electronics

MDPI

Article

Machine Learning Techniques for Cyberattack Prevention in IoT Systems: A Comparative Perspective of Cybersecurity and Cyberdefense in Colombia

Emanuel Ortiz-Ruiz [1,2], Juan Ramón Bermejo [2,*], Juan Antonio Sicilia [2] and Javier Bermejo [2]

[1] EAN University, Bogotá 110221, Colombia; eeortiz@universidadean.edu.co
[2] Engineering and Technology School, Internacional de La Rioja University, 28040 Madrid, Spain; juanantonio.sicilia@unir.net (J.A.S.); javier.bermejo@unir.net (J.B.)
* Correspondence: juanramon.bermejo@unir.net

Abstract: This study investigates the application of machine learning techniques for cyberattack prevention in Internet of Things (IoT) systems, focusing on the specific context of cyberattacks in Colombia. The research presents a comparative perspective on cyberattacks in Colombia, aiming to identify the most effective machine learning methods for mitigating and preventing such threats. The study evaluates the performance of logistic regression, naïve Bayes, perceptron, and *k*-nearest neighbors algorithms in the context of cyberattack prevention. Results reveal the strengths and weaknesses of these techniques in addressing the unique challenges posed by cyberattackers in Colombia's IoT infrastructure. The findings provide valuable insights for enhancing cybersecurity measures in the region and contribute to the broader field of IoT security.

Keywords: machine learning techniques; cyberattack prevention; Internet of Things (IoT) systems; cyberattacks in Colombia; comparative perspective; IoT security and privacy

Citation: Ortiz-Ruiz, E.; Bermejo, J.R.; Sicilia, J.A.; Bermejo, J. Machine Learning Techniques for Cyberattack Prevention in IoT Systems: A Comparative Perspective of Cybersecurity and Cyberdefense in Colombia. *Electronics* **2024**, *13*, 824. https://doi.org/10.3390/electronics13050824

Academic Editors: Christos J. Bouras, Mohiuddin Ahmed and Abebe Diro

Received: 9 January 2024
Revised: 7 February 2024
Accepted: 18 February 2024
Published: 20 February 2024

1. Introduction

The rapid proliferation of Internet of Things (IoT) networks in Colombia [1] and various global regions has unquestionably resulted in a significant increase in cyberattacks [2]. This rapid expansion of interconnected devices carries significant implications for individuals, organizations, and governing entities [3]. As a result, deficiencies in security updates and transparency regarding IoT device security, coupled with unsafe internet deployment practices, leave these devices vulnerable to cybercriminal activities. Periodic assessments conducted by cybersecurity firms highlight the suboptimal security landscape of IoT infrastructure. Given the widespread deployment of IoT devices not only in private environments [4], but also in a variety of settings, including critical infrastructure installations [1], it is imperative to protect these devices and their associated infrastructures. Numerous techniques are currently available to identify cyberattacks targeting IoT infrastructures. Machine-based methodologies offer distinct advantages over signature-based analysis, such as enhanced detection precision and reduced false positives [3] (p. 5), while also facilitating the identification of both anomalies and novel attack characteristics. Nevertheless, these approaches are not devoid of drawbacks [5], such as the demand for supplementary hardware resources and diminished data-processing velocities.

A viable countermeasure to this burgeoning menace involves the employment of machine learning methodologies. By harnessing intelligent algorithms, detection and prevention of cyber-intrusions upon IoT infrastructures become feasible. Nevertheless, it is essential to recognize these techniques' limitations and necessitate scrupulous supervision and perpetual modification to outpace the everchanging tactics employed by cyber-malefactors [1].

- Colombia's vulnerability to cyber-intrusions underscores the immediacy for more robust cybersecurity provisions [1]. As the nation's IoT network experiences continue growth, the likelihood of more severe breaches intensifies, jeopardizing personal privacy and national security.
- To effectively attenuate these hazards, it is imperative for individuals, organizations, and governing authorities to engage in close cooperation towards devising comprehensive cybersecurity protocols tailored explicitly for IoT systems [1]. Unswerving vigilance and anticipatory endeavors are crucial in maintaining a strategic advantage over nefarious actors seeking to capitalize on susceptibilities within Colombia's rapidly developing digital terrain [1].
- This document provides an explanation of the utilization of machine learning algorithms within the Internet of Things (IoT) security landscape [6]. The transparency and traceability afforded by blockchain facilitate secure and verifiable data transactions, enabling ML algorithms to operate on trustworthy datasets [7]. This synergy not only fortifies the reliability of AI-driven decisions in IoT applications but also establishes a resilient defense against potential security breaches, providing a meticulous overview of fundamental principles and key attributes technology, elucidating how these features can be strategically harnessed to augment the capabilities of AI in application issues.
- In general, each cyber-attack has a solution that improves the impact on the economy [8] and the introduction of good practices for globalization change:
    1. Symbiotic integration for decentralized economies: Delving into the synergistic potential of integrating AI and blockchain [9], elucidating how this union could cultivate a novel ecosystem characterized by decentralized economic structures. Additionally, outlining the inherent benefits derived from this transformative integration.
    2. Comprehensive taxonomy of IoT ecosystem: Offering a meticulous taxonomy encompassing diverse dimensions, including blockchain platforms, architectural frameworks, infrastructure typologies, and consensus protocols. This taxonomy is complemented by an exploration of existing applications of decentralized AI within this comprehensive framework.
    3. Examination of practical applications: Presenting a thorough examination and discussion of multiple practical use cases wherein AI applications leverage blockchain technology across various vertical domains.

The ongoing research on Internet of Things (IoT) anomaly detection [10,11] is a rapidly expanding field. This growth necessitates an analysis of application trends and current gaps. The vast majority of publications focus on areas such as network and infrastructure security, sensor monitoring [12], and applications for smart homes and smart cities, and are extending into even more sectors. Recent advancements in the field have increased the need to study the numerous applications of anomaly detection [3,7,13] in IoT. This paper commences with a summary of detection methods based in ML and applications into the Colombia ecosystem, followed by a discussion of categorizing anomaly detection algorithms in IoT. Subsequently, we scrutinize current publications to identify distinct application domains, examining selected documents based on our search criteria. Different surveys encompass 64 documents among recent publications released between January 2019 and July 2021. In these recent publications, we observe a shortage of anomaly detection methodologies in IoT [14], for instance, when confronting the integration of systems with various sensors obtained with distributive data analytics, changes in data and conceptual shifts, and data augmentation where there is a scarcity of ground truth data.

In the context of cyberattacks on IoT [13], we identify significant gaps in current methodologies. The integration of systems with multiple sensors becomes more vulnerable, and changes in data and concepts present considerable challenges. The lack of ground truth data in data augmentation situations further complicates anomaly detection in IoT. This perspective underscores the critical need for more effective strategies in the cybersecurity context of IoT. Despite advances, substantial challenges persist that necessitate innovative solutions and more robust detection methods such as machine learning [13].

In 2019, Symantec recorded a 600% rise in attacks on the IoT platform [15]; in response to these challenges, we propose new perspectives and areas for future research. Enhancing anomaly detection in complex system integration scenarios, adapting to changes in data and concepts, and generating cyberattacks with different tactics, techniques, and procedures (TTPs) based MITRE ATT&CK attack techniques [16]. In augmentation environments are key areas requiring immediate attention. Furthermore, we advocate for a collaborative approach between the IoT research community and the cybersecurity industry to comprehensively address these challenges and progress towards a safer and more resilient IoT environment. Regarding detection, there are different studies of necessary comprehension aboutever-expanding IoT networks, both in Colombia and globally, which have brought about a profound transformation in the technological landscape. However, this rapid growth has also given rise to an alarming surge in cyberattacks. As a result, the need for effective detection and prevention measures has become vital to safeguard critical infrastructure and ensure cybersecurity.

In numerous investigations, efforts have been exerted to fathom the intricacies of cyberattacks on Internet of Things (IoT) systems. Understanding the methodologies employed by cyberattackers is pivotal for the development of comprehensive strategies aimed at identifying and mitigating potential threats such as DDOS [17] attacks. Grasping the nuances inherent in cyberattacks targeting critical infrastructure is essential for the implementation of suitable countermeasures, thereby safeguarding these systems against security breaches.

It is imperative for both individuals and organizations to remain abreast of the latest developments in cybersecurity and augment their comprehension of detection methods. Addressing the escalating concern surrounding the vulnerability of IoT networks to cyberattacks on critical infrastructure necessitates a commitment to continuous research, thorough analysis, and the implementation of proactive measures. Only through these concerted efforts can we effectively confront the challenges posed by the evolving landscape of cyberthreats.

In response to the above existing problems in anomaly detection, this paper proposes and evaluates the performance of logistic regression, naïve Bayes, perceptron, and *k*-nearest neighbors algorithms in the context of cyberattack prevention. The model uses a method adaptable in multiple scenarios of anomaly detection in IoT systems. The contributions of this paper are as follows.

This study employs a diverse array of algorithms to discern optimal features and validate their relevance within the context of a behavioral cyberattack dataset in Colombia. Addressing the intricate challenge posed by high-dimensional data, particularly in the context of cyberattack pattern analysis, our methodology proves effective in overcoming this obstacle. Conventional utilization of supervised learning algorithms for classification and regression tasks has demonstrated inadequacies in achieving desired efficacy. In response, we introduce a novel strategy, amalgamating cybersecurity attack categories through the extension of related methods. This approach is systematically applied in conjunction with advanced techniques to elevate cybersecurity measures and diagnostic stages, offering a transformative solution to the intricate analysis of cyberattack patterns.

We outline the model of this article as follows. Section 2 summarizes background and relative work. Section 3 describes the methodology and research about different methods of analysis and cyberattacks. Section 4 contains the experimental evaluation with machine learning algorithms applied in IoT Systems. Section 5 summarizes our paper and gives conclusions.

2. Background and Relative Work

This section provides a concise summary of Advanced Persistent Thread (APT) comparison and actual perspective in Section 2.1, common systems behavioral (CSB) sources used for attack detection in Section 2.2, and APT and cyberattacks detection (ML) and cyberphysical systems (CPSs) approaches in Section 2.3.

2.1. APT Comparison and Actual Perspective

Currently, advanced persistent threats (APTs) known to be associated with cyberattacks around the world have an imminent vision related with an actual perspective [18]. ENISA establishes a point of reference for all threats applying to assets related to information and communication technology (ICT) [19] and different scenarios such as IoT and Industrial Internet of Things (IIoT), where collaborative efforts among governments, industry leaders, and organizations are imperative to formulate comprehensive strategies that effectively tackle the expanding threat landscape surrounding IoT networks. Through strategic investments in cutting-edge security solutions and the adoption of proactive measures, including robust encryption protocols and advanced intrusion detection systems, we can successfully mitigate the risks associated with cyberattacks on critical infrastructures. Ref. [19] (p. 11) indicates an attacks incidents taxonomy related with a combination of cybercrime subject (motive), action (method), object (outcome). EUROPOL's European Cybercrime Centre has categorized incidents based on their type and event, which is a crucial aspect of incident management along two vectors where organizations can effectively analyze and respond to various types of events in a structured manner [20]. With regard to APT events in a critical cyber infrastructure (CCI) in a digital ecosystem, APT types are defined by NIST (US National Institute of Standards and Technology) [21]. TTPS enabled at the IoT layers (MI-TRE ATT&CK methodology) and consequently areas of interest [22] such as industry, smart grids, transport and medical services involve the implementation of restricted requirements based on security, privacy and trust.

The cyber-physical perspective with detection and prevention systems has some paradigms in IoT systems related to perimeters. I. Stellios et al. [23] describe *edge nodes* as (i) edge nodes (e.g., RFIDs, sensors), (ii) edge computing (fog), and (iii) communications. Other paradigms to consider are the OT/IoT reference security architecture NIST SP 800-82r3 [24] or scenarios, such as Mirai, and DDOS cyber-attacks [25] when there are undefined parameters and constant changes, high heterogeneity, autonomous entities, inclusion of non-traditional devices, and limited permission granularity. Along these lines, the expansion of DDOS cyber-attacks is significant among hyper-connected devices as follows [21]:

- Computers in general;
- Network nodes;
- Mobile devices;
- Wearable items;
- Video games;
- Home automation items;
- Storage devices;
- Surveillance items;
- Work devices;
- Domestic virtual assistants;
- Cars;
- Media and TV items;
- Appliances examples;
- Other generic items.

The cyberspace context change of cyberattacks detection [26] related to scenarios and taxonomy involve APT features considered in the attack cycle (see Figure 1):

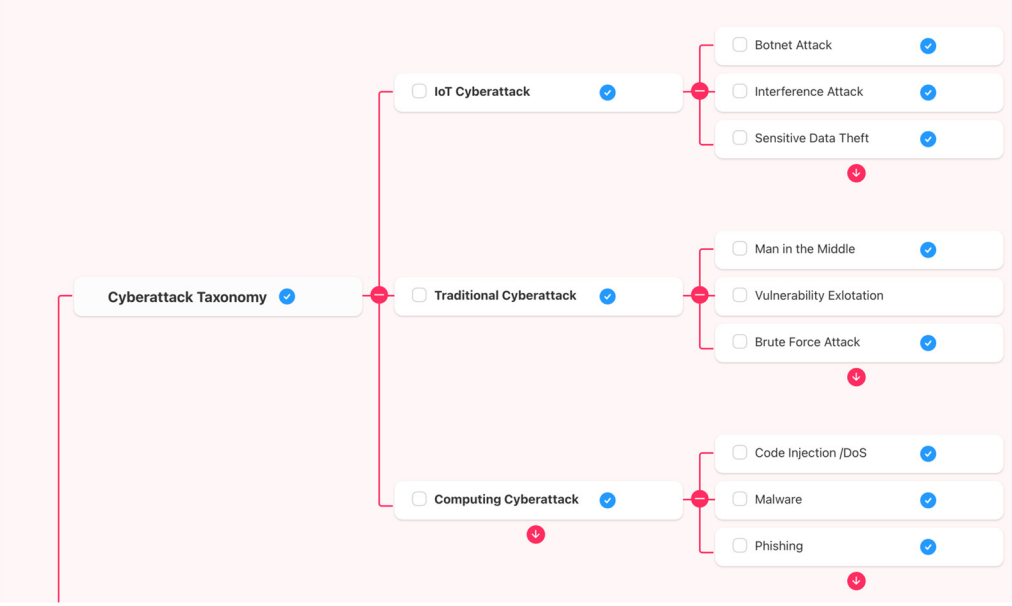

Figure 1. Cyberattack taxonomy based in [27] the APT lifecycle.

During a given campaign about advanced persistent process, Kim, Gihoom et al. [28] refer to the APT attack process related with *preparation*, *intrusion*, *inside activity*, and *achievement*. In the preparation phase, the adversary acts by analyzing and collecting the information of the victim and gaining control with the command and control server (C and C) with a malicious code. In the inside phase, the insider collects IT infrastructure information of the targeted object to achieve its final goal of the system becoming infected in the intrusion phase. In the example explained in [27], an APT using malicious software tailored to a specific target establishes a communication network and facilitates the injection of malicious code by attackers. Employing a covert approach, this specialized malware traverses laterally within the system, meticulously scanning for security vulnerabilities and leveraging them to infect additional systems within the network. The phases related to TTPs conducted over a long time in infiltration and exploitation and include (1) initial access, (2) defense, evasion, (3) credential access, (4) discovery, (5) collection, (6) and exfiltration. The taxonomy used in the APT lifecycle context [26,27] is different to common analysis and actual perspectives; according to [27], by 2025, 64 billion IoT devices will be connected to various cutting-edge environments, including smart cities, Industry 4.0, and crowdsensing. Military and governments in the world, except Latin-American, indicate [29] growth of connections to 5G, and there is a problematic focus for Colombia to advance in IoT [30] implementation of a 5G plan. In harmony with world trends, the MinICT (Ministry of Information Communications and Technology) proposes possible uses in each of the 5G spectrum areas. Frequencies below 1 GHz will preferably be used for high-speed mobile broadband in urban, suburban, and rural areas; in this context, many potentialities of IoT framed in 5G are analyzed for the advancement of the regions such as agro-industrial, manufacturing, tourism, E-health, and environment.

2.2. Common Systems Behavioral (CSB) Sources Used for Attack Detection

The consequence related with taxonomy use, according to [27], is that the lifecycle of APT and IoT (implementation 5G in Colombia) must contribute an Annual Cybersecurity Survey 2022–2023 [31]. Consequently, a methodology applicable and adaptable in Colombia is created to identify the root cause of approximately 157,000 daily cyberattacks [32] suffered by the country, a large percentage of which are aimed at IoT technologies, which can be implemented with the emergence and implementation of the 5G network. The consequence related to taxonomy use, according to [28], including the lifecycle of APT and description of some main factors on IoT (5G implementation in Colombia) is attributed in part to the increase

of cybercrime in Colombia, according to Annual Cybersecurity Survey 2022–2023 [32], with the balance in 2021–2022 of 62% of cybercrime infractions in Colombia. The context in this order of ideas indicates that the cost could exceed USD 90 million by 2025. In [31] (page 19) the same report, it indicates that AI for early detection and anticipation of such actions is one of the major challenges in the development of innovative solutions against cyberattacks.

Indeed, Colombia has been facing an adversarial context since 2019 [33], particularly in relation to behavioral patterns associated with identified advanced persistent threats (APTs) such as APT-related Cyberattacks in Colombia:

- (APT-C-36) [34] "*Attackers coming from South America carried out continuous targeted attacks against Colombian government institutions as well as important corporations in financial sector, petroleum industry, professional manufacturing*", as per the Spalax Operation in 2021 with a RAT (remote-access Trojan) scuh as: Remcos, njRAT, AsyncRAT, and others (see Figure 2).

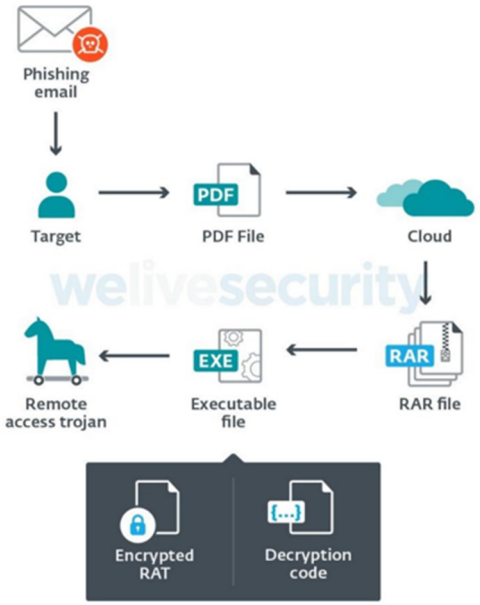

Figure 2. Behavioral execution AsyncRAT (payload) obtained by the ESET [35] campaign APT—C-36 [33].

In relation to CSB, the common system behavioral is a set of TTPs related with APT actions on objectives [28], but in IoT–I–IoT [36] ecosystems, there are different behavioral sources (external and in-device behavior) [27]. Other methods use detection systems which are related to malicious domain names, such as extracting the readability of domain names, time to live (TTL) [37], features to establish classifiers for detection or methods [36] related to recognition of phases, e.g., reconnaissance, weaponization, delivery, exploitation, installation, com-mand and control, and action [16,18,36]. With respect to an attack modeling, a large cyberphysical system (CPS) is characterized by integration, across technologies, industrial domains, and the lifecycle, and by "smartness as IoT". CPSs can be described using a set of characteristics: technical emphasis, cross-cutting aspects, level of automation and lifecycle integration. But in relation with common system behavioral detection, there are different components regarding comprehension for IoT systems. In Ref. [26] (page 4), the characteristics have three instances for detection: (1) IoT terminal devices are deployed, especially scattered; (2) the majority of IoT terminal devices have limited hardware resources which are unsuitable for installing large-scale attack detection software; (3) it is impossible to deploy security detection hardware directly on the resource-constrained IoT terminal devices.

In this sense, the layers for detection and prevention have has different elements (data sources of data and construct of ontology) like a comparison and structure [26], e.g.,

Alerts instances correlation:

- Alert instances filters.
- Alert instances clustering.
- APT attack scenario.

Victim–host IP

- Preliminary correlation.
- Log instances community detection.

2.3. APT and Cyberattacks Detection (ML) and (CPS) Approaches

The first step for detection and identification of an APT behavior [27,38] is to connect sources by origin and destiny. Depending on the design stage, the TTP fingerprint was specifically developed and configured to detect advanced persistent threat (APT) attacks. This fingerprint leverages the correlation between the attack tree and MITRE framework, which serves as a comprehensive reference for mobile tactics and techniques associated with APT attacks. According to reference [38], the attack tree implemented in the fingerprint encompasses all relevant mobile tactics and techniques outlined by the MITRE framework. This ensures that a wide array of potential APT attack vectors can be accurately identified and mitigated through this sophisticated detection system.

Related with APT-C 36, there are components for evaluating [33,35] a CSB where detection systems can identify [26] anomalies according to [4] behavioral needs:

- *Anomaly detection:* [7,13]: Construct a taxonomy [23] or ontology [28] based [18] correlation (Table 1) and convert a BE (behavioral execution) in an input coordinated with a CSB detected. The inclusion analyzes the interrelationships of machine learning use in cybersecurity and CPS [13], e.g., *virtual MAC spoofing detection* [39].
- *Misuse detection:* The second dataset is the distance from the data to the nearest neighbor within the cluster or CPS (correlated). Furthermore, the approach rebuilds the data features by using the distances, and formats the data features as logistic regression, naïve Bayes, perceptron, and *k*-nearest neighbor (*k*-NN) classifier.

Table 1. The attributes of advanced persistent threat (APT) [26] with CSB detected and associated.

Number	Attributes	Description
1	Timestamp	The time the alert occurs.
2	Alert_Type	The type of alert.
3	Src_Ip	The source IP of the attack step.
4	Dest_Ip	The destination IP of the attack step.
5	Src_Port	The source port number of the attack step.
6	Dest_Port	The destination port number of the attack step.
7	Victim_HostIp	The IP of the host victimized by the attack step.
8	APT type (associate) Classic attack	The CSB detected (1).
9	TTP associate	The CSB detected (2).
10	Vector associate	The CSB detected (3).
11	Actor associate	The CSB detected (4).
12	Purpose associate	The CSB detected (5).
13	APT attribution	

Expanding upon the current context of the work and correlating it with each approach to the primary anomaly detection models in real time [40], there is a compelling interest in thoroughly optimizing the study's results, compensating for the volume of transformed data and effectively applying the methods mentioned in [41], particularly in convolutional neural networks (CNNs) with recurring features; several strategies can be employed as continuous review of these features also allows them to effectively complement real-time intrusion detection systems, thereby enhancing intrusion detection [42] accuracy and performance.

The purpose of this methodology is to conduct a systematic revision of the state of cybersecurity in Colombia [31] and evaluate the potential impact of AI in preventing future IoT cyberattacks [43]. Currently, there is a lack of technical expectation regarding the effectiveness of AI in this field [44]. By analyzing the existing literature and relevant data sources [8], this study aims to provide insights into the implementation and potential benefits of AI technologies for enhancing cybersecurity measures in Colombia. Through a structured approach to reviewing available information and identifying key findings, this research seeks to contribute to the development of strategies and practices that can effectively mitigate cyberthreats in the country.

3. Methodology and Research

3.1. OODA-Diamond-CKC and TTPs Methodology

In the application of the methodology of [45], related to [23], the authors created a novel dataset and conducted experiments to analyze the robustness of federated models against different types of attacks, malicious participants, and aggregation-functions-focused APTs in cyberattacks in IoT. This information analyzed at the comprehension level will affect the state of the future operating environment over time in obtaining analysis related to applied supervised learning [46] with training data or layering data oriented based on the operation of frameworks as OODA, explained by Thulfiqar Jabar et al. [27] (p. 6):

For complex IoT ecosystems, the utilization of machine learning (ML) techniques [46] can greatly enhance the efficiency and effectiveness of data collection in the context of IoT cyberattacks [4,27,45]. ML algorithms can be designed to identify patterns and anomalies in network traffic data [19,23], enabling researchers to detect potential attacks and understand the TTPs employed by threat actors. The different key factors (KFs) to consider when designing ML models for data collection include selecting appropriate feature sets, determining relevant metrics for performance *evaluation*, *choosing suitable algorithms for classification or clustering tasks*, and establishing a scalable infrastructure for data processing.

The second part is related to externally collected behavior, in relation to events of IDS/IPS systems, but in device behaviors, sources have many events related with a possible cyberattack (Figure 3). This distinction has a motivation and purpose pathway [20,36] related with APT approaches and the below figure shows this purpose:

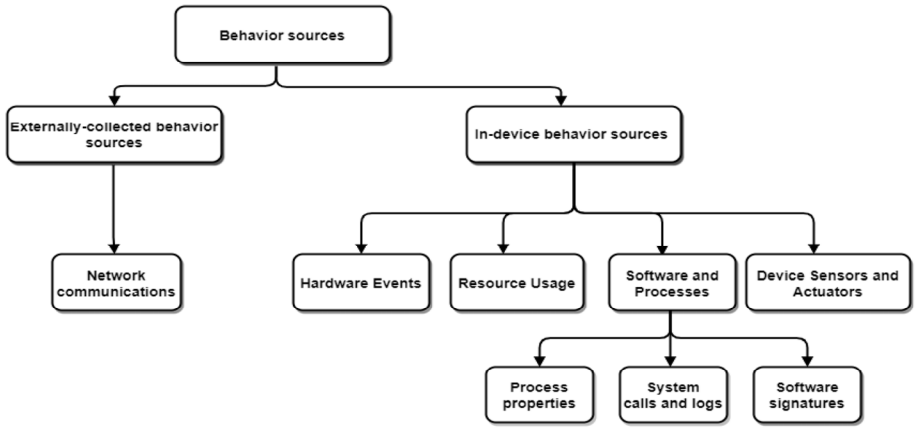

Figure 3. Behavioral sources [27] with comparison of OODA and Cyber Kill Chain [47] methods by obtaining intelligence of cyberattacks.

3.2. Diamond Model [48]

- *Actor analysis:* Identify threat actors involved in APTs against IoT systems.
- *Capability analysis:* Assess the capabilities of threat actors [49] in exploiting IoT vulnerabilities.
- *Infrastructure analysis:* Investigate the infrastructure utilized by threat actors in IoT-related APTs.
- *Victimology analysis:* Understand the targeted entities and motives behind IoT APTs.

3.3. OODA Loop [47]

- *Observe:* Collect real-time data on IoT network activities and potential security events.
- *Orient:* Analyze collected data to discern patterns, anomalies, and potential indicators of APTs.
- *Decide:* Formulate decisions based on the analysis, prioritizing potential threats [20].
- *Act:* Implement proactive and responsive measures against identified threats in the IoT ecosystem.

3.4. Cyber Kill Chain [50]

- *Reconnaissance:* Identify reconnaissance activities targeting IoT systems.
- *Weaponization*: Analyze the development and deployment of malicious tools tailored for IoT APTs.
- *Delivery:* Investigate methods employed for delivering malicious payloads to IoT devices.
- *Exploitation:* Assess the exploitation of vulnerabilities in IoT devices to establish persistence.
- *Installation:* Examine the installation of APT [18] components within IoT systems.
- *Command and control:* Analyze the establishment of communication channels between threat actors and compromised IoT devices.
- *Actions on objectives:* Understand the final objectives and impact of APTs on IoT ecosystems.

3.5. TTPs Attack Patterns [51]

- Investigate tactics, techniques, and procedures [52] employed by threat actors in executing APTs against IoT environments.
- Classify TTPs based on specific attack patterns observed in IoT-related incidents or based on external information of external data sources [27].

Furthermore, with regards to the dissemination of sensitive information pertaining to cyberattacks, stringent measures must be implemented to bolster the safeguarding of identities and uphold the highest level of confidentiality for primary data contributors.

Table 1, included in Section 2.2, enumerates the different attributes that resolve an input for identification (Diamond analysis—DA, OODA analysis, CKC analysis, and (MITRE ATT&CK patterns)... These atributtes can store [19] each of the related features in IoT and IIoT cyberattack scenarios according to industry functions based on these aspects [53], as opposed to the early detection [11,54] of an IDS when faced with an imminent cyberattack against the infrastructure. The following are the methodologies' paths:

First, the attackers use networks to perform denial of service (DoS) attacks and distributed denial of service (DDoS) attacks as a way to attack network availability. DoS and DDoS attacks [17,53] include ping of death and others, synchronize (SYN) flooding, and Hypertext Transfer Protocol (HTTP). In this case, by obtaining detection information, e.g., APT alert output from different attack detection sensors in APTALCM [26]: *timestamp, alert type, Src Ip, Dest_Ip, Src_Port, Dest Port, and Victim_HostIp*, they make a vector dimensional A(I(alert)m) = (a1, a2, a3, a4, a5, a6, a7).

The second scenario is when the attackers use an offensive cybersecurity (OS); in detail, the encryption is a method to encrypt information, and also use substitution, symmetric, asymmetric, and hash algorithms. According to Kyoungoon Kim [52], this selection has an experimental model based on OS according to threat actor capacity against cyberphysical systems (CPSs) and persons including encryption, networks, web, malware, and systems (Figure 4).

The offensive methodology based on OWASP organization (Figure 5) allows the understanding of vulnerabilities and related attacks such as cross-site scripting (XSS), broken authentication and session management, broken authentication A3 and session management, sensitive data exposure A4, insecure direct object references, XML external entities (XXE) A5, cross-site request forgery (CSRF), security misconfiguration, broken access control A6, security misconfiguration, sensitive data exposure, security misconfiguration A7,

insecure cryptographic storage, missing function level access control, failure to restrict URL access, insecure deserialization A9, insufficient transport layer protection, use of known vulnerable components, use of components with known vulnerabilities A10, unvalidated redirects, forwards unvalidated redirects and forwards insufficient logging and monitoring, each of which responds to a special feature of an APT's malicious techniques. Some anomaly detection systems are associated with current components of active observation against malware traffic in central systems [55], but, undoubtedly, it is necessary to be able to identify the actors of such threats beforehand.

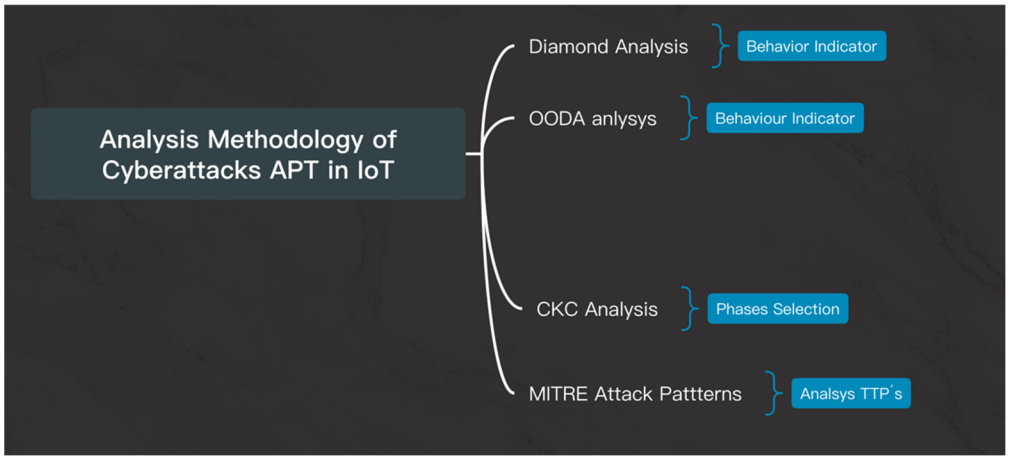

Figure 4. Combination behavioral sources (AMC–APT–IoT) [27] with comparison between OODA and Cyber Kill Chain [47] methods by obtaining intelligence of cyberattacks.

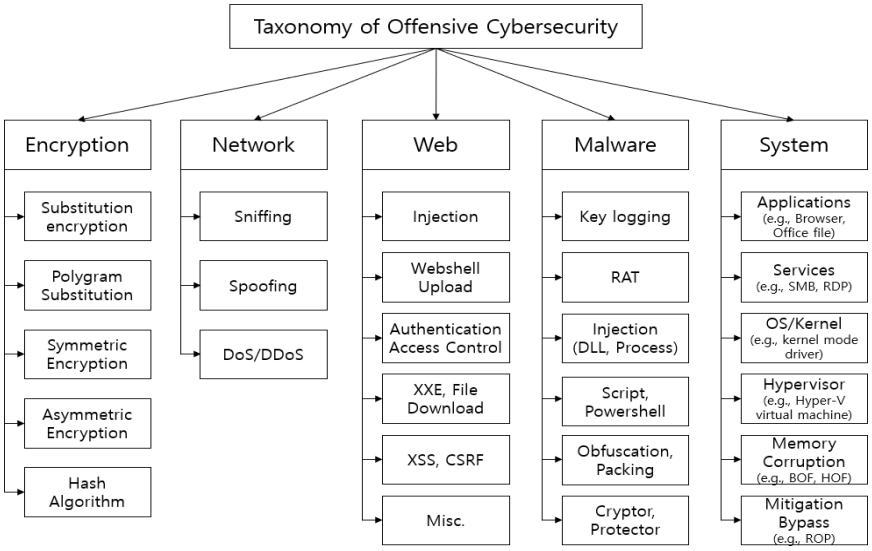

Figure 5. Taxonomy of offensive cybersecurity [52].

3.6. Select a Combination Methodology

In essence, the Colombian panorama for these methodologies does not change, but there is, in this case, an active presence of the following modalities, described as follows:

The notable surge in Figure 6 of 180,873 cases related with cyberattacks concerning this behavioral pattern stands out significantly when compared to the offensive traits [31,33,35]. Consequently, this nomenclature prompts the imperative selection of a methodology to procure pertinent information linked to malicious activities. In line with the comprehensive overview in the framework of this article, the strategy of cybersecurity measures is based on the comparative analysis carried out. This purpose is defined in three parts:

1. Alert and detection process: This method is oriented with a commentary on different positions on PSCs and protection devices, an example related to the defense of PSCs against an eventual attack affecting their availability and integrity.
2. The second process is the identification and analysis of the characteristics indicated in Figure 5, which identifies the actual identification, assessment, and capability aspects of the threat actor.
3. The relationship of the two previous parts regarding the selection of variables to determine the impact that the threat could generate in Colombia, considering the comparison in perspective related to the CPS.

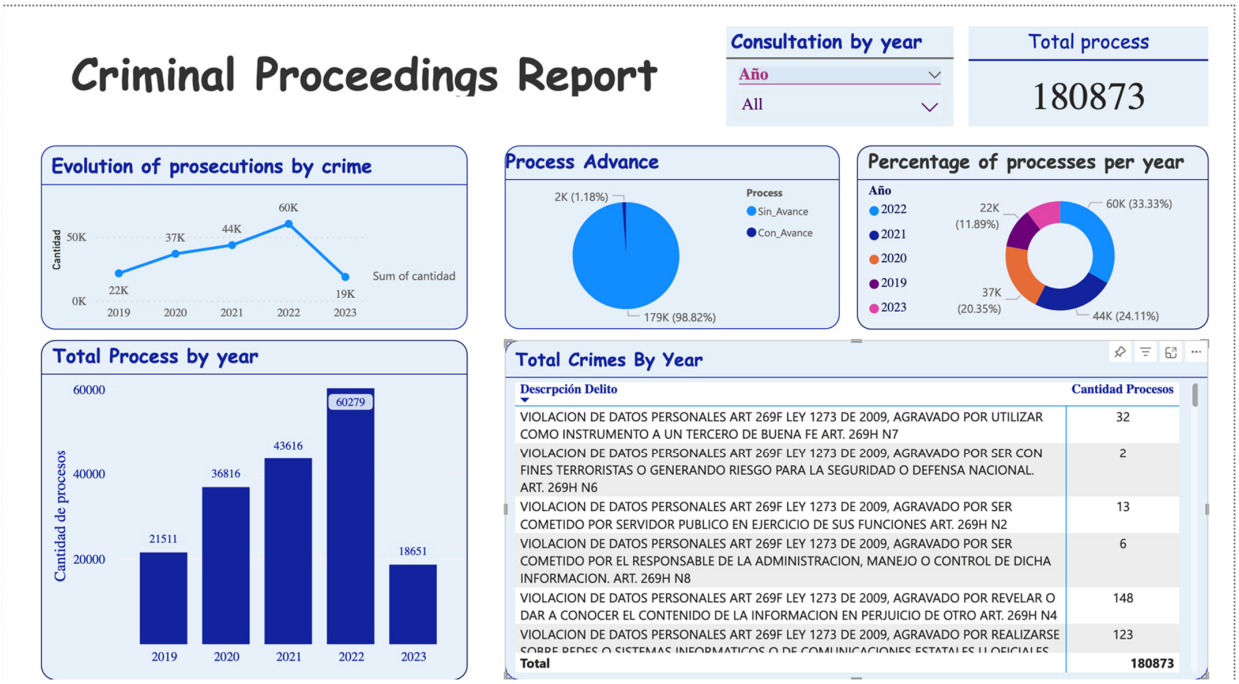

Figure 6. Power BI report (2019–2013) related to cybercriminal proceedings in Colombia collected and processing of public information [56].

In the tree phase, the threat actors are discussed as BLIND EAGLE (APT-C-36) and relationships are focused on IoT technologies based in CPS attributes compared with the APT lifecycle [57]. Brahim Ghafir indicates the entry point in an APT related to MANDIANT and explains the hyperlink and executable files that make a simple document (pdf, doc, ppt, or Excel). Many factors will be identified with a perimeter tool, but they do not know the origin of the threat. Among other factors, it is related to co-communications and traffic, which in this case have already been studied by several authors to the aspects indicated (pp. 7–8).

In relation to these two previous parts in terms of the selection of variables to determine the impact that the threat could generate in Colombia, considering the comparison in perspective related to the SPI, it is convenient to identify the differential factors before the creation of the differential factors of the "associate" components between the Diamond model (DM), OODA, CKC, and the related APT attack patterns (MITRE ATT&CK).

3.7. Attack Evaluation Patterns and Select Data Sources

A comprehensive technical analysis was carried out of the malware arsenal employed by the suspected South American espionage group, APT Blind Eagle [58]. Notably, Blind Eagle has set its sights on governmental institutions and corporations in Colombia, specifically within the financial, oil, and manufacturing sectors. Employing sophisticated attack techniques, including custom malware, social engineering tactics, and zero-day exploits, the group exhibits a high level of cyber-espionage capabilities. The report scrutinizes Blind

Eagle's multistage attack chain, offering indicators of compromise crucial for the detection and defense against their cyber-assaults. The initial stage involves a JavaScript downloader utilizing ActiveX Object to execute PowerShell commands. The second stage features a PowerShell script that loads a DLL into memory from a Base64-obfuscated and encoded string. This DLL, in turn, serves as a .NET executable file.

System Layers and Vulnerabilities

- The four layers of a system—applications, services, OS and kernel, and hypervisor—are explored, emphasizing their significance in both traditional IoT and cloud environments.
- Application categories, including browsers, Microsoft Office, and Adobe programs, play pivotal roles in the system's vulnerability landscape.
- Services, such as the server message block (SMB) and the remote desktop protocol (RDP), represent external functionalities and are potential targets for attacks.
- Operating system and kernel levels are consistently targeted, underscoring the need for robust defensive measures.
- Memory corruption techniques:
- Memory corruption is a prevalent category of vulnerabilities, encompassing techniques such as buffer overflow (BOF), heap overflow (HOF), and integer overflow.
- Various mitigation techniques, such as data execution prevention (DEP) in Windows and the no-execute (NX) bit, aim to safeguard against the execution of malicious shell code in the stack area.

Address space layout randomization (ASLR) adds an additional layer of defense by dynamically changing the stack address after each execution.

In relation with a table of input indicators (Table 1), many factors in the detection (CPS) phase complement this combined aspects to analyze the recurrent and historical data. It is necessary to take the four aspects DM/OODA/CKC/APT, and assign a numbering and aggregate weight to each of them to establish their proximity to the selected data; in this case, from 2019 to 2023, about (479.098 records obtained of ICT-OSINT) were filtered, running a script of Internet (Shodan [59], ZoomEye [60], Exploit DB [61], Packet Storm [62]), related and structured to Colombia, finding the following:

Certain distinct categories (Tables 2 and 3) enable the manifestation of the impact on mobile devices, specifically within the realms of IoT, IIoT [24], and related contexts. According to Ref. [63], with a differential schema-oriented perspective model, the purpose is related to analyze the traffic obtained from internal sources (CPS or CSB) about anomaly behavior (feature extraction); this separation has bad or good traffic-related, e.g., DDoS, attacks, in general, MIRAI Botnet. This is achieved (Table 2) through the application of sophisticated techniques, such as those outlined in the MITRE ATT&CK framework [16]. These techniques facilitate the categorization of threat actors, delineating their modus operandi (DM [48]), infrastructure, advanced persistent threat (APT) capabilities, decision-making values (OODA loop [49]), and the intricacies of their activities across each phase of the Cyber Kill Chain [47].

Table 2. An Example for associate components APT Blind EAGLE [33].

Phases ***MITRE ATT&CK***	***DM/OODA/CKC/APT***
Execution	PowerShell, scripting.
Persistence	Registry run keys, scheduled task, new service.
Privilege escalation	Process injection, exploit, access token manipulation, new service.
Defense evasion	Hidden files, modify registry, permission modify, process injection, packing, deletion, obfuscate, masquerade, de-obfuscate, disable tools, Mshta, Indicator rm, SandBox evasion.
Credential access	Credential dumping, brute force, credential in files, Pass the hash.
Hash lateral movement	WA Share, exploit remote, remote file copy.
Command and control	Common ports, multilayer encryption, remote file copy, uncommon ports, data encoding, Data obfuscation.
Exfiltration	Automated over alt. protocol, data encrypted, over C&C impact disk, encrypt data.

Table 3. Data source of behavioral methodology.

Total Source IP Addresses	Total IP Addresses Targeted	Command and Control Communications	Origin Ports	Port Destination	Malware Families	IoT Attacked	Vulnerable Devices
32.434	198.434	160.834	49.123	28.343	10	350	9.570

Each stratum enables the retrospective scrutiny of the advanced persistent threat (APT), establishing connections between its distinct techniques. This approach facilitates the examination of proximity elements, delineates attack actions, and furnishes essential elements for analysis within an MISP [64] (malware information sharing platform) or focused on incident management. Moreover, it serves as a preemptive measure for blue and red teams (Figure 6) operating within infrastructures that encompass IoT environments or infrastructures aligned with prospective analyses.

According to Figure 7, each item for orientation in CTI (cyberthreat intelligence) or intelligence gathering is necessary and the external and internal behavioral can be condensed in a combined methodology during a scenario attack [26]. In situational cyberattacks, the artificial intelligent (AI) [44] can be combined [43] with an IoT infrastructure [13]; in this case, the correlation between macro "Items" (DM [48]), infrastructure, advanced persistent threat (APT) capabilities, decision-making values (OODA loop [49]), and the intricacies of their activities across each phase of the Cyber Kill Chain are important [47]. The correlation has an effect on the measurement of parameters in relation to machine learning selection [65]; this supervised information undergoes preprocessing or data normalization.

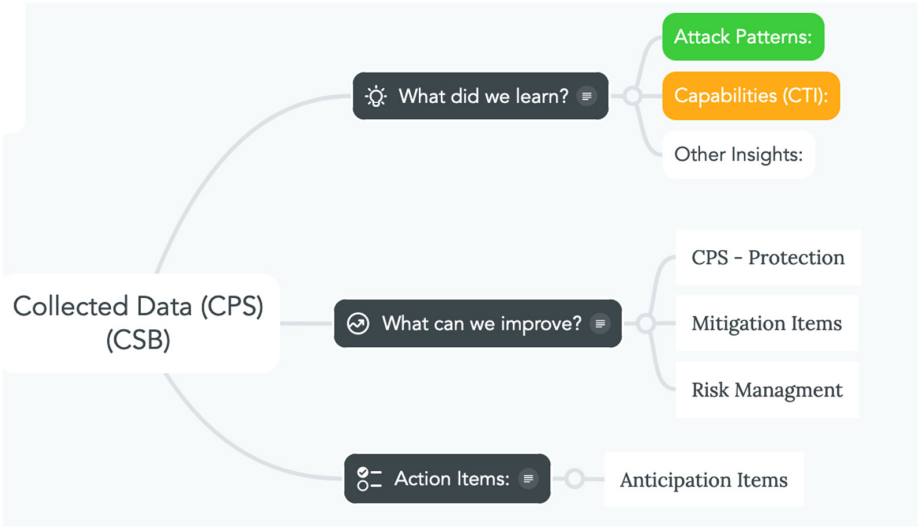

Figure 7. Structure primary data and data labeling applied.

3.8. Adjustable Supervised ML Models (Categorical Variables [63])

In the context of a binary response variable, denoted as f (with 1 representing the occurrence of the event of interest and 0 for the nonevent), and a set of predictor variables, in this case, P is conditional probability.

$$P = f\,(X) \tag{1}$$

3.9. Handling Categorical Variables and Data Times

The select model of supervised algorithm depends on IoT infrastructures in other circumstances. In this case, logistic regression was selected for normalized data regarding attack patterns, malware type and CKC phase, using the extraction of temporal components from variables, such as timestamps associated with OODAto facilitate their inclusion in the model.

3.10. Feature Selection (CVSS [66]-APT Associate)

Evaluation of the importance of features through techniques like correlation analysis or feature selection uses additional models such as naïve Bayes, perceptron, and *k*-nearest neighbor algorithms.

3.11. Creation of Derived Features

Normalization of CVSS [24] scores maintains consistency and comparability, addressing potential imbalances in class distribution, especially in the APT associate variable, to prevent biases in the model.

3.12. Training and Testing Models

The proposed model involves four algorithms (logistic regression (LR), naïve Bayes (NB), perceptron (P), and *k*-nearest neighbors (*k*-NN)).

Figure 8 proposes a combination with algorithms of external data based in complex IoT platforms [67–69], and data collection herein stems from the examination of the correlation between criminal behavior (Figure 5, Table 3) and the dataset designed for the refinement of the prevailing methodology (Figures 6 and 7). This methodology is intended to establish the nexus between the control variables and the ascription of the APT (Table 2 and [19,36] works) encompassing the requisite information for generating the categorization of each indicator outlined in the data model.

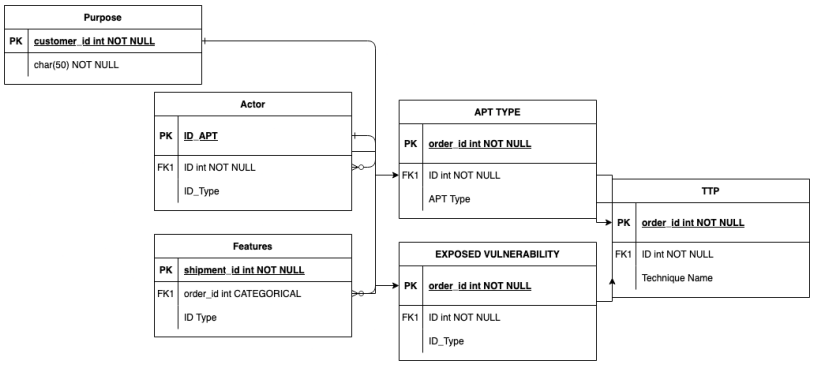

Figure 8. Database diagram (purpose, actor, features, vector, APT type, exposed vulnerability, TTP) based on Table 1, Figures 3–5.

The primary objective of the training data is to delineate an actor, encompassing features, vectors, APT types, exposed vulnerabilities, and tactics, techniques, and procedures (TTPs). These are associated with scalar attributes, including protocols, ports, and communications (command and control, C&C), all of which are incorporated within the performance of the dataset.

4. Precision of Data Frame and Accuracy Tasks

For this data model, the following good practices [70] were used to calculate the precision and accuracy of the data as clear definition of metrics, accurate ground truth labels, randomized data split, confusion matrix analysis, model complexity, feature importance. In this case of analysis, using a Python [71] sequence gives a categorical distribution (Table 3) and transforms each one to predict and calculate model accuracy in the following form.

The features in Figure 9 and the target variable (Y) are extracted from the dataset. It is assumed that "APT attribution" is the target variable with a split training and testing data using an 80–20 ratio. This score of 0.28 means that, compared to the calculated variable, APT attribution allowed a volume of 28% accuracy to subsequently carry out predicted actions and correctly assess an adjusted model against the trained dataset. In the case of obtaining other results [72], the accuracy is a common metric used to evaluate classification models, and it represents the ratio of correctly predicted instances to the total instances.

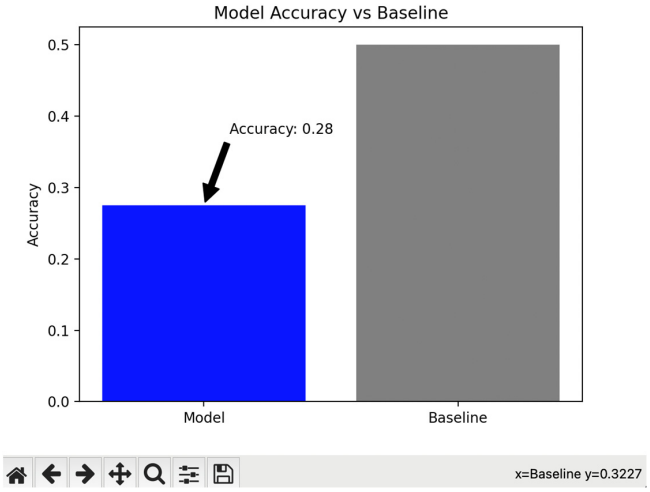

Figure 9. Accuracy diagram of DataFrame (DataSet).

In this case, the accuracy is compared with a "baseline" of the dataset.

4.1. Precision, Recall, and F1 Score (Separated)

1. Precision, or positive predictive value (PPV):
   - Precision (Figure 10) is defined as the proportion of accurately predicted positive instances relative to the total instances predicted as positive. In the given context, a precision value of 0.25 indicates that only 25% of instances predicted as positive were correctly identified as true positives.
2. Recall, or sensitivity/true positive rate (TPR):
   - Recall (Figure 10) represents the ratio of accurately predicted positive instances to the total number of actual positive instances. A recall value of 0.25 implies that the model successfully captured only 25% of all positive instances.
3. F1 Score:
   - The F1 score (Figure 10) is the harmonic mean of precision and recall, offering a balanced evaluation that accounts for both false positives and false negatives. With an F1 score of 0.25, the model attains a moderate equilibrium between precision and recall, reflecting a trade-off between these two metrics.

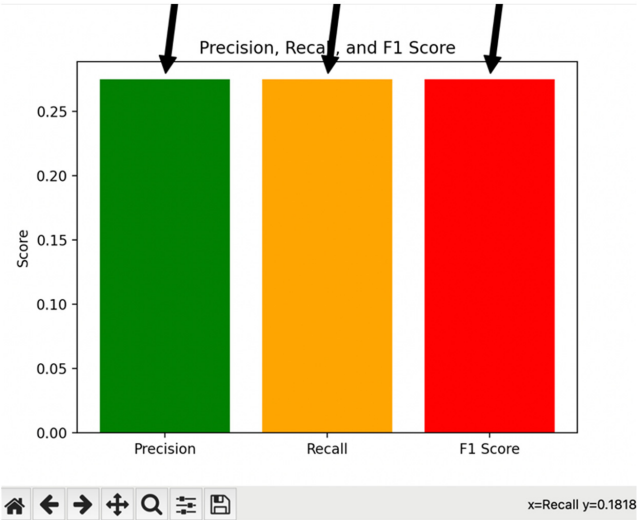

Figure 10. Precision, recall and F1 score of DataFrame (DataSet).

4.2. Application of Algorithms and Results

In parallel, an assessment of the veracity with an accuracy of 0.28 is conducted. Furthermore, a comprehensive examination of precision, recall, and F1 score is undertaken to juxtapose the computational outcomes across various classifiers or algorithms. The ensuing results are as follows:

Accuracy

The overall accuracy across the classifiers is 0.28, indicating a relatively low performance in predicting the target variable.

Precision, Recall, and F1 score:

1. **Random forest:**
   - Precision: 0.1707.
   - Recall: 0.275.
   - F1 Score: 0.2038.
   - *Interpretation:* The model exhibits low precision, capturing only a small proportion of true positive predictions. However, it achieves a balance between precision and recall (*excluded*).
2. **Logistic regression:**
   - Precision: 0.3898.
   - Recall: 0.325.
   - F1 Score: 0.3418.
   - *Interpretation:* The model demonstrates moderate precision and recall, striking a balance between each one.
3. **Naïve Bayes:**
   - Precision: 0.3057.
   - Recall: 0.275.
   - F1 Score: 0.2589.
   - *Interpretation:* The model exhibits moderate precision and recall, with a trade-off between each one.
4. **Perceptron:**
   - Precision: 0.4086.
   - Recall: 0.35.
   - F1 Score: 0.3658.
   - *Interpretation:* The model achieves a relatively higher precision and recall, demonstrating balanced performance.
5. ***k*-NN:**
   - Precision: 0.3144.
   - Recall: 0.275.
   - F1 Score: 0.27.
   - *Interpretation:* The model exhibits moderate precision, capturing a moderate proportion of true positive predictions with a trade-off between precision and recall.

4.2.1. Logistic Regression (RL) Validation

The contribution of application of logistic regression (LR) in anomaly detection audit trail data and any intrusion [3,12] is related with [20] ICS on IoT platforms; in this case, for this model, the hypothesis is defined as follows:

$$h\theta(x) = 1 + e - (\theta 0 + \theta 1x1 + \theta 2x2 + \ldots + \theta nxn)1 \quad (2)$$

The binary classification distributive $h\theta(x)$ is the probability of the positive class; $\theta 0, \theta 1, \ldots, \theta n$ is the model parameters (Figures 6 and 7); and $x1, x2, \ldots, xn$, are inputs of new features or established variables.

In a binary logistic regression classification (Figure 10), the model's predicted probability [72] distribution is visually depicted by a dashed line, anchored at an x-coordinate of 0.8668 and a y-coordinate of 0.56. This presentation effectively communicates the model's confidence in predicting the positive class [27]. The selection of the specific x-value, 0.8668, suggests a defined decision-making threshold, potentially optimized through an analysis of the receiver operating characteristic (ROC) curve. The corresponding y-value, 0.56, signifies the associated probability, offering valuable insights into the certainty of the classification [46,58]. This graphical representation significantly enhances the interpretability of the model's output, facilitating a clearer understanding of decision thresholds and supporting performance evaluation. These visualizations play a pivotal role in conveying the predictive behavior and performance of the model to a wider audience, promoting transparency and enabling more informed decision making. Figure 11 plots related "predicted probability" (mean true positive rate (TPR)) against the false positive rate (FPR) at various threshold values, providing a comprehensive visualization of the model's performance. The ROC curve allows for the evaluation of the model's discriminative ability and helps in selecting an appropriate threshold that balances between sensitivity and specificity based on the specific application's requirements in the dataset selected.

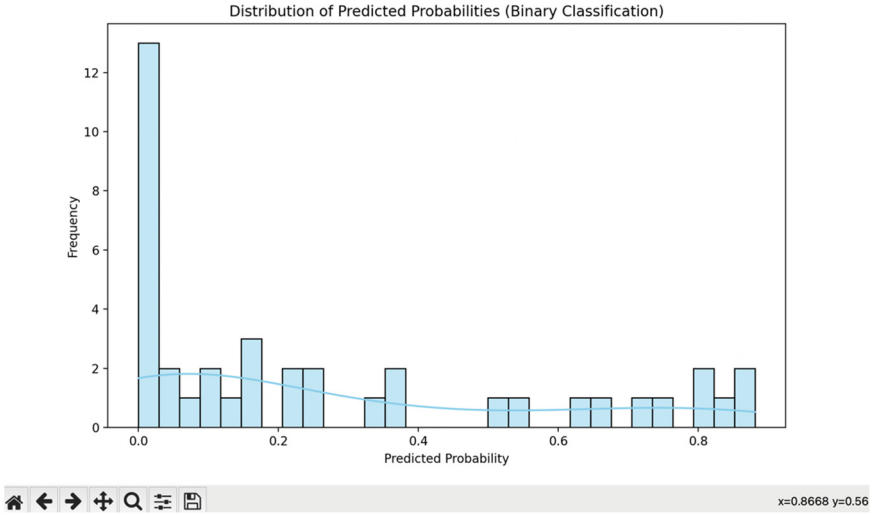

Figure 11. Logistic regression calculated in DataFrame (DataSet).

4.2.2. Naïve Bayes Confrontation

The problem of classification uses a feature vector X and a binary target variable Y (e.g., spam or not spam [73]) or detection network. The naïve Bayes classifier calculates the conditional probability of a specific class given the observed features using Bayes' theorem.

The formula applied with the model (Table 3, Figure 7) for naïve Bayes can be expressed as follows:

$$\mathrm{P(Y \mid X)} = \mathrm{P(X)P(X \mid Y) \cdot P(Y)} \tag{3}$$

The P value posterior probability of class Y given features X, $\mathrm{P}(y \mid \mathrm{X})$, indicate that $\mathrm{P}(\mathrm{X} \mid y)$ is the likelihood representing the probability of observing features X , given class Y, and $P(Y)$ is the prior probability of class Y. $P(X)P(X)$ is the evident probability, ensuring that the probabilities sum up to 1.

The "naïve" assumption (data model) is that features are conditionally independent given the class label, simplifying the likelihood term:

$$\begin{aligned} P(X \mid Y) &= P(X1 \mid Y) \cdot P(X2 \mid Y) \cdot \ldots \cdot P(Xn \mid Y)P(X \mid Y) \\ &= P(X1 \mid Y) \cdot P(X2 \mid Y) \cdot \ldots \cdot \end{aligned} \tag{4}$$

The explanation data in "naïve Bayes" [69] indicate (data model with a feature_1) that the Gaussian distribution, or Gaussian curve, is often employed to characterize the density of subcategories within a given data model. In this case, the distribution is derived from the density of each subcategory, and the exposed curve, defined by the coordinates X = 0.4950 and Y = 5.10, provides valuable insights into the probability factors associated with predictive modeling [12,37,43,45], and IoT [27] environments for coordinates (X = 0.4950, Y = 5.10) on the that curve represent key parameters that influence the probabilities involved in the detective and predictive process (Figure 12). The values presented in Figure 12 outline subcategory variables within the dataset, each of which is represented by the minimum calculations used to derive the representative sample. These interrelations are evident in the curvature observed in the Bayesian analysis.

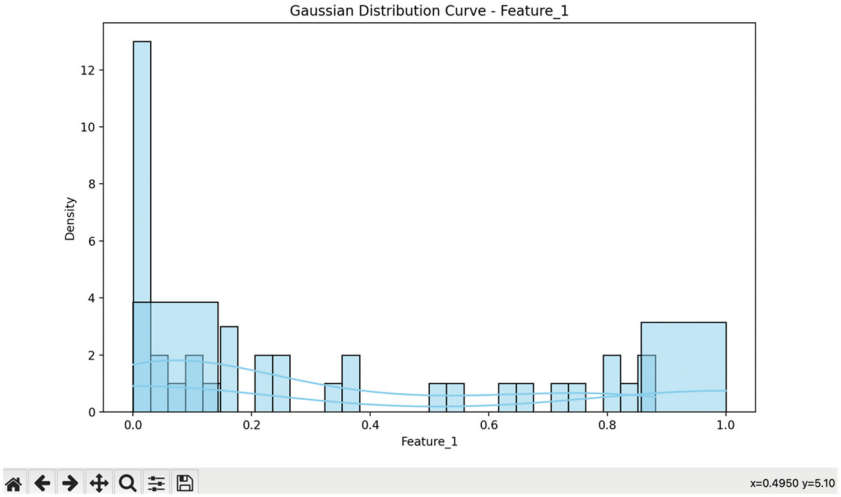

Figure 12. Naïve Bayes calculated in DataFrame (DataSet).

In relation to this calculated performance, various applications in deep packet inspection (DPI) utilize the characteristics extracted from the dataset. By comparing them with different traffic inputs in IoT networks, they enable the training of a neural network similar to the Message Passing Neural Network (MPNN) algorithm [74], thereby assessing the significance of the naïve Bayes calculation.

4.2.3. Perceptron Relations

In this perceptron algorithm, the model takes an input vector X and produces an output Y based on the weighted sum of its inputs as follows:

$$\begin{array}{l} \text{Input}: \ x = [x1, x2, \ldots, xn] \\ \text{Weight}: \ w = [w1, w2, \ldots, wn] \\ \text{Weight}: \ w = [w1, w2, \ldots, wn] \\ \text{Weighted sum}: \ z = \sum i = 1nwi \cdot xi \ \textit{Weighted Sum} : z = \sum i = 1nwi \cdot xi \end{array} \quad (5)$$

The function uses parameters (Figures 6 and 7) assessed through an activation function $f(z)$ for a perceptron that is the step function:

$$y = f(z) = \begin{cases} 1, & if\ z > threshold \\ 0, & Otherwise \end{cases} \quad (6)$$

In this case (Figure 13), Y involves the aggregation of weighted inputs (Zi) to generate a prediction. The perceptron, a foundational element within neural networks [21], employs weights (f) applied to input features (W) and computes their sum, e.g., as in Figures 2–4 and Tables 2 and 3, oriented with behavioral data [19,63] that are multivector factors of cyberattacks. In addition to precision of calculation assigned to these perceptron analyses,

many more inputs are required to facilitate the orientation through the sum of TTPs and attack patterns. In addition, various deep learning architectures are used, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) and multilayer perceptrons (MLPs) [75].

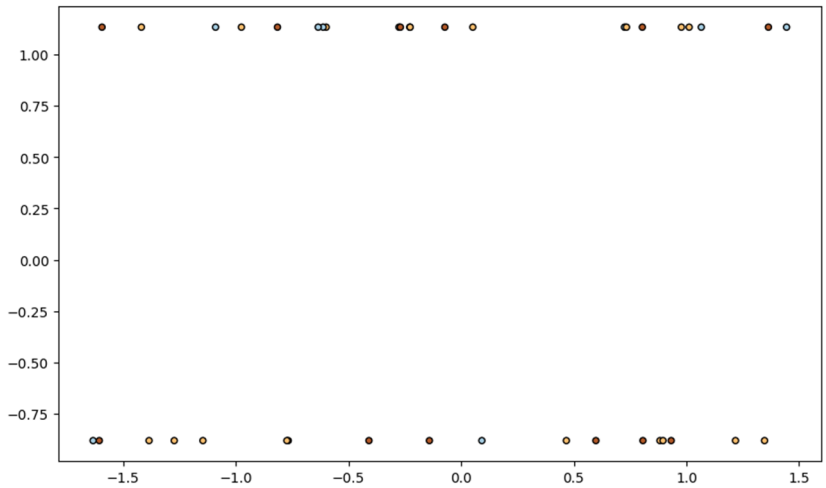

Figure 13. Perceptron calculated in DataFrame (DataSet).

4.2.4. K-Nearest (*k*-NN) Analyses and Results

The benefits of using the *k*-NN algorithm, e.g., for images analysis [76] or attack patterns of TTPs to determine risk [47] in models of APT patterns or APT attributions [23], can be the answer to guide the actions of cyberattackers against other types of cybercriminal behavior. In this scenario, the dataset has the necessary qualities and parameters to perform a validation, where $D = \{(X_1, y_2), (X_2, y_2), \ldots, (X_N, y_{nN})\}$, where X_i is a data point and y_i is its corresponding label.

We test data point X_q, for which we want to predict the label and convert in this formula:

$$d_{ij} = \sqrt{\sum_{k=1}^{d} (X_{ik} - X_{ik})^2} \quad (7)$$

This is in relation with a positive integer, k, representing the number of nearest neighbors to consider during prediction.

The (sample) *k*-NN calculated the following (Figure 14): (1) APT attack alerts that are correlated generating the APT attack scenarios and labeling APT attribution; (2) data generated in devices infected by APT attacks are essentially correlated to detect the aggressive malicious activities. The analysis has different observations [26] in the same timeline. According to these two CPSs cited in Figure 1, results of cyber-situation comprehension are composed of APT attack typologies and converting the relations in actions to detect an attacking scenario in real time. The last component aligns with the concept of threat hunting, where analysis of the behavior of compromised devices provides insights into the tactics, techniques, and procedures (TTPs) [33,51,59,60,62] with a family's features [26] and adversary's modus operandi.

According to Sections 4.2.1–4.2.3, each of these analyses aims to provide a perspective on using algorithms for calculating the characteristics obtained through the combined methodology outlined in Section 3.6. This section not only encompasses a combination of best practices but also enables the utilization of machine learning algorithms to enhance and strengthen the early detection of cyberattacks within the Colombian ecosystem, as illustrated in Figure 6. This process involves training models for future multivariate data analysis, thereby enabling the acquisition of more precise information for security and detection tools against anomalous patterns in cyberattacks targeting the digital infrastructure in Colombia.

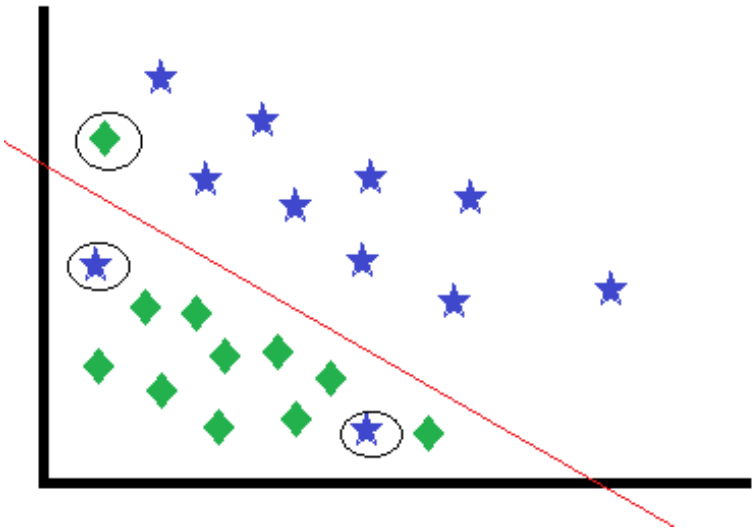

Figure 14. *k*-NN algorithm calculated in DataFrame (DataSet).

5. Conclusions and Limitations

As discussed in Sections 3 and 4, the APT behavior [27,38] connects sources by origin and destiny; depending on the design stage and history logs related with an RL, the TTP fingerprint is specifically developed and configured to detect advanced persistent threat (APT) attacks. In this case, the results were compared with a composition of each of the factors taken into account for analysis (Table 2 and Figure 4); once these two complementary elements were obtained, comparable parameters were obtained. Each one categorizes variables discussed in [48,49]; in this regard, they provide valuable information to compare the frequency, timing, and value of the cyberattacker's action and depth of the cyberattack in an IoT environment. In many cases, anomaly detection can also be derived from the types of IoT systems involved [77]. This proposal can be applied to multiple scenarios involving malicious traffic sensors.

The correlation of macro "items" (DM [48]), infrastructure, advanced persistent threat (APT) capabilities, decision-making values (OODA loop [49]), and the intricacies of their activities were discussed across each phase of the Cyber Kill Chain [47]. The correlation has an effect on measurement parameters in relation to machine learning selection [65]. As observed in the naïve Bayes (NB) analysis, by means of model sufficiency and data quantity (Table 3), positive prediction curvature mechanisms were obtained (Figures 11 and 12); this implies an element of confidence in the accuracy of the model and the applicability against a CPS in prevention stage or in detection stage for an IDS-IPS.

This study discusses the validation of each vulnerable IoT system and the innovative methodology for determining and facilitating the case study being validated, thereby expanding knowledge of the root causes of cyberattacks, the entry modality into defense or perimeter tools, and the model's capacity to receive combined information from various sources and enhance the comparative perspective of each anomalous pattern in cyberattacks in Colombia.

According to the discussion on the perceptron [46] and neighboring *k*-NN algorithms in both situations, data related to foresight and perspective actions were obtained [46]. The data model had an adequate accuracy [10] record (Figures 12 and 13) related to the number of inputs for information assimilation to be able to support larger sources of variables and parameters in the extension of supervised data and later unsupervised inputs approaches (deep learning).

This research contributes to the purpose of the criminal and cyberdefense investigation in Section 2.1 by meeting the required parameters in terms of technology validation in the context of cyberattack prevention [6,37] and the technical IoT [5] environment that addresses each of the influencing factors. The results show the strengths and weaknesses

of these techniques to address [34,64,66] the unique challenges posed by cyberattackers in Colombia's IoT infrastructure (Figures 6–8), in order to balance and contribute to the analysis of APT attribution cases, techniques, tactics, procedures, TTPs, OODA components, and Cyber Kill Chain (CKC) cited Section 3.3–3.5, and the new attributes of attack patterns in Colombia shown in Figure 6.

Limitations

The multivariate nature of the data means that machine learning methodologies are more limited in terms of the process of assigning input variables to train the model, and the data must be cleaned very frequently to obtain normalized and standardized data for each of the algorithms calculated.

Author Contributions: Conceptualization, J.R.B., J.A.S. and J.B.; Methodology, J.R.B. and J.A.S.; Software, E.O.-R.; Validation, J.R.B., J.A.S. and J.B.; Formal analysis, J.R.B., J.A.S. and J.B.; Investigation, E.O.-R., J.A.S. and J.B.; Writing—original draft, E.O.-R.; Writing—review & editing, J.R.B. and J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Parra, D.T.; Talero-Sarmiento, L.H.; Ortiz, J.D.; Guerrero, C.D. Technology readiness for IoT adoption in Colombian SMEs. In Proceedings of the 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), Chaves, Portugal, 23–26 June 2021; pp. 1–6. [CrossRef]
2. Russell, B. IoT Cyber Security. In *Intelligent Internet of Things: From Device to Fog and Cloud*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 473–512. [CrossRef]
3. Seifousadati, A.; Ghasemshirazi, S.; Fathian, M. A Machine Learning Approach for DDoS Detection on IoT Devices. *arXiv* **2021**, arXiv:2110.14911.
4. Pokhrel, S.; Abbas, R.; Aryal, B. IoT Security: Botnet detection in IoT using Machine learning. *arXiv* **2021**, arXiv:2104.02231.
5. Hassija, V.; Chamola, V.; Saxena, V.; Jain, D.; Goyal, P.; Sikdar, B. A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures. *IEEE Access* **2019**, *7*, 82721–82743. [CrossRef]
6. Jain, V.K.; Gajrani, J. IoT Security: A Survey of Issues, Attacks and Defences. In *Proceedings of the Intelligent Learning for Computer Vision: Proceedings of Congress on Intelligent Systems, New Delhi, India, 5–6 September 2020*; pp. 219–236. [CrossRef]
7. Xiao, L.; Wan, X.; Lu, X.; Zhang, Y.; Wu, D. IoT Security Techniques Based on Machine Learning: How do IoT devices use AI to enhance security? *IEEE Signal Process. Mag.* **2018**, *35*, 41–49. [CrossRef]
8. Dodda, R.; Gaddam, V.; Prasad, J.R.; Rao, B.V. The Evolution of Internet Of Things (IOT) And Its Impact on Existing Technology. *Int. J. Sci. Technol. Eng.* **2016**, *2*, 96–103.
9. Pennino, D.; Pizzonia, M.; Vitaletti, A.; Zecchini, M. Blockchain as IoT Economy Enabler: A Review of Architectural Aspects. *J. Sens. Actuator Netw.* **2022**, *11*, 20. [CrossRef]
10. An, Y.; Yu, F.R.; Li, J.; Chen, J.; Leung, V.C. Edge Intelligence (EI)-Enabled HTTP Anomaly Detection Framework for the Internet of Things (IoT). *IEEE Internet Things J.* **2020**, *8*, 3554–3566. [CrossRef]
11. Chatterjee, A.; Ahmed, B.S. IoT Anomaly Detection Methods and Applications: A Survey. *Internet Things* **2022**, *19*, 100568. [CrossRef]
12. Liang, F.; Hatcher, W.G.; Liao, W.; Gao, W.; Yu, W. Machine Learning for Security and the Internet of Things: The Good, the Bad, and the Ugly. *IEEE Access* **2019**, *7*, 158126–158147. [CrossRef]
13. Bharati, S.; Podder, P. Machine and Deep Learning for IoT Security and Privacy: Applications, Challenges, and Future Directions. *arXiv* **2022**, arXiv:2210.13547. [CrossRef]
14. Rashid, M.M.; Kamruzzaman, J.; Imam, T.; Kaisar, S.; Alam, M.J. Cyber Attacks Detection from Smart City Applications Using Artificial Neural Network. In Proceedings of the Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 16–18 December 2020; pp. 1–6. [CrossRef]
15. Chierzi, V.; Mercês, F. Evolution of IoT Linux Malware: A MITRE ATT&CK TTP Based Approach. In Proceedings of the 2021 APWG Symposium on Electronic Crime Research (eCrime), Boston, MA, USA, 1–3 December 2021; pp. 1–11.
16. Kolias, C.; Kambourakis, G.; Stavrou, A.; Voas, J.M. DDoS in the IoT: Mirai and Other Botnets. *Computer* **2017**, *50*, 80–84. [CrossRef]
17. Friedberg, I.; Skopik, F.; Settanni, G.; Fiedler, R. Combating Advanced Persistent Threats: From Network Event Correlation to Incident Detection. *Comput. Secur.* **2015**, *48*, 35–57. [CrossRef]

18. Kharchenko, V.; Sklyar, V. ENISA Documents in Cybersecurity Assurance for Industry 4.0: IIoT Threats and Attacks Scenarios. In Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, Metz, France, 18–21 September 2019; Volume 2, pp. 1046–1049.
19. Tsakalidis, G.; Vergidis, K.; Madas, M. Decision and Information Technologies (CoDIT)—Cybercrime Offences: Identification, Classification and Adaptive Response. In Proceedings of the 2018 5th International Conference on Control, Decision and Information Technologies (CoDIT), Thessaloniki, Greece, 10–13 April 2018; pp. 470–475. [CrossRef]
20. da Rocha, B.C.; de Melo, L.P.; de Sousa, R.T., Jr. A Study on APT in IoT Networks. In Proceedings of the 18th International Conference on e-Business (ICE-B 2021), Nanjing, China, 3–7 December 2021.
21. Stellios, I.; Kotzanikolaou, P.; Psarakis, M.; Alcaraz, C.; López, J. A Survey of IoT-Enabled Cyberattacks: Assessing Attack Paths to Critical Infrastructures and Services. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 3453–3495. [CrossRef]
22. NIST. *Guide to Operational Technology (OT) Security*; NIST: Gaithersburg, MD, USA, 2023.
23. Available online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r3.pdf (accessed on 1 January 2024).
24. Bertino, E.; Islam, N. Botnets and Internet of Things Security. *Computer* **2017**, *50*, 76–79. [CrossRef]
25. Cheng, X.; Zhang, J.; Chen, B. Cyber Situation Comprehension for IoT Systems Based on APT Alerts and Logs Correlation. *Sensors* **2019**, *19*, 4045. [CrossRef] [PubMed]
26. Singh, T.; Jabar, M.M. Exploration of Mobile Device Behavior for Mitigating Advanced Persistent Threats (APT): A Systematic Literature Review and Conceptual Framework. *Sensors* **2022**, *22*, 4662.
27. Kim, G.; Choi, C.; Choi, J. Ontology modeling for APT attack detection in an IoT-based power system. In Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems—RACS '18, Adaptive and Convergent Systems (RACS), Honolulu, HI, USA, 9–12 October 2018; pp. 160–164. [CrossRef]
28. OMDIA. 01 de Diciembre de 2023. 5G Forecast 2023–2028. 2023. Available online: https://www.5gamericas.org/resources/charts-statistics/latin-america/ (accessed on 1 January 2024).
29. Barrios, A.; Cama, D.; Mardini, J.; Díaz, J. Projections of IoT Applications in Colombia Using 5G Wireless Networks. *Sensors* **2021**, *21*, 7167. [CrossRef] [PubMed]
30. (TicTac), ICT Analysis and Creativity Tank. AI for Protection and Threat Prevention. 2023. Available online: https://www.ccit.org.co/estudios/estudio-anual-de-ciberseguridad-2022-2023/ (accessed on 1 January 2024).
31. Kaspersky, «Impacto TIC» 25 01 2024. [En Línea]. Available online: https://impactotic.co/tecnologia/157-000-ciberataques-diarios-en-colombia-en-el-2023/ (accessed on 30 January 2024).
32. QiAnXin Threat Intelligence Center 2023. Available online: https://ti.qianxin.com/blog/articles/apt-c-36-continuous-attacks-targeting-colombian-government-institutions-and-corporations-en/ (accessed on 1 January 2024).
33. Malpedia. 2023. Available online: https://malpedia.caad.fkie.fraunhofer.de/actor/apt-c-36 (accessed on 1 January 2024).
34. ESET WeliveSecurity. 2021. Available online: https://www.welivesecurity.com (accessed on 1 January 2024).
35. Available online: https://www.welivesecurity.com/2021/01/12/operation-spalax-targeted-malware-attacks-colombia/ (accessed on 1 January 2024).
36. Javed, S.H.; Ahmad, M.B.; Asif, M.; Almotiri, S.H.; Masood, K.; Ghamdi, M.A.A. An Intelligent System to Detect Advanced Persistent Threats in Industrial Internet of Things (I-IoT). *Electronics* **2022**, *11*, 742. [CrossRef]
37. Ma, Z.; Li, Q.; Meng, X. Discovering Suspicious APT Families Through a Large-Scale Domain Graph in Information-Centric IoT. *IEEE Access* **2019**, *7*, 13917–13926. [CrossRef]
38. Al-Kadhimi, A.A.; Singh, M.M.; Jabar, T. Fingerprint for Mobile-Sensor APT Detection Framework (FORMAP) Based on Tactics Techniques and Procedures (TTP) and Mitre. In Proceedings of the 8th International Conference on Computational Science and Technology: ICCST 2021, Labuan, Malaysia, 28–29 August 2022; Springer: Singapore, 2022; pp. 515–533.
39. Jiang, P.; Wu, H.; Wang, C. Virtual MAC Spoofing Detection Through Deep Learning. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
40. Pelletier, C.; Webb, G.I.; Petitjean, F. Deep learning for the classification of Sentinel-2 image time series. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.
41. Dong, S.; Xia, Y.; Peng, T. Network abnormal traffic detection model based on semi-supervised deep reinforcement learning. *IEEE Trans. Netw. Serv. Manag.* **2021**, *18*, 4197–4212. [CrossRef]
42. Di Mauro, M.; Galatro, G.; Liotta, A. Experimental review of neural-based approaches for network intrusion management. *IEEE Trans. Netw. Serv. Manag.* **2020**, *17*, 2480–2495. [CrossRef]
43. Ahanger, T.A. Defense Scheme to Protect IoT from Cyber Attacks Using AI Principles. *Int. J. Comput. Commun. Control* **2018**, *13*, 915–926. [CrossRef]
44. Kuzlu, M.; Fair, C.; Guler, O. Role of Artificial Intelligence in the Internet of Things (IoT) Cybersecurity. *Discov. Internet Things* **2021**, *1*, 7. [CrossRef]
45. Li, S.; Zhang, Q.; Wu, X.; Han, W.; Tian, Z. Attribution Classification Method of APT Malware in IoT Using Machine Learning Techniques. *Secur. Commun. Netw.* **2021**, *2021*, 9396141. [CrossRef]
46. Raschaka, S.; Mirjalili, V. *Python Mavhine Learning*, 1st ed.; Marcombo: Barcelona, Spain, 2019; ISBN 978-84-267-2720.
47. Hämäläinen, T.; Bodström, T. A Novel Method for Detecting APT Attacks by Using OODA Loop and Black Swan Theory. In Proceedings of the Computational Data and Social Networks: 7th International Conference, CSoNet 2018, Shanghai, China, 18–20 December 2018; Proceedings 7. Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 498–509.

48. Caltagirone, S.; Pendergast, A.; Betz, C. The Diamond Model of Intrusion Analysis. *Threat Connect* **2013**, *298*, 1–61.
49. Choi, J.J.; Choi, C.; Lynn, H.M.; Kim, P. Ontology-Based APT Attack Behavior Analysis in Cloud Computing. In Proceedings of the 2015 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), Krakow, Poland, 4–6 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 375–379.
50. Mohsin, M.; Anwar, Z. Where to Kill the Cyber Kill-Chain: An Ontology-Driven Framework for IoT Security Analytics. In Proceedings of the 2016 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 19–21 December 2016; pp. 23–28. [CrossRef]
51. Noor, U.; Shahid, S.; Kanwal, R.; Rashid, Z. A Machine Learning Based Empirical Evaluation of Cyber Threat Actors High-Level Attack Patterns over Low-Level Attack Patterns in Attributing Attacks. *arXiv* **2023**, arXiv:2307.10252.
52. Kim, K.; Alfouzan, F.A.; Kim, H.K. Cyber-Attack Scoring Model Based on the Offensive Cybersecurity Framework. *Appl. Sci.* **2021**, *11*, 7738. [CrossRef]
53. ElKashlan, M.; Aslan, H.; Azer, M.A. DDoS Attack Detection in IoT Using Machine Learning-Based Intrusion Detection System (IDS). In Proceedings of the 2022 18th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 28–29 December 2022; pp. 19–24.
54. Ngo, M.V.; Luo, T.; Chaouchi, H.; Quek, T.Q. Contextual-Bandit Anomaly Detection for IoT Data in Distributed Hierarchical Edge Computing. In Proceedings of the 2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS), Singapore, 29 November–1 December 2020; pp. 1227–1230.
55. Xia, Q.; Dong, S.; Peng, T. An Abnormal Traffic Detection Method for IoT Devices Based on Federated Learning and Depthwise Separable Convolutional Neural Networks. In Proceedings of the 2022 IEEE International Performance, Computing, and Communications Conference (IPCCC), Austin, TX, USA, 11–13 November 2022; IEEE: Piscataway, NJ, USA, 2022.
56. Open Data of General Prosecutor's Office of the Nation 2019–2023. Available online: https://www.datos.gov.co/browse?q=fiscalia%20spoa&sortBy=relevance/ (accessed on 12 January 2024).
57. Ghafir, V.; Prenosil, V.; Hammoudeh, M.; Aparicio-Navarro, F.J.; Rabie, K.; Jabban, A. Disguised Executable Files in Spear-Phishing Emails: Detecting the Point of Entry in Advanced Persistent Threat. In Proceedings of the 2nd International Conference on Future Networks and Distributed Systems, Amman, Jordan, 26–27 June 2018; pp. 1–5.
58. ThreatMon. *APT Blind Eagle's Malware Arsenal: Technical Analysis of the New Attack Chain*; TMRansommonitor, Vancouver: Sterling, TX, USA.
59. Available online: https://www.shodan.io/ (accessed on 1 January 2024).
60. Available online: https://www.zoomeye.org/ (accessed on 1 January 2024).
61. exploit-db. Available online: https://www.exploit-db.com/ (accessed on 1 January 2024).
62. packetstormsecurity.com. Available online: https://packetstormsecurity.com/ (accessed on 1 January 2024).
63. Lysenko, S.; Bobrovnikova, K.; Kharchenko, V.; Savenko, O. IoT Multi-Vector Cyberattack Detection Based on Machine Learning Algorithms: Traffic Features Analysis, Experiments, and Efficiency. *Algorithms* **2022**, *15*, 239. [CrossRef]
64. MISP. MISP Project. Available online: https://github.com/MISP (accessed on 1 January 2024).
65. Bout, E.; Loscrí, V.; Gallais, A. How Machine Learning Changes the Nature of Cyberattacks on IoT Networks: A Survey. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 248–279. [CrossRef]
66. CVSS, NIST–CVS-. Available online: https://nvd.nist.gov/Vulnerability-Metrics/Calculator-Product-Integration (accessed on 1 January 2024).
67. Turber, S.; Vom Brocke, J.; Gassmann, O.; Fleisch, E. Designing Business Models in the Era of Internet of Things: Towards a Reference Framework. In Proceedings of the 9th International Conference on Advancing the Impact of Design Science: Moving from Theory to Practice, (DESRIST 2014), Miami, FL, USA, 22–24 May 2014; Springer International Publishing: Berlin/Heidelberg, Germany, 2014; Volume 9, pp. 17–31.
68. McKinsey. Making Sense of Internet of Things Platforms. Available online: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/making-sense-of-internet-of-things-platforms (accessed on 1 January 2024).
69. Joyanes, L. *Internet of the Things*; AlphaEditoria: Madrid, Spain, 2021.
70. Nitin, G.; Shashank, M.; Hima, P.; Satoshi, M.; Naveen, P.; Sambaran, B.; Sameep, M.; Shanmukha, C.; Guttula, S.; Afzal, R.; et al. Data Quality for Machine Learning Tasks. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Virtual Event, 14–18 August 2021; pp. 4040–4041. [CrossRef]
71. Raschka, S.; Mirjalili, V. *Machine Learning and Deep Learning with Python, Scikit-Learn and TensorFlow*; Packt Publishing: Birmingham, UK, 2017; ISBN 9781787125933.
72. Sarker, I.H. IntruDTree: A Machine Learning-Based Cyber Security Intrusion Detection Model. *Symmetry* **2020**, *12*, 754. [CrossRef]
73. Huang, Y. Network Intrusion Detection Method Based on Naive Bayes Algorithm. In Proceedings of the 2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT) IEEE, Changzhou, China, 4–6 November 2022; pp. 1–10. [CrossRef]
74. Dong, S.; Li, R. Traffic identification method based on multiple probabilistic neural network model. *Neural Comput. Appl.* **2019**, *31*, 473–487. [CrossRef]
75. Naeem, H.; Cheng, X.; Ullah, F.; Jabbar, S.; Dong, S. A Deep Convolutional Neural Network Stacked Ensemble for Malware Threat Classification in Internet of Things. *J. Circuits Syst. Comput.* **2022**, *31*, 2250302. [CrossRef]

76. Ori, N.; Ayellet, T. k-NNN: Nearest Neighbors of Neighbors for Anomaly Detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops, Waikoloa, HI, USA, 4–8 January 2024.
77. Dong, S.; Su, H.; Xia, Y.; Zhu, F.; Hu, X.; Wang, B. A Comprehensive Survey on Authentication and Attack Detection Schemes That Threaten It in Vehicular Ad-Hoc Networks. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 13573–13602. [CrossRef]