Prediction of Cyber Attacks During Coronavirus Pandemic by Classification Techniques and Open Source Intelligence

Shannon Wass, Sina Pournouri, and Gregg Ibbotson

Abstract Over the years, technology has grown rapidly and become a major part of everyday life. Due to the increased presence of technology, cybercrime is on the rise and the number of cyber-attacks has increased significantly, this has made data mining techniques an important factor in detecting security threats. This research proposes that Classification techniques can be used to reliably classify and predict cyber-attacks. This paper proposes a classification framework using data collected from Hackmagedon, a blog which contains timelines and statistics for cyber-attacks. The dataset includes cyber-attacks which occurred between 2017 and 2019 within countries in Europe. The purpose of this research is to investigate how Classification techniques can be used to better understand and predict future cyber-attacks. Different Classification techniques will be applied to the dataset to determine which technique produces the most accurate results. The model will be validated using a dataset containing COVID-19 cyber-attacks from Hackmagedon.

Keywords Cyber attack · Prediction · OSI · Pandemic · COVID-19

1 Introduction

Over the years, cyber-attacks have become an increasingly prevalent part of today's society. This is due partly to the growth in technology and it's use and dependence in everyday life. Due to the increased presence of technology, cybercrime is on the rise and the number of cyber-attacks has increased significantly.

S. Wass · S. Pournouri (✉) · G. Ibbotson
Sheffield Hallam University, Sheffield, UK
e-mail: s.pournouri@shu.ac.uk

S. Wass
e-mail: shannonwass@hotmail.co.uk

G. Ibbotson
e-mail: g.ibbotson@shu.ac.uk

H. Jahankhani et al. (eds.), *Cybersecurity, Privacy and Freedom Protection in the Connected World*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-68534-8_6

In 2019, a third of businesses (32%) reported having a cyber security breach or attack. In winter 2019 and early 2020, almost half of businesses (46%) reported having a cyber breach or attack [8]. This demonstrates cyber-attacks are evolving and becoming more frequent.

Since the COVID 19 pandemic began, cyber attackers have taken advantage of the uncertainty and anxiety of the general populace as an opportunity for financial gain. In March and April alone, brute-force attacks have risen by 400% [12]. It is likely this is due to the increase in remote desktop connections, as employees have had to push technology out at an unprecedented rate, mistakes have been made, and systems are left unsecure. Under half of businesses have experienced at least one business impacting cyber attack related to COVID-19 in April 2020, such as: loss of customer, employee and confidential data, ransomware pay-outs and financial loss and theft.

In August 2020, it was discovered Blackbaud, a software provider, fell victim to a ransomware attack in May. The company is used by numerous other universities, such as Birmingham, Leeds, York, and Reading, who also suffered a data breach. The information which was breached included names and contact details for donors, alumni and stakeholders. Blackbaud paid the ransom and the data was destroyed [16].

Alongside this, in May, two companies involved in building coronavirus hospitals were hit by cyber-attacks: Bam Construct and Interserve. Bam Construct shut down its website and other systems as a precaution after being hit by a computer virus, whereas Interserve stated some operational services were affected [1].

The two attacks mentioned above were both on different work sectors: education and healthcare. This indicates cyber attackers do not discriminate on the sectors, they merely target the sector where the greatest amount of disruption and financial gain can be made.

The top sectors most likely to be affected by a cyber-attack during COVID-19 are healthcare and banking. Within the healthcare sector, malicious actors are exploiting the situation to deliver malware payloads at medical facilities to compromise the infrastructure and demand ransom to restore functionality. Due to the desperation and necessity of the compromised equipment, healthcare sectors are likely to pay the ransom as it would be more costly to not pay the ransom. In the banking sector, cyber criminals are taking advantage of the COVID-19 pandemic by creating phishing campaigns using COVID-19 misinformation to steal user data [18]. The increase in cyber-attacks and the elaborate methods used to conduct them demonstrate the need for more advanced cyber-attack detection methods. Currently, methods such as Intrusion Detection Systems are used to detect anomalies, however, false positives are frequent and IP packets can be faked. Therefore, this study aims to develop a framework using classification techniques alongside historical data to better understand and predict cyber-attacks in Covid-19 era.

1.1 Background

Many researches have been done in the field of cyber security and data mining specifically classification techniques. In this section we have reviewed few most recent studies about how classification techniques can be used for prediction of different elements in cyber security.

1.1.1 Support Machine Vector

Sarker et al. [21] proposed an Intrusion Detection model using SVM. This experiment used security datasets which represented a collection of information records consisting of security features which could be used to train the intrusion detection model. The dataset contained more than 25,000 records which were collected from a variety of intrusion detection systems simulated in a military network environment. To collect this data, an environment was created by simulating a US Air Force Local Area Network. The dataset consisted of both nominal and numerical security features:

1. Duration logged in
2. Number of failed logins
3. Server error rate
4. Protocol type.

This experiment ranked the security features and assigned them a score, this was so significant features could be chosen for further processing and irrelevant features could be removed. This made the model more effective in terms of prediction accuracy for unseen test cases. The effectiveness of the model was calculated in terms of accuracy, true positive, and false negative. To calculate the results for unseen test cases, 80% of the data was used for training and 20% for testing. The results were calculated using a matrix which reported the number of false positives, false negatives, true positives and true negatives. However, the limitations of this study were the features in the security dataset were not equally significant to build a data driven security model.

Kim Donghoon et al. [9] proposed a SVM framework to detect a class of cyber-attacks which redistribute loads, such as DDOS attacks. The dataset consisted of captured network packets, 70,000 were used for the training set and 30,000 were used for the validation set. The limitations of this experiment were the DDOS attack methods were not complicated enough compared to real world attacks, meaning the results may be inaccurate. To reliably evaluate the performance, raw data from client machines affected by a DDOS attack is needed.

Kinan Ghanem et al. [13] proposed a Network Intrusion Detection model using SVM to reduce the number of instances used during the computation of the SVM which also reduces the training time of the model. The model was evaluated using different network traffic datasets from wired and wireless networks at Loughborough University. The SVM categorised the traffic as malicious or normal, if it was

malicious, it was placed into one of the four classes: Denial of Service, Remote to Local, User to Root and Probing.

A positive point to Kinan Ghanem et al. (2017)'s approach was the classifier had an accuracy of 100% for detection rate. However, the problem with this technique is the SVM required labelled training data to accurately predict the class of the data. The limitation of this study is the SVM only reached an accuracy of 81% and generated false positives of 19% for the Probing class, this could be a result of the data consisting of different factors; for example, wired and wireless connections, the different attack types and a possible imbalance in the dataset.

1.1.2 Random Forest

A Random Forest classifier was proposed by Nabila Farnaaz [11] to predict malicious intrusions using a host-based intrusion detection system. The data used for this experiment was a NSL-KDD data set, this contained records of internet traffic detected by a simple intrusion detection network, the data contained 43 features per record, 41 of these features refer to the traffic input and 2 are the labels of whether it is normal or an attack and the score given to the severity of the traffic input. Within the dataset, there were 4 different classes of attacks: Denial of Service, Probe, User to Root and Remote to Local. Their proposed approach and the method was split into 8 steps:

1. Load the dataset—the data was imported into WEKA.
2. Apply pre-processing technique—WEKA was used to replace the missing values.
3. Cluster the data into datasets—the data was clustered into groups based on their similarities.
4. Partition the data into training and test sets—the data was split into different sets.
5. Select the best set using feature selection—the best set was used based on their features e.g. type of attack.
6. Dataset was applied to the Random Forest classifier for training—the data was trained using RF.
7. Test data was given to the Random Forest classifier for classification—the data was classified using RF.
8. Calculate accuracy, detection and false alarm rate.

The proposed model yielded a high detection rate and a low false alarm rate with an overall accuracy of 99.97%

A positive point of Nabilia Farnaaz (2016)'s approach is multiple performance measures were used to evaluate the classifier: accuracy, detection rate, false positive and true negative. For example, if the classifier were only evaluated using the detection rate and everything was detected successfully, the classifier would have a success rate of 100%. However, this would not take into account the accuracy and

amount of false positives vs true negatives, meaning the results of the experiment would be inaccurate.

The problem with Random Forest is many noisy trees are created which affect the accuracy and decisions for future samples. However, Nabila Farnaaz (2016) takes this into account in the feature selection process, having irrelevant features decreases the accuracy of the classifier. The features chosen were:

1. Accuracy—the ratio of correctly classified samples.
2. Detection rate—the ratio between total numbers of attacks detected by the system to the total number of attacks present in the dataset.
3. False alarm rate—this is defined as: FP/TN + FP.

An additional approach was suggested by Leyla Bilge et al. [3], a system called RiskTeller. This system used a Random Forest classifier to predict which machine was at higher risk of a cyber-attack. The dataset used for this experiment was binary file logs from 600,000 machines belonging to 18 enterprises. For each machine, 89 features were used. These features were based on the number of events, application categories, rarity of files, patching behaviour and past threat history. The features were categorised into 6 groups:

1. Volume based—percentage of events, fraction of events from top file hashes, percentage of applications.
2. Temporal—Monthly percentage of events—median/standard deviation.
3. Vulnerability/patching—percentage of patched vulnerabilities/applications.
4. Application categories—top 5 application categories with most events.
5. Infection history—fraction of events for malicious/benign/unknown files.
6. Prevalence-based—fraction of events seen in only one enterprise.

This method consisted of:

1. Data pre-processing—the file and directory names were normalized to identify those which were likely to perform the same.
2. Feature extraction—features were computed into the classifier.
3. Feature labelling—features were labelled as clean or infected machines.
4. Feature discovery—for each machine, a profile was created containing 89 different features based on events which will be used to predict the machines risk of future infection.

The problem with this method was whereas RiskTeller predicts the general risk that a machine is at risk from malware, it did not classify specific malware categories. However, a positive point is Leyla Bilge et al. (2017)'s experiment proves the concept of using machine learning to predict malware with a high accuracy rate. Risk Teller achieved a 96% true positive rate and 4% false positive rate, no previous work has been able to achieve a 96% true positive rate with a 4% false positive rate at a machine level granularity.

Rupa Ch et al. [5] proposed a computational system to detect and classify cyber-crimes using Random Forest. The data was gathered from Kaggle, a data science community and CERT-IN, an Indian Computer Emergency Response Team. The

information collected from each data source consisted of 2097 records with 8 features: the type of incident, victim, access violation, harm, year, location and age of offender. After the data was pre-processed, features were extracted and used to classify the cybercrimes: Incident, Offender, Harm, Access Violation, Year and Victim.

A positive point on Rupa Ch et al. (2020)'s approach was the proposed model classified with a 99% accuracy rate. The limitations of the study was currently it only classified cybercrimes into certain groups: Identity theft, hacking, copyright attacks and other. A feature extension is required to provide countermeasures to the crime agencies, in order to reduce the frequency of cybercrimes in specific locations.

1.1.3 Naïve Bayes

Prajakta Yerpude [20] proposed a framework for predicting crime using supervised machine learning methods, such as Naïve Bayes. The communities and crime dataset from the UCI repository (a collection of databases) was used, which consisted of crime data in Chicago, containing a total of 1994 records. Included in the dataset were features such as:

1. Population—urban, rural etc.
2. Race—asian, Caucasian etc.
3. Sex—female/male
4. Police—percentage of police officers.

To select the most important features for the classifier, Feature Importance was used to assign each feature a score. The higher the score, the more significant the feature was to the classifier. The features extracted according to their feature importance scores were:

1. NumUnderPov: Number of people under the poverty line.
2. NumbUrban: Number of people in Urban Areas.
3. HousVacant: Number of vacant houses.
4. RacePctHisp: Percentage of race Hispanic.
5. LemasPctOfficDrugUn: Percentage of officers assigned to drug unit.
6. PctNotSpeakEnglWell: Percentage of people who did not speak English well.
7. acePctAsian: Percentage of race Asian.

The classifier used both clean and dirty data. The accuracy of the result for the clean data, 77.64%, was higher than the dirty data, 75.42%. This demonstrates that missing data creates inconsistencies and affects the performance of the model. A positive point of Prajakta Yerpude's approach is the feature importance score proved to be highly predictive, specifically "NumUnderPov" and "NumbUrban". The limitation of this study is the dataset consisted of every crime,this analysis can be narrowed down to categories of crime to yield more accurate results.

1.1.4 K Nearest Neighbour

Ben Abdel Ouahab Ikram et al. [2] proposed a malware classification framework using K-Nearest Neighbor alongside visualisation techniques. Using visualisation techniques, the malware binary was converted into a grayscale image, afterwards, an image descriptor was computed to classify the malware using K-Nearest Neighbour with features extracted from the data. The dataset used for this framework was Malimg, this is a malware classification dataset from the website Kaggle, containing 9339 malware samples from 25 different families in the form of grayscale images. This dataset was split into training data and test data: TestDB and TrainDB. The features exported from the dataset were: malware families names, number of images in each family.

The model was trained with different k values and a score was calculated for each case, for example: $k = 2$, $k = 5$, $k = 10$. Since $k = 10$ gave the highest training score, this was saved for future predictions. The test score yielded an accuracy of 97.92%. The model was saved and used on the TestDB with completely new, unlabelled data. The purpose of this was to see how the model would perform on unknown samples. The model yielded an accuracy of 92% with the unknown samples. Overall, the model accurately classified 46 out of 50 malware samples. The limitation of this study was the data was not cleaned before being imported into the model for training, this means predictions from the data may be inaccurate and misleading.

Masoumeh Zareapoor [25] proposed a model for predicting credit card fraud using KNN. A credit dataset was used containing e-commerce transactions with 100,000 records labelled as legitimate or fraudulent, 2293 of the records (2.8%) were fraudulent and 97,707 (97.2%) were legitimate. The model was evaluated using the following metrics: Fraud Catching Rate, False Alarm Rate and Balanced Classification Rate. Incoming transactions were classified by calculating the nearest point to the newest incoming transaction. If the nearest neighbour was fraudulent, the transaction was classified as fraudulent.

A positive point of Masoumeh Zareapoor (2015)'s approach was the possible bias in the unbalanced data was taken into account, the evaluation metrics were changed from accuracy and error rate to False Alarm Rate and Balanced Classification rate.

1.1.5 Neural Networks

Moshe Kravchik et al. [15] proposed a framework for detecting cyberattacks in industrial control systems using convolutional Neural Networks. The dataset used was a Secure Water Treatment testbed which represented a real-world industrial water plant and included 36 different cyber-attacks, consisting of 7 days of recording under normal conditions (benign) and 4 days when the 36 attacks were performed (attack) and logged on a server. The entire dataset contained 946,722 records labelled as attack or benign. The features used in this experiment were attributes from the water test sensors: flow meters, water level meters, conductivity and pH analysers. The data from the sensors was logged and used for training and testing the model. A

positive point of Moshe Kravchik et al. (2018)'s approach is the model successfully detected 32 out of 36 attacks. The limitation of this study is convolutional neural networks are stateless, this means they lack the ability to learn beyond what was used as a sample.

Ihor Tereikovskyi [24] proposed a model using deep neural networks to detect cyber-attacks. The model was written using the programming language Python. The NSL-KDD dataset was used, this is a modification of KDD-99. The dataset contained a set of data to be audited, including a variety of intrusions in a military environment, the number of records was 25,192. The features used in the NSL-KDD dataset were:

1. duration—time of connection in seconds.
2. protocol_type—UDP, TCP etc.
3. service—network service.
4. flag—connection status.
5. src_bytes—Data amount transferred from source to recipient in bytes.
6. dst_bytes—Data amount transferred from recipient to source in bytes.

These features served as input variables for the neural network model, the number of input neurons was 4 which relates to the number of cyberattacks.

The features were combined into groups:

1. Basic attributes.
2. Content attributes.
3. Host traffic attributes.

The dataset contained values of each feature to detect the following types of cyber-attacks:

1. Distributed Denial of Service.
2. Probe.
3. Remote to Local.
4. User to Root.

The classifier had a 90% detection rate. A positive point of Ihor Tereikovskyi (2017)'s model is it was written in the programming language Python. Python consists of code libraries, making it easier to use and the preferred language for machine learning. The limitations of this study were the dataset only contained 25,192 records, in order for deep neural networks to perform better over other techniques, a larger dataset is required.

1.2 Summary

In this section, 2 concepts were investigated: cyber security and data mining techniques. The different components of cyber security were discussed: cyber threats, threat actors and cyber activities. Alongside this, the different data mining techniques were discussed, including their definitions and how they can be applied to cyber security.

2 Data Pre-processing

2.1 Introduction

This section aims to demonstrate the data type and pre-processing stage. Different tools which have been used for pre-processing the data will be explained.

2.2 Data Collection

This section aims to demonstrate the data collection process. The main source of data used in this research is Hackmagedon (https://www.hackmageddon.com/). Hackmagedon is a blog which contains timelines and statistics for cyber-attacks. The data collected from Hackmagedon includes cyber-attacks which occurred between 2017 and 2019 within countries in Europe. The dataset includes 1989 records of cyber-attack incidents where the type of attack is known and unknown.

2.3 Data Categorize

At this stage, the initial dataset needs to be explored in more detail. Each incident of cyber-attack comes with 11 features:

1. ID—Identifying number given to the record.
2. Date—the date the attack was recorded.
3. Author—the person behind the attack e.g. Lizard Squad.
4. Target—the name of the target.
5. Description—a brief description of the attack.
6. Attack—the type of cyber-attack e.g. Brute-Force.
7. Target Class—the business sector affected by the attack.
8. Attack Class—the category the attack falls into e.g. cyber criminals.
9. Country—the country affected by the attack e.g. UK.
10. Link—a link to a news article about the attack.
11. Tags—key words relating to the attack.

Table 1 shows an example of the structure of the dataset:

The columns that are not relevant to the research will be removed; this is because they are not useful to the predictive model. The ID, Date, Target, Description, Link and Tags columns will be removed. These columns are not an area of interest and will not improve the model. Finally, the Author column will be renamed to Attackers. The remaining columns that will be used as part of the model are: Attackers, Attack, Attack Class, Target Class and Country. After the columns are removed, the dataset is restructured, and 5 different columns remain.

Table 1 Dataset row example

Date	Author	Target	Description	Attack	Target class	Attack class	Country	Link	Tags
01/01/17	APT28 AKA fancy bear	Unnamed TV station	N/A	Targeted attack	Information and communication	CE	GB	N/A	APT2, secure works

OpenRefine will be used to cleanse the data. The data will be inputted into OpenRefine and pre-processing will begin, pre-processing includes the following steps:

1. Removing Duplicates: This will be done using facet text. In OpenRefine, Facet values provide an overview of the values in a column, making it easier to detect duplicates. For example, Account Hijacking was present twice due to a spelling mistake.
2. Capital and lower-case letters: OpenRefine can also cluster the text so lower-case and capital letters are automatically integrated. Brute-force and Brute-Force were both values in the Attack column, the difference was the capital F. The lower-case F became a capital, this merged the values into one.
3. Removing irrelevant records: This is done manually in Excel by filtering out the records which are Unidentified and contain no information about an attack.
4. Combining Values: Values need to be combined for them to be categorised for further analysis, multiple columns must be categorised:
5. Attack: The attack is categorised based on the nature of the attack. Table 2 shows the Attack categories, abbreviations and examples:

(a) Target Class: The target class is categorised based on the sector attacked. Table 3 shows the Target Class categories and examples:

(b) Attack Class: The attack class is categorised based on the class of the attack (Table 4)

Table 2 Type of threat and abbreviations

Attack	Example	Abbreviation
Account Hijacking	Account Hijacking occurs when a person's account is stolen by a hacker	AH
Brute-Force	An attacker submits multiple passwords until they gain access to a victim's account	BF
DDoS	An attacker floods a service with internet traffic to disrupt users from accessing the service as normal	DDOS
Injection	An attacker inserts malicious code into an application to manipulate the application into working a certain way	INJ
Malware	A type of software designed to cause damage to a computer: viruses, worms, trojan horses	M
MITM	An attacker listens into the data sent between two computers	MITM
Phishing	An attacker attempts to gain sensitive information from a user by pretending to be a known, trustworthy source	PH
Social Bots	Automated social media accounts	SB
Targeted Attack	An anonymous attacker actively trying to infiltrate a victim's system	TA
Unidentified	The type of attack is unknown	UID

Table 3 Type of target abbreviations

Target class	Example	Abbreviation
Administration Activities	Preparing budget, personnel management, maintenance of computer data	AA
Education and Social Work	Schools, universities and social	EASW
Financial and Communication Activities	Banking companies and communications related activities	FACA
Multiple Individuals	Several individuals	MI
Production Related Activities	Includes manufacturing companies, mining etc	PRA
Retail and Transport	Retail shops and transportation companies	RAT
Science and Technology	Companies providing technology and scientific research	SAT
Single Individual	Single individual	SI
Unidentified	The target class is unknown	UID
Utilities	Water, gas etc	UT

Table 4 Attack class abbreviations

Attack class	Example	Definition
CE	Stealing/spying on classified information from a government entity or organization	Cyber Espionage
H	The use of computer systems for politically motivated reasons	Hacktivists
CC	A criminal which uses a computer to commit crime	Cyber Criminals
CW	The use of computer systems to attack a country	Cyber Warfare

(c) Attackers: The attack is categorised based on the group or person carrying out the attack (Table 5)

2.4 Data Statistics

2.4.1 Cyber Attacks by Attack Type

Malware attacks accounted for 43.06% of all cyber-attacks which occurred between 2017 and 2019. This is possibly due to the fact that ransomware attacks are on the rise. Whereas customer ransomware attacks are decreasing, enterprise ransomware attacks are increasing (Fig. 1).

One factor which has played a role in the increase is the number of ransomware authors, ransomware authors know there is a good chance the business will pay the ransom as it will be more costly for the business to suffer an outage than to pay the

Table 5 Attackers abbreviations

Attackers	Abbreviation
Anonymous	A
Unidentified	UID
Chinese Hackers	CH
Iranian Hackers	IH
Multiple Attackers	MA
Nigerian Hackers	NH
North Korean Hackers	NKH
Russian Hackers	RH
Turkish Hackers	TH
USA Hackers	USAH

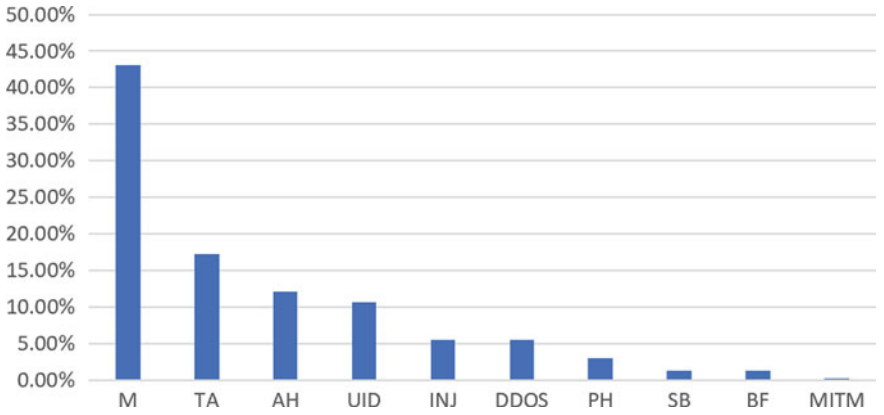

Fig. 1 Cyber attacks by type of attack

author [19]. In 2017, 39% of businesses hit by ransomware paid the author, in 2018, this increased to 45%. Finally, in 2019 this figure increased to 58% [7].

2.4.2 Cyber Attacks by Target Class

The industry most affected by cyber-attacks was Education and Social Work between 2017 and 2019, 30.98% of attacks were on this sector. In the past two years, cyber-attacks on higher education institutions exposed over 1.35 million student identities [17].

Universities are specifically targeted due to the sensitive information stored within their systems. Additionally, cyber-attacks on universities occur because the systems are large and complex, which makes implementing protections correctly difficult [4] (Fig. 2).

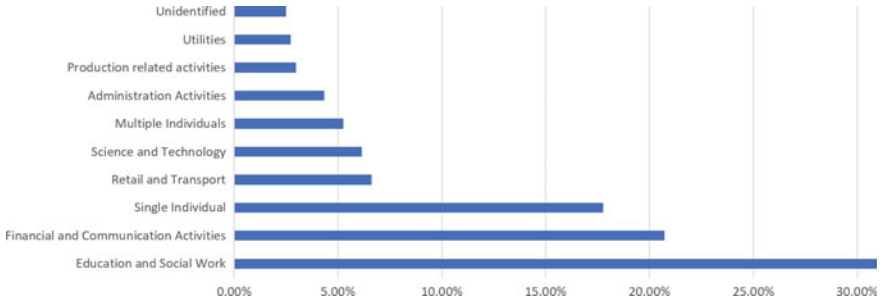

Fig. 2 Cyber attacks by industry

2.4.3 Cyber Attacks by Attack Class

Cyber Criminals were responsible for 75.70% of cyber-attacks which took place between 2017 and 2019. In 2017, the number of cyber incidents was up by 46% and in 2018, this decreased to 43% and finally, in 2019, this number decreased to 32%. One of the reasons for this decrease is between September 2017 and September 2018, the number of computer misuse incidents decreased from 1.5 million to 1 million. Another explanation for fewer businesses identifying breaches is organisations are becoming more secure and aware of cyber incidents. Since 2018, organisations have increased their defences against cyber-attacks (GOV UK Department for Digital, [14]) (Fig. 3).

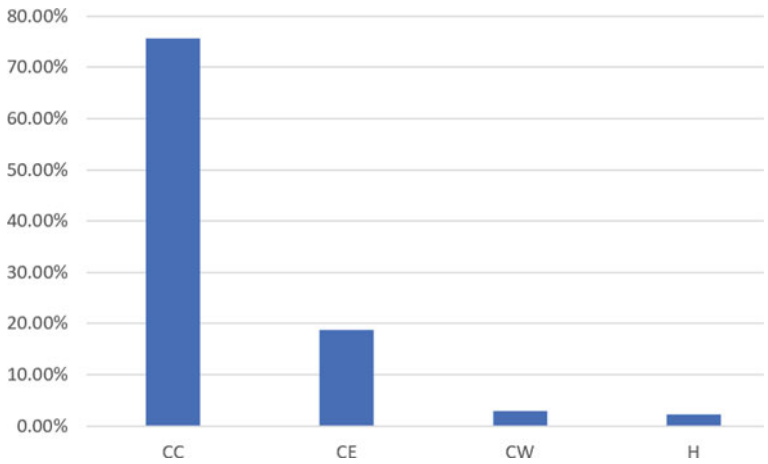

Fig. 3 Cyber attacks by attack class

2.4.4 Cyber Attacks by Country

The country most affected by cyber-attacks across all industries and attack types was the United Kingdom. In 2019, almost 50% of UK businesses suffered a security breach or cyber-attack; in the last 12 months, only half of the businesses had completed an internal and external security audit, it is highly likely they did not have the correct security protections in place to prevent a cyber-attack [23] (Fig. 4).

2.5 Summary

This section discussed the data collected and the different tools and techniques used to structure the data to prepare it for further analysis, alongside the data statistics.

3 Data Analysis

3.1 Introduction

This section will focus on investigating different classification algorithms, such as Support Vector Machine, Random Forest, K-Nearest Neighbour, Naïve Bayes and Neural Networks. Each algorithm will be applied to the dataset and the results will be analysed and compared to discover which one produces the most accurate results in predicting the type of attack used in a cyber-attack.

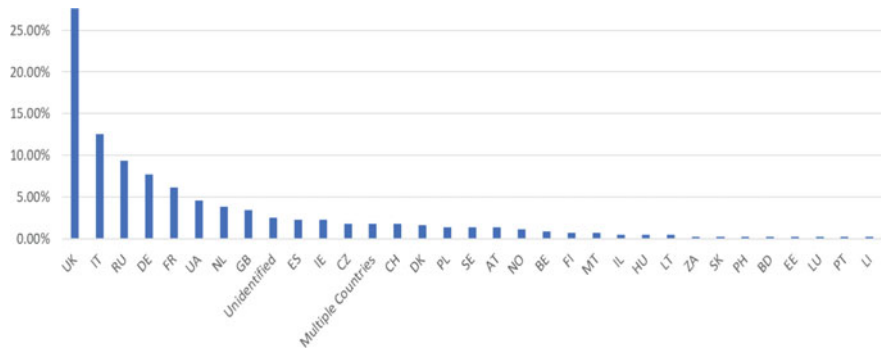

Fig. 4 Cyber attacks by country

3.1.1 Performance Evaluation Metrics

The following are the performance evaluation metrics the algorithms are going to be measured against (Table 6):

3.2 Support Vector Machine Analysis

Now data pre-processing has been completed, the analysis stage can begin. In this part of the analysis, Support Vector Machine, which was discussed in Sect. 1.1.1, will be applied to the dataset to train the data in predicting the different types of attacks used in a cyber-attack. 10-fold cross validation will be applied.

Table 6 Explanation of performance evaluation metrics

Name	Definition	
TP rate	Correctly classified values where the result is correct and is actually correct [10]	
FP rate	Incorrectly classified instances where the result is correct and is actually incorrect [10]	
Precision	The percentage of results returned which are more relevant than irrelevant [22]	Precision = TP/TP + FP
Recall	The percentage of results returned which are mostly relevant [22]	Recall = TP/TP + FN
F-measure	A weighted average of the Recall and Precision. Takes false positives/false negatives into account [10]	F-Measure = 2*(Recall*Precision) / (Recall + Precision)

```
Correctly Classified Instances        1125               63.3089 %
Incorrectly Classified Instances       652               36.6911 %
Kappa statistic                          0.3761
Mean absolute error                      0.0815
Root mean squared error                  0.2855
Relative absolute error                 52.1119 %
Root relative squared error            102.1648 %
Total Number of Instances             1777
```

Fig. 5 SVM summary by attack

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.147	0.010	0.485	0.147	0.225	0.243	0.568	0.124	DDOS
	0.834	0.059	0.772	0.834	0.802	0.753	0.887	0.676	TA
	0.958	0.591	0.601	0.958	0.739	0.434	0.684	0.596	M
	0.009	0.001	0.333	0.009	0.018	0.046	0.504	0.064	INJ
	0.004	0.003	0.200	0.004	0.008	0.010	0.501	0.135	AH
	0.000	0.000	?	0.000	?	?	0.500	0.015	BF
	0.000	0.000	?	0.000	?	?	0.500	0.034	PH
	0.000	0.000	?	0.000	?	?	0.500	0.015	SB
	0.000	0.000	?	0.000	?	?	0.500	0.002	MITM
Weighted Avg.	0.633	0.297	?	0.633	?	?	0.668	0.449	

Fig. 6 SVM attack accuracy by class

3.2.1 Prediction of Attack by Support Vector Machine

Figure 5 demonstrates the Support Vector Machine classifier correctly classified 63.3% of instances and incorrectly classified 36.6% of instances based on the type of attack.

Figure 6 demonstrates the Support Vector Machine classifier yielded a TP rate of 0.633%, a FP rate of 0.297% and a Recall rate of 0.633%.

The M class has a TP rate of 0.958%, a FP rate of 0.591%, a Precision of 0.601%, a Recall of 0.958% and a F-Measure of 0.739. As the M class has a significantly higher Recall (0.958%) compared to the Precision (0.601%), this means the classifier is overclassifying instances as M. The classifier correctly classified 820 instances of M and incorrectly classified 33 as TA. Out of all the classes, M and TA were classified the most frequently.

TA has a TP rate of 0.834%, a FP rate of 0.059%, a Precision of 0.772%, a Recall of 0.834% and an F-Measure of 0.802%. The Recall and Precision are not drastically different, meaning whilst the classifier is recalling a high number of instances, the instances recalled are correct the majority of the time due to the high Precision. 287 instances of TA were classified correctly and 56 were misclassified as M. It is possible M and TA share certain patterns which makes the classifier believe TA is M and vice versa. INJ, AH and DDOS were the only other classes aside from M and TA to have correct classifications, however, the majority of the instances belonging to the classes were also misclassified as TA or M.

3.2.2 Discussion and Interpretation

Table 7 demonstrates the accuracy of the Support Vector Machine classifier:

The SVM classifier accurately classified the type of attack 63.3% of the time. The classifier performed the greatest when predicting Malware or Targeted Attack instances. It is possible this is because there are patterns within the Malware and

Table 7 SVM accuracy in predicting

Type of prediction	SVM accuracy rate
Attack	63.3%

Targeted Attack classes which the other classes do not have. The Targeted Attack class demonstrated equally high Recall and Precision, this demonstrates the classifier is selecting relevant instances which are also accurate to the class, whereas the Malware class shows a higher Recall than Precision, this means a portion of instances classified as Malware belong to another class. Aside from the classes Malware and Targeted Attack, the classifier accurately classified instances of Injection, Account Hijacking and DDOS into the correct class. However, several instances from each class were misclassified as DDOS, Targeted Attack and Malware. Out of all the classes, Malware contains the highest number of misclassified instances.

3.3 Random Forest Analysis

Random Forest, which was discussed in Sect. 1.1.2, will be applied to the dataset to train the data in predicting the different types of attacks used in cyber-attacks. tenfold cross validation will be applied.

3.3.1 Prediction of Attack by Random Forest

Figure 7 demonstrates the Random Forest classifier correctly classified 67.4% of instances and incorrectly classified 32.5% of instances based on the type of attack.

Correctly Classified Instances	1199	67.4733 %
Incorrectly Classified Instances	578	32.5267 %
Kappa statistic	0.4883	
Mean absolute error	0.1041	
Root mean squared error	0.2352	
Relative absolute error	66.5507 %	
Root relative squared error	84.1378 %	
Total Number of Instances	1777	

Fig. 7 RF summary by attack

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.330	0.014	0.610	0.330	0.429	0.424	0.853	0.383	DDOS
	0.817	0.042	0.824	0.817	0.820	0.778	0.929	0.815	TA
	0.900	0.415	0.668	0.900	0.767	0.507	0.787	0.710	M
	0.264	0.013	0.569	0.264	0.360	0.361	0.759	0.260	INJ
	0.279	0.049	0.472	0.279	0.351	0.290	0.693	0.347	AH
	0.000	0.001	0.000	0.000	0.000	-0.004	0.675	0.039	BF
	0.000	0.003	0.000	0.000	0.000	-0.010	0.548	0.036	PH
	0.593	0.005	0.640	0.593	0.615	0.610	0.908	0.608	SB
	0.000	0.000	?	0.000	?	?	0.595	0.003	MITM
Weighted Avg.	0.675	0.216	?	0.675	?	?	0.796	0.598	

Fig. 8 RF attack accuracy by class

Figure 8 demonstrates the Random Forest classified yielded a TP rate of 0.675%, a FP rate of 0.216% and a Recall of 0.675%

The M class is classified more than any other class; however, it also has the highest Recall rate, 0.900%, it is likely there is a pattern within the M instances which makes the classifier believe the selected instance is M and misclassify. 770 instances of M were classified correctly, however, throughout the other classes, there are several misclassifications as M. For example, 75 instances of INJ and 45 instances of DDOS were misclassified as M. Aside from the M class, the classifier accurately classified instances of TA, DDOS, AH, INJ and SB into the correct class. However, there were also a high number of misclassifications. The classes which were misclassified the most were M, TA, DDOS and AH. The only class to have no correct classifications was MITM, all the instances of MITM were misclassified as DDOS, M and TA.

3.3.2 Discussion and Interpretation

Table 8 demonstrates the accuracy of the Random Forest classifier in terms of the different attacks within the dataset which apply to cyber-attacks:

The RF classifier accurately classified the type of attack 67.1% of the time. The classifier performed the greatest when predicting Targeted Attack, Social Bot and Malware instances, it is likely this is due to these classes sharing certain characteristics which the other classes do not have. The Targeted Attack class demonstrates equally high Recall (0.817%) and Precision (0.824%), this demonstrates the classifier is selecting relevant instances which are also accurate to the class, whereas the Malware class shows a higher Recall (0.900%) than Precision (0.668%), this means a portion of instances classified as Malware belong to another class. The classifier accurately classified instances of TA, DDOS, AH, INJ and SB. Out of all the classes, Man in the Middle, Brute Force and Phishing were the only classes to not have a correct classification.

3.4 Naïve Bayes Analysis

Naïve Bayes, which was discussed in Sect. 1.1.3, will be applied to the dataset to train the data in predicting the type of attack used during a cyber-attack. tenfold cross validation will be applied.

Table 8 RF accuracy in predicting

Type of prediction	RF accuracy rate
Attack	67.4%

3.4.1 Prediction of Attack by Naïve Bayes Classifier

Figure 9 demonstrates the Naïve Bayes classifier correctly classified 64.3% of instances, and incorrectly classified 35.6% of instances based on the type of attack.

Figure 10 demonstrates the Naïve Bayes classifier yielded a TP rate of 0.643%, FP rate of 0.224% and a Recall rate of 0.643%.

The TA class yields a TP rate of 0.831%, a FP rate of 0.53%, a Precision of 0.790%, a Recall of 0.877% and a F-Measure of 0.810%. The recall is slightly higher than the precision, this means the classifier is selecting relevant instances which are correct a majority of the time. For example, the classifier accurately classified 286 instances of TA. There are a few misclassifications of TA in the other classes, for example, there are 23 misclassifications of TA in the AH class. However, unlike the M class, TA has a low number of misclassifications. The M class yields a TP rate of 0.877%, a FP rate of 0.421%, a Precision of 0.659% and a Recall of 0.877%, as the recall is higher than the precision, this indicates there are patterns within the M class which causes the classifier to misclassify instances as M. The classifier believes the instances to be relevant to the M class and classifies the instances. In the INJ class, 14 instances were correctly classified and 76 instances were misclassified into the M class, it is possible this is due to M containing similar patterns which the other classes do not have which makes the classifier more likely to classify the instances as M. SB and DDOS were the only classes to have a higher Precision rate over Recall, this shows the instances recalled were relevant to the class.

Correctly Classified Instances	1143	64.3219 %
Incorrectly Classified Instances	634	35.6781 %
Kappa statistic	0.4399	
Mean absolute error	0.0997	
Root mean squared error	0.2395	
Relative absolute error	63.7098 %	
Root relative squared error	85.7043 %	
Total Number of Instances	1777	

Fig. 9 NB summary by attack

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.312	0.026	0.436	0.312	0.364	0.334	0.844	0.377	DDOS
	0.831	0.053	0.790	0.831	0.810	0.764	0.943	0.840	TA
	0.877	0.421	0.659	0.877	0.753	0.475	0.793	0.717	M
	0.127	0.011	0.424	0.127	0.196	0.207	0.741	0.236	INJ
	0.183	0.064	0.310	0.183	0.230	0.151	0.696	0.283	AH
	0.000	0.000	?	0.000	?	?	0.615	0.029	BF
	0.000	0.001	0.000	0.000	0.000	-0.004	0.519	0.037	PH
	0.519	0.005	0.636	0.519	0.571	0.569	0.856	0.509	SB
	0.000	0.000	?	0.000	?	?	0.410	0.002	MITM
Weighted Avg.	0.643	0.224	?	0.643	?	?	0.797	0.593	

Fig. 10 NB attack accuracy by class

Table 9 NB accuracy in predicting

Type of prediction	NB accuracy rate
Attack	64.3%

3.4.2 Discussion and Interpretation

Table 9 demonstrates the accuracy of the Naïve Bayes classifier:

The NB classifier accurately classified the type of attack 64% of the time. The classifier was more likely to recall and classify instances as Targeted Attack or Malware. Both of these classes have a higher Recall than Precision, this means the classifier is recalling instances and misclassifying. Unlike Targeted Attack or Malware, Social Bot had a higher Precision than Recall, meaning the instances recalled were accurate to the class. The classifier classified no correct instances of Man in the Middle, Phishing and Brute Force, the classifier misclassified these instances as DDOS, Targeted Attack and Malware.

3.5 K-Nearest Neighbour Analysis

K-Nearest Neighbour, which was discussed in Sect. 1.1.4, will be applied to the dataset to train the data in predicting the type of attack used during a cyber-attack. tenfold cross validation will be applied.

3.5.1 Prediction of Attack by KNN

Figure 11 demonstrates the KNN classifier correctly classified 67.2% of instances, and incorrectly classified 32.7% of instances based on the type of attack.

Figure 12 demonstrates the KNN classifier yielded a TP rate of 0.672%, a FP rate of 0.222%, and a Recall of 0.672%.

The classifier shows the TA class has a TP rate of 0.826%, a FP rate of 0.039, a Precision of 0.835% and a Recall rate of 0.826%. As the Precision is slightly higher than the Recall, this demonstrates the classifier is selecting relevant instances

```
Correctly Classified Instances        1195               67.2482 %
Incorrectly Classified Instances       582               32.7518 %
Kappa statistic                          0.482
Mean absolute error                      0.1025
Root mean squared error                  0.2391
Relative absolute error                 65.5001 %
Root relative squared error             85.5589 %
Total Number of Instances             1777
```

Fig. 11 KNN summary by attack

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.321	0.019	0.530	0.321	0.400	0.384	0.825	0.374	DDOS
	0.826	0.039	0.835	0.826	0.830	0.790	0.915	0.818	TA
	0.903	0.428	0.662	0.903	0.764	0.500	0.779	0.696	M
	0.227	0.015	0.500	0.227	0.313	0.309	0.753	0.249	INJ
	0.275	0.043	0.500	0.275	0.355	0.302	0.697	0.333	AH
	0.000	0.001	0.000	0.000	0.000	-0.003	0.646	0.035	BF
	0.000	0.002	0.000	0.000	0.000	-0.009	0.535	0.035	PH
	0.444	0.003	0.706	0.444	0.545	0.555	0.897	0.508	SB
	0.000	0.000	?	0.000	?	?	0.689	0.005	MITM
Weighted Avg.	0.672	0.222	?	0.672	?	?	0.787	0.586	

Fig. 12 KNN attack accuracy by class

and correctly classifying them as belonging to the TA class. Whereas M shows a Precision of 0.662% and a Recall of 0.903%, in this case the Recall is higher than the Precision, although the classifier is selecting relevant instances, a high proportion of the instances are being misclassified as M. For example, 50 instances of DDOS were misclassified as M.

The SB class has a TP rate of 0.444%, a FP rate of 0.003%, a Precision of 0.706% and a Recall of 0.444%. This means the classifier does not believe many instances belong in the SB class and a high proportion of the SB instances are being misclassified. For example, 12 instances of SB were correctly classified, however, 14 instances were misclassified into DDOS, TA, M and AH. This could be because the SB class shares similarities with the DDOS, TA, M and AH classes which makes the classifier believe instances of SB belong to those classes.

3.5.2 Discussion and Interpretation

Table 10 demonstrates the accuracy of the KNN classifier:

The K-Nearest Neighbour classifier using the type of attack feature accurately classified the type of attack 67% of the time. The classifier performed the greatest when predicting Targeted Attack instances. The Recall (0.826%) and Precision (0.835%) are equally as high, this demonstrates the classifier is selecting relevant instances which are also accurate to the class. The Malware class also had a high number of classifications; however, the Recall is higher than the than Precision, this a significant portion of instances classified as Malware belong to another class. Similar to Targeted Attack, Social Bot shows a higher Precision (0.706%) than Recall (0.444%), however, as the Recall is significantly lower, this demonstrates the classifier is not recalling Social Bot instances.

Table 10 KNN accuracy in predicting

Type of prediction	KNN accuracy rate
Attack	67%

3.6 Neural Networks

Neural Networks, which was discussed in Sect. 1.1.5, will be applied to the dataset to train the data in predicting the type of attack used during a cyber-attack. tenfold cross validation will be applied.

3.6.1 Prediction of Attack by NN

Figure 13 demonstrates the NN classifier correctly classified 66.8% of instances, and incorrectly classified 33.1% of instances based on the type of attack.

Figure 14 demonstrates the NN classifier yielded a TP rate of 0.669%, a FP rate of 0.212% and a Recall of 0.669%.

The M class yields a TP rate of 0.897%, a FP rate of 0.403%, a Precision of 0.674%, a Recall of 0.897% and an F-Measure of 0.770%. The Recall is higher than the Precision, this means the classifier is selecting instances it believes to be relevant to the class, however, as the Precision is also high, this is a strong indicator that the majority of the instances recalled are accurate to the M class. The classifier accurately classified 768 instances of M. Out of all of the class, the AH class had the highest number of M classifications. This indicates M and AH share similar patterns which lead the classifier to believe instances of M are AH.

Correctly Classified Instances	1188	66.8542 %
Incorrectly Classified Instances	589	33.1458 %
Kappa statistic	0.481	
Mean absolute error	0.098	
Root mean squared error	0.2402	
Relative absolute error	62.6439 %	
Root relative squared error	85.9269 %	
Total Number of Instances	1777	

Fig. 13 NN summary by attack

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.330	0.023	0.486	0.330	0.393	0.369	0.823	0.401	DDOS
	0.808	0.040	0.827	0.808	0.818	0.775	0.917	0.819	TA
	0.897	0.403	0.674	0.897	0.770	0.515	0.794	0.712	M
	0.227	0.015	0.500	0.227	0.313	0.309	0.752	0.231	INJ
	0.279	0.054	0.447	0.279	0.344	0.277	0.685	0.309	AH
	0.000	0.002	0.000	0.000	0.000	-0.005	0.666	0.042	BF
	0.000	0.002	0.000	0.000	0.000	-0.008	0.568	0.042	PH
	0.519	0.005	0.636	0.519	0.571	0.569	0.885	0.507	SB
	0.000	0.000	?	0.000	?	?	0.511	0.003	MITM
Weighted Avg.	0.669	0.212	?	0.669	?	?	0.793	0.592	

Fig. 14 NN attack accuracy by class

3.6.2 Discussion and Interpretation

Table 11 demonstrates the accuracy of the Neural Network classifier:

The Neural Network classifier accurately classified the type of attack 66.8% of the time. The classifier performs the best when predicting Targeted Attacks, the Precision (0.827%) and Recall (0.808%) for this class are equally as high, this means the instances recalled are relevant and accurate. However, the classifier believes a high number of instances to be Malware, the Malware class has a Recall of 0.897% and a Precision of 0.674%. This means the classifier is recalling and misclassifying instances which do not belong to the Malware class. Unlike Targeted Attack were the Precision and Recall are equally as high, Injection shows a higher Precision (0.500%) than Recall (0.227%), as the Recall is significantly lower, this demonstrates the classifier is not recalling Injection instances.

3.7 Summary

In this section, data analysis was carried out. Support Vector Machine, Random Forest, Naïve Bayes, K-Nearest Neighbour and Neural Networks were applied to the data and various performance metrics were discussed and analysed. For each algorithm, a discussion was carried out discussing the key points made from the analysis.

4 Model Comparison

4.1 Introduction

In this section, the different algorithms are compared together in order to assess which one performs the greatest in predicting the type of attack.

Table 11 NN accuracy in predicting

Type of prediction	NN accuracy rate
Attack	66.8%

4.2 Comparison of Models

5 different models were trained for predicting the type of attack used in a cyber-attack, these models are: Support Vector Machine, Random Forest, Naïve Bayes, K-Nearest Neighbour and Neural Networks.

Out of all the models trained, the Random Forest classifier demonstrated the highest accuracy in terms of TP (0.675%) and FP (0.216%). Instances were classified correctly 67.4% of the time. The classifier successfully classified instances correctly into each class, aside from MITM. In second place, K-Nearest Neighbour demonstrated a lower TP rate (0.672%) and a higher FP rate (0.222%). However, instances were classified correctly 67.2% of the time. This is 0.2% less than the Random Forest Classifier. The RF model had less misclassifications than KNN and was more successful in classifying Social Bots, Injection and Account Hijacking instances than KNN.

Compared to Random Forest, Support Vector Machine had significantly more misclassifications, the only classes to have correct classifications were DDOS, Targeted Attack and Malware instances. The FP rate was higher, 0.297%, when compared to Random Forest's FP rate of 0.216% and the TP rate was lower, 0.633%. Instances were classified correctly 63.3% of the time, 4.1% less than Random Forest. Neural Networks and Naïve Bayes both had a lower TP rate. Compared to Random Forest, Neural Networks and Naïve Bayes demonstrated a higher number of DDOS misclassifications.

Due to the RF model demonstrating a higher TP rate, a lower FP rate and a higher accuracy in classifying instances correctly when compared against the other models, Random Forest was chosen as the most suitable algorithm for the final model (Table 12).

As shown in the table above, Random Forest demonstrated the highest accuracy in predicting the type of attack. Therefore, the model will be built using the Random Forest classification algorithm. It is possible Random Forest yielded the highest accuracy rate due to Random Forest being an ensemble model. The Random Forest model creates trees on the subset of the data and combines the output from all the trees. This reduces overfitting and reduces the variance, subsequently, this improves the accuracy of the classifier.

Table 12 Comparison of model accuracy rate

Algorithm	Accuracy rate (%)
RF	67.4
KNN	67.2
NN	66.8
NB	64.3
SVM	63.3

4.3 Summary

Based on the model comparison, Random Forest performed the best and was chosen for the final model. In the next section, in order to evaluate how successful the model is, validation data will be applied which contains unseen records of cyber-attacks which occurred during January and May 2020 during the COVID-19 pandemic.

5 Discussion

5.1 COVID 19 Data Statistics

5.1.1 Cyber Attacks by Attack Type during COVID 19

Since the COVID 19 pandemic began in January, Malware attacks have accounted for 54.54% of cyber-attacks, this is an increase of 11.48% from the end of 2019, prior to the pandemic. Similarly, Account Hijacking attacks have increased by 3%. It is likely the increase in cyber-attacks is due to the increase in remote working since the pandemic began (Fig. 15).

For example, the coronavirus pandemic has forced organisations to roll out new technologies at an unprecedented rate. Employees who have been working from home have been using virtual private networks which lack the necessary protections, this opens up a point of entry for a potential attacker.

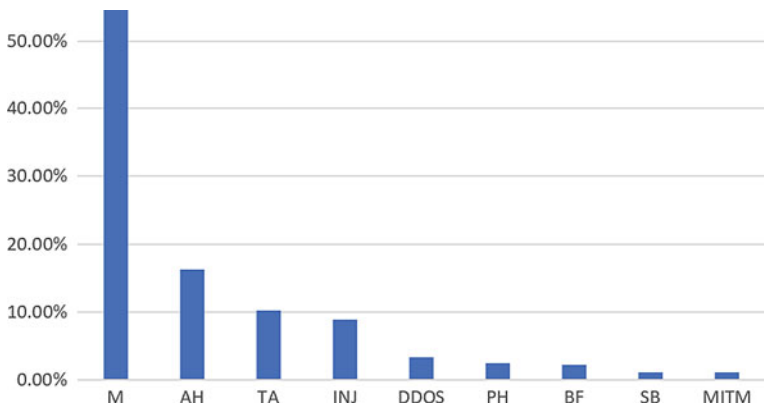

Fig. 15 Cyber attacks by type of attack during COVID19

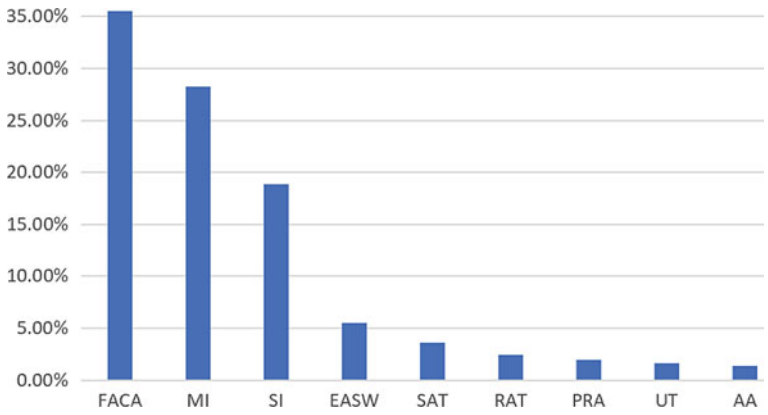

Fig. 16 Cyber attacks by target class during COVID19

5.1.2 Cyber Attacks by Target Class during COVID 19

The Financial and Communication Activities industry suffered more cyber-attacks than any other industry, accounting for 35.5% of all cyber-attacks. This is an increase of 24.4% from the end of 2019 (Fig. 16).

From February to April, banks have seen an increase in cyber-attacks by 238%. It is likely attackers are exploiting the situation of the pandemic, for example, the excessive demand for goods, the anxiety of the general population and the fact a high proportion of the population are working from home. Alongside this, attackers have been using coronavirus related information to harvest data and compromise personal data. As there is a large amount of misinformation spreading regarding the coronavirus pandemic, this means the general public are more likely to believe a phishing scam.

5.1.3 Cyber Attacks by Attack Class during COVID 19

Cyber Criminals were responsible for 89% of cyber-attacks which took place during the COVID 19 pandemic. This is an increase of 13.3% since 2019, prior to the pandemic, demonstrating that cyber-attacks undertaken by cyber criminals have significantly risen. Subsequently, cyber-attacks by the threat actor group, Cyber Espionage, have decreased by 9.5%, Cyber Warfare has also decreased by 2% and finally, Hacktivism has seen a decrease of 1.75%. However, the rate of Hacktivist attacks has fluctuated over the years. For example, in 2015 Hacktivist attacks saw a decrease of 95% and have been decreasing ever since, in 2017, IBM reported 5 incidents, two in 2018 and none in the first few month of 2019 [6] (Fig. 17).

It is possible that the decrease in Hacktivist attacks is partly due to groups such as Anonymous fading from mainstream view. Alongside this, organisations are more

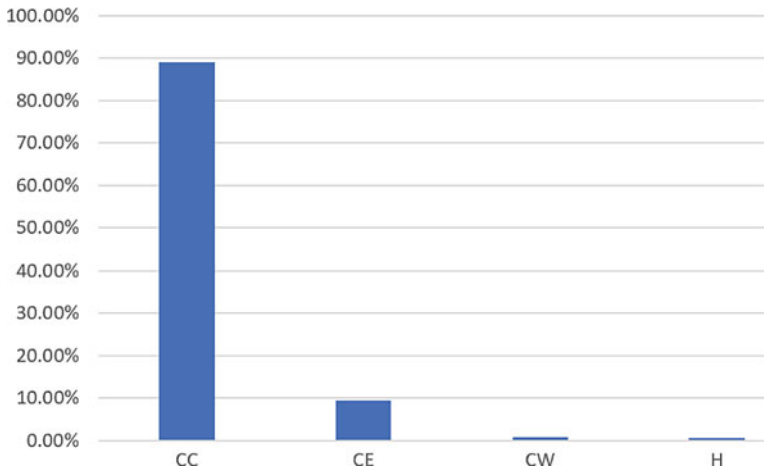

Fig. 17 Cyber attacks by attack class during COVID19

aware of Hacktivism than they were a decade ago and have the adequate protections in place to protect against attacks such as DDOS.

5.1.4 Cyber Attacks by Country During COVID 19

Based on the data, no single country has significantly been affected more than others during the COVID 19 pandemic. 73.5% of cyber-attacks affected multiple countries during one attack (Fig 18).

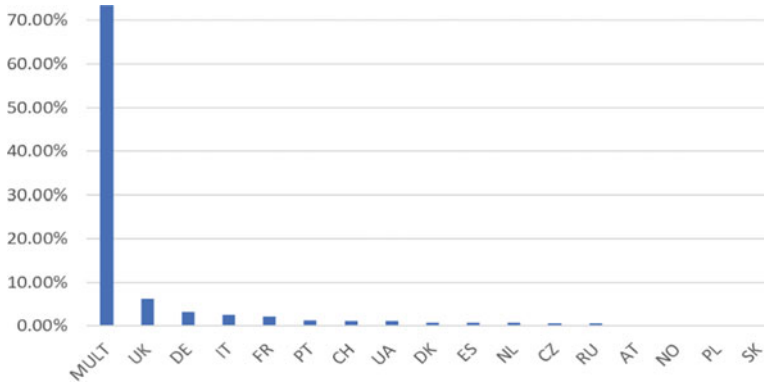

Fig. 18 Cyber attacks by country during COVID19

5.2 Evaluation of Model

To evaluate the effectiveness of the chosen model, a validation dataset is applied, the purpose of this is to see how the model performs in predicting the type of attack on unknown data. The validation dataset includes cyber-attacks which occurred during the COVID-19 pandemic between January and May 2020. One of the main differences between the validation set and the training set is the number of instances, the validation set has 35 instances and the training set has 1777. The likely reason for this difference is the time span of the data collected, for the training dataset, the data spans from 2017 to 2019, whereas the validation set spans from January to May 2020. This was taken into account when analysing the results of the validation set.

5.2.1 Validation of the Type of Attack Predictive Model

In this stage, the Type of Attack will be evaluated by applying the validation dataset to the training set. In the validation dataset, there are 35 instances where the attacks are known.

Figure 19 shows the results of the Random Forest model when the validation dataset is applied. The Random Forest correctly classified instances 62.8% of the time and incorrectly classified instances 37.1% of the time. These figures alone do not provide a clear indicator of the accuracy of the model,therefore, the results need to be analysed further.

The model demonstrated an overall TP rate of 0.629%, a FP of 0.349% and a Recall of 0.629%. The overall results of the model do not demonstrate a reliable and

```
Correctly Classified Instances          22               62.8571 %
Incorrectly Classified Instances        13               37.1429 %
Kappa statistic                          0.3309
Mean absolute error                      0.1166
Root mean squared error                  0.2499
Relative absolute error                 77.8484 %
Root relative squared error             93.5289 %
Total Number of Instances               35

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.000     0.029     0.000       0.000   0.000      -0.029    0.765     0.071     DDOS
               0.714     0.000     1.000       0.714   0.833       0.816    0.893     0.887     TA
               0.800     0.600     0.640       0.800   0.711       0.219    0.603     0.641     M
               0.333     0.031     0.500       0.333   0.400       0.364    0.885     0.487     INJ
               0.000     0.029     0.000       0.000   0.000      -0.029    0.706     0.063     AH
               ?         0.000     ?           ?       ?           ?        ?         ?         BF
               0.000     0.030     0.000       0.000   0.000      -0.042    0.394     0.062     PH
               0.000     0.000     ?           0.000   ?           ?        0.853     0.091     SB
               ?         0.000     ?           ?       ?           ?        ?         ?         MITM
Weighted Avg.  0.629     0.349     ?           0.629   ?           ?        0.688     0.596
```

Fig. 19 Validation of type of attack predictive model

strong model, however, the individual results based on the classes provide a clearer picture.

The Targeted Attack class has the highest portion of correctly classified cyber threats, this means the model will be more accurate in predicting Targeted Attacks. Based upon the Precision, the instances recalled and classified were accurate to the class. The second highest class is Malware; however, the high FP rate suggests the model is selecting instances as Malware and classifying them as such when they belong to another class, another indicator this is the case is the higher Recall over Precision. The third highest class is Injection with a significantly lower TP rate and FP rate than Malware and Targeted Attack. However, TP/FP rate by themselves does not provide a strong indication of accuracy in terms of prediction, the Precision and Recall needs to be taken into account. For example, the Injection class has a lower TP rate, however, when the model does classify instances into the Injection class, the majority of them are correct. This is shown by the Precision being higher than the Recall.

Figure 20 shows the comparison of TP, FP and Precision for each of the different classes. The model failed to classify any instances into the Brute Force, Phishing, Social Bot and Man in the Middle classes.

Figure 21 shows the comparison of Precision and Recall for the different classes. Targeted Attack and Injection are the only classes to have a higher Precision than Recall. The Precision was not calculated for the DDOS class or the overall weight of average for the model.

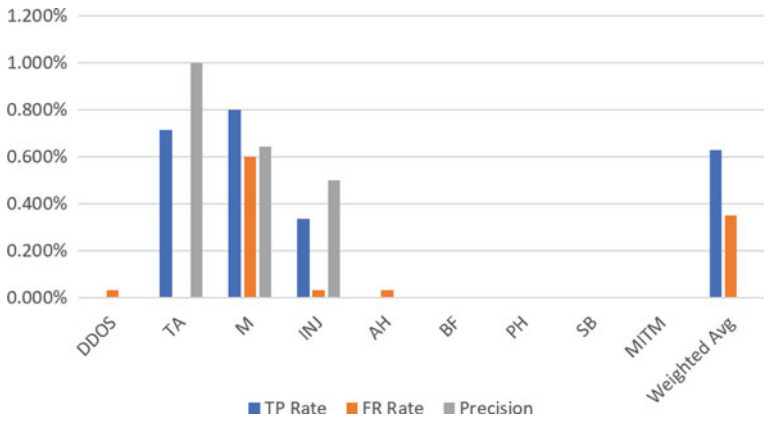

Fig. 20 TP, FP, and precision for prediction of type of attack

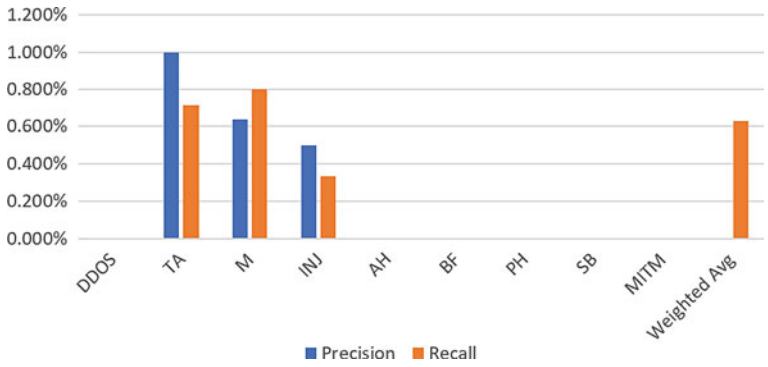

Fig. 21 Precision and recall for prediction of type of attack

5.3 Summary

In this section, the model was evaluated by applying validation data to the final model. Alongside this, variable importance was investigated to discover which variable had the largest impact on the model and the statistics for the COVID 19 validation data were discussed.

6 Conclusion

6.1 Contribution to Knowledge

Over the years, technology has been advancing at an alarming rate, alongside this, the complexity of cyber-attacks has also been progressing beyond comprehension. This means there is an increased urgency for specialists to evolve their methods in how they detect, prevent and manage cyber-attacks. Data mining is a growing technology and it is being frequently utilised in multiple sectors, such as fraud detection in the banking sector, email filtering and marketing. In recent years, data mining has been used to predict trends in cybercrimes and current or past cyber-attacks.

We have developed a predictive model based on data prior to Covid-19 pandemic, the model was applied to a COVID 19 validation dataset, this is data the model had never seen before. Based on the results of the validation data set, it was evident the model detected Malware more than any other class. However, the model demonstrated a high false positive rate when detecting Malware with the validation data. It is likely this is because there is some factor within the Malware class which makes the model inclined to overclassify instances as Malware.

Since the beginning of the COVID 19 pandemic in January, there has been a noticeable increase in cybercrime. Based on the COVID 19 data from January to

May, there is a growing trend in Malware attacks, more than half of cyber-attacks since COVID 19 began were Malware attacks, accounting for 54.5% of all cyber-attacks. Prior to the pandemic, between 2017 to 2019, Malware attacks accounted for 43.06% of cyber-attacks. As mentioned previously, enterprise ransomware is on the rise due to the increasing number of ransomware authors and the fact the attackers know there is a good chance the organisation will pay the ransom.

However, since COVID 19, attackers have been using coronavirus misinformation as a lure, such as financial scams offering financial assistance and free downloads for technology solutions which are in high demand. It is likely that due to the financial desperation COVID 19 has caused, ransomware scams have a higher likelihood of succeeding. Attackers are taking advantage of the general populations existing anxiety which was created by the pandemic and exploiting it for financial gain. Alongside this, remote working has also increased the risk of a successful ransomware attack, this is likely due to a combination of factors, such as weaker security controls and phishing scams via email.

Another observation made on the COVID 19 data is the attackers change in focus on industry. Prior the pandemic, Education and Social Work, such as universities, were more affected than any other industry, accounting for 30.98% of all attacks. However, during COVID 19, this reduced drastically to 6.3%. The highest affected industry during the first three months of the COVID 19 pandemic was Financial and Communication Activities, such as banks, accounting for 35.5% of all cyber-attacks.

Overall, using the model with the COVID 19 data, the model was successful in detecting Targeted Attacks, Malware and Injection attacks based on the high Recall and Precision rates. However, the model was unsuccessful in detecting any instances of Account Hijacking, DDOS, Phishing, Social Bot and Brute Forcing attacks. This demonstrates the model is effective in detecting certain types of attacks during the COVID 19 pandemic.

6.2 Limitations of Study

The limitations in this investigation are related to the data used. As mentioned previously, the data was collected from the Hackmageddon blog. One of the issues with this is the data came with a significant amount of irrelevant data. During the data cleansing phase of this investigation, irrelevant records were removed from the dataset and the various attack categories were reduced. The reduction of the categories was done manually and was also time consuming, this was because this was done manually. Another issue with the data is it is possible there are records stated to be a Malware attack, whereas in actuality they belong to another class. Occasionally, when the description of the attack wasn't very clear and not much information was provided, an educated guess had to be given. Due to this and the data being unbalanced as the dataset contained significantly more Malware attacks than any other class, it is likely this has added to the disproportionate amount of records identified as Malware.

6.3 Future Work

To improve and extend this research further, the following can be accomplished:

1. As mentioned in Sect. 6.2 Limitations of Study, the dataset used was unbalanced. There were a higher number of Malware instances compared to the other classes, the dataset contained 1777 records of cyber-attacks and 856 of these were Malware instances. In order to provide an accurate interpretation across the classes, the number of records in each class would need to be balanced. Alongside this, this study can be improved by using more than one attribute when predicting cyber-attacks, therefore, rather than the framework only predicting the type of attack which occurred, the framework could predict which country it's most likely to occur in.
2. An additional attribute could be used alongside the type of attack, such as the target class or country. This would allow for trends to be analysed and preventative measures put in place in specific areas, for example, which industry is more affected in France by Malware attacks compared to the United Kingdom?
3. The Target Class attribute could be used as the primary feature alongside the type of attack or country to discover which sector is affected the most by specific types of cyber attacks during the COVID-19 pandemic in the UK compared to other countries in Europe, such as Germany or France.
4. Data can be collected and used from client machines to detect whether certain attacks are more likely to occur on a wireless or wired network, or if the machine is using a remote connection.

In this investigation as per the aim and objectives, classification techniques were used. In future research, other data mining techniques can be applied alongside classification, such as clustering. K Means can be applied alongside Random Forest, K Means would classify and assign observations in the dataset into multiple groups, such as attack type or country, based on their similarities. Once the data has been separated into multiple clusters, labels will be applied based on what the clusters are, for example, a cluster representing the different types of malware. By using clustering algorithms alongside classification, clustering detects patterns within the data and assigns them into clusters based on the similarity, once multiple clusters have been created, classification is used to label and classify the data. Due to there being multiple clusters, it is likely more accurate results will be produced.

References

1. BBC (2020) Coronavirus: cyber-attacks hit hospital construction companies. Retrieved from BBC News. https://www.bbc.co.uk/news/technology-52646808
2. Ben Abdel Ouahab Ikram BM (2018) Machine learning application for malwares classification using visualization techniques
3. Bilge L (2017) RiskTeller: predicting the risk of cyber incidents

4. Carfagno D (2019) Why is higher education the target for cyber attacks? Retrieved from Cyber Shark. https://www.blackstratus.com/why-is-higher-education-the-target-for-cyber-attacks/
5. Ch R (2020) Computational system to classify cyber crime offenses using machine learning
6. Cimpanu C (2019) Hacktivist attacks dropped by 95% since 2015. Retrieved from ZDNet. https://www.zdnet.com/article/hacktivist-attacks-dropped-by-95-since-2015/
7. Coble S (2020). Ransomware payments on the rise. Retrieved from InfoSecurity. https://www.infosecurity-magazine.com/news/rise-in-ransomware-payments/#:~:text=%22This%20rise%20is%20arguably%20fueled,up%20to%2058%25%20in%202019.
8. Department for Digital Culture Media & Sport Official Statistics (2020) Cyber security breaches survey 2020
9. Donghoon K (2017) Detection of DDoS attack on the client side using support vector machine
10. Exsilio Solutions (2016) Accuracy, precision, recall & F1 score: interpretation of performance measures. Retrieved from Exsilio
11. Farnaaz N (2016) Random forest modeling for network intrusion detection system
12. Gewirtz D (2020) COVID cybercrime: 10 disturbing statistics to keep you awake tonight. Retrieved from ZDNet. https://www.zdnet.com/article/ten-disturbing-coronavirus-related-cybercrime-statistics-to-keep-you-awake-tonight/
13. Ghanem K (2017) Support vector machine for network intrusion and cyber-attack detection
14. GOV UK Department for Digital CM (2019) Cyber security breaches survey 2019
15. Kravchik M (2018) Detecting cyberattacks in industrial control systems using convolutional neural networks
16. Loader J (2020) Sheffield Hallam University confirms data breach following cyber attack. Retrieved from The Sheffield Tab. https://thetab.com/uk/sheffield/2020/07/30/sheffield-hallam-university-confirms-data-breach-following-cyber-attack-44944
17. Oskoui R (2017) The 5 industries most vulnerable to cyber-attacks. Retrieved from CDNetworks. https://www.cdnetworks.com/cloud-security-blog/the-5-industries-most-vulnerable-to-cyber-attacks/
18. Panda SK (2020) Top five sectors prone to cyber threat amid COVID-19 lockdown. Retrieved from Entrepreneur. https://www.entrepreneur.com/article/350502
19. Posey B (2019) Why enterprise ransomware attacks are on the rise. Retrieved from ITProToday. https://www.itprotoday.com/security/why-enterprise-ransomware-attacks-are-rise
20. Prajakta Yerpude VG (2017) Predictive modelling of crime dataset using data mining
21. Sarker IH (2020) IntruDTree: a machine learning based cyber
22. Saxena S (2018) Precision vs recall. Retrieved from Towards Data Science. https://towardsdatascience.com/precision-vs-recall-386cf9f89488
23. Scroxton A (2020) Almost half of UK businesses suffered a cyber attack in past year. Retrieved from ComputerWeekly. https://www.computerweekly.com/news/252480582/Almost-half-of-UK-businesses-suffered-a-cyber-attack-in-past-year
24. Tereikovskyi I (2017) Deep neural networks in cyber attack detection systems
25. Zareapoor M (2015) Application of credit card fraud detection: based on bagging ensemble classifier