International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-11 Issue-6, August 2022

Framework for Cyber Threats in Social Networks

Sheba Pari N., Senthil Kumar K.

Abstract*: Social networking is the most common way of communication nowadays. Maintaining the information's confidentiality, integrity and availability becomes a very critical aspect. As the number of users on social media keep increasing, the amount of data about the users are available on the network is also increasing. Attacks on these networks are currently at an all-time high which can be by Phishing attacks, Botnets, Sybil Attack, Profile Cloning, Spam, Denial of service to name a few of them. There are a number of threats possible on social networks. Data in social networks must be protected from various types of cyber-attacks. The main requirement is providing security to such networks. Maintaining the information's confidentiality, integrity and availability becomes a very critical aspect. As and when security is being provided to these networks, attacks are also evolving. Cyber-attacks are becoming complex which means that sometimes the threat for which the solution needs to be found is unknown. Threats are becoming automated, hence, using less efficient algorithms for cyber security is not the optimal solution. Hence, machine learning is used to support cyber security to social networks. A framework is built which comprises of the steps such as Data Collecting, Data Preparing, Applying Machine Learning Techniques, Post-processing by applying domain specific knowledge to build a secure system for social networks using machine learning techniques.

***Keywords*: Attacks, Botnets, Cyber Security, Machine Learning, Phishing, Social Networks, Spam.**

I. INTRODUCTION

Social Networks are a means of communication to stay Social Networks are a means of communication to stay connected with family, friends, or any other person of interest. This network structures persons or group of persons who may be connected by a number of interdependency factors. These factors includes common interests, same place of work, school mates, college friends, colleagues, business partners or any other criteria. Social networks are also used productively, probably by researchers so that they can share their similar thoughts and ideas in developing something new. Traditional methods of exchanging ideas through e-mail were rather a very slow process, where ideas exchanged would take a long time to be actually implemented. With the existence of social networks, enormous amount of data could be exchanged at a faster pace. The number of users using social media is increasing drastically. The Global Digital growth as of January 2021 as per the sources, the UN, Local Government Bodies, GSMA Intelligence, ITU, GWI, EUROSTAT, CNNIC, APJII, Social media platforms self-service advertising tools, Company Earnings Reports, Mediascope is shown in Fig. 1. Social Networks have become exceedingly popular. A few years before, no one would have anticipated this tremendous growth of social networks. Today more than half of the world's population uses social media [2]. In the first half of the year 2020, most of the countries around the world were observing a partial state of lockdown to prevent the spread of an epidemic. This resulted in even more number of users using social networks. As per the research by Global Web Index (worldwide), Pew Internet Surveys (US) and Of Com (UK) and Data reportal Global Overview report, users using social media spend an average of around 2 hours per day using multi social media networks and messaging apps. The total number of users in each of the social networks as of the year 2020 is shown below in Fig. 2.

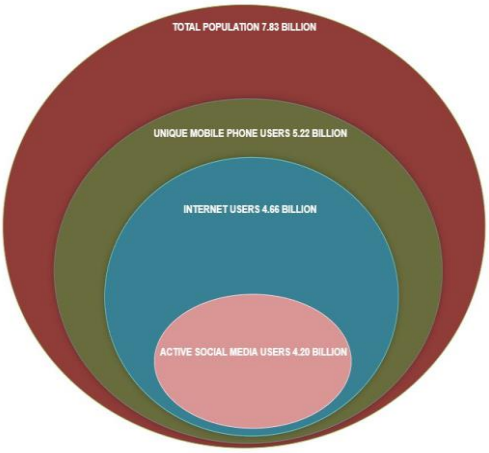

Fig. 1: Global Digital Growth

[Sources: The UN, Local Government Bodies, GSMA Intelligence, ITU, GWI, EUROSTAT, CNNIC, APJII, Social media platforms self-service advertising tools, Company Earnings Reports, Mediascope]

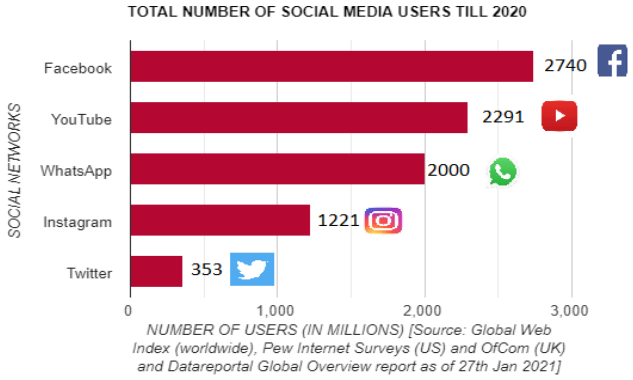

Fig. 2: Total Number of Social Media users till 2020

Manuscript received on 28 July 2022 | Revised Manuscript received on 02 August 2022 | Manuscript Accepted on 15 August 2022 | Manuscript published on 30 August 2022.

* Correspondence Author
Sheba Pari N., Research Scholar, SITE, VIT University, Vellore (Tamil Nadu), India. Email: shebapari.n2017@vitstudent.ac.in
Dr. Senthil Kumar K.*, Associate Professor, SCOPE, VIT University, Vellore (Tamil Nadu), India. Email: ksenthilkumar@vit.ac.in

Users continue to create social network profiles. Most of the users end up giving all the confidential information while creating profiles and usually don't go through the privacy policy of each of such social networks. Once all the information regarding a person has been given, no one is sure as to how secure it is [10]?

II. THREATS ON SOCIAL MEDIA

There are a number of threats possible on social networks, a few of them are

A. Phishing attacks

This is a falsified attempt at obtaining the social network profile users personal information. This personal data could be sensitive information of a person such as where he/she works, place of residence, universities where education has been completed and so on. Phishing is mainly considered as an identity theft where an attacker is trying to act like the genuine user, since he has gotten hold of all the sensitive information. There are different ways in which the attacker can get hold of the sensitive information. A few of them are

- A spyware could be installed on a user's system driver or browser and every action made by the user could be sent to the attacker without the user's knowledge.
- A malicious content could be added to a website which the user thinks is genuine and visits frequently. Through the genuine website, the user could be directed to the attacker's website so the attacker can obtain confidential information.
- The attacker could create some fake Web Pages. These web pages could be indexed in any search engine. Web pages could have certain keywords through which they would occur first when searched in a search engine and further users could enter confidential information.

B. Privacy Leak

This a type of attack in which all the information in a profile of a user in a social network would be accessed by the attacker. This information obtained could be used for any malicious doings [11]. On social network pages, the friends of the users have access to almost every confidential information posted by the user. Hence the profile user must have to change his/her security settings and see who is actually trustworthy, based on which he can add his friends list.

C. Profile Cloning

Whenever a user creates a new account on social media, he/she needs to provide a few confidential information to create a profile. This information can be used by a malicious user to create similar profiles of the same user [14]. This is called as Profile cloning. This could be done in the same social network and act like the legitimate user or take all the user details from one social network and create a user profile on another social network.

D. Spam

This type of attack is a scenario where a huge number of messages using electronic messaging are sent extensively in the network. Most common form of sending spam is via e-mail. The spam could be either sent in a broadcast form of spam where the attacker does not know the identification of the social network user but uses some combinations of words to generate some probable social network user identifications and send spam messages to those ids. Spam is basically a concept wherein the email technology is misused in a drastic way.

E. Denial of service

In this type of attack, genuine users are prevented access to social networks [13][15]. These attacks flood the genuine target with sufficient amount of traffic which sends the user login to be blocked, to trigger a crash or completely shut down. There was a cyber-attack called "WannaCry", in the year 2017 which affected in almost 150 countries around the world. This attack firstly penetrates into a computer and then encrypts all files present in that computer so it would not be readable to the actual users. If the users needed to decrypt their own files they had to buy special decryption software which were sold by the attackers themselves, hence the attackers themselves attacked and made money by providing a solution to this problem. There was another important cyber-attack called the "ransomware", which affected individual users and large organizations as well. The total loss incurred with respect to this attack was close to 4 billion US dollars.

III. MACHINE LEARNING TO PROVIDE CYBER SECURITY

Cybersecurity as defined by "are a set of tools, practices, and guidelines that can be used to protect computer networks, software programs, and data from attack, damage, or unauthorized access". Cybersecurity is a set of technologies and processes designed to protect computers, networks, programs and data from attacks and unauthorized access, alteration, or destruction. In summary cyber security can be summarized as dealing and understanding the cyber-attacks or the threats that affect our social networks. This could be done by conceiving a number of ways to preserve 3 predominant goals of security, the CIA triad (Confidentiality, Integrity and Availability). Providing cyber security in social networks main objective should be preserving confidentiality, integrity and availability [9][12].

- Confidentiality- Confidentiality is one of the most important parts of providing security to social networks. It signifies keeping the user's details private, as in all the personal information shared by the social network user should be given only restricted access or limited access. The objective is to safeguard all the confidential information on a social media users profile against any actions that may affect the confidentiality of information.
- Integrity- In a social media user's account integrity has to be maintained. This implies that only the user who is a valid user should be able to modify his/her private information or any information that is made public. The posts posted by the user in his/her name should be only by the valid user. Any unauthorised posts should not be possible in the name of the user's account.

- Availability- This goal of security means that social media user's account should be accessible to authorized users. When a valid user tries accessing his/her account, access should be granted. If an attacker has taken over the user's account then the chances of the account been accessible by the authorized user is minimised.

Machine learning has become a very important part of providing cybersecurity. When attacks are evolving drastically, the solutions to solve these problems cannot be the same traditional methods of solving. Machine learning tries to solve the attacks before they even occur (pre-emptively). Different types of pattern detection can be implemented wherein automated recognition of patterns and regularities in data are done. Usually detection of patterns is done by using data which is labelled but there are a number of scenarios when labelled data is not available, hence detection of patterns would be difficult. In such scenarios, there are different types of algorithms available for this. Mainly we can use knowledge discovery in databases processes or also known as KDD (Data Mining). When data is labelled, different types of supervised algorithms can be used.

In Supervised learning, the most commonly used techniques are Classification algorithms and Regression algorithms [16]. Both these techniques can be used to prevent a cyber threat from happening. These techniques can predict if the cyber threat is going to happen or not. When we want to predict numerical or continuous values, regression methods could be used. Examples of cyber threats that could be detected using regression methods are to find out continuous values such as total phishing attacks over a period of time. This type of analysis could also be used to fine out the main reason why the social networks are attacked. The most common type of regressions used is linear regression. This regression establishes a relationship between dependent variable which is usually continuous and one or more independent variables which can be continuous or discrete using a best straight line (regression line) where the independent variable is considered as X and dependent variable considered as Y. Logistic regression is mainly used for classification problems even though it is a regression. This does not apply a linear relationship between dependent and independent variables. Regression Analyses are used for detection of certain types of cyber threats.

In Classification, predicting or identifying different classes of cyber threats could be done. Classification identifies if a particular attack is present or not. The predicted output for this type of algorithm is categorical or discrete. The most commonly used classification algorithms are Naïve Bayes, K-Nearest Neighbours, Decision tree, random forest and support vector machine. Most of the real time scenarios where a particular cyber threat is resolved, a mix of different simple models are used. This sort of learning is ensemble learning. Unsupervised learning is used when the data is unlabelled [17]. In social networks, malware is usually unknown for a long duration. During that time the behaviour of this malware keeps changing drastically, which makes it difficult to be detected using supervised learning. In such scenarios, unsupervised learning is preferred. Unsupervised algorithms are mainly clustering or association.

Clustering algorithms usually processes all the data available and tries to group into clusters. It tries to find natural grouping among the data. Clustering algorithms are of different types such as K means clustering, K nearest neighbours, hierarchical clustering and so on.

Association is usually used when we have large amount of data and when relationships are established between the data available.

In addition to supervised and unsupervised learning techniques, there are other types of learning techniques such as semi-supervised. Semi-supervised learning is a mix of 2 techniques, supervised and unsupervised techniques. These techniques can work in both the labelled and unlabelled data. For providing cyber security for social networks, data is usually unlabelled and labelling the data has to be done automatically. In such scenarios it is better to use semi-supervised learning techniques. This helps in finding threats in a much faster approach. Another type of learning technique is the Reinforcement techniques [1]. This works based on a reward approach, wherein the algorithm learns for providing cyber security in social networks, especially for detecting malicious data on the social networks or detecting unwanted data in the traffic, reinforcement could be used. There are a number of researchers who have spoken about the social networks and security with machine learning in these networks. S. Rathore et al. discusses about the security threats that occur mainly with respect to multimedia data in social networks and how to resolve these threats [2]. Sarker et al. discusses about the different issues in security and provides a framework which is based on machine learning techniques to provide a complete security solution [1]. Gao et al. also discusses about the different security threats in social networks. Each of the issue is explained in detail and measures to provide security for these threats are discussed [3]. Novak et al. discusses the existing techniques that protect social networks data. This research also talks about the necessity of link prediction and user attributes [4]. Jin et al. discusses about the existing methods for providing security in social networks. This research also discusses about the social networks with respect to user behaviour [5]. Kayes et al.'s discusses about the different threats on social networks with respect to the stakeholders and solutions to eliminate these threats, but this research also spoke about the challenges in using these solutions [7]. Deliri et al. discussed about the most common attacks affecting social networks such as malware, Sybil attack and phishing and ways in providing security from these attacks [8]. Fire et al.'s research was about a broader perspective of different threats present in the social networks. This research classified all the threats into a broader category such as classic threats, modern threats and discussed about existing ways in providing security to these threats [6]. In addition to these researches, many researchers have discussed about providing security in social networks, most of these papers have specified a traditional way of providing solution to these threats. This paper mainly talks about providing automated solution to the same threats by making use of different machine learning techniques.

Table 1 summarizes a few of the concepts considered by the researchers in their researches with respect to threats present currently and solutions to these threats.

Table- I: Comparison of related works with respect to providing security.

Research Work	Year	Threats in Security	Existing Security Solutions	Use of Machine learning techniques	With respect to Social Networks
Sarker et al.	2020	Yes	Yes	Yes	—
S. Rathore et al.	2017	Yes	Yes	—	Yes
Deliri et al.	2015	Partial	Partial	—	Yes
Kayes et al.	2015	Yes	Yes	—	Yes
Fire et al.	2014	Yes	Yes		Yes
Jin et al.	2013	Partial	Partial	—	Yes
Novak et al.	2012	Partial	Yes	—	Yes
Gao et al.	2011	Partial	Yes		Yes

IV. FRAMEWORK FOR PROVIDING CYBER SECURITY

To provide social networks with security against cyber threats, a model needs to be built. This model firstly collects all the data, applies machine learning techniques to find cyber threats and tries to build a more secure system. The following is an outline of the steps required for providing cyber security. Fig. 3 shows the steps involved in this framework [1].

A. Data Collecting

This is foremost, the most important step. In this step we need to collect data from different social media. Based on the type of security provided by the system to protect against cyber threats, the security data is collected. Data collected should be valuable in building the model.

B. Data Preparing

As per the above step, data is collected. Post the collection step, raw data should be prepared. This step is done by applying various processes. A huge amount of data is usually collected from the data collection step. Every data in that is not useful data. Only data which helps in building the model should be preserved. The quality of the data should be high. Noisy data, missing data or incorrect data should not be used as we need to build a model of high efficiency. The raw data collected can either be structured, semi- structured or unstructured. Whatever be the type, data should be converted to structured data. Also data should be pre-processed so as to handle the missing values and clean the data. The more efficiently this step is done, the more efficient is the secure model built.

C. Data Preparing

As per the above step, data is collected. Post the collection step, raw data should be prepared. This step is done by applying various processes. A huge amount of data is usually collected from the data collection step. Every data in that is not useful data. Only data which helps in building the model should be preserved. The quality of the data should be high. Noisy data, missing data or incorrect data should not be used as we need to build a model of high efficiency. The raw data collected can either be structured, semi- structured or unstructured. Whatever be the type, data should be converted to structured data. Also data should be pre-processed so as to handle the missing values and clean the data. The more efficiently this step is done, the more efficient is the secure model built.

D. Applying Machine Learning Techniques

Once the data is prepared, machine learning algorithms are applied to build the secure model to find the cyber threats in social media. From the data prepared in the above step, data needs to be analysed to see which data helps in generating security features. There are numerous ways in building a model suitable for building a secure system for social networks. One of the models is to make the available raw data into data which is more informative to build a security model. This is called as security feature engineering. Several techniques can be used in this such as feature transformation and normalization, feature selection or feature generation.

The second model can be used to find out the cyber threats present in the huge amount of data. This could be done using data clustering which basically groups related data into clusters. How would clusters enable in solving cyber threats? This can find out which of the data doesn't follow the security policies or any irregularities in data. Any malicious behaviour of the data can also be identified by clustering. Another modelling which can be considered is classification, to classify the different cyber threats. Prediction can also be considered to predict if any cyber threats such as denial of service can occur. In reality machine learning models can work in a combined fashion for better efficiency. Models could be modified according to security problem.

E. Post processing by applying domain specific knowledge

The security model is completed in this step. In this step the security model is kept up to date. The training data got above to build a model is able to build a security model but to keep the system up to date, recently detected patterns should also be used. Recently used security patterns are more likely to be more effective in finding out the cyber threats in social media than the older ones [18]. Thus this framework comprises of different layers. To provide security to social networks, getting data is the main concern. After that machine learning models are applied to find out the different cyber threats involved. Basically this model builds secure system from raw data.

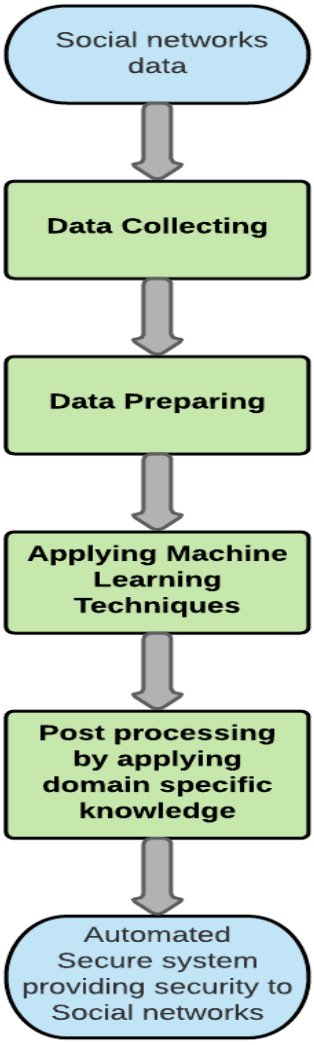

Fig. 3: Framework for providing Cyber Security

V. CONCLUSION

Social networks are the most prominent way of communication these days, especially during the epidemic the world is going through. Social networks consist of huge amount of data with loads of private information of millions of users. Maintaining the information's confidentiality, integrity and availability becomes a very critical aspect. Users end up giving loads of their private information during communication and profile creation. Hardly a very few number of users go through the privacy policy before agreeing to create profiles. There are a number of threats possible on social networks. Data in social networks must be protected from various types of cyber-attacks such as Phishing attacks, Botnets, Sybil Attack, Profile Cloning, Spam, Denial of service to name a few of them. There are various methods in which cyber security can be provided to these attacks. Threats are evolving hence the methods used to provide security also needs to be advanced. Machine learning is used to support cyber security to social networks. Most of the machine learning models is based on supervised learning, unsupervised learning or reinforcement learning. To provide social networks with security against cyber threats, a model needs to be built. In this paper, the model to be developed for providing security was discussed using machine learning techniques. Most of the researches in this area usually concentrated without using machine learning techniques for social networks or using a single technique. It is ideal to use machine learning techniques as a combination of techniques rather than a single ideal model. This model firstly collects all the data, applies machine learning techniques to find cyber threats and tries to build a more secure system. The framework required to build a secure system in social networks comprises of the steps such as Data Collecting, Data Preparing, Applying Machine Learning Techniques, Post- processing by applying domain specific knowledge. While applying machine learning techniques, the models can work more efficiently in a combined fashion.

REFERENCES

1. Iqbal H. Sarker, A. S. M. Kayes, Shahriar Badsha, Hamed Alqahtani, Paul Watters and Alex N. "Cybersecurity data science: an overview from machine learning perspective", Journal of Big Data, July 2020, Vol. 7, Issue 41. [CrossRef]
2. Shailendra Rathore, Pradip Kumar Sharma, Vincenzo Loia, Young-Sik Jeong, Jong Hyuk Park, "Social network security: Issues, challenges, threats, and solutions", Elsevier, 2017, pp. 43-610 [CrossRef]
3. H. Gao, J. Hu, T. Huang, J. Wang, Y. Chen, Security issues in online social networks, IEEE Internet Computing, 2011, Vol. 15, Issue 4, pp. 56–63. [CrossRef]
4. E. Novak, Q. Li, "A Survey of Security and Privacy in Online Social Networks", College of William and Mary Computer Science, 2012, pp. 1–32.
5. L. Jin, Y. Chen, T. Wang, P. Hui, A.V. Vasilakos, "Understanding user behavior in online social networks: a survey", IEEE Communications Magazine, 2013, Vol. 51, Issue 9, pp. 144–150. [CrossRef]
6. M. Fire, R. Goldschmidt, Y. Elovici, "Online social networks: threats and solutions", IEEE Communications Survey & Tutorials, 2014, Vol. 16, Issue 4, pp. 2019–2036. [CrossRef]
7. I. Kayes, A. Iamnitchi, "A Survey on Privacy and Security in Online Social Networks", arXiv preprint arXiv, 2015, pp. 1–40.
8. S. Deliri, M. Albanese, "Security and privacy issues in social networks", Data Management in Pervasive Systems, Data-Centric Systems and Applications, Springer, 2015, pp. 195–209. [CrossRef]
9. Jang-Jaccard J, Nepal S., "A survey of emerging threats in cybersecurity", Journal of Computer & System Sciences-Elseiver, 2014, Vol. 80, Issue 5, pp. 973–993 [CrossRef]
10. Li S, Da Xu L, Zhao S., "The Internet of things: a survey", Information Systems Frontiers, 2015, Vol.17, Issue 2, pp. 243–259. [CrossRef]
11. Sun N, Zhang J, Rimba P, Gao S, Zhang LY, Xiang Y., "Data-driven cybersecurity incident prediction: a survey", IEEE Communications Survey & Tutorials, 2018, Vol. 21, Issue 2, pp. 1744–1772. [CrossRef]
12. McIntosh T, Jang-Jaccard J, Watters P, Susnjak T., "The inadequacy of entropy-based ransomware detection", International conference on neural information processing, New York, Springer, 2019, pp. 181–189. [CrossRef]
13. Gao, H., Hu, J., et al., "The status quo of online social network security: A survey", IEEE Internet Computing, 2011, pp. 1–6.
14. Makridakis, A., Athanasopoulos, E., Antonatos, S., Antoniades, D., Ioannidis, S., & Markatos, "Understanding the behavior of malicious applications in social networks", IEEE Network, 2010, Vol. 24, Issue 5, pp. 14-19. [CrossRef]
15. Kontaxis, G., Polakis, I., Ioannidis, S., & Markatos, "Detecting social network profile cloning", IEEE Third International Conference on Security and Social Networking, 2011, pp. 295–299. [CrossRef]
16. Sarker I. H. , Kayes A., Watters P., "Effectiveness analysis of machine learning classifcation models for predicting personalized context-aware smartphone usage", Journal of Big Data, 2019, Vol. 6, Issue 1, pp. 1–28. [CrossRef]
17. Sarker I. H., "Context-aware rule learning from smartphone data: survey, challenges and future directions", Journal of Big Data, 2019, Vol. 6, Issue 1. [CrossRef]
18. Sarker I. H., Colman A., Han J., "Recencyminer: mining recency-based personalized behavior from contextual smartphone data", Journal of Big Data, 2019, Vol. 6, Issue 1, pp. 49. [CrossRef]

AUTHORS PROFILE

Sheba Pari N., She is a Ph. D. Research Scholar in the School of Information Technology & Engineering (SITE) in VIT, Vellore, India. Her research interest includes Information and Cyber Security, Social Networks, Artificial Intelligence and Network Engineering. She received her M.Tech in Software Engineering from Visvesvarya Technological University, Belgaum, India and B.E. in Information Science & Engineering from Visvesvarya Technological University, Belgaum, India. She has over 11 years of teaching experience as an Assistant Professor in various Engineering colleges.

Dr. Senthil Kumar K., received his Ph.D degree from the Department of Computer Science and Engineering at Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India, in the year 2016. He is currently an Associate Professor in VIT Vellore India. His research interest area includes Knowledge in Data Engineering, Mobile Computing, Software Engineering, Mobile Data Management, Wireless Networks, Cloud Computing, Data Science, Machine Learning.

Retrieval Number: 100.1/ijeat.F37620811622
DOI: 10.35940/ijeat.F3762.0811622
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)