Followed by this, the model is trained and tested k number of times, of the k samples, one sub sample is maintained as validation knowledge of the testing model and remaining k-1subsamples area unit used as coaching set it's determined that tree based mostly [18], SVM and supply classifiers classify most accurately. Performance totally different| of various classifiers is evaluated exploitation different performance metric that area unit delineate during this section. Its determined that SVM and Bayes classify the data set with the highest accuracy of ninety-nine, 89%. the subsequent performance metrics area unit used for evaluating our model.

Legitimate: positive

Phishing: negative

True positive (Tp) = No. of samples identified correctly as legitimate
False positive (Fp) = No. of incorrectly identified samples as legitimate
True negative (Tn) = No. of correctly identified samples as phishing
False negative (Fn) = No. of incorrectly samples identified as phishing

Accuracy: The test samples exactness is its capacity to separate the genuine and phishing samples accurately. To figure out the precision of this framework, we ought to consistently ascertain the level of factual positive & factual negative out & out assessed samples. Logically, this will be unequivocal as:

Accuracy=Tp+Tn/(Tp+Tn+Fp+Fn)

Sensitivity: Sensitivity defined is the ability to see the legitimate samples properly. To estimate it, we should always compute the percentage of factual positive in legitimate samples. Scientifically, this is declared as:

Sensitivity=Tp/Tp+Fn

Specificity: The specificity of a take a look at is its capacity to work out the legitimate samples properly. To evaluate it, we must always calculate the percentage of real negative in legitimate cases. Scientifically, this may be represented as:

Specificity=Tn/Tn+Fp

In the Table 2 and Table 3, we have shown the performance of various other methods compared with our proposed hybrid method and found that our proposed method performs better in terms of accuracy, sensitivity, and specificity of spam mail prediction from other algorithms. In Figure 2 we have shown the neural network parameter setting in MATLAB nn tool. In Figure 3 we have compared the true positive and false negative rate i.e., ROC of our proposed hybrid method with SVM & NN methods. In Figure 4 we have shown the comparison between percentage a sensitivity, accuracy & specificity among the full hybrid and half hybrid methods.

Table 2. Performance with hybrid method

Method	Accuracy	Sensitivity	Specificity
KNN	86	85	89
BAYES	89	88	91
RF	91	89	92
LOGIC R	93	91	94
Hybrid(proposed)	98.3	98	98.5

Table 3. Performance with hybrid method vs separate

Method	Accuracy	Sensitivity	Specificity
SVM	87	88.5	91
NN	90.5	92	93.5
Hybrid (proposed)	98	97	97.5

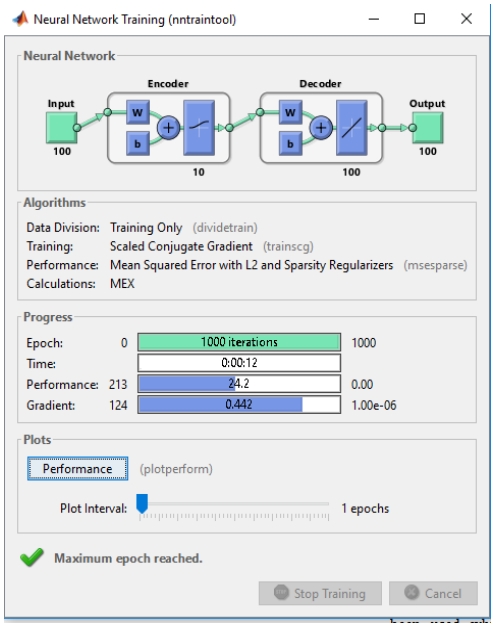

Figure 2. Neural network output

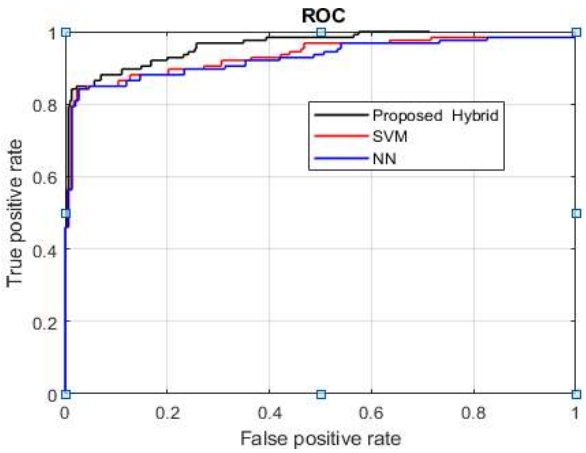

Figure 3. ROC for hybrid vs SVM vs NN

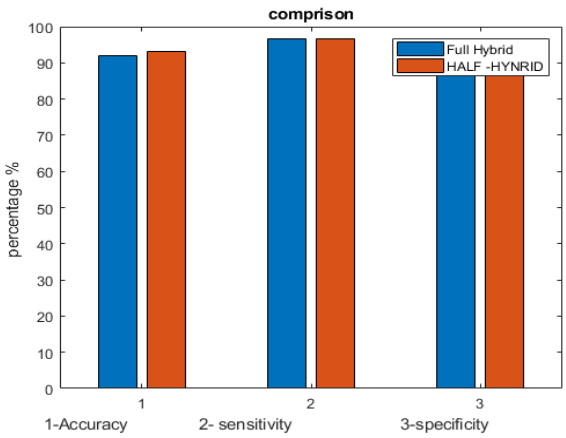

Figure 4. Feature SEMI vs FULL (hybrid)

6. CONCLUSION

Latest advances in phishing vulnerability research have extended flow comprehension of how individuals settle on choices with respect to suspicious messages. Be that as it may, the exact job of different message-explicit elements, including how and why they impact individuals' decisions and choices, stays hazy. We have used a method of feature extraction, extracted some special kind of features from the text and images then the feature selection done over the extracted features and the support vector machine classifier acting as a binary classifier and classify the legitimate and phished emails with methods like Text parsing, word tokenization and Stop words removal along with Bayes classification method. In this work we are proposing a hybrid methodology for phishing detection incorporating feature extraction and classification of the mails using support vector machines (SVM). While compared our proposed hybrid method along with Support Vector Machine (accuracy- 87%, sensitivity-88.5 % & specificity-91%) & Neural Network (accuracy-90.5%, sensitivity-92%, specificity-93.5%) method, we found our proposed hybrid method (accuracy-98%, sensitivity-97%, specificity-97.5%) performed better. In future we will try to n future works, the projected system will be improved by increasing the dataset. By adding a range of emails each of kind phished and ham, the system would be nearer to the real-life situation wherever fraudsters area unit day by day rising their techniques. Victimization real world samples would change areas to deploy a formal system that will be used across organization and in private to forestall users from being victims to phishing attacks.

REFERENCES

[1] APWG. Anti-phishing working: http://www.antiphishing.org

[2] Phishing Activity Trends Report 2014: http://docs.apwg.org/reports/apwg_trends_report_q1_2014.pdf.

[3] Hamid, I. R. A., & Abawajy, J., "Hybrid feature selection for phishing email detection," In *International Conference on Algorithms and Architectures for Parallel Processing* (pp. 266-275), Springer, Berlin, Heidelberg, October, 2011.

[4] Zareapoor, M., & Seeja, K. R. "Feature extraction or feature selection for text classification: A case study on phishing email detection," *International Journal of Information Engineering and Electronic Business*, *7*(2), 60, 2015.

[5] Dong, X., Clark, J. A., & Jacob, J., *"Modelling user-phishing interaction," In 2008 conference on human system interactions* (pp. 627-632), IEEE, May, 2008.

[6] Yasin, A., & Abuhasan, A., "An intelligent classification model for phishing email detection," *arXiv preprint arXiv:1608.02196,* 2016.

[7] Akinyelu, A. A., & Adewumi, A. O, "Classification of phishing email using random forest machine learning technique," *Journal of Applied Mathematics*, 2014.

[8] Smadi, S., Aslam, N., Zhang, L., Alasem, R., & Hossain, M. A., "*Detection of phishing emails using data mining algorithms,"* In 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA) (pp. 1-8). IEEE, December, 2015.

[9] Stembert, N., Padmos, A., Bargh, M. S., Choenni, S., & Jansen, F, *"A study of preventing email (spear) phishing by enabling human intelligence,"* In 2015 European Intelligence and Security Informatics Conference, pp. 113-120, IEEE, September 2015.

[10] Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A., "Breaching the human firewall: Social engineering in phishing and spear-phishing emails," *arXiv preprint arXiv:1606.00887,* 2016.

[11] Volkamer, M., Renaud, K., Reinheimer, B., & Kunz, A, "User experiences of TORPEDO: tooltip-powered phishing email detection," *Computers & Security*, *71*, 100-113, 2017.

[12] Hu, H., & Wang, G., "End-to-end measurements of email spoofing attacks," In *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pp. 1095-1112, 2018.

[13] Williams, E. J., & Polage, D., "How persuasive is phishing email? The role of authentic design, influence and current events in email judgements," *Behaviour & Information Technology*, 38(2), 184-197, 2019.

[14] Ferreira, A., & Teles, S., "Persuasion: How phishing emails can influence users and bypass security measures," *International Journal of Human-Computer Studies*, 125, 19-31, 2019.

[15] Verma, R., & Rai, N., *"Phish-idetector: Message-id based automatic phishing detection*," In *2015 12th* International Joint Conference on e-Business and Telecommunications (ICETE), Vol. 4, pp. 427-434. IEEE, July 2015.

[16] Hand, D. J., Mannila, H., & Smyth, P. "Principles of data mining (adaptive computation and machine learning)," MIT Press, 2001.

[17] Basnet, R., Mukkamala, S., & Sung, A. H., "Detection of phishing attacks: A machine learning approach," *In Soft Computing Applications in Industry*, pp. 373-383. Springer, Berlin, Heidelberg, 2008.

[18] Kohavi, R., "A study of cross-validation and bootstrap for accuracy estimation and model selection," In *Ijcai,* Vol. 14, No. 2, pp. 1137-1145, August, 1995.

[19] Ruskanda, Fariska Zakhralativa, "Study on the Effect of Preprocessing Methods for Spam Email Detection," *Indonesian Journal on Computing (Indo-JC),* vol. 4, no. 1 109-118, 2019.

[20] Fang, Yong, Cheng Zhang, Cheng Huang, Liang Liu, and Yue Yang. "Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism," *IEEE Access,* Volume 7, 56329-56340, 2019.

[21] Litvak, Marina, "Deep Dive into Authorship Verification of Email Messages with Convolutional Neural Network." In *Annual International Symposium on Information Management and Big Data*, Springer, Cham, pp. 129-136, 2018.
[22] Chatterjee, Jyotir Moy, Srijani Ghatak, Raghvendra Kumar, and Manju Khari, "BitCoin exclusively informational money: a valuable review from 2010 to 2017," Quality & Quantity 52, no. 5 (2018): 2037-2054.
[23] Tripathy, Hrudaya Kumar, Biswa Ranjan Acharya, Raghvendra Kumar, and Jyotir Moy Chatterjee, "Machine learning on big data: A developmental approach on societal applications," In Big Data Processing Using Spark in Cloud, Springer, Singapore, pp. 143-165, 2019.
[24] Chatterjee, Jyotir, "IoT with Big Data Framework Using Machine Learning Approach," *International Journal of Machine Learning and Networked Collaborative Engineering*, 2(02), pp. 75-85, 2018.
[25] Hiransha, M., Nidhin A. Unnithan, R. Vinayakumar, K. Soman, and A. D. R. Verma, "Deep learning-based phishing e-mail detection," In *Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal. (IWSPA).* Tempe, AZ, USA, 2018.

BIOGRAPHIES OF AUTHORS

Abhishek Kumar is currently working as an Assistant Professor in CSE department at Chitkara University and is pursuing his PhD in computer science from University of Madras and research is going on face recognition using IOT concept and done M. Tech in Computer Sci. & Engineering from Government engineering college Ajmer, Rajasthan Technical University, Kota India. He has a total Academic teaching experience of more than 8 years with more than 50 publications in reputed, peer reviewed National and International Journals, books & Conferences like Wiley, Taylor & Francis Springer, Elsevier Science Direct, Inderscience, Annals of Computer Science, Poland, and IEEE. My research area includes- Artificial intelligence, Image processing, Computer Vision, Data Mining, Machine Learning. I have been in International Conference Committee of many International conferences. I have been the reviewer for IEEE and Inderscience Journal. He has authored 5 books published internationally and edited 8 books with Wiley, IGI GLOBAL Springer, Apple Academic Press and CRC etc. He is also member of various National and International professional societies in the field of engineering & research like Member of IAENG (International Association of Engineers), Associate Member of IRED (Institute of Research Engineers and Doctors), Associate Member of IAIP (International Association of Innovation Professionals), Member of ICSES (International Computer Science and Engineering Society), Life Member of ISRD (International Society for research & Development), Member of ISOC (Internet Society),Editorial Board member in IOSRD I have got Sir CV Raman life time achievement national award for 2018 in young researcher and faculty Category. He is serving as an Associate Editor of Global Journal on Innovation, Opportunities and Challenges in Applied Artificial Intelligence and Machine Learning.

Jyotir Moy Chatterjee is currently working as an Assistant Professor (IT) at Lord Buddha Education Foundation (Asia Pacific University of Technology & Innovation), Kathmandu, Nepal. Prior to this I have worked as an Assistant Professor (CSE) at GD Rungta College of Engineering & Technology (CSVTU), Bhilai, India. I have completed M. Tech in Computer Science & Engineering from Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha and B. Tech in Computer Science & Engineering from Dr. MGR Educational & Research Institute, Chennai. I have 36 international publications, 2 authored books, 2 edited volume books & 2 book chapters into my account. My research interests include the Cloud Computing, Big Data, Privacy Preservation, Data Mining, Internet of Things, Machine Learning & Blockchain Technology. I am member of various professional societies and international conferences.

Dr. Vicente García Díaz is an Associate Professor of Department of Computer Science, Languages and Information Systems at University of Oviedo (School of Computer Science). He is in editorial board of various International journals and having various international research paper publication