International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 11 ǁ November 2022 ǁ PP. 184-187

Cyber Crime Estimation: A Hybrid Key Indicator Using Data Analytics and Machine Learning

Mrs. P. Yamuna[1], Y. Saideep [2], K.Tejaswini[3], C. Vishal[4], P. Sagarika[5]
Assistant Professor, Department of Computer Science and Engineering[1]
IV B. Tech Students, Department of Computer Science and Engineering[2,3,4,5]
ACE Engineering College, Hyderabad, Telangana, India

ABSTRACT
Social media text analytics is the process of deriving information from text sources. Text analysis can be applied to any text-based datasets, including social media.
Crime is a major problem faced today by society, even the routes have turned all over to social media. The quality of life and economic growth have been severely impacted by crime.
By detecting and analysing historical data, we can uncover crime patterns and make predictions about future crimes. However, some crimes go unreported and unsolved because there isn't enough proof.
As a result, finding criminals is still a difficult task. We can monitor social media for criminal activity. Because people who utilize social media occasionally post statements about their surroundings on those platforms.
In this paper, we proposed a Machine Learning approach is used to detect the crimes and analyze it on its type. As the first step, we fetch the text messages using predefined keywords relating to the crimes. Then, after the preprocessing, we applied a support vector machine- based filtering approach to eliminate the noise. And then Random Forest is used for classification. Then in the final stage, we analyze and categories the crime type.
Keywords: *Cyber Crime, Cyber Attacks, Security, SVM, Malware, Data Analytics, Machine Learning .*

Date of Submission: 05-11-2022 Date of acceptance: 19-11-2022

I. INTRODUCTION

A crime such as spamming, passing on computer viruses, harassment, Cyber stalking, and others have becomecommon in our modern world. While these issues do not carry potential monetary loss, they are just as harmful in the possibility of losing files, information and access to your computer. Because of this, cyber security is essential. Cyber-security is the process of preventing unauthorized access to, use of, disclosure, interruption, modification, or destruction of information held on computers, electronic devices, and other hardware and software.

Additionally, it aids in defending the computer system from many hazardous technologies and shields the PC from harm (viruses, worms, bugs and bacteria). Additionally, it aids in network monitoring and safeguards it against various dangers. Therefore, to some extent, we need utilize computer security solutions to secure our data from various types of sniffing theft problems.

Crime is a social issue that has a significant negative impact on many facets of our society. For both local authorities and individuals, the capacity to recognize high-crime zones and locate the most recent crimes in a certain region is of significant concern.

On the other hand, when residing in a busy environment, people are constantly engaged in enhancing safety and developing trusting connections with neighbour. One of the biggest problems facing civilizations around the world, especially in urban areas, is the frequency of crime. Although social crimes are the subject of more studies than any other category, social media has only been used in a small number of studies involving crimes and criminal behaviour.

As a result, the study seeks to offer (Random Forest algorithm), which makes use of machine learning and is intended to have a strong capability to detect crimes by features of social media dataset utilizing the concept of data mining.

Social media is the main source of our data. The basic objective is to locate every hidden data source and forecast outcomes.

II. LITERATURE SURVEY

The application recognizes and detects the crimes and analyze it on its type using a machine learning approach. This application includes,fetching the text messages using predefined keywords relating to the crimes.Then, after the preprocessing, applying a support vector machine- based filtering approach to eliminate the noise. And then Random forest is used for classification. SVM based filters are applied for analyzing and classification.

S. Dixon [1] Social-media are generally considered to be online platform for younger populations however people of all ages use such platforms for business purpose, politics, socializing and daily communication.

M. A. Khan and K. Salah [2] It aims to improve security professionals' understanding of threats and increase their skill levels in defending and mitigating them. It focuses on improving, understanding of the latest threats and increasing skill levels in defending and mitigating them.This paper analyzes the unclassified literature related to cyber ranges and safety test beds to gain a better understanding of the concept.

D. Halder and K. Jaishankar [3]Cybercrimes are defined as crimes performed over computer networks, such as the internet. "Transgressions that are obligated against individuals or groups of individuals with a criminal intent to intentionally damage the victim's reputation or cause the victim's psychological or physical harm directly or indirectly, using modern telecommunication networks such as the Internet (Discussion forums, e - mails, bulletin boards and groups) and mobile phones (Short message service)."

J. Srinivas, A. K. Dasand N. Kumar [4] It examines the importance of different cyber defense standards and cyber security framework architectures. Additionally, we discuss government strategies to protect cyberspace as well as national information security policies.Also discussed are national information security policies for securing cyberspace and government strategies for protecting cyberspace.

W. Clay[5] Random cyber attacks on computers on the Internet continue to plague computers around the world, and their perpetrators remain unclear.Most of the random attacks on computer resources are now increasingly implemented through the use of automated tools, called "bots",that direct large numbers of compromised computers to launch attacks through theInternet as swarms.The growing trend toward the use of more automated attack tools has alsooverwhelmed some of the current methodologies used for tracking cyberattacks on the Internet. People benefit from the Internet's anonymity to make their digital lives easier. It is anticipated but unavoidable that criminals will take advantage of the anonymity offered by the Internet.

I. M. Venter, R. J. Blignaut, K. Renaud, and M. A. Venter [6] The requirements for the Federal Government's evaluation of the US Department of Health and Human Services' cyber security policies are covered in this paper. Compliance with enacted Federal regulations and standards facilitates the overarching goal of cyber security policies and procedures, which is to safeguard the operational assets and objectives of the US Department of Health and Human Resources and to promote best security practices in the defence of information systems against unauthorized actors and cyber threats.

K. F. Cheung and M. G. H. Bell [7] This automation improves order delivery efficiency and decreases human error in order processing. Attacks from the internet, in particular, can jeopardize that, though. In this research, we present a novel attacker-defender paradigm Using the defensive budget and the reliance on features, we defend critical assets against a quantum response (QR) opponent. The solution's level of asset protection demonstrates how desirable it is to secure each asset.

Computer Attack Article accessed on 12/03/2013 [8] For analysis detection abnormal activities. According to this paper, based on various characteristics different types of anomalies and their categorizations are discussed. These anomalies creates various problems, which is to be handled carefully. These gives an idea of number of data mining techniques to detect anomalies. For Eg. Some of the malicious users may use false identities and use them communicate with large number of innocent users.

M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke [9] We define 35 well-known cyber datasets and classify them into seven categories: network traffic dataset, electric network dataset, internet traffic dataset, virtual private network dataset, android device dataset, IoT traffic dataset, and internet link. The dataset is crucial for intrusion detection.

R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman [10]The development of an intrusion detection system (IDS) that can quickly and automatically identify and categorise cyberattacks at network and host rates frequently makes use of machine learning techniques. However, a number of issues arise when harmful attacks are constantly changing and happen in a huge number, necessitating a scalable solution. The information security community has access to a large number of publicly accessible malware databases for further research.

M. A. Khan, S. K. Pradhan, and H. Fatima [11] The researchers started from case reports, extracting and determining the attributes/variablesof the cases. The a priori algorithm was applied to the set of variables to identify frequent item sets. Although the proposed algorithm was not implemented, the algorithm is useful for detecting the attributes and variables of cybercrime case reports. The researchers utilized visualizations, such as bar chart or graphs, to make analysis easier for investigators.

J. Yan, D. Jin, C. W. Lee, and P. Liu [12]This study explores the viability of off-line deep learning based NIDSes by developing the detection engine with various highly advanced deep learning models and undertaking a quantitative and comparative evaluation of these models. First, we discuss deep learning's general methodology and its theoretical implications for the problem of network intrusion detection. Next, we examine a number of machine learning approaches for two network intrusion detection tasks.
B. Zhu, A. Joseph and S. Sastry [13]Security is crucial in a digital age when it permeates both our public and private lives on a daily basis. If there is no security, everything will crumble. Attacks like WannaCry have devastated unprepared individuals, companies, and organizations, placing their operations in danger.

Kemal Hajdarevic,Adna Kozic and Indira Avdgic [14]Researchers uses these models to calculate the potential impact of disasters on crucial global social media platforms by looking at how assets are interconnected. A threat-based model is created, with each danger having unique destructible processes, weak spots, and consequences.

N. Virvilis, A. Mylonas, N. Tsalis and D. Gritzalis [15]For example, new technologies are continually being developed to combat hazards. Anyone who has been keeping up with the news is aware of how companies are handling cyber security threats. Files at businesses and institutions around the world have been encrypted until ransom demands are met. Cyber-security is a problem that transcends the IT industry. Its breadth is actually quite extensive. Nowadays, everyone is aware of the internet.

Vijay B, G. Ajay and A. Ala [16]Phishing attacks (Phishing attacks use emails sent from a reliable source to deceive the receiver into providing personal information. Identity theft can be readily caused by phishing. An email campaign known as a "phish assault" directs consumers to websites where they are prompted to enter sensitive data such bank account details, credit card numbers, or passwords.

INFOCOM [17]Identity-based attacks (IBAs) are one of the most serious threats to wireless networks. There have been several attempts to detect network attacks (Cyber Crime) using various techniques. Recently, received signal strength (RSS) based detection mechanisms were proposed to detect IBAs in static networks.

Software	Our Ratings	Best For	Category	Features	Free Trial	Price
Solar Winds Security Event Manager	5 Stars	Small to large businesses	Cloud based tool for SIEM	Threat Intelligence, SIEM Security and Monitoring, Log correlation and Analysis, Network and Host Intrusion Detection	Available for 14 days	It starts at $4500
Indeni	4.5 Stars	Small to large businesses	Behavioral Analytics, Incident Management	Indeni is an automated crowd-sourced cyber security platform for network and security infrastructure	Free for 90 days	Get a quote
Intruder	5 Stars	Small to large businesses	Cloud-based Vulnerability Scanner	Over 9000 security vulnerabilities, checks for web application flaws.	Available for 30 days	Get a quote
CIS	5 Stars	Small to large businesses	Cyber security tools	Securing Organization, Securing a specific platform and tracking specific threats	No	Free as well as paid subscription tools.
Cyber Control	4 Stars	Small and medium-sized businesses	Vulnerability scanning	Fraud detection reporting suite and file security review for data privacy and GDPR	Free trial available	Annual License - £29.99

III. CONCLUSION

We proposed an approach to improving the accuracy of the detection of Crime-related posts from Social Media text messages . Here, first, we applied a keyword-based filter, and then to remove the noise, we applied the SVM based filter and random forest classification. According to the existing works, SVM has the best accuracy among the classifiers. So, we used SVM for the research.

ACKNOWLEDGEMENT

We would like to thank our guide Mrs. P. Yamuna and Mrs. Soppari Kavitha for dedicating their valuable time and guidance. Also, we are extremely grateful to Dr. V. VIJAYA SARADHI, Head of Computer Science and Engineering Department for his invaluable time, support, ACE Engineering College.

REFERENCES

[1]. S. Dixon, ACM SIGCOMM Computer Communication Review, Volume 39, Number 5, October 2009, "A Brief History of the Internet"

[2]. M. A. Khan and K. Salah, Review of IoT security, block-chain solutions, and unresolved issues 2018's Future Generation Computer Systems.

[3]. Cybercrime and the Victimization of Women: Laws, Rights, and Regulations, by D. Halder and K. Jaishankar. USA: IGI Global, Hershey, PA, 2011.

[4]. J. Srinivas, A. K. Das, and N. Kumar, Future Generation Computer Systems, 2019. "Government rules in cyber security: Framework, standards, and suggestions."

[5]. W. Clay. (2005), Threats and Policy Concerns for Congress Aspect Of computer Attack and Cyber Terrorism Congressional Research Service report to Congress, 2005. reached on February 3rd, 2013 at http://www.history.navy.mil/library/online/computerattack.htm#summ.

[6]. I. M. Venter, R. J. Blignaut, K. Renaud, and M. A. Venter, "The three R's and cyber security education are equally important," 2019; Heliyon

[7]. K. F. Cheung and M. G. H. Bell, Eur. J. Operat. Res., 2019. "Attacker-defender paradigm for cyber security in logistics management against attackers with quantal responses: An introductory research.

[8]. http://www.fema.gov/pdf/onp/toolkit_app_d.pdf. ComputerAttack.Accessed on 12/03/2013.

[9]. M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, J. Inf. Secur. Appl., 2020, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative analysis."

[10]. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," IEEE Access, 2019.

[11]. M. A. Khan, S. K. Pradhan, and H. Fatima, "Applying data mining techniques in cyber crimes," in 2017 2nd International Conference on Anti-CyberCrimes (ICACC), 2017, pp. 213-216: IEEE.

[12]. J. Yan, D. Jin, C. W. Lee, and P. Liu, The International Conference on Ubiquitous and Future Networks published a paper titled "A Comparative Study of Off-Line Deep Learning Based Network Intrusion Detection" in 2018.

[13]. B. Zhu, A. Joseph and S. Sastry,"A taxonomy of cyber-attacks on SCADA systems," 4th International Conference on Cyber Physical and Social Computing and 2011 International Conference on Internet of Things, pp. 380-388, 2011.

[14]. Kemal Hajdarevic,Adna Kozic and Indira Avdgic,"Using the GNS3 Simulator to Train Network Managers in Ethical Hacking Techniques to Manage Resource Starvation Attacks,Communication and Automation Technologies (ICAT),International Conference on Information, Sarajevo, Bosnia-Herzegovina, pp. 1-6 , Oct 26- 28, 2017

[15]. N. Virvilis, A. Mylonas, N. Tsalis and D. Gritzalis, "Security Factors: Web browsing security vs. Blackguard sites", Comput. Secur., vol. 52, pp. 90-105, 2015

[16]. B. Vijay B, G. Ajay and A. Ala, Detection of masquerade attacks on Wireless Sensor Networks, 2010. Available at http://www.ists.dartmouth.edu/library/343.pdf.Accessed on 13/03/2013.

[17]. INFOCOM,Identity-based Attack Detection in Mobile Wireless Networks. Proceedings of the IEEE held in Shanghai, April 10-15, 2011, pp. 1880-1888.Available at http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp. Accessed on 02-04-2014.