International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

THREATS AND POTENTIAL RISK DETECTION OF NETWORK INBOUND DATA USING MACHINE LEARNING

Auday Qusay Sabri[1], Halina Binti Mohamed Dahlan[2]
[1] Faculty of Computing Universiti Teknologi Malaysia
Johor, Malaysia
qusay@graduate.utm.my
[2] Faculty of Management
Universiti Teknologi Malaysia
Johor, Malaysia
halina@utm.my

Abstract

In this research, an attempt is made to detect threads and potential risks caused by incoming data through the network. Our research inspects three different types of risks the environmental, the operation risk, and the technical risk. Two Machine Learning methods have been used for the sake of research comparison and the rigidity of the results. Naive-Bayes and K-Nearest Neighbor algorithms have been applied to a structured data set with 15 input features representing the incoming risk data set to the network and a target prediction column of three different risk categories, namely environmental, operation, and technical risk. Results from Naive-Bays obtained an accuracy of 84% in risk detection, while K-Nearest Neighbor with 5 neighbors produced a 75% accuracy in risk detection.

Keywords— Naive-Bays, K-Nearest Neighbor (K-NN), Minkowski distance, Manhattan distance, Euclidean distance**.**

I. INTRODUCTION

Due to the rapid advancement of technology and the exponential growth of data collection, the field of network security has undergone significant transformation. The proliferation of digital devices and the expansion of the Internet of Things (IoT) expose networks to an increasing number of security risks. From more commonplace intrusions like malware and phishing to more sophisticated ones like Advanced Persistent Threats (APTs) and zero-day vulnerabilities, these threats range in complexity. Because of this, the need for robust and intelligent threat detection systems is now higher than ever [1].

This study will discuss the ability of machine learning algorithms to identify threads and possible threats from incoming network data is demonstrated. Also, this study will compare Naive-Bayes and K-Nearest Neighbor to reveal the advantages and disadvantages of each algorithm and shed light on how each might be used in various risk situations, Our goal is to improve the accuracy and dependability of network risk identification by utilizing the Naive-Bayes and K-Nearest Neighbor algorithms.

Currently, there is an increase in the number of terrorist attacks committed by organized terrorist communities with a network and a disorganized organization, as well as by lone terrorists influenced by propaganda and extremist ideology. The Internet that is, web resources, social networks, and email—is the primary tool for information sharing, hiring, and promoting such structures. This makes it necessary to

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

detect and identify communication topics and connections, as well as to keep an eye on user behavior and forecast potential threats originating from individuals, groups, and network communities that produce and disseminate extremist and terrorist content on the Internet [1].

One of the most challenging tasks for security administrators is identifying insider threats, which makes it challenging to recognize these internal risks. To identify the most accurate classifier to predict these insider risks, this study used a range of supervised machine learning classifiers with particular criteria [2].

Numerous cybersecurity applications, such as anomaly detection, intrusion detection, and vulnerability assessment, have extensively used machine learning methods. Popular algorithms Naive-Bayes and K-Nearest Neighbor have demonstrated potential in these areas. Based on Bayes' theorem, the probabilistic classifier Naive-Bayes is renowned for its ease of use and effectiveness when processing big datasets. It has been useful in network intrusion detection, virus detection, and spam filtering [3].

Machine learning techniques have been widely applied in a variety of cybersecurity applications, including anomaly detection, intrusion detection, and vulnerability assessment. Naive Bayes and K-Nearest Neighbor, two well-known algorithms, have shown promise in these domains. The probabilistic classifier Naive-Bayes, which is based on Bayes' theorem, is well known for being simple to use and efficient in handling large datasets. It has been helpful in spam filtering, virus detection, and network intrusion detection [4].

ML can create diverse models with different algorithms, and there are significant differences in how these models are worked with as well. When there is a large amount of labeled data, the network operator can use supervised learning to train a predictor based on the existing dataset; if there is a limited amount of labeled data, they can use a semi-supervised learning model. The results vary even when using the same model to identify the same kind of attack, based on the parameters you want machine learning to take into account [4, 5].

The use of machine learning techniques to leverage Internet data to address the challenge of combating extremism and terrorism, This problem entails monitoring and modeling information flows in these communities, identifying the structure of user groups and online communities that disseminate this information, retrieving electronic messages, documents, and web resources that may contain information of a terrorist or extremist nature, assessing threats and predicting risks based on monitoring results [6].

K-nearest neighbor and Naïve Bayes are the machine learning methods used in this work. These algorithms' performance is documented along with a comparative analysis. The project is carried out in Python, and the algorithms' effectiveness is evaluated using accuracy, sensitivity, specificity, and precision. Based on these observations, a model based on logistic regression was found to be more effective in predicting fraudulent activity than other models derived from Naïve Bayes and K-nearest neighbor.

Additional arguments for choosing these algorithms such as, the bulk of tree structure machine learning models use ensemble learning, which often leads to better results than single models such as KNN. Also, the network data that has been suggested falls under the area of non-linear, high dimensional data that they can manage, there is an advantage for feature selection in that the feature importance calculations are done during the building of those models [3].

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Despite much research, safeguarding networks from malevolent intrusions remains a formidable task. The increasing versatility of network assaults can be attributed to the rapid growth of linked devices and the expansion of networks in new technologies. In contrast to conventional detection techniques, machine learning offers a new and adaptable way to find network intrusions, working with any kind of network architecture [7].

Subsequent research endeavors will center on augmenting the dataset with a wider range of representative traits and investigating the potential utilization of sophisticated machine learning techniques, such as deep learning and reinforcement learning, to detect network danger. Enhancing the resilience and scalability of machine learning (ML)-based risk detection systems will need the incorporation of real-time data streams and the creation of adaptive learning models [8].

There is great potential for improving risk identification and mitigation through the use of machine learning in network security. Through the utilization of Naive-Bayes and K-Nearest Neighbor algorithms' advantages, this research adds to the current endeavors to create more dependable and efficient methods for detecting network risks [9, 10].

II. Related Works

In recent years, there has been a significant increase in interest in the use of machine learning for threat and risk identification in network systems. This section examines important contributions made by a range of field researchers, highlighting the efficacy and limitations of various machine learning-techniques. Gradient boosting decision trees (GBDT) are a popular method for detecting cyber security threats using network event data. GBDT is a potent machine learning method that enhances accuracy by combining the predictions of several decision trees. Researchers' study examines how GBDT can be used to identify security vulnerabilities in network logs and shows how, by learning from past data, GBDT can efficiently identify anomalies and possible dangers. This approach has a lot of potential because it can handle big datasets and detect threats with a high degree of accuracy [11].

A different study examines how different machine learning algorithms perform in intrusion detection systems (IDS). The efficacy of several methods, such as support vector machines (SVM), random forests, and neural networks, in detecting intrusions is compared by the researchers. The results show that while every algorithm has advantages, neural networks frequently perform better in challenging infiltration scenarios because of their capacity to recognize minute patterns in the data. This study emphasizes how crucial it is to choose the right algorithm depending on the particular needs and network environment type [12].

Machine learning is also essential for the key field of large-scale malware classification. Neural networks and random projections have been used extensively by researchers to categorize malware. Through random projections, the study achieves great computing economy without sacrificing accuracy by lowering the dimensionality of the feature space. The malware is then classified using neural networks, which show excellent accuracy and resilience against different kinds of malware. This method emphasizes how dimensionality reduction strategies may be combined with strong classifiers to improve malware detection systems [13].

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Because there are so many linked devices in the Internet of Things (IoT) setting, it is imperative to detect fraudulent communications. Using machine learning techniques, the CorrAUC approach detects harmful Bot-IoT traffic in IoT networks. This approach achieves high identification rates by using correlation-based features and traffic pattern analysis. The study illustrates the particular difficulties presented by IoT environments, including resource limitations and a variety of communication protocols, and shows how machine learning may successfully resolve these difficulties [14].

Another major cybersecurity difficulty is detecting insider threats. The potential of deep learning approaches in this field has been thoroughly examined. Key obstacles are identified by the review, such as the necessity for real-time detection and the difficulty of acquiring labeled data. Deep learning techniques, especially those based on long short-term memory (LSTM) and recurrent neural networks (RNNs), show promise in spite of these obstacles since they can identify small irregularities that may be signs of insider threats and model temporal correlations [15].

Intrusion detection systems based on machine learning have also improved risk management in network security. Safety methods are incorporated into the Safety-Augmented Network Intrusion Detection System (S-NIDS) in order to reduce hazards. This method offers a complete security solution by combining safety protocols and machine learning models to identify intrusions and control related risks. In settings where upholding system integrity and safety is crucial, this integration is very helpful [16].

Machine learning has been used to optimize network attack detection through the introduction of novel class probability features. With this method, the feature space is improved by adding probability-based characteristics that indicate the likelihood of specific network actions. This improves detection rates by enabling the detection models to differentiate between benign and malicious activity more accurately. This study highlights how crucial feature engineering is to improving machine learning models' effectiveness in network security [17].

The special difficulties in integrating cyber and physical components are addressed by the dependency-based security risk assessment for cyber-physical systems (CPS). With this method, the dependencies between various components are modeled, and the security risks are evaluated according to these dependencies. The study offers a thorough risk assessment framework that can anticipate possible vulnerabilities and their effect on the system as a whole by utilizing machine learning techniques. This approach is especially pertinent to industrial control systems and critical infrastructure where CPS is widely used [18].

An intrusion detection system (IDS) designed for edge-envisioned environments has been created in the field of smart agriculture. In severe agricultural situations, where standard detection methods might not be effective, this intrusion detection system (IDS) uses machine learning to detect intrusions. The study demonstrates how edge computing can effectively and promptly identify intrusions, protecting intelligent agricultural systems from cyberattacks [19].

The creation of an autonomous fuzzy decision support system has advanced risk assessment through big data analytics. Big data is used by this system to evaluate risks and make timely, well-informed judgments. Fuzzy logic and machine learning are combined to give a system that can manage uncertainty and produce reliable risk evaluations. This method works especially well in complicated and dynamic network situations where dangers are always changing [20].

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The creation of the machine learning-based system Smart Sentry has improved cyber threat intelligence in industrial IoT contexts. By gathering and evaluating threat intelligence data, this system offers useful insights for reducing cyber threats. The research highlights the significance of instantaneous data analysis and the function of machine learning in converting unprocessed data into insightful knowledge for anticipatory threat mitigation [21].

Industrial control systems (ICS) anomaly detection has been tackled through the creation of specific datasets and machine learning models. Researchers can efficiently train and assess machine learning models by creating an extensive anomaly detection dataset. By offering a standardized dataset that makes it easier to design and assess novel anomaly detection methods for ICS, this study advances the discipline [22].

Because IoT devices are used so widely, it is imperative to detect botnet assaults in the IoT ecosystem. It has been suggested to use a hybrid machine learning approach to effectively identify botnet attacks. To improve detection accuracy, this model integrates a number of machine learning approaches, including clustering and classification. The study emphasizes the necessity of hybrid strategies to deal with botnet assaults in IoT environments because they are varied and constantly changing [23].

Side-channel attacks are a serious risk to Internet of Things deployments. By examining side-channel data like power usage and electromagnetic emissions, machine learning approaches have been used to identify these attacks. This method adds an extra degree of security for Internet of Things devices by using the minute variations in side-channel data to detect malicious activity [24].

The implementation of machine learning models on fog devices has enabled the early identification of cyberattacks. Real-time detection is made possible by this framework's data processing at the network edge, which lowers latency and speeds up reaction times. The study shows how fog computing and machine learning may be used together to improve early threat detection in a variety of network contexts, both practically and effectively [25].

To assess their performance, a comparison of machine learning-based models with intrusion detection systems (IDS) has been done. In terms of accuracy, efficiency, and scalability, the study contrasts machine learning-based models—such as decision trees and SVM—with conventional IDS. The results imply that, in general, machine learning-based models perform better than conventional IDS, particularly when managing big and complicated information [26].

One of the biggest challenges in network intrusion detection is managing large and unbalanced data sets. Stacking feature embedding, feature extraction, and oversampling approaches are used in a study to overcome this problem. By balancing the dataset and identifying pertinent characteristics, these techniques improve machine learning models' performance and increase detection rates. In real-world situations, this strategy is essential for guaranteeing the dependability of intrusion detection systems [27].

Recent research has offered a thorough analysis of machine learning-based intrusion detection systems (IDS), threats, and network security models. The use of machine learning techniques in IDS, as well as popular attack vectors, are all covered in this review. The study offers insightful information on how network security is currently doing as well as how machine learning may be used to counter new threats [28].

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Empirical research on machine learning-based intrusion detection systems has been conducted to discover ways to defend smart-home IoT devices against MQTT attacks. This paper offers a paradigm for improving the security of smart-home Internet of Things devices by assessing the efficacy of several machine learning algorithms in identifying MQTT-based attacks. The results highlight the necessity of resilient and flexible intrusion detection systems to protect against changing Internet of Things threats [29].

Finally, an analysis of the relative performances of malware detection systems based on different classifiers and texture features has been carried out. This research examines the effectiveness of several classifiers and investigates the usage of texture features in virus detection. The findings suggest that specific texture properties might greatly increase malware detection accuracy when paired with strong classifiers. This study demonstrates how new feature sets can be used to improve malware detection systems' efficacy [30].

In summary, the evaluated works highlight how important machine learning is for identifying possible dangers and threats in network contexts. The continuous efforts to improve the precision, effectiveness, and dependability of security systems are reflected in the developments in algorithms, feature engineering, and deployment techniques. Maintaining strong network defenses will require integrating machine learning with conventional security methods as cyber threats continue to grow.

A summary of related works can be shown in Table 1, which contains the reference, target of research, and methodology applied.

TABLE 1. SUMMARY OF RELATED WORKS

reference	***Research Target***	***Method***	***Data Set used***
12	E-nose	Naive Bayesian	gas sensors
13	E-nose	SVM, ANN	chemical sensor
14	Evaluation	K-NN	heterogeneous datasets
15	E-nose	Gaussian Naive Bayes, ANN, SVM	volatile organic compounds
3	Human scent	SVM, Naive Bayes	Human odour
2	E-nose	ANN	human scents

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

From Table 1, no previous research tackled the scientific solution of our aim of this paper. Several attempts to use K-NN in E-nose technology but none of them work on human odor.

III. Risk Detection – inbound data threads inspection using KNN and Bayes :

A. *Data Sampling procedure: (Technical, Operational, and Environmental Risks) :*

Our data set has been collected consisting of 15 features, each feature is a fractional number, and collectively these 15 features represent a risk thread. Our data set covering three types of risk threads may represent inbound data to the data hub, the Technical risk, the Operational risk, and the environmental risk. The first type of risk is the technical risk contains 15 features, these features represent the thread data which is a positive potential risk representing the technical part. A predictive column has been added to the data set with the value of 1 representing the technical risk. Table 2 represents the sample of the real data from technical risk features:

Table 2. 15 features technical risk with the predictive column with value 1.

Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Feature7	Feature8	Feature9	Feature10	Feature11	Feature12	Feature13	Feature14	Feature15	Risk
0.1	0.4	0.4	0.1	0	0.07	0.07	0.24	0.34	0.28	0.31	0.31	0.02	0.01	0.35	1
0.3	0.2	0.1	0.1	0.3	0.14	0.2	0.41	0.1	0.15	0.09	0.23	0.43	0.14	0.11	1
0.4	0.1	0.1	0.2	0.2	0.03	0.11	0.2	0.43	0.23	0.15	0.14	0.05	0.03	0.63	1
0.2	0.2	0.2	0.2	0.2	0.33	0.43	0.12	0.08	0.04	0.17	0.33	0.38	0.03	0.09	1

The second type of risk is the operation risk contains 15 features, these features represent the thread data which is a positive potential risk representing the technical part. A predictive column has been added to the data set with the value of 2 representing the operation risk. Table 3 represents the sample of the real data from operation risk features:

Table 3. 15 features operation risk with the predictive column with value 2.

Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Feature7	Feature8	Feature9	Feature10	Feature11	Feature12	Feature13	Feature14	Feature15	Risk
0.11	0.4	0.42	0.01	0.06	0.12	0.01	0.09	0.43	0.35	0.41	0.25	0.21	0.09	0.04	2
0.32	0.32	0.11	0.12	0.13	0.14	0.24	0.15	0.23	0.24	0.47	0.31	0.05	0.12	0.05	2
0.14	0.1	0.1	0.32	0.34	0.2	0.12	0.18	0.25	0.25	0.4	0.2	0.14	0.05	0.21	2
0.1	0.4	0.21	0.1	0.19	0.29	0.18	0.14	0.25	0.14	0.41	0.02	0.17	0.09	0.31	2

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The third type of risk is the environmental risk contains 15 features, these features represent the thread data which is a positive potential risk representing the environmental part. A predictive column has been added to the data set with a value of 3 representing the environmental risk. Table 4 represents the sample of the real data from environmental risk features:

TABLE 4. 15 FEATURES ENVIRONMENTAL RISK WITH THE PREDICTIVE COLUMN WITH VALUE 3.

Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Feature7	Feature8	Feature9	Feature10	Feature11	Feature12	Feature13	Feature14	Feature15	Risk
0.4	0.13	0.19	0.15	0.13	0.46	0.04	0.21	0.21	0.08	0.14	0.33	0.21	0.31	0.01	3
0.34	0.14	0.21	0.21	0.1	0.29	0.29	0.21	0.01	0.2	0.32	0.14	0.16	0.01	0.37	3
0.42	0.25	0.09	0.13	0.11	0.33	0.11	0.14	0.21	0.21	0.26	0.33	0.01	0.12	0.28	3
0.09	0.09	0.02	0.43	0.37	0.5	0.08	0.03	0.31	0.08	0.41	0.07	0.09	0.14	0.29	3

Combing all the features together in a single table to compose the input data set to the machine learning model. The data is fractional and does not require any normalization or scaling tools. The same data set will be applied to KNN and Bayes for comparison and efficiency.

B. *Apply Naïve Bayes algorithm:*

Naïve Bayes is one of the most popular data mining algorithms. Its efficiency comes from the assumption of attribute independence, although this might be violated in many real-world data sets. The classification task that Naïve Bayes solves is assumed as follows. Given a training data sample $Dtrain$ of t classified objects, we are required to predict the probability P($y|x$) that a new example x=x1,x2…,xa belongs to some class y, where xi is the value of the attribute Xi and $y\in\{1,\ldots,c\}$ is the value of class variable Y [31].

Bayes theorem helps determine the likelihood that one event will occur with unclear information while another has already happened. The mathematical formulation of the Bayes theorem is [32, 33, 34, 35]:

$$\rho(A|B) = \frac{\rho(B|A)\rho(A)}{\rho(B)} \quad \square\square\square$$

Where A and B are events and $\rho(B) \neq 0$

$\rho(A|B)$ is a conditional probability in which the probability of event A occurring given that B is true called posterior probability of A given B [32, 33, 34, 35]. $\rho(A|B)$ is also a conditional probability in which the probability of event B occurring given that A is true which can be interpreted as the likelihood L of A given a fixed B because $\rho(A|B)$=L(A|B) [32, 33, 34, 35]. $\rho(A)$ and $\rho(B)$ are the probabilities of observing A and B respectively without any given conditions, they are known as prior probability and marginal probability [32, 33, 34, 35].

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Results obtained from the Naive-Bayes algorithm have been applied to a structured data set with 15 input features representing the incoming risk data set to the network and a target prediction column of three different risk categories, namely environmental, operation, and technical risk. The best results from Naive-Bays obtained an accuracy of 84%. Several mathematical equations have been tested using Naïve Bayes for accuracy enhancement and research inspection. Results obtained from Bernoulli Naïve Bayes was 47%, for Categorical Naïve Bayes accuracy is 47%, Complement Naïve Bayes bring 81% accuracy, and Multinominal Naive Bayes bring 47% accuracy. The best accuracy obtained from Gaussian Naïve Base, figure 1 summarizes the results obtained from a variety of Naïve Bayes equations:.

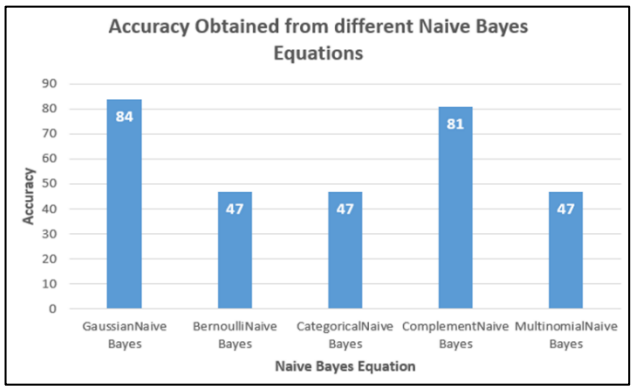

Fig. 1. Results Obtained from a variety of Naïve Bayes equations.

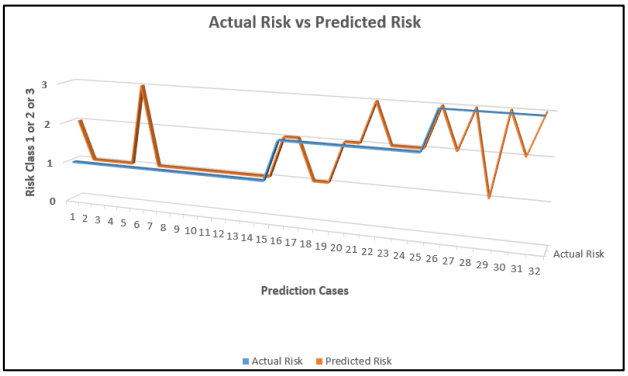

Fig. 2. Matching between actual and predicted risk using Gaussian Naïve Bayes equations.

C. *Apply the KNN algorithm:*

The k-nearest neighbor (KNN) algorithm is a novel machine learning method proposed in this article based on the modified, which can extract more features from the data sets through the advanced workflow and simulation techniques [36]. For the KNN algorithm, we can implement a KNN method by following the steps [36]:

1. Calculate the distance between the new point and each training point.
2. The closest k points are selected based on the distance.
3. The wighted average value of these selected data points serves as the final prediction for the new point.

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Mathematically, assume a value k for the KNN and a prediction point x_0 and then use N_0 to denote the k closest training observations to the prediction point x_0. The KNN returns the estimation

f (x_0) using the average of all the responses in N_0 as shown in equation 1 [3**6**]:

$$f(x_0) = \frac{1}{k} \sum_{x_i \in N_0} y_i \qquad \square\square\square\square\square$$

Distance between neighbor points can be calculated using many methods to calculate the distance in the KNN. The most used three methods are Euclidian, Manhattan, and Minkowski distances [36]. The Minkowski distance or Minkowski metric is a measurement in a normed vector space which can be considered as a speculation of both the Euclidean distance and the Manhattan distance. Calculate the Minkowski distance between two factors. The situation where p = 1 is identical to the Manhattan distance and the situation where p = 2 is comparable to the Euclidean distance. Minkowski distance calculation is as shown below in equation2 [36]:

$$\sqrt[p]{\sum_{i=1}^{n} |x^i - y^i|} \qquad (2)$$

Manhattan distance calculation is as shown below in equation2 [3**6**]:

$$d(X,Y) = \sum_{i=1}^{n} |X_i - Y_i|$$

Euclidean distance calculation is as shown below in equation2 [3**6**]:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n} (Y_i - X_i)^2}$$

Results obtained from the KNN algorithm using Minkowski with 5 neighbors have been applied to a structured data set with 15 input features representing the incoming risk data set to the network and a target prediction column of three different risk categories, namely environmental, operation, and technical risk. Results from KNN having k=5 with the Minkowski, Manhattan, and Euclidean distance equation obtained an accuracy of 75%. Results from KNN having k=7 and the Minkowski distance got 72%, the Manhattan distance equation got 63%, and the Euclidean distance equation obtained an accuracy of 72%. Figure Fig. 3. Shows the results obtained from KNN with variety of K values and different distance measurement equations.

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

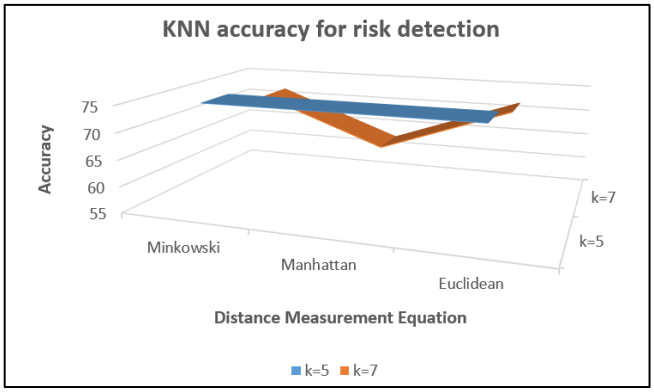

Fig. 3. Results Obtained from the KNN algorithm with k=5 and 7 using Mikowski, Manhattan, and Euclidean.

CONCLUSION

The research aims to identify potential risks and threads resulting from incoming data passing across the network. Three distinct risk categories are examined in our research: technical, operational, and environmental risks. To ensure that the results are rigid and for the purpose of study comparison, two machine learning techniques have been applied. A structured data set of 15 input characteristics that represent the entering risk data set to the network and a target prediction column of three distinct risk categories—environmental, operating, and technical risk—has been subjected to the Naive-Bayes and K-Nearest Neighbor algorithms. The risk detection accuracy from Naive-Bays results was 84%, and the risk detection accuracy from K-Nearest Neighbor with five neighbors was 75%.

REFERENCES

[1] Mussiraliyeva, S., Bolatbek, M., Omarov, B., Bagitova, K. (2020). Detection of Extremist Ideation on Social Media Using Machine Learning Techniques. In: Nguyen, N.T., Hoang, B.H., Huynh, C.P., Hwang, D., Trawiński, B., Vossen, G. (eds) Computational Collective Intelligence. ICCCI 2020. Lecture Notes in Computer Science(), vol 12496. Springer, Cham. https://doi.org/10.1007/978-3-030-63007-2_58.

[2] Farouk, M., Sakr, R.H. & Hikal, N. Identifying the most accurate machine learning classification technique to detect network threats. Neural Comput & Applic 36, 8977–8994 (2024). https://doi.org/10.1007/s00521-024-09562-9

[3] Itoo, F., Meenakshi & Singh, S. Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. Int. j. inf. tecnol. 13, 1503–1511 (2021). https://doi.org/10.1007/s41870-020-00430-y

[4] S. Wang, J. F. Balarezo, S. Kandeepan, A. Al-Hourani, K. G. Chavez and B. Rubinstein, "Machine Learning in Network Anomaly Detection: A Survey," in IEEE Access, vol. 9, pp. 152379-152396, 2021, doi: 10.1109/ACCESS.2021.3126834.

[5] Handa, A., Sharma, A., & Shukla, S. K. (2019). Machine learning in cybersecurity: a review. WIREs Data Mining and Knowledge Discovery, 9(4). https://doi.org/10.1002/widm.1306

[6] Mashechkin, I.V., Petrovskiy, M.I., Tsarev, D.V. et al. Machine Learning Methods for Detecting and Monitoring Extremist Information on the Internet. Program Comput Soft 45, 99–115 (2019). https://doi.org/10.1134/S0361768819030058

Received: August 09, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[7] L. Yang, A. Moubayed, I. Hamieh and A. Shami, "Tree-Based Intelligent Intrusion Detection System in Internet of Vehicles," *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013892.

[8] Waqas, M., Tu, S., Halim, Z. et al. The role of artificial intelligence and machine learning in wireless networks security: principle, practice and challenges. Artif Intell Rev 55, 5215–5261 (2022). https://doi.org/10.1007/s10462-022-10143-2

[9] Aziz, S., Dowling, M. (2019). Machine Learning and AI for Risk Management. In: Lynn, T., Mooney, J., Rosati, P., Cummins, M. (eds) Disrupting Finance. Palgrave Studies in Digital Business & Enabling Technologies. Palgrave Pivot, Cham. https://doi.org/10.1007/978-3-030-02330-0_3

[10] Ardabili, S., Mosavi, A., Várkonyi-Kóczy, A.R. (2020). Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods. In: Várkonyi-Kóczy, A. (eds) Engineering for Sustainable Future. INTER-ACADEMIA 2019. Lecture Notes in Networks and Systems, vol 101. Springer, Cham. https://doi.org/10.1007/978-3-030-36841-8_21

[11] Q. H. Vu, D. Rutan and L. Cen, "Gradient boosting decision trees for cyber security threats detection based on network events logs," 2019 IEEE International Conference on Big Data (Big Data),Los Angeles, CA, USA, 2019, pp.5921-5928,doi: 10.1109/BigData47090.2019.9006061.

[12] T. Saranya, S. Sridevi, C. Deisy, Tran Duc Chung, M.K.A.Ahamed Khan, Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review, Procedia Computer Science, Volume 171,2020, Pages 1251-1260,ISSN 1877-0509, https://doi.org/10.1016/j.procs .2020.04.133.

[13] G. E. Dahl, J. W. Stokes, L. Deng and D. Yu, "Large-scale malware classification using random projections and neural networks," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013, pp. 3422-3426, doi: 10.1109/ICASSP.2013.6638293.

[14] M. Shafiq, Z. Tian, A. K. Bashir, X. Du and M. Guizani, "CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques," in IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3242-3254, 1 March1, 2021, doi: 10.1109/JIOT.2020.3002255.

[15] Shuhan Yuan, Xintao Wu, Deep learning for insider threat detection: Review, challenges and opportunities, Computers & Security, Volume 104, 2021,102221, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2021.102221.

[16] B. Brenner *et al*., "Better Safe Than Sorry: Risk Management Based on a Safety-Augmented Network Intrusion Detection System," in *IEEE Open Journal of the Industrial Electronics Society*, vol. 4, pp. 287-303, 2023, doi: 10.1109/OJIES.2023.3297057.

[17] A. Raza, K. Munir, M. S. Almutairi and R. Sehar, "Novel Class Probability Features for Optimizing Network Attack Detection With Machine Learning," in *IEEE Access*, vol. 11, pp. 98685-98694, 2023, doi: 10.1109/ACCESS.2023.3313596.

[18] Akbarzadeh, A., Katsikas, S.K. Dependency-based security risk assessment for cyber-physical systems. Int. J. Inf. Secur. 22, 563–578 (2023). https://doi.org/10.1007/s10207-022-00608-4

Received: August 09, 2025

[19] D. Javeed, T. Gao, M. S. Saeed and P. Kumar, "An Intrusion Detection System for Edge-Envisioned Smart Agriculture in Extreme Environment," in *IEEE Internet of Things Journal*, vol. 11, no. 16, pp. 26866-26876, 15 Aug.15, 2024, doi: 10.1109/JIOT.2023.3288544.

[20] M. Siami, M. Naderpour, F. Ramezani and J. Lu, "Risk Assessment Through Big Data: An Autonomous Fuzzy Decision Support System," in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 8, pp. 9016-9027, Aug. 2024, doi: 10.1109/TITS.2024.3392959.

[21] S. Sadhwani, U. K. Modi, R. Muthalagu and P. M. Pawar, "SmartSentry: Cyber Threat Intelligence in Industrial IoT," in IEEE Access, vol. 12, pp. 34720-34740, 2024, doi: 10.1109/ACCESS.2024.3371996.

[22] A. Dehlaghi-Ghadim, M. H. Moghadam, A. Balador and H. Hansson, "Anomaly Detection Dataset for Industrial Control Systems," in *IEEE Access*, vol. 11, pp. 107982-107996, 2023, doi: 10.1109/ACCESS.2023.3320928.

[23] M. Ali, M. Shahroz, M. F. Mushtaq, S. Alfarhood, M. Safran and I. Ashraf, "Hybrid Machine Learning Model for Efficient Botnet Attack Detection in IoT Environment," in *IEEE Access*, vol. 12, pp. 40682-40699, 2024, doi: 10.1109/ACCESS.2024.3376400

[24] A. D. Campos, F. Lemus-Prieto, J. -L. González-Sánchez and A. C. Lindo, "Intrusion Detection for IoT Environments Through Side-Channel and Machine Learning Techniques," in IEEE Access, vol. 12, pp. 98450-98465, 2024, doi: 10.1109/ACCESS.2024.3362670.

[25] K. U. Aditya, P. N. Kamath, Y. Poral, B. D. Mallika and V. Acharya, "Framework for Early Cyber Attack Detection Using ML Models Deployed On Fog Devices," *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, San Antonio, TX, USA, 2024, pp. 1-6, doi: 10.1109/ISDFS60797.2024.10527351.

[26] Z. Azam, M. M. Islam and M. N. Huda, "Comparative Analysis of Intrusion Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree," in *IEEE Access*, vol. 11, pp. 80348-80391, 2023, doi: 10.1109/ACCESS.2023.3296444.

[27] Talukder, M.A., Islam, M.M., Uddin, M.A. et al. Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. J Big Data 11, 33 (2024). https://doi.org/10.1186/s40537-024-00886-w

[28] D. Kapil, N. Mehra, A. Gupta, S. Maurya and A. Sharma, "Network Security: Threat Model, Attacks, and IDS Using Machine Learning," *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India, 2021, pp. 203-208, doi: 10.1109/ICAIS50930.2021.9395884.

[29] R. Alasmari and A. A. Alhogail, "Protecting Smart-Home IoT Devices From MQTT Attacks: An Empirical Study of ML-Based IDS," in IEEE Access, vol. 12, pp. 25993-26004, 2024, doi: 10.1109/ACCESS.2024.3367113.

[30] I. T. Ahmed, B. T. Hammad and N. Jamil, "A Comparative Performance Analysis of Malware Detection Algorithms Based on Various Texture Features and Classifiers," in *IEEE Access*, vol. 12, pp. 11500-11519, 2024, doi: 10.1109/ACCESS.2024.3354959.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[31] Shenglei Chen, Geoffrey I. Webb, Linyuan Liu, Xin Ma, A novel selective naïve Bayes algorithm, Knowledge-Based Systems, Volume 192,2020, 105361, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2019.105361.

[32] Stuart, A.; Ord, K. (1994), Kendall's Advanced Theory of Statistics: Volume I – Distribution Theory.

[33] F. -J. Yang, "An Implementation of Naive Bayes Classifier," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 301-306, doi: 10.1109/CSCI46756.2018.00065.

[34] Chen, H., Hu, S., Hua, R. et al. Improved naive Bayes classification algorithm for traffic risk management. EURASIP J. Adv. Signal Process. 2021, 30 (2021).

[35] Peng, F., Schuurmans, D. & Wang, S. Augmenting Naive Bayes Classifiers with Statistical Language Models. Information Retrieval 7, 317–345 (2004).

[36] Sabri, A.Q., Al-Nuaimi, Z. “Inverse-Distance Weighted K- Nearest Neighbor for Raw and Scaled Data set in Human Detection using Odour”, 2023 10th International Conference on Soft Computing and Machine Intelligence, ISCMI 2023, 2023, pp. 161–165.

Received: August 09, 2025