IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI and ML techniques to Analyze Communication Emails and Text patterns To Secure from Attacks

Lakshmisri Surya, *Data Scientist, Department of Information Technology, USA*

Abstract— *My main focus on this research is to showcase the importance of analyzing communication emails and text patterns to maintain security with artificial intelligence and machine learning. This research shows the importance of artificial intelligence and machine learning and their impact on the present world. Artificial intelligence is one of the most innovative technologies found across the globe. It has revolutionized every field, and one of such areas is security. This research primarily deals with explaining how machine learning and artificial intelligence can generate results based on past emails scanning on which emails are crucially essential and attached with malware and scams. Machine learning algorithms can act as a great judge of filtering through emails and deciding which one is better and not right. This research also reviews a few of the literature written already about this topic assesses how our research improves on those points already mentioned in the past literature.*

INTRODUCTION

Artificial Intelligence is the most arising innovation on the planet at present. It is present in each field of life, whether it is clinical, training, media transmission, domestic, mechanical, etc. We can see various AI items all over the world. The world currently relies on AI creation. Artificial intelligence can be explained as developing the system and machine capacities to perform tasks and grasp requests much like a person would do it (Charniak, 1985). It suggests that PCs are adjusted to think and deal with issues like human mental capacity. Joining human agreement and understanding into machines involves Artificial Intelligence (Michalski, Carbonell, & Mitchell, 1983). Artificial intelligence's critical intention is to create useful smart gadgets to make decisions with no assistance from individuals. Facial acknowledgment, fingerprint scanner, motion and light sensor, voice affirmation, etc., are a segment of the renowned, standard, and crucial artificial intelligence techniques used in nearly every mechanical thing nowadays. It is the day and time of smart things. Electronic things with artificial intelligence in them are called smart items or things. These things can go from phones to washing machines, coolers, to vehicles. Smart things are expected to update the essential working of a device. Smart things often play out the necessary tasks they are intended for yet likewise proficiently play out a couple of various functionalities. In the current day and age, smart things are open worldwide and have gotten a genuine norm in our ordinary day-to-day existence.

The possible destiny of artificial intelligence is enormously extended and gainful. Associations that are placing assets into artificial intelligence or making smart things are the most significant. Google, Facebook, Tesla, Amazon, Apple, etc., are driving associations planning artificial intelligence techniques into their items. The destiny of AI shows extraordinarily high financial fledgling. In 2017, artificial intelligence regulated $8 billion (Columbus, 2017). This demonstrated a boundless beneficial future in placing assets into AI. By 2025, Artificial information is needed to have a market of $126 Billion (Bughin, et al., 2017). These estimations show a superb necessity for computerized reasoning today on the planet. Likewise, subsequently, AI gets enormous money-related benefits and advantages.

The rule of artificial intelligence will just spread way quicker and all the more broadly thought out the world later on. This procedure is simple in its early phases, and there will come when artificial intelligence assumes control over the whole world. All the instruments and processes we use will have smart strategies or techniques consolidated. The impact of artificial intelligence on various fields can be seen even at this point. In certain zones, artificial intelligence is only at the start, and in some areas, it is a legitimate veteran.

Machine learning, in any case, is a subfield of artificial intelligence. Machine learning is an interaction where machines are modified to gain from encounters and experiences (Mitchell & Jordan, 2015). Much the same as individuals, machine learning likewise centers around causing machines to learn by giving them truthful information instead of entering the data inside the devices. AI depends on the idea of predictive learning. It is expected that due to the machine learning training, the machine would have the option to settle on substantial choices all alone with no human help (Webb & Sammut, 2011).

AI and ML for email and text analysis

Artificial intelligence is useful in making smart gadgets and machines, and there are multiple other benefits of artificial intelligence. The most significant advantage of artificial intelligence is that it has dramatically reduced the work of human beings needed in any field. Now for any work, artificial intelligence is used. It has truly changed the landscape of what can be achieved through computers and technology.

Artificial intelligence has also become crucial in accessing security and providing the topmost safety to the users. Security of any information is an essential need. Any correspondence between two frameworks ought to be completely safe. The data moved electronically between two frameworks ought to be secured and guaranteed no misusing of important information. Numerous noxious programmers can hack the correspondence and can meddle with the framework. They can likewise gain admittance to delicate data that is being moved from the framework. To guarantee the most extreme security of information being transferred, one can apply the different security standards to the correspondence between emails.

The company must keep up its optimal security for an organization's accomplishment. Various risks can smash the system's operability and handiness. This way, it is essential for the organization to keep up its security reliably. Specific computer agents are seen as the most serious threat to the organization's computer systems' safety. When one system of an organization gets infected, it ruins every organization's system. Security is perhaps the most prominent part that every item should have. An application can work gainfully if it is ensured, and no danger can hurt the item's essential convenience and one of the most common ways security attacks occur is through mails. Emails become a hub of malicious software that transfers easily. This malicious software can get installed on a user's computer who opens the emails and can ruin the data and integrity o that computer. A user can also lose some precious data because of this malware. Malware is a little code of a program that can impact the computer and debilitate its working. Malware antagonistically influences information stored away in a computer as it degrades and eradicates essential information. Malware kills or changes the windows customers' capabilities to make a couple of individuals not sign into their profiles (Wafer, 2018). To stop malware, it is essential to keep up with the most recent antivirus programming that will alert the customer when malware is found, and then the same software will try to get rid of the said malware (Chung, 2016). Antivirus software can clear and fight off viruses whether or not the contamination has quite recently affected the computer or on the off chance that it is an old habitant of the framework.

Malware is overhauled continuously, and there is in every case, some new malware is found in the frameworks that are obscure and difficult to find. So, it is much preferred and always suggested by security officials to stop malware from ever entering the system. This can be done if the emails are scanned and analyzed accurately to judge any error or danger attached to that email. Often, malware is attached to the emails, and the users who received these emails never seem to know that these emails are corrupted and are infected with malware. So, they unabashedly open these emails and even discard them when they found nothing important of the sort in that email. Until then, it is too late, and the malware already infects the system without the user's consent.

Another primary way security is in danger through email communication is through phishing, which is a very common attack technique done through emails. Perhaps the most significant threat looked at by the world right currently is phishing. Phishing is a sort of attack technique where a cybercriminal impersonates a trusted source and contact the client of that source through email or instant message and attempts to gather and collect individual data like financial and banking private information from them (Lord, 2019). Phishing is a significant assault on organizational security. A digital criminal contacts client under fake personality through email or instant message and attempts to extricate essential data that can antagonistically influence the client.

Phishing is done when the cybercriminal gets data from the phishing units. These phishing packs are the assortment of programming that has made phishing amazingly basic. Phishing pack causes the cybercriminal to recreate a known association, which he can use to trick individuals into accepting the messages' legitimacy. These packs are planned in a manner that turns out to be very hard for security heads to perceive. The cybercriminals then send trick messages to clients' rundown that they have regularly got from the dark web. There are numerous locales on the dark web where this sort of data is handily sold and purchased. On many occasions, these spam messages are only for managing individuals into entering private data. Once in a while, these are just requesting individuals to open a specific fake site that can bring about downloading infectious code of programs or malware.

These phishing and malware messages are becoming very persuading. They regularly utilize the name of enormous organizations like Microsoft, Apple, and other such organizations to fool individuals into conveying data. Spam email envelope is the greatest answer for these phishing assaults. Consequently, the spam envelopes identify any unpredictable or dubious email and keep it from entering the inbox, which causes the client to see these hazardous messages. All the email specialist organizations like Gmail or Outlook have an in-constructed spam inbox envelope. They provide the basic functionality of filtering every email, and they decide whether or not some emails are dangerous or spam. If those emails are not considered spam, they are sent to the inbox. But if the computer thinks it is spam or harmful to the customer, it sends it to the spam email.

These spam filters are a result of artificial intelligence and machine learning. These techniques are utilized to filter through emails. Artificial intelligence bots are trained to go through hundreds of thousands of emails to recognize fake ones' genuine emails. After getting the training on how a fake email looks, they check every email that comes to the server and decides whether it is a spam email or a genuine one. It is important to note that every seldom does happen that the machine learning bot is wrong in its prediction and can declare a simple email as a spam one, but that happens very seldom. At most times, the machine correctly distinguished the wrong email from the right.

There are continuously new concepts and algorithms introduced in machine learning, making it possible to distinguish emails more quickly. There are even new algorithms that filter the emails with malware hidden behind them and find scam emails that may look normal at first look but can also be deceptively tricky in their writing. In this way, text analysis patterns are utilized to make the decision easy for the machines. Test analysis is a process in which the text is analyzed, and a definite answer is given based on that text. Machine learning software performs text mining and text analysis to conclude results on whether an email is corrupted or not.

Text is considered valuable data. And data is given tremendous importance in machine learning and artificial intelligence because, in this procedure, it is only because of

data that these machines have gathered through experience and learning. The data based on which they make valuable decisions like whether an email is fake or not. As portrayed in the article" Business data mining — a machine learning perspective," data mining has become a vital piece of a business advancement (Bose & Mahapatra, 2001).

Text mining is a part of data mining. The only difference that separates the two procedures from each other is that the vast collection of data is usually organized or semi-organized in data mining. In contrast, in text mining, information is generally scattered all-around social media platforms and emails, and it is unorganized (Boulicaut & Masson, 2005). So, it becomes more difficult to conclude valuable information from that disorganized data collection in text mining. In text mining, the AI bots are responsible for generating valid data from the enormous collection of text, images, videos, and other information gathered from an extensive collection of emails. Afterward, the AI bot decides a conclusion about that massive collection of data. It becomes valuable for AI bot and ML algorithms to find what kind of emails have been seen as right and what kind of emails have been considered fake. So, it has only because of the help from data mining done through artificial intelligence that email analysis and text patterns have been crucial in ensuring security.

There are a few famous machine learning algorithms used for supervised learning. Naïve Bayes classifier is one such type of algorithm based on Bayes theorem, and it is explicitly used to solve classification problems. It is considered one of the most straightforward and most easy-to-use machine learning algorithms (Gandhi, 2018). Another critical method or algorithm used in machine learning to filter spam emails is Multilayer Perceptron (MLP) classifier. It is an algorithm that uses a technique known as backpropagation for training. It is crucial in the way that it can diagnose multiple layers and nonlinear data as well (W.BauerJr. & Belue, 1995).

Literature review

There has been a lot of research already done on this topic. Some of these have been crucial for me in understanding the importance of machine learning and artificial intelligence. I have focused on discussing few key literature sources in this research.

One such article is titled "Classification of Phishing Email Using Random Forest Machine Learning Technique," written by Andronicus A. Akinyelu1 and Aderemi O. Adewumi. It was published in the Journal of applied mathematics, first published in 2014. This article focuses on a machine learning algorithm used to determine phishing emails and recognize phishing attacks. It also discusses the aspects that an improved phishing classifier can have. It takes the data from around 2000 phishing emails (Akinyelu & Adewumi, 2014).

Another major article that was helpful in my research was "Machine learning for email spam filtering: review, approaches, and open research problems" by Emmanuel Gbenga Dada, Joseph Stephen Bassi, Haruna Chiroma, Shafi Muhammad Abdulhamid, Adebayo Olusola Adetunmbi, and Opeyemi Emmanuel Ajibuwa published in 2019 in Heliyon volume no 5 and issue no 6. This article reviews multiple machine learning algorithms and techniques that are important in filtering spam and phishing emails. It also efficiently compares each machine learning approach (Abdulhamid, et al., 2019).

"Email Spam Filtering using Supervised Machine Learning Techniques" by V. Christina, S. Karpagavalli, G. Suganya is another important article dealing with email spam filtering techniques used by email providers to filter spam emails from the genuine ones. It explicitly names many email spam filtering techniques (S.Karpagavalli, Suganya, & Christina, 2010).

Another important literature article that provided significant help in my research is "Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets" by Nurul Fitriah Rusland1, Norfaradilla Wahid1, Shahreen Kasim, and Hanayanti Hafit1. It teaches the Naive Bayes algorithm's analysis taking datasets and performing email filtering techniques using that algorithm to analyze its accuracy (Rusland, Wahid, Kaism, & Haift, 2017).

"Determining input features for multilayer perceptrons" is another article that I studied to understand this topic. This article dealt with understanding the primary method through which the machine learning technique of multilayer perceptron can be improved to give better results, and more qualified answers were. Lisa Belue and K. W Bauer wrote this article, and it was published in the 1995 issue of Neurocomputing journal (W.BauerJr. & Belue, 1995).

The article "Comparing Manual Text Patterns and Machine Learning for Classification of Emails for Automatic Answering by a Government Agency" by Hercules Dalianis, Jonas Sjöbergh, and Eriks Sneiders was extremely beneficial in my studies. It delves deeper into the positive application of using machine learning and text patterns to generate how well the overall procedure helps the government agencies respond and answer emails more uniformly. The researchers of this article focused on 4,404 emails that Swedish social insurance agencies received from citizens. And after evaluating these emails, the researchers found out that text-based patterns proved more beneficial than traditional machine learning methods in finding answerable emails (Dalianis, Sjöbergh, & Sneiders, 2011).

Finally, the article that I considered for increasing my knowledge on this topic was from, and it is titled "Classifying Phishing Email Using Machine Learning and Deep Learning." This article specifically dealt with how emails can be determined as phishing or non-phishing emails. The research team is working on this article focused on applying deep semantic analysis, deep learning techniques, and machine learning algorithms to determine an email's validity to decide whether it's a phishing email or not (Nandi, Bagui, White, & Bagui, 2019).

Conclusion

My research above can be concluded from the fact that artificial intelligence and machine learning have been developed as ground-breaking and innovative strategies that have changed the landscape of all computer techniques. In addition to new techniques, it has also helped update every field, and security is one such field. Now, it has become possible to maintain maximum security all around the system through the help of artificial intelligence.

My research specifically focused on how artificial intelligence has been fruitful for the safety of emails and how analyzing communication emails and text patterns have become crucial for maintain top security. Malware and phishing emails are a serious threat to the sanctity of emails, and they can be prevented with the help of machine learning.

REFERENCES

[1] Abdulhamid, S. M., Dada, E., Bassi, J., Chiroma, H., Adetunmbi, A., & Ajibuwae, O. (2019). Machine learning for email spam filtering: review, approaches, and open research problems. *Heliyon*.

[2] Akinyelu, A. A., & Adewumi, A. (2014). Classification of Phishing Email Using Random Forest Machine Learning Technique. |*Journal of Applied Mathematics*.

[3] Bose, I., & Mahapatra, R. (2001). Business data mining — a machine learning perspective. *Information & Management Volume 39, Issue 3*, 211-225.

[4] Boulicaut, J. F., & Masson, C. (2005). Data Mining Query Languages. In O. Maimon, & L. Rokach, *Data Mining and Knowledge Discovery Handbook* (pp. 715-726). Boston: Springer.

[5] Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlstrom, P., . . . Trench, M. (2017). *ARTIFICIAL INTELLIGENCE THE NEXT DIGITAL FRONTIER?* McKinsey Global Institute.

[6] Charniak, E. (1985). *Introduction to Artificial Intelligence.* Pearson.

[7] Chung, E. (2016, July 8). *Antivirus software is 'increasingly useless' and may make your computer less safe*. Retrieved from CBC: https://www.cbc.ca/news/technology/antivirus-software-1.3668746

[8] Columbus, L. (2017, July 9). *McKinsey's State Of Machine Learning And AI, 2017*. Retrieved from Forbes: https://www.forbes.com/sites/louiscolumbus/2017/07/09/mckinseys-state-of-machine-learning-and-ai-2017/#2899ab8275b6

[9] Dalianis, H., Sjöbergh, J., & Sneiders, E. (2011). Comparing Manual Text Patterns and Machine Learning for Classification of Emails for Automatic Answering by a Government Agency. *International Conference on Intelligent Text Processing and Computational Linguistics*, (pp. 234-243).

[10] Gandhi, R. (2018, May 5). *Naive Bayes Classifier*. Retrieved from Towards Data Science: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

[11] Manishaben Jaiswal "RISK ANALYSIS IN INFORMATION TECHNOLOGY" , International Journal of Scientific Research and Engineering Development (IJSRED) , ISSN:2581-7175, Vol 2-Issue 6, P110, pp. 857-860, November - December 2019 Available at: http://www.ijsred.com/volume2/issue6/IJSRED-V2I6P110.pdf

[12] Lord, N. (2019, July 12). *Phishing Attack Prevention: How to Identify & Avoid Phishing Scams in 2019*. Retrieved from Digital Guardian: https://digitalguardian.com/blog/phishing-attack-prevention-how-identify-avoid-phishing-scams

[13] Michalski, R., Carbonell, J., & Mitchell, T. (1983). *Machine Learning An Artificial Intelligence Approach.* Berlin: Springer.

[14] Mitchell, T. M., & Jordan, M. (2015). Machine learning: Trends, perspectives, and prospects. *Science Vol. 349, Issue 6245*, 255-260.

[15] Nandi, D., Bagui, S., White, R. J., & Bagui, S. (2019). Classifying Phishing Email Using Machine Learning and Deep Learning. *2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, (pp. 1-2). The United Kingdom.

[16] Rusland, N. F., Wahid, N., Kaism, S., & Haift, H. (2017). Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets. *IOP Conference Series: Materials Science and Engineering, Volume 226.* Melaka, Malaysia.

[17] S.Karpagavalli, Suganya, G., & Christina, V. (2010). Email Spam Filtering using Supervised Machine Learning Techniques. *International Journal on Computer Science and Engineering 2(9)*.

[18] W.BauerJr., K., & Belue, L. (1995). Determining input features for multilayer perceptrons. *Neurocomputing*, 111-121.

[19] Wafer, A. (2018, January 20). *FIX: Credential Manager Not Working in Windows 10*. Retrieved from Windowsreport: https://windowsreport.com/11939-credential-manager-no-working-windows-8/

[20] Webb, G. I., & Sammut, C. (2011). *Encyclopedia of Machine Learning.* Springer Science & Business Media.