Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (53-63)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.42815

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

Challenges and Strategies in Forensic Investigation: Leveraging Technology for Digital Security Using Log/Event Analysis Method

Ammar Yasir Nasution[1*], Hartono[2], Rika Rosnelly[3]

[1,3]Computer Science, Computer Science and Engineering, Potensi Utama University
[2]Computer Science, Computer Science and Engineering, Medan Area University
[1,3]Jl. K.L. Yos Sudarso KM 6,5 No. 3A Tj. Mulia, Indonesia
[2]Jl. Kolam Nomor 1 Medan Estate, Medan, Indonesia

Article:
Accepted: January 02, 2025
Revised: December 02, 2024
Issued: April 30, 2025

***Correspondence Address:**
yasirnasti0396@gmail.com

ABSTRACT

Cybersecurity threats continue to evolve, necessitating advanced techniques for network anomaly detection. This study developed a comprehensive methodology for detecting network anomalies by leveraging sophisticated log and event analysis using machine learning algorithms. By employing a Naive Bayes classification approach on a synthetic cybersecurity dataset comprising 40,000 entries with 25 unique features, the research aimed to enhance anomaly detection precision. The methodology involved meticulous data preprocessing, feature selection, and strategic model validation techniques, including cross-validation and external benchmarking. Comparative analysis with K-Nearest Neighbors and Support Vector Machine algorithms demonstrated the Naive Bayes method's superior performance, achieving a classification accuracy of 94.8%, an Area Under the Curve (AUC) of 0.949, and a Matthews Correlation Coefficient of 0.896. The study identified critical parameters influencing anomaly detection, such as source port characteristics and attack signatures. These findings contribute significant insights into machine learning-based network security strategies, offering a robust framework for early threat identification and mitigation.

Keywords : *network anomaly detection; naive bayes classification; machine learning; cybersecurity; log analysis.*

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (53-63)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.42815

1. INTRODUCTION

The rapid technological advancement has been accompanied by increasing digital vulnerabilities and cyber threats, thereby necessitating more sophisticated digital security approaches and forensic investigations. In the context of escalating cybercrime, digital forensics has emerged as a critical discipline for addressing complex technological challenges [1].

The contemporary digital landscape presents intricate security challenges, particularly as organizations increasingly rely on interconnected digital infrastructures[2][3]. Cybercrime has evolved into a sophisticated domain where technological innovations are simultaneously employed as both tools and targets of malicious activities [4]. The continuously escalating complexity of digital threats demands investigative methodologies capable of effectively analyzing, detecting, and mitigating potential security breaches [5] [6].

Log and event analysis represents a promising approach in digital forensic investigations, offering unique capabilities in anomaly detection and comprehensive security assessment [7]. This method, through data integration from diverse sources and advanced techniques such as machine learning algorithms, enables forensic experts to reconstruct event timelines, identify potential actors, and validate investigative hypotheses [8], [9], [10], [11], [12].

However, despite its significant potential, log and event analysis confronts several substantial challenges, including massive data volumes, diverse log formats, complex data interpretation, and increasingly sophisticated anti-forensic techniques [13] [14]. To address these challenges, technologies such as Security Information and Event Management (SIEM) have emerged as crucial tools facilitating real-time security analysis and effective log management [15].

This research aims to explore the complex landscape of digital forensic investigations with several primary focal points: evaluating the effectiveness of log and event analysis in malware detection, identifying investigative challenges across various forensic stages, optimizing SIEM technology utilization for comprehensive security analysis, and developing effective forensic reporting and documentation strategies[16] [17][18].

By concentrating on leveraging technological innovations to enhance digital security, this research is expected to provide meaningful contributions in confronting increasingly sophisticated cyber threats and strengthening organizational digital defense mechanisms[19]. The study adopts a comprehensive approach utilizing the Naive Bayes algorithm for anomaly detection and log dataset analysis, with the objective of providing practical recommendations for enhancing digital forensic investigation strategies [20].

2. METHODS

The log/event analysis method refers to a structured approach used to review, interpret, and analyze logs or event data generated by systems, networks, software applications, or devices. This process is essential in identifying patterns, troubleshooting issues, ensuring system security, and gaining insights into system behavior over time. Event logs typically contain timestamped records of activities, transactions, system operations, errors, or warnings, which can be mined for useful information. The methode behind log/event analysis spans several key principles and methodologies, involving data collection, aggregation, normalization, correlation, visualization, and interpretation.

This research develops a comprehensive methodology for malware detection by leveraging in-depth log and event analysis to address critical challenges in cybersecurity forensic investigations [21]. [22] By integrating advanced data processing techniques and machine learning algorithms, the study aims to create a robust approach for identifying potential security threats within complex forensic technology environments.

2.1. Data Collection and Preprocessing

The approach commenced with meticulous data collection, utilizing a secondary dataset from the Kaggle platform [23]. The dataset encompassed event logs from diverse sources, including servers, routers, firewalls, and cookie systems, providing a rich and complex analytical landscape. Recognizing the dataset's variability and limitations, a staged preprocessing strategy was implemented to

transform raw data into a structured and analyzable format [24].

2.1.1. Data Analysis

a. Preprocessing Data

Data cleaning: Removes invalid or duplicate entries; Data transformation; Change the data format to fit the analysis needs. ; Feature extraction: Identify and select relevant features for malware analysis.

b. Implementation

Implementation of the Naive Bayes Algorithm Dataset sharing: Divide the data into training and test sets. ;P etraining model: Training the Naive Bayes model using training data. ;P Model testing: Testing the model using test data.

c. Model Evaluation

Calculate performance metrics: Accuracy, precision, recall, and F1-score; Confusion matrix analysis to understand the types of classification errors; Cross-validation

to assess model generalizations.

d. Anomaly Analysis

Application of statistical techniques for the identification of outliers and abnormal patterns; The use of unsupervised learning for clustering and anomaly detection.

e. SIEM Optimization

Configure correlation rules to connect events from multiple sources; Dashboard implementation for real-time visualization of network activity ;P alert settings for quick notification of potential threats.

f. Data Analysis of Strategic Approaches to Research

This study adopts a strategic approach to overcome challenges in IT forensic investigation using log/event analysis methods. Each stage of the research is designed with specific strategies to optimize the process and results.

2.1.2. Data Collection and Preparation Strategies

a. Dataset Selection:

Strategy: Conduct a comprehensive and relevant selection of Kaggle datasets.; Tactics: a. Evaluate multiple datasets based on criteria of relevance, novelty, and completeness. b. Cross-referencing with the latest literature to ensure the suitability of datasets with current IT forensic practices.

b. Data Integrity Validation:

Strategy: Ensuring the authenticity and integrity of the dataset.; Tactics: a. Using hashing techniques to verify the integrity of dataset files. b. Check the internal consistency of the data to identify anomalies.

c. Bias Mitigation Strategies:

Strategy: Identify and mitigate potential biases in the dataset; Tactics: a. Conduct statistical analysis to uncover potential biases; Implementation of resampling or weighting techniques if class imbalances are found.

2.1.3. Data Preprocessing and Normalization Strategy

a. Log Format Standardization:

Strategy: Convert various log formats into a consistent standard format. ; Tactics: a. Custom parser development for each log type in the dataset. b. Implementasi pipeline ETL (Extract, Transform, Load) menggunakan tools seperti Logstash.

b. Feature Engineering:

Strategy: Extracting and creating the most informative features for malware analysis; Tactics: a. Expert domain analysis to identify key features in malware detection. b. The application of automatic feature selection techniques such as PCA or feature importance from the decision tree model.

c. Lost Data Handling Strategies:

Strategy: Minimize the impact of lost or incomplete data. ; Tactics: a. Implementation of intelligent imputation techniques based on the context of the event log. b. Development of models to predict missing values based on patterns in existing data.

2.1.4. Log/Event Analysis Implementation Strategy

a. Naive Bayes Algorithm Optimization:

Strategy: Improve Naive Bayes' performance for malware detection; Tactics: a. Experiment with Naive Bayesian variations (e.g., Gaussian, Multinomial, Bernoulli) to find the most suitable. b. Implementation of smoothing techniques to overcome the zero probability problem.

b. Anomaly Detection Strategy:

Strategy: Integrate anomaly detection methods to improve the accuracy of malware identification; Tactic; Application of Isolation Forest algorithm for outlier identification. b. The use of clustering techniques such as DBSCAN to classify normal vs. anomalous behavior.

c. Multi-source Event Correlation:

Strategy: Develop methods to connect events from various log sources; Tactics: a. Implementation of graph-based analysis to map the relationship between events. b. Development of time- and context-based correlation rules to connect related events.

2.1.5. Model Evaluation and Validation Strategies

a. Robust Cross-Validation:

Strategy: Ensuring good model generalization; Tactics: a. Implementation of k-fold cross-validation with stratification to maintain class distribution. b. The use of bootstrapping techniques for estimating the confidence interval of model performance.

b. Comprehensive Evaluation Metrics:

Strategy: Provides a complete picture of the model's performance; Tactics: a. Analysis of ROC and PR curves for understanding trade-off sensitivity vs. specific; Implementation of custom metrics that reflect cost-benefit in the context of IT forensics.

c. External Validation:

Strategy: Validating results against external benchmarks.; Tactics: a. Compare the results with similar published studies. b. If possible, validate with an IT forensic expert to gain actionable insights.

2.1.6. Reporting and Visualization Strategies

a. Stakeholder-Oriented Narrative:

Strategy: Drafting reports that can be understood by different levels of technical stakeholders; Tactics: a. Development of multi-level report templates (executive, managerial, technical). b. Use of language tailored to every level of technical understanding.

b. Interactive Visualization:

Strategy: Improve understanding of results through interactive visualizations. ; Tactics: a. Interactive dashboard development using tools such as Tableau or D3.js. b. Implementation of drill-down capability for deeper data exploration.

2.1.7. G.Strategies to Mitigate Dataset Limitations

a. Synthetic Data Augmentation:

Strategy: Enrich datasets with synthetic data to increase variationTactics: a. The use of generative modeling techniques to create synthetic log events. b. Validate synthetic data with domain experts to ensure realism.

b. Transfer Learning:

Strategy: Leverage knowledge from related domains to improve performance on limited datasets; Tactics: a. Exploration of pre-trained models of relevant cybersecurity domains. b. Fine-tuning models on Kaggle datasets for adaptation to specific contexts.

2.2. Analysis Framework and Algorithm

The analytical framework centered on the Naive Bayes classification algorithm, selected for its probabilistic approach and capability to handle high-dimensional categorical data [25]. Employing supervised learning techniques, the model was trained to distinguish between normal and suspicious system behaviors. The dataset was strategically divided into training and testing subsets to ensure comprehensive model validation.

To enhance detection capabilities, the methodology integrated multiple anomaly detection strategies [26], including:

1. Statistical Analysis: Establishing system behavior baselines.
2. Dynamic Behavioral Analysis: Leveraging machine learning techniques.
3. Rule-Based Detection: Adding an additional oversight layer to identify threats.

2.3. Model Performance Evaluation and Validation

Model evaluation extended beyond mere accuracy measurements. Researchers utilized various performance metrics, including precision**,** recall**,** F1-score, and confusion matrix analysis. This holistic evaluation ensured a comprehensive understanding of the model's strengths and limitations.

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (53-63)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.42815

The research also implemented sophisticated validation strategies, including:

1. Cross-Validation Techniques: To assess model generalizability.
2. External Benchmark Comparisons: Involving domain expert validation.
3. Limitation Documentation: Ensuring transparency regarding dataset and methodological constraints.

2.4. Advanced Techniques and Practical Implementation

Techniques such as data augmentation and transfer learning to enhance the model's adaptability and performance across diverse computing environments. Practical implementation was supported by tools like ManageEngine EventLog Analyzer and advanced machine learning and statistical analysis libraries.

3. RESULTS AND DISCUSSION

3.1. Result

3.1.1. Cybersecurity Dataset Characteristics

This research utilized a synthetic cyber log/event dataset published in January 2023, sourced from Kaggle.com. The dataset comprises 25 unique features and 40,000 data rows, meticulously designed to realistically represent computer network data traversal history.

a. Data Preprocessing Process

During the preprocessing stage, the research identified and addressed missing data challenges in critical features, including:

1. Malware Indicators: Empty data supplemented with "No IoC Detected"
2. Alerts/Warnings: Empty data populated with "No Alert Triggered"
3. Proxy Information: Empty data replaced with zero values
4. Firewall Logs: Empty data appended with "No Log Data"
5. IDS/IPS Alerts: Empty data filled with "No Alert Data"

This approach aimed to:

1. Enhance dataset quality
2. Minimize analysis bias
3. Maximize machine learning model performance

b. Analysis Technique

The research employed a stochastic approach using Naive Bayes Machine Learning Algorithm to:

1. Detect attack patterns
2. Identify network anomalies
3. Analyze network activities based on logs/events

Table 1. *Data preprocessing summary*

Feature Category	Initial Data Challenge	Preprocessing Resolution Strategy
Malware Indicators	Partial data omissions	Standardization with "No IoC Detected" placeholder
Alerts/Warnings	Intermittent data losses	Uniform imputation with "No Alert Triggered"
Proxy Information	Incomplete data entries	Replacement with numerical zero representation
Firewall Logs	Fragmented log records	Supplementation with "No Log Data" default value
IDS/IPS Alerts	Inconsistent data points	Consistent filling with "No Alert Data" descriptor

c. Feature Selection

The feature selection process was conducted to optimize the predictive model by selecting the most relevant features. From the total 26 input features, the research successfully identified:

1. 17 primary features used in modeling
2. 1 target feature for Anomaly Detection
3. 8 features excluded due to minimal contribution

Table 2. *Feature selection summary*

Feature Category	Quantity	Descriptive Characterization
Total Input Features	26	Encompassing primary and meta-attribute dimensions
Core Model Features	17	Fundamental features utilized in model construction
Target Detection	1	Binary anomaly classification objective
Removed Features	8	Minimal contribution to detection probabilistic space

d. Data Distribution

To ensure classification accuracy, the research utilized a dataset with:

1. Total 40,000 data rows
2. Balanced distribution between "Anomaly" and "No Anomaly" categories
3. Each category: 20,000 data rows

e. Modeling with Naive Bayes

The research implemented the Naive Bayes algorithm with:

1. Probability calculations based on Bayes' Theorem
2. Prior probability analysis for each feature
3. Training data division: 80% (32,000 rows)
4. Test data division: 20% (8,000 rows)

Table 3. Dataset Compositional Distribution

Classification Category	Data Volume	Probabilistic Representation
Anomaly	15,960	0.499 (49.9%)
Non-Anomaly	16,040	0.501 (50.1%)

Prediction and Validation In the test data, the model was capable of:

1. Classifying anomaly potential with high accuracy
2. Distinguishing between attacks and normal activities
3. Demonstrating precise detection capabilities

3.1.2. Anomaly Detection Model Performance Evaluation

This research developed an anomaly prediction model using the Naïve Bayes Algorithm with a data division technique of 80% training and 20% test data, involving a total of 160,000 data points. The model evaluation results demonstrated exceptional performance across various measurement metrics:

Table 4. *Model evaluation metrics summary*

Evaluation Metrics	Value	Interpretation
Classification Accuracy (CA)	0.948 (94.8%)	Exceptionally high classification accuracy level
Area Under the Curve (AUC)	0.949	Excellent model discrimination capability
Precision	0.948 (94.8%)	High accuracy in positive class prediction
Recall	0.948 (94.8%)	Capability to identify entire positive cases
F1 Score	0.948 (94.8%)	Balance between precision and recall
Matthews Correlation Coefficient (MCC)	0.896	Classification quality approaching perfection

a. Confusion Matrix Analysis

Prediction results using the Confusion Matrix revealed:

1. 75,662 Anomaly data points correctly predicted
2. 76,056 No Anomaly data points correctly predicted
3. Prediction error rate of 5.2% for both classes

b. ROC Analysis and Model Optimization

The ROC Analysis curve generated a probability score of 0.934, demonstrating the Naïve Bayes model's exceptional capability in differentiating positive and negative classes.

c. Optimal Parameter Identification

Utilizing Stochastic Gradient Descent, the research identified 37 critical parameters influencing anomaly detection. Significant attributes included:

1. Packet Type: Control
2. Malware Indicators: IoC Detected
3. Firewall Logs: No Log Data

3.2. Discussion

3.2.1. Comprehensive Network Anomaly Detection Analysis

a. Feature Selection: Dataset Optimization

The feature selection process in this research generated a significant optimization strategy for network security analysis. From the initial 19 input features, the study successfully extracted 17 critical features focusing on crucial aspects of network traffic and attack indicators.

Table 5. *Feature selection summary*

Category	Number of Features	Description
Initial Input	19	Total features before selection
Output Features	17	Critical features retained
Removed Features	8	Features with minimal contribution
Target	1	Anomaly Detection

Methodological Advantages

1. Computational Efficiency: Substantial data dimensionality reduction significantly accelerates the analysis process.
2. Critical Information Preservation: Retaining key features such as:
   a. Source/destination ports
   b. Protocols
   c. Packet length
   d. Attack indicators
3. Ethical Considerations: Elimination of sensitive information like specific IP addresses and individual user data.

b. Data Imbalance Analysis

The dataset distribution demonstrates an extraordinary balance:

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (53-63)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.42815

1. No Anomaly: 20,050 samples (50.1%)
2. Anomaly: 19,950 samples (49.9%)

Balance Significance

1. Difference of only 100 samples (0.2%)
2. Reduction of model bias risk
3. Enhancement of metric evaluation representativeness

c. Feature Correlation Analysis

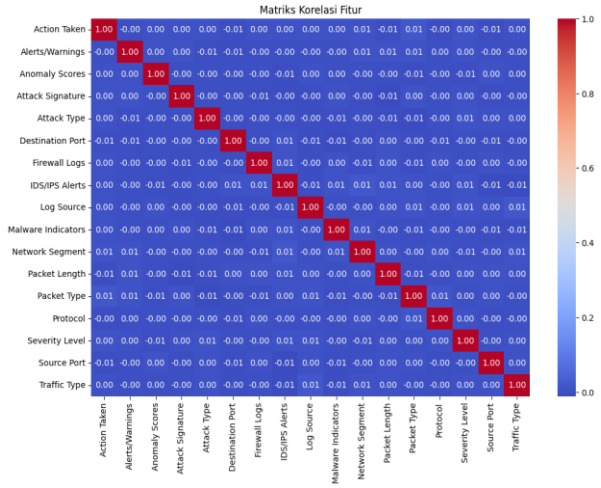

Figure 1. *Correlation matrix insights (17x17)*

Feature correlations were extremely low, ranging from -0.01 to 0.01, indicating feature independence. Each feature potentially offers unique insights into network anomaly detection.

The Confusion Matrix revealed high precision, with 94.8% accurate predictions for both anomalous and non-anomalous cases. This demonstrates the model's exceptional capability to distinguish between normal and potentially suspicious network activities.

Evaluation using diverse statistical indicators consistently demonstrated superior model performance. The metrics included Accuracy, Precision, Recall, F1 Score, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC).

Table 6. Model Evaluation Metrics Summary for Anomaly Detection

Metric	Value	Interpretation
Accuracy	94.8%	Model prediction success rate
Precision	94.7%	Anomaly prediction accuracy
Recall	94.8%	Actual anomaly detection capability
F1 Score	94.8%	Precision and recall balance
MCC	0.896	Binary classification quality
AUC	0.949	Class discrimination capability

The model's primary advantage lies in minimizing false positives and false negatives. With only 5.2% prediction errors, the model offers high reliability in detecting potential network threats. This capability is critically important in cybersecurity, where early detection can prevent significant potential losses.

Interestingly, the Naive Bayes algorithm demonstrated effectiveness despite assuming feature independence. The model's performance indicates that this assumption does not substantially reduce detection accuracy in network anomaly contexts.

The research compared the performance of three machine learning algorithms for network log/event anomaly detection: Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). A comprehensive analysis using various evaluation metrics yielded significant findings about each algorithm's performance.

Table 7. *Classification algorithm performance comparison*

Algorithm	AUC	Accuracy	F1-Score	Precision	Recall	MCC
Naive Bayes	0.949	0.948	0.948	0.948	0.948	0.896
KNN	0.511	0.508	0.508	0.508	0.508	0.015
SVM	0.853	0.771	0.769	0.771	0.771	0.552

Naive Bayes demonstrated superior performance with remarkable consistency across evaluation metrics. This superiority stems from the algorithm's ability to assume feature independence and its effectiveness on datasets with stable probability distributions. This aligns with the characteristic of network log/event datasets that exhibit relatively structured probability patterns.

Support Vector Machine (SVM) ranked second, displaying solid performance, particularly in handling complex data through optimal hyperplane class separation. The algorithm proves especially effective for datasets with non-linear characteristics and demonstrates excellent generalization capabilities.

In contrast to the previous two algorithms, KNN exhibited the weakest performance, with accuracy approaching random guessing. These limitations suggest the algorithm's inability to capture complex patterns in network datasets, likely due to the intricate data distribution characteristics.

The Receiver Operating Characteristic (ROC) analysis further confirmed these findings, with Naive Bayes displaying the highest Area Under Curve (AUC) of 0.934, indicating exceptional discrimination capability in distinguishing anomaly classes.

3.2.2. Classification Algorithm Performance Evaluation in Anomaly Detection

This research utilized three classification algorithms (Naive Bayes, k-Nearest Neighbors/kNN, and Support Vector Machine/SVM) to detect anomalies in network log/event data. The analysis revealed significant variations in each algorithm's ability to classify anomalous data.

a. Classification Algorithm Characteristics: Comprehensive Analysis

The investigation delves into three sophisticated machine learning algorithms, each presenting unique methodological approaches to data classification and anomaly detection.

1. **Naive Bayes:** A Probabilistic Classification Paradigm The Naive Bayes algorithm represents a fundamental probabilistic approach to classification, distinguished by its computational simplicity and probabilistic foundation. Despite its seemingly simplistic assumption of feature independence, the algorithm demonstrates remarkable efficacy in data classification. By calculating the conditional probability of each class and leveraging Bayes' theorem, it provides a robust mechanism for categorizing complex datasets with computational efficiency.
2. **k-Nearest Neighbors (kNN):** Distance-Based Pattern Recognition k-Nearest Neighbors emerges as a flexible, non-parametric classification methodology centered on proximity-based pattern recognition. The algorithm's core mechanism involves measuring the spatial relationships between test data points and training instances, typically utilizing Euclidean or Manhattan distance metrics. By identifying the k most proximate neighbors and determining class membership through majority voting or weighted contribution, kNN offers exceptional adaptability in recognizing intricate and non-linear data patterns.
3. **Support Vector Machine (SVM):** Hyperplane Optimization for Complex Classification Support Vector Machine represents an advanced classification algorithm specifically engineered to maximize class separation through optimal hyperplane identification. By strategically transforming input features into higher-dimensional spaces via kernel functions, SVM excels in handling datasets characterized by complex, non-linear relationships. The algorithm's primary objective involves constructing a decision boundary that maximizes the margin between distinct classes, thereby enhancing predictive generalization and minimizing classification errors.

b. Critical Factors in Anomaly Detection: Stochastic Gradient Descent Analysis

The Stochastic Gradient Descent (SGD) analysis unveiled a nuanced exploration of parameters influencing anomaly detection, categorized into positive and negative impact domains.

Table 8. *Parameters with positive significant influence*

Parameter	Coefficient	Interpretative Insight
Source Port	0.097849	Potential indicator of suspicious network communication patterns
Attack Signature	0.0625674	Identification of recognized malicious activity patterns
Alerts/Warnings	0.0311252	Elevated probability of anomalous event occurrence

Table 9. *Parameters with Negative Significant Influence*

Parameter	Coefficient	Interpretative Insight
IDS/IPS Alerts	-0.686991	Reduced anomaly probability during active monitoring
Severity Level = Medium	-0.0439915	Inverse correlation with anomalous event likelihood

c. Analytical Implications

The comprehensive analysis illuminates the intricate interactions between network parameters and their probabilistic manifestations in anomaly detection. By systematically examining these multidimensional factors, researchers can develop increasingly sophisticated classification models capable of distinguishing

between normative and potentially malicious network behaviors.

The statistical coefficients provide profound insights into the complex mechanisms underlying network security assessment. Notably, the negative coefficients—particularly for IDS/IPS alerts—suggest that robust monitoring systems significantly mitigate the potential for undetected anomalous activities.

This sophisticated methodology, integrating advanced machine learning algorithms with rigorous statistical analysis, represents a pivotal approach to enhancing network security and threat detection capabilities. The research underscores the importance of multi-algorithmic strategies in developing comprehensive anomaly detection frameworks.

The findings highlight the necessity of continuous refinement in machine learning techniques, emphasizing that no single algorithm provides a universal solution. Instead, a nuanced, context-aware approach that leverages the strengths of diverse classification methodologies emerges as the most promising strategy in contemporary network security research.

CONCLUSION

The research presents a sophisticated approach to network anomaly detection through advanced machine learning methodologies, demonstrating significant insights into computational cybersecurity strategies. By employing a meticulously preprocessed synthetic dataset comprising 40,000 data rows with 25 unique features, the study effectively addresses critical challenges in network security analysis. The Naive Bayes algorithm emerged as the most outstanding classification technique, achieving an exceptional classification accuracy of 94.8% and consistently outperforming alternative machine learning approaches such as k-Nearest Neighbors and Support Vector Machine.

The comparative analysis of machine learning algorithms revealed nuanced performance variations, with Naive Bayes demonstrating superior capabilities in handling network log/event datasets. The algorithm's effectiveness, despite its fundamental assumption of feature independence, highlights the complexity of anomaly detection in network security contexts. The Stochastic Gradient Descent analysis further enriched the research by identifying critical parameters influencing anomaly detection, such as source ports, attack signatures, and monitoring alerts, thereby providing a multidimensional understanding of network security dynamics.

Ultimately, the study underscores the evolving landscape of cybersecurity research, emphasizing that no single algorithmic approach provides a universal solution. Instead, the findings advocate for a sophisticated, context-aware methodology that integrates diverse classification techniques and continuous computational refinement. By minimizing false positives and false negatives with only a 5.2% prediction error rate, the research offers a promising framework for early threat detection, potentially mitigating significant potential cybersecurity risks across complex network environments.

REFERENCES

[1] N. Fadila, G. Goso, R. Hamid, and I. Ukkas, “Pengaruh Literasi Keuangan, Financial Technology, Persepsi Risiko, dan Locus of Control Terhadap Keputusan Investasi Pengusaha Muda,” *Owner*, vol. 6, pp. 1633–1643, Mar. 2022, doi: 10.33395/owner.v6i2.789.

[2] S. Sheoran and D. Mahna, “Enhancing Forensic Voice Analysis with the Aid of Noise Cancellation: A Forensic Approach,” in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–7. doi: 10.1109/ICCCNT56998.2023.10308303.

[3] B. K. Sharma, M. A. Joseph, B. Jacob, and B. Miranda, “Emerging trends in Digital Forensic and Cyber security- An Overview,” in *2019 Sixth HCT Information Technology Trends (ITT)*, 2019, pp. 309–313. doi: 10.1109/ITT48889.2019.9075101.

[4] G. Suarez-Tangil, J. E. Tapiador, P. Peris-Lopez, and A. Ribagorda, “Evolution, detection and analysis of malware for smart devices,” *IEEE Commun. Surv. tutorials*, vol. 16, no. 2, pp. 961–987, 2013.

[5] F. Jimmy, "Emerging threats: The latest cybersecurity risks and the role of artificial intelligence in enhancing cybersecurity defenses," *Val. Int. J. Digit. Libr.*, pp. 564–574, 2021.

[6] S. Agostinelli, F. Chiariello, F. M. Maggi, A. Marrella, and F. Patrizi, "Process mining meets model learning: Discovering deterministic finite state automata from event logs for business process analysis," *Inf. Syst.*, vol. 114, p. 102180, 2023.

[7] A. Almusayli, T. Zia, and E.-H. Qazi, "Drone Forensics: An Innovative Approach to the Forensic Investigation of Drone Accidents Based on Digital Twin Technology," *Technologies*, vol. 12, no. 1, p. 11, 2024.

[8] J.-S. Kim, D.-G. Kim, and B.-N. Noh, "A fuzzy logic based expert system as a network forensics," in *2024 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)*, 2024, pp. 879–884 vol.2. doi: 10.1109/FUZZY.2004.1375521.

[9] V. Pooryousef, M. Cordeil, L. Besançon, R. Bassed, and T. Dwyer, "Collaborative Forensic Autopsy Documentation and Supervised Report Generation Using a Hybrid Mixed-Reality Environment and Generative AI," *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 11, pp. 7452–7462, 2024, doi: 10.1109/TVCG.2024.3456212.

[10] A. M. Bade, S. H. Othman, S. Z. M. Hashim, and S. H. M. Yusof, "Expert Validation of Online Social Networks Forensic Investigation Metamodel (OSNFIM)," in *2023 International Conference on Data Science and Its Applications (ICoDSA)*, 2023, pp. 500–505. doi: 10.1109/ICoDSA58501.2023.10277048.

[11] S. Benkerroum and K. Chougdali, "Enhancing Forensic Analysis Using a Machine Learning-based Approach," in *2023 6th International Conference on Advanced Communication Technologies and Networking (CommNet)*, 2023, pp. 1–6. doi: 10.1109/CommNet60167.2023.10365260.

[12] A. Sreekumar, R. V Jayaram, and H. N. A.G, "Weapons and Related Object Classification in Digital Forensic Using Machine Learning," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–5. doi: 10.1109/ICCCNT56998.2023.10307988.

[13] S. U. Qureshi *et al.*, "Systematic review of deep learning solutions for malware detection and forensic analysis in IoT," *J. King Saud Univ. Inf. Sci.*, p. 102164, 2024.

[14] M. M. Alshabibi, A. K. Bu dookhi, and M. M. Hafizur Rahman, "Forensic Investigation, Challenges, and Issues of Cloud Data: A Systematic Literature Review," *Computers*, vol. 13, no. 8, p. 213, 2024.

[15] D. L. Bhatt *et al.*, "A controlled trial of renal denervation for resistant hypertension," *N. Engl. J. Med.*, vol. 370, no. 15, pp. 1393–1401, 2014.

[16] Z. F. Hapsah and M. I. P. Nasution, "ANALISIS TINGKAT KEAMANAN DATA PERUSAHAAN YANG RENTAN TERHADAP SERANGAN CYBER DALAM SISTEM INFORMASI MANAJEMEN," *WANARGI J. Manaj. Dan Akunt.*, vol. 1, no. 2, pp. 338–343, 2024.

[17] A. P. Kehista *et al.*, "Analisis Keamanan Data Pribadi pada Pengguna E-Commerce: Ancaman, Risiko, Strategi Kemanan (Literature Review)," *J. Ilmu Manaj. Terap.*, vol. 4, no. 5, pp. 625–632, 2023.

[18] L. F. Sikos, "Packet analysis for network forensics: A comprehensive survey," *Forensic Sci. Int. Digit. Investig.*, vol. 32, p. 200892, 2020.

[19] A. AL-Hawamleh, "Cyber resilience framework: Strengthening defenses and enhancing continuity in business security," *Int. J. Comput. Digit. Syst.*, vol. 15, no. 1, pp. 1315–1331, 2024.

[20] G. Horsman, "Digital evidence strategies for digital forensic science examinations," *Sci. Justice*, vol. 63, no. 1, pp. 116–126, 2023.

[21] M. Zhang, "Forensic imaging: a powerful tool in modern forensic investigation," *Forensic Sci. Res.*, vol. 7, no. 3, pp. 385–392, 2022.

[22] I. H. Sarker, "Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects," *Ann. Data Sci.*, vol. 10, no. 6, pp. 1473–1498, 2023.
[23] C. S. Bojer and J. P. Meldgaard, "Kaggle forecasting competitions: An overlooked learning opportunity," *Int. J. Forecast.*, vol. 37, no. 2, pp. 587–603, 2021.
[24] P. Dhawas, A. Dhore, D. Bhagat, R. D. Pawar, A. Kukade, and K. Kalbande, "Big Data Preprocessing, Techniques, Integration, Transformation, Normalisation, Cleaning, Discretization, and Binning," in *Big Data Analytics Techniques for Market Intelligence*, IGI Global, 2024, pp. 159–182.
[25] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Variable selection for Naïve Bayes classification," *Comput. Oper. Res.*, vol. 135, p. 105456, 2021.
[26] N. Jeffrey, Q. Tan, and J. R. Villar, "A review of anomaly detection strategies to detect threats to cyber-physical systems," *Electronics*, vol. 12, no. 15, p. 3283, 2023.