

# Automatika

Journal for Control, Measurement, Electronics, Computing and Communications



ISSN: (Print) (Online) Journal homepage: [www.tandfonline.com/journals/taut20](http://www.tandfonline.com/journals/taut20)

## Integrated threat intelligence platform for security operations in organizations

K. U. Abinesh Kamal & S. V. Divya

To cite this article: K. U. Abinesh Kamal & S. V. Divya (2024) Integrated threat intelligence platform for security operations in organizations, *Automatika*, 65:2, 401-409, DOI: 10.1080/00051144.2023.2295146

To link to this article: <https://doi.org/10.1080/00051144.2023.2295146>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 10 Jan 2024.



Submit your article to this journal [↗](#)



Article views: 1175



View related articles [↗](#)



View Crossmark data [↗](#)



# Integrated threat intelligence platform for security operations in organizations

K. U. Abinesh Kamal<sup>a</sup> and S. V. Divya<sup>b</sup>

<sup>a</sup>Department of Information Technology, Noorul Islam Centre for Higher Education, Kumaracoil, India; <sup>b</sup>Department of Computer Science & Engineering, V.S.B College of Engineering Technical Campus, Coimbatore, India

## ABSTRACT

Organizations have to establish strong security operations to protect their digital assets since cyberattacks are becoming more prevalent and sophisticated. Integrating threat intelligence into security operations is a fundamental strategy for enhancing an organization's security posture. However, the precision and dependability of the underlying machine learning classifiers employed for analysis determine how successful such platforms really are. In this paper, we leverage the UNSW-NB15 dataset to propose an integrated threat intelligence platform for security operations in organizations. In order to determine which machine learning classifier performs best, we run a variety of classifiers to the dataset, including Ensemble Learning, Stochastic Gradient Descent (SGD), Logistic Regression, and Ridge Classifier. Our findings demonstrate that the Ensemble Learning classifier beats the other classifiers, with accuracy, precision, recall, and F1 score of 97.02%, 98.34%, 99.02% and 98.17% respectively. This suggests that our proposed system is quite good at detecting potential threats and may offer insightful information for security operations in organizations.

## ARTICLE HISTORY

Received 22 September 2023  
Accepted 8 December 2023

## KEYWORDS

Threat intelligence platforms; cyberattacks; security; UNSW-NB15; ensemble learning

## 1. Introduction

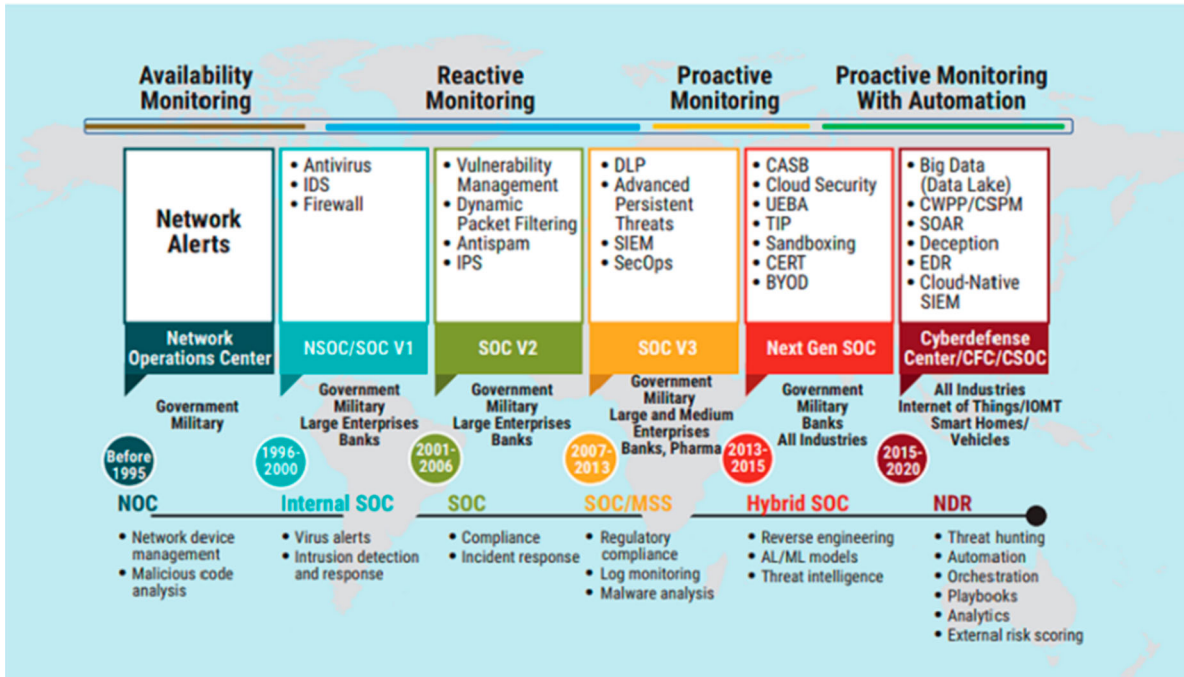
The importance of cyber security is rising on the national level. The requirement for systems for monitoring and detecting attacks which is a challenge with an ever-increasing depth and complexity, is one of the difficulties with cyber security. A trustworthy and effective security organization that can defend businesses and nations from cyberattacks is becoming more and more crucial to security defence. The Security organizations Centre (SOC) is an organization that daily gathers and analyses security information from networks, particular servers, and databases, keeps an eye out for unusual activity, and offers security services like safety precautions to the targeted users. The evolution of SOC is seen in Figure 1. SOC are capable of monitoring and responding to cyberattacks, but it is difficult to combat the ever-more complex threats by relying just on conventional heuristic and signature-based defences [1].

Attack times are shorter, there are more threat variations than previously, and harms that belong to the exact same threat category frequently begin with invaders employing identical techniques that take advantage of similar system vulnerabilities and result in significant losses on a broad scale. The simultaneous attacks on several Microsoft Windows-based systems by the ransomware infections WannaCry and Petya [2] are well-known examples. In addition, the burden on the defence is increased by multi-vectored and

multi-staged cyberattacks such as advanced persistent threats (APT), polymorphic threats, zero-day threats, and composite threats. Companies and nations that are not acquainted with the characteristics of current and emerging cyberattacks are more susceptible to the threat. As a result, sharing threat intelligence and active defence are receiving more attention from SOCs as well as additional security organizations.

Threat intelligence (TI) analysis allows for recognizing existing and upcoming cyberattacks more effectively, allowing targeted users to take prompt countermeasures to safeguard their systems and crucial information. This technique involves provides early warning of potential assaults. Threats are behaviours that may adversely affect an organization's priceless resources. Threats often take the use of a system's flaws to cause harm or the destruction of an asset. TI keeps track of static threat parameters such alias, reporting time, MD5 hashes, impacted systems, and so on. Threat dynamic features like particular attack behaviours are also contained in TI. NLP (natural language processing) and information retrieval (IR) methods can be combined with TI to extract threat activities such as tactics, techniques, and procedures (TTPs) of cyberattacks in order to comprehend the attack cycle [3]. It can additionally be used to efficiently gather unstructured TI text by extracting the indications of compromise (IOCs) of attacks [4].

**CONTACT** K. U. Abinesh Kamal [abineshkamalku.it@gmx.com](mailto:abineshkamalku.it@gmx.com) Department of Information Technology, Noorul Islam Centre for Higher Education, Kumaracoil 629180, India



**Figure 1.** A security operations centre's evolution.

Numerous countries are open to sharing successful detection techniques and exchanging security knowledge in order to successfully combat cyberattacks and mitigate hidden costs. In order to make consumers' lives easier, a lot of security firms and organizations publish blogs and reports on threat information on open source. The SOC gains from these TI papers as well since they provide analysts with a new resource that makes it much simpler and quicker for them to understand the characteristics of various assaults. They also provide targeted organizations with direction for making early security defence decisions.

A category of artificial intelligence (AI) called machine learning (ML) is widely and successfully utilized to strengthen decision-making systems across a variety of disciplines [5]. During the training phase, ML models extract and discover useful patterns from past data. The learned semantics are then applied by the models to categorize or predict unknown samples of data into the appropriate classes or values. The intelligence of ML has driven its use in various sectors to give a greater degree of analysis to enable automation and assistance in difficult decision-making processes [6]. Overall, ML improves systems' performance and effectiveness without explicit programming [7] by understanding complicated patterns that are hard for domain specialists to spot. As a result, ML has been embraced in the creation of TIP, which uses an intelligent defence layer to increase cyberattack detection [8]. To identify zero-day and sophisticated cyberthreats, organizations have widely implemented ML-based TIP capabilities. Therefore, attention has been drawn to the establishment of machine learning (ML) TIP to identify threats due to its emphasis on the behavioural patterns

of network assaults and the absence of dependence on recognized IOCs [9].

Due to the lack of publicly available datasets, this study trains and verifies the proposed approach using the UNSW-NB15 dataset as, considered as a whole, these sorts of data are those that would be gathered within an organization to develop threat intelligence. A few of the paper's major contributions are as follows:

- To present a threat intelligence method for identifying cyberthreats in organizations.
- To assess the effectiveness of the suggested method using several machine learning classifiers, such as Ensemble Learning, Stochastic Gradient Descent (SGD), Logistic Regression, and Ridge Classifier.
- Performance is measured in terms of precision, recall, accuracy, and F1-score.

## 2. Literature review

### 2.1. Background of study

#### 2.1.1. Cyberthreat intelligence (CTI)

CTI is a collection of data that has been evaluated by a company to identify the goals and attack methods of a cyberthreat. However, the majority of businesses nowadays are primarily concentrated simply on basic applications, such as firewalls, while adding intelligence boost with popular Intrusion Prevention Systems (IPS) as well as Security Information and Event Management Systems (SIEMs). Organizations can learn about the subsequent actions of the adversary with the use of the CTI System. As a result, the business can proactively defend itself against assaults in the future.

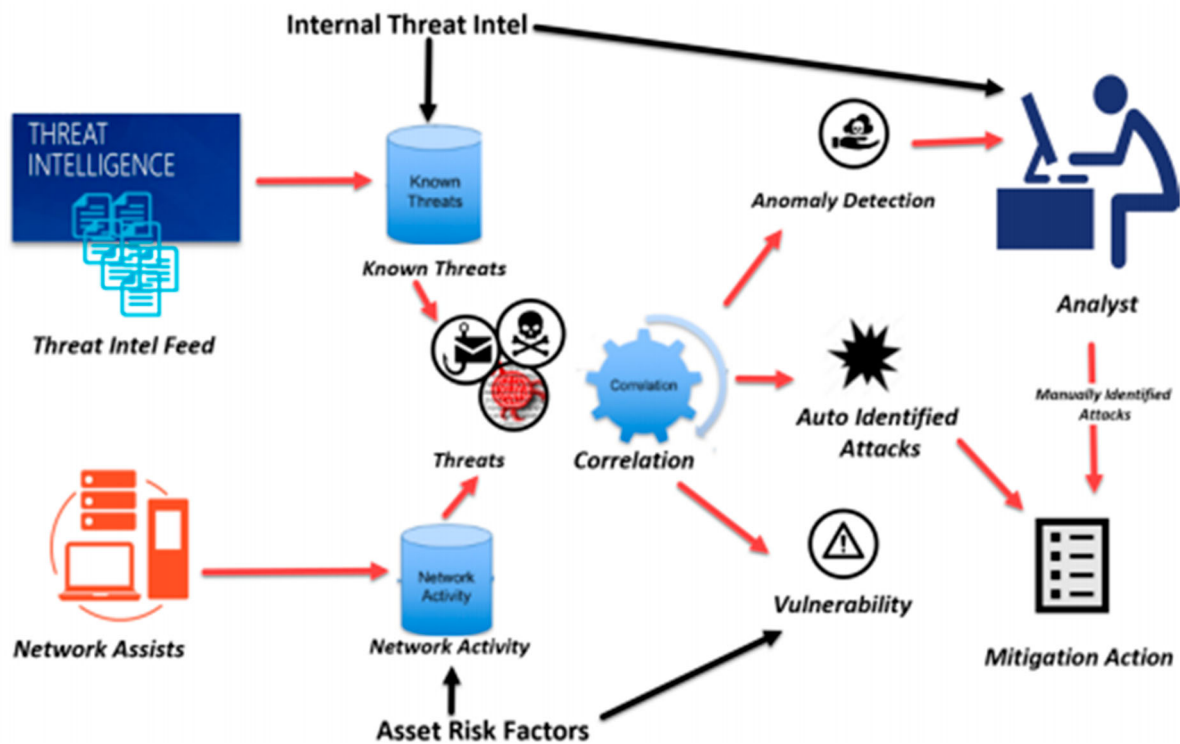


Figure 2. Threat intelligence platform.

Both CyBOX and Trusted Automated exchange of Indicator Information (TAXII) are regarded as viable alternatives, Structured Threat Information Expression (STIX) [10] is thought to be the most widely utilized CTI standard [11]. The modular framework offered by STIX can effectively include other standards [12]. STIX is used in a variety of situations with various characteristics. Instead, Sadique et al. [13] offer a novel method for creating STIX documents from raw threat data to automatically produce cyberthreat intelligence. Whereas Li and Xue [15] and Chia et al. [16] employ a system based on blockchain technology to distribute CTI data in STIX format, Ko et al. [14] utilize STIX to share risks and security-related data in IoT settings. Narayanan et al. [17] use STIX as a source of threat information and combine it with other sources of information of similar kinds to create an integrated cognitive system that can identify risks by fusing several collaborative agents that cover host and network details.

### 2.1.2. Threat intelligence platform (TIP)

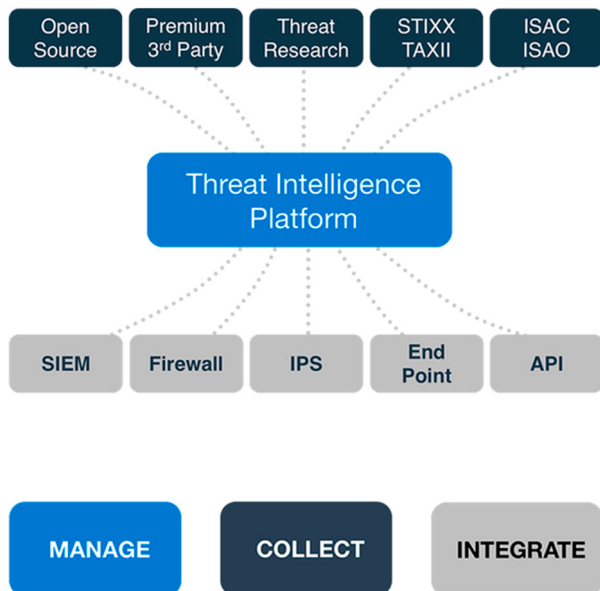
The purpose of a TIP is to gather, process, analyse, and distribute data on possible cyber dangers to an organization. Through the provision of real-time threat information, monitoring, and detection capabilities, it helps organizations defend their internet connections, applications, and systems. Threat intelligence has been described as the procedure of gathering information from several sources concerning cyberthreats that may be utilized for identifying malicious behaviour with the intention of safeguarding the assets of companies [18].

Security experts employ TIPs to monitor and find threats to their systems, networks, and data. A TIP often compiles information from a variety of sources, including internal data sources like logs and warnings produced by security technologies, as well as public and commercial sources of threat intelligence. The platform then examines this data to look for patterns and abnormalities that could point to the existence of a threat, frequently utilizing machine learning along with other innovative analytical approaches. Once an issue has been discovered, a TIP can notify security analysts and give them background knowledge about the threat, including the nature of the threat, the attacker's strategy and tactics, and any other pertinent information that might aid in an effective response (Figure 2).

In Figure 3, a TIP automatically gathers data from numerous sources and formats and reconciles it. An effective security architecture must be able to take in data from a number of sources. A TIP's architecture generally consists of a number of crucial elements, using multiple sources inside the organization's network architecture, SIEM is a security platform that gathers and examines security-related data. Real-time security risks and events may be detected and handled because of it.

TIPs include both free and premium third-party threat data feeds to provide users a complete picture of potential dangers. While premium third-party feeds need a subscription, free open-source feeds are accessible to everyone without charge. Entering and exiting network traffic is monitored and managed by firewalls, which are network security tools. TIPs gather firewall





**Figure 3.** Components in TIP.

records and analyse them to find possible threats. An IPS security system scans network traffic for possible security threats and takes appropriate action to stop them. Cyberattacks frequently target endpoints, including mobile and laptop computers. TIPs gather endpoint information, which is then used to spot potential threats. Firewalls, SIEMs, and IPSs are just a few examples of security programmes and solutions that may be integrated with one another via an application programming interface (API).

TIPs include both free and premium third-party threat data feeds to provide users a complete picture of potential dangers. While premium third-party feeds need a subscription, free open-source feeds are accessible to everyone without charge. Managing and Collecting Interface are user-friendly interfaces used by the TIP to administer and gather data. They offer a mechanism to produce reports, establish security rules, and access and analyse threat intelligence data.

## 2.2. Related works

In their exploratory investigation of software suppliers and research views of threat intelligence exchanging platforms, Sauerwein et al. [19] draw the conclusion that the market for sharing threat intelligence is still in its early stages. It has been demonstrated that SOCs can help in strengthening an organization's safety record while preventing, detecting, analysing, and mitigating to cybersecurity issues [20]. Serious security issues are still regular and widespread, though, and the SOC system as it stands is unable to stop these online attacks. According to security analysts and managers surveyed by Faris Kokulu et al. [21], contemporary SOCs have issues with weak defence against some types of threats, insufficient threat information, slow reaction times, and

a low level of automation. The authors [22] explored how to strengthen threat analysis and categorization, including new threats. To improve and standardize feature selection for threat classification, the authors suggested an approach based on stacked autoencoders.

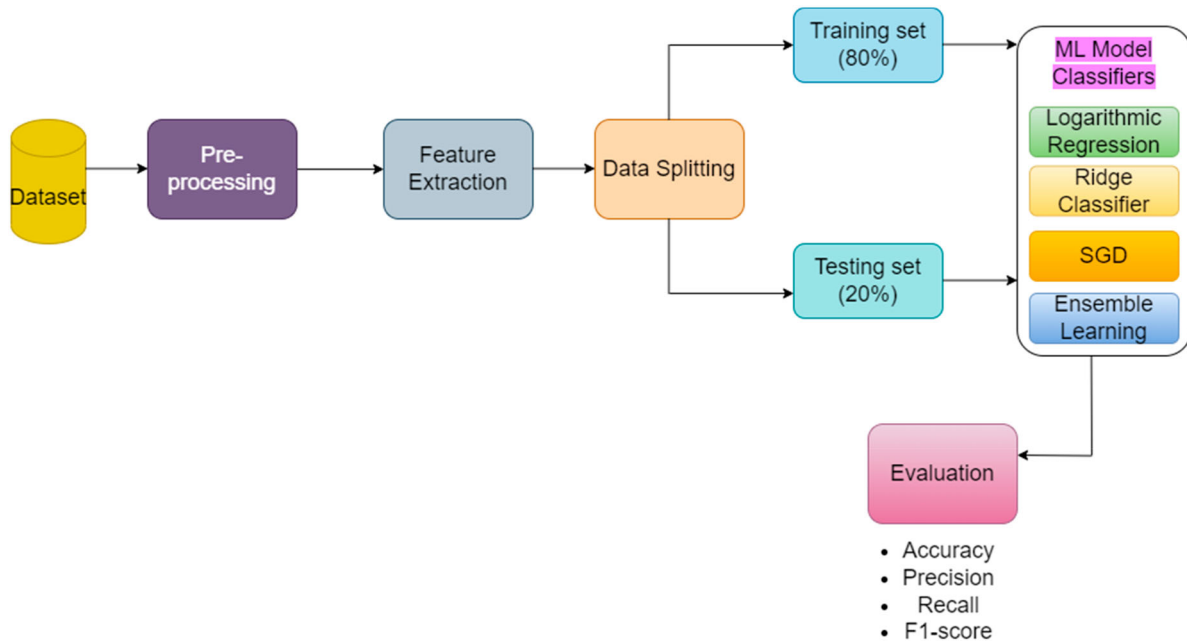
To enhance threat analysis and categorization, a hybrid DL model has been suggested in a number of research publications. Grey wolf optimization (GWO) and a CNN model were suggested by the authors in [23]. The initial GWO model is employed to choose the features in the suggested hybrid model, while the second CNN framework is used to classify threats. To enhance attack categorization, several researchers have extracted spatial and temporal features using a hybrid DL model based on CNNs and RNNs. Since a CNN might provide quick feature selection to facilitate real-time analysis, the authors [24] employed it for feature selection. The weight-dropped LSTM (WDLSTM) model was one of the LSTM variations that the authors employed for threat categorization. In regard to execution time, the suggested hybrid model performed well.

The impact of CNN on threat categorization and the operation of intrusion detection systems (IDS) was researched by Vinayakumar et al. [25]. The model using CNN-LSTM beat the other models when the authors examined several hybrid DL approaches with CNNs. The authors also made notice of the fact that using a minimal number of criteria for threat categorization reduced the classification's effectiveness. Because of this, feature selection may be accomplished by DL models with good results. Using the XGBoost and Random Forest algorithms for danger prediction, Yeboah-Ofori et al. [26] suggested a cyber supply chain threat analysis. The work takes into account threat data and forecasts the TTP used in a cyberattack, displaying remarkable accuracy in their empirical assessment.

Zonget et al. [27] provided a way to assess the seriousness of CS threats by using a DL methodology to examine the language used in CS-related tweets. The tests made use of a collection of 6000 tweets that described software vulnerabilities and were annotated with the authors' assessments of how serious they were. The collected results showed a high degree of predicting accuracy for high-severity vulnerabilities and also highlighted the correlation between reports of high-severity vulnerabilities taken from web sources and actual exploits.

## 3. Methodology

In Figure 4, the pre-processing of the dataset to make it acceptable for machine learning is the first stage in creating a threat intelligence platform. Both category and numerical characteristics are present in the UNSW-NB15 dataset, but a few of the features lack values.



**Figure 4.** Procedure of the proposed methodology.

In order to do this, the data must be cleaned, missing values eliminated, categorical data transformed into numerical data, and the data scaled. To determine the distinctive properties of the threats, the data must first undergo pre-processing before the pertinent information can be recovered. The procedure of feature extraction is crucial since it aids in recognizing the distinctive traits of various threats. A set for training and a set for testing are produced from the dataset. The test set is used to assess the effectiveness of the machine learning approach, whereas the set for training is employed to train the model. In order to reduce the error rate, the parameters of the model, including the rate of learning and the number of layers, are changed during training several classification methods, including ensemble learning, the Logistic Regression, Ridge classification algorithm, and SGD. The relevant characteristics collected from the dataset are used to train the chosen machine learning algorithm on the training set. The performance of the trained model in successfully identifying assaults is then assessed on the testing set. A number of performance criteria, including precision, recall, precision, accuracy, and F1 score, can be used in the evaluation.

### 3.1. Dataset

The UNSW-NB15 dataset was used to assess the effectiveness of the proposed method since it contains a significant number of recent acceptable and anomalous actions. By recognizing possible threats and taking preventative action to avoid them, this information is utilized in an intelligent threat platform to improve the security of an organization. The dataset includes 4 CSV

files with 2,540,044 feature vectors and nearly 100 GB of connected packets. The class label and 47 characteristics are both included in each vector. It has 49 properties, including a class name, and 2,540,044 occurrences that are individually classified as either normal or threatening. The data is divided into ten classes, as indicated in Table 1, with one class representing regular operations and nine classes representing security incidents and malware activities.

Figure 5 shows how both the training dataset and the test data set are distributed based on various sorts of normal and threat activities. The X-axis depicts the captured typical threat patterns and the Y-axis which indicates the number of recordings in all the graphs.

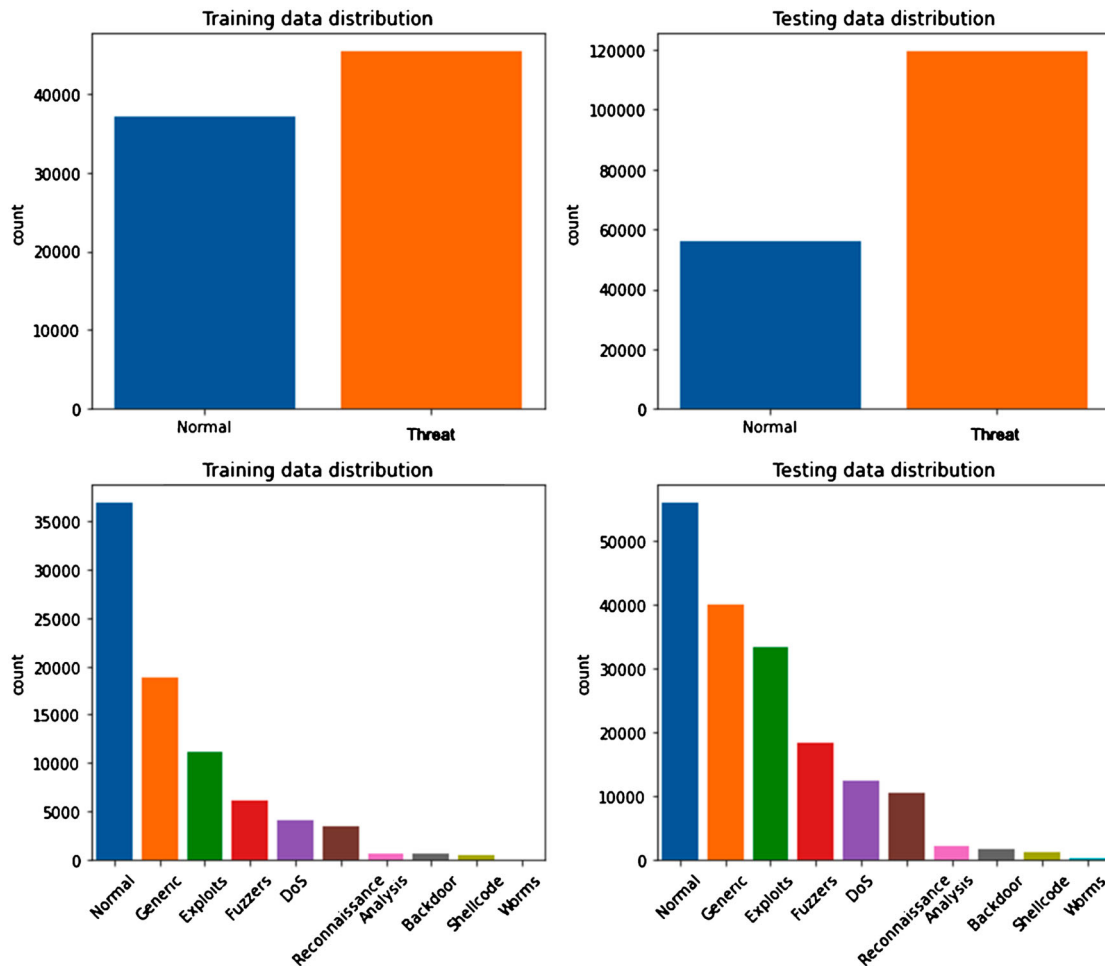
### 3.2. Classification of algorithms with machine learning

#### 3.2.1. Linear regression

Finding the functional relationship between two or more variables is done using regression. The linear regression is defined by how the straight line matches

**Table 1.** Types of threats in the dataset.

Type	Whole No. of records	Training No. of records
Normal	2,218,761	56,000
Fuzzers	24,246	18,184
Analysis	2677	2000
Backdoors	2329	1746
DOS	16,353	12,264
Exploits	44,525	33,393
Generic	215,481	40,000
Reconnaissance	13,987	10,491
ShellCode	1511	1133
Worms	174	130
		175,341



**Figure 5.** Description of the dataset.

over the variables. The interaction among a dependent variable and one or more independent variables is modelled using the statistical technique of linear regression. To determine the likelihood of a threat based on certain characteristics or factors, we employ linear regression in this study.

### 3.2.2. Ridge classifier

A linear classification technique called the Ridge Classifier performs binary classification using Ridge regression. Threat detection is a common application for it.

### 3.2.3. Stochastic gradient descent (SGD)

A popular iterative optimization approach to develop machine learning models, particularly those used for threat detection, is stochastic gradient descent (SGD). SGD is used in threat detection to reduce the loss function of a model of classification that has been trained to differentiate between normal and abnormal behaviour. SGD adjust the parameters of an algorithm that receives information about potential threats as inputs and predicts whether a threat will materialize as an output. When it comes to threat detection, the loss function is described as a binary cross-entropy between the true label for every instance in the training data and the

projected likelihood for each event. In order to minimize the loss and enhance the model's effectiveness on the training data, SGD updates the model's parameters during training via small adjustments in the path of the loss function's negative gradient. Up till convergence or a certain range of epochs is reached, the algorithm iterates through the training data, adjusting the model's parameters for each instance.

### 3.2.4. Ensemble learning classifier

Multiple models are integrated in an ensemble learning approach to outperform each one alone. In this situation, combining the three models Logistic Regression, Ridge Classifier, and Stochastic Gradient Descent (SGD) will allow us to identify threats. Utilizing each model's unique advantages to make up for the shortcomings of the others is the notion. For example, the Ridge Classifier excels at managing sparse data, logistic regression is capable of handling both nonlinear as well as linear connections, and SGD has a rapid convergence rate. To categorize whether a particular input constitutes a threat or not with regard to identifying threats, we can apply this ensemble model. For instance, the input can be a set of features that describe a particular activity or behaviour. The output would be a binary

**Table 2.** Hyperparameters.

Hyper parameters	Values
Epoch	30
Learning rate	0.0001
Optimizer	ADAM
Batch size	32

**Table 3.** Performance metrics.

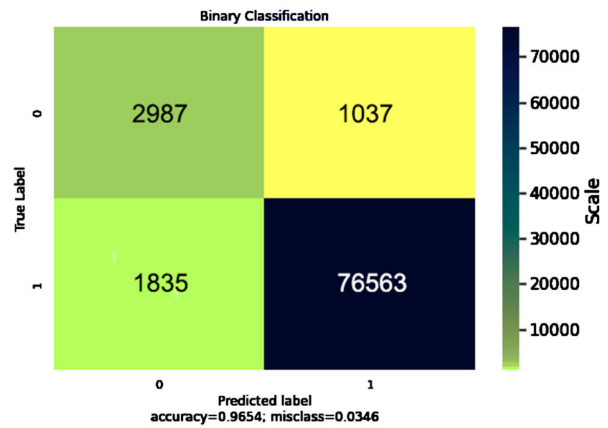
Performance metrics	Formulas
Accuracy	$\frac{(TP + FP)}{(TP + FP + TN + FN)}$
Precision	$\frac{(TP)}{(TP + FP)}$
Recall	$\frac{(TP)}{(TP + FN)}$
F1-Score	$2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$

Note: TP (True positives) are the number of correctly predicted positive instances (i.e. instances where a security threat is present), TN (True Negatives) are the number of correctly predicted negative instances (i.e. instances where a security threat is not present), FP (False Positives) are the number of incorrectly predicted positive instances, and FN (False Negatives) are the number of incorrectly predicted negative instances.

classification, where 1 indicates the presence of a threat and 0 indicates no threat.

### 3.3. Model implementation

Hyperparameters are variables that the user sets to direct the learning process rather than ones that are learned during training. Hyperparameters may be employed to tweak an organization’s security threat intelligence platform’s settings and algorithms in order to get better outcomes. The hyperparameters utilized are shown in Table 2.



**Figure 6.** Confusion matrix for binary classification.

### 3.4. Evaluation tools and metrics

The performance of the suggested model was assessed in this study using a variety of assessment criteria, including precision, recall, accuracy, and F1-Score. Accuracy measures how accurate a threat classification model is by comparing the proportion of properly categorized threats to all threats. Recall measures how well a model can categorize threats (Table 3).

## 4. Results and discussions

Using the Python programming language, we programmed with Google Collab Notebook software. The suggested model is implemented using the scikit learn and Keras data pre-processing programmes. A 1.6 GHz



**Figure 7.** Learning curves.





- [2] Symantec. Petya ransomware outbreak: here's what you need to know; 2017 Dec. [cited 2019 Aug 25]. Available from: <https://www.symantec.com/blogs/threat-intelligence/petya-ransomware-wiper>
- [3] Husari G, Al-Shaer E, Ahmed M, et al. TTPDrill: automatic and accurate extraction of threat actions from unstructured text of CTI sources. In: Proceedings of the 33rd Annual Computer Security Applications Conference; Orlando, FL, USA. ACM; 2017. p. 103–115.
- [4] Liao X, Yuan K, Wang X, et al. Acing the IOC game: toward automatic discovery and analysis of open-source cyber threat intelligence. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; Vienna, Austria. ACM; 2016. p. 755–766.
- [5] Goodfellow I, Bengio Y, Courville A. Machine learning basics. *Deep Learn.* 2016;1(7):98–164.
- [6] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science.* 2015;349(6245):255–260. doi:10.1126/science.aaa8415
- [7] Mahesh B. Machine learning algorithms – a review. *Int J Sci Res.* 2020;9:381–386.
- [8] Tsai C-F, Hsu Y-F, Lin C-Y, et al. Intrusion detection by machine learning: a review. *Expert Syst Appl.* 2009;36(10):11994–12000. doi:10.1016/j.eswa.2009.05.029
- [9] Bhuyan MH, Bhattacharyya DK, Kalita JK. Network anomaly detection: methods, systems and tools. *IEEE Commun Surv Tutor.* 2013;16(1):303–336. doi:10.1109/SURV.2013.052213.00046
- [10] M. STIX. Available from: <https://oasis-open.github.io/cti-documentation/stix/intro>
- [11] Shackleford D. Who's using cyberthreat intelligence and how? SANS Institute; 2015.
- [12] Burger EW, Goodman MD, Kampanakis P, et al. Taxonomy model for cyber threat intelligence information exchange technologies. In: ACM Workshop on Information Sharing & Collaborative Security; 2014.
- [13] Sadique F, Cheung S, Vakili I, et al. Automated structured threat information expression (STIX) document generation with privacy preservation. In: 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (IEEE UEMCON 2018); 2018.
- [14] Ko E, Kim T, Kim H. Management platform of threats information in IoT environment. *J Ambient Intell Humaniz Comput.* 2018;9(4):1167–1176. doi:10.1007/s12652-017-0581-6
- [15] Li J, Xue Z. Distributed threat intelligence sharing system: a new sight of P2P botnet detection. In: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS); 2019.
- [16] Chia V, Hartel P, Hum Q, et al. Rethinking blockchain security: position paper. In: 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData); 2018.
- [17] Narayanan SN, Ganesan A, Joshi K, et al. Early detection of cybersecurity threats using collaborative cognition. In: 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC); 2018.
- [18] Bromiley M. Threat intelligence: what it is, and how to use it effectively. North Bethesda (MD): SANS Inst.; 2016. (Tech. Rep. 37282).
- [19] Sauerwein C, Sillaber C, Mussmann A, et al. Threat intelligence sharing platforms: an exploratory study of software vendors and research perspectives. In: Conference on Wirtschaftsinformatik; 2017.
- [20] Kan M. Boeing's wannacry run-in is a reminder to patch your systems; 2018.
- [21] Kokulu FB, Soneji A, Bao T, et al. Matched and mismatched SOCs: a qualitative study on security operations center issues. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019; 2019 Nov 11–15; London, UK. ACM; 2019. p. 1955–1970. doi:10.1145/3319535.3354239
- [22] Thing VLL. IEEE 802.11 network anomaly detection and attack classification: a deep learning approach. In: Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC); 2017 Mar 19–22; San Francisco, CA, USA. p. 1–6.
- [23] Garg S, Kaur K, Kumar N, et al. A hybrid deep learning-based model for anomaly detection in cloud datacenter networks. *IEEE Trans Netw Serv Manag.* 2019;16:924–935. doi:10.1109/TNSM.2019.2927886
- [24] Hassan MM, Gumaei A, Alsanad A, et al. A hybrid deep learning model for efficient intrusion detection in big data environment. *Inf Sci.* 2020;513:386–396. doi:10.1016/j.ins.2019.10.069
- [25] Vinayakumar R, Kp S, Poornachandran P. Applying convolutional neural network for network intrusion detection. In: Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 2017 Sept 13–16; Udupi, India. p. 1222–1228.
- [26] Yeboah-Ofori A, Mouratidis H, Ismai U, et al. Cyber supply chain threat analysis and prediction using machine learning and ontology. In: Proceedings of the Artificial Intelligence Applications and Innovations—17th IFIP WG 12.5 International Conference, AIAI 2021; 2021 Jun 25–27; Crete, Greece. Cham: Springer; 2021. Vol. 627, p. 518–530.
- [27] Zong S, Ritter A, Mueller G, et al. Analyzing the perceived severity of cybersecurity threats reported on social media. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2019 Jun 2–7; Minneapolis, MN, USA. Stroudsburg (PA): Association for Computational Linguistics; 2019. Vol. 1, p. 1380–1390.