Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(RESEARCH ARTICLE)

Check for updates

Machine learning enabled system for intelligent classification of host-based intrusion severity

Anthony Effiong Edet * and Godwin Okon Ansa

Department of Computer Science, Akwa Ibom State University, Mkpat Enin, Nigeria.

Global Journal of Engineering and Technology Advances, 2023, 16(03), 041–050

Publication history: Received on 13 July 2023; revised on 01 September 2023; accepted on 04 September 2023

Article DOI: https://doi.org/10.30574/gjeta.2023.16.3.0171

Abstract

Intrusion severity classification or the analysis of the impact of intrusion is a much needed solution to effectively manage intrusion events in an organization. A lot of intrusion scenarios have been carried out by systems administrators or the internal workers over the years in different organizations and the external hackers are berated for it. Many deliberate inversions have happened from the internal actors with top management board members only swinging into actions to manage the effect of it without digging into the inversion to apprehend the actors or the source of the intrusion. So, this work has been designed to assist IT firms to effectively carry out the analysis of the impact of intrusion, especially those from the internal workers. In this work, we proposed a Machine Learning Enabled System for Intelligent Classification of Host-based Intrusion Severity. The proposed model is aimed at detecting the severity of intrusion problems, carryout source analysis and give security recommendation for effective management of intrusion problems. The model is divided into three phases; the detection of intrusion severity, source analysis and security recommendation using counterfacatual reasoning.We built a system that aided us to gather user interaction over time, we captured these interaction in the activity log, our dataset was extracted from these activity log data.We used Bayesian Network to design the intrusion severity classification system, source analysis is carried out immediately, then counterfactual model is employed to give security recommendation. The accuracy of Bayesian Network in the intrusion severity classification model is 82%. An API was generated and deployed to allow scalability.

Keywords: Intrusion; Severity; Bayesian Network; Causal Reasoning; Privilege abuse; Internal and External intrusion

1. Introduction

Intrusion is defined as a security incident that consists of one or more security incidents in which an unauthorized person has access to, or attempts to gain access to, a system or system resource [1]. An illegal entrance into your company's software or computer system is referred to as a host or software intrusion [2]. In this work, we categorize the degree of host-based intrusion, pinpoint the intrusion's origin, and then provide security advice on how to handle the problem. That is, our attention is on privilege abuse since insider actors frequently misuse their access to the system, which constitutes intrusion [3]. Most businesses experience ongoing attacks. The majority of businesses are routinely targeted by outside users coming from the internet space [2]. The majority of data leaks are caused by employees abusing their power. Although employee sabotage of one's own firm does occur occasionally and on occasion with deliberate intent, most of the time it is entirely unintentional [4]. The main goal of cybercriminals is to obtain an employee's or administrator's login information so they may move through the system with full access to everything. Employee cybersecurity training is now more important than ever [5]. Knowing how an incursion affects the system and the damage it causes is one thing; understanding that both internal and external intrusion are equally destructive is quite another[6],[7]. However, because it is carried out by authorized system users, intrusion from an internal source is more dangerous than one from an external source [8]. The reputation of the company and profits could suffer long-lasting and more severe harm if an employee decides to deface the company's website or leak trade secrets to a rival

* Corresponding author: Anthony Effiong Edet

[9]. This makes intrusion from an internal source potentially more dangerous than from an external source. External intruders frequently seek out information they can sell or use for financial gain, so external hacks may be financially more damaging if they compromise your software, conceal important data, and then demand payment in exchange for the information's release. The worst part of internal intrusions is that, when the news breaks, system administrators often beret black hat hackers because they abuse their access to information and privileges to cause major issues. Organizational administrators frequently sell their systems' login information to outside hackers they collaborate with in cybercrimes, who then use the information to harm the target [10]. Over the years, research has proven that this is possible, but there is still a serious contention on how to rightly identify an intrusion that is externally perpetrated and the one that is from the insiders [4]. Internal intrusion can happen in two forms, by mistake and willfully abusing privileges [10]. However, it happens, the need to identify with clear cut evidence that it is internally precipitated is highly necessary[11]. However, we are not detecting network intrusion in this work. Our system is not designed to monitor or detect any cybercrime or network intrusion, the system is rather designed to classify the severity of host-based intrusion. Even so, we are putting up a classification scheme, one that distinguishes between the seriousness of user behavior from the system activity record. In order to see how power abuse occurs in practice and to collect data from the system's activity log to generate the dataset for the primary classification system, we first need to design a system that grants users privileges and tracks their interactions with it. To model the intrusion severity classification system, we employ a Bayesian network. This is an example of how artificial intelligence is typically used to solve computer security issues [12]. Bayesian Network is a model that shows causal relationship[13], hence, it is compatible with Causal reasoning technique such as the Counterfactual model.

2. Methodology

Our methodology is divided into two parts, we would use the principle of Bayesian Network to build the first part which would cover intrusion scenario and source detection. The second part would use the principle of Counterfactual Reasoning model; which would focus on proffering counterfactual solutions to the intrusion scenario in part one.

Bayesian Network as a machine learning model is based on the principles of Bayes theorem. To establish the proposed model, it is pertinent to state what Bayes theorem says since Bayesian Network depends on its foundation.

Bayes' Theorem states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.

The main objective of the method is to model the posterior conditional probability distribution of outcome (often causal) variable(s) after observing new evidence. A posterior probability, in Bayesian statistics, is the revised or updated probability of an event occurring after taking into consideration new information. In statistical terms, the posterior probability is the probability of event 'A' occurring given that event 'B' has occurred. The formula to calculate a posterior probability of A occurring given that B occurred:

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)\times P(B|A)}{P(B)},$$

Where:
A, B are Events P(B|A) = The probability of B occurring given that A is true.
P(A) and P(B) = The probabilities of event A occurring and event B occurring independently of each other.

The posterior probability is thus the resultant distribution, P(A|B).Prior probability represents what is originally believed before new evidence is introduced, and posterior probability takes this new information into account. Posterior probability distributions should be a better reflection of the underlying truth of a data generating process than the prior probability since the posterior included more information. We specifically apply this principle to the problem at hand, given that θi is the data sample.

2.1. Model formation

2.1.1. Modeling What Constitute Intrusion

In this work, we are to use the dataset of intrusion to specify and categorize intrusion, but in order to match an intrusion scenario to its source, we have to build a model that classifies intrusion and its source. Given the data set θi , class intrusion (YES) is denoted by YES. hence, the posterior probability of an activity being suspicious is given by P(YES|θi). The posterior probability of an activity not being suspicious is also given by P(NO|θi). Therefore, for us to say a given data tuple from activity log of any host in the organization is intrusion it must obey the following model

$$P(YES|\theta i) = \frac{P(\theta i|YES).P(YES)}{P(\theta i)}.$$

Also for an activity to be certified as not intrusion it must follow $P(NO|\theta i) = 1\text{-}P(YES|\theta i)$. Where evidence that the activity or data tuple is intrusion is given by: $P(\theta i) = P(\theta i|YES).P(YES)+P(\theta i|NO).P(NO)$

2.1.2. Modeling The Source of Intrusion

Let the joint probability distribution of the conditional variables be P(x1,x2,x3,……,xn) where Xi is the data features in the data set. Given that we have two classes namely, Internal and External classes denoted by X and β respectively, with the data sample denoted by θi, we design our conditional probability as thus

$$P(X|\theta i) = \frac{P(\theta i|¥).P(¥)}{P(\theta i)}.$$

Where $P(X|\theta i)$ is posterior, $P(\theta i|X)$ is likelihood of the intrusion being from the internal source and P(X) is the prior probability of the intrusion being from the internal source and the denominator $P(\theta i)$ being evidence can be expressed as $P(\theta i) = P(\theta i|X).P(X)+P(\theta i|\beta).P(\beta)$, Hence, externally caused intrusion would be $P(\beta|\theta i) = 1\text{-}P(X|\theta i)$

2.1.3. Final Combined Model for Intrusion and Its Source

It has been established that an intrusion is given by $P(YES|\theta i)$, if it is from the external source, it is given by $P(\theta i|\beta)$, while internal intrusion is given by $P(\theta i|X)$. Therefore, to model intrusion and its source (internal first), we combine these equations :

$$P(YES.X|\theta i) = P(\theta i|X)+P(\theta i|YES),$$

$$\text{Meaning that } P(\theta i|YES.X) = \frac{P(\theta i|YES).P(X)}{P(\theta i)}+\frac{(P(\theta i|YES).P(YES)}{P(\theta i)}.$$

Therefore, an intrusion from external source is given by $P(\theta i|YES.\beta) = 1 - P(\theta i|YES.X)$.

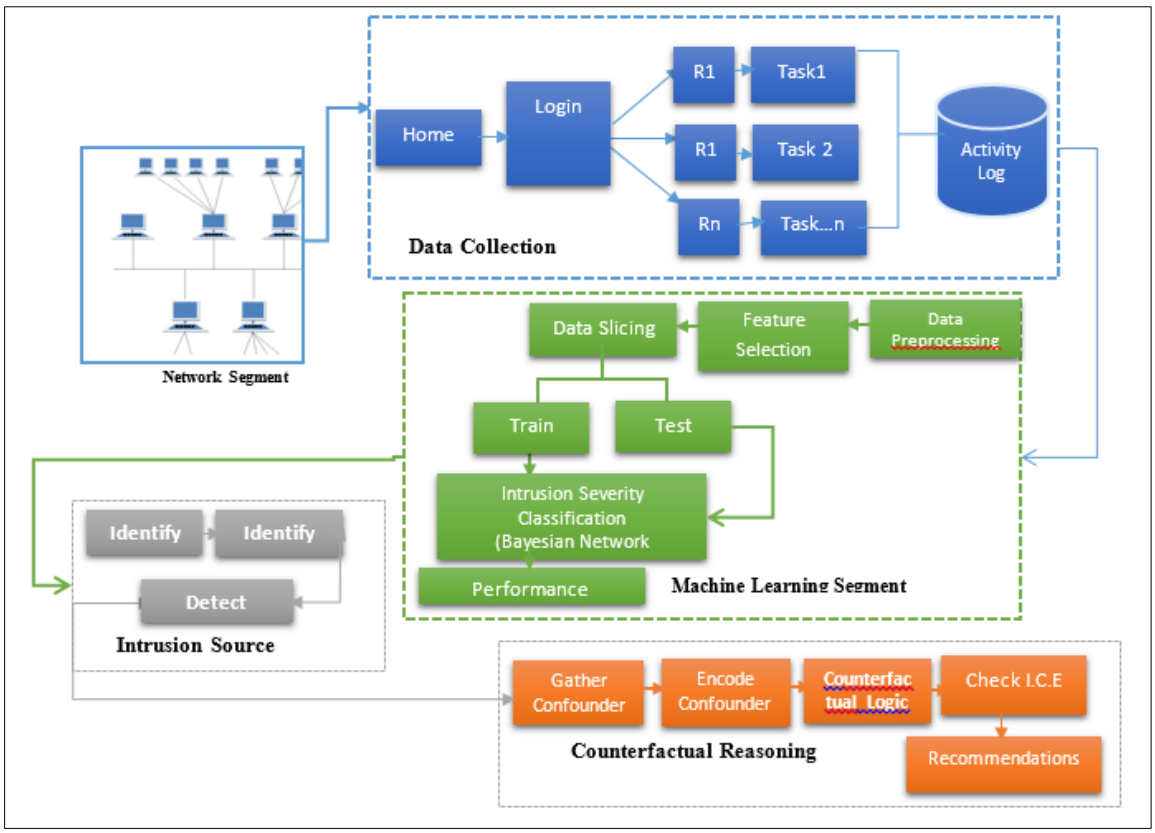

Figure 1 Conceptual framework of the System

2.1.4. Intrusion Severity Model

We have multiple classes for severity classification. The classes are Low, Medium and High. Given the data set θi , the posterior probability of intrusion severity being Low is given by P(Low|θi). The posterior probability of intrusion being Medium is given by P(Medium|θi). The posterior probability of intrusion severity being High is given by P(High|θi) Therefore, for us to say a given data tuple from activity log of any host in the organization is intrusion, its severity index can be determined as follows;

$$P(Low|\theta i) = \frac{P(\theta i|Low).P(Low)}{P(\theta i)} \text{.....(A)}$$

$$P(Medium|\theta i) = \frac{P(\theta i|Medium).P(Medium)}{P(\theta i)} \text{.....(B)}$$

$$P(High|\theta i) = \frac{P(\theta i|High).P(High)}{P(\theta i)} \text{.....(C)}$$

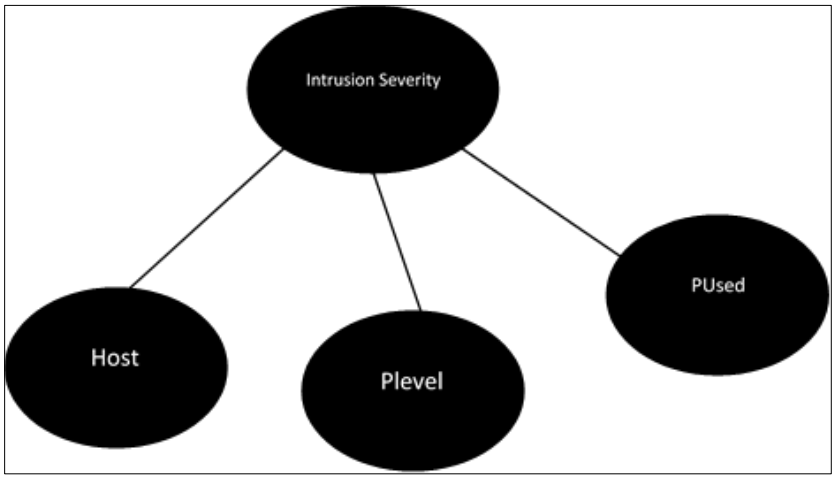

Figure 2 Direct Acyclic Graph of Conditional Variables

Direct acyclic graph here shows the conditional variables in our Bayesian Nework model for intrusion severity classification.

2.2. Algorithmic Conditions For Counterfactual Inference

- **Condition 1:** Proportionality - The likelihood that an intrusion is caused by a particular source should be proportional to the posterior likelihood of that intrusion M(Evidence, Intrusion. Source) ∝ P(Intrusion = T|E). The likelihood that a source explains the host's log is proportional to the likelihood that the host is intruded by the source. This means that, for intrusion to be from the internal actors such as the employees P(θi|YES.X)∝P(θi|YES) and for an external intrusion P(θi|YES.β)∝P(θi|YES). This type describes the direct relationship between two quantities. In simple words, if one quantity increases, the other quantity also increases and vice-versa.
- **Condition 2:** No Causal Evidence - An intrusion attack A that cannot cause any of the host's indications/symptoms or evidence can not constitute a counterfactual outcome or diagnostic measure, M(intrusion, θi) = 0 (causality). If there is no causal evidence that a user has abused a privilege, then the intrusion cannot constitute causal explanation and should be disregarded.
- **Condition 3:** - Simplicity - intrusion that has a greater number of the host's log evidence should be more likely-(simplicity).

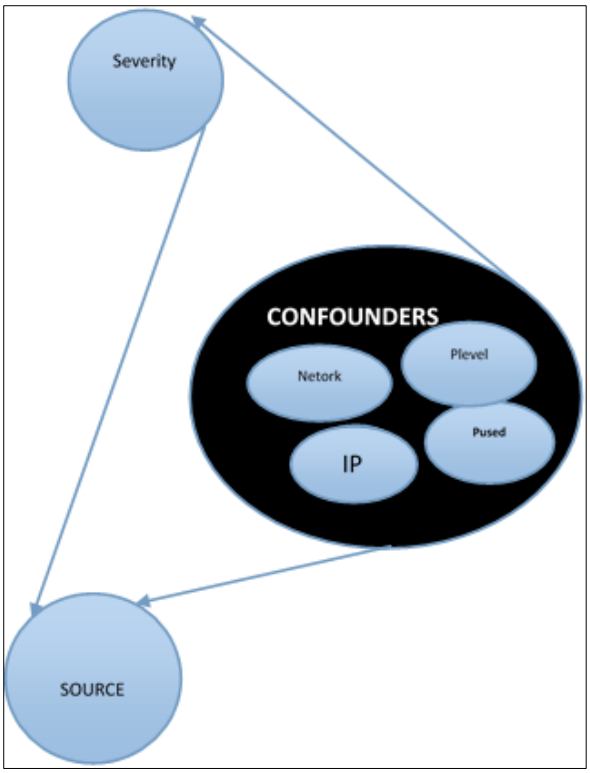

Figure 3 Counterfactual confounders

The image above shows confounders (evidence) for counterfactual inference leading to security recommendation.

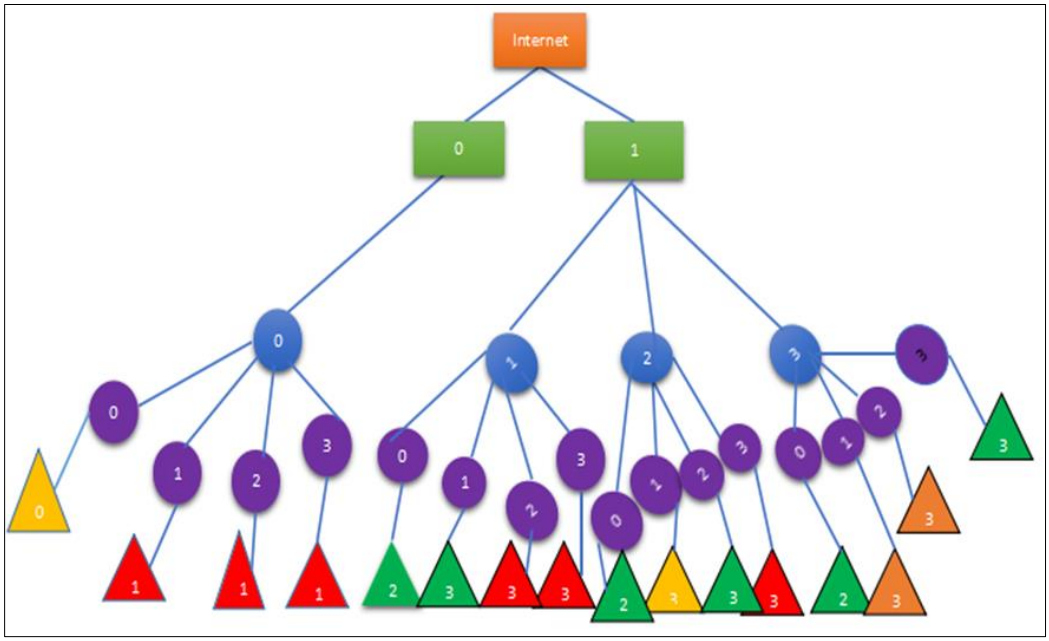

Figure 4 Counterfactual Inferential Network Tree Diagram

Fig. 4 illustrates a counterfactual inferential network tree used for deriving security recommendations based on user interactions recorded in an activity log. The tree consists of different levels, including the INTERNET level where everyone has access to the system, IP address groups represented by 0 and 1, privilege assigned (Plevel) represented by levels numbered 0-3 (blue buttons), privilege used (Pused) represented by a purple-colored level, and individual counterfactual effect (ICE).Based on this tree network, a table and corresponding recommendations were derived, which capture specific conditions in causal reasoning.

2.2.1. Systems Actors Privileges

Privilege is a legitimate authority or authorization one or an actor has to access, read and write to a file. It is what defines the boundary of actions of system actors. In this work, we are classifying the severity of intrusion. This intrusion is the one happening at the host, that is, software level. From there, we are moving to determine if the intrusion was carried out by an insider or outsider for effective recommendation on what to do to mitigate future occurrence. It is in this view that we have to design a system that we could use to monitor the activity of users (Internal and external(visitors)), more of this will come later in this chapter. In this work, we have four levels of privileges ranging from 0-4. We have assigned privilege zero (0) to visitors. Privilege one (1) is for normal users who are part of the insider team. These set of people may include the Industrial Attachment students, and other junior staff who may wish to access the system to carryout their specific duties. We have level two (2), who are the systems Administrators. These people have access to limited functions and are monitored and controlled by those who are in level three(3) known as the Directors. However, the Administrators control and monitors the junior staff and students. At level three(3), we have the Directors. These set of actors have unrestricted access to every part of the system. They are fully in charge of the administration. They have access to all systems functions and can login to the system to act as both the normal users, and administrators. In this work, we have assigned specific functions to each actors. At level zero (0), every actor starts from there. It is the landing page of the software. Every actor visits there to login to use their assigned privileges. So, at level zero, every user is seen as a visitor until they login to their respective profiles using their correct systems credentials.

Table 1 Privilege Encoding Matrix

		Privilege Used			
		Level 0	**Level 1**	**Level 2**	**Level 3**
Privilege Assigned	Level 0				
	Level 1				
	Level 2				
	Level 3				

Legend: Explains table 1

Low	1
Medium	2
High	3

Green color signifies a state where there is no security issues, the orange color indicates a mild security concerns calling on the appropriate authorities to do a quick security checks, while the red color signifies a serious security breaches that must be swiftly investigated and addressed

Table 2 Intrusion Severity Threshold

		Privilege Used			
		Level 0	**Level 1**	**Level 2**	**Level 3**
Privilege Assigned	Level 0	2	3	3	3
	Level 1	1	1	3	3
	Level 2	1	2	1	3
	Level 3	1	2	2	1

Legend: Explains table 2

Low	1
Medium	2
High	3

In table 2, adding the to explanation in Table 1, the numbers in the table stand for different intrusion severity thresholds. Number one (1) means there is no problem, number two (2) means there is need for system assessment while three (3) calls for a serious security checks.

3. Results and evaluation

Each condition is mapped to a recommendation which is determined by putting several factors into consideration as we have explained above.

The plots below show the behaviour of our model on the dataset. The overall performance of the model is summarized in the visualization below;

Table 3 Counterfactual Inference/Security recommendation

S/n	Observation	Recommendation	Condition
1	(1,2,1),(1,3,1), (1,3,2)	There was no privilege abuse, however, an internal system user with higher Role performed the function described in the role of an internal system user with lower Role.	Condition 1 holds: Proportionality
2	(1,3,3),(1,3,0),(1,2,2), (1,2,0),(1,1,1),(0,0,0)	There was no privilege abuse, hence, there is no causal evidence or explanation	Condition 2 holds: No Causal Evidence
3	(0,0,1),(0,0,2),(0,0,3), (1,1,2),(1,1,3),(1,2,3)	There was intrusion, hence, the system credentials should be reviewed	Condition 3 holds: - Simplicity

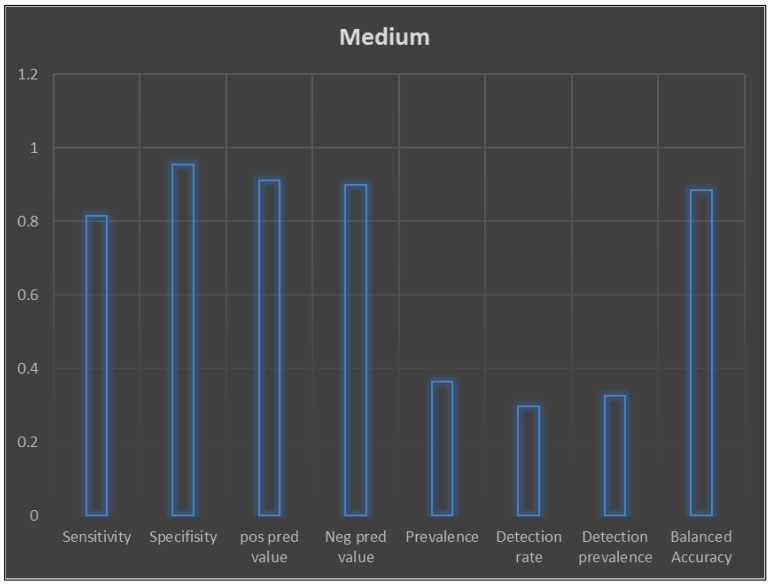

Figure 5 Overall Performance of model on severity class Medium

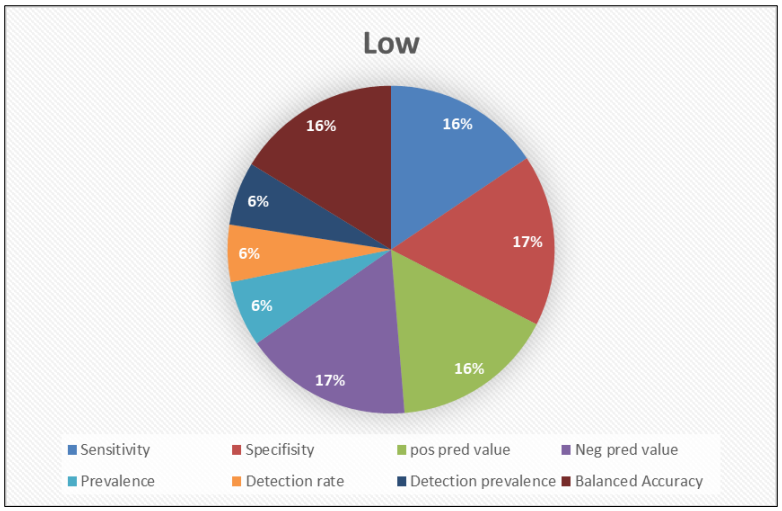

Figure 6 Overall model performance on Severity Class LOW

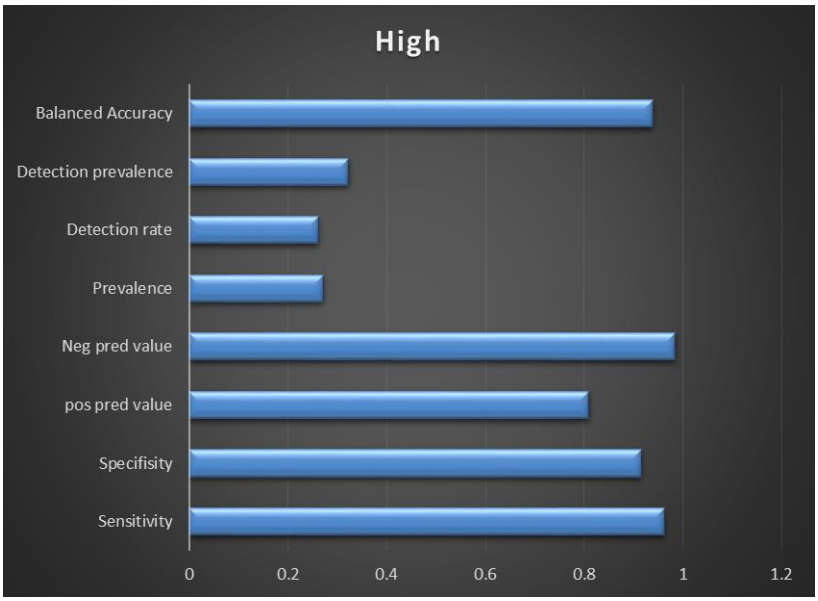

Figure 7 Overall model performance on severity class High

Figure 8 All model characteristics and class performance

4. Conclusion

Accordingly, as we stated in our problem statement, not all intrusion inversions are from the external source. Some inversions are perpetrated by the administrators and other internal users of the system. This can happen by mistake sometimes and can also be a deliberate act. The severity of intrusion determines the impact of intrusion on the system. So, we first of all classified intrusion severity, determines its source and give a helpful security recommendation using causal reasoning technique. In this work, we have successfully modeled intrusion severity classification system. This work has its application in every sector that makes use of computers, networks and software that assigns privileges to its users. The advantage of this work is that minimize the cost of fighting intrusion because it first of all shows the effect or the impact (I.e how serious) of the intrusion on the organization. From there, one can decide to tackle it or just ignore it if the severity suggests so. However, what may be seen as the weakness of the work is that it does not function in real time.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] P. Roshni Mol, Dr. C. Immaculate Mary. (2021). Classification of Network Intrusion Attacks Using Machine Learning and Deep Learning. Annals of the Romanian Society for Cell Biology, 25(2), pp-1927–1943.

[2] Emad E. Abdallah, Wafa' Eleisah, Ahmed Fawzi Otoom, (2022). Intrusion Detection Systems using Supervised Machine Learning Techniques: A survey. International Conference on Ambient Systems Networks and Technologies, 8(3), pp-205–212.

[3] Albasheer, H.; Md Siraj, M.; Mubarakali, A.; Elsier Tayfour, O.; Salih, S.; Hamdan, M.; Khan, S.; Zainal, A.; Kamarudeen, S.(2022). Cyber-Attack Prediction Based on Network Intrusion Detection Systems for Alert Correlation Techniques: A Survey. Sensors, 22(44), pp-14-94.

[4] Patil, S.; Varadarajan, V.; Mazhar, S.M.; Sahibzada, A.; Ahmed, N.; Sinha, O.; Kumar, S.; Shaw, K.; Kotecha, K.(2022). Explainable Artificial Intelligence for Intrusion Detection System. Electronics, 11(4),pp-30-79.

[5] Shadman Latif, Faria Farzana Dola, MD. Mahir Afsar, Ishrat Jahan Esha(2022). Investigation of Machine Learning Algorithms for Network Intrusion Detection. International Journal of Information Engineering and Electronic Business. DOI: 10.5815/ijieeb.2022.02.01 ,1-22.

[6] Erukala Suresh Babu, Mekala Srinivasa Rao,Rambabu Pemula,Soumya Ranjan Nayak,Achyut Shankar(2022). A Hybrid Intrusion Detection System against Botnet Attack in IoT using Light Weight Signature and Ensemble Learning Technique. Research Square Journals,4(2),pp- 1-17.

[7] Gurbani Kaur, Dharmender Kumar (2020). Classification of Intrusion using Artificial Neural Network with GWO. International Journal of Engineering and Advanced Technology (IJEAT),9(4), pp-599-606.

[8] Amit Singh,Jay Prakash, Gaurav Kumar(2022). Intrusion Detection System: A Comparative Study of Machine Learning-based IDS. Journal of Research Square, DOI: https://doi.org/10.21203/rs.3.rs-1634802/v1.

[9] Hidayat, Imran; Ali, Muhammad Zulfiqar; Arshad, Arshad (2022). Machine learning based intrusion detection system: an experimental comparison. Journal of Computational and Cognitive Engineering, 5(3),pp-33-52.

[10] Jawad Ahmad, Syed Aziz Shah, Shahid Latif, Fawad Ahmed, Zhuo Zou, Nikolaos Pitropakis,(2022). A deep random neural network with particle swarm optimization for intrusion detection in the industrial internet of things. Journal of King Saud University – Computer and Information Sciences,3(3),pp-1-10.

[11] Ruohao, Z.; Condomines, J.-P.; Lochin, E.(2022). A Multifractal Analysis and Machine Learning Based Intrusion Detection System with an Application in a UAS/RADAR System. Drones, 6(21). https://doi.org/10.3390/ drones6010021.

[12] Imeh Umoren, Saviour Inyang, and Onukwugha Gilean (2021). Bayesian Network Algorithm for Predictive Modeling of Cyber Security for Efficient Bank Channels Digitalization, 5(2), pp-54-68.

[13] Anietie Ekong, Blessing Ekong and Anthony Edet (2022), Supervised Machine Learning Model for Effective Classification of Patients with Covid-19 Symptoms Based on Bayesian Belief Network, Researchers Journal of Science and Technology(2022),2, pp-27-33.