International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Explore the Application of Predictive Analytics and Machine Learning Algorithms in Identifying and Preventing Cyber Threats and Vulnerabilities within Computer Systems

Ranadeep Reddy Palle

Independent Researcher
Software Engineer

Abstract: ***In the introduction section, the study provided a well - developed aim as well as research objectives and research questions associated with the topic. In that case, the rise in cybersecurity issues throughout the world is a problem statement for this study. In the literature review, this study provides several examples of Machine Learning Models that help in the security effectiveness enhancement segment. In that case, Predictive analytics as well as the contribution of different algorithms has been discussed in the Cybersecurity landscape. Different strategies have been implemented to increase the prevention rate against cybersecurity threats in the modern era. In the methodology section, this study used a primary approach where primary data collection has been performed. In that case, the online survey offered all the data and these data were offered by 70 participants as per the asked questions. Quantitative data analysis method performed for data analysis. In the findings and analysis section, the IBM SPSS tool was used for demographic analysis. Similarly, this tool offered a hypothesis test by performing regression analysis and correlation analysis. In the discussion section, all the findings are discussed in which collaborating ML models with predictive analytics enhances cybersecurity measures. In the end, it can be concluded that different applications of predictive analytics improve the security of organisations in the evolving digital era.***

Keywords: Cybersecurity, Predictive Analytics, Machine Learning, Threat Detection, Vulnerability Prevention, Cyber Threats, Computer Systems etc.

1. Introduction

In this 21st century, the digital revolution has brought about connectivity, efficiency and innovation. However, along with these advancements, there has also been a rise in the threat of cyber - attacks and vulnerabilities. As a result, ensuring the security and protection of computer systems and sensitive information has become crucial in the field of cybersecurity. On the other hand, Chinedu et al. (2021) stated that traditional measures such as firewalls and antivirus software have proven to be ineffective against the evolving tactics of cybercriminals. Therefore, there is a need for a paradigm shift in security strategies to address the growing complexity of threats like malware, ransomware and advanced persistent threats (APTs). Revolutionary technologies like analytics and machine learning have emerged as solutions for enhancing the detection and prevention of cyber threats.

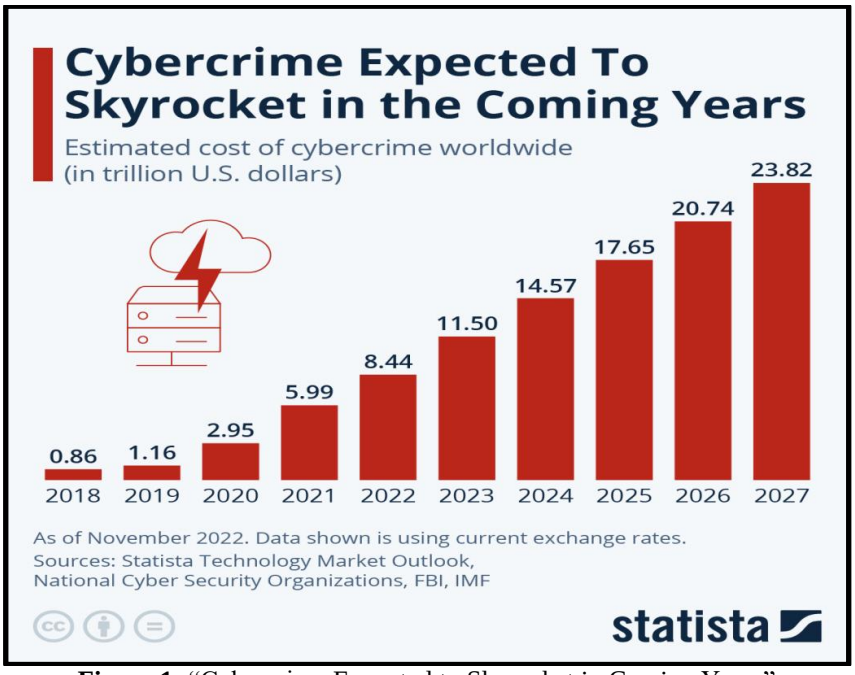

Figure 1: "Cybercrime Expected to Skyrocket in Coming Years" (Source: Statista. com, 2022)

Statista's Cybersecurity outlook has estimated that cybercrime will become more expensive over the next five years. The cost is predicted to rise from 8.44 trillion US dollars in the financial year 2022 to 23.84 trillion US dollars by fiscal year 2027 (statista. com, 2022). In that case, by using historical data as well as statistical algorithms and machine learning models, predictive analytics identify future outcomes. When it comes to cybersecurity, machine learning has the ability to identify risks by identifying patterns and irregularities within datasets (Alloghani et al.2020). By recognizing diversified patterns and anomalies, machine learning enables systems to gain knowledge from data, adjust accordingly and enhance their performance. On the other hand, predictive analytics and machine learning have a wide range of applications in cybersecurity. These applications are related to proactive threat intelligence and mitigation. These applications are not only reacting against known threats but also provide strengthened defences.

The aim of this research is to explore and assess how predictive analytics and machine learning algorithms can be used to detect and mitigate cyber threats and vulnerabilities in computer systems.

The objectives of this research are;

- To analyse the principles and techniques of predictive analytics in cybersecurity contexts.
- To explore different machine learning algorithms that can be used for cybersecurity purposes.
- To create predictive models and devise machine learning strategies to prevent cyber vulnerabilities.
- To evaluate the effectiveness of predictive models and machine learning algorithms in real - world scenarios.

The research Questions are;

- What are the key principles underlying predictive analytics in cybersecurity that enhance threat detection?
- Which machine learning algorithms are most effective in adaptive threat detection?
- What are the strategies that would help in preventing cyber vulnerabilities?
- What is the accuracy, efficiency and effectiveness of the developed models in identifying and preventing cyber threats?

2. Literature Review

The key principles underlying predictive analytics in cybersecurity that enhance threat detection

In this modern era, the use of predictive analytics in cybersecurity is based on important principles that can help in detecting threats. The process starts with gathering and analysing data from sources including time and historical information (Husák et al.2021). This comprehensive approach allows for the identification of patterns and anomalies that could indicate security risks. To ensure models it is essential to carefully select and transform relevant attributes through feature engineering enabling the generation of meaningful insights.

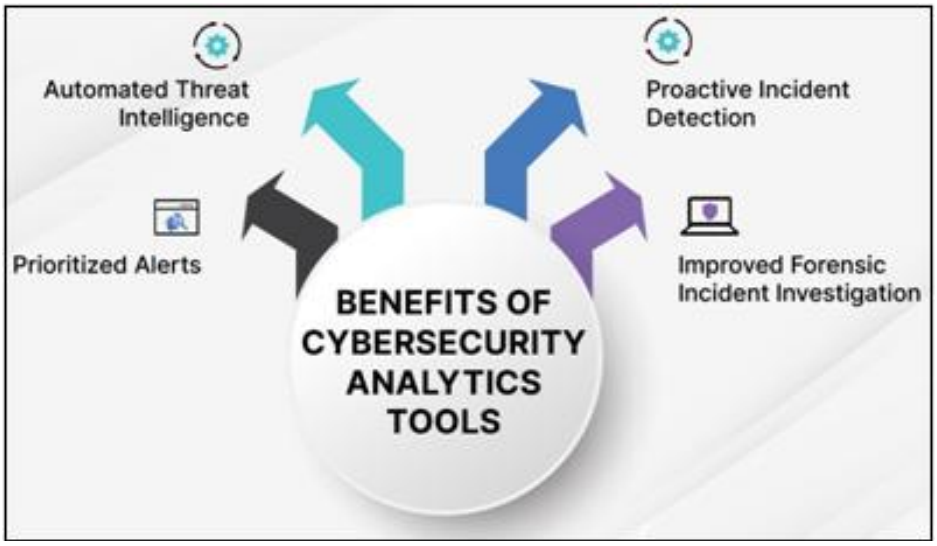

Figure 2: Predictive analytics in cybersecurity
(Source: Benzaïd & Taleb, 2020)

When it comes to cybersecurity data choosing the machine learning algorithms is extremely important. Different types of ML models such as Decision trees as well as forests, support vector machines and neural networks all have their distinct advantages in capturing different patterns of data threats (Trivedi et al.2020). To identify activities effectively, it is crucial to be able to detect anomalies or deviations from patterns. This principle of anomaly detection plays a role in this regard. On the other hand, to detect patterns it is possible to use analysis in combination with predictive analytics. This approach offers an effective understanding of both user and system behaviours. To effectively counter evolving cyber threats, constant learning and adaptation are important (Benzaïd & Taleb, 2020). Similarly, this ensures that the predictive models associated with machine learning stay up - to - date and effective. By integrating threat intelligence feeds, it can enhance the different models with

real - time data that not only increases their capability but also identifies arising threats.

Effective Machine learning algorithms in adaptive Threat detection

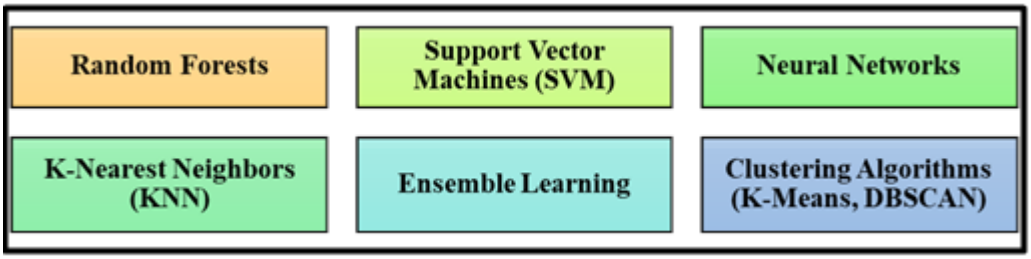

Figure 3: Most effective machine learning algorithms in adaptive threat detection
(Source: Naicker, Adeliyi & Wing, 2020)

In terms of threat detection, there are machine learning models that play a role. According to a study conducted by González et al. (2020), an ensemble learning approach called Random Forests has proven to be highly effective in handling datasets. This method not only detects anomalies and patterns but also identifies threats within the dataset. On the other hand, Naicker, Adeliyi & Wing (2020) mentioned that Support Vector Machines (SVM) are renowned for their ability to identify relationships in data and excel in binary classification tasks. Additionally, when it comes to capturing patterns and evolving cyber threats neural networks are indispensable. KNN also known as K Nearest Neighbours is a classification algorithm known for its simplicity and effectiveness. It categorises data points by analysing the majority class of their neighbours making it particularly valuable in scenarios where the normal behaviour structure is clearly defined. Similarly, decision trees serve as a machine learning model that's easily interpretable and comprehensible. Decision trees are a tool for recognizing patterns in cybersecurity data and aiding decision - making by analysing data based on features.

Ensemble learning involves utilising a combination of models to build a more robust classifier. Clustering algorithms are employed to group data points based on their similarities aiding in the identification of anomalies. On the other hand, Autoencoders leverage unsupervised learning techniques to detect threats and extract meaningful features from the data (Wei, Chow & Yiu, 2020).

The strategies that would help in preventing cyber vulnerabilities

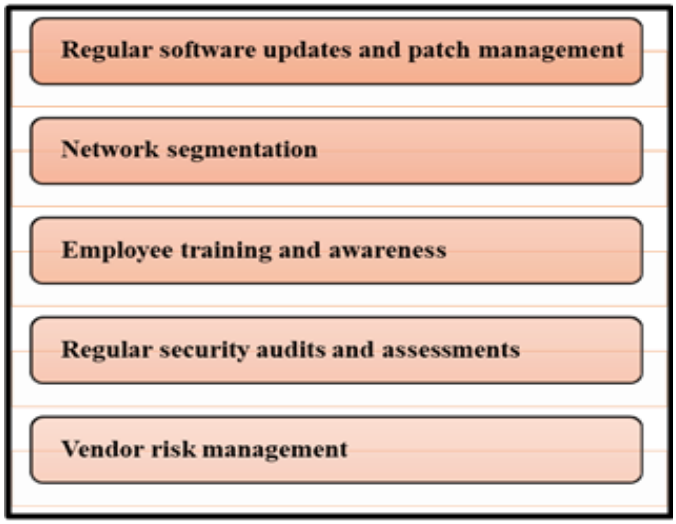

Figure 4: Strategies to reduce cyber security threats
(Source: Upadhyay & Sampalli, 2020)

There are several strategies that help in preventing cyber vulnerabilities. As per the opinion of Mugarza, Flores & Montero (2020), regular software updates and patch management help to eliminate potential entry points for attackers. In that case, regular software updates not only help provide information about the latest security threats but also help to build robust prevention against cyber threats. On the other hand, Basta et al. (2022) stated that dividing a network into isolated segments known as network segmentation is an effective strategy to prevent cyber threats that reduce the impact of a security breach. Strong and effective access controls following the principle of privilege play an important role in minimising the potential avenues for attackers to gain unauthorised access to a system. Organisations can proactively prevent attacks by conducting security audits and vulnerability assessments (Upadhyay & Sampalli, 2020). These assessments help to mitigate any weaknesses. Furthermore, training programs and awareness initiatives for employees are resources that enable them to recognize and report activities. This proactive approach helps mitigate the risks associated with error.

Developed models for identifying and preventing cyber threats

In the field of cybersecurity evaluating models is based on three factors; accuracy, efficiency and effectiveness. These metrics are crucial for identifying and preventing threats. As per the review of Wu et al. (2021), accuracy measures the correctness of predictions by comparing the number of instances to the number of instances. While high accuracy is desirable in prediction when dealing with imbalanced datasets it's important to consider metrics like recall and F1 score. On the other hand, Zhou, Jadoon & Shuja (2021) stated that efficiency plays a role in real - time threat detection as it reflects the computational resources and processing time required. Efficient models enable the analysis of datasets resulting in prompt responses. Therefore, achieving a balance between accuracy and efficiency is essential for implementation within an organisation. Lastly, in combating cyber threats the effectiveness of a model is vital for performance. It encompasses adaptability to evolving attack vectors and detection capabilities for both unknown threats and contributes to reducing cyber risk. Therefore, the machine learning model needs to be effective as well as accurate and unbiased which helps in the reduction of cyber risk.

3. Methodology

In the methodology section, the primary research approach was used due to the nature of numerical data. In that case, positivism research philosophy is used to understand the different applications associated with predictive analysis and different ML models that help in cyber risk. On the other hand, all the relevant information is measured by using a deductive research approach and this approach tests existing hypotheses that help to measure the integrity of the data (Murray et al.2019). This study used a random sampling method where all the participants were selected from different organisations who are from IT departments. In terms of the data collection process, the primary data collection method was used. In that case, all the data was collected from 70 participants who participated in the online survey. As per the asked questions, all the respondents offered their valuable experiences that helped in the data interpretation aspects.

In terms of data analysis method, this study used the primary data analysis method. By using the IBM SPSS tool, this study performed statistical analysis against collected data (Babbie, Wagner III & Zaino, 2022). In this analysis, different tests were performed that include descriptive analysis as well as Pearson correlation analysis and linear regression analysis that helps to understand the relationship between different factors.

4. Finding and Analysis

Demographic analysis

Age Distribution

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	31	44.3	44.3	44.3
	2	18	25.7	25.7	70.0
	3	10	14.3	14.3	84.3
	4	11	15.7	15.7	100.0
	Total	70	100.0	100.0	

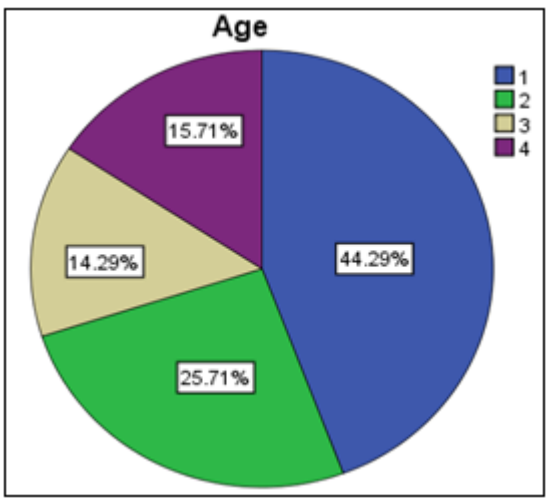

Figure 5: Age Distribution of participants in the survey
(Source: IBM SPSS)

The majority of survey respondents fall into the 30 - 37 age bracket which accounts for 44.3 percent of the total. The second largest group consists of individuals aged 38 - 44 years representing 25.7 percent. On the other hand, the

remaining participants are divided between the 45 - 50 years and 51 - 60 age groups comprising 14.3 percent and 15.7 percent of the sample respectively. The total number of participants was 70.

Gender Distribution

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	42	60.0	60.0	60.0
	2	23	32.9	32.9	92.9
	3	5	7.1	7.1	100.0
	Total	70	100.0	100.0	

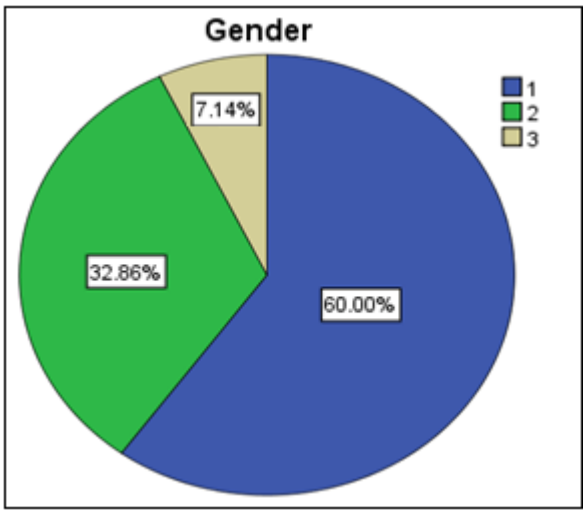

Figure 6: Gender Distribution of participants in the survey
(Source: IBM SPSS)

A significant portion of the survey participants were men comprising 60 percent of the respondents. The female category was also well represented accounting for 32.9 percent of the participants. Moreover, 7.1 percent of those surveyed identified with gender categories that go beyond the options and identified themselves as "Others. " This diverse mix of gender identities within the survey sample enables an analysis of cyber security perspectives and practices inclusively.

Educational Qualification

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	13	18.6	18.6	18.6
	2	15	21.4	21.4	40.0
	3	11	15.7	15.7	55.7
	4	31	44.3	44.3	100.0
	Total	75	100.0	100.0	

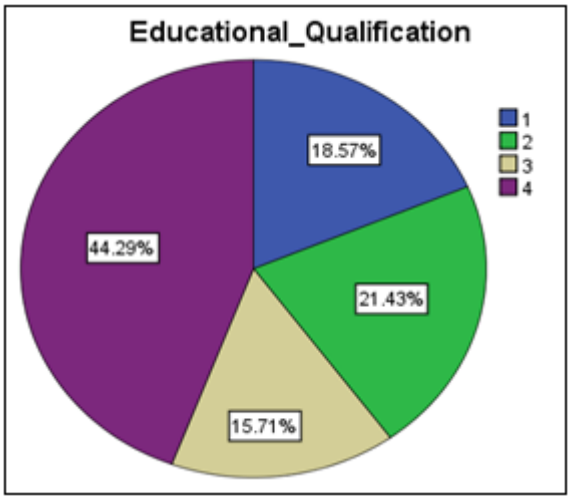

Figure 7: Qualification distribution among respondents
(Source: IBM SPSS)

The educational backgrounds of the survey participants show a range of qualifications. Most of the respondents have completed B. Tech degrees making up 44.3 percent of the total indicating a presence of individuals, with education. Following that 21.4 percent have graduation qualifications while 15.7 percent are post - graduates. A smaller but still significant group, comprising 18.6 percent of the participants hold diplomas.

Hypothesis 1

Table: Regression analysis for Hypothesis 2

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.768[a]	.589	.583	2.80107	.355

a. Predictors: (Constant), IV1_Predictive_Analytics

b. Dependent Variable: DV_Cybersecurity_Effectiveness

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	765.844	1	765.844	97.610	.000[b]
	Residual	533.528	68	7.846		
	Total	1299.371	69			

a. Dependent Variable: DV_Cybersecurity_Effectiveness

b. Predictors: (Constant), IV1_Predictive_Analytics

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.899	.662		1.358	.179
	IV1_Predictive_Analytics	.816	.083	.768	9.880	.000

a. Dependent Variable: DV_Cybersecurity_Effectiveness

(Source: IBM SPSS)

The regression analysis indicates a statistical relationship between DV_Cybersecurity_Effectiveness and IV1_Predictive_Analytics. In that case, the R - squared value is 0.589 indicating that 58.9 percent of role play by that IV1 on the dependent variables. On the other hand, 0.355 is the Durbin - Watson value that shows the autocorrelation ship between these factors. Similarly, in the ANOVA test, the F value is 97.610 shows a strong positive effect of predictive analytics on cybersecurity.

Hypothesis 2

Table 2: Regression analysis for Hypothesis 2

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.741[a]	.550	.543	2.93354	.439

a. Predictors: (Constant), IV2_Machine_Learning_Algorithms

b. Dependent Variable: DV_Cybersecurity_Effectiveness

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	714.188	1	714.188	82.991	.000[b]
	Residual	585.183	68	8.606		
	Total	1299.371	69			

a. Dependent Variable: DV_Cybersecurity_Effectiveness

b. Predictors: (Constant), IV2_Machine_Learning_Algorithms

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.255	.678		1.851	.068
	IV2_Machine_Learning_Algorithms	1.457	.160	.741	9.110	.000

a. Dependent Variable: DV_Cybersecurity_Effectiveness

(Source: IBM SPSS)

The results of the analysis indicate that the model is significant where the F value is 82.991. The signified coefficient value is 0.741 indicating an effective positive correlation between IV2_Machine_Learning_Algorithms and effectiveness of cybersecurity. On the other hand, the Durbin - Watson value is 0.439 suggests that a potential autocorrelation has been observed. This positive relation suggests that the one - unit increase in the use of machine learning algorithms increases the effectiveness of cybersecurity.

Correlation analysis

Table 3: Correlation analysis

Correlations				
		DV_Cybersecurity_ Effectiveness	IV1_Predictive _Analytics	IV2_Machine_ Learning_Algorithm
DV_Cybersecurity_ Effectiveness	Pearson Correlation	1	.768**	.741**
	Sig. (2- tailed)		.000	.000
	N	70	70	70
IV1_Predictive _ Analytics	Pearson Correlation	.768**	1	.894**
	Sig. (2- tailed)	.000		.000
	N	70	70	70
IV2_Machine_ Learning_Algorithm	Pearson Correlation	.741**	.894**	1
	Sig. (2- tailed)	.000	.000	
	N	70	70	70

**Correlation is significant at the 0.01 level (2- tailed)

(Source: IBM SPSS)

The table above indicates that the correlation analysis demonstrates robust and statistically significant relationships between the variables. In that case, the dependent variable makes strong relationships with both independent variables that signified at 0.01 level. The above data shows strong and unbiased predictive analytics as well as machine learning algorithms enhance the effectiveness of cybersecurity. This enhancement not only reduces cyber threats but also increases the productivity of organisations.

5. Discussion

The above findings indicate that predictive analytics applications as well as ML models play a vital role in mitigating cyber threats. In that case, analysing different historical data helps to understand the cybersecurity threats. Here, different principles of predictive analytics are associated with data collection and pre - processing where different ML models play vital roles in threat detection. On the other hand, behavioural analysis is used by predictive analytics to understand typical user and system behaviours in the dataset. Similarly, continuous learning helps an organisation to understand new types of threats whereas real - time analysis helps in accurate threat detection. Therefore, an increase in the use of these technologies corresponds to an improvement in overall cybersecurity effectiveness (Smith & Dhillon, 2020). By collaborating with machine learning models and predictive data analytics methods, an organisation gets a proper indication of the biases in the data that not only improve the cybersecurity measure but also brings improvement in the cybersecurity prevention segment.

On the other hand, it is important that organisations need to focus on regular software updates. In that case, organisations get information about the new threats and block those threats in positive ways. The correlation test indicates that a strong relationship has been observed where if an organisation maintains their security audits that would help to understand the current scenario of the cyber security aspects effectively. Employee training is an important factor in that segment where employees get in hand experiences with the different types of ML models. These models are Decision trees as well as SVM, KNN and K - means clusters that offer effective data analysis for different kinds of data. Therefore, by incorporating these technologies organisations can strengthen their cybersecurity frameworks enabling them to detect and mitigate cyber threats.

6. Conclusion

In the end, it can be concluded that the exploration of predictive analytics and machine learning algorithms for identifying and preventing cyber threats provides a complete understanding of their significant roles in enhancing cybersecurity. In that case, with the rise of cybersecurity threats throughout the world, organisations struggled in terms of addressing security threat issues effectively. On the other hand, by implementing different M models in the business analysis segment, an organisation not only increases outcomes but also understands the pattern of threats in positive ways. Here, by offering training as well as education to the employees, an organisation creates a robust cybersecurity ecosystem. Similarly, human awareness as well as effective strategies also help in the context. Therefore, the importance of analytics and machine learning in the field of cybersecurity is expected to grow more as technology advances. This advancement will play a role in safeguarding assets and ensuring the integrity of computer systems.

References

[1] Alloghani, M., Al - Jumeily, D., Hussain, A., Mustafina, J., Baker, T., & Aljaaf, A. J. (2020). Implementation of machine learning and data mining to improve cybersecurity and limit vulnerabilities to cyber attacks. *Nature - inspired computation in data mining and machine learning*, 47 - 76. Retrieved on: 23rd December 2022, from: https: //link. springer. com/chapter/10.1007/978 - 3 - 030 - 28553 - 1_3

[2] Babbie, E., Wagner III, W. E., & Zaino, J. (2022). *Adventures in social research: Data analysis using IBM SPSS statistics*. Sage Publications. Retrieved on: 23rd December 2022, from: https: //books. google. com/books?hl=en & lr= & id=amdjEAAAQBAJ & oi=fnd & pg=PT22 & dq=By+using+the+IBM+SPSS+tool, +this+study+performed+statistical+analysis+against+collected+data. + & ots=jVMUQ9lgVu & sig=6jiTzYetdjbSU2VtgXN3E99KrVU

[3] Basta, N., Ikram, M., Kaafar, M. A., & Walker, A. (2022, April). Towards a zero - trust micro - segmentation network security strategy: an evaluation framework. In *NOMS 2022 - 2022 IEEE/IFIP Network Operations and Management Symposium* (pp.1 - 7). IEEE. Retrieved on: 23rd December 2022, from: https: //arxiv. org/pdf/2111.10967

[4] Benzaïd, C., & Taleb, T. (2020). AI for beyond 5G networks: a cyber - security defense or offense enabler?. *IEEE network*, *34* (6), 140 - 147. Retrieved on: 23rd December 2022, from: https: //arxiv. org/pdf/2201.02730

[5] Chinedu, P. U., Nwankwo, W., Masajuwa, F. U., & Imoisi, S. (2021). Cybercrime Detection and Prevention Efforts in the Last Decade: An Overview of the Possibilities of Machine Learning Models. *Review of International Geographical Education Online*, *11* (7). Retrieved on: 23rd December 2022, from: https: //www.researchgate. net/profile/Paschal - Chinedu/publication/355668267_Cybercrime_Detection_and_Prevention_Efforts_in_the_Last_Decade_An_Overview_of_the_Possibilities_of_Machine_Learning_Models/links/61792620a767a03c14bbc798/Cybercrime - Detection - and - Prevention - Efforts - in - the - Last - Decade - An - Overview - of - the - Possibilities - of - Machine - Learning - Models. pdf

[6] González, S., García, S., Del Ser, J., Rokach, L., & Herrera, F. (2020). A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion*, *64*, 205 - 237. Retrieved on: 23rd December 2022, from: https: //www.sciencedirect. com/science/article/pii/S1566253520303195

[7] Husák, M., Bartoš, V., Sokol, P., & Gajdoš, A. (2021). Predictive methods in cyber defense: Current experience and research challenges. *Future Generation Computer Systems*, *115*, 517 - 530. Retrieved on: 23rd December 2022, from: https: //www.researchgate. net/profile/Martin - Husak/publication/344649158_Predictive_methods_in_cyber_defense_Current_experience_and_research_challenges/links/5f86d2fa92851c14bcc6b6c2/Predictive -

methods - in - cyber - defense - Current - experience - and - research - challenges. pdf

[8] Mugarza, I., Flores, J. L., & Montero, J. L. (2020). Security issues and software updates management in the industrial internet of things (iiot) era. *Sensors*, *20* (24), 7160. Retrieved on: https: //www.mdpi. com/1424 - 8220/20/24/7160/pdf 23rd December 2022, from:

[9] Murray, D. L., Bastille - Rousseau, G., Beaty, L. E., Hornseth, M. L., Row, J. R., & Thornton, D. H. (2019). From research hypothesis to model selection. *Population ecology in practice*. Retrieved on: 23rd December 2022, from: https: //books. google. com/books?hl=en & lr= & id=PlTGDwAAQBAJ & oi=fnd & pg=PA17 & dq=all+the+relevant+information+is+measured+by+using+a+deductive+research+approach+and+this+approach+tests+existing+hypotheses+that+help+to+measure+the+integrity+of+the+data. & ots=MJaxv - EXSZ & sig=XIIsuL8E1HRKbOt2aoXfcQebhZo

[10] Naicker, N., Adeliyi, T., & Wing, J. (2020). Linear support vector machines for prediction of student performance in school - based education. *Mathematical Problems in Engineering*, *2020*, 1 - 7. Retrieved on: 23rd December 2022, from: https: //www.hindawi. com/journals/mpe/2020/4761468/

[11] Smith, K. J., & Dhillon, G. (2020). Assessing blockchain potential for improving the cybersecurity of financial transactions. *Managerial Finance*, *46* (6), 833 - 848. Retrieved on: 23rd December 2022, from: https: //www.emerald. com/insight/content/doi/10.1108/MF - 06 - 2019 - 0314/full/html

[12] Statista. com, 2022. Cybercrime Expected To Skyrocket in Coming Years. Retrieved on: 23rd December 2022, from: https: //www.statista. com/chart/28878/expected - cost - of - cybercrime - until - 2027/

[13] Trivedi, N. K., Simaiya, S., Lilhore, U. K., & Sharma, S. K. (2020). An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology*, *29* (5), 3414 - 3424. Retrieved on: 23rd December 2022, from: https: //www.researchgate. net/profile/Dr - Kumar - Lilhore/publication/341932015_An_Efficient_Credit_Card_Fraud_Detection_Model_Based_on_Machine_Learning_Methods/links/5ee4a477458515814a5b891e/An - Efficient - Credit - Card - Fraud - Detection - Model - Based - on - Machine - Learning - Methods. pdf

[14] Upadhyay, D., & Sampalli, S. (2020). SCADA (Supervisory Control and Data Acquisition) systems: Vulnerability assessment and security recommendations. *Computers & Security*, *89*, 101666. Retrieved on: 23rd December 2022, from: https: //www.researchgate. net/profile/Darshana - Upadhyay/publication/342084000_SCADA_Supervisory_Control_and_Data_Acquisition_systems_Vulnerability_assessment_and_security_recommendations/links/5ee181e792851ce9e7d915fe/SCADA - Supervisory - Control - and - Data - Acquisition - systems - Vulnerability - assessment - and - security - recommendations. pdf

[15] Wei, Y., Chow, K. P., & Yiu, S. M. (2020). Insider threat detection using multi - autoencoder filtering and unsupervised learning. In *Advances in Digital Forensics XVI: 16th IFIP WG 11.9 International Conference, New Delhi, India, January 6–8, 2020, Revised Selected Papers 16* (pp.273 - 290). Springer International Publishing. Retrieved on: 23rd December 2022, from: https: //inria. hal. science/hal - 03657238/file/503209_1_En_15_Chapter. pdf

[16] Wu, X., Zheng, W., Xia, X., & Lo, D. (2021). Data quality matters: A case study on data label correctness for security bug report prediction. *IEEE Transactions on Software Engineering*, *48* (7), 2541 - 2556. Retrieved on: 23rd December 2022, from: https: //ink. library. smu. edu. sg/cgi/viewcontent. cgi?article=8439 & context=sis_research

[17] Zhou, S., Jadoon, W., & Shuja, J. (2021). Machine learning - based offloading strategy for lightweight user mobile edge computing tasks. *Complexity*, *2021*, 1 - 11. Retrieved on: 23rd December 2022, from: https: //www.hindawi. com/journals/complexity/2021/6455617/