Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/170622

A Predictive Technique using Random Forest Classifier for Phishing Malicious Attack

Abdullah Alhaj[1,*], *Mua'ad Abu-Faraj*[2] *and Basel J. A. Ali*[3]

[1]Department of Information Technology, Faculty of Information Technology and Systems, The University of Jordan, Aqaba 77110, Jordan
[2]Department of Computer Information Systems, Faculty of Information Technology and Systems, The University of Jordan, Aqaba 77110, Jordan
[3]College of Economics and Management (CoEM), Al Qasimia University, Sharjah, United Arab Emirates

Received: 10 Apr. 2023, Revised: 2 Jul. 2023, Accepted: 28 Aug. 2023
Published online: 1 Nov. 2023

Abstract: A person with an email account is always vulnerable to fraud. An email account can be exploited by using a type of social engineering attack where the attackers trick the victims to steal user credentials by masquerading as a trusted entity. Email phishing has become the major action performed in various sectors such as banking, business, any enterprise or social media etc. While the action of phishing, the attackers make use of another technique called email spoofing. Email spoofing is not much different from email phishing since the email spoofing involves the usage of forged email header pretending as an entity created by a person of a trusted source. Phishing always has a malicious intent which means the person behaves knowingly or purposefully to cause them harm without a legal reasoning. Since the globe has more victims, we come across a large dataset. The major objective of the study is to determine the performance factors based on the phishing using random forest classifiers. For analysing a predictive model, we need a proper technique or an algorithm. In this case the random forest algorithm is accurate because it is built with many decision trees that produce a predictive model about the error rate.

Keywords: Phishing, Email Spoofing, Decision Trees, Classification, Malicious Attack.

1 Introduction

Phishing is an attack that aims to take the user's information such as credit card numbers and login credentials. It is a computer security term for the process of sending emails that appear to be from a source that occurs when an attacker, this is done by masking themselves as a trusted source, fooling a victim to open an email, text message, or instant message. The receiver is then misled to click a malicious website, which intent leads to the installation of malicious software consequently, halting of the system [1]. Nowadays, phishing attacks have changed quite a lot, and become more harmful than earlier. Through the login process in social platforms, an unauthorized person can occasionally perform a data breach on a person by some data like a phished password, which leads to the vulnerability called ransomware attack. Now, many new technologies are also used. For an example, In U.K. an energy firm CEO thought that they were talking to their boss on the phone, meanwhile they were tricked by the influence of the CEO's chief management from the main concern through an AI feature who asked money to a peculiar distributor, which is a phishing scenario [2]. It is uncertain whether the attackers made use of machines like bots to operate the victim's enquiries. The investigation might be tough, suppose the phisher has utilized both for the attack automation.

Typically, the Phishing attacks pass on social networking approach put on to the email or any electronic communication ways. The ways consist of direct messages which are sent through SMS text messages and social networks. The public resources of information are used by phishers to collect background information such as private and work reports, activities and interests of the victim. Generally, to craft a believable email some social networks such as Twitter, Facebook and LinkedIn are usually utilized to disclose information like the names, job descriptions and email address of possible users [16].

Realizing Phishing Email

Fortunately, phishing emails are very complicated to differentiate commencing actual emails. Because they

*Corresponding author e-mail: aa.alhaj@ju.edu.jo

appear with similar commercial logos as well as others gathered recognizing information, pretending themselves as an entity from a well-known company [3].

Although, there are diverse hints which can specify that an email is a phishing attempt. Some such clues are given below:

- The sub domains, misspelled URLs (typo squatting) or otherwise suspicious URLs are used in the message.
- The Gmail and public email addresses are used other than a commercial address of the email by recipients.
- The written email might invoke panic or else a temptation of urgency.
- A demand to authenticate personal information, such as economic details or a password is included in the communication.
- The message may have poor content, spelling mistakes and grammatical errors.

Moreover, a phishing attacker doesn't target a single employee from a company [7]. Alternatively, an attacker will send a certain number of emails, potentially even sorting through the entire organization's directory. For a successful attack, the attacker barely needs a single person to sink for a scam; executing such an extensive attack increases their possibilities.

Commonly, a victim gains a message which pretends to be sent by a recognized person or a company. The attack is accomplished then either during a malicious attachment of the file, or through other links navigating to the cruel sites. No matter what in both cases, the intention is to enable the malware on the victim's machine or to deviate the user to a false website [12]. These fake sites are prepared to fool users into disclosing individual data and economic information, like account IDs, passwords, or credit card details. Usually, the emails of phishing are horribly enveloped and obviously fake, cybercriminal collections progressively employ the similar methods expert vendors use to find the mainly successful kinds of emails.

Mostly the susceptible emails can be identified by thorough reading of the mail. For detecting phishing emails, IT professionals typically use a three-stage method. The email reader attempts to grasp the email's meaning and determine how it connects to certain other aspects of their life in the initial stage. As they go, they detect inconsistencies, or little aspects of the email that seem "wrong." The more anomalies the receiver finds, the more they feel the need for a different justification for the email. At certain times, they become aware that phishing is a potential other explanation due to some aspect of the email, typically the existence of a link asking for a response. Now that they have turned suspicious, they dig into the email further by searching for technical information that would positively identify it as phishing. After discovering this data, they go to the next phase and decide whether to delete or report the email [15].

Contribution of the study:
The current study demonstrates the following contributions:

- The study provides a detailed explanation of the cyber-attacks mainly phishing, the applications, how the attackers commit these attempts, and how to identify the suspicious emails.
- The countermeasures are also discussed in this study, which is one of the main advantages of this research. This also discusses how expert people identify malicious mails and how common people identify them.
- Random forest classifier is explained from a critical angle. As a result, the study determines the performance factors of the classifier.

2 Application of Phishing Attacks

Phishing scams come in all variants with different shapes and sizes. Users must stay secure, aware and be organized by studying several of the most relevant methods that scammers execute phishing. Some applications of further new attacks of phishing as follows:

Digital Payment Based Phishing Attack

This happens whilst foremost fee programs and websites are used as an entity to benefit touchy facts from phishing victims. In this swindle, a phisher pretends to be a web fee provider along with Venmo, PayPal or Transfer wise. Basically, those assaults are executed thru email, wherein a faux model of a depended-on fee provider asks a consumer to affirm login credentials and different figuring out facts [14]. Generally, they declare that that is crucial for you to remedy a problem with the consumer's account. Regularly, those phishing tries encompassing a hyperlink to a deceptive "spoof" web page.

PayPal is properly knowledgeable of those threats as well as has launched informational substances for his or her consumers to refer to the material for you to live organized in opposition to phishing assaults. They advocate that all and sundry who get a suspicious e-mail from an account putting

forward to be PayPal have to know no longer click on any links, instead, employ the soaring approach mentioned to look if the hyperlink copes with PayPal's real domain [11]. PayPal additionally advocates logging in to their account to make certain the whole thing seems like it has to. If a consumer is unaware of an imitation online disbursement phishing email, there are a small number of former details to look out for. Commonly, a phishing email from PayPal has been recognized to contain:

- Dodgy salutations that don't encompass the sufferer's call. The legit emails of PayPal will constantly deal with customers through their real call or enterprise title. Phishing tries in this area have a tendency to start with "Dear user," or use an e-mail address instead.
- In PayPal and in lots of different on-line charge services, maximum of those frauds supplies an "alert" to their feasible victim as their account might be hanged soon. And others claim that their customers have been accidentally "overpaid" and now want to ship cash again to an unauthorised account.
- The downloadable connected documents aren't sent through PayPal to their customers. If a person obtains an email from PayPal or some other comparable provider that includes a connected document, they ought to no longer try and download it. If someone gets one among the emails, they must release their imbursement web page on a different browser tab or window and check whether their mail account has any alert notifications. If a customer overpays or is meeting deferment in their account, it will say so there. Incrementally, PayPal tends the consumers to convey any apprehensive activity to them, so they can proceed to observe these endeavours and avoid their consumers from being scammed [17].

Finance-based Phishing Attack

There are traditional forms of phishing as well as it plays on the belief that sufferers may be anxious to offer the private records to the scammer. Generally, in those situations, the scammer behaves as a financial institution or a few different monetary institutes [9]. In an e-mail or a telecall smart phone call, the scammer conveys that their possible suffering along with their safety has been demolished. Frequently, scammers will occupy the hazard of identification robbery to efficiently do simply that. A small number of examples of this scam comprise:

- There are apprehensive emails containing cash transports, a good way to confuse the sufferer. In this phishing operation, the ability sufferer receives an electronic mail which has an acceptance or refusal of electronic mail concerning the transfer of AHC. Repeatedly, the sufferer who observes this electronic mail will immediately expect that in the event that they click on an awful hyperlink with inside the message a fraudulent price may be made of their account which would depart their private information liable to being mined.
- Direct deposit scams are again and again utilised on new personnel of an organisation or business. The sufferers get hold of notification that their login data isn't always working. They emerge as demanding, approximately now no longer getting compensated, the sufferers click on a "phishy" hyperlink within the electronic mail which leads the sufferer to spoof an internet site that establishes malware into their system. Commencing there, the banking data is liable to be declared easily, mainly due to deceitful charges [18].

Work Based Phishing Attack

Varieties of rip-off are more often than not very customized and tough to identify and are in particular alarming [7]. In this operation, an enemy claim to be the recipient's head; The contacts of CEO or CFO the sufferer, and appeals for a twine switch or a faux purchase.

Work-associated rip-off has been cracking up round corporations with inside the closing depend on years is a tactic to attain passwords. This rip-off time and again targets the executive-degree of employees, due to the fact they're optimistically now no longer thinking that an email from their head might be a rip-off. The fraudulent email regularly operates due to the fact, apart from being alarmist, it simply talks about ordinary places of business subjects. Generally, it informs the sufferer that an organised assembly wishes to be changed.

From there, the worker is assigned to fill out a ballot that once a terrific time to reorganisation can be through a hyperlink. That hyperlink will then be delivered to the sufferer to a spoof login web page for Microsoft Outlook or Office 365. As they received and entered your login information, the scammers acquired their password [19]. Recognizing Phishing Attacks

- The preliminary step to have the least chance of becoming a victim of phishing is to recognize the signs. The most common red flags which indicates a phishing email are as follows:
- A wrong, unknown, unfamiliar, long, or else suspicious-looking sender address.

- Requests for private or enterprise information in any uninvited message.
- Requests that you manipulate account information or perform other online tasks.
- Requests for personal credentials from apparently official institutions. (Government agencies and legitimate financial institutions never do this.)
- Any message that uses dread, warning, a perception of urgency, or an offer too good to be true as a motivator.
- Any message with a title consisting of words like "Action required!" or "Urgent!"
- Any message that consists of an odd or unexpected subject content, punctuation errors, incorrect grammar, misspellings etc.
- Messages from trusted groups such as co-workers, friends, enterprise executives which requests personal data like phone number, personal email alias, address or that claim "off" in subject content, grammar, tone and structure, embedded links, time of day sent, etc.
- Abbreviated URLs, which repeatedly directs users to phishing web pages.
- Messages sent to an unfamiliar or unlikely set of receivers.
- Messages with an ungainly spelled or non-specific greeting (Hello Customer, Hi Dear) or usage of a full name (Hi John Doe) [19].

3 Email Spoofing

Email spoofing is a sort of cyber-assault wherein a hacker mails an electronic mail that has been operated to seem like it's far advanced from a depended-on source [10]. An electronic mail spoofing is a well-known plan utilized in junk mail campaigns and phishing considering humans are greater credible to reveal an electronic mail after they assume this has become despatched through an acknowledged sender. The electronic mail spoofing is to ruse receivers into disclosing or reacting to the message. The objective of spoofing is to rude customers into trusting the e-mail is despatched from somebody they recognize or can believe in maximum conditions it is able to be a brand, seller or colleague. Utilizing that trust, the attacker enquires the receiver to expose data or obtain a few different operations.

The act of electronic mail spoofing is sending emails with incorrect sender addresses, typically as a part of a phishing assault created to pilfer your data, infect your laptop with malware or simply ask for money. Common payloads for malicious emails consist of adware, ransomware, Trojans (like Emotet), cryptohackers or malware that enslaves your laptop in a botnet (see DDoS).

Email spoofing is the development of an electronic mail header with inside the hopes of deceiving the receiver to assume as the email is advanced from someone or an area apart from the deliberated source. Since a centre electronic mail protocol doesn't have an integrated approach of authentication, it's a normal region for junk mail and phishing emails to apply the spoofing to trick the receiver to agree with the beginning of the message [20].

Email spoofing can occur due to the fact the Simple Mail Transfer Protocol (SMTP) doesn't have a mechanism for coping with authentication [11]. Even though electronic mail cope with authentication protocols and mechanisms is being advanced to combat in opposition to electronic mail spoofing, nomination of those mechanisms is being unhurried. From the maximum broadly used attacks, electronic mail spoofing takes region while the sender produces electronic mail headers to that patron software program which indicates the crooked sender copes with, wherein maximum customers seize at face value. But they discover the header closely; electronic mail receivers think the located sender has despatched the message. If it's a call they know, they could agree with it. Spoofed emails usually request permission to get admission to a gadget or cash transfer.

Incrementally, they are able to at instances incorporate attachments which deploy malware like Trojans or viruses while opened. In maximum cases, the malware is advanced to head over infecting your pc and unfold to the entire network. Email spoofing is based deeply on social engineering the functionality to meet a human consumer to consider that what they're seeing is legitimate, tempting them to do so and open the connected file, switch cash, and so on [21].

How Email Spoofing Works

An attacker might also additionally expand an email that looks as if it arrives from PayPal. The communication conveys to the receiver that their account might be suspended in the event that they no longer click on the link, substantiate it into the webpage and alternate the account's password. If the receiver efficiently does this and brands in

information, the attacker receives the information to authenticate into the objected PayPal account consumer, and attempts to steal the cash from the consumer account. Most compound assault's purpose monetary personnel and employee online reconnaissance and social engineering to swindle a designated consumer into transferring big quantities to the account of attacker's. For the consumer, a spoofed e-mail message appears as if justifiable, and plenty of attackers would possibly obtain factors from their reputable internet site to create a view of a plausible message [22].

Outgoing email is returned, and its miles routed the usage of the Simple Mail Transfer Protocol (SMTP). Whenever the consumer clicks "Send" in an email to purchaser that despatched message is first of all despatched to the extroverted SMTP server that's configured inside the purchaser's software. The SMTP server recognizes the receiver's area and directs it to the area's e-mail server. The receiver's email server then routes the communication to the precise consumer inbox. For each and every "hop" an email takes because it moves via the net from server wise, and the IP deal with every server is blanketed along with logged within the email headers. Like these headers monitor the genuine course and sender, however earlier than interacting with an e-mail sender, many customers does now no longer take a look at headers.

Generally, three foremost sections of an email are:

- Address of Sender
- Address of recipient
- Body of email

An additional detail regularly utilized in phishing is the reply-to area. This is very much like configurable from the sender and may be utilized in an attack of phishing. Besides, the SMTP protocol and e-mail servers no longer validate whether or not this email is cast or legitimate. It's as much as the person discovering that the response is dispatched to the incorrect recipient.

In 2014, the Sender Policy Framework (SPF) was ready as a well-known protection protocol. This works in conjunction with the usage of DMARC (Domain-primarily based totally Message Authentication, Reporting and Conformance) to terminate phishing assaults along with malware. SPF can locate with spoofed email, and it has emerged as regular with nearly many email offerings to struggle with phishing. In SPF, the area holder must organize a DNS TXT access that specifies all IP addresses legal to ship email on welfare of the area. With this DNS access configured, receiver's email servers select out the IP cope with while accepting a message to make certain that it fits the email area's legal IP addresses. If there's a contest, the Received-SPF area presents a PASS repute. If it isn't always equal, its presentations have a FAIL reputation within the area [22,23].

Email Spoofing Techniques

The email authentication technologies like SPF, DMARC and DKIM are free to use for an organization. These technologies might become very useful for that organization and for the domains out in the cyber world.SPF (Sender Policy Framework): It inspects whether or not a positive IP is permitted to ship mail from a given area. SPF may cause a false positive, but the challenge of analysing SPF file and validating the sender are accomplished via the means of the receiving server.

The SPF mechanism utilises the area with inside the return-direction deal with to pick out the SPF file. When a sender tries to switch an email to an email "receiving" server for delivery, the server inspects to test whether or not the sender is at the area's listing of allowed senders. In that case, then a hyperlink is being hooked up among the piece of e mail and the e-mail area. Otherwise, the server proceeds processing the email as an acknowledged one without this hyperlink, wherein sever matters will be going on.

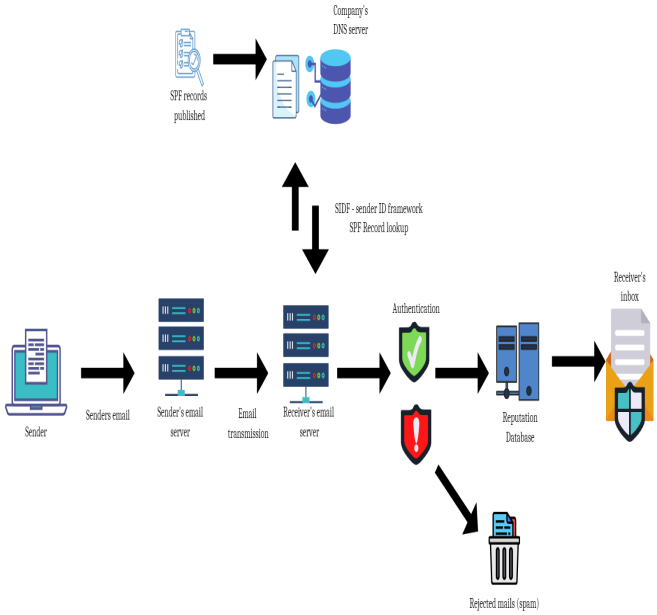

Fig. 1: Mechanism of Email Spoofing.

In figure 1 the mechanism of email spoofing is illustrated. Utilizing fraudulent emails, attackers frequently perform spear phishing and social engineering operations against organisations. An attacker might raise the chance that their victim would accede to a request, like downloading a malicious attachment or revealing credentials, by altering the offender's identity or other elements of a header section to make it seem as though the email was sent from a different

source. The e-mail can be an actual one, however the file of senders won't be definite. Real email might also additionally have forwarded which broadcasts the email may have come from everywhere and the file of accepted senders do now no longer assist that plenty. Or else, the email exists undesirable and fake. Too many possible effects make it tougher to connect that means to the absence of the hyperlink that SPF ought to provide. DKIM fills the distance within the DMARC technical framework as an additional technique to try and hyperlink a chunk of e-mail lower back to an area [24].

SPF and DMARC for Email

SPF can join a chunk of email with an area. With the DNS facts in place, DMARC binds the effects of SPF to the content material of e-mail, specially to the area determined within the go back course or From: the header of the e-mail. And for SPF to paint well within the context of DMARC, the go back-course deal needs to be relevant to the area of the From: header, that is the object that binds collectively DMARC alignment.

Importance of SPF

SPF has turned out to be enormously important to assist in showing which sending structure can impart email on behalf of your area. Accomplishing SPF for email materials wonderful benefits:

- Enlarges email deliverability and area reputation.
- Battles the email spoofing and domain impersonation to protect your brand reputation
- One of the fundamental methods of email authentication for DMARC.

Domain Keys Identified Mail in short DKIM, while this is an email authentication method. This authentication method is used to detect fake or spoofed sender email addresses [25].

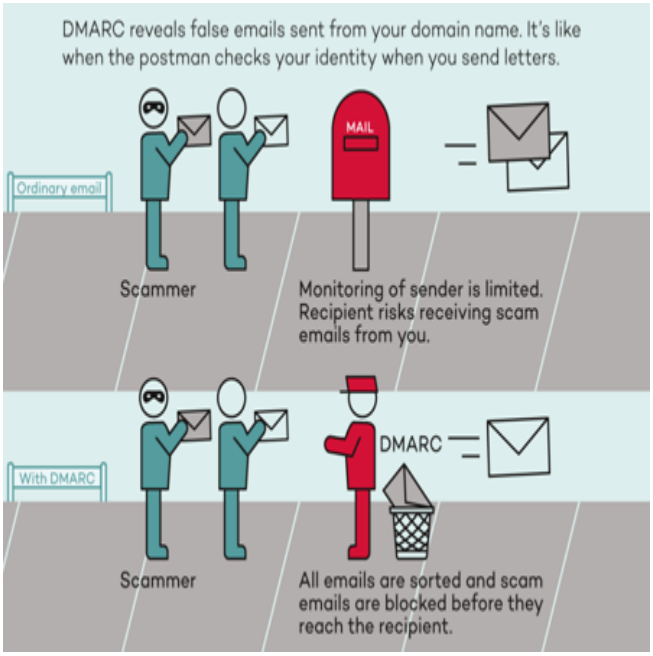

Fig. 2: Secure Email with DMARC.

At a glance from figure2, it can be postulated that while making use of DKIM method, a sender might attach DKIM signatures to an email enclosed with a header that is incremented to the message and is protected with encryption, and when the receiver acquires the email, they can inspect that it is literally you who sent it. The sizable reason why DKIM is so indispensable for your enterprise is since the spoofing emails from a believed domain is a well-liked technique for phishing campaigns, and DKIM builds it harder to spoof emails from the domains that utilize it.

DKIM signing can be via forwarding. If your area is enfolded with DKIM, dmarcian's capability to hit upon forwarding increases. SPF doesn't function within the context of forwarding, as SPF is absolutely a report of servers that are legal to ship on behalf of your area, and it's far not possible to keep a listing of forwarders for a website owner [26].

- DMARC-Domain-Based Message Authentication, Reporting, and Conformance: DMARC technique provides a dispatcher the opportunity to allow the recipient to understand whether or not its electronic mail is covered with the aid of using DKIM or SPF, and what are the measures to be taken whilst an electronic mail fails authentication process. DMARC isn't so popularly used.
- DMARC gets rid of guesswork from the recipient's management of those failed messages, lowering or putting off the user's vulnerability to in all likelihood fraudulent & dangerous messages. DMARC additionally provides a scheme for the email recipient to file lower back to the sender

approximately messages that fail and/or by skipping DMARC evaluation [26].

4 Random forest Classifier

A supervised learning algorithm called Random Forest is employed for mutually classifications together with regression. As we are familiar that a forest is built of trees and many more trees which means additional robust forest. Similarly, a random forest algorithm enlarges decision trees on data samples and then gains the prediction from each and every single data and in the end; it chooses the most excellent solution by mode of voting. It is a composite method which is progressive than a particular decision tree due to averaging the result, it reduces the over-fitting.

A random forest is a technique used in machine learning which is employed to work out classification and regression issues. It makes use of ensemble studying, while this is a procedure that collaborates numerous classifiers to provide explanations to complex issues. A random forest algorithm is made up of many decision trees. The 'forest' is brought by the random forest algorithm which is instructed by bootstrap aggregating or bagging.

The random forest algorithm exhibits the result which depends upon the decision tree's estimations. It is predicted by finding the regular or mode of the results from different trees. Improving the number of trees enlarges the correctness of the result. A random forest completely destroys the restriction of a decision tree algorithm. It decreases the overfitting of datasets and improves accuracy. It brings predictions without using many configurations in packages such as scikit-learn [27].

This is more error free compared to the decision tree algorithm.

- Can provide a sensible estimation lacking with hyper-parameter tuning.
- It will also resolve the trouble called overfitting in decision trees.
- However, at the node's splitting point, a subset of features is chosen randomly in every random forest tree.

Working structure of Random Forest algorithm

Step 1: Initially, begin with the selection of random samples from a provided dataset.

Step 2: Then, this algorithm will develop a decision tree for each sample. And it will obtain the forecast consequence from the complete decision tree.

Step 3: Through this step, voting will be carried out for all estimated results.

Step 4: Finally, choose the majority voted estimate outcome as the final estimation effect. The following figure will demonstrate its working (see figure 3):

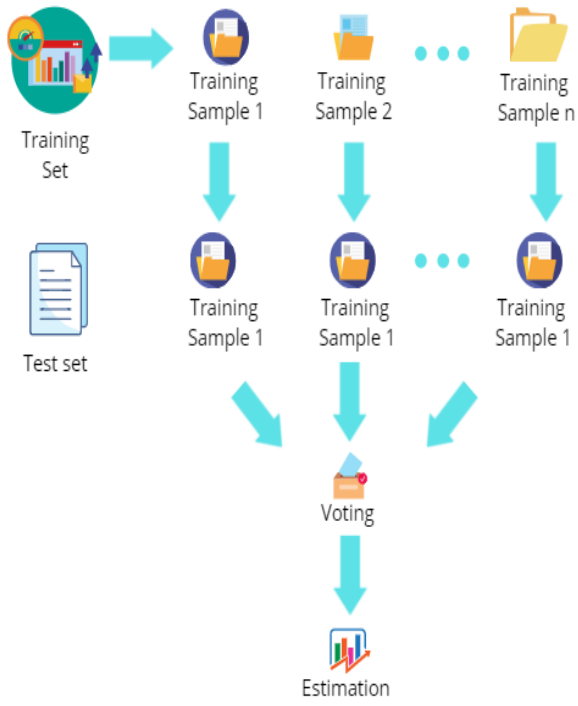

Fig.3: Working Structure of Random Forest Classifier.

While understanding the building structures of the random forest algorithm are known as decision trees. The technique of decision support like decision trees is being utilized, where it creates a tree-like architecture. A summary of decision trees would hold up us to study the performance of random forest algorithms [6].

In a decision tree it is composed of three parts such as decision nodes, leaf nodes, and a root node. The decision tree algorithm bisects a training dataset to diverse branches, which additionally divide into other branches. Such progression is carried out until a leaf node is achieved. The leaf node is not separated additionally. The nodes of the decision tree constitute attributes that are operated to forecast the consequences. Decision nodes furnish a connection to the leaves [30]. Figure 4 demonstrates the three kinds of nodes in a decision tree.

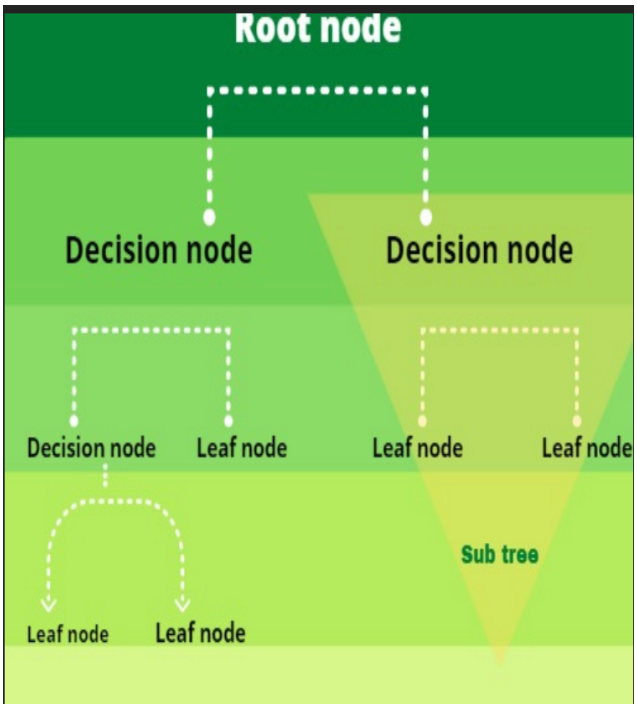

Fig.4: Decision Tree with Nodes.

5 Experimental Results

With the assistance of the Weka tool the Random Forest Classifier prediction technique is derived along with a malicious hacking database. 499 malicious hacking instances are considered for improving the error rate.

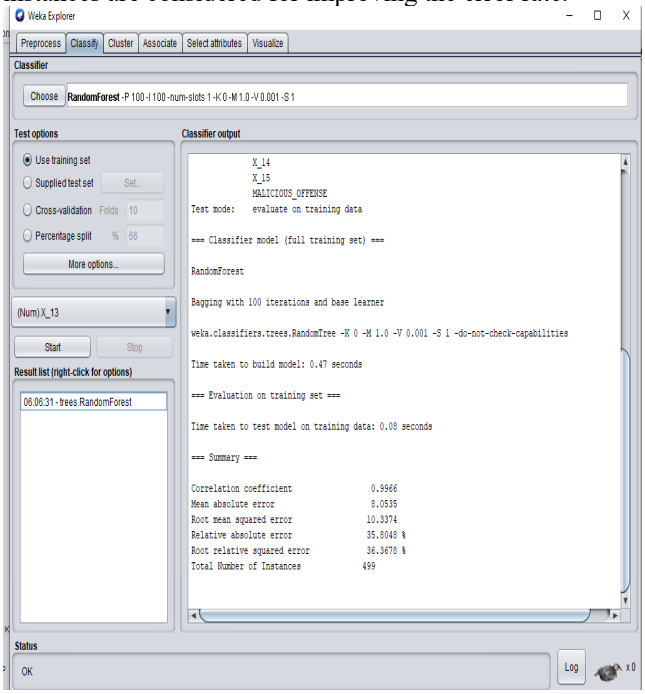

Fig.5: Performance of Random Forest Classifier Predictive Technique.

Figure 5 portrays the Time taken to investigate a model on training data is 0.08 seconds and time taken to build a model 0.47 seconds. Unique Identifier for an incident log, Date of Incident occurrence, Animalized logging parameters with various direction and Malicious offence are the attributes of malicious hacking dataset. These mentioned attributes are considered to discover the performance factors of correlation coefficient, Mean absolute error, Root mean squared error, Relative absolute error and Root relative squared error.

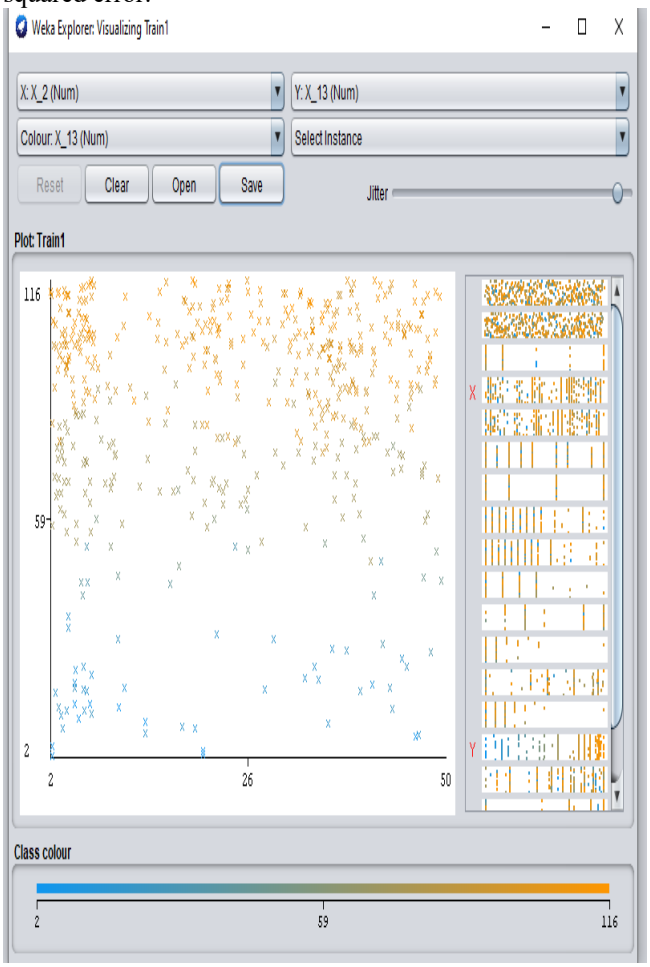

Fig.6: Outcome of Random Forest Classifier.

In the figure 6, the above said classifier results are plotted with the assistance of various animalized logging parameters along with the parameter of the malicious offence. Hence the data objects of malicious hackings are classified with the values of x and y axis respectively.

Figure 7 is a graphical representation of a particular dataset consisting of malicious offences. This dataset contains the incidents and a column specifying whether that site has experienced any malicious offense or not. Every record is specified with its unique incident id. From this data set we can predict the error rate for the upcoming few years.

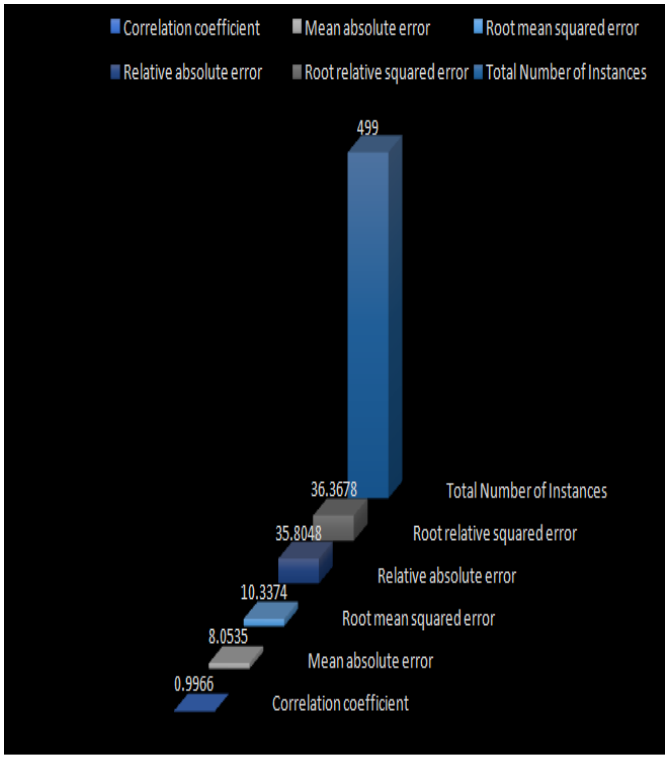

Fig.7: Illustration of Malicious Dataset.

The figure 7 graphically represents the predictive performance factors such as Correlation coefficient, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error and Total Number of Instances. This analysis is done by taking the information about malicious attacks occurring since the year 1991 till the present date. However, with this we can have an approximate estimation about sites, which intent leads to the forecasting of attacks.

Through the Random Forest technique, we can conclude with the following predictive performance factors:

Correlation coefficient	0.9966
Mean absolute error	8.0535
Root mean squared error	10.3374
Relative absolute error	35.8048
Root relative squared error	36.3678
Total Number of Instances	499

According to Akinyelu & Adewumi, (2014), With an accuracy rate of 99.7%, an FP rate of 0.06% and a FN rate of 2.50%, the algorithm significantly outperformed when evaluated on the dataset with the biggest size. This suggests that their technique will indeed be successful when used on real-world datasets, which are frequently huge in size [28]. Similarly, here each record is identified by a distinct event id. We can forecast the error rate over several years using this data set. In the study of Huo et al., (2021), The difference in prediction accuracy between the feature data outside the bag before and after the interference is discovered. The noise of major characteristics would occur in the random forest method, which would result in considerable changes for classification accuracy [30]. As a result, one of the important criteria for choosing characteristics is their relevance.

6 Conclusion

Phishing is one of the common cyber-attacks faced by the majority of people. It has become a mandatory situation to put forward a solution. For that, we have proposed a remedy to conflict in this era when a person is being misled by an attacker through any social engineering, that person is not aware of what he is going to lose. Therefore, we have utilized a predictive method known as Random Forest algorithm, where we can estimate the upcoming chance of cyber-attacks. The main objective of the analysis is to reduce risk. Although, prediction supports an organisation to recognise the security threats before they could perform any damage to our sector. Alternatively, we can stay carefree about the "damaging stage" of a cyber-attack. Our organisation can point out the enterprises can detect expected attacks and improvise prevention of attacks. As a closure statement we have put up a dataset analysis with an improved error rate.

References

[1] European Research Consortium for Informatics and Mathematics. (2012, July). "ERCIM NEWS: Cybercrime and Privacy Issues". Retrieved from http://ercim-news.ercim.eu/images/stories/EN90/EN90-web.pdf

[2] Huda, S., Abawajy, J., Alazab, M., Abdollalihian, M., Islam, R., & Yearwood, J. (2016). Hybrids of support vector machine wrapper and filter based framework for malware detection. Future Generation Computer Systems, 55, 376-390. http://dx.doi.org/10.1016/j.future.2014.06.001

[3] Kaggle. (2015). "Microsoft Malware Classification Challenge (BIG 2015)". Retrieved from https://www.kaggle.com/c/malware-classification

[4] A. Sharma, S. K. Sanjay, "Improving the detection accuracy of unknown malware by partitioning the executables in groups", Proceedings of the 9th international confrence on advanced computing and communication technologies, Nagpur, India, (2015).

[5] P. Mell, K. Kent, J. Nusbaum, "Guide to malware incident prevention and handling", US Department of Commerce, Technology Administration, National Institute of Standards and Technology, (2005).

[6] Govindaraju, "Exhaustive statistical analysis for detection of metamorphic malware", Master's thesis, San Jose State University (2010).

[7] A. Sharma, S. K. Sahay, "Evolution and detection of polymorphic and metamorphic malware: A survey", International Journal of Computer Applications., vol. 90, no. 2, (2014), pp 7-11.

[8] V. Weafer, Editor, "Mcafee labs threats report june 2014", McAfee Labs threat report, (2014).
[9] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood, "Random Forests and Decision Trees", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012
[10] Arnu Pretorius, Surette Bierman and Sarel J. Steel, "A Meta-Analysis of Research in Random Forestsfor Classification", 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech) Stellenbosch, South Africa.
[11] Vrushali Y Kulkarni, Pradeep K Sinha, "Effective Learning and Classification using Random Forest Algorithm", International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, May 2014.
[12] Fayez Tarsha Kurdi, Wijdan Amakhchan and Zahra Gharineiat, "Random Forest Machine Learning Technique for Automatic Vegetation Detection and Modelling in LiDAR Data", International Journal of Environmental Sciences & Natural Resources, Volume 28 Issue 2 - June 2021 DOI: 10.19080/IJESNR.2021.28.556234.
[13] Siji Geroge C G et al., "Genetic Algorithm Based Hybrid Model of Convolutional Neural Network and Random Forest Classifier for Sentiment Classification, Vol 12, No 2, 2021.
[14] Prajwala T R, "A Comparative Study on Decision Tree and Random Forest Using R Tool", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 1, January 2015.
[15] Wash, R. (2020). How experts detect phishing scam emails. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1-28.
[16] Kaur, R., Singh, S., & Kumar, H. (2018). Rise of spam and compromised accounts in online social networks: A state-of-the-art review of different combating approaches. Journal of Network and Computer Applications, 112, 53-88.
[17] He, M., Horng, S. J., Fan, P., Khan, M. K., Run, R. S., Lai, J. L., ... & Sutanto, A. (2011). An efficient phishing webpage detector. Expert systems with applications, 38(10), 12018-12027.
[18] Thomas, J. (2018). Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. Thomas, JE (2018). Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. International Journal of Business Management, 12(3), 1-23.
[19] Aung, E. S., Zan, C. T., & Yamana, H. (2019). A survey of URL-based phishing detection. In DEIM Forum (pp. G2-3).
[20] Hu, H., & Wang, G. (2018). {End-to-End} Measurements of Email Spoofing Attacks. In 27th USENIX Security Symposium (USENIX Security 18) (pp. 1095-1112).
[21] Shen, K., Wang, C., Guo, M., Zheng, X., Lu, C., Liu, B., ... & Yang, M. (2020). Weak links in authentication chains: A large-scale analysis of email sender spoofing attacks. arXiv preprint arXiv:2011.08420.
[22] Hu, H., & Wang, G. (2018). {End-to-End} Measurements of Email Spoofing Attacks. In 27th USENIX Security Symposium (USENIX Security 18) (pp. 1095-1112).
[23] Hu, H., & Wang, G. (2018). Revisiting email spoofing attacks. arXiv preprint arXiv:1801.00853.
[24] Babu, P. R., Bhaskari, D. L., & Satyanarayana, C. H. (2010). A comprehensive analysis of spoofing. International Journal of Advanced Computer Science and Applications, 1(6).
[25] Kambourakis, G., Gil, G. D., & Sanchez, I. (2020). What email servers can tell to Johnny: an empirical study of provider-to-provider email security. IEEE Access, 8, 130066-130081.
[26] Nanaware, T., Mohite, P., & Patil, R. (2019, March). Dmarcbox–corporate email security and analytics using dmarc. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.
[27] Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.
[28] Akinyelu, A. A., & Adewumi, A. O. (2014). Classification of phishing email using random forest machine learning technique. Journal of Applied Mathematics, 2014.
[29] Huo, W., Li, W., Zhang, Z., Sun, C., Zhou, F., & Gong, G. (2021). Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and random forest feature selection. Energy Conversion and Management, 243, 114367.

Abdullah Alhaj received B.Sc. and M.Sc. degree in computer engineering from Lviv polytechnic institute - USSR, in 1988, PhD in Computer Science from Bradford University UK, in 2008. Currently, he is an associate professor in the Information Technology department at The University of Jordan, Aqaba branch. His research interests include computer architecture, networks, IT and network security, performance evaluation of communication networks, machine learning, Big Data and AI. E-mail: aa.alhaj@ju.edu.jo

Mua'ad M. Abu-Faraj received the B.Eng. degree in Computer Engineering from Mu'tah University, Jordan, in 2004, the M.Sc. degree in Computer and Network Engineering from Sheffield Hallam University, UK, in 2005, and the M.Sc. and Ph.D. degrees in Computer Science and Engineering from the University of Connecticut, Storrs, USA, in 2012. He is, at present, an Associate Professor at The University of Jordan, Jordan. His research interests include computer architecture, reconfigurable hardware, image processing, cryptography, and wireless networking. Dr. Abu-Faraj is a member of the IEEE, and JEA (Jordan Engineers Association).

Basel J. Ali earned a Bachelor of Commerce from Aligarh Muslim University, India, in 2001, a Master of Commerce -Accounting from Jai Narain Vays University, India, in 2012, and a Ph.D. in Accounting from University Malaysia Perlis, Malaysia, in 2017. He is currently an Assistant Professor at Applied Science University in Bahrain. His research interests include digital accounting, artificial intelligence in accounting, AIS, and digital accounting.