ASTONJADRO

pISSN 2302-4240
eISSN 2655-2086

Volume 13, Issue 2, June 2024, pp.587-596
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

Global Network Cyberattack Classification Using Naive Bayes Method Time Range 2020 – 2023

Acep Sandi Mutia, Irawan Irawan, Christina Juliane

STMIK Likmi, Bandung, INDONESIA

E-mail: acepsandimutia707@gmail.com, irawan.skom@gmail.com, christina.juliane@likmi.ac.id

| Submitted: December 18, 2023 | Revised: December 19, 2023 | Accepted: May 20, 2024 |

| Published: May 22, 2024 |

ABSTRACT

This study focuses on developing a classification model for cyberattacks on global networks during the time span of 2020 to 2023 using the Naive Bayes method. The main objective of the study is to analyze and classify the frequent severity of cyber, which helps in improving network security and reducing vulnerabilities. The Naive Bayes method was chosen for its efficiency in handling large datasets and its ability to make predictions based on probabilities. Collecting cyberattack data from a variety of reliable and up-to-date sources, the study covers attacks such as ransomware, phishing, DDoS, and other malware. The classification process includes data pre-processing, feature extraction, and finally the application of Naive Bayes algorithms to identify patterns in such attacks. The classification results are then evaluated using the Apply Model and Performance validation methods to assess the effectiveness of the model. The results of this study show that Naive Bayes is able to accurately classify cyberattacks, providing a useful tool for cybersecurity professionals to understand attack trends and respond proactively. The study also suggests areas for further research, including the integration of the Naive Bayes model with other artificial intelligence systems for improved cyberattack detection. The study provides new insights into the application of the Naive Bayes method in cybersecurity and paves the way for improved data-driven cyber defense strategies.

Keywords: data mining; classification of cyberattacks; naive bayes; network security; security data analysis.

INTRODUCTION

In today's digital age, cybersecurity has become a top priority for both individuals, organizations, because the security of every digital device that is part of a global network with millions of computer nodes is critical (Veeramanickam et al., 2022). The rapid growth in information and communication technology has opened up opportunities for global economic activity. However, on the other hand, it also poses opportunities for different types of cyberattacks. From ransomware to DDoS (Distributed Denial of Service) attacks. DDoS attacks flood a network with traffic from many infected computers to disrupt or disable access to a target's device or service (Shukla et al., 2023).

The years 2020 to 2023 recorded a significant increase in the number and complexity of cyberattacks. The COVID-19 pandemic, for example, became a catalyst for cyberattacks as it transitioned massively to remote work, widening the attack surface for cybercriminals. The sophistication of AI and machine learning technology has also been leveraged by cyber attack actors to create threats that are more adaptive and difficult to detect (Amin et al., 2020; Bécue et al., 2021; Kuhn et al., 2021).

lassifying cyberattacks is critical because it helps in identifying, analyzing, and handling threats more effectively. Classification allows cybersecurity experts to recognize patterns and distinctive features of different types of attacks, which can speed up the process of detection and response to security incidents. With a better understanding of the nature and characteristics of attacks, security strategies can be tailored to be more resilient in the face of specific threats (Lin et al., 2022; Syrmakesis et al., 2022).

At the global level, effective classification of cyberattacks enables better collaboration and coordination among countries and international organizations in the fight against cybercrime. This is important given that cyberattacks are often transnational in nature, crossing national borders and legal jurisdictions. Therefore, international cooperation, supported by a deep understanding of cyberattacks, is key to responding effectively to cyber threats and minimizing their impact on global security (Sarker, 2023; Sarker et al., 2020).

Through efficient classification, we can also understand the evolution of cyber threats and anticipate future trends. This helps in the development of cybersecurity solutions that are proactive, rather than just reactive, ensuring that security measures remain relevant and effective in the face of evolving threats (Abu Al-Haija & Al-Fayoumi, 2023).

Overall, a thorough understanding of cyberattacks and accurate classification of these threats are critical components in a global cybersecurity strategy, ensuring data security, critical infrastructure, and maintaining the integrity of information systems around the world.

The Naïve Bayes method is a statistical classification technique based on Bayes' Theorem. This is one of the most simple and effective machine learning models, especially in the case of large-dimensional data. Naive Bayes operate under the assumption that features in the dataset are independent of each other, an approach known as "naivety". Although this assumption is sometimes unrealistic in practice, Naive Bayes proved efficient in a wide range of practical applications, especially in natural language processing and text classification (Kim & Lee, 2022).

In the context of cyberattack classification, Naive Bayes was chosen for several reasons. First, its ability to handle large amounts of data closely matches the vast and diverse characteristics of cyberattack data. Secondly, this method is efficient in terms of computational time, which is very important in the detection of cyberattacks, where response speed is key. Third, Naive Bayes can operate well even using datasets that have noise and incomplete features, which often occur in cyberattack data (Shi et al., 2021).

In the classification of cyberattacks, Naive Bayes are used to identify and categorize different types of attacks based on features present in the data, such as network traffic patterns, payload types, and other abnormal behaviors. The model is trained with a dataset that includes known instances of attacks, allowing the algorithm to 'learn' the different patterns and characteristics of each type of attack.

The main advantages of Naive Bayes are ease of implementation and efficiency in computing. In addition, these models tend to work well even in imperfect data conditions. However, there are also some challenges. Assumptions of feature independence are often unrealistic in cyberattack data, where features can be interrelated. In addition, Naive Bayes can be less effective if the distribution of data within classes is not uniform or if there are dominant features that can give rise to bias (Chen et al., 2021; Chu et al., 2020; Redivo et al., 2023).

Overall, despite its limitations, Naive Bayes remains a popular choice in the classification of cyberattacks due to its practicality and efficiency. Especially in situations where speed and the ability to manage large volumes of data are a priority (Blanquero et al., 2021).

Information systems that support network planning activities are technology platforms used to collect, process and analyze data to assist in the network planning and management process. The system integrates various types of data, including geographic, demographic and infrastructure data, to provide a comprehensive picture of field needs and conditions (Sari OL et.al, 2024; Nur A et.al, 2024).

In the data processing process, this system carries out needs analysis and forecasting to predict future network demand. Modeling and simulation are used to test various network design scenarios and select the best option based on performance, cost, and scalability. Furthermore, this information system supports network topology design by determining the optimal structure and allocating resources efficiently. Project management processes, including scheduling, monitoring, and budget management, are also facilitated by this system to ensure network implementation runs smoothly and according to plan. Additionally, the system provides reporting and documentation tools to monitor network performance and manage technical documentation. Collaboration and communication between teams is made easy through collaboration portals and real-time

pISSN 2302-4240
eISSN 2655-2086
Volume 13, Issue 2, June 2024, pp.587-596
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

communication tools, while data security is maintained through encryption and compliance with privacy regulations (Pradnyana IM et.al, 2023; Prastowo FI et.al, 2023).

RESEARCH METHOD

This study applies algorithm modeling techniques using the Knowledge Discovery in Database (KDD) approach. The purpose of this modeling technique is to extract previously unknown information and understanding from the database. The explanation can be seen in figure 1 (Fahd et al., 2022).

1. Collecting Data
   Collecting data is the stage of data that has been collected. The data used is sourced from kaggle.com, which is opensource (Schoenenwald et al., 2021).
2. Cleansing
   The collected data will be cleaned and some will be deleted. This process includes cleaning bad data, data that has empty attributes, abnormal data, as well as unused attributes when modeling. In this dataset we cleanse ip addresses that do not give a *ping response* (Hosseinzadeh et al., 2021).
3. Transformation
   Transformations are used to produce optimal performance in data modeling, some data that is not directly related is reduced to improve accuracy. The data transformation process is carried out on the dataset so that the data can be used in this research when modeling. This process can also have an effect on the modeling results displayed at the evaluation stage (Damayunita et al., 2022).
4. Modeling
   Modeling After completing the data cleansing and transformation phase, the modeling stage is run. At this stage, the results of classification and predictions are established. In this study, the Bayesian Naive modeling algorithm was used (Damayunita et al., 2022).
5. Evaluation
   The results of modeling experiments are displayed in the form of a fusion matrix or error matrix. This fusion matrix represents actual information about the modeling performed and also provides information in the form of accuracy results (Damayunita et al., 2022).

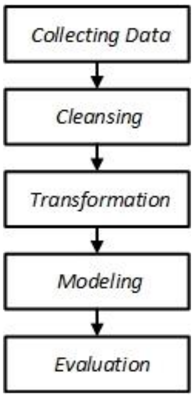

Figure. 1 Method

RESULTS AND DISCUSSION

Collecting Data

At this stage, data collected from kaggle.com about *the cyber attack global network* between 2022 to 2023 amounted to 40,000 data records and 25 attributes. The explanation of each attribute can be seen in table number 1, while the sample dataset can be seen in figure 2 with the attribute name placed in column 2 so that the example record can be read properly.

Table 1. Attribute Description

No	Attribute Name	Deskirption
1	timestamps	Logging date
2	source_ip_adress	The origin of the IP Address that carried out the attack
3	destination_ip_address	IP Address korban serangan
4	source_port	Asal port
5	destination_port	Purpose of the attacked port
6	protocol	Protocol type
7	packet_length	Large data packets sent
8	packet_type	Package type
9	traffic_type	Traffic by type
10	payload_data	Fill in the data sent
11	malware_indicators	Malware indicator
12	anomaly_scores	Nilai anomaly
13	alertswarnings	Automatic Alerts are called
14	attack_type	Types of attacks
15	attack_signature	Attack identifier
16	action_taken	Interceptor steps of attacks performed
17	severity_level	Attack damage level
18	user_information	User information
19	device_information	Application information used
20	network_segment	Network segments
21	Geolocation_data	Where the attack came from
22	proxy_information	Proxy information
23	firewall_logs	Firewall notes
24	idsips_alerts	IDS record
25	log_source	Logging origin

1	**timestamps**	2021-04-17 04:37:18	2020-02-19 04:10:17	2023-09-23 19:07:33	2023-02-20 06:41:55	2023-09-13 02:42:05
2	**source_ip_address**	195.52.158.206	105.83.233.209	203.171.62.228	19.14.168.54	44.24.112.64
3	**destination_ip_address**	71.162.236.14	71.10.113.172	27.5.94.221	68.144.93.235	71.28.47.114
4	**source_port**	52720	3394	41615	23870	27274
5	**destination_port**	35946	52170	15184	21385	28937
6	**protocol**	ICMP	TCP	ICMP	TCP	ICMP
7	**packet_length**	1076	100	1346	619	340
8	**packet_type**	Data	Data	Data	Data	Data
9	**traffic_type**	DNS	DNS	FTP	FTP	HTTP
10	**payload_data**	Earum aperiam ipsa n	Ad dolore nisi sequi	Magni blanditiis ver	Et magnam magnam v	Unde dolore vero dol
11	**malware_indicators**	IoC Detected		IoC Detected	IoC Detected	
12	**anomaly_scores**	85.75	64.63	67.73	67.56.00	27.25.00
13	**alertswarnings**	Alert Triggered			Alert Triggered	Alert Triggered
14	**attack_type**	Malware	Malware	Intrusion	Intrusion	Intrusion
15	**attack_signature**	Known Pattern A	Known Pattern A	Known Pattern A	Known Pattern B	Known Pattern A
16	**action_taken**	Logged	Blocked	Blocked	Logged	Blocked
17	**severity_level**	High	Low	Low	High	High
18	**user_information**	Lagan Butala	Vritika Andra	Bhavin Chaudhari	Siya Singhal	Indrajit Chahal
19	**device_information**	Opera/8.82.(Windows NT 5.1; pa-IN) Presto/2.9.174 Version/12.00	Mozilla/5.0 (Linux; Android 2.2.3) AppleWebKit/531.2 (KHTML, like Gecko) Chrome/28.0.857.0 Safari/531.2	Mozilla/5.0 (Android 7.1; Mobile; rv:22.0) Gecko/22.0 Firefox/22.0	Opera/9.85.(Windows NT 6.0; cmn-TW) Presto/2.9.161 Version/11.00	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/532.1 (KHTML, like Gecko) Chrome/62.0.877.0 Safari/532.1
20	**network_segment**	Segment A	Segment C	Segment A	Segment B	Segment B
21	**geolocation_data**	Patiala, Chhattisgarh	Jorhat, Andhra Prade	Jaunpur, Uttar Prades	Tiruppur, Bihar	Giridih, Himachal Pradesh
22	**proxy_information**		178.147.154.55			
23	**firewall_logs**				Log Data	
24	**idsips_alerts**	Alert Data	Alert Data	Alert Data		
25	**log_source**	Firewall	Firewall	Server	Server	Server

Figure 2. Datasets displayed horizontally

pISSN 2302-4240
eISSN 2655-2086
Volume 13, Issue 2, June 2024, pp.587-596
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

Cleaning Data

Once data collection is complete, the next step is data cleansing. At this stage, the data that is considered invalid is eliminated, bringing the total number of data to 8751 data to be retrieved. Of the 8751 data, 80% will be used as a training dataset and 20% will be used as a test dataset. The following script is used to perform the data cleaning process, created using PHP and run in the Linux console prompt.

```
<?php
$pgsql_host     = 'localhost'; //host
$pgsql_username = 'postgres'; //username
$pgsql_password = '123456'; //password
$pgsql_database = 'mandat'; //db
$connection = pg_connect("host='$pgsql_host' port=5432 dbname=$pgsql_database user='$pgsql_username' password='$pgsql_password'");

if (!$connection) {
        echo "Gagal koneksi database ";
        die();
}
function pinger($address){
        if(strtolower(PHP_OS)=='winnt'){
                $command = "ping -n 1 $address";
                exec($command, $output, $status);
        }else{
                $command = "ping -c 1 $address";
                exec($command, $output, $status);
        }
        if($status === 0){
                return true;
        }else{
                return false;
        }
}
$query = "select  ip_address from cyberattack_cleansing  order by ip_address";
$result = pg_query($query);
$i = 1;
while ($row = pg_fetch_assoc($result)) {

                $ip = $row["ip_address"];
                $live = pinger($ip);
                if ($live)
                        {  echo "$ip...ip address on";
                          $queryx = "update cyberattack set online = 1 where ip_address = '$ip'";
                          $resultx = pg_query($queryx);
                }
                else
                         {echo "$ip...ip address offline";}
                echo "\n";
}
?>
```

Figure 3. Cleansing

From the cleansing program shown in figure 3 after running, all records that have IP addresses that cannot be pinged will be deleted from the dataset, so that the dataset we use is only records whose ip addresses are active at the time this study is made. An example of the dataset we used can be seen in Figure 4.

1	timestamps	2022-05-14 14:06:23	2021-09-09 08:48:36	2023-09-11 00:24:22	2021-06-12 14:02:44	2021-12-12 21:20:42
2	source_ip_address	89.16.154.228	86.34.177.196	60.17.114.154	187.218.72.50	178.0.248.243
3	destination_ip_address	125.249.214.230	189.230.36.218	59.137.203.254	58.151.17.120	217.55.61.63
4	source_port	64395	49362	44776	63723	1229
5	destination_port	58080	4932	53896	19177	52816
6	protocol	TCP	ICMP	TCP	UDP	TCP
7	packet_length	711	1348	660	739	760
8	packet_type	Control	Data	Control	Control	Control
9	traffic_type	FTP	DNS	HTTP	DNS	FTP
10	payload_data	Maxime cupiditate e	Vero culpa et vel un	Placeat dolorum deb	Nihil omnis neque. A	Magni enim dolor ne
11	malware_indicators	IoC Detected			IoC Detected	IoC Detected
12	anomaly_scores	53.55.00	96.70	05.17	23.44	48.10.00
13	alertswarnings	Alert Triggered	Alert Triggered	Alert Triggered		Alert Triggered
14	attack_type	Intrusion	DDoS	Malware	DDoS	DDoS
15	attack_signature	Known Pattern B	Known Pattern A	Known Pattern A	Known Pattern B	Known Pattern A
16	action_taken	Blocked	Ignored	Ignored	Blocked	Blocked
17	severity_level	Low	Low	Low	Medium	Medium
18	user_information	Raunak Kapadia	Misha Chadha	Yashvi Garde	Kashvi Srivastava	Charvi Bora
19	device_information	Mozilla/5.0 (Windows 98; as-IN; rv:1.9.1.20) Gecko/5324-04-12 21:25:46 Firefox/3.6.2	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT 5.01; Trident/5.0)	Mozilla/5.0 (Windows NT 4.0; xh-ZA; rv:1.9.2.20) Gecko/9969-01-18 18:33:59 Firefox/5.0	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT 4.0; Trident/3.1)	Mozilla/5.0 (Windows 95; bs-BA; rv:1.9.2.20) Gecko/7318-12-19 22:50:07 Firefox/3.8
20	network_segment	Segment B	Segment A	Segment C	Segment A	Segment C
21	geolocation_data	Farrukhabad, Uttara	Akola, Telangana	Gulbarga, Mizoram	Mau, Mizoram	Fatehpur, Uttarakha
22	proxy_information		77.160.189.251			69.250.227.209
23	firewall_logs		Log Data	Log Data	Log Data	
24	idsips_alerts	Alert Data			Alert Data	Alert Data
25	log_source	Server	Server	Server	Server	Server

Figure 4. Results Dataset Cleansing

Transformation Data

The data transformation process establishes a performance evaluation of the algorithms used to convert data, with the features used as encoders being destructive_power, severity_t, attack_t, action_t, malware_indicators_t, and anomaly_scores.

Table 2. Data Transformation

Destructive_power character varying (20)	Severity_t integer	Attack_t integer	Action_t integer	Malwere_indicators_t integer	Anomaly_scores numeric(5,2)
Low	1	3	3	2	7,26
Low	1	3	3	2	44,98
Low	1	3	3	2	2,41
Low	1	3	3	2	40,58
High	3	2	3	2	63,83
High	3	2	3	2	77,35
High	3	3	3	2	65,92
Medium	2	2	3	2	15,81
Medium	2	2	3	2	51,51
Medium	2	2	3	2	12,66

Modeling

This stage aims to find the classification and prediction results from Syber. This study used Naive Bayes algorithm modeling. The formula used to describe Bayes' Naive theorem is listed in formula (1).

$$P(H|X) = \frac{p(X|H)p(H)}{P(X)} \quad (1)$$

Information:

pISSN 2302-4240
eISSN 2655-2086

Volume 13, Issue 2, June 2024, pp.587-596
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

X = Data whose class is still not identified.

H = Hypothesis that data X belongs to a certain class.

p(H|X)= Possible hypothesis H taking into account condition X.

p(X|H)= The probability of X given the conditions present in hypothesis H.

p(H)= Possible of hypothesis H.

p(X)= Possible occurrence of X.

Data Editor

Row No.	destructive_power (polynominal) label	severity_t (integer) regular	attack_t (integer) regular	action_t (integer) regular	malware_indicators_t (integer) regular	anomaly_scores (real) regular
42	3	3	3	3	2	77.190
43	2	3	3	1	2	79.290
44	3	3	3	3	2	60.040
45	3	3	1	3	2	77.900
46	2	2	1	3	2	50.170
47	3	3	3	3	2	62.430

Repository Location: //Local Repository/processes/attack

Figure 5. Dataset Attack

From the results of data transformation, it is continued by entering the dataset in Figure 5 into the datamining design process in the rapidminer application as shown in figure 6.

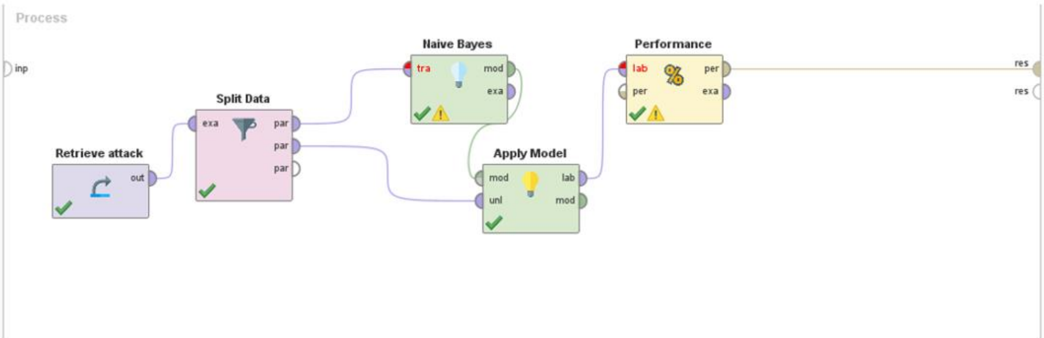

Figure 6. Datamining Design

Evaluation

The next step in determining the accuracy value of modeling carried out on test data is to test the test data that has been separated before.

accuracy: 84.53%

	true High	true Medium	true Low	class precision
pred. High	450	41	90	77.45%
pred. Medium	0	815	119	87.26%
pred. Low	0	156	954	85.95%
class recall	100.00%	80.53%	82.03%	

Figure 7. Naïve Baiyes Classification Results

The results of execution in datamining design can be seen in figure 8 which shows the accuracy of this study of 84.53% with detailed results as follows:

- Predicted High and Turned High as much as 450
- High prediction and it turns out Medium as much as 0
- High prediction and it turns out Low as much as 0
- Medium prediction and it turns out High as much as 41
- Medium prediction and it turns out that Medium is 815
- Medium prediction and it turns out Low as much as 156
- Low predicted results and it turned out to be High as much as 90
- Low predicted results and it turned out to be Medium as much as 119
- Low's prediction results and it turned out that Low was 954

Table 3. Value Confidence

No	attack_n	action_n	severity_n	anomaly_scores	destructive_power	Confidence (High)	Confidence (Medium)	Confidence (Low)	Prediction (destructive_power)
1	3	3	3	59,5	High	0,7	0,2	0,0	High
2	3	3	3	61,2	High	0,7	0,2	0,0	High
3	1	2	2	92,2	Medium	0,0	0,9	0,1	Medium
4	1	3	2	56,5	Medium	0,0	1,0	0,0	Medium
5	1	2	2	89,6	Low	0,0	0,9	0,1	Medium
6	1	2	2	59,3	Low	0,0	0,6	0,4	Medium
7	3	2	2	9,3	Medium	0,0	0,3	0,7	Low
8	1	3	2	63,4	Medium	0,0	1,0	0,0	Medium
9	1	1	1	49,0	Low	0,0	0,2	0,8	Low
10	..								

Table 3 describes statistical data from the application of the Naïve Baiyes algorithm for the calculation of prediction results on destruktive_power labels from attack_n, action_n, severity_n, and anomaly_scores data, for example High's accurate prediction results are highly dependent on the high value of attack_n, action_n, severity_n and anomali_score data, which shows a considerable damage impact on the system if this data is of high value, And vice versa the impact of damage is low if the data is attack_n, action_n, severity_n and anomali_score of low value.

CONCLUSION

From the results of this study, the author can draw several conclusions as follows: 1) the use of the Naïve Baiyes classification model has proven to be effective in classifying cyberattack datasets with an accuracy rate of 84.53%. 2) the level of accuracy in testing with high prediction results reached 77.45%, Medium prediction results reached 87%, Low prediction results reached 85.95%. The results confirm the accuracy of Naive Bayes in classifying cyberattacks and recommend their combination with other methods for a more effective cyber defense strategy.

REFERENCES

Abu Al-Haija, Q., & Al-Fayoumi, M. (2023). An intelligent identification and classification system for malicious uniform resource locators (URLs). *Neural Computing and Applications*, 1–17.

Amin, B. M. R., Taghizadeh, S., Maric, S., Hossain, M. J., & Abbas, R. (2020). Smart grid security enhancement by using belief propagation. *IEEE Systems Journal*, *15*(2), 2046–2057.

Bécue, A., Praça, I., & Gama, J. (2021). Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities. *Artificial Intelligence Review*, *54*(5), 3849–3886.

Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Constrained Naïve Bayes with application to unbalanced data classification. *Central European Journal of Operations Research*, 1–23.

Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, *2021*(1), 1–12.

Chu, S.-C., Dao, T.-K., Pan, J.-S., & Nguyen, T.-T. (2020). Identifying correctness data scheme for

pISSN 2302-4240
eISSN 2655-2086
Volume 13, Issue 2, June 2024, pp.587-596
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

aggregating data in cluster heads of wireless sensor network based on naive Bayes classification. *EURASIP Journal on Wireless Communications and Networking*, *2020*, 1–15.

Damayunita, A., Fuadi, R. S., & Juliane, C. (2022). Comparative Analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) Algorithms for Classification of Heart Disease Patients. *Jurnal Online Informatika*, *7*(2), 219–225.

Fahd, K., Miao, Y., Miah, S. J., Venkatraman, S., & Ahmed, K. (2022). Knowledge graph model development for knowledge discovery in dementia research using cognitive scripting and next-generation graph-based database: a design science research approach. *Social Network Analysis and Mining*, *12*(1), 61.

Hosseinzadeh, M., Azhir, E., Ahmed, O. H., Ghafour, M. Y., Ahmed, S. H., Rahmani, A. M., & Vo, B. (2021). Data cleansing mechanisms and approaches for big data analytics: a systematic study. *Journal of Ambient Intelligence and Humanized Computing*, 1–13.

Kim, T., & Lee, J.-S. (2022). Exponential loss minimization for learning weighted naive bayes classifiers. *IEEE Access*, *10*, 22724–22736.

Kuhn, K., Bicakci, S., & Shaikh, S. A. (2021). COVID-19 digitization in maritime: understanding cyber risks. *WMU Journal of Maritime Affairs*, *20*(2), 193–214.

Lin, C.-J., Huang, M.-S., & Lee, C.-L. (2022). Malware Classification Using Convolutional Fuzzy Neural Networks Based on Feature Fusion and the Taguchi Method. *Applied Sciences*, *12*(24), 12937.

Redivo, E., Viroli, C., & Farcomeni, A. (2023). Quantile-distribution functions and their use for classification, with application to naïve Bayes classifiers. *Statistics and Computing*, *33*(2), 55.

Sarker, I. H. (2023). Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. *Annals of Data Science*, *10*(6), 1473–1498.

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, *7*, 1–29.

Schoenenwald, A., Kern, S., Viehhauser, J., & Schildgen, J. (2021). Collecting and visualizing data lineage of Spark jobs: Digesting Spark execution plans to surface lineage graphs via a full-stack application. *Datenbank-Spektrum*, *21*, 179–189.

Shi, Y., Lu, X., Niu, Y., & Li, Y. (2021). Efficient jamming identification in wireless communication: Using small sample data driven naive bayes classifier. *IEEE Wireless Communications Letters*, *10*(7), 1375–1379.

Shukla, P., Krishna, C. R., & Patil, N. V. (2023). EIoT-DDoS: embedded classification approach for IoT traffic-based DDoS attacks. *Cluster Computing*, 1–20.

Syrmakesis, A. D., Alcaraz, C., & Hatziargyriou, N. D. (2022). Classifying resilience approaches for protecting smart grids against cyber threats. *International Journal of Information Security*, *21*(5), 1189–1210.

Veeramanickam, M. R., Khullar, V., Salunke, M. D., Bangare, J. L., Bhosle, A. A., & Ingavale, A. (2022). Streamed Incremental Learning for Cyber Attack Classification using Machine Learning. *2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT)*, 1–5.

Sari, O. L., Basyaruddin, B., & Khasanah, U. (2024). Building Maintenance Priority Decision Support System Using the Method Profile Matching. ASTONJADRO, 13(1), 125–137. https://doi.org/10.32832/astonjadro.v13i1.14495

Nur Aulia, A., Tsani, M., & Suharso, W. (2024). Design a Web-Based Library Information System Using the Waterfall Method (Case Study of SMA Muhammadiyah 2). ASTONJADRO, 13(1), 169–181. https://doi.org/10.32832/astonjadro.v13i1.14562

Pradnyana, I. M., Widyantara, I. M. O., & Pramaita, N. (2023). Evaluation of DVB-T2 Digital TV Propagation Performance in the Bali Broadcast Area. ASTONJADRO, 12(3), 886–896. https://doi.org/10.32832/astonjadro.v12i3.14311

Prastowo, F. I., Husin, A. E., & Amalia, N. (2023). Improving Project Performance Based on Building Information Modelling 6D & LCCA in High-Rise Office Building. ASTONJADRO, 12(2), 368–378. https://doi.org/10.32832/astonjadro.v12i2.8787