A · P · U

ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

E-Investigations
Individual Assignment

Thomas MacKinnon TP066728
Intake Code: APDMF2204CYS(PR)
Module Code: CT109-3-M
Dr. Thomas Patrick O'Daniel
Date Assigned: 14/06/2022
Date Completed: 26/07/2022
Word Count: 4768

Contents

List of Figures

1 Introduction

Investment into Cyber Security has seen a massive rise over the last twenty years, as the public and businesses incorporate more technology into everyday life the need for protecting this new frontier as risen accordingly. However, Cyber crime has risen at an increasing rate too, it is only natural for new technology to lead to new methods of exploiting victims. Large organisations build Security Operations Centres (SOC) to counter Cyber attacks, and invest in top of the line defence technology to ensure their assets, reputation, and customers are safe (Casola et al. 2019). As much as Security Specialists invest and build better defense they can never truly patch the one greatest weakness to a whole organisations security, that being the human element (Ezhova et al. 2021).

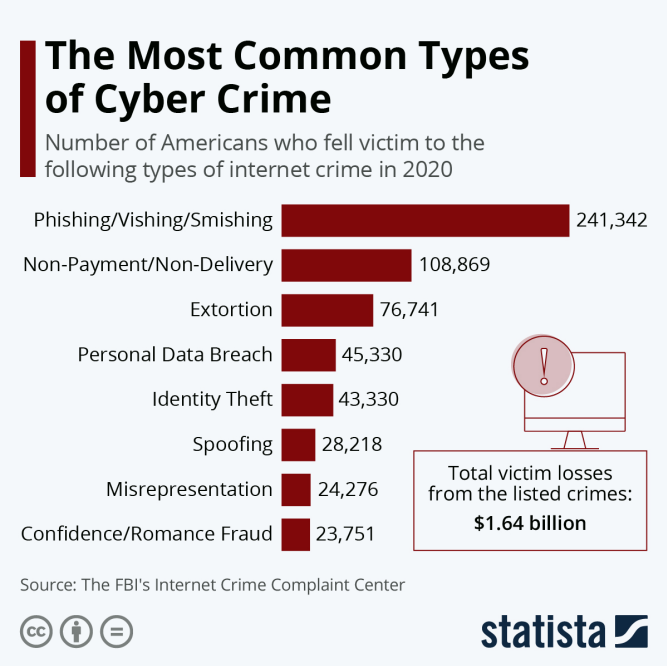

Figure 1: Graph showing Social Engineering to be the most common attack types in Cyber Crime in America throughout 2020 (Ritcher, 2021)

The majority of the population is naturally trusting, and can be manipulated in many ways into revealing sensitive information or giving unauthorised access to malicious hackers. A whole category of attack has been birthed from this hard coded human weakness, aptly named Social Engineering. The Federal Bureau of Investigation's (FBI) Internet Crime Report for 2020 listed Social Engineering as the most common attack type used by Malicious Hackers, as seen in Figure 1. Phishing scams are a subset of the Social Engineering umbrella term (also includes Vishing, Smishing), which involves an attacker sending fake messages to trick the user into revealing sensitive information or unintentionally installing malware onto their device. This is often seen in a email prompt to login into a social media account or a banking inquiry needing urgent attention. A recent notable Phishing Scam occurred after residents of the United Kingdom began receiving text warning that they had been in close proximity with someone infected with the Omicron variant, and that they should order some Rapid tests using the provided link (Vallance, 2022). However, this was not being operated by the British

National Health Service, and was instead a phishing scam aiming to harvest banking card information after asking for payment for Postage and Packaging, as seen in Figure 2. This shows the prevalence of this type of Cyber crime, and opportunism many Malicious Hackers employ when it comes to exploiting victims.

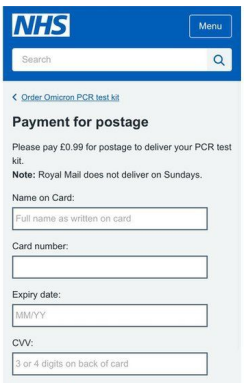

Figure 2: Fake NHS payment request for Covid-19 tests (Vallance, 2022)

Digital Forensics Investigators aim to put these criminals behind bars, and deter future Phishing scams from creating more victims. This is done through a robust methodology aiming to properly investigate a suspect without compromising the integrity of evidence seized, through preservation methods and detailed documentation. This is never easy, especially with Phishing cases, as the amount of evidence can vary greatly, from a single email to months and months of manipulating a victim (Kaspersky, N.D.). Investigators face many legal and ethical issues when conducting their analysis of evidences, especially in regard to handling digital evidence. Emerging technology and techniques like Artificial Intelligence and Machine Learning offer new ways to detect patterns that wouldn't be discernible to the human eye, but what use can they be in a Digital Forensics Investigation?

This report will cover the various challenges Investigators encounter when dealing with a digital evidence in a Phishing case. The aspects of a typical Phishing attack will be discussed as to give a better understanding of what Investigators are tackling, followed by the standard methodology used in an investigation. The chain of custody and various stages of handling Digital evidence will be covered, with a look into the Legal and Ethical Issues present in a Phishing Investigation.

Profiling techniques used for Phishing cases will also be researched and detailed, which is followed by a discussion on what Artificial Intelligence/Machine Learning tools can have on an investigation. The report will conclude with a Proposed Framework for future Phishing cases using emerging technology, with an evaluation on the effectiveness of this new model.

2 Phishing

Christopher Hadnagy describes Social Engineering as "The art and science of skillfully maneuvering human beings to take action in some aspect of their lives that may or may not be in the target's best interest" (Lekati, 2018). Successful Phishing attacks can have incredibly damaging effects on victims and their associated organisations. Figure 3 shows a typical phishing email, from a spoofed email with a urgent call to action, as to manipulate the victim before they have time to verify the request (Impreva, N.D.). Phishing can not only leak sensitive information such as payment card or account details, but also be used as entry point for further attacks like ransomware or spyware. Phishing attacks are often small operations, as it requires little skill to operate.

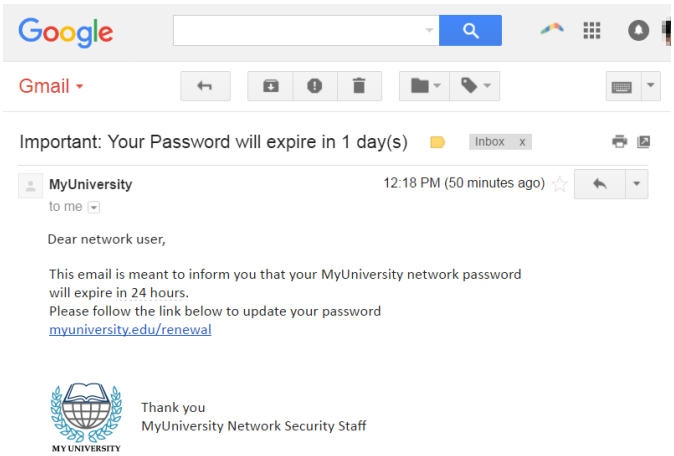

Figure 3: Example Phishing email (Impreva, N.D.)

Phishing attacks can also operate on a larger scale too, as Kwon et al. (2021) details how the Kimsuky hacker group operate social engineering attacks with governmental backing. The group is allegedly funded by the North Korean government, and makes persistent attacks at South Korean targets, like National Security Organisations, the fields of Defense, and Government. This type of group is known as an Advanced Persistent Threat (APT). This disparity leads to great difference in evidence available to investigators, further adding to challenge of getting a conviction. Typical evidence from a Phishing case can be: email/sms/phone logs, digital/physical files, device activity, etc.

3 Handling Evidence in Phishing Cases

3.1 International Guidelines

The High Technology Crime Investigation Association (HTCIA) held the International High-Tech Crime Conference back in 1999, were many of the guidelines Forensic Investigators follow today. The Guidelines were made to be timeless, and are still applicable to modern day cases (Ryan & Shpantzer, 2011). The three core guidelines are as follows:

- No changes are to be made to evidence (digital or physical) after Seizure.
- Only Forensically Competent Investigators are allowed access to the data found on seized evidence.
- All activity relating to Digital and Physical evidence (Seizure, Storage, Transfer) must be documented, and later reviewed. (Takes the form of a "Chain of Custody" in modern Investigations)

These Guidelines are still fit for purpose in current day cases, but does present several legal grey zones. The technique of Live Image Searching is one of these tricky areas in Digital Forensics, as the Investigator is technically altering the state of the machine by searching through the live machine. The digital evidence that can be retrieved from Live Image Searching is very valuable to getting a conviction, especially in Phishing cases. Finding a suspect logged into the Email account that sent Phishing emails is as conclusive evidence can get in a Digital Forensics Investigation. Investigators must prove that their actions on the suspect machine did not change the evidence, therefore maintaining Integrity in handling the digital evidence.

3.2 Chain of Custody

Description of Evidence		
Item #	**Quantity**	**Description of Item** (Model, Serial #, Condition, Marks, Scratches)

Chain of Custody				
Item #	**Date/Time**	**Released by** (Signature & ID#)	**Received by** (Signature & ID#)	**Comments/Location**

Figure 4: Chain of Custody Form (Fortuna, 2018)

A Chain of Custody (COC) typically takes the form of a document which details the journey that evidence takes from seizure at the crime scene to eventual release after the investigation has concluded (Longley, 2022). The evidence has to be accounted for and documented for the entire Investigation

length, such as which Investigators handled it, for how long, and what was done to evidence (checked for fingerprints, imaged memory, etc). Evidence that has not been handled correctly will be rejected from use in Court, and serve as effective rhetoric against the Investigation by the suspects defense (by claiming the team was not competent and therefore breaking the second guideline).

The Chain of Custody is usually a form, as seen in Figure 4, and specifically for Digital Evidence Hash Signatures would be included to verify no changes have been made (Fortuna, 2018). Chain of Custody was implemented to thwart claims that evidence was planted by the Investigation team, and is one of the core legal aspects Investigators have to care for during the case.

3.3 Digital Forensics Methodology

There are many different Methodologies used in Digital Forensics cases, as each methodology is suited to different crime types. Mobile device investigations require different preparations, whilst different law enforcement agencies have different stages. Notably, the FBI takes particular care in avoid unnecessary harm or trauma to innocents and even the suspect by having a stage dedicated to Approach Strategy (Beckett et al. 2009). A standard Digital Forensics Methodology looks something like this:

1. **Authorisation and Preparation -** Involves acquiring the legal permissions to conduct the Investigation and Seize evidence. The court approves a Search Warrant and Affidavit created by the Investigation team, which contains all the evidence that is predicted to be seized from the crime scene. Any evidence not listed cannot be collected, no matter how valuable it is, as it will simply be thrown out of court. Plans are drawn up for how the evidence will be handled, specifically on how digital evidence will be preserved to avoid losing the integrity of evidence. The Chain of Custody will be created at this stage, and usually assigned to one Investigator as their responsibility to simplify it.

2. **Identification -** This stage involves identifying potential areas where hidden evidence could be, that could prove valuable to the case. In Phishing this could be the aforementioned Live Image Searching, email logs, and possibly paper documents (which could have printed evidence of a Phishing attack).

3. **Collection and Preservation -** Care must be taken when collecting evidence from a crime scene, digital evidence is prone to water or heat damage, so the right storage must be used, such as anti static bags. Faraday bags have also become standard in collection, as they block outside signals from reaching devices seized. This is particularly useful when it comes to mobile devices with the ability to be remotely wiped by the owner, potentially losing a goldmine of evidence in a Phishing case. This is also the stage were Live Image Searching would be performed. The Preservation side comes in the form of Imaging digital evidence, so the original will never be changed by the Investigators actions. To further validate this a hash signature is recorded so that after examination on digital evidence Investigators can verify that no changes have been made. Original copies of Digital Evidence is stored safely in evidence lockers away from any potential damages, and any moving or use is recorded in the Chain of Custody.

4. **Examination -** First evidence is filtered so only important digital evidences remains, items like system files and duplicate data is removed so only user created files remain. An interesting technique that Cyber criminals use is hiding incriminating evidence in the slack space between partitions on a hard disc, so it is necessary for investigators to examine these areas.

5. **Analysis -** This stage involves analysing the evidence to be use for court, specifically by creating a narrative of events.

6. **Reconstruction -** Which is further expanded upon in the Reconstruction phase, where a timeline of events is created using Kernel timings, as traditional creation times can easily be edited. This serves as a way to display the relationship between pieces of evidence to jurors, linking everything to the suspect and the victims. In Phishing this would take the form of email time logs compared to the suspects activity.

7. **Reporting -** Reporting is the most crucial aspect of any Investigation, as it involves serving the final document to the courtroom and jury. The report must be well written in a professional manner, without any mistakes (even as small as a spelling mistake), with a glossary of terms and in easy to understand language. There should be no conclusive statements or opinions, simply a narrative of factual information retrieved from evidence with a scale of likelihoods on the suspects involvement.

4 Legal and Ethical Issues for Phishing Cases

The biggest issue facing Investigators in Phishing cases is legal jurisdiction. The majority of Phishing attacks come from other countries to that of the victim, as in the example of Kimsuky, so Investigators have no legal jurisdiction over the Malicious Hackers. In these scenarios local authorities take over but face similar issues. Only International bodies like Interpol can really handle these issues, but often have greater concerns other than small scale phishing scams. The laws regarding Cyber Crime are not universal either, meaning that an illegal online activity in Malaysia might be legal in China, further complicating investigations.

An interesting legal issue comes in the form of the protection granted by the Fifth amendment to American Citizens. The amendment protects people from self-incrimination, usually in the form of not commenting in court or discuss case details with law enforcement. In digital forensics it means suspects do not have to give up passwords or unlock devices for investigators, however, biometrics do not fall under this amendment. It's a standard trick for police to offer a glass of water to suspects that will later be dusted for their prints. There is existing research into creating artificial thumbs to unlock devices, such as the paper by Kovalchik & Maro (2018). The pair developed a gelatin thumb using a 3D printer that was able to unlock a smart phone 60% of the time. This type of unethical unlock has seen some practical application in a case from 2016. Anil Jane, a researcher at Michigan State University, was asked by police to unlock the phone of a dead man to uncover more information about his death. Anil was able to unlock the device using fingerprints from law enforcement (that were recovered from a previous arrest of the victim) and made an artificial thumb print using conductive paper (Lumb, 2016).

This application of artificial fingerprints to unlock suspect devices is obviously ethically questionable but not illegal in the United State (biometrics are protected in other countries). Of course there is many other legal and ethical issues that investigators must avoid, mainly in handling evidence correctly as covered.

5 Profiling Techniques

The Federal Bureau of Investigation (FBI) defines Profiling as a " technique used to identify the perpetrator of a violent crime by identifying the personality and behavioral characteristics of the offender based upon an analysis of the crime committed" (FBI, 1986). Profiling has been used for decades in criminal investigations, aiding greatly in narrowing down the list of suspects. Interestingly, the first real profiling methodology used was the "Malleus Maleficarum" used for identifying suspected Witches during the Salem Witch Trials (Balogun & Zuva, 2018). Profiling uses extensive knowledge of the law, sociology, criminology, and psychology to build up a criminal profile, and around 77% of cases that use profiling have successful outcomes (Angelopoulo & Slide, 2014).

There are two types of Profilers being:

- **Inductive -** Uses statistical information about the crime, through previouse cases and convicted criminals to build a profile. Excellent for unsolved cases or where limited evidence is available.
- **Deductive -** Forms a conclusion from evidence and the crime scene to build a criminal profile.

Godwin (2012) describes how inductive tends to be better for real world crimes, as there is far more historical data to develop a criminal profile. However, Angelopoulou & Slide (2014) state that for digital forensics cases deductive profiling is often better, as there is far more evidence available to analyse, and a smaller array of potential suspects.

A commonly used Profiling methodology is "Behavioural Evidence Analysis (BEA)", which is a deductive Profiler fit for a phishing case. The model is composed of four stages, being:

- **Equivocal Analysis -** Reviews all sources of evidence, whilst removing any irrelevant evidence and highlighting particularly important pieces. This could be removing system files so only user created files can be analysed for hints of phishing.
- **Crime Scene Characteristics -** What methods were used to hide the crime? how was the crime committed? In phishing it could be analysis on how incriminating evidence was deleted or hidden in partitions, or what machines were used.
- **Victimology -** How did the victim behave before/during/after the attack? What does the choice of Victim say about the attacker? is the victim related to the attacker in any way? What role did the victim play? These questions help build up a profile on the victim to better understand the attackers motives. This is particularly important in phishing cases as the attacker often has a lot of communication with the victim.
- **Offender Characteristics -** What was the attackers goal? What was the activity of the attacker? How did the attacker act (demeanor, personality)? Is any revealing detail known about the attacker (such as hobbies or off comments revealing about their life)? As there is much communication between victim and attacker it is likely some personal information slipped out during manipulation. This can all come together to build a profile on the attacker.

The FBI present a different form of Profiler, being a hybrid of Inductive and Deductive methods to produce the best result. The methodology involves much of the same steps as BEA but includes data from existing cases to build a criminal profile (Angelopoulou & Slide 2014). This seems like an obvious improvement to aid in securing a conviction.

Profiling is not only used by Law enforcement to find criminals, but also by attackers to find suitable victims. Malicious hackers target individuals with little to no technical skill, as to avoid wasting the attackers time, which is why phishing emails often have glaring mistakes, as the target audience will simply ignore or not notice the mistakes (Lekati, 2018). Conviction in Cyber Crime tends to be low, being less than 9% in the United States, due to obsolete laws, jurisdictional problems, long investigation lengths, and lack of skilled investigators (Balogun & Zuva, 2014). The repetitive nature of phishing scams is just begging to be used in profiling, but the large volumes of data can be challenging for investigators to sift through to build up a criminal profile.

6 Machine Learning in Phishing cases

Machine Learning, as defined by Ellrodt et al. (2018), is the process of recognizing patterns in large volumes of data in order to perform a specific task. Machine Learning has been around since the early 1950s, originally titled by Arthur Samuel when developing a self learning checkers program (Foote, 2019). Artificial Intelligence is different from machine learning, as it focuses on simulating human behaviour (like chat bots) rather than learning from data to give accurate outputs. There is a vast array of machine learning algorithms being used in everyday life, Netflix constantly refines their recommended content depending on a users watching habits to keep customers satisfied. Machine learning allows for solutions to difficult problems whilst continuously improving, saving time and resources.

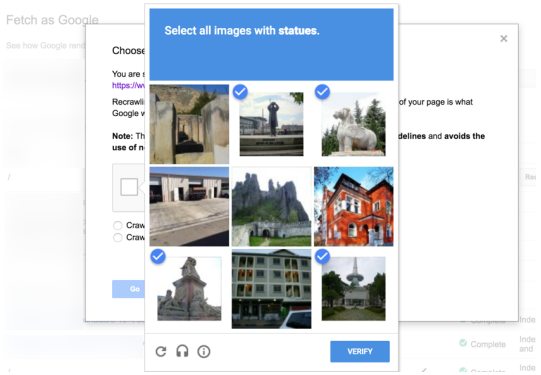

Figure 5: reCaptcha acquiring Training data of statues (o'Malley, 2018)

Machine learning works through feeding training data for the algorithm to find underlying patterns and classify the data. However, the type of data being fed has a great impact on the algorithm. There are three main types of Machine Learning, being:

- **Supervised Learning -** data in supervised learning is labeled, and often is small in size, but will be better understood by the machine learning solution. The data fed in training is very similar to the data given once deployed. A good example of a Supervised Machine Learing algorithm is Google's reCaptcha, as users label images to authenticate they are not a bot. This data is then fed back into the algorithm to become better at distinguishing similar items, like statues from buildings as seen in Figure 5, and improve their users experience of using their products(o'Malley, 2018).
- **Unsupervised Learning -** this type of algorithm uses unlabeled data for training, which saves a great deal of time in making data more machine readable, and allows for greater volume of training data. Unsupervised solutions also can find hidden structures to data that wouldn't be found otherwise. Clustering often uses Unsupervised Machine Learning to take unorganised data and cluster them into specific groups (Potentia, 2021).
- **Reinforcement Learning -** is slightly different to the previous two, as it has a defined end goal that the algorithm has to achieve. Reinforced Learning uses trial and error to find the most effective solution to a task, praising positive outcomes and discouraging less effective solutions. This is notably used in shortest path finding algorithms like Google Maps to present the best option between to destinations.

There is already use of Machine Learning within Digital Forensics, Qadir & Varol (2020) discuss a classifier algorithm that detects skin in images, which can greatly aid Investigators in eliminating irrelevant evidence in cases that deal with illegal explicit images. The researchers state that the digital revolution has caused a huge increase in potential evidence for cases, and that machine learning is the best solution to filtering out useless data. As mentioned there is a shortage in Digital Forensics talent, and cases take a very long time to come to completion, utilising Machine Learning greatly aids an investigation.

6.1 Support Vector Machines (SVMs)

SVM is a binary classifier, that uses supervised learning, which works by creating two categories for data to fall into, and then divide the data using a hyperplane. Data is mapped to coordinates as seen in Figure 6, with a hyperplane being as far away from near data points as possible as to avoid data being misidentified into the wrong category (Ray, 2017). In the case of phishing SVM machine learning could be used to categories data into "Normal" and "Suspicious" to save Investigators time.

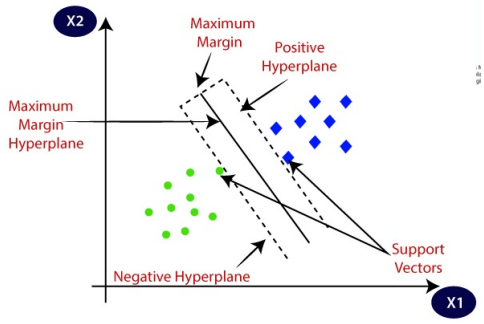

Figure 6: SVM positive/negative hyperplane margins

SVM is not a perfect solution, as it does struggle with larger datasets, since it is harder to draw a hyperplane with more data points and more outliers. SVM also struggles with overlapping classes for data, which is an issue for something as vague as likelihood of phishing email. As phishing cases tend to have a large amount of data this algorithm is probably not suited to the job (Ali et al. 2021).

6.2 Decision Trees:

Decision Trees also use supervised data, and is a binary classifier, however, it does operate differently from SVM. Decision Trees use binary recursive partitioning to filter data on previously set parameters until is placed into one of the two categories. Figure 7 (Bakos, 2010) shows a decision tree for the likely hood that a partner will cheat during a relationship, taking several factors from the labeled data and finding the underlying pattern. Decision Trees are visually appealing and easy to understand, and the amount of filtering increases as more data is fed, meaning the machine learning algorithm is more likely to spot suspicious emails overtime.

There is an issue of "overfitting" with Decision Trees, where too many hypothese are made about the training data that wouldn't apply to real data. The algorithm becomes too good at categorising the training data, which could have quirks not seen in most evidence. Pruning is the process of removing unnecessary filtering rules to make the tree simpler, so this problem is avoided. Underfitting can also be an issue if too much pruning occurs.

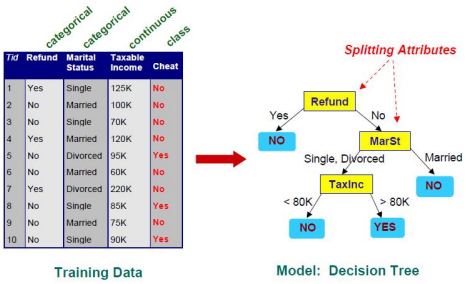

Figure 7: Decision Tree Example

6.3 k-Means Clustering:

This Machine learning algorithm uses unsupervised learning to cluster data into different groups. K-Means Clustering (k-MC) saves investigators a lot of time as they can simply dump data onto the algorithm to teach it. Figure 8 shows k-MC clustering the data points into three distinct groups, meaning that $k=3$ (Jain, 2021). This provides an advantage over the other two solutions presented, as emails can be categorised into risk likelihood rather than a binary option of "normal" and "suspicious", which help minimize false negatives in evidence that might be missed otherwise. k-MC scales excellently with large datasets making it ideal for phishing cases, and can display data in graphs, which would be useful in a courtroom.

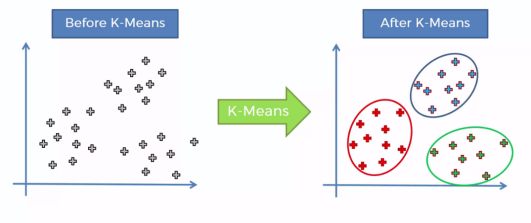

Figure 8: k-Means Clustering on random data points

Oversized clusters is an issue for k-Mean clustering as it assumes all clusters will be around the same size. As the majority of digital evidence will be irrelevant this would cause an issue with k-MC. The algorithm also struggles with overlapping clusters from outliers, however, in the case of phishing this is not a big (Google Developer, 2021).

7 Proposed Framework

Clearly Digital Forensics needs an upgrade to deal with the ever increasing rate of Cyber Crime, and to help raise the low conviction rate. Therefore a new framework has been designed and proposed, using modern machine learning and profiling to better aid Investigators in putting criminals behind bars, faster and easier than ever before.

7.1 The Phishing Framework

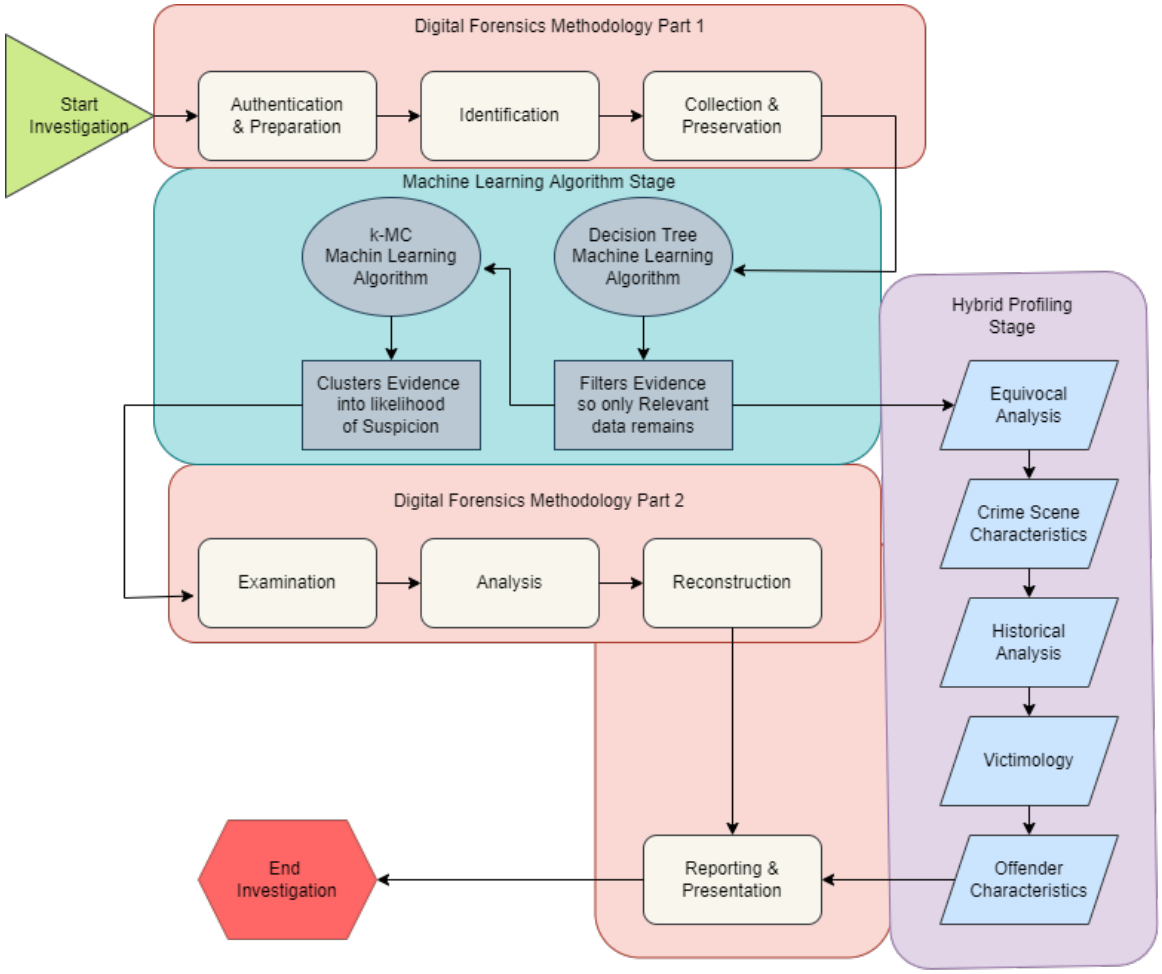

Figure 9: Phishing Framework for Digital Forensics cases

The framework proposed takes the original Digital Forensics Methodology and improves it by adding in a Profiling and Machine learning stage. The process begins with the standard steps, acquiring legal permissions in "Authentication and Preparations", followed by "Identification" and the "Collection & Preservation" of evidence from the crime scene.

Once evidence is collected it can be filtered down with a Decision Tree machine learning algorithm,

removing all irrelevant files, such as system files so only user created files remain. The relevant data can then be used in a k-Means Cluster algorithm to categories the digital evidence into four categories, being: Normal, Unusual, Suspicious, Dangerous. The data is now much easier for Investigators to examine, saving a great deal of time and hopefully going to court a lot sooner than without these Machine Learning Algorithms.

The Relevant data separated by the Decision Tree can be used in the Profiling section in conjunction with the other stages of this framework. The usual stages apply for profiling as seen in the "Behavioural Evidence Analysis" but with the added "Historical Evidence" stage taken from the FBI profiler. As it is using only relevant data the Profiler will be far more efficient and save time in building a criminal profile.

The second half of the Digital Forensics methodology is then conducted, with the Profiling results joining the rest of the flowchart in the Reporting stage. This concludes the framework.

7.2 Evaluation of Framework

The choices made for the proposed framework were reached from research presented in this paper and existing methodologies from academic/governmental sources. However, further discussion is needed on some choices in the design.

The choice of Machine Learning algorithm is very important, as the amount of data being analysed is large and complex. The choice of a decision tree was important for separating irrelevant files from actual useful evidence, the filtering system present in this algorithm would serve perfectly at this job. System files would be easy to add to the trees filter and create an excellent classifier that would present only the user created files. There would be little need for pruning as the difference between categories would be good enough. Furthermore, unlike SVM algorithms, Decision Trees work well with larger amounts of data, making it superior for this initial stage of evidence classifying.

This leads nicely to k-Means Clustering, which as mentioned, cannot deal well with large differences between cluster sizes. Since the evidence has been greatly filtered the k-MC algorithm should have no problem clustering the remaining data into risk likelihood. These two Machine Learning algorithms used together provide Investigators with an excellent solution to reducing and organising the amount of evidence, which would save a lot of time and resources otherwise.

The Profiling section is also important, as Angelopoulou & Slide (2014) mention that cases that use profiling are successful 77% of the time. The BEA profiler is excellent and was recommended by several authors for Digital Forensics cases, so was used as the backbone for the Profiling section. However, a small amendment was made from the FBI's influence, adding a "Historical Analysis" to use data from existing cases to build a criminal profile. This means the Profiler type is Hybrid rather than Inductive or Deductive. This decision was made as historical data could add a lot to the profiler and makes it future proof for when more successful phishing cases have concluded.

7.3 Evaluation of Machine Learning Algorithms

To guarantee that the Machine Learning Algorithm is performing well and making accurate predictions an evaluation metric must be put in place. This assessed the accuracy and effectiveness of the model, visualising results to the Investigative team and later to the courtroom.

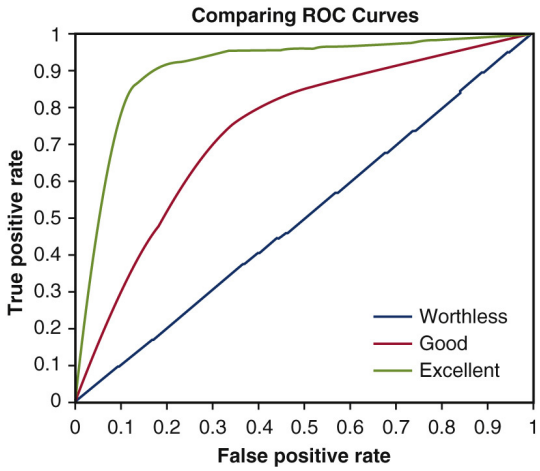

Figure 10: ROC curve comparison

Receiver Operating Characteristic Curves (ROC) provides an excellent evaluation metric for binary classifiers, like SVM and Decision Trees. ROC curves maps out the True Positive rate against the False Positive rate, as seen in figure 10, allowing the user to monitor the trade-off between sensitivity (true positive rate) and specificity (false positive rate) of results.

Confusion Matrix is a table used to show the performance of classification algorithms, this is done through test data were the true values are already known. The table shows areas of weakness for the algorithm, by displaying which results were predicted correctly or incorrectly through percentages in a matrix, as seen in figure 11. This evaluation metric is very visually informative and can be a great aid in finding under performing areas of the algorithm.

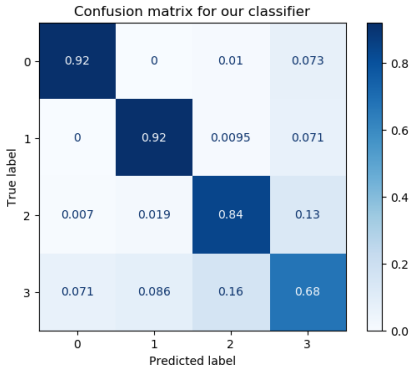

Figure 11: Confusion Matrix Example

8 Conclusion

In Conclusion, this report presents the current impact Phishing has on the digital world, and the challenges Investigators face when dealing with this type of case. Investigators have to take particular care in avoiding legal issues and ensuring that digital evidence is handled in the correct manner to avoid a mistrial. The use of Profiling methods is explained thoroughly, with their benefit to Phishing cases very clear. This is further added to with the potential of Machine Learning algorithms to cases, such a k-Mean Clustering to categorise evidence. The proposed framework seeks to upgrade the existing digital forensics massively through the use of Machine Learning and Profiling, with all choices made backed up in reviewed literature. The solution presented would greatly speed up investigations and help get more Cyber criminals convicted, thus preventing future victims.

9 References

Ali, M., Gadekallu, T.R., Hina, M., Jalil, Z., Javed, A.R., Srivastava, G. 2021. Email Classification and Forensics Analysis using Machine Learning. *2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI).* pp. 630-635.

Beckett, J., Lin, Y., Slay, J., Turnbull, B.P. 2009. Towards a Formalization of Digital Forensics. *Advances in Digital Forensics V - Fifth IFIP WG 11.9 International Conference on Digital Forensics, Orlando, Florida, USA, January 26-28, 2009, Revised Selected Papers.* pp. 37-47.

Bakos, Y.J. 2010. Decision Tree Classifier. [online] Colorado School of Mines. Available at: http:
//mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/lguo/decisionTree.html [Accessed 26/07/2022]

Casola, V., Catelli, R., Debertol, D., Meda, E., Mokalled, H., Zunino, R. 2019. The applicability of a SIEM solution: Requirements and Evaluation. *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE).* Pp.132-137.

Ellrodt, L.R., Fields, T.L., Freeman, I.C., Haigler, A.J. & Schmeelk, S.E. 2018. What are they Researching? Examining Industry-Based Doctoral Dissertation Research through the Lens of Machine Learning. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA).* pp. 1338-1340.

Ezhova, A.A., Kotelyanets, O.S., Leonov, P.Y., Morozov, N.V., Vorobyev, A.V., Zavalishina, A.K. 2021. The Main Social Engineering Techniques Aimed at Hacking Information Systems. *2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT).* pp. 471-473.

FBI. 2020. Internet Crime Report 2020. Pennsylvania, United States of America. Available at: https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf [Accessed 24/07/2021]

FBI. 1986. Investigative Tool Against Violent Crime. *FBI Law Enforcement Bulletin Volume: 55 Issue: 12.* pp. 9-13.

Foote, K.D. 2019. A Brief History of Machine Learning. [online] Dataversity. Available at: https://www.dataversity.net/a-brief-history-of-machine-learning/# [Accessed 26/07/2022]

Fortuna, A. 2018. Digital Forensic: the Chain of Custody. [online] Blog. Available at: https://andreafortuna.org/2018/04/09/digital-forensic-the-chain-of-custody/

Godwin, M. 2012. Brief Discussion on Inductive/Deductive Profiling. [online] Blog. Available at: http://www.drmauricegodwin.com/inductiveprofiling.html [Accessed 24/07/2022]

Google Developer. 2021. k-Means Advantages and Disadvantages. [online] Google Developer.

Availabe at:
https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages [Accessed 26/07/2022]

Imperva. N.D. Phishing attacks. [online] Imperva. Available at: https://www.imperva.com/learn/application-security/phishing-attack-scam/ [Accessed 24/07/2022]

Jain, T. 2021. K-Means Clustering. [online] Medium Data Driven Investor. Available at: https://medium.datadriveninvestor.com/k-means-clustering-ac3ff1d3463d [Accessed 26/07/2022]

Kovalchik, M. & Maro, E. 2018. Bypass Biometric Lock Systems With Gelatin Artificial Fingerprint. *Proceedings of the 11th International Conference on Security of Information and Networks (SIN '18)*. pp. 23-24.

Longley, R. 2022. What Is Chain of Custody? Definition and Examples. [online] ThoughtCo. Available at: https://www.thoughtco.com/chain-of-custody-4589132 [Accessed 24/07/2022]

Lumb, D. 2016. Police get dead man's finger 3D-printed to unlock his phone. [online] Engadget. Available at: https://www.engadget.com/2016-07-21-police-get-dead-man-s-finger-3d-printed-to-unlock-his-phone.html [Accessed 24/07/2022]

O'Malley, J. 2018. Captcha if you can: how you've been training AI for years without realising it. [online] techradar. Available at: https://www.techradar.com/uk/news/captcha-if-you-can-how-youve-been-training-ai-for-years-without-realising-it [Accessed 24/07/2022]

Potentia. 2021. What Is Machine Learning: Definition, Types, Applications And Examples. [online] Potentia Analytics. Available at: https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/ [Accessed 26/07/2022]

Qadir, A.M. & Varol, A. 2020. The Role of Machine Learning in Digital Forensics. *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*. pp. 1-5.

Ray, S. 2017. Understanding Support Vector Machine(SVM) algorithm from examples (along with code). [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/ [Accessed 26/07/2022]

Ritcher, F. 2021. The Most Common Types of Cyber Crime. [online] Statista. Available at: https://www.statista.com/chart/24593/most-common-types-of-cyber-crime/ [Accessed 24/07/2022]

Vallance. 2022. NHS warns of scam Covid-test texts. [online] BBC News. Available at: https://www.bbc.com/news/technology-61882239 [Accessed 24/07/2022]