Methodologies for Predicting Cybersecurity Incidents

Yaser M. A. Abualkas, D. Lalitha Bhaskari

***Abstract:** Data science may be used to detect, prevent, and address ever-evolving cybersecurity risks. CSDS is a fast developing field. When it comes to cybersecurity, CSDS emphasises the use of data, concentrates on generating warnings that are specific to a particular threat and uses inferential methods to categorise user behaviour in the process of attempting to enhance cybersecurity operations. Data science is at the heart of recent developments in cybersecurity technology and operations. Automation and intelligence in security systems are only possible through the extraction of patterns and insights from cybersecurity data, as well as the creation of data-driven models that reflect those patterns and insights An attempt is made in this work to describe the various data-driven research approaches with a focus on security. In accordance with the phases of the technique, each work that anticipates cyber-incidents is thoroughly investigated to create an automated and intelligent security system.*

***Keywords:** Cybersecurity, Machine Learning, Data Science, Decision Making, Cyber-Attack, Security Modeling, Intrusion Detection, Cyber Threat Intelligence, Data-Driven, Cybersecurity Incidents.*

I. INTRODUCTION

A. Overview of Cybersecurity Data Science

Data science has been used successfully in a number of fields to speed up the prediction and discovery of probabilistic outcomes. AI and machine learning are a mix of data science since the main purpose is to uncover patterns and extract correlations from raw data, as well as to process vast volumes of information to extract correlations with estimated probabilities and mistakes. Machine-based methods for extracting relevant information from large volumes of data. In other words, a branch of study that focuses on acquiring, storing, analysing, and extracting information about individuals and the larger environment. An advanced training programme in Data Science [1] teaches advanced data mining and programming abilities. There are several stages to Data Science. Data scientists must have a thorough awareness of the whole data science life cycle, as well as the ability to adapt and change as the process progresses in order to get the most out of their efforts.

For competent workers, Data Science [2] has emerged as one of the most promising and sought-after job pathways in the industry. Data Science [3] is a new branch of study that combines computer science and statistical technique to provide predictions and insights that may be used across a wide range of conventional academic disciplines. Data Science [4] is the study of data and the process by which information may be derived from it. A Data Scientist's primary responsibility is to build prediction models. Today, data science is relevant because we have so much data at our disposal. We weren't worried about a lack of information. Data is now at our fingertips. Previously, there were no clearly defined algorithms; now, there are. Because of the high cost of the programme in the past, it could only be used by industries with large budgets, but today it's open-source and free. The cost of data storage was prohibitive until now, but now that it is available for a fraction of the cost, we will have access to vast quantities of data at a very cheap cost. In addition, Internet access was scarce and prohibitively expensive. For example, the ability to work with numbers, as well as the capacity to store and analyse data, is crucial. Everything you need for connectivity is here, and it's inexpensive, accessible, and pervasive. This is the best time in history to be a knowledge scientist.

The most important aspect of any company's security measures will always be protecting its most sensitive data.. The world we live in now is one in which every piece of information is stored electronically or digitally. A secure haven to communicate with loved ones is provided by social networking sites. Cybercriminals still target social networking platforms to obtain personal information from home users. Security precautions must be taken not only on social media but also when making financial transactions. In some respects, the internet has shrunk the world, but it has also exposed us to a broader range of influences and responsibilities than ever before. The hacking world has recently grown as quickly as the security industry has grown. On the one hand, you can see how tough it is to protect yourself against cyberattacks. It's important to note that cloud computing firms are solely focused on providing this service, so they'll have the most up-to-date encryption technology at their disposal. Systems linked to the internet guard it against cyber-attacks and safeguard its data. Businesses use physical and cyber security to guard against unwanted access to data centres and other computerised systems when addressing security concerns. A subset of physical security, cyber security is concerned with protecting the confidentiality, integrity, and availability of data.

Manuscript received on 27 June 2022 | Revised Manuscript received on 21 March 2023 | Manuscript Accepted on 15 May 2023 | Manuscript published on 30 May 2023.

* Correspondence Author(s)

Yaser M.A. Abualkas*, Research Scholar, Department of Computer Science and Systems Engineering, College of Engineering, Andhra University, Visakhapatnam (A.P), India. E-mail: ykasking@gmail.com, ORCID ID: https://orcid.org/0009-0006-3338-1980

D. Lalitha Bhaskari, Professor, Department of Computer Science and Systems Engineering, College of Engineering, Andhra University, Visakhapatnam (A.P) India.

Retrieval Number:100.1/ijcns.F36770811622
DOI: 10.54105/ijcns.F3677.053123
Journal Website: www.ijcns.latticescipub.com

Published By:
Lattice Science Publication (LSP)
© Copyright: All rights reserved.

In the last ten years, the IT industry's primary emphasis has shifted to cyber security. Cybercrime is a huge concern for anyone in today's globe. People are extremely concerned about cyber-security breaches, which may result in everything from wholesale theft to blackmailing large corporations, as hackers are stealing important information from the government and a select enterprise enterprises. It is important for everyone to be aware of scams and the many methods and technologies that may be utilised to protect themselves against cyber-crimes. All enterprises that are concerned about the security of their sensitive data. Isn't only the loss of personal data, but the loss of client connections that comes with it? (Bendovschi, 2015).

Data-driven cybersecurity data science uses machine learning methods, aims to quantify cyber threats and promote inferential approaches to study behavioural patterns, focuses on creating security response alerts, and ultimately aims to optimize cybersecurity operations... Assault, damage or unauthorised access to computer systems is the goal of cybersecurity [5]. Protection of data on the internet The emergence of data science as a paradigm for discovery is one of the most significant breakthroughs of the early twenty-first century. cybersecurity data science is being used to every human effort that can be supported by sufficient data. Despite the impressive results, even more grandiose promises have been made. There are several advantages, difficulties, and dangers to consider. In order to create a security system that is both automated and intelligent, the science behind cybersecurity data science focuses on predicting and protecting data and systems. Data science is having a profound impact on the world's businesses. To ensure the long-term viability of intelligent cybersecurity systems and services, "Security is all about data." When trying to define cyber risks, we look at the safety data contained in files, logs, network packets, and other relevant sources. Traditionally, security professionals have not used data science approaches to develop detections that support these data sources. An alternative was to employ file hashes, special criteria like signatures, or manually set algorithms. Despite the fact that these tactics have their virtues in some situations, it requires an excessive amount of human labour to keep up with the always evolving cyber threat landscape. Data science, on the other hand, has the potential to fundamentally alter technology and how it is used. For example, machine learning algorithms frequently seek to discover and avoid security issue trends in the training data. These approaches can be used, for example, to detect malware or suspicious patterns, or to defend against it, or to extract policy rules.

II. CYBERSECURITY INCIDENT PREDICTION

The definition of a cybersecurity incident and a research technique for prediction are discussed in the first section of this chapter, brief overview of cybersecurity incident prediction in the second part. The key situation that motivated us to undertake this study is one that is described in the following section. While defining cybersecurity incident prediction, it also establishes the scope of this research. As an added benefit, the definition aids researchers in other subjects by outlining the approach used in this specific area.

A. Cybersecurity Incident Definition

Numerous meanings of the phrase "cybersecurity incident" exist in the scientific literature. As a result, it becomes more difficult to provide a standard description and taxonomy for occurrences while yet keeping the length of the survey report manageable. This is especially true for the term "event," which is used differently by different teams and projects. The following are some examples of definitions found in the literature: Australian Computer Emergency Response Team (AusCERT) defined an event as "any kind of network attack, computer-related crime, and therefore the misuse or abuse of network resources or access" "Any real or suspected adverse event relating to the security of computer systems or computer networks," according to SANS Institute [18] and Department of the Navy [19]. There are various ways to describe "incidents," yet they all have a lot in common, as seen by the preceding definitions. As a result, we have defined cybersecurity in this study.

The term "incident" refers to any occurrence, activity, or collection of data that is done with the goal of causing harm to cybersecurity. [26–[22] also define "cybersecurity incident" using the correct terminology. A taxonomy based on a list of single and defined words is a widely used strategy. For example, "viruses and worms," "unauthorized data copying," "logic bombs," and "denial-of-service" are all included in a list of 24 phrases used to describe a cybersecurity event by David and Karl [20]. Although this categorization is simple to execute, it is necessary to include encyclopedic volumes of specified terminology in order to cover all forms of cyber occurrences. Aside from that, accepting the meanings of some phrases is difficult. Because of this, some literature has used a set of categories to describe cyber events. The NIST Computer Security Incident Handling Guide lists several types of security incidents, including denial of service (DoS), malicious code, and unauthorized or improper use. [21]. In addition, other taxonomies focus on the action of a cybersecurity event [23] or the effect generated by the occurrences [24]. [23] Figure 1 depicts our suggested classification strategy for cybersecurity events based on the above taxonomies, which helps us better understand and describe occurrences, as well as explain their extent. NIST's cyber incident taxonomy is used to categorize each occurrence as either improper use, DoS, malicious code, or unauthorized access. This document also provides references to the relevant papers by David and Karl [20] in order to organize the episodes in each evaluated work into a "list of terms".

III. METHODOLOGIES IN CYBERSECURITY INCIDENT PREDICTION

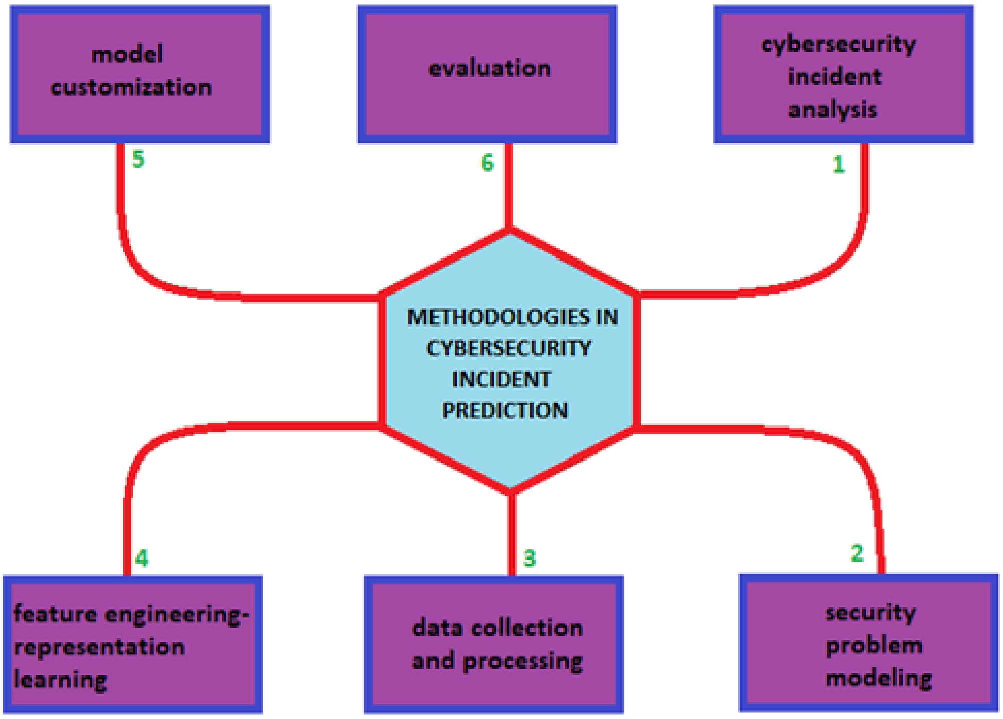

model customization
evaluation
cybersecurity incident analysis
5
6
1
METHODOLOGIES IN CYBERSECURITY INCIDENT PREDICTION
4
3
2
feature engineering-representation learning
data collection and processing
security problem modeling

Fig. 1. METHODOLOGIES IN CYBERSECURITY INCIDENT PREDICTION

Figure 1 depicts a popular methodology for predicting and discovering cybersecurity incidents. Six phases make up the model, which is an iterative and ongoing process: Cybersecurity incident analysis, security problem modeling and data collection and processing, feature engineering and representation learning, model customization, and assessment are all stages in the process.

1-Cybersecurity Incident Analysis:

The phrase "cyber incident analysis" refers to the process of establishing what happened, why it happened, and what can be done to prevent it from happening again. The degree of the damage caused by a cyber-attack may typically be assessed through a cyber incident analysis report. It is a crucial phase in responding to electronic events, and it sets the stage for later, opposing actions in the process. This means that the reaction plan is regarded a failure if it does not include an analysis of the situation. One of the most important things a soldier can do is "Get to know their enemy". The first thing we need to do in the face of an overwhelming number of cybersecurity events is to target one or more specific sorts of issues. Many efforts have been made to locate new indications that may link to events by completely evaluating an occurrence in order to reach the objective of anticipating and discovering cybersecurity incidents. Understanding a cybersecurity event is also useful in order to identify some component of the issue at which to focus our attention.

2-Security Problem Modeling:

In this stage, the research challenge molded by the project's needs should be established after analyzing previous target cybersecurity events. Defining the problem of cybersecurity incident prediction and discovery may be approached in one of two ways: either by developing a technique that can be used universally to all types of security events, or by focusing on specific types of security incidents and the data they contain. There are two ways to conduct proof of concept: one is to select a specific incident type, and the other is to investigate a specific prediction or discovery problem while analyzing an incident, like predicting whether a benign website will become malicious [6] or revealing black keywords employed by the online underground economy. In the basement, we'll build a model for resolving the problem at hand. As a means of establishing whether or not the idea is viable and practicable, a number of tests are conducted on a small number of implementation choices. A basic introduction to modeling methodologies and techniques for predicting cybersecurity incidents follows below. ML and DM are two of the most common techniques used to facilitate the development of security models. ML and DM have a lot in common since they both use a lot of the same techniques [9]. With regard to the prediction and identification of cybersecurity incidents, machine learning focuses on predictions based on training data. Current methods, such as those described in [8] and [10], can be used to break this problem down into two smaller problems that can be treated as binary classification problems. Multi-label classification is used when more than two labels can be assigned to each observation. As described in [14], regression is an option when the prediction issue is aimed to investigate the connection between variables..

On the other hand, data mining (DM) aims to uncover new information in the databases. To put it another way, a clustering problem is one in which a collection of items is grouped in such a way that objects within that group are more similar to one another than to those inside other groups [7], [16]. In addition to ML/DM, various approaches should be taken into account while building a model and addressing specific problems. Using NLP to analyse data like [12], [15]–[16] and [16–[17], for example, is common since security events are typically reported in natural language. Some projects also use statistical and graph mining [13] methods to attain their objectives. [13].

3) Data Collection and Processing:

Obtaining adequate data is the final stage after completing the first two processes outlined above. Data plays a critical role in predicting cybersecurity events. Gathering data is an important phase in this process because it serves as a link between the steps that come before and after it. The quality and amount of knowledge determine whether or not the research challenge suggested in the previous phase may be effectively solved.. The accuracy of a prediction model may also be affected by data, which serves as a source of ground truth. Since 2010, the number and diversity of security-related datasets have increased dramatically, giving us the ability to better anticipate and detect cyberattacks. Big data management begins with obtaining data from a wide variety of sources, depending on the research question and the goals of the project. Once we've gathered all the raw data, we need to figure out how to store it so that we can undertake additional analysis. In order to store data in the right databases or storage services, a physical foundation and cloud storage services are normally necessary [25]. When data is arranged and processed, knowledge may be discovered from databases. This comprises data cleanup, data mapping, data merging, and data conversion to structured forms, among other things. To get the best results, it's vital to have high-quality labels for data when using supervised learning, which may be the process of learning a function that maps an input to an output using examples [26]. Experts and professionals may be required to label some data. Our investigation of earlier work on data gathering has yielded some ideas on how to collect data for cybersecurity incident prediction. Certain types of information are readily available to researchers, while others are more challenging. In particular, data published on the Internet is frequently accessible using web crawlers. According to Verizon's yearly Data Breach Investigations Reports (DBIR), we may access the historical event report in the National Vulnerability Database (NVD) [27]. Social media data will be mined through the use of APIs given by major social media platforms like Twitter. Some data, especially when it comes to personal information, is more difficult to obtain.

4) Feature engineering:

is transforming data into features/attributes that better represent the underlying structure of your data, usually done by domain experts. Feature Extraction transforms data into the specified form. Features (characteristics, qualities, attributes, etc.) can be extracted from data using domain knowledge, or feature engineering (or feature extraction). Machine learning processes may be improved by providing more information, rather than simply providing data for them to work with. Feature Engineering / Representation Learning: Feature extraction from the obtained data is an essential step in the methodology's fourth phase, which is necessary for both generating optimum prediction outcomes and using machine learning. When using ML techniques, the choice of features or representation of knowledge has a significant impact on the method's performance[28]. The act of manually creating features by integrating domain knowledge with data is known as "feature engineering," and it significantly relies on this expertise. The process of feature engineering often begins with a brainstorming session during which data is analyzed and research challenges are considered. In some cases, we might draw inspiration from other works of literature or creative endeavors. There has been an overall reduction in the length of both automatic and manual feature extraction algorithms and methodologies. However, it might be difficult, time-consuming, and lacking in professional expertise [29] to come up with relevant features. It's difficult to extract precise features from real-world examples like photos or videos using typical feature engineering approaches. In contrast to feature engineering, representation learning seeks to identify and disengage any potential explanatory components hidden within the data [28]. Knowing how the data is represented may help classifiers and other prediction algorithms extract relevant information. In actuality, capturing the posterior distribution of the hidden components given the observed data may be a suitable representation of the probabilistic model. Several predictor types, such as supervised dictionary learning [30] and principal component analysis [31], as well as unsupervised and deep learning predictors, may all use an efficient representation as an input. Feature engineering or representation learning is a frequent technique in machine learning and deep learning. Examining components that may have been present before to an event is crucial for cybersecurity incident prediction. In other words, although the traits may not be directly related to the cybersecurity issue, they nonetheless need to be a warning sign of the danger. Liu et al. [8] forecasted a cybersecurity event on an organization-level granularity using criteria that indicated mismanagements on the network and infrastructure rather than recognized features. Representation learning may improve one's capacity for seeing cybersecurity issues and, therefore, for making accurate predictions based on certain prediction criteria.

5) Model Customization:

The simplest model possible is built using machine learning and data processing techniques, and it is then optimized to fit the data. Traditional machine learning and data mining (ML/DM) algorithms can perform well in some instances. To be sure, the same cannot be said for a specific research question related to cybersecurity incident prediction. Instead of relying on prepackaged solutions to overcome this issue, researchers may get the most out of their data by customizing standard ML/DM algorithms to fit their specific needs. Consequently, the predictive model's efficiency and efficacy can be greatly improved as a result of these changes.

Retrieval Number:100.1/ijcns.F36770811622
DOI: 10.54105/ijcns.F3677.053123
Journal Website: www.ijcns.latticescipub.com

Published By:
Lattice Science Publication (LSP)
© Copyright: All rights reserved.

Deep learning (DL), a machine learning strategy that depends on learning data representations, is progressing past its early successes in pattern recognition and toward inventive applications in other fields and industries [32]. There is a potential trend of using great computing efficiency to evaluate enormous amounts of data [33]. Researchers have recently used DL to identify spam [34], cover vulnerability, and have introduced DL for IoT-based systems [35], [36] in the cybersecurity arena. In order to solve cybersecurity incident prediction concerns, we must use and tailor deep learning in a sensible way. In classical computer science, the mastery of basic ML/DM methods and data structures provides a foundation for exploring the approaches and insights of customizing models. Improvements also tend to be much more specialized when linked with a particular research problem.

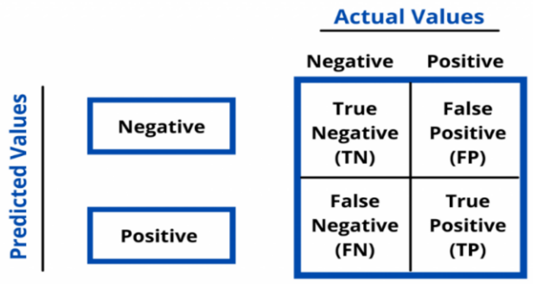
6) Evaluation:

In the final phase, we evaluate the model to see if the findings are in line with our goals. In other words, using proper measurements to verify that the study objectives are met. Confidence matrices are used to calculate evaluation metrics, such as FP, FN, TR, and TN, which are represented in the figure below

The evaluation metrics frequently used for reviewed work are:

• Accuracy — It is calculated as (TP + TN)/(TP + TN + FP + FN) as the percentage of items correctly predicted out of the total number of items.

• Recall — In other words, it is known as sensitivity or TPR. TP/(TP + FP) is the proportion of class X values that were correctly predicted as belonging to class X.



		Actual Values	
		Negative	Positive
Predicted Values	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

CONFUSION MATRIX

• Precision — It shows the percentage of all items classified as X that were correctly predicted as having the TP/(TP + FN) ratio.

• False Positive Rate (FPR) — Indicating the percentage of items incorrectly classified as a class n, or X, which is calculated as FP/(TN+FP).

• F-measure— Precision and recall are combined into a second measurement of accuracy, which is calculated as 2 Precision Recall/ (Precision + Recall). Use the ROC curve to show how well a binary predictor performs in predicting future results. There are relative trade-offs between benefits and costs (TP) shown by plotting TPR as the y-axis and FPR as a the x-axis of a ROC space (FP). The Receiver Operating Characteristic (ROC) curve can also be used to demonstrate the predictive power of the binary predictor. For example, by plotting the x-axis against the y-axis, we can see how benefits (TP) and costs (FPR) are traded off in a ROC space (FP).

When the FPR is present in a cybersecurity environment, it can markedly reduce the effectiveness of security technologies [44, 45]. When it comes to the detection problem, FPs invariably lead to significant costs. For example, if malware is discovered in a certain piece of software, the operation must be halted. As a result, the detection problem's primary objective is to increase TPR while minimizing overall size. However, in the prediction task, FPs are more difficult to prevent, but their value is smaller than in the detection problem. Additionally, this is in accordance with the purpose of prediction: to uncover all probable problems, prioritize warnings, and provide security training in advance. According to recent research, an insurance firm can accept 20% of false positives. [11].

• True Positive: truth value is positive and therefore the test predicted that it had been positive, or an individual is sick and therefore the test shows it.

• True Negative: truth value is negative and therefore the test predicted that the result was negative, or the person isn't sick and therefore the test shows it.

• False Negative: truth value is positive, and therefore the test predicted that the result's negative. The person is sick, but the test incorrectly says they're not. This is what's known in statistics as type II error

• False Positive: The true value is negative, and the test predicted that the result is positive. The person isn't sick, but the test incorrectly tells us that he's.

IV.METHODOLOGIES IN CYBERSECURITY DATA SCIENCE INCIDENT PREDICTION.

Data analysis: the gathering, transformation, and organization Know lessons to conclude make predictions, and drive informed decision-making. in this section IV, methodologies of data science incident prediction. Data analytics is such a lot quite just plugging information into a platform to seek out insights. It is about solving problems. To get to the basis of those problems and find practical solutions, there are many opportunities for creativity. No matter the matter, the primary and most vital step is knowing it. From there, it is good to take a problem-solver approach to your analysis to help you decide what information needs to be included, how you can transform the data, and how the data will be used.

Data analysis incident prediction methodology typical work with six approach types as described below.

1. Making predictions

Using data to make educated judgments on how things could be in the future. An example of a drag that necessitates analysts making projections is a business that wants to know the easiest advertising approach for bringing in new clients. Analysts with data on location, sort of media, and the number of latest customers acquired as a result of past ads can't guarantee future results, but they will help predict the best placement of advertising to succeed in the audience.

2. Categorizing things

Grouping data based on common features. An example of a drag requiring analysts to categorize things may be a company's goal to enhance customer satisfaction. Analysts might classify customer service calls supported by certain keywords or scores. This could help identify top-performing customer service representatives or help correlate certain actions to higher customer satisfaction scores.

3. Spotting something unusual

Identifying data that is different from the norm. It would be in the best interest of a firm that sells smartwatches to people to assist them in tracking their health to construct their software in such a way that it could detect anything that was out of the usual. Analysing aggregated health data may assist product developers establish the best algorithms to recognize and reject warnings when specific data does not trend normally.

4. Identifying themes

Recognizing broader concepts and trends from categorized data. Many UX designers may mistakenly assume that analysts are tasked with gathering data on user interactions. To help prioritize which product aspects should be improved, usability enhancement initiatives may necessitate the assistance of analysts in identifying themes. In most cases, themes are not useful for academics who want to investigate specific areas of knowledge. There are several topics that may be found in a user research. By now, you're probably thinking if there's a difference between classifying things and recognizing themes. In other words, the best way to think of it is this: classifying things entails assigning them to categories, and defining themes requires taking those categories and combining them into broader themes.

5. Discovering connections

Identifying similar challenges across different entities—and using data and insights to find common solutions. Analysts may need to establish links with a third-party logistics provider working with another firm to ensure timely delivery of shipments to clients. By analyzing the wait times at shipping hubs, analysts can determine the acceptable schedule changes to extend the amount of on-time deliveries.

6. Finding patterns

Using facts from the past to predict how likely it is that the same event will occur in the future. Minimizing downtime caused by machine failure is an example of a drag requiring analysts to seek out patterns in data. For example, by analyzing maintenance data, they could discover that the majority of failures happen if regular maintenance is delayed by quite a 15-day window.

In this section, a six phases methodology is explained which will be suitable for incident prediction approaches in cybersecurity. Cybersecurity research methodology in cybersecurity data science passed in six phases as explained below mention from Ask questions to make Data-Driven Decision course by Google[37].

Phase 1

Ask: Define the problem and confirm stakeholder expectations, prepare the infrastructure for future work

Phase 2

Prepare: Collect from different resources and store data for analysis and make it available when needed, authentication and authorization

Phase 3

Process: Clean and transform data to ensure integrity to improve your data quality and in doing so, increase overall productivity.

Phase 4

Analyze: Use data analysis tool conclusions using algorithms by taking the data after cleaning to get the result

Phase 5

Share: Interpret and communicate results to others and compare the results to define the proper algorithm to make data-driven decisions.

Phase 6

Act: Put your insights into a tool to solve the original problem and LEM, and train the algorithms for unusual behavior such iron as block action or another action.

V. FUTURE WORK

Recommended to researchers to discuss how to protect companies from cyber-attacks, which are increasing day by day, and to clarify all the most common attacks and vulnerabilities that are likely to affect individuals and organizations, through the creation of highly efficient algorithms to counter the attacks that occur from time to time.

VI. CONCLUSION

In view of the growing significance of cybersecurity and data science techniques and methods, this research examines data science, machine learning, and artificial intelligence techniques and methods. The method for forecasting each sort of cybersecurity event is discussed above for cybersecurity, data science, and cybersecurity data science, and it is based on the details given in those sections. The goal is to figure out how to forecast cybersecurity events using data science approaches by looking at actual security incident data and the services that go along with it. The goal of this article is to provide readers with an overview of how cybersecurity, data science, and incident prediction are conceptualized, understood, modeled, and thought about.

ACKNOWLEDGMENT

The authors appreciate the anonymous reviewers for their insightful critique, and they give particular recognition to Dr. D. Lalitha Bhaskari (Andhra University) for her assistance in preparing this work.

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Yes, It is relevant. Availability of Data and Material tells the reader where the research data associated with an article is available, and under what conditions the data can be accessed as mentioned in reference.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. What is data science?: a complete data science tutorial for beginners [Blog]. Retrieved 8. 10. 2019 from https://data-flair.training/blogs/what-is-datascience/. 3. Dhar, V. (2013).
2. Brodie, M.L. (2015). Understanding Data Science: An Emerging Discipline for Data-Intensive Discovery, in Shannon Cutt (ed.), Getting Data Right: Tackling the Challenges of Big Data Volume and Variety, O'Reilly Media, Sebastopol, CA, USA, June 2015.
3. Gregory Piatetsky, KDnuggets, https://www.kdnuggets.com/tag/data-science
4. Harvard Data Science Initiative https://datascience.harvard.edu
5. ax J, Sanders H. Malware data science: Attack detection and attribution, 2018.
6. K. Soska and N. Christin, "Automatically detecting vulnerable websites before they turn malicious," in Proc. USENIX Security Symp., 2014, pp. 625–640.
7. K. Borgolte, C. Kruegel, and G. Vigna, "Delta: Automatic identification of unknown Web-based infection campaigns," in Proc. ACM SIGSAC Conf. Comput. Commun. Security, 2013, pp. 109–120. [CrossRef]
8. Y. Liu et al., "Cloudy with a chance of breach: Forecasting cyber security
9. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Commun. Surveys Tuts., vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016. [CrossRef]
10. G. Lin et al., "Cross-project transfer representation learning for vulnerable function discovery," IEEE Trans. Ind. Information., vol. 14, no. 7, pp. 3289–3297, Jul. 2018. [CrossRef]
11. L. Bilge, Y. Han, and M. Dell'Amico, "RiskTeller: Predicting the risk of cyber incidents," in Proc. ACM SIGSAC Conf. Comput. Commun. Security, 2017, pp. 1299–1311. [CrossRef]
12. J. Zhang, Z. Durumeric, M. Bailey, M. Liu, and M. Karir, "On the mismanagement and maliciousness of networks," in Proc. Symp. Netw. Distrib. Syst. Security (NDSS), 2014, pp. 1–12. [CrossRef]
13. R. A. Rossi, B. Gallagher, J. Neville, and K. Henderson, "Modeling dynamic behavior in large evolving graphs," in Proc. 6th ACM Int. Conf. Web Search Data Min., 2013, pp. 667–676. [CrossRef]
14. S. Banescu, C. Collberg, and A. Pretschner, "Predicting the resilience of obfuscated code against symbolic execution attacks via machine learning," in Proc. 26th USENIX Security Symp., 2017, pp. 661–678. [CrossRef]
15. D. Kong, L. Cen, and H. Jin, "AUTOREB: Automatically understanding the review-to-behavior fidelity in Android applications," in Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Security, 2015, pp. 530–541. [CrossRef]
16. R. P. Khandpur et al., "Crowdsourcing Cybersecurity: Cyber attack detection using social media," in Proc. ACM Conf. Inf. Knowl. Manag., 2017, pp. 1049–1057. [CrossRef]
17. AusCERT Team. Russert Is a Leading Cyber Emergency Response Team (CERT) in Australia and the Asia/Pacific Region. Accessed: Apr. 3, 2018. [Online]. Available: http://www.auscert.org.au/
18. TS Institute. Computer Security Incident Handling Step-by-Step. Accessed: Apr. 3, 2018. [Online]. Available: https://www.sans.org/ reading-room/whitepapers/incident/incident handlers-handbook-33901
19. G. Killcrece, K.-P. Kossakowski, R. Ruefle, and M. Zajicek, "State of the practice of computer security incident response teams (CSIRTs)," CSIRT Develop. Team, Waldorf, MD, USA, Rep. CMU/SEI-2003-TR-001, 2003. [CrossRef]
20. I. David and S. Karl, Computer Crime: A Crime Fighter's Handbook. Sebastopol, CA, USA: O'Reilly Assoc., 1995.
21. T. Grance, K. Kent, and B. Kim, "Computer security incident handling guide," document SP 800-61, NIST, Gaithersburg, MD, USA, 2004. [CrossRef]
22. W. R. Cheswick, S. M. Bellovin, and A. D. Rubin, Firewalls and Internet Security: Repelling the Wily Hacker. Boston, MA, USA: Addison-Wesley, 2003.
23. W. Stallings, Network and Internetwork Security: Principles and Practice, vol. 1. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.
24. F. B. Cohen and F. B. Cohen, Protection, and Security on the Information Superhighway. New York, NY, USA: Wiley, 1995.
25. J. Li, Y. Zhang, X. Chen, and Y. Xiang, "Secure attribute-based data sharing for resource-limited users in cloud computing," Comput. Security, vol. 72, pp. 1–12, Jan. 2018. [CrossRef]
26. S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach. Kuala Lumpur, Malaysia: Pearson Edu. Ltd., 2016.
27. 2018 Verizon Annual Data Breach Investigations Report. Accessed: Sep. 13, 2018. [Online]. Available: https://www.verizonenterprise.com/ verizon-insights-lab/dbir/
28. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, Aug. 2013. [CrossRef]
29. A. Ng. (2013). Machine Learning and AI Via Brain Simulations. Accessed: May 3, 2018. [Online]. Available: http://ai.stanford.edu/Ëoeang/slides/DeepLearning-Mar2013.pptx
30. J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach, "Supervised dictionary learning," in Proc. Adv. Neural Inf. Process. Syst., 2009, pp. 1033–1040.
31. S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," Chemometrics Intell. Lab. Syst., vol. 2, nos. 1–3, pp. 37–52, 1987. [CrossRef]
32. S. Tokui, K. Oono, S. Hido, and J. Clayton, "Chainer: A next-generation open-source framework for deep learning," in Proc. Workshop Mach. Learn. Syst. (LearningSys) 29th Annu. Conf. Neural Inf. Process. Syst. (NIPS), vol. 5, 2015, pp. 1–6.
33. M. M. Najafabadi et al., "Deep learning applications and challenges in big data analytics," J. Big Data, vol. 2, no. 1, p. 1, 2015. [CrossRef]
34. B. Feng, Q. Fu, M. Dong, D. Guo, and Q. Li, "Multistage and elastic spam detection in mobile social networks through deep learning," IEEE Netw., vol. 32, no. 4, pp. 15–21, Jul./Aug. 2018. [CrossRef]
35. H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," IEEE Netw., vol. 32, no. 1, pp. 96–101, Jan./Feb. 2018. [CrossRef]
36. L. Li, K. Ota, and M. Dong, "When weather matters: IoT-based electrical load forecasting for smart grid," IEEE Commun. Mag., vol. 55, no. 10, pp. 46–51, Oct. 2017. [CrossRef]
37. Ask questions to make a Data-Driven Decision course by Google.

AUTHOR PROFILE

Yaser M A Abualkas Was born in Palestine – Gaza – on 1995 16th of April, a Ph.D. research scholar at Andhra University in the department of computer science and system engineering in College of Engineering (A) – Visakhapatnam – India. I finished my M.Tech at Andhra University in the department of computer science and system engineering at the College of Engineering (A) – Visakhapatnam – India. I finished my B.Tech at Al-Azhar University – Gaza –Palestine in the college of Engineering in Software System Engineering.
2 Journal Articles published

D. Lalitha Bhaskari, Professor & Dean, IQAC at Andhra University, around 70 Journal articles, 2 Book chapters, 3 books, 27 Conference Proceedings, and 2 Projects. Interests mainly include Security, Theory of Computation, Image Processing, Network Security & Cryptography.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Lattice Science Publication (LSP)/ journal and/ or the editor(s). The Lattice Science Publication (LSP)/ journal and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.