Acceleration, Quantum, Information Technology and Algorithm Journal (AQILA)
Vol. 2, No. 2, December 2025, pp 85-89
ISSN: 3062-8555, DOI: 10.62123/aqila.v1i2.123

Cyber Attack Prediction Using Machine Learning: A Comparative Study of Bayesian Network and Support Vector Machine

Cut Try Utari[1*], Indri Sulistianingsih[1], Diva Rofsyahfitri[1], Nurul Rizkina Kalsum Batubara[1], Wizdanil Yumna Nawar[1]

[1]Politeknik Negeri Medan, Medan, Indonesia

*Corresponding Email: cuttry@polmed.ac.id

DOI : 10.6213/aqila.v1i2.123

Received : November 28, 2025
Revised : December 17, 2025
Accepted : December 20, 2025

Keywords:
Cyber Attack Prediction
Machine Learning
Bayesian Network
Support Vector Machine

ABSTRACT

Cybersecurity is becoming a critical issue with the increasing reliance on digital systems that are vulnerable to attacks. Proactive cyberattack prediction is one of the main approaches in early detection systems, where machine learning plays a strategic role. This research compares two popular machine learning algorithms, namely Bayesian Network and Support Vector Machine (SVM), to determine the most effective algorithm in predicting cyberattacks. This research uses two benchmark datasets, namely UNSW-NB15 and KDD99, as well as real attack data from Elazığ, Turkey. The analysis shows that the Bayesian Network implemented through the MCVAE_PBNN approach achieves up to 96% accuracy on the UNSW-NB15 dataset, with the advantage of detecting distributed and uncertain attacks. On the other hand, the SVM linear (SVML) algorithm showed a prediction accuracy of 95.02% in attack method classification, excelling in the case of data with clearly defined features. This study also analyzes the advantages and limitations of both algorithms, and provides implementation recommendations based on the needs of the detection system. The findings reinforce the urgency of developing adaptive predictive models in modern cybersecurity.

1. INTRODUCTION

The rapid growth of digital technology has brought serious consequences to information security, given the increasing frequency and complexity of cyberattacks. These threats target not only countries' critical infrastructure, but also business systems and individuals, causing huge economic and social losses [1]. To anticipate and proactively mitigate such attacks, machine learning-based approaches have become a widely developed solution in intrusion detection systems[2].

Two algorithms that are often used in this context are Bayesian Network and Support Vector Machine (SVM). Bayesian Network is a probabilistic graph model that allows modeling uncertainty between variables, as well as being able to uncover causal relationships in data. This approach is well suited for complex and uncertain environments, such as in distributed cyber-attacks[3]. In contrast, SVM is known as an effective classification algorithm in handling high-dimensional data, and has stable performance in separating clearly defined classes [4].

This study aims to compare the accuracy of Bayesian Network and SVM in predicting cyberattacks based on benchmark datasets and real-world data, analyze the strengths and weaknesses of each algorithm in the context of attack prediction, and provide practical recommendations for the optimal application of the algorithm in cyberattack detection systems.

By comparing the two algorithms based on standard evaluation metrics such as accuracy, sensitivity, and AUC, this research is expected to provide guidance for policy makers and cybersecurity practitioners in choosing the algorithm that best suits their system needs.

2. RELATED WORKS

The contemporary cybersecurity paradigm [5] has undergone a fundamental transformation with the emergence of machine learning technology as a key instrument in threat detection and prediction. This evolution is driven by the inherent limitations of conventional detection systems that rely on signature-based detection, which has proven ineffective in dealing with sophisticated attack vectors such as zero-day exploits and advanced persistent threats that evolve dynamically. Recent research by Siva et al. [6] developed a comprehensive cyberattack detection and prediction system that integrates multiple machine learning

algorithms including AdaBoost, Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Classifier, and Logistic Regression. The system enables automated data preprocessing and adaptive model selection based on dataset characteristics, providing significant flexibility for various threat landscape scenarios. The contribution of this research lies in the holistic approach that combines data preprocessing, model selection, and threat forecasting in one integrated framework.

Verma and Thakur [7] conducted an in-depth investigation into the effectiveness of four distinct classifiers in cyberattack prediction using the UNSW-NB15 dataset. They used a 9-fold cross-validation methodology to ensure the robustness of the evaluation results, with Random Forest showing superior performance with 95.167% accuracy, 96.252% true positive rate, and 6.749% false positive rate. This finding indicates that ensemble methods such as Random Forest have a superior ability to identify complex attack patterns compared to individual classifiers. Swaminathan et al. [8] explored the application of machine learning algorithms for cyberattack prediction with a focus on real-time analysis of data from forensic units. Their research implemented three main methodologies namely Logistic Regression, Random Forest, and K-Nearest Neighbor for cybercrime investigation and assessment[9] of the impact of various attributes in the identification of attack methods and perpetrator profiling[10]. This approach provides a unique perspective by integrating forensic analysis with predictive modeling for enhanced proactive threat detection [11].

In the context of Bayesian Network applications, research shows superior effectiveness in handling uncertainty and incomplete information scenarios often found in cybersecurity environments. The probabilistic framework of Bayesian Networks enables quantification of uncertainty in predictions, providing valuable insights for security analysts in making informed decisions about potential threats. The ability to incorporate prior knowledge and update beliefs based on new evidence makes Bayesian Networks particularly suitable for adaptive threat detection systems.

The implementation of Support Vector Machine in the cybersecurity domain has shown remarkable performance in handling high-dimensional feature spaces that are characteristic of network traffic data. Extensive research has proven that SVMs exhibit strong performance across a wide range of attack types, especially when combined with appropriate feature selection techniques and kernel functions that are able to capture non-linear relationships in complex network data patterns. The mathematical foundation of SVM based on optimal hyperplane discovery with maximum margin separation provides a theoretical guarantee for generalization performance [12].

3. RESEARCH METHODS

3.1 Dataset

This research uses a multi-dataset approach to ensure the validity and generalizability of the results. The benchmark datasets used are UNSW-NB15 and KDD99[13], both of which contain various types of cyberattacks, including DoS, phishing, ransomware, and malware, as well as features such as protocols, services, and flags relevant in attack classification [14]. The real-world dataset uses the Elazığ Cyber Crime Dataset collected from the Turkish Police for five years, including demographic features of perpetrators and victims, attack methods, and losses due to attacks. This dataset is used in attack method prediction models and offender identification [15].

Comparative studies show that the use of UNSW-NB15 dataset provides a more comprehensive representation for the evaluation of machine learning algorithms in cybersecurity. Research by Verma and Thakur [7] used 9-fold cross-validation on the UNSW-NB15 dataset[16] to evaluate the performance of various classifiers, showing that proper dataset selection and validation methodology can significantly affect the results of algorithm evaluation.

3.2 Algorithm

The Bayesian Network model is implemented with the Multi-Connect Variational Auto-Encoder with Probabilistic Bayesian Networks (MCVAE_PBNN) approach. This architecture utilizes probabilistic relationships between variables to efficiently detect attack patterns, including under hidden or distributed attack conditions [14]. This approach is in line with recent trends in cybersecurity that integrate deep learning with probabilistic modeling for enhanced threat detection capabilities.

Bayesian networks (BNs) are fundamental analytical tools in probability modeling, with their functional core lying in Bayes' Theorem. This theorem allows updating the probability of a hypothesis (H) based on new evidence or information (e) observed [17]. The generic formulation of Bayes' Theorem is expressed as follows:

$$P(H|e) = \frac{P(e|H) \cdot P(H)}{P(e)} \quad (1)$$

Where H represents a hypothesis, and e is evidence associated with an event. The probability of hypothesis H in the presence of evidence e (P(H|e)) is calculated by multiplying the initial hypothesis probability (P(H)) by the posterior probability P(H|e) [17].

In a BN structure, represented as a Directed Acyclic Graph (DAG), there are a set of variables (nodes) and interdependencies (edges) that show the conditional relationships among the variables. For example, for a set of variables S = {M1, M2, M3, M4, M5} with edges depicting conditional interdependencies (as illustrated in Figure 1), the joint probability

distribution decomposition of these variables can be expressed as: P(M1, M2, M3, M4, M5) = P(M1) P(M2|M1) P(M3|M1) P(M4|M2, M3) P(M5|M4)
In general, this decomposition can be formulated concisely as:

$$\prod_{i=1}^{n} P\left(M_i \mid \text{Parents}(M_i) \right) \quad (2)$$

Where P(Mi|Parents(Mi)) denotes the conditional probability of node Mi given the values of its parent nodes [17]. Support Vector Machine is used in two main variants: SVM linear (SVML) and SVM kernel (SVMK). SVML is proven to be effective in the classification of attack methods based on input features such as victim's age, education, and loss type. The algorithm is implemented using the scikit-learn library on the Python platform [15]. Recent research shows that the combination of SVM with other algorithms such as Random Forest and K-Nearest Neighbors can provide superior performance in analyzing real-time forensic data [8].

3.3 Evaluation Metrics

To assess the performance of each algorithm, this research uses several comprehensive evaluation metrics. Accuracy is used to measure the percentage of correct predictions compared to total predictions. False Alarm Rate (FAR) [18] measures the proportion of false positive predictions, while sensitivity (Recall) evaluates the model's ability to correctly identify attacks. Specificity measures the model's ability to recognize normal traffic, and Area Under Curve (AUC)[19] provides a measure of overall classification performance. Precision and F1-score are used specifically in SVM to evaluate the accuracy of multi-class classification. All models were evaluated using a cross-validation and testing approach on data that had been split into 80% training and 20% testing, following established best practices in cybersecurity machine learning research [7].

4. DISCUSSION AND RESULT

4.1 Experiment Results on Benchmark Datasets

The implementation of Bayesian Network through MCVAE_PBNN model produces excellent performance in detecting attacks on UNSW-NB15 and KDD99 datasets. Based on the results obtained by Mouti et al.[14], this model achieves in Table 1:

Table 1. Benchmark Datasets

Metrics	UNSW-NB15	KDD9
Accuracy	96%	95%
False Alarm Rate (FAR)	71%	68%
Sensitivity	92%	92%
Specificity	82%	84%
AUC	75%	78%

The main advantage of the Bayesian Network is its ability to model uncertainty and relationships between variables probabilistically. This model has proven effective for detecting complex and stealthy attacks, such as distributed attacks that are difficult to track with deterministic approaches.

4.2 Experimental Results on Real World Dataset (Elazığ, Turkey)

In a study by Bilen and Özer [15], the SVM Linear (SVML) algorithm showed superior performance in predicting attack methods based on victim and incident features. The accuracy results of some algorithms are as follows in Table 2.

Table 2. Results on Real World Dataset

Algorithm	Akurasi (%)	Precision (%)	Recall (%)	F1-score (%)
SVML	**95.02**	95.43	95.03	95.16
RF	94.48	94.48	94.48	94.48
LR	93.92	94.41	93.92	94.10
SVMK	92.82	92.99	92.82	92.88

NB	81.77	81.79	81.77	81.23

SVML is the best algorithm in predicting attack methods such as phishing, social engineering, and malware based on victim demographic features. This result indicates that SVM is suitable for explicit feature-based classification and high linearity.

4.3 Comparative Analysis

The comparison results show that both Bayesian Network and SVM have their respective advantages in the context of cyberattack prediction. Bayesian Network excels in handling incomplete or noise data, modeling dependencies between features suitable for complex attacks, and showing solid performance in distributed environments [14]. Meanwhile, SVM (Linear) excels in high prediction accuracy on structured and clean datasets, computational efficiency when the number of features is high and linearity is clear, and is easy to implement and tune for large-scale classification [15].

Recent research has shown that cyberattack detection and prediction systems that integrate multiple algorithms can provide enhanced capabilities [6]. This approach enables adaptation to different types of attacks and different operational environments, in line with the trend of developing adaptive cybersecurity systems.

Several studies have shown that a hybrid approach or combination of machine learning techniques can improve the effectiveness of cyberattack detection, especially if accompanied by appropriate feature selection [20]. The integration of probabilistic reasoning from Bayesian Networks with classification efficiency from SVM has the potential to produce a more robust and adaptive system.

4.4 Practical Implications

These findings provide important implications for the development of cyberattack early detection systems in various organizations. Algorithm selection should consider the type of data (structured vs. incomplete), real-time requirements, computing resources, and the complexity of relationships between features. With the right approach, machine learning-based cybersecurity systems can significantly improve predictive and responsive capacity to evolving digital threats.

Forensic-based analysis research shows that the combination of demographic profiling with behavioral analysis can significantly enhance threat prediction capabilities [8]. This provides direction for the development of systems that not only focus on technical indicators, but also incorporating human factors in threat assessment.

5. CONCLUSION

This study examined and compared two machine learning algorithms Bayesian Network and Support Vector Machine (SVM) to assess their effectiveness in predicting cyberattacks using both benchmark datasets and real-world data. The results indicate that both algorithms perform well, but each excels in different contexts. Bayesian Network, particularly when implemented with the MCVAE_PBNN approach, demonstrated strong capabilities in handling complex, incomplete, and uncertain data. This makes it particularly suitable for detecting hidden or distributed cyberattacks, which are common in modern network environments. The model achieved an accuracy of up to 96% on the UNSW-NB15 dataset and effectively mapped probabilistic relationships between variables—an essential feature in security risk analysis.

On the other hand, Linear SVM showed high performance on datasets with clear structures and explicit features, as demonstrated in the Elazığ dataset. With a prediction accuracy of 95.02%, SVM proved highly reliable in classifying attack methods based on variables such as victim age, education, or the type of incident. It also stood out in terms of computational efficiency and scalability for large-scale applications. From these findings, it can be concluded that no single algorithm is universally superior. The choice of algorithm should be based on the characteristics of the data and the specific needs of the security system being developed. In some cases, a hybrid approach that combines the strengths of both algorithms may provide the most effective solution. Leveraging the Bayesian Network's ability to model uncertainty and SVM's classification strength for structured data can result in a more adaptive and responsive detection system. Overall, this study highlights the importance of a flexible and context-aware approach in building predictive cyberattack systems. Amid evolving digital threats, the appropriate use of machine learning technologies can be a key factor in developing smarter and more resilient security defenses.

REFRENCES

[1] D. Dasgupta, Z. Akhtar, and S. Sen, "Machine learning in cybersecurity: a comprehensive survey," *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, vol. 19, no. 1, pp. 57–106, Jan. 2022, doi: 10.1177/1548512920951275.

[2] G. Kocher and G. Kumar, "Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges," *Soft comput*, vol. 25, no. 15, pp. 9731–9763, 2021.

[3] S. Mouti, S. K. Shukla, S. A. Althubiti, M. A. Ahmed, F. Alenezi, and M. Arumugam, "Cyber Security Risk management with attack detection frameworks using multi connect variational auto-encoder with probabilistic Bayesian networks," *Computers and Electrical Engineering*, vol. 103, p. 108308, 2022.
[4] A. Bilen and A. B. Özer, "Cyber-attack method and perpetrator prediction using machine learning algorithms," *PeerJ Comput Sci*, vol. 7, p. e475, Apr. 2021, doi: 10.7717/peerj-cs.475.
[5] E. Pleshakova, A. Osipov, S. Gataullin, T. Gataullin, and A. Vasilakos, "Next gen cybersecurity paradigm towards artificial general intelligence: Russian market challenges and future global technological trends," *Journal of Computer Virology and Hacking Techniques*, vol. 20, no. 3, pp. 429–440, 2024.
[6] O. V Siva, K. Neeraja, D. Kalyan, and K. S. Naga, "Cyber Attack Detection and Prediction System," in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2024, pp. 1–8. doi: 10.1109/ACCAI61061.2024.10602219.
[7] R. Verma and B. Thakur, "Machine Learning Techniques for the Prediction of Cyber-Attacks," in *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2023, pp. 978–985. doi: 10.1109/ICCCIS60361.2023.10425542.
[8] A. Swaminathan, B. Ramakrishnan, K. M, and S. R, "Prediction of Cyber-attacks and Criminality Using Machine Learning Algorithms," in *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2022, pp. 547–552. doi: 10.1109/3ICT56508.2022.9990652.
[9] K. Veena, K. Meena, Y. Teekaraman, R. Kuppusamy, and A. Radhakrishnan, "C SVM classification and KNN techniques for cyber crime detection," *Wirel Commun Mob Comput*, vol. 2022, no. 1, p. 3640017, 2022.
[10] I. Kotenko, E. Fedorchenko, E. Novikova, and A. Jha, "Cyber attacker profiling for risk analysis based on machine learning," *Sensors*, vol. 23, no. 4, p. 2028, 2023.
[11] A. R. P. Reddy, "The role of artificial intelligence in proactive cyber threat detection in cloud environments," *NeuroQuantology*, vol. 19, no. 12, pp. 764–773, 2021.
[12] A. R. Lubis, Y. Y. Lase, D. A. R, and D. Witarsyah, "Optimization of SVM Classification Accuracy with Bayesian Optimization Utilizing Data Augmentation," in *2023 6th International Conference of Computer and Informatics Engineering (IC2IE)*, IEEE, Sep. 2023, pp. 169–174. doi: 10.1109/IC2IE60547.2023.10331580.
[13] Z. Ruan, Y. Miao, L. Pan, N. Patterson, and J. Zhang, "Visualization of big data security: a case study on the KDD99 cup data set," *Digital Communications and Networks*, vol. 3, no. 4, pp. 250–259, Nov. 2017, doi: 10.1016/j.dcan.2017.07.004.
[14] S. Mouti, S. K. Shukla, S. A. Althubiti, M. A. Ahmed, F. Alenezi, and M. Arumugam, "Cyber Security Risk management with attack detection frameworks using multi connect variational auto-encoder with probabilistic Bayesian networks," *Computers and Electrical Engineering*, vol. 103, Oct. 2022, doi: 10.1016/j.compeleceng.2022.108308.
[15] A. Bilen and A. B. Özer, "Cyber-attack method and perpetrator prediction using machine learning algorithms," *PeerJ Comput Sci*, vol. 7, pp. 1–21, 2021, doi: 10.7717/PEERJ-CS.475.
[16] A. D. Vibhute, M. Khan, C. H. Patil, S. V. Gaikwad, A. V. Mane, and K. K. Patel, "Network anomaly detection and performance evaluation of Convolutional Neural Networks on UNSW-NB15 dataset," *Procedia Comput Sci*, vol. 235, pp. 2227–2236, 2024, doi: 10.1016/j.procs.2024.04.211.
[17] N. U. I. Hossain, M. Nagahi, R. Jaradat, C. Shah, R. Buchanan, and M. Hamilton, "Modeling and assessing cyber resilience of smart grid using Bayesian network-based approach: A system of systems problem," *J Comput Des Eng*, vol. 7, no. 3, pp. 352–366, Jun. 2020, doi: 10.1093/jcde/qwaa029.
[18] H. J. Kang, K. L. Aw, and D. Lo, "Detecting false alarms from automatic static analysis tools: How far are we?," in *Proceedings of the 44th International Conference on Software Engineering*, 2022, pp. 698–709.
[19] N. R. Datta *et al.*, "Quantification of thermal dose in moderate clinical hyperthermia with radiotherapy: a relook using temperature–time area under the curve (AUC)," *International journal of hyperthermia*, vol. 38, no. 1, pp. 296–307, 2021.
[20] R. Ben Said, Z. Sabir, and I. Askerzade, "CNN-BiLSTM: A Hybrid Deep Learning Approach for Network Intrusion Detection System in Software Defined Networking with Hybrid Feature Selection.," *IEEE Access*, vol. PP, p. 1, Jan. 2023, doi: 10.1109/ACCESS.2023.3340142.