iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40663

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Analyse Cyberattack at Organizations using Logistic Regression Algorithm

Saurabh Kumar Sen[1], Anuradha Deolase[2]
[1]*Saurabh Kumar Sen, Department of Information Technology (Cyber security), Vikrant Institute of Technology and Management, Indore (M.P) India*
[2]*Anuradha Deolase, Professor, Dept. of Information Technology, Vikrant Institute of Technology and Management, Indore (M.P) India*

***Abstract:** Ransomware cyberattacks have grown in severity, effectiveness to cause damage, and ease of execution during the last decade. Advanced ransomware detection technologies must be included with traditional anti-malware procedures. The results of a study and analysis of ransomware attack risk are presented in this work, with the goal of identifying the characteristics that separate ransomware from other malware and benign executable files with the help of detected logs. The ransomware's normal behaviour and structure are determined by statically and dynamically analysing the executable binaries. Ransomware-specific features are extracted from executable files using dynamic and static analysis techniques. This study shows that graph representation of attacks with a collection of datasets for malware detection improves when using machine learning techniques.*
***Keywords:** Ransomware, Malware Detection, Static Analysis, Dynamic Analysis, Anti-malware, Machine learning etc.*

I. INTRODUCTION

Ransomware is a type of software that encrypts files on a computer and prevents the user or organisation from accessing them. This malware encrypts files and demands a ransom payment for the decryption key, putting businesses in a position where paying the ransom is the simplest and cheapest method to recover access to their data. Some ransomware variations have introduced extra capabilities, such as data theft, to entice ransomware victims to pay the ransom.

Ransomware has quickly risen to prominence as the most visible and well-known sort of malware. Recent ransomware attacks have harmed hospitals' capacity to offer critical services, paralysed city government systems, and wreaked havoc on a variety of enterprises.

The modern ransomware craze began with the WannaCry outbreak in 2017[1]. This large-scale, highly configurable attack has proven that ransomware attacks are both possible and profitable. Since then, dozens of ransomware variants have been developed and used in various attacks. The COVID19 pandemic has also contributed to the recent rise in ransomware. As organizations rapidly transition to working remotely, holes have been created in their cyber defences. Cybercriminals exploited these vulnerabilities to launch ransomware, leading to a wave of ransomware attacks. In Q3 2020, ransomware attacks increased by 50% compared to the first half of that year.

II. WORKING STEPS OF RANSOMWARE

In order to be successful, ransomware needs to gain access to a target system, encrypt the files there, and demand a ransom from the victim [2]. While the specifics of implementation differ from one ransomware variation to the next, all three stages remain the same.

1) Step 1. Infection and Distribution Vectors

Ransomware, like any other type of malware, can acquire access to a company's systems in a variety of methods. Ransomware authors, on the other hand, tend to favour a few unique infection vectors.

Phishing emails are one of them. A malicious email may include a link to a website that hosts a malicious download or an attachment with built-in downloader functionality. The ransomware is downloaded and executed on the PC if the email recipient falls for the scam. Another common ransomware infection vector makes use of services like the Remote Desktop Protocol (RDP) (RDP). An attacker can utilise RDP to authenticate to and remotely access a machine on the company network if they have stolen or guessed an employee's login credentials [4]. With this access, the attacker can download malware directly and run it on the machine they control. Others may try to actively infect computers, as WannaCry did with the Eternal Blue vulnerability [1]. The majority of ransomware versions have a variety of infection routes.

2) Step 2. Data Encryption

After gaining access to a machine, ransomware might begin encrypting its files. Because an operating system includes encryption, all that is required is accessing files, encrypting them with an attacker-controlled key, and replacing the originals with the encrypted copies. To maintain system stability, most ransomware strains pick carefully which files to encrypt. To make recovery without the decryption key more difficult, certain variants would erase backup and shadow copies of files [4].

3) Step 3. Ransom Demand

Once the files have been encrypted, the ransomware is ready to demand a payment [3]. Different ransomware variations implement this in a variety of ways, but it's fairly uncommon for the display background to be changed to a ransom note or for text files to be inserted in each encrypted directory with the ransom note. In exchange for access to the victim's files, these messages usually demand a specific sum of cryptocurrency. The ransomware operator will either supply a copy of the private key used to safeguard the symmetric encryption key or a copy of the symmetric encryption key itself if the ransom is paid. This information can be entered into a decryptor tool (also provided by the cybercriminal) to reverse the encryption and give the user access to their data again [2].

While all ransomware variations have these three essential phases, individual ransomware can have various implementations or additional steps. Ransomware variants like as Maze, for example, perform file scanning, registry information theft, and data theft before encrypting data, and WannaCry hunts for more vulnerable machines to infect and encrypt [1].

How to Protect Against Ransomware

A. Utilize Best Practices

The cost and impact of a ransomware assault can be drastically reduced with proper planning. Using the following best practises, an organization's susceptibility to ransomware can be reduced and its effects minimised:

1) *Cyber Awareness Training and Education:* Phishing emails are frequently used to distribute ransomware. It is critical to educate people on how to recognise and avoid ransomware attacks. Because many recent cyber-attacks begin with a targeted email that contains no malware at all, but rather a socially-engineered message that induces the user to click on a harmful link, user education is frequently seen as one of the most critical defences a company can implement [1].
2) *Continuous Data Backups:* Ransomware's definition says that it is malware designed to make it so that paying a ransom is the only way to restore access to the encrypted data. A company can recover from an assault with minimal data loss and without paying a ransom if it uses automated, protected data backups. Maintaining regular data backups as a routine procedure is critical for preventing data loss and ensuring data recovery in the case of corruption or disc hardware failure. Backups that are functional can also assist firms in recovering from ransomware infections.
3) *Patching:* Critical component in defending against ransomware attacks as cyber-criminals will often look for the latest uncovered exploits in the patches made available and then target systems that are not yet patched. As a result, it's vital for businesses to make sure that all of their systems are patched, as this decreases the number of potential vulnerabilities that an attacker could use.
4) *User Authentication:* Accessing services like RDP with stolen user credentials is a favourite technique of ransomware attackers [1]. Strong user authentication makes it more difficult for an attacker to use a guessed or stolen password.

B. Reduce Attack Surface

Because of the huge potential cost of a ransomware attack, the best ransomware mitigation technique is prevention. This can be achieved by reducing the attack surface by addressing [3]:

1) Phishing Messages
2) Unpatched Vulnerabilities
3) Remote Access Solutions
4) Mobile Malware

C. Deploy Anti-ransomware Solution

The need to encrypt all of a user's files means that ransomware has a unique fingerprint when running on a system. Anti-ransomware solutions are built to identify those fingerprints. Common characteristics of a good anti-ransomware solution include [3]: Wide variant detection, Fast detection, Automatic restoration, Restoration mechanism not based on common built-in tools (like 'Shadow Copy', which is targeted by some ransomware variants).

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

III. METHODS TO REMOVE RANSOMWARE

A ransom note on a computer is not something anyone wants to see since it indicates that a ransomware infection has been effective. At this point, several responses to an active ransomware outbreak can be taken, and a company must decide whether or not to pay the ransom.

A. How to Mitigate an Active Ransomware Infection

After data encryption is complete and a ransom letter is shown on the infected computer's screen, many successful ransomware operations are only identified. The encrypted files are likely unrecoverable at this stage; however, some procedures should be performed right away:

1) *Quarantine the Machine:* Some ransomware strains will attempt to propagate to other PCs and associated drives. Limit the malware's spread by denying it access to additional potential targets.
2) *Leave the Computer On:* Encryption of files can make a computer unstable, and turning down a computer can cause volatile memory to be lost. Keep the computer turned on to increase the chances of a successful recovery.
3) *Create a Backup:* Some ransomware variations allow you to decrypt your files without paying the ransom. Make a backup of encrypted files on removable media in the event that a solution becomes available in the future or if an unsuccessful decryption attempt causes the files to be damaged.
4) *Check for Decryptors:* To see if a free decryptor is available, contact the No More Ransom Project. Whether that's the case, try running it on a copy of the encrypted data to see if it can recover the files.
5) *Ask For Help:* Backup copies of files kept on computers are sometimes stored on computers. If the infection hasn't removed these copies, a digital forensics expert may be able to recover them.
6) *Wipe and Restore:* Reinstall the operating system or restore the machine from a clean backup [5]. This verifies that all malware has been eliminated from the computer.

IV. MACHINE LEARNING ALGORITHM: LOGISTIC ALGORITHM

A. Let's take a Closer Look at this Diagram

First and foremost, as previously stated, Logistic Regression models are classification models; specifically, binary classification models (they can only be used to distinguish between two categories, such as whether a person is obese or not based on their weight, or whether a house is large or small based on its size). This suggests that our data contains two types of observations (Category 1 and Category 2), as shown in the diagram.

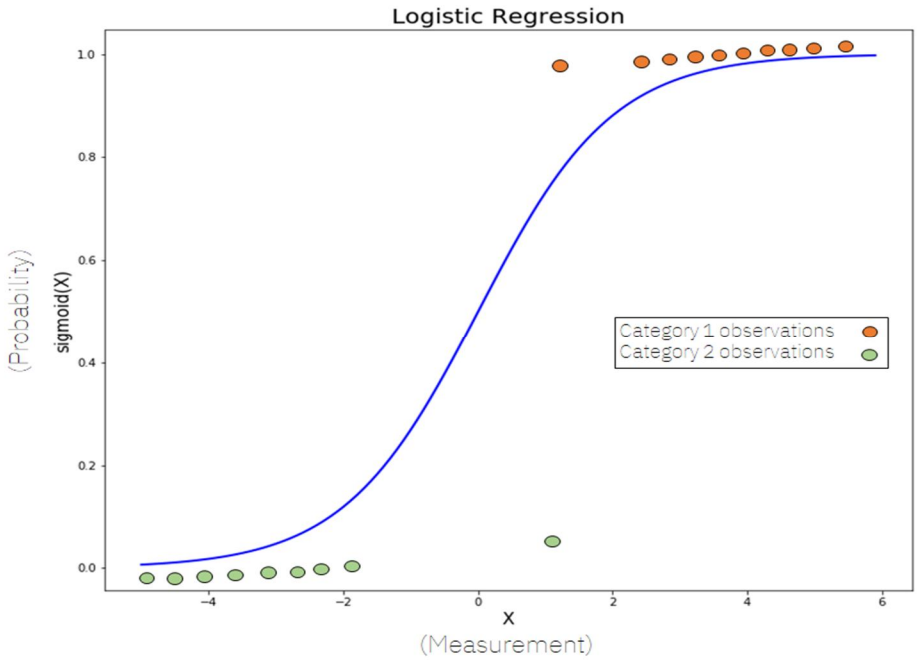


Figure 1: two observations

Second, the Y-axis goes from 0 to 1 as can be seen. This is because the sigmoid function always uses these two values as maximum and minimum, which is ideal for our purpose of categorising data into two groups. We acquire a probability (between 0 and 1 obviously) of an observation belonging to one of the two categories by computing the sigmoid function of X (that is, a weighted sum of the input features, much like in Linear Regression).

The sigmoid function has the following formula:

$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

V. IMPLEMENTION OF LOGISTIC REGRESSION ALGORITHM

During the implementation, require to import dataset or libraries. The dataset was taken manually and describes information about a Malware being detected through a malware detection tools such as Symantec, Trend micro.

We will be predicting the value of *detected ransomware attack and* consider a single feature, *threat* to predict the values of *Malware*. You can have multiple features as well.

We would import the following modules:

make_classification: available in sklearn.datasets and used to generate dataset

matplotlib.pyplot: for plotting

LogisticRegression: this is imported from sklearn.linear_model. Used for performing logistic regression

train_test_split: imported from sklearn.model_selection and used to split dataset into training and test datasets

First, Import the libraries and dataset to analyse cyberattack risk in the organization.

```
# Import required modules
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Import dataset
df = pd.read_csv('Ransom.csv')
# Show plot
sns.countplot('Security Threat', data=df)
plt.show()
```

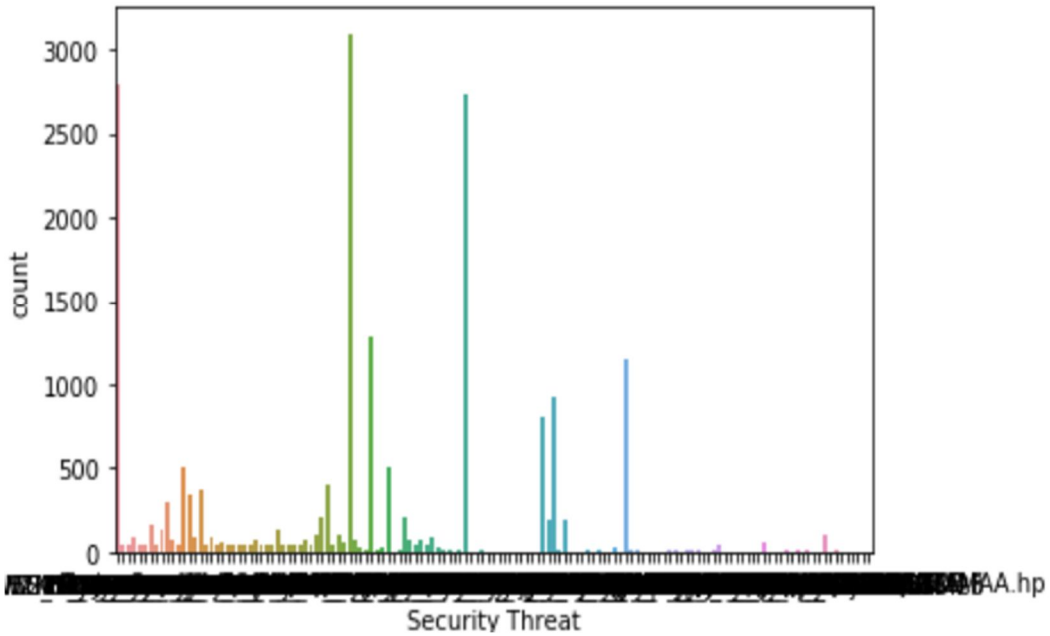


Figure 2: Dataset Graph

Import the train test split, Logistic regression and classification report.

1) Step 1: Import Packages, Functions, and Classes

```
from sklearn.model selection import train_test_split
from sklearn.linear_model import Logistic Regression
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix, accuracy_score
```

2) Step 2: Generate the dataset

Now you need to generate the dataset using the make classification () function. You need to specify the number of samples, the number of features, number of classes and other parameters.

The code for the make_classification is given below:

```
# Generate and dataset for Logistic Regression
x, y = make_classification(
    n_samples=100,
    n_features=1,
    n_classes=2,
    n_clusters_per_class=1,
    flip_y=0.03,
    n_informative=1,
    n_redundant=0,
    n_repeated=0)
```

3) Step 3: Visualize the Data

```
# Create a scatter plot
plt.scatter(x, y, c=y, cmap='rainbow')
plt.title('Scatter Plot of Logistic Regression')
plt.show()
```

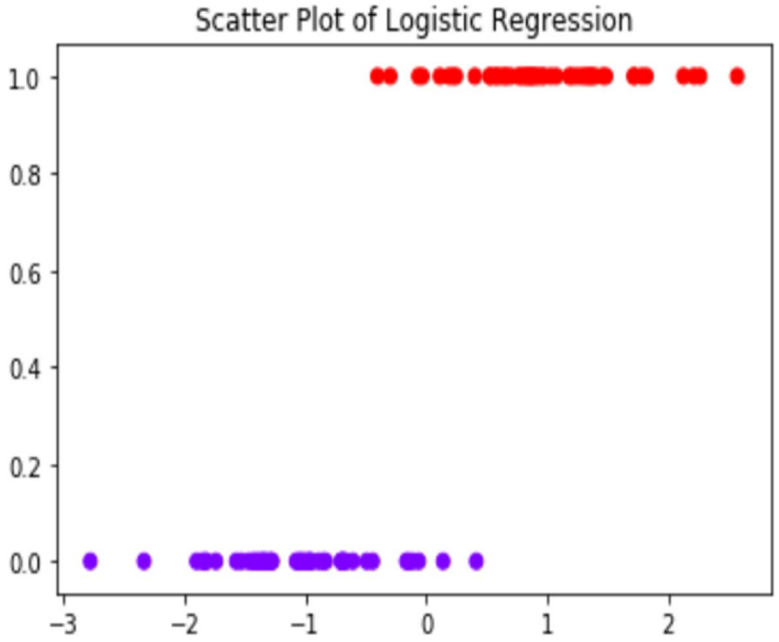


Figure 3: Plotting logistic regression

4) Step 4: Split the Dataset

Now we would split the dataset into training dataset and test dataset. The training dataset is used to train the model while the test dataset is used to test the model's performance on new data.

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.33, random_state=1)
logmodel= LogisticRegression()
```

5) Step 5: Perform Logistic Regression

Here we would create a LogistiRegression object and fit it without dataset.

```
#Step 3: Create a Model and Train It
model = LogisticRegression(solver='liblinear', random_state=0)
model.fit(x, y)
```

The logistic regression output is given below:

LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=100, multi_class='auto', n_jobs=None, penalty='l2', random_state=0, solver='liblinear', tol=0.0001, verbose=0, warm_start=False)

You can view the logistic regression coefficient and intercept using the code below:

6) Step 6: Make prediction using the model

We now use the model to predict the outputs given the test dataset.

```
In [22]: >>> model.predict(x)
Out[22]: array([0, 0, 0, 1, 1, 1, 1, 1, 1, 1])

In [23]: >>> model.score(x, y)
Out[23]: 0.9
```

score () takes the input and output as arguments and returns the ratio of the number of correct predictions to the number of observations.

7) Step 7: Display the Confusion Matrix

The accuracy of the model with a **confusion matrix**. In the case of binary classification, the confusion matrix shows the numbers of the following:

a) True negatives in the upper-left position
b) False negatives in the lower-left position
c) False positives in the upper-right position
d) True positives in the lower-right position

To create the confusion matrix, you can use confusion matrix() and provide the actual and predicted outputs as the arguments:

```
#Show the Confusion Matrix
>>> confusion_matrix(y, model.predict(x))
```

The output is:

```
array ([[3, 1],
        [0, 6]], dtype=int64)
```

We can deduce from the confusion matrix that:

```
# True positive: 3 (upper-left) – Number of positives we predicted correctly
# True negative: 6 (lower-right) – Number of negatives we predicted correctly
# False positive: 1 (top-right) – Number of positives we predicted wrongly
# False negative:  0 (lower-left) – Number of negatives we predicted wrongly
```

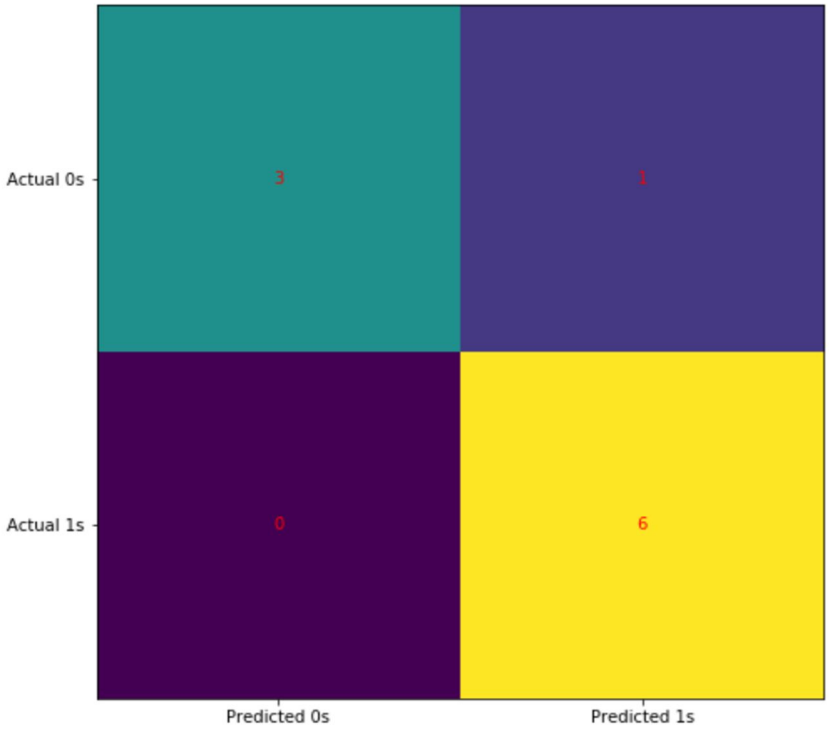


Figure 4: Shows Confusion Matrix

A. Classification Report

	precision	recall	f1-score	support
0	1.00	0.75	0.86	4
1	0.86	1.00	0.92	6
accuracy			0.90	10
macro avg	0.93	0.88	0.89	10
weighted avg	0.91	0.90	0.90	10

B. Advantages

Because of its simple and efficient nature, it does not require a lot of computing power, is simple to execute and analyse, and is extensively utilised by data analysts and scientists. It also does not necessitate feature scaling. For each observation, logistic regression generates a probability score.

C. Disadvantages

Logistic regression can't handle a large number of categorical features/variables. It's susceptible to being overfitted. Also, logistic regression cannot address a non-linear problem, which is why non-linear characteristics must be transformed. In logistic regression, independent variables that are not associated with the target variable but are very similar or correlated to each other will not perform well.

D. Observation

We analyse a lot of details about logistic regression where we build respective models, to visualize results and some of the theoretical background information. Also covered the concepts such as sigmoid function, maximum likelihood, confusion matrix, F1-Score, accuracy. It also supports multiple features. Because the input values must be in a specified format, they were reshaped before being trained with the fit approach. The accuracy using this is 90%, which is very close to the accuracy of our model.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

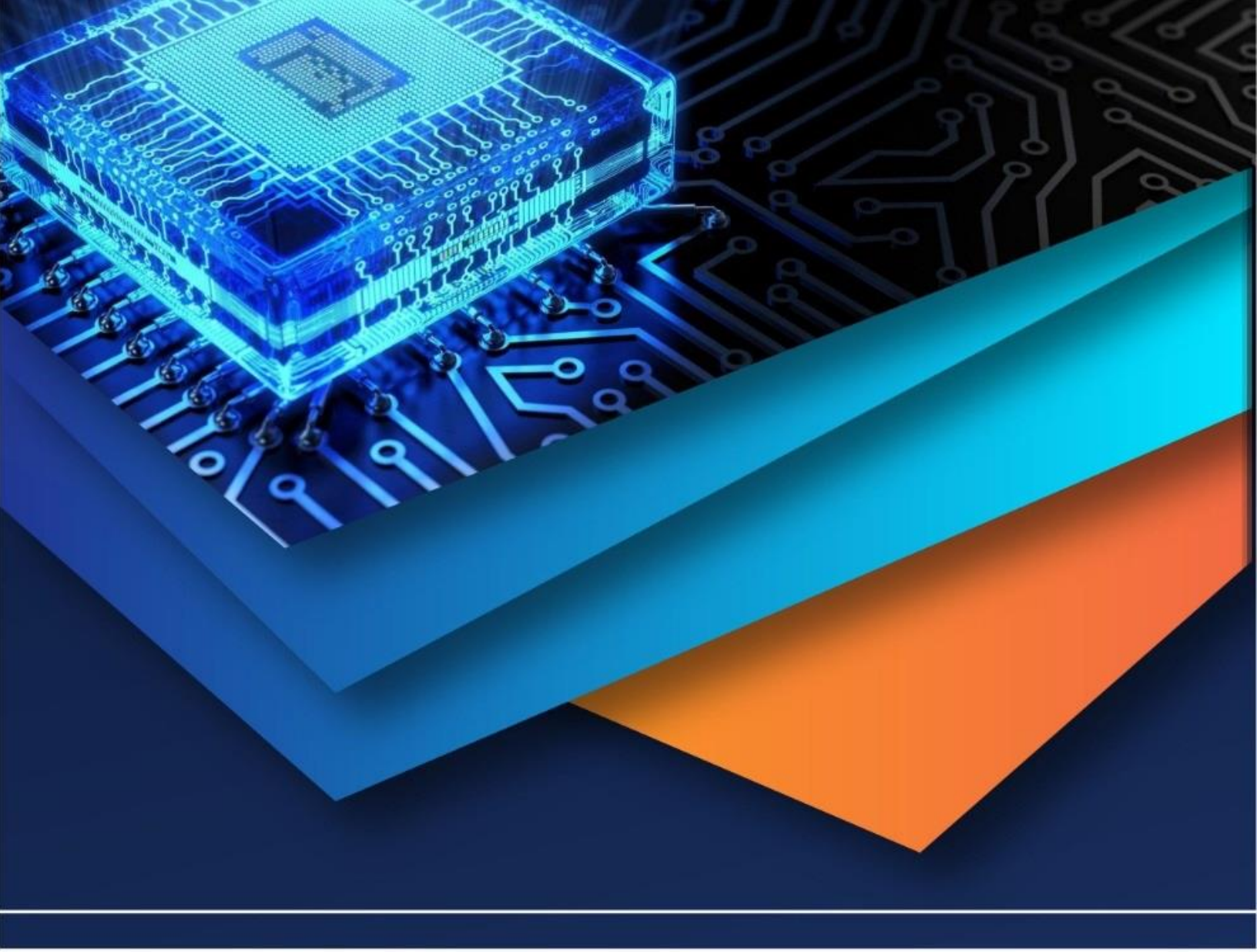
VI. CONCLUSION

We analyse the malware risk at organization with the help of occurred ransomware dataset where we found that there is lots of ransomware signature to infect the system and encrypt the data of any organization. After the analysis, we got 90% chances of cyber-attack or infect the network.

Therefore, required to implement the major security infrastructure in the organization to secure the organisation with the employee awareness program. Also, aware against the cyber-attack and mitigation procedure.

REFERENCES

[1] A Study of Ransomware Detection and Prevention at Organizations-IRJET, Saurabh Kumar Sen, Nidhi Chourey, VITM Indore (M.P)
[2] A Brief Study of Wannacry Threat: Ransomware Attack 2017- IJARCS Savita Mohrule,Manisha patil, MITACSC, Alandi, Pune, India.
[3] Fundamental of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies by John Kelleher, Brian Mac Namee, and Aoife D'Arcy.
[4] Ransomware: A Research and a personal case study of dealing with this nasty malware-Azad Ali, Indiana University of Pennsylvania.
[5] Detection of ransomware on windows operating systems-Jaan Priisalu, TALLINN UNIVERSITY OF TECHNOLOGY.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY