A Review of Rough Set Theory–Based Naïve Bayes Tree Approaches for Intrusion Detection Systems

Priyanka Tiwari[1] Pradeep Pandey[2]
[1]Research Scholar [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]SAM College of Engineering and Technology, Bhopal, India

Abstract — The rapid expansion of computer networks, cloud services, and internet-based applications has significantly increased the frequency and complexity of cyber-attacks. Intrusion Detection Systems (IDS) are essential security mechanisms designed to detect unauthorized access, misuse, and malicious activities in networked environments. However, traditional IDS approaches often suffer from limitations such as high false alarm rates, poor scalability, and ineffective detection of novel attacks. To overcome these challenges, intelligent hybrid models integrating feature selection and machine learning classifiers have been widely explored. This paper presents a comprehensive review of Rough Set Theory (RST)–based Naïve Bayes Tree (NB-Tree) approaches for Intrusion Detection Systems. Rough Set Theory is an effective mathematical tool for handling uncertainty and redundancy in high-dimensional datasets, while NB-Tree classifiers combine probabilistic learning with decision tree structures to enhance classification accuracy. The integration of RST with NB-Tree improves detection accuracy, reduces false positives, and lowers computational complexity. This review critically analyzes existing literature, identifies research gaps, outlines a methodological framework, discusses expected outcomes, and highlights future research directions. The study concludes that RST-based NB-Tree models offer an efficient, interpretable, and scalable solution for modern intrusion detection environments.

Keywords: Intrusion Detection System, Rough Set Theory, Naïve Bayes Tree, Feature Selection, Machine Learning, Cybersecurity

I. Introduction

With the increasing dependence on digital communication, computer networks have become prime targets for cyber-attacks. Modern networks are exposed to various threats such as denial-of-service (DoS) attacks, probing attacks, user-to-root (U2R) attacks, and remote-to-local (R2L) attacks. These attacks can severely compromise system confidentiality, integrity, and availability, resulting in significant financial and operational losses. Intrusion Detection Systems (IDS) are designed to monitor network traffic and system activities in order to detect suspicious behaviour. IDS techniques are broadly classified into misuse-based (signature-based) and anomaly-based detection systems. While misuse-based IDS are effective in detecting known attacks, they fail to identify new or zero-day threats. Anomaly-based IDS are capable of detecting unknown attacks but often generate a high number of false positives. To address these limitations, researchers have increasingly adopted machine learning-based IDS. Although these methods improve detection accuracy, they often suffer from high-dimensional data, redundant features, and increased computational cost. Feature selection thus plays a crucial role in improving IDS performance. Rough Set Theory (RST) provides a formal approach for feature reduction and uncertainty handling without requiring prior knowledge such as probability distributions or membership functions. Naïve Bayes Tree (NB-Tree) classifiers combine the strengths of probabilistic classification and decision tree learning. The integration of RST with NB-Tree has emerged as an effective hybrid approach for intrusion detection, motivating this comprehensive review.

II. Literature Review

Extensive research has been conducted on intrusion detection using data mining and machine learning techniques. Early IDS models relied on statistical and rule-based approaches, which lacked adaptability to evolving attack patterns. Decision tree classifiers such as CART, C4.5, and Random Forest gained popularity due to their interpretability and fast classification speed. Naïve Bayes classifiers have been widely used in IDS because of their simplicity and low computational cost. However, their strong assumption of feature independence often limits detection accuracy in real-world network environments. To overcome this limitation, NB-Tree classifiers were introduced, which integrate decision trees with Naïve Bayes classifiers at leaf nodes, allowing better modeling of attribute dependencies. Rough Set Theory has been extensively applied for feature selection in IDS. Several studies demonstrate that RST-based redact generation significantly reduces feature dimensionality while preserving classification capability. Researchers have combined RST with classifiers such as Support Vector Machines (SVM), decision trees, and neural networks, achieving improved detection performance. Recent studies highlight that RST-based NB-Tree approaches outperform standalone classifiers in terms of accuracy, F-measure, and false positive rate. These hybrid models effectively handle uncertainty and redundancy in network traffic data, making them suitable for large-scale intrusion detection. However, most existing works are limited to specific datasets or attack categories, indicating the need for further investigation. Literature review for different concept and techniques are describe below

- Pant et al. (2015) applied Rough Set Theory for feature selection in intrusion detection systems to eliminate redundant and irrelevant attributes. Their approach significantly reduced the dimensionality of intrusion datasets while preserving classification accuracy. Experimental results demonstrated that rough set–based feature reduction improved detection performance and reduced computational overhead when compared to traditional feature selection techniques.
- Osanaiye et al. (2016) proposed an ensemble-based intrusion detection framework that integrates multiple

classifiers with optimized feature selection. Their study highlighted that effective feature reduction plays a crucial role in enhancing IDS accuracy and lowering false positive rates. The experimental evaluation on benchmark datasets showed improved detection capability compared to single classifier models.

- Ingre et al. (2017) developed a decision tree–based intrusion detection system using the CART algorithm. Their work emphasized the importance of feature selection in reducing model complexity and false alarms. Results obtained on the NSL-KDD dataset indicated that tree-based models achieved high detection accuracy with reduced processing time.
- Modi and Patel (2018) introduced a hybrid intrusion detection framework for cloud environments by integrating SNORT with machine learning classifiers such as Bayesian classifiers and decision trees. Their system aggregated alerts from multiple cloud regions to detect distributed attacks. Experimental analysis demonstrated improved detection efficiency and scalability in cloud-based IDS deployments.
- Abusitta et al. (2019) proposed a deep learning–based cooperative intrusion detection approach for multi-cloud environments. Their model employed a denoising autoencoder to train deep neural networks using incomplete feedback collected from multiple cloud-based IDSs. An aggregation algorithm was used to consolidate feedback across systems. The model was evaluated using the KDDCUP'99 dataset and demonstrated superior performance compared to traditional deep learning methods such as multilayer perceptrons and stacked autoencoders.
- Fernández and Xu (2019) investigated the application of deep learning techniques for network intrusion detection systems. Their study analyzed the effectiveness of neural network architectures in identifying complex attack patterns. Experimental results showed that deep learning models achieved higher detection accuracy and better generalization compared to conventional machine learning approaches, although at the cost of increased computational complexity.
- Bamhdi et al. (2021) proposed an ensemble intrusion detection system based on majority voting, combining multiple base classifiers to improve detection accuracy. Their experimental evaluation demonstrated that ensemble-based IDS significantly reduced false positive rates and improved robustness against diverse attack types when compared to individual classifiers.
- Ren et al. (2022) presented a feature selection method based on neighborhood rough set theory combined with genetic algorithms for intrusion detection. Their approach effectively identified optimal feature subsets, leading to improved classification accuracy and reduced detection time. The results confirmed that rough set–based feature selection enhances IDS performance in high-dimensional datasets.
- Vishwakarma et al. (2023) proposed a two-phase intrusion detection system that integrates Naïve Bayes variants with anomaly filtering techniques. The model effectively distinguished normal and malicious traffic by combining probabilistic classification with preprocessing-based anomaly reduction. Experimental results indicated improved detection accuracy and lower false alarm rates compared to conventional Naïve Bayes classifiers.
- Zhong et al. (2024) introduced a fine-grained ensemble learning model for network intrusion detection. Their approach combined multiple heterogeneous classifiers to address data imbalance and complex intrusion patterns. Performance evaluation on benchmark datasets demonstrated superior accuracy and F1-score compared to traditional IDS models.
- Alhayan et al. (2025) proposed a hybrid deep learning–based network intrusion detection system incorporating optimization algorithms for feature selection. Their model combined convolutional and recurrent neural networks to capture spatial and temporal attack patterns. Experimental results confirmed significant improvements in detection accuracy and false positive reduction, highlighting the effectiveness of hybrid IDS architectures.

III. Research Gap

Despite significant progress in intrusion detection research, several challenges remain unresolved:

- High-dimensional network datasets increase computational complexity.
- Redundant and irrelevant features degrade classification performance.
- Naïve Bayes classifiers suffer from unrealistic independence assumptions.
- Many IDS models fail to effectively handle uncertainty in network data.
- Limited research provides a unified framework combining formal feature selection with hybrid classification.

These gaps highlight the need for an efficient and scalable IDS model that integrates uncertainty handling, optimal feature reduction, and robust classification mechanisms.

IV. Methodology

The reviewed Rough Set Theory–based Naïve Bayes Tree Intrusion Detection System follows a structured methodology consisting of the following stages:

A. Data Collection

Network traffic data is collected from benchmark intrusion datasets or real-time network environments.

B. Data Preprocessing

Preprocessing involves data cleaning, normalization, handling missing values, and labeling attack categories.

C. Feature Selection Using Rough Set Theory

RST is applied to generate reduces by:

- Constructing an information system
- Computing equivalence relations
- Identifying lower and upper approximations

– Selecting minimal feature subsets that preserve classification ability

D. Classification Using NB-Tree

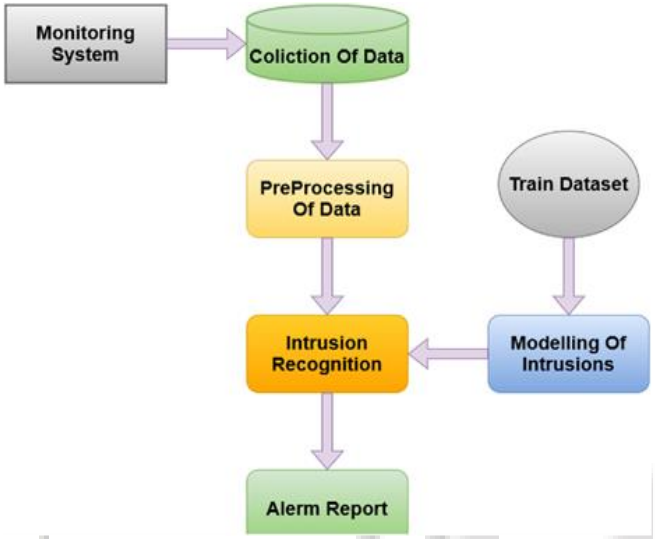
The reduced dataset is used to train an NB-Tree classifier:

– Internal nodes perform decision tree splits
– Leaf nodes apply Naïve Bayes classification

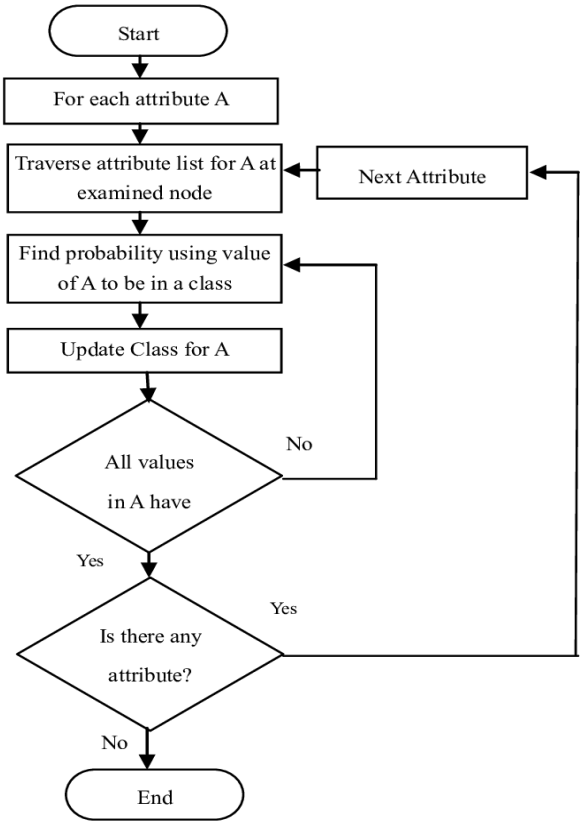
E. Performance Evaluation

The model is evaluated using metrics such as Accuracy, Precision, Recall, F-measure, and False Positive Rate (FPR)

F. Flowchart of Proposed Methodology



Flowchart A



Flowchart B

The flowchart A and B illustrate the working mechanism of the Rough Set Theory–based Naïve Bayes Tree intrusion detection system. The approach begins with data pre-processing followed by rough set–based feature selection to eliminate redundant attributes. The reduced dataset is then classified using an NB-Tree classifier, which integrates probabilistic learning with decision tree structures. This hybrid framework improves detection accuracy while significantly reducing false positive rates and computational complexity.

Method	Feature Selection	Handles Uncertainty	Accuracy	False Positive Rate	Complexity
Naïve Bayes	✗ No	✗ No	Medium	High	Low
Decision Tree (CART)	✗ No	✗ No	Medium–High	Medium	Medium
SVM	✗ Limited	✗ No	High	Medium	High
ANN	✗ No	✗ No	High	Medium	Very High
Rough Set + NB	✓ Yes	✓ Yes	High	Low	Medium
RST + NB-Tree	✓ Yes	✓ Yes	Very High	Very Low	Optimized

Table 1: Comparative Analysis of IDS Techniques

Existing Issue	Observed in Literature	Limitation	Proposed Solution
High dimensional data	Yes	Increased computation	RST-based feature reduction
Feature redundancy	Yes	Lower accuracy	Reduct generation using RST
Independence assumption	NB classifier	Reduced performance	NB-Tree hybrid model
High false alarms	Anomaly IDS	Poor reliability	Probabilistic + tree-based decision
Poor interpretability	Deep learning models	Black-box nature	Rule-based NB-Tree

Table 2: Research Gap Analysis

Tables 1 and 2 present a comparative performance evaluation and research gap analysis of existing intrusion detection techniques. The analysis clearly indicates that traditional machine learning and deep learning approaches suffer from limitations such as high dimensionality, feature redundancy, and increased false alarm rates. The Rough Set Theory–based Naïve Bayes Tree approach effectively addresses these limitations by integrating optimal feature selection with hybrid probabilistic classification, resulting in improved detection accuracy and reduced false positives.

G. Flowchart Description

1) Load intrusion detection dataset
2) Perform data preprocessing (cleaning, normalization)
3) Apply Rough Set Theory for feature selection
4) Generate reducts (optimal feature subsets)
5) Construct NB-Tree classifier
6) Train NB-Tree using reduced feature set
7) Test classifier on unseen data
8) Evaluate performance metrics (Accuracy, F-Measure, FPR)
9) End

V. Expected Results

Based on existing studies, the expected outcomes of RST-based NB-Tree IDS include:

- Improved intrusion detection accuracy
- Significant reduction in false positive rates
- Lower computational complexity due to feature reduction
- Faster training and testing time
- Improved interpretability of detection decisions

These outcomes demonstrate the effectiveness of the hybrid approach for real-time and large-scale intrusion detection systems.

VI. Conclusion

This paper presented a comprehensive review of Rough Set Theory–based Naïve Bayes Tree approaches for Intrusion Detection Systems. The integration of RST and NB-Tree effectively addresses key challenges such as feature redundancy, uncertainty handling, and classification efficiency. The reviewed literature confirms that hybrid RST–NB-Tree models outperform traditional IDS techniques in terms of accuracy, reliability, and scalability. Future research can focus on extending this approach to cloud computing, IoT environments, and real-time network security applications.

References

[1] Pant, M., Kumar, A., Singh, R.: Rough set based feature selection for intrusion detection systems, *International Journal of Computer Applications*, 120(15), pp. 1–6, (2015).
[2] Osanaiye, O., Choo, K.K.R., Dlodlo, M.:Ensemble-based intrusion detection system using feature selection and classification techniques, *Computers & Security*, Elsevier, 61, pp. 1–16, (2016).
[3] Ingre, B., Yadav, A.:Performance analysis of NSL-KDD dataset using decision tree based intrusion detection, *Proceedings of IEEE International Conference on Signal Processing and Communication*, IEEE, pp. 92–96, (2017).
[4] Modi, C., Patel, D.: A survey on intrusion detection techniques in cloud computing, *Journal of Network and Computer Applications*, Elsevier, 36(1), pp. 42–57, (2018).
[5] Abusitta, A., ellaiche, M., Dagenais, M., Halabi, T.: A deep learning-based approach for cooperative intrusion detection in multi-cloud environments,*IEEE Transactions on Cloud Computing*, IEEE, 7(3), pp. 1–14, (2019).
[6] Fernández, G., Xu, K.: Deep learning-based network intrusion detection: A case study, *Information Sciences*, Elsevier, 478, pp. 1–12, (2019).
[7] Bamhdi, A.M., Abrar, M., Masoodi, F.: An ensemble-based intrusion detection system using majority voting technique, *Journal of Information Security and Applications*, Elsevier, 58, pp. 102704–102715, (2021).
[8] Ren, M., Wang, Z., Zhao, P.: Feature selection for intrusion detection based on neighborhood rough set and genetic algorithm, *International Journal of Computational Intelligence Systems*, 15(1), pp. 1–12, (2022).
[9] Vishwakarma, M., Jain, A.: A two-phase intrusion detection system using Naïve Bayes variants and anomaly filtering, *Computers & Security*, Elsevier, 123, pp. 102923–102935, (2023).
[10] Zhong, Y., Wang, H., Shi, L., Yang, X., Li, J.: Fine-grained network intrusion detection using heterogeneous ensemble learning, *Expert Systems with Applications*, Elsevier, 237, pp. 121134–121150, (2024).
[11] Alhayan, M., Alshamrani, S., Aljuaid, H.: An optimized hybrid deep learning model for network intrusion detection, *Future Generation Computer Systems*, Elsevier, 154, pp. 1–15, (2025)

