Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)
IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4

A Semantic Approach for Cyber Threat Prediction Using Machine Learning

Yojana Goyal
MTech Student, CSE Deptt.
SET-MUST, Lakshmangarh, Sikar, India
e-mail: yojanagoyal1995@gmail.com

Anand Sharma
Asst. Prof. , CSE Deptt.
SET-MUST, Lakshmangarh, Sikar, India
e-mail: anand_glee@yahoo.co.in

***Abstract*—Computers that are connected to the smart systems or Internet infrastructure is incited with various security threats, ranging from Computer infection to the drive-by downloads and botnets. The decent variety and measure of security threats in the system has fundamentally expanded. These threats are utilized by an underground economy for illicit exercises, for example, circulation of spam messages, refusal of-benefit assaults and robbery of MasterCard information. Using such calculations the higher level of security and respectability can be acted efficiently in the client level system framework. This issue of unprotected Computer framework can be analyzed by machine learning models. Machine learning is an order of science in which we are worried about the plan and extension of calculations that enable PCs or any machine to learn the required information, for instance, from sensor information or databases. The main aim of this research work is to utilize machine learning idea for the acknowledgment of complex examples and expectation of threats dependent on the information. Here in this paper, machine learning is applied to cyber security, with a reason to predict, identify and avert the complex cyber-security threats.**

Keywords—Cyber Security; Security Threats; Prediction; Machine Learning; Cyber Threats

I. INTRODUCTION

Artificial Intelligence (AI), and Machine Learning, has technically influenced all parts of society and business processes. Affiliations are dynamically watching the potential of machine learning in exploring the required data, arranging the information at scale, and even more suitably demonstrating various information threats.

Recently, four different ways of machine learning are associated with the prior knowledge about the threats to empower the associations, including changing unstructured substance, making officially cloud affiliations, and perceiving key terms that relate to malware, or other relevant threats.

- To understand how machine learning model is grabbing detectable quality and how it efficiently influences all the business processes.
- To learn how phenomenal man-made consciousness methodology are integrated and associated in a more agent convincing manner.
- See how insightful examination can uncover future threats.

A. Cyber Threat

For a cyber security master, the Oxford Dictionary meaning of cyber threat is a touch of coming up short on: "the likelihood of a pernicious Endeavour to harm or disturb a PC system or framework." This definition is inadequate without including the Endeavour to get to records and penetrate or take information.

In this definition, the risk is characterized as plausibility. Nonetheless, in the cyber security network, the threat is all the more firmly related to the performer or foe endeavoring to access a framework. Or then again a risk may be recognized by the harm being done, what is being stolen or the Tactics, Techniques and Procedures (TTP) being utilized.

B. Types Of Cyber Threats

In 2012, Roger A. Grimes gave this rundown, distributed in Info world, of the main five most basic cyber threats:

- Social Engineered Trojans
- Unfixed Software, (for example, Java, Adobe Reader, Flash)
- Phishing
- Network travelling worms
- Advanced Persistent Threats

Be that as it may, since the production of this rundown, there has been far reaching reception of a few distinct kinds of amusement evolving innovation: distributed computing, enormous information, and appropriation of cell phone utilization, to give some examples.

In September 2016, Bob Gurley shared video containing remarks from Rand Corporation declaration to the House Homeland Security Committee, Subcommittee on Cyber security, Infrastructure Protection and Security Technologies in regards to developing cyber threats and their suggestions. The video features two innovation drifts that are driving the cyber threat scene in 2016:

- Internet of Things – Smart objects interfacing with web of different systems

II. Related Work

Anderson HS, Roth P. Coal, Presents the open dataset with named perspectives with name EMBER so the preparation of machine learning models and prescient investigation is emerged as a possible one. The dataset is having the threats just as start malware with the explicit infiltration levels for the preparation of models and expectations. [1]

Shalaginov A, Banin S, Dehghantanha A, Franke K, Underlines the definite overview with the strategies and methodologies which can be incorporated for the malware location utilizing static assessment. The methodologies presents the profoundly solid techniques and calculations for the higher level of exactness and expectations [2].

Mahindru A, Singh P. Presents the work on malware expectation and the information disclosure for the android gadgets and related system condition with the machine learning. The methodology of dynamic authorization is incorporated with the proposed work. [3]

Milosevic N, Dehghantanha A, Chook KK. july 1, Presents the noxious expectation components with the anticipated methodology for the Android based malware so future forecasts on the versatile traffic should be possible. [4]

Buczak AL, Guven E. Oct 18, Assesses the machine learning and information revelation approaches with the joining of machine learning for malware interruption identification and the cyber security. [5]

P.V. Shijo, A salim,, Clarifies the work on the utilizations and usage of the positive and great perspective for both of static just as powerful techniques for grouping and examination of arranged malignant innovation programming. [6]

S.Alam, R. N. Horspool a d I. Tarore,"MARD, The creators in this work apply machine learning put together methodology utilizing classifier with respect to android dataset having 48919 records. The key target was to discover and remove the irregular timberland and assess the assaults expectation. [7]

M. Damshenas, A. Dehghantanha, R. Mahmoud, Takes a shot at the malware identification and examination utilizing proliferation strategies and structures. The proposed methodology is extremely compelling in terms of malware identification and expectation. [8]

David S. Anderson, Chris Fleizach, Stefan Savage, and Geoffrey M.Voelker, Foresee the conduct of malware relying on Markov chain based Graphs and found that the methodology is giving powerful outcome. In this work, the creators take a shot at a one of a kind engineering having staggered diagrams and expectation. [9]

F. Shahzad, M. Shahzad and M. Farooq, Proposed the procedure control board based usage with the goal. In this proposed methodology, the gadget based examination and forecast can be made possible. [10]

Fiore U, Palmieri F, Castiglione An and Santis AD, Take note of the fundamental issues identified with the execution of the DRBM classifier as the "haphazardness of the traffic conduct." joined with the absence of preparing information speaking to the real typical traffic. These issues to a great extent don't influence, or ought not, influence ICS systems. Confined Boltzmann Machines (RBM), or all the more explicitly, Discriminative RBM (DRBM), is explored for peculiarity discovery. The outcomes gave point to a promising methodology. They utilize a comparable non-marked methodology as is actualized by the MBM, no past data on atypical traffic is accessible. [11]

K. Allix, Q. Jerome, T.F.Bissyande, J. Kelvin,R. State and Y.L. Traon, Breaks down a lot of uses having malware for profound examination. In this methodology, the creators present a powerful model for parcels development and systematic. [12]

K, Harrison , B. Bordbar ,S.T.T. Ali ,C. I. Dalton , A. NormanV, In this work demonstrates the one of a kind and compelling design for the acknowledgment of side effects identified with noxious example when contrasted with the malware in the cloud based framework. [13]

III. Objective Of The Present Work

The Existing or Classical Methodologies of the Malware Forecast Investigation and Recognition Ought To Be Progressed. The Established Forecast technique has significantly created for achieving higher exactness and viability. Following are the examination destinations to be executed in the proposed research work so the general situations and points of view of anticipated research can be displayed.

- To play out a detailed audit on the arrange threat.
- To examine the susceptibilities and vulnerabilities with security viewpoints and to propose a methodology for the risk expectation by utilizing Machine Learning models.

IV. Methodology

Execution Strategy/Research Methodology

- Data Set / Informational index Formation from Assorted Sources

The various data sets can be fetched from the malignant dataset portal. The dataset can be brought and created utilizing machine learning devices.

- Execution of machine learning apparatuses in methodology

The methodology Researcher will use in this work will worked out for the usage of existing work.

- Features Extracting utilizing Machine learning

The features or imperative parts of the fetched dataset will be explored and afterward can be utilized for prediction.

- Profound Investigation and Predictive Analysis.
- Training and Modeling the dataset of Malicious
- A Unique approach of machine learning will be formulated and prepared from the dataset.

A. *Dataset For Malware Predictions and Analytics*

- https://archive.icu.uci.edu/ml/datasets/Detect+Malacious+Executable(AntiVirus)
- https://archive.icu.uci.edu/ml/datasets/phising+websites
- https://archive.icu.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaloT
  - And some more

Other than above connections, the datasets on malware in cell phones likewise accessible for innovative work and these will be utilized for the proposed research to prepare the model. These datasets can be utilized for preparing and testing and further prescient investigation.

In our work, the test set of system information will be taken and will be coordinated with these prepared models to discover the likelihood of malware in the explicit system condition

In the traditional methodology, there is no powerful philosophy for the assessment of infiltration dimension of the malware. In the current work, there is tremendous extent of research which can be worked out for powerful forecast examination in the malware datasets.

- There are number of open datasets accessible from arranged cyber measurable gatherings which ought to be worked out and utilizing meta heuristic methodology, these can be streamlined.
- The proposed model will be prepared dependent on the exploration datasets of arranged sources.
- The prepared dataset will be utilized to anticipate and maintain a strategic distance from the related or comparable malware in the virtual condition.

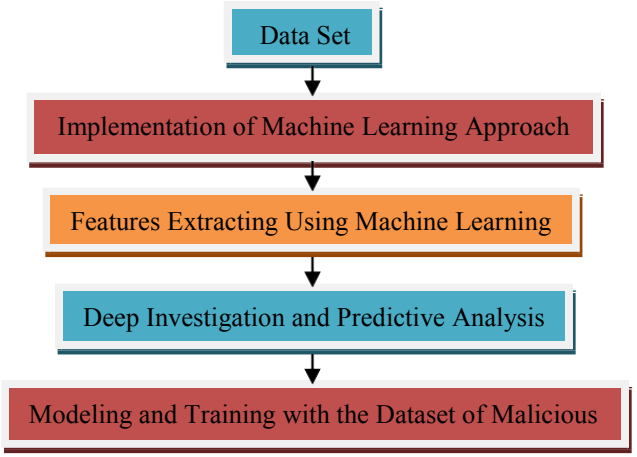

Figure 1. Flowchart

V. TOOLS FOR IMPLEMENTATION

This section will cover the various tools which are used in cyber threats prediction through machine learning.

A. *Anaconda*

Anaconda is utilized for logical processing, information science, measurable investigation, and framework examining. The ultra-present day model of Anaconda 5.0.1 is propelled in October 2017.The discharged model 5.0.1 tends to a couple of minor bugs and gives useful capacities, for example, forward R dialect help. Those abilities weren't accessible in the bona fide 5.0.0launch.This bundle bargain administrator is likewise a domain chief, a Python appropriation, and an accumulation of open supply programs and joins additional than 1000 R and Python Data Science Packages.

B. *Spyder*

Spyder is ground-breaking medicinal surroundings written in Python, for Python, and structured by method for and for researchers, architects and actualities investigators. It capacities a one of a kind blend of the unrivaled improving, investigation, troubleshooting and profiling usefulness of an entire improvement device with the actualities investigation, intuitive execution, profound assessment and wonderful perception abilities of a logical package. Besides, Spyder gives worked in combination with numerous mainstream logical applications, comprehensive of NumPy, SciPy, Pandas, IPython, QtConsole, Matplotlib, SymPy, and additional.

C. *Python*

Python is an inordinate dimension programming dialect concocted with the guide of Guido van Rossum and first discharged in 1991. It's the most acclaimed coding dialect used by programming designers to fabricate, control, control and for testing. In Python, gathered code is spared with the report augmentation .py For example new.py. It is in like manner a mediator which executes Python applications. The python translator is called python.Exe on Windows. To execute new.py program

D. SIEM

One of the greatest difficulties in cyber security is managing the amazing volume of data that originates from movement on frameworks and comprehending it so as to transform crude information into knowledge – to determine cautioning indications of assaults, comprehend the idea of deficiencies or give confirm reports to partners. In 2005, Gartner instituted the term 'security data occasion the board' (SIEM). They utilized it to portray a customary security observing framework that meets review and consistence needs. Be that as it may, as data security has developed so too have the requests of the SIEM. An expansion to streamlining your consistence detailing, you need.

- Security threat location
- Timely alarming and revealing

VI. CONCLUSION AND FUTURE SCOPE

Applying machine learning in threat prediction gives two critical accomplishments in the space of risk predication. Security threats predication and machine taking in are a long way from being "most exceedingly terrible foes". Rather, there is great expectation to make them "closest companions" sooner rather than later. Enlarging the machine with a sensibly competent human implies you're more effectively outfitted than any time in recent memory to uncover and react to threats expectation. This pinpoints the pertinent result of threats and focal points of connecting the machine learning goes for encouraging fascinating security research to risk expectation. Analyst will proceed with the work in future on use of machine learning and recognition expectation to robotized and semi-mechanized location of threats.

REFERENCES

[1] Anderson HS, Roth P. EMBER:"An Open Dataset for Training Static PE Malware Machine Learning Models". arXiv preprint arXiv: pp 1804.04637,April 2018.

[2] Shalaginov A,Banin S,Dehghantanha A, Franke K. "Machine learning aided static malware analysis:A survey and tutorial Cyber Threat Intelligence".pp7-45,2018.

[3] Mahindru A, Singh P. Dynamic permissions based Android malware detection using machine learning techniques.In Proceedings of the 10th innovations in Software Engineering Conference.pp. 202-210. ACM,Feb 5 2017.

[4] Milosevic N, Dehghantanha A, Chook KK. Machine Learning aided android malware classification; 61: pp 266-74 Computers & Electrical Engineering, jul 1 2017.

[5] Buczak AL, Guven E."A survey of data mining and machine learning methods for cyber Security intrusion detection". IEEE Communications Surveys and Tutorials.(2): pp 1153-76,Oct;18 2016.

[6] P.V. Shijo, A salim."Integrated Static and Dynamic Analysis for Malware Detection",International Conference on Information and Communication Technologies, Kochi India, pp.804-811, December 2015.

[7] S.Alam, R. N. Horspool a d I. Tarore,"MARD: "AFRAMEWORK for metamorphic malware analysis and real-time detection", 28th IEEE International Conference on Advanced Information Networking and Application (AINA), Victoria, pp.480-489,May 2014.

[8] M. Damshenas, A. Dehghantanha, R. Mahmoud,"A Survey on Malware Propagation, Analysis, and Detection", International Journal of Cyber Security and Digital Forensics (IJCSDF), Vol. 2 , Issue 4,pp. 10-29, 2013.

[9] David S. Anderson, Chris Fleizach, Stefan Savage, and Geoffrey M.Voelker.Spam scatter:" Characterizing Internet scam hosting infrastructure".In Proceedings of the Sixteenth USENIX Security Symposium,2007.

[10] F. Shahzad, M. Shahzad and M. Farooq,"In-execution Dynamic Malware Analysis and Detection by Mining Information in Process Control Blocks of Linux Operating System". Data Mining for information Security, Vol. 231, pp.45-63,2013.

[11] Fiore U, Palmieri F, Castiglione A & Santis AD"Network anomaly detection with the restricted boltzmann machine. Neuro computing 122(0):.Advances in cognitive and ubiquitous computing.pp 13 – 23,2013.

[12] K. Allix, Q. Jerome, T.F.Bissyande, J. Kelvin,R. State and Y.L. Traon,"A Forensic Analysis of Android Malware—How is Malware Written and How it Could Be Detected".38th IEEE Annual Computer Software and Applications Conference (COMPSAC), Vasteras, pp. 384-393,July 2014.

[13] K, Harrison , B. Bordbar ,S.T.T. Ali ,C. I. Dalton , A. NormanV, "A framework for detecting malware in cloud by identifying symptoms", 16th IEEE International Enterprise Distributed Object Computing Conference (EDOC), Beijing". pp. 164-172,September 2012.