RESEARCH ARTICLE OPEN ACCESS

Intelligence Forensics: Ai- Powered Classification of Cyber Attacks

Ms. Hinduja J*, Ms. Elampirai Gopika S **

*Department of Computer Science and Engineering, Dr. M.G.R. Educational and Research Institute
** Assistant Professor, Centre Of Excellence in Digital Forensics,
Chennai, India

ABSTRACT

As the prevalence of cyber crime increased in the era of digital technology, emergency demands for complex tools for effective classification and analysis of cyber attacks have been created. This paper provides an AI intelligence systems that categorize different types of cybercrime, including phishing, malware and distributed denial of service (DDOS). Using machine learning methods like logistics regression algorithm, decision tree algorithm and random forest algorithm, here the logistics regression algorithm provides accuracy of 59.162832871295954%, the decision tree algorithm provides the accuracy of 66.09636637953459% and the random forest algorithm provides the accuracy of 99.68487316888272%. Comparing the accuracy of these three algorithm random forest provides the best accuracy. The system recognizes cyber attacks and increases accuracy when classified. The proposed solution integrates data from diverse sources to identify patterns and anomalies that characterize different types of threats. This approach supports ethical cyber experts and law enforcement in comforting cybercrime with accurate and efficient attack classification that ultimately contributes to a healthy cyber safety infrastructure.

Keywords — Cybercrime, Classification, Artificial Intelligence, logistic regression, Decision Tree, Random Forest.

I. INTRODUCTION

In this cyber world, the cyber attacks has become the most challenges in the digital era. Now-a-days people rely on technologies for personal, financial and business operations, which makes the cyber criminals continuously evolve their attack. Malware, short for malicious software is a program code that is hostile and often used to corrupt or misuse a system [1]. This malware has become a significant tool for the attackers which they use in their attacks. Detecting, classifying and preventing malware is the complicated issue for cyber security department due to several problems. According to International Telecommunication Union (2014), defeating cybercrime constitutes an integral part of any nations' cybersecurity protection strategy for its critical information infrastructure [2].

The cyber crime with malware and cyber threats challenges the traditional security measures. This paper directly points the key problems engaged with cyber attacks for their detection, classification, and its prevention. The traditional methods, like signature based anti-virus solutions struggles to detect and classify the type of cyber attacks based on their behaviour and predictive modelling using AI and machine learning. It makes more easy for the cyber professionals to detect, classify and prevent as it provides the preventive for detected attack. AI algorithms can process vast amounts of data in real-time, identifying patterns and anomalies indicative of cyber threats, this capability surpasses traditional methods that often rely on predefined signatures, enabling the detection of both known and emerging threats [3].

This system focuses on the behaviour of cyber attacks to improve the system security by detection and classification, to make it possible the system processes the large amount of data from vast sources like network traffic, system protocols and event reports. It identifies the attack patterns using machine learning algorithms and detects the threat more faster and accurate. This system will also give importance to continuous learning, so that it can adapt to new and existing threats. By employing AI algorithms, it can detect and respond to threats in real-time, enhancing the organization's ability to mitigate potential damages [4]. The AI-driven system will continuously learn from new data, enabling it to adapt to emerging cyber threats, this adaptive learning ensures that the system remains effective and reduces the reliance on manual updates [5].

II. LITERATURE REVIEW:

Muhsiena K H and Amrutha [6] had proposed When AI inherits existing security concerns in computer systems, concerns increase due to its own cyber attacks supported by AI. He defines AI crimes and divides them into two categories: tool crimes and target crimes. It also deals with AI crime property, both in the past and in the present. Traditional forensic approaches cannot affect difficult-to-solve problems.

Raed S. A. Faqir [7] had proposed that multifaceted approach, including qualitative, descriptive and analytical methods, and supports data primarily from many legal documents and scientific literature. He describes the key role in playing AI within law enforcement, and encompasses aspects with advanced audio analytics techniques. The results analyse data about the organization and highlight the potential to change machine learning techniques.

Hareesh Kumar C, Trisha B [8] had proposed that a deep dive into how AI is shaping digital forensics, enhancing the ability to tackle complex cyber threats and manage vast amounts of digital data. It explores key AI technologies, like machine learning and deep learning, and highlights their growing importance in forensic research. He considers the use of machine learning and artificial intelligence (AI) in automatic analysis and threat classification and highlights how AI and machine learning can drive digital justice research while emphasizing the vital role of cyber forensics in maintaining security and fairness .

Christopher Rigano [9] had proposed that AI is the reality that forms our daily life deeper. From smartphones and self-driving cars to finance, healthcare, and beyond, AI is transforming how we live and work. It plays a role in everything from agriculture and manufacturing to education, government, and public safety.

Abraham Okandeji Omokanye , Akintayo Micheal Ajayi and et al., [10] had proposed that the bank case is digitized, financial crimes are more sophisticated and traditional prevention methods are less effective. It looks at how machine learning models, real-time detection systems, and advanced analytics are transforming fraud prevention. He also identifies research gaps and proposes future directions to further effectively prevent AI-controlled financial crime prevention.

Amin and et al [11] proposes that the rapidly rising usage of telecommunication and information networks which inter-connect modern society through computers, smart phones and other electronic devices has led to security threats and cyber-crimes (CC) activities. He proposed that this technique on publicly available dataset about intrusion attacks, the results show that the proposed approach can fully predict all intrusion attacks and also provides prior useful information to the security engineers or developers to conduct a mandating action.

Gordon and et al [12] proposed that the breadth of computer-based crime, providing a definition of the emerging terms “Cybercrime” and “crimeware”. He divide Cybercrime into two distinct categories: Type I Cybercrime, which is mostly technological in nature, and Type II Cybercrime, which has a more pronounced human element. And use two case studies to illustrate the role of crimeware in different types of Cybercrime, and offer some observations on the role of cognition in the process of Cybercrime. Finally we provide several suggestions for future work in the area of Cybercrime.

III. PROPOSED METHODOLOGY:

In this paper, the proposed methodology first a dataset is collected from the Kaggle website which contains historical data. The collected data are pre-processed using python's numpy and Pandas library to organize the dataset. The NumPy package ensures the efficient storage and processing of numerical arrays [13]. The Pandas gives Python the ability to work with spreadsheet-like data for fast data loading, manipulating, aligning, merging, etc [14]. The effective preprocessing techniques play a pivotal role in enhancing the efficiency and accuracy of large data [15]. Next, you can visually examine the data as it is displayed in visual format using Seaborn and Matplotlib. Matplotlib is a graphics library for data visualization package in Python which encompasses as an integral aspect in the python and it is easily supported with NumPy, Pandas and other relevant libraries [16]. Seaborn is a graphic visualization library that is built on the primary configurations of Matplotlib, it provides accessibility to the users with some of the most commonly provides data visualizations processes with certain data visualizations necessities such as mapping colour to a variable or using faceting requirements across the globe [16].

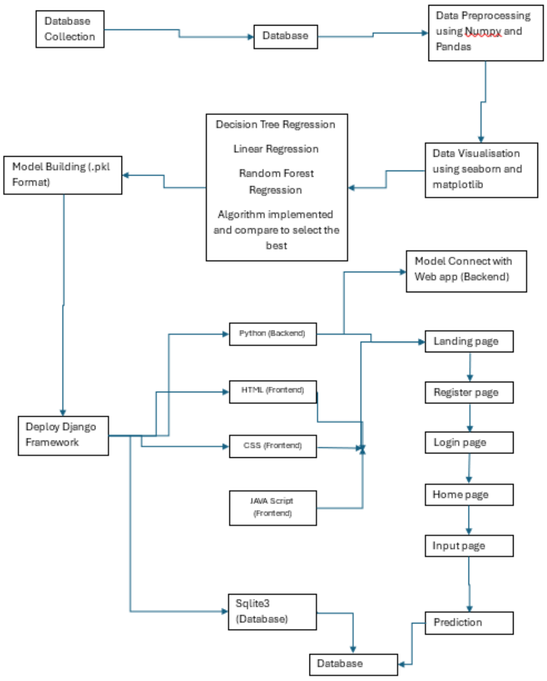


Fig 1: Architecture diagram

Now using logistics regression algorithm, decision tree regression algorithm and random forest algorithm we start building the model and train them by using the pre-processed dataset. Logistic regression is an machine learning approach used mainly for classification in machine learning that employs supervised learning to estimate the likelihood of a target variable [17]. Decision tree regression is used to find data, to extract text, to find missing data [18]. Random forests can be used either for a categorical response variable, such as "classification", or a continuous response, referred to as "regression" [19]. This process evaluates each model's working by the accuracy level they provide at output and then we select the best performing model. Here the logistics regression algorithm provides accuracy of 59.162832871295954%, the decision tree algorithm provides the accuracy of 66.09636637953459% and the random forest algorithm provides the accuracy of 99.68487316888272%. Comparing the accuracy of these three algorithm random forest provides the best accuracy. Then the chosen model is saved in .pkl format using Joblib. After the train and test process, the trained model (i.e., model.pkl) is loaded into the Django framework in the backend. This process creates views and logic to handle the predictions and input data. It also handles the routing process and API integration for the model, for this backend process the python language is used.

Here we develop the frontend for the users interaction with the model using HTML, CSS and JS, it contains landing page, registration page, login page, home page, input page and output page. Here we use the sqlite3 database to store user data from registration page, input values, and its prediction, it was also managed future tracing. These details can be seen in the created database page.

Finally, the Django framework combines the backend (Python) and the frontend (HTML, CSS and JS), and connects the model with the database. Django defines itself as a free and open-source web framework based on the Python programming language and adhering to the model–template–views architectural paradigm [20]. The application can be used locally or accessed on cloud platform. The final output is the users can register/login in the registration page, input the values of the network and, predicted attack type, its status and the preventive measures are displayed according to the input values.

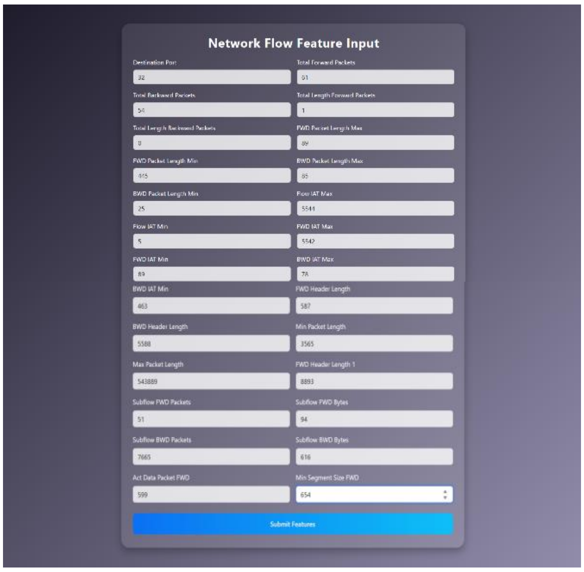
IV. FINDINGS :

This paper primarily focuses on AI – Powered classification of cyber crime attacks using machine learning and forensic analytics. Where as the prior researchers discusses about the cover broader areas, including AI in legal enforcement [7], financial crime prevention [10], and general criminal justice applications [9].

Compared to various AI applications including biometric identification [7], deep learning in digital forensics [8] and predictive policing [9], this paper focuses on machine learning techniques like random forest and logistics regression for cyber crime classification and detection.

Algorithm	Accuracy	Percision	Recall	F1 – Score	Support
Logistic Regression	59.162832871295954 %	0.68	0.59	0.59	134549
Decision Tree	66.09636637953459 %	0.55	0.66	0.58	134549
Random Forest	99.68487316888272 %	1.00	1.00	1.00	134549

Some explorations like financial fraud detection [10] and AI given legal framework [7 – 9], this paper concentrates cyber security threats like phishing, distributed denial of services (DDOS), and malware. While deeper insights into AI powered surveillance [7] and forensic data management [8], Intelligence Forensics: Ai- Powered Classification Of Cyber Crime Attacks discusses AI in cyber security and forensics analysis but lacks in – depth exploration of AI's role in biometric security and blockchain – based crime prevention and it also lacks dedicated discussion on AI ethics privacy and regulatory challenges.


Network Flow Feature Input

Destination Port
32

Total Forward Packets
51

Total Backward Packets
54

Total Length Forward Packets
1

Total Length Backward Packets
0

FWD Packet Length Max
89

FWD Packet Length Min
445

BWD Packet Length Max
85

BWD Packet Length Min
25

Flow IAT Max
5541

Flow IAT Min
5

FWD IAT Max
5542

FWD IAT Min
89

BWD IAT Max
78

BWD IAT Min
463

FWD Header Length
587

BWD Header Length
5588

Min Packet Length
3565

Max Packet Length
543889

FWD Header Length 1
8893

Subflow FWD Packets
51

Subflow FWD Bytes
94

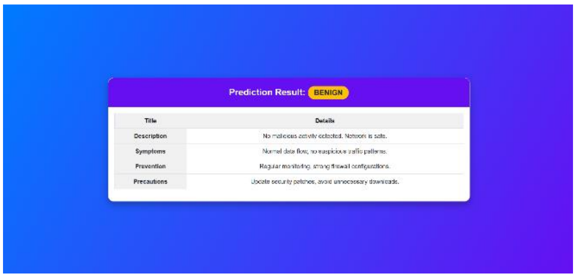
Subflow BWD Packets
7665

Subflow BWD Bytes
616

Act Data Packet FWD
599

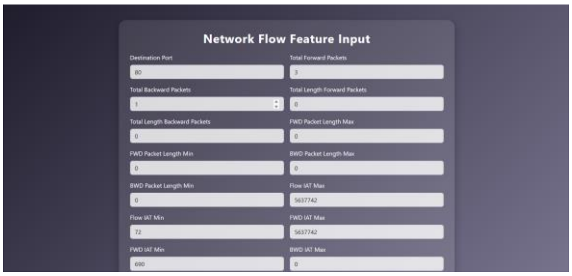
Min Segment Size FWD
654

Fig 2: prediction 1 Input page


Prediction Result: BENIGN

Title	Details
Description	No malicious activiy detected. Network is safe.
Symptoms	Normal data flow, no suspicious traffic patterns.
Prevention	Regular monitoring, strong firewall configurations.
Precautions	Update security patches, avoid unnecessary downloads.

Fig 3: Prediction 1 output


Network Flow Feature Input

Destination Port
80

Total Forward Packets
3

Total Backward Packets
1

Total Length Forward Packets
0

Total Length Backward Packets
0

FWD Packet Length Max
0

FWD Packet Length Min
0

BWD Packet Length Max
0

BWD Packet Length Min
0

Flow IAT Max
5637742

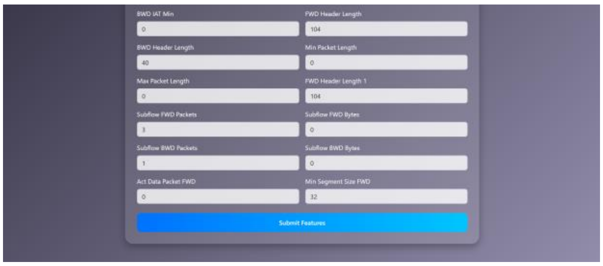
Flow IAT Min
72

FWD IAT Max
5637742

FWD IAT Min
690

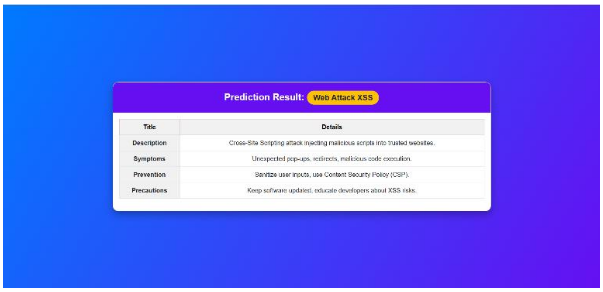
BWD IAT Max
0

Fig 4: Prediction 2 input page (a)



BWD IAT Min	0	FWD Header Length	104
BWD Header Length	40	Min Packet Length	0
Max Packet Length	0	FWD Header Length 1	104
Subflow FWD Packets	3	Subflow FWD Bytes	0
Subflow BWD Packets	1	Subflow BWD Bytes	0
Act Data Packet FWD	0	Min Segment Size FWD	32

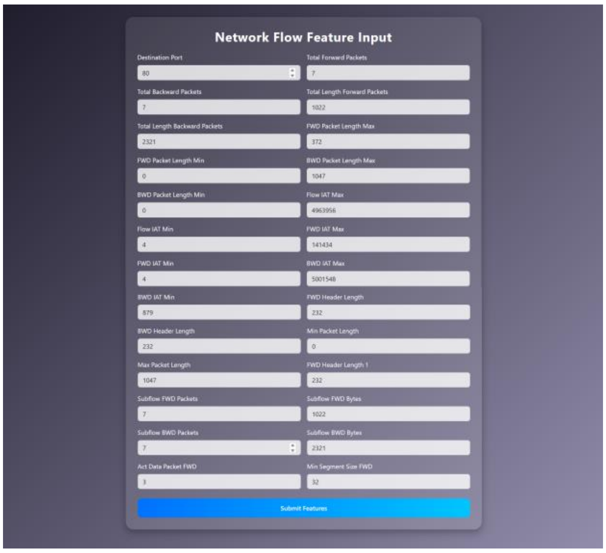
Fig 5: Prediction 2 input page (b)



Prediction Result: Web Attack XSS

Title	Details
Description	Cross-Site Scripting attack injecting malicious scripts into trusted websites.
Symptoms	Unexpected pop-ups, redirects, malicious code execution.
Prevention	Sanitize user inputs, use Content Security Policy (CSP).
Precautions	Keep software updated, educate developers about XSS risks.

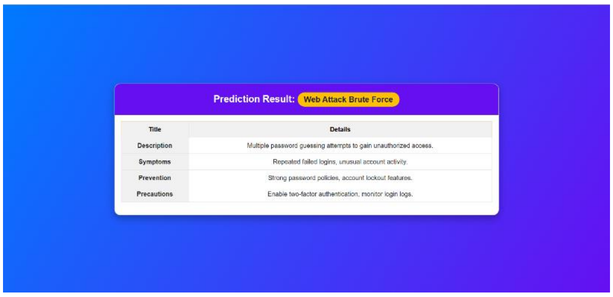
Fig 6: Prediction 2 output



Network Flow Feature Input

Destination Port	80	Total Forward Packets	7
Total Backward Packets	7	Total Length Forward Packets	1022
Total Length Backward Packets	2321	FWD Packet Length Max	372
FWD Packet Length Min	0	BWD Packet Length Max	1047
BWD Packet Length Min	0	Flow IAT Max	4963956
Flow IAT Min	4	FWD IAT Max	141434
FWD IAT Min	4	BWD IAT Max	5001548
BWD IAT Min	879	FWD Header Length	232
BWD Header Length	232	Min Packet Length	0
Max Packet Length	1047	FWD Header Length 1	232
Subflow FWD Packets	7	Subflow FWD Bytes	1022
Subflow BWD Packets	7	Subflow BWD Bytes	2321
Act Data Packet FWD	3	Min Segment Size FWD	32

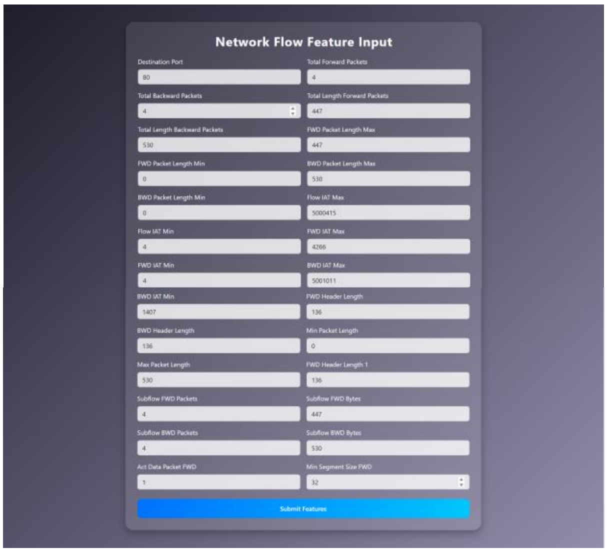
Fig 7: Prediction 3 input page



Prediction Result: Web Attack Brute Force

Title	Details
Description	Multiple password guessing attempts to gain unauthorized access.
Symptoms	Repeated failed logins, unusual account activity.
Prevention	Strong password policies, account lockout features.
Precautions	Enable two-factor authentication, monitor login logs.

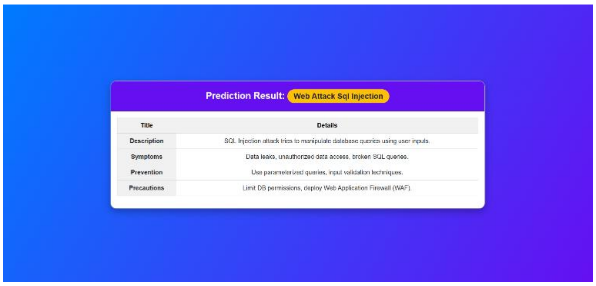
Fig 8: Prediction 3 output



Network Flow Feature Input

Destination Port	80	Total Forward Packets	4
Total Backward Packets	4	Total Length Forward Packets	447
Total Length Backward Packets	530	FWD Packet Length Max	447
FWD Packet Length Min	0	BWD Packet Length Max	530
BWD Packet Length Min	0	Flow IAT Max	5000415
Flow IAT Min	4	FWD IAT Max	4266
FWD IAT Min	4	BWD IAT Max	5001011
BWD IAT Min	1407	FWD Header Length	136
BWD Header Length	136	Min Packet Length	0
Max Packet Length	530	FWD Header Length 1	136
Subflow FWD Packets	4	Subflow FWD Bytes	447
Subflow BWD Packets	4	Subflow BWD Bytes	530
Act Data Packet FWD	1	Min Segment Size FWD	32

Fig 9: Prediction 4 input page



Prediction Result: Web Attack Sql Injection

Title	Details
Description	SQL Injection attack tries to manipulate database queries using user inputs.
Symptoms	Data leaks, unauthorized data access, broken SQL queries.
Prevention	Use parameterized queries, input validation techniques.
Precautions	Limit DB permissions, deploy Web Application Firewall (WAF).

Fig 10: Prediction 4 output

V. CONCLUSION:

In conclusion, the Intelligence Forensics AI – Powered classification of cyber attacks helps in automating the detection and classification of cyber crime using machine learning algorithms like logistic regression, decision tree and random forest algorithm. In this paper the logistic regression provides 59.1628328712959554% of accuracy, the decision tree algorithm provides 66.09636637953459% of accuracy and the random forest algorithm 99.68487316888272% of accuracy. By comparing the accuracy of these algorithms, the random forest algorithm provides the best accuracy. By using the best accuracy algorithm it improves the efficiency and speed of identifying the cyber attacks. This system has the capability of continuous learning so that it can adapt to the emerging threats. It also generates the detailed report which aids in forensic investigations and decision making. This system offers a strong solution for strengthening cyber forensics, it also enhances the security measures for organisations, and supports the law enforcement in defending the cyber crime.

REFERENCES:

[1] Namanya, A. P., Cullen, A., Awan, I., & Pagna Diss, J. (2018). The world of malware: An overview. *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, 318–323.

[2] 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. *Rigeo*, *11*(7).

[3] Salem, A. H., Azzam, S. M., Emam, O. E., & others. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. *Journal of Big Data*, *11*, 105.

[4] Ofoegbu, K. D. O., Osundare, O. S., Ike, C. S., Fakeyede, O. G., & Ige, A. B. (2024). Real-Time Cybersecurity threat detection using machine learning and big data analytics: A comprehensive approach. *Computer Science & IT Research Journal*, *4*(3).

[5] Sudaryono, S., Pratomo, R., Ramadan, A., Ahsanitaqwim, R., & Fletcher, E. (2025). Artificial Intelligence in Predictive Cybersecurity: Developing Adaptive Algorithms to Combat Emerging Threats. *Journal of Computer Science and Technology Application*, *2*(1), 1-13.

[6] Muhsiena K H and Amrutha N. (2022) "An Exploration into Artificial intelligence of Security threat, Crime and Forensics". *International Journal for Modern Trends in Science and Technology*, 8(03), pp. 25-33.

[7] Faqir, R. S. (2023). Digital criminal investigations in the era of artificial intelligence: A comprehensive overview. *International Journal of Cyber Criminology*, *17*(2), 77-94.

[8] Hareesh Kumar C, Trisha B. (2024), "Cyber Forensic Analytics With Ai. *International Journal For Multidisciplinary Research".*

[9] Rigano, C. (2019). Using artificial intelligence to address criminal justice needs. *National Institute of Justice Journal*, *280*(1-10), 17.

[10] Omokanye, A. O., Ajayi, A. M., Olowu, O., Adeleye, A. O., Chianumba, E. C., & Omole, O. M. (2024). AI-powered financial crime prevention with cybersecurity, IT, and data science in modern banking. *International Journal of Science and Research Archive*, *13*(3).

[11] Amin, A., Anwar, S., Adnan, A., Khan, M. A., & Iqbal, Z. (2015, November). Classification of cyber attacks based on rough set theory. In *2015 First International Conference on Anti-Cybercrime (ICACC)* (pp. 1-6). IEEE.

[12] Gordon, S., & Ford, R. (2006). On the definition and classification of cybercrime. Journal in computer virology, 2, 13-20.

[13] Gupta, P., & Bagchi, A. (2024). Introduction to NumPy. In *Essentials of Python for Artificial Intelligence and Machine Learning* (pp. 127-159). Cham: Springer Nature Switzerland.

[14] Gupta, P., & Bagchi, A. (2024). Introduction to Pandas. In *Essentials of Python for Artificial Intelligence and Machine Learning* (pp. 161-196). Cham: Springer Nature Switzerland.

[15] Bala, B., & Behal, S. (2024, October). A Brief Survey of Data Preprocessing in Machine Learning and Deep Learning Techniques. In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 1755-1762). IEEE.

[16] Sial, A. H., Rashdi, S. Y. S., & Khan, A. H. (2021). Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python. *International Journal*, *10*(1), 277-281.line

[17] Gonaygunta, H. (2023). Machine learning algorithms for detection of cyber threats using logistic regression. *Department of Information Technology, University of the Cumberlands*.

[18] Bobirovich, M. A. (2022). Regression Based on Decision Tree Algorithm. *Вестник науки и образования*, (4-1 (124)), 25-30.

[19] Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69-79.

[20] Sunday, S. M. (2023). Phishing website detection using machine learning: Model development and django integration. *Journal of Electrical Engineering, Electronics, Control and Computer Science*, *9*(3), 39-54.