IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, May 2022
Impact Factor: 6.252

Investigation and Classification of Cyber Crime using Deep Learning Approach

Tai. S. Pawar[1], Dr. M. D. Rokade[2],
PG Student, Department of Computer Engineering [1]
Asst. Professor, Department of Computer Engineering[2]
Sharadchandra Pawar College of Engineering, Dumbrewadi, Otur, Maharashtra, India

Abstract: *An intrusion detection system (IDS) is software that monitors a single or a network of computers for hostile behaviours such as data theft, censorship, or network protocol corruption. The majority of intrusion detection techniques used today is incapable of dealing with the dynamic and complicated nature of cyber-attacks on computer networks. Despite the fact that effective adaptive methods, such as various Deep learning algorithms, can result in higher detection rates, lower false alarm rates, and cheaper computing and communication costs. Data mining can result in frequent pattern mining, classification, clustering, and micro data streams when used correctly. This proposal proposes an enhanced technique for intrusion detection based on data mining and deep learning. The two types of intrusion detection systems are host-based IDS and network-based IDS. Network based IDS is utilised in this proposal to safeguard the computer network and its resources from harmful attacks. Papers representing each approach were located, reviewed, and summarised based on the number of citations or the relevance of a developing method. Well-known cyber data sets are employed in Deep learning and data mining because data is so vital in these approaches.*

Keywords: Cybercrime, cyber-attacks, Deep Learning, Data Mining, and Intrusion Detection Systems

I. INTRODUCTION

All countries in the globe are connected by computer networks, and bad actors can wreak significant harm anywhere in the world without ever leaving their workstation. Individuals not being able to access their personal computers for a few hours or coming across racist or obscene material on the Internet, companies' internal networks being inaccessible for 24 hours or trade secrets being stolen, and the government's public websites being blocked or seeing state secrets appear on the web are all examples of potential harm. Financial losses can range from extortion of a few hundred dollars to multi-million dollar losses as a result of cyber fraud or cyber sabotage. Cross-border crime is exemplified by cybercrime. External intrusions from outside the company, as well as internal invasions, are examples of security breaches. Recommendation The methods of Deep Learning and Data Mining are discussed, as well as many applications of each method to cyber intrusion detection problems. The proposal discusses the complexity of various Deep learning and data mining algorithms, and it includes a set of comparison criteria for Deep learning and data mining methods, as well as recommendations for the best ways to apply based on the features of the cyberspace. a problem to be resolved Cyber security refers to a combination of technologies and procedures for defending computers, networks, programmes, and data from assault, unauthorised access, modification, or destruction.

Misuse-based, anomaly-based, and hybrid cyber analytics are the three main forms of cyber analytics used to support intrusion detection systems. Misuse-based approaches are intended to detect known attacks by analysing their signatures. They are effective at identifying known types of attacks while producing a low number of false alarms. They necessitate manual database changes with rules and signatures on a regular basis. Misuse-based detection approaches are incapable of detecting novel assaults. Anomaly-based techniques identify anomalies as deviations from normal behaviour by modelling normal network and system activity. Their capacity to identify zero-day assaults makes them appealing. Another benefit is that typical activity profiles are customised for each system, application, or network, making it harder for attackers to figure out which activities they can engage in. Anomaly-based techniques can also be used to define the signatures for abuse detectors using the data that they detect. Because previously undiscovered system actions may be classified as anomalies,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, May 2022
Impact Factor: 6.252

the fundamental downside of anomaly-based approaches is the possibility for high false alarm rates. As the use of Internet services grows, so do the hazards posed by the internet to computer systems and data.

Attackers can simply gain access to our systems' critical data resources. It is critical to protect data from such attackers because they can use and sell the information for their own personal gain, or it could fall into the wrong hands. Large amounts of data are saved on company servers and PCs. As a result, it is critical to ensure that the vital data is kept safe and secure. This can be accomplished with the help of a real-time intrusion detection system that identifies any unusual behaviour and alerts the user. Prevents assaults by sending an alert to the administrator. This can be accomplished by employing a variety of Deep Learning and Data Mining techniques.

The researcher believes that cyberspace's growing dominance and wide-ranging breadth, as well as cybercrime, represent a major breach of international law, and that courts will have jurisdictional issues in addressing cybercrime.

II. Need for Research

As the use of Internet services grows, the hazards posed by the internet to computer systems and data grow as well. Attackers can simply gain access to our systems' critical data resources. It is critical to protect data from such attackers because they can use and sell the information for their own personal gain, or it could fall into the wrong hands. Large amounts of data are saved on company servers and PCs. As a result, it is critical to ensure that the vital data is kept safe and secure. This can be accomplished with the use of a real-time intrusion detection system, which identifies and alerts to any unusual behaviour. These cyber-attacks have the potential to steal your information and corrupt your system. After gaining access to a specific PC through a cyber-attack, cyber attackers can use the internet to mine data. or some other criminal activity, such as crypto currency mining. This will have an impact on the entire network, and the business will ultimately suffer a significant loss of data or money.

III. Proposed System

The application files were created with Python 3.6. It is necessary to ensure that Python 3.6 and the following libraries are installed before running the files. Learn to Sklearn: - Library of Machine Learning Numpy (laughing): - Mathematical Procedures Pandas are a type of panda. - Tools for Data Analysis Mat -plotlib is a library for plotting data. - Visualization and graphics The implementation phase is divided into five steps: 2- Statistics 1- Pre-processing 3- Filtering of Attacks 4- Feature Picking Implementation of 5-machine Learning One or more Python files are included in each of these steps. Both the "py" and "ipynb" extensions were used to save the same file. They both have the same code in them. The ipynb file has the benefit of storing the state of the last run of that file as well as the screen output.

IV. Python

Python is a high-level, general-purpose programming language with an interpreter. Python has a design philosophy that prioritises code readability, which includes a lot of whitespace. It has structures that allow for clear programming at both small and large sizes. Until July 2018, Van Rossum was the language community's leader. Python is garbage-collected and dynamically typed. Procedural, object-oriented, and functional programming are among the programming paradigms supported. Python comes with a large standard library and is known as "batteries included."

For a wide range of operating systems, Python interpreters are available. CPython, the standard Python implementation, is open-source software with a community-driven development strategy. The Python Software Foundation, a non-profit organization, oversees Python and CPython.

V. Support Vector Machine

The Support Vector Machine (SVM) technique is a simple yet effective Supervised Machine Learning algorithm that may be used to create both regression and classification models. Both linearly separable and non-linearly separable datasets can benefit from the SVM method. Even with a small amount of data, the support vector machine method performs admirably. An easy way to comply with the Journal paper formatting requirements is to use this document as a template and simply type your text into it.

Step 1: Consider the situation depicted in Figure 1 to determine the best hyper plane for separating the two classes. We strive to maximise the distance between the hyper plane and the closest data point in SVM.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, May 2022
Impact Factor: 6.252

Step 2: All decision boundaries are split classes in this case.
Step 3: Data is not uniformly distributed on the left and right in this case.
Step 4: When choosing a hyper plane, SVM will disregard the data point and choose the hyperplane with the best performance.
Step 5: Linear classifiers are highlighted in this case, and data can be segregated using any straight line.

VI. PROCEDURE METHODOLOGY

1. Data Set: There are various steps involved in the data preparation process. Different data preparations have different changes in the steps provided.
2. Pre-process and visualise: Data pre-processing is a data mining approach that entails converting raw data into a format that can be understood. Cleaning, selection, normalising, transformation feature extraction, and selection are all examples of data preparation.
3. SVM-based Training Model: Machine learning entails predicting and classifying data, as well as employing various machine learning methods based on the data set.
4. User Input: The raw data that is processed is referred to as input.
5. Save Model: Using sklearn, save and load a user machine learning model in Python. This enables you to save the user model to a file and then load it later to perform predation.
6. Predict: - When an assault is discovered, it is checked.

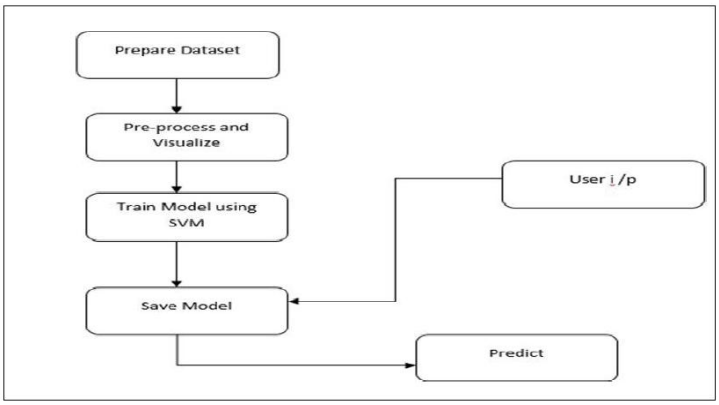

Figure: System Architecture Block Diagram

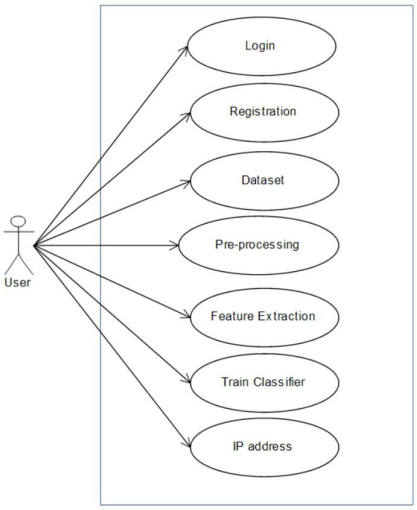

Figure: UML Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, May 2022
Impact Factor: 6.252

VI. CONCLUSION

In the suggested system, we presented a system that utilised a dataset comprising information about cyber-attacks in businesses. Following that, certain Machine Learning techniques will be used to do pre-processing and standardisation on these datasets. In comparison to previous methodologies, the proposed system can provide the highest level of accuracy. This project's functionalities can be grown significantly in the future. These features could include: Crime data analysis in real time: Using real-time datasets, we may be able to obtain crime patterns and forecasts in the future. Data mining social media to generate datasets, which are then pre-processed and analysed to identify trends in the present criminal scenario in a certain location or region. Compare and display the outcomes of all relevant and available tests.

REFERENCES

[1] Songnian Li, Suzana Dragicevic ,Francesc Anton Castro , Monika ester, Stephan ´ Winter, Arzu Coltekin, Christopher Pettit, ”Geospatial big data handling theory and methods: A review and research challenges”.
[2] Deepak A Vidhate, Parag Kulkarni, 2019, “Performance comparison of multiagent cooperative reinforcement learning algorithms for dynamic decision making in retail shop application”, International Journal of Computational Systems Engineering, Inderscience Publishers (IEL), Volume 5, Issue 3, pp 169-178.
[3] Yang C, Goodchild M, Huang Q, Nebert D, Raskin R, ”Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?”, International Journal of Digital Earth, pp. 305-329,Vol. 4, No. 4, July 2011.
[4] Deepak A Vidhate, Parag Kulkarni, 2019, “A Framework for Dynamic Decision Making by Multi-agent Cooperative Fault Pair Algorithm (MCFPA) in Retail Shop Application”, Information and Communication Technology for Intelligent Systems, Springer, pp 693-703.
[5] Duffy DQ, Schnase JL, Thompson JH, Freeman SM, Clune TL, “Preliminary Evaluation of Map Reduce for High-Performance Climate Data Analysis”, NASA new technology report white paper, 2012.
[6] Deepak A Vidhate, Parag Kulkarni, 2018,” A Novel Approach by Cooperative Multiagent Fault Pair Learning (CMFPL)”, Communications in Computer and Information Science, Springer, Singapore, Volume 905, pp 352-361.
[7] Gema Bello-Orgaza, Jason J. Jungb, David Camacho, ”Social big data: Recent achievements and new challenges”, Journal of Information Fusion, Science Direct, pp. 45– 59,Volume 28, March 2016.