RESEARCH PAPER

# Enhancing the Efficiency of IDS through Machine Learning by Providing Authentication

**Pratik Jain[1], Sonali Abhijeet Padalkar[2], Khushbu Tikhe Konde[3], Megha Shrivastav[4], Harleen Kaur Malhotra[5], Suhani Nimkar[6]**

[1]IPS Academy, Institute of Engineering & Science, Indore, Email id: Pratikjain@ipsacademy.org
[2]Shree L. R. Tiwari College of Engineering, Thane, Email id: sonalipadalkar88@gmail.com
[3]Shree L. R. Tiwari College of Engineering, Thane, Email id: tikhe.khushbu@gmail.com
[4]IPS Academy, Institute of Engineering & Science, Indore, Email id: meghashrivastav@ipsacademy.org
[5]IPS Academy, Institute of Engineering & Science, Indore, Email id: harleenkaurmalhotra@ipsacademy.org
[6]IPS Academy, Institute of Engineering & Science, Indore, Email id: nimkarsuhani@gmail.com

**ABSTRACT**

Alerts are sent by intrusion detection systems when they observe unusual activity while keeping an eye on known threats and network traffic. Intrusion detection is a hot topic in today's cyberspace security. It might be impartial. However, the duration is longer. We must first define what an incursion is in order to comprehend intrusion detection. Oxford Learners Dictionary defines intrusion as "the act of intruding." an unwelcome or private location. Because of this, this article describes an unauthorized system or network intrusion that is connected to one or more systems or networks. Here's an illustration of a devoted user attempting to log in to the system. Access takes longer to finish than usual during the trial time. The Aim is to establish a connection to a certain account or an unapproved user's remote server port. Fired former worker Real Workers May Also Be Triggered by Late Intruders I made it by my actions of provoking or invading people from the outside. According to this clause, attacks are regarded as false positive cases on average. The examples and solutions provided for the same problem are the main topic of this paper. The KDD CUP 1999 record is used here. The anomaly class, based on the results, is an A class that has a greater count than this class. When a genuine user attempts to access it, strange things happen. This is a result of the large number of classes. This study presents a method to identify genuine people, Eliminate false positives.

**Keywords:** Intrusion Detection Systems (IDS), Cybersecurity, Anomaly Detection, False Positives, Network Security, Unauthorized Access.

**How to cite this article:** Jain P, Padalkar SA, Konde KT, Shrivastav M, Malhotra HK, Nimkar S. Enhancing the Efficiency of IDS through Machine Learning by Providing Authentication. Int J Drug Deliv Technol. 2026;16(19s): 617-622. DOI: 10.25258/ijddt.16.19s.70

## INTRODUCTION

In the last 20 years, technology has seen significant transformation. Furthermore, it is ideal right now in terms of security, given the rising risk of theft. The importance of network systems is growing. And then there have been instances of abuse. Although networks have existed for a very long time, they have recently become more widespread due to computer technology in many parts of the world. Finding the best place to secure data has become essential, as it contains vital information. The methods used today to secure data include data encryption, firewalls, and VPNs. You don't get any more effective in data breaches since you don't catch any intrusions. However, intrusion detection is more flexible. It additionally provides network defense against incursion.

Attack/Counterattack:-A network intrusion detection system, or "NIDS," often employs one of the three architectural paradigms. Anomaly bases, signature-based detection systems, and protocol modeling are examples of this. The devices in each model have different capabilities and drawbacks. Here, we'll combine the three models.

**Signature-based NIDS**

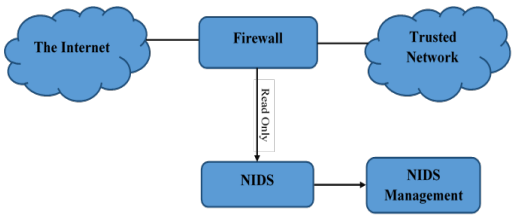


**Figure 1(a) Signature – Based NIDS**

The very broad design of all NIDS devices, which is seen in figure 1, is somewhat dependent on signature-based detection (a). These technological exposition packets pertain to specific patterns associated with well-known attacks. It is appropriate for accurately recognizing known assaults and is suitable for unzipping, observing, and updating. One disadvantage, despite its great effectiveness, is that it might not identify an unknown or changed attack. To some extent, almost all

NIDS devices are stopped. Dependence on detection based on signatures. This happens a lot. Figure I shows the design (a). This limited-edition technology exhibition pack is connected to the attack by familiar. It works rather well for unpacking, observing, and updating; it is also appropriate for accurately recognizing known attacks. Incredibly efficient, but there are downsides Does not recognize unknown or modified attacks.

**Categories of intrusion detection systems**:

In the figure 1. (b) below, Two types of intrusion detection systems (IDSs) exist. detection systems based on signatures, anomaly-based detection systems, and protocol modeling. The intrusion detection system searches for unusual activity when monitoring your network and other recognized dangers, and it notifies you when something is amiss.

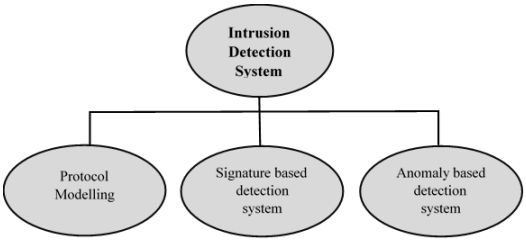


**Fig 1(b) Categories of intrusion** detection systems

1. **Protocol modeling**: This method checks for unusual network activity by watching for strange patterns in network traffic and analyzing logs. Alerts are sent out if the traffic doesn't match known protocols or if it's something the system doesn't recognize. To do this effectively, multiple sources of data are used to understand what normal network behavior looks like. This involves looking at network traffic, using protocol guidelines like RFCs, and reviewing applications that use the same protocols. Essentially, it's about comparing current traffic with established norms to spot any irregularities.
2. **Signature-based recognition system**: Recognized signatures are used by a detection system based on signatures (SBIDS) to pinpoint threats.The effectiveness of this approach comes from its robust and frequently updated signatures. It is designed to detect known threats and avoid engaging with recognized hostile activities. However, it may struggle with identifying unusual or new types of attacks, which is a common limitation. Despite this, its detection rate for known threats remains reliable.
3. **An Anomaly detection system:** The purpose of an anomaly-based intrusion detection system (ABIDS) is to detect novel and unidentified threats that conventional techniques could miss. Its ability to identify new dangers that aren't covered by current machine learning models is what makes it so strong. Compared to Signature-Based Intrusion Detection Systems (SBIDS), ABIDS has two primary advantages. Initially, by contrasting recent activity with established patterns and spotting anomalous behaviors, it can recognize external threats and zero-day attacks. Secondly, it builds a unique profile of typical behavior that is unique to your system, making it easier to spot any odd behavior without setting off alarms. How well ABIDS handles various rule sets and tests all pertinent protocols determines how well it performs. Managing the division of rules for identifying anomalies is a major difficulty.

**LITERATURE SURVEY**

In this paper entropy merging is compared. Vector machines and network functions both one definition of a hybrid process is an individual process. A technique for detecting anomalous traffic that compares the entropy of Support Vector Machines (SVM) and Network Functions [1]. This study demonstrates how the reduction technique is used with symbolic dynamic filtering (SDF) [2]. The article discusses feature selection and intrusion data classification. The hybrid intrusion method used in the swarn optimization Base Coarse Set (IDR-IR), an intelligent dynamic swan detection system [3]. B. Data reduction and Fussy Adaptive Resonance Theory (Fussy ART) will be covered in the study. Networks of anomalies for categorization Principal Component Analysis (PCA)-based intrusion detection technique is applied [4]. To become Functions in the SDF in order to build model taxonomy for probability calculations Time series data provide the foundation of finite state automata (PFSA). In the process of creating symbol sequences, a distinction is made [5]. or the writers clarify that the atypical application of cluster analysis Utilizing a basic K-Mean clustering technique for detection 2. K-Mean clustering is a well-known approach that is straightforward and primarily utilized for giants. Note the computer's depth [6]. This is the writer of concentrating on data mining with the Embattle Tree and data from the 1999 KDD Cup Techniques During his investigation, Face Direct his device According to his experimental findings, the SVM is less effective than the C4.5 algorithm in detecting anomalies and false alarm rates using 1999 KDD Cup data [7]. The authors of this study adjusted the false alarm rate decrease and The primary method aims to increase the detection rate. The underlying IDS is seen from a hybrid perspective in clustering analysis. mining of data [8]. The article provides two explanations. Why do deterministic datasets exist in high-dimensional datasets? The number of clusters the user specified does not match Due to the fact that it produces erroneous dates on a variety of

irregular dates, grades are evaluated [9]. In order to facilitate the process of detecting anomalies without human intervention, novel density-based clustering algorithms and grid-based clustering algorithms are presented in this study [10]. This document provides justification for this bundle. Two elements are data mining taxonomy and technique. Its framework for hybrid detection is dependent upon [11]. This document Lighting is the capacity to generate the intended result. measured using three matrices: maximum Trust rule, recall, and precision. The behavior of the end method with increased noise may be the cause of changes in the capacity to extract the rule accurately [12]. The document will be supplied in a way that makes it possible to identify different intruders. You conduct a number of extensive and contrasting investigations. Program for detecting anomalies [13]. The following is made clear by the author: Production management decision-making truly takes into account current results. is a workable technique that uses work order data to build the Astir network [14]. This essay provides all justifications. Our tool, Weka, uses an algorithm [15]. BoraBardük et al [16]. is more advantageous for companies to preserve the interests of customers than to win new customers. Identifying customers who are at risk of churn and churn can enable companies to change their strategy in advance to achieve the best results. This study proposes a new method of non-contractual business. One of the most widely used methods in the literature for these types of companies is to use Beta Geometric Negative Binomial Distribution (BG / NBD) to model customer behavior. Although this method performs well in predicting the overall churn rate, it does not perform well in marking individual customers. This research aims to improve the BG / NBD model by incorporating machine learning into the decision-making process. In the proposed method, the mathematical definition of the BG /NBD model is used for feature of features and is used through decision trees. Chiaki Doi et al [17]. Projected a technique to predict customer value by focusing on purchasing behavior. This method generates a correlation model for the number of days and the purchase amount in each period between customer value and purchase history based on a 17 consumer group survey in advance. The author uses the random forest method to produce predictable model. The proposed method helps to provide intelligent customer management to each customer based on e.g. the level of recommended products or services. When testing two data sets with different characteristics, it can be observed that the proposed method can significantly improve the performance. Dehua Kong et al [18]. Predicting the customer's point of interest is the key to improving the accuracy of e-commerce recommendations in a big data environment. Current technology is primarily to predict existing customer interests, but does not fully take into account the impact of customer behavior, time series, and time on product recommendations. Research in the prediction method of points of interest based on the customer network's spatial-temporal behavior, it is necessary to build a customer network spatiotemporal behavior super net model that includes four subnets of customer, time, behavior and point of interest. At the same time, the behavioral factors are introduced, and an algorithm for predicting points of interest based on super-edge similarity is proposed to address the influence of customers' multiple behaviors and timing on recommended products, thereby improving the recommendation's success rate. Harsh Valecha et al [19]. in the ultra-modern era of technology, the prediction of market trends is very important to observe consumer behavior in this highly competitive world because trends are volatile. Based on the development of machine learning and previous work in the science of predicting behavior, we built a model for predicting consumer behavior. The purpose of this research report is to examine the relationship between consumer behavioral parameters and buying intentions. First, we conduct a study to discover the relationship between consumer buying behavior and environmental factors, organizational factors, personal factors, and interpersonal factors. Therefore, this paper proposes a time-varying random forest classifier that uses unique functional techniques to predict consumer behaviors that affect product selection. The results of the random forest classifier are more accurate than other machine learning algorithms.

**PROBLEM RECOGNITION**

Intrusion detection involves identifying unauthorized access or activities that occur in various stages. An intrusion detection system (IDS) is used to monitor and protect against these unauthorized entries by recognizing deviations from established patterns of normal behavior. There are two primary approaches to detecting anomalous activity:

1. **Predefined Normal Behavior:** This technique involves setting a baseline of normal activities. Any behavior that significantly deviates from this baseline is flagged as potentially abnormal. The system observes and evaluates activities to determine if they fall outside the established norms, distinguishing them from known malicious patterns.
2. **Predefined Intrusion Behavior:** The main goal of this strategy is to identify and document known harmful or invasive activities. In order to detect and react to threats more quickly and accurately, it entails comparing ongoing activity with known patterns of intrusion. This technique keeps the detection rate higher while lowering false

positives. The accuracy (AC), detection rate (DR), and false alarm rate (FAR) of intrusion detection systems (IDS) are typically used to evaluate them.

Formula: For calculating it
Suppose,
T.P = True Positive, T.N = True Negative,
F.P = False Positive, F.N = False Negative

1) Accuracy = (T.N + T.P) / (F.N+ T.N + F.P + T.P)
2) Rate of detection = (T.P)/ (F.P+ T.P)
3) Rate of False alarm = (F.P)/(T.N+ F.P)

**Table 1 Intrusion detection general behavior data**

| **Prediction** | **Actual** | |
|---|---|---|
| | **Intrusion** | **Normal** |
| **Predicted Intrusion** | True positive | False Positive |
| **Predicted Normal** | False Negative | True Negative |

The data from the 1999 KDD Cup are used in this paper. When Lincoln Labs prepared and managed his program in 1998. It's called the DARPA Intrusion Detection Assessment Program. The primary subjects of this article are the examination, evaluation, and conclusions of intrusion detection research. Here, we are employed. Technically speaking, the issue where regular data is interpreted as an intrusion is known as a false positive. On these dates, the set is set up in a military network. A number of intrusions have been made to look at the outcomes. For the duration of this study, this will be used as the survey default record. The 1999 KDD Intrusion Detection Contest makes use of the same version of the dataset. When anomalies are compared to the normal class and 41 attribute class, one can determine if the object is an input. Be a part of the typical or unusual class. 41 characteristics include: allocate 1:time act for F1, allocate 2:protocol_type act for F2, allocate 3:service act for F3, allocate 4:flag act for F4, allocate 5:src_bytes act for F5, allocate 6:dst_bytes act for F6, allocate 7:land represented as F7, allocate 8:wrong_fragment act for F8, allocate 9:urgent act for F9, allocate 10:hot act for F10, allocate 11:num_failed_logins act for F11, allocate 12:logged_in act for F12, allocate 13:num_compromised act for F13, allocate 14:root_shell act for F14, allocate 15:su_attempted act for F15, allocate 16:num_root act for F16, allocate 17:num_file_creations act for F17, allocate 18:num_shells act for F18, allocate 19:num_access_files act for F19, allocate 20:num_outbound_cmds act for F20, allocate 21:is_host_login act for F21, allocate 22:is_guest_login act for F22, allocate 23: ‘count’ act for F23, allocate 24:srv_count act for F24, allocate 25:serror_rate act for F25, allocate 26:srv_serror_rate act for F26, allocate 27:rerror_rate act for F27, allocate 28:srv_rerror_rate act for F28, allocate 29:same_srv_rate act for F29, allocate 30:diff_srv_rate act for F30, allocate 31:srv_diff_ host_rate act for F31, allocate 32:dst_host_count act for F32, allocate 33:dst_host_srv_count act for F33, allocate 34:dst_host_ same_srv_rate act for F34, allocate 35:dst_host_diff_srv_rate act for F35, allocate 36:dst_host_same_src_port_rate act for F36, allocate 37:dst_host_srv_diff_host_rate act for F37, allocate 38:dst_host _serror_rate act for F38, allocate 39:dst_host_srv_serror_rate act for F39, allocate 40:dst_host_ rerror_rate act for F40, allocate 41:dst_host_srv_rerror_rate act for F41 Class: Normal: Class A; anomaly:Class B It will now be determined by these 41 allocates whether the data is anomalous or normal. As an illustration: Let's look at four KDD Cup 1999 dataset examples.

**EXPERIMENTS AND RESULTS**

The increase in the count attribute can lead to a higher rate of false positives. When upgrading systems, it is crucial to focus on two key factors: the recognition rate and the false alarm rate. The recognition rate is calculated by dividing the number of true intrusions identified by the system by the total number of actual intrusion patterns in the dataset. This measure reflects how effectively the system detects intrusions. Conversely, the false alarm rate measures how often normal instances are incorrectly identified as intrusions.

To enhance system performance, it is essential to evaluate these metrics using sample data. An ineffective count attribute can be removed and replaced with One-Time Passwords (OTPs) or unique passwords to address authentication issues. Integrating OTPs with email addresses or phone numbers is an effective method to improve security and reduce authentication problems.

**Algorithm 1: Registration**

1. Start the registration process.
2. Enter your details and complete the registration form with all required fields.
3. If any required fields are missing, display an "Error message" dialog box.
4. If all required fields are filled, show a confirmation message indicating successful registration.
5. End the registration process.

**Algorithm 2: Registration**

1. Begin the registration process.
2. Fill out the CAPTCHA or choose the "I'm not a robot" checkbox after entering the username and password.
3. The registration is successful if the password, username, and CAPTCHA are correct.

4. If the entries are invalid, allow up to 10 attempts to correct the information. Repeat steps 1 and 2 as needed.
5. Create an OTP and send it to the user's mobile or email address.
6. If the OTP is accurate, repeat steps 1 through 4 as needed.

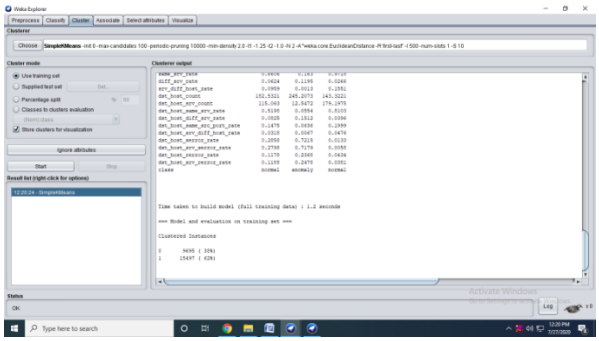

**Fig.2 Result of Experiment with count attribute**

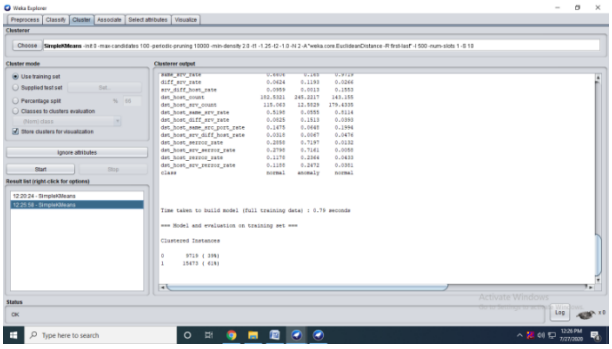

**Fig 3Result of Experiment without count attribute**

**Table 2: A comparison of the Simple K-mean algorithm's outcomes.**

| Algorithm | Time taken with count attribute | Time taken without count attribute |
|---|---|---|
| Simple k mean | 1.2 seconds | 0.79 seconds |

## CONCLUSION

To address this issue, this document suggests removing the count attribute. By doing so, performance improvements are observed, as shown in Table 2. Specifically, the K mean value decreased from 1.2 to 0.79. These changes contribute to increased efficiency, reduced false positives, and faster detection of issues, ultimately lowering the risk of system damage from delayed responses.

## REFERENCES


1. Kapil Wankhade, Mrudula Gudadhe, Prakash Prasad, "A New Data Mining Based Network Intrusion Detection Model", In Proceedings of ICCCT 2010, IEEE, 2010, pp.731-735.
2. Dorothy E. Denning. 1986 IEEE "An Intrusion-Detection Model" Computer Society Symposium on Research in Security and Privacy, pp 118-31.
3. Shu Wu, Member, and Shengrui Wang "Information- Theoretic Outlier Detection for Large-Scale Categorical Data" VOL. 25, NO. 3, MARCH 2013.
4. Bhavani Thuraisingham "Data Mining for Malicious Code Detection and Security Applications" 2009 IEEE/WIC/ACM 2009.
5. S. K. Chaturvedi1 , Prof. Vineet R. , Prof. Nirupama T. "Anomaly Detection in Network using Data mining Techniques" International Journal ISSN 2250-2459 Volume 2, Issue 5, May 2012.
6. T. Bhavani et al., "Data Mining for Security Applications," Proceedings of the 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing Volume 02, IEEE Computer Society 2008.
7. Francesco Mercaldo, "Identification of anomalies in processes of database alteration" IEEE 2013.
8. Detection & amp;quot; jcai , pp.275- 277, 2009 International Joint Conference Artificial Intelligence, 2009.
9. Shih-Wei Lina, Kuo-Ching Yingb, Chou-Yuan Leec, Zne-Jung Leed "An intelligent algorithm with Feature selection and decision rules applied toanomaly detection" Elsevier 2011.
10. Jonathan J, Davis , Andrew J. Clark "Data preprocessing for anomaly based network intrusion detection: A review" Elsevier 2011.
11. V. Chandola,A.Banerjee,V.Kumar, "Anomaly detection as a survey" ACM Comput. Surv. 41(3) (2009)15:1–15:58.
12. UgoFiore , Francesco, Aniello "Network anomaly detection with the restricted Boltzmann machine" Neurocomputing 122 (2013) 13–23.
13. Abdul Samad bin Haji Ismail "A Novel Method for Unsupervised Anomaly Detection using Unlabeled Data" IEEE 2008.
14. Bharat Singh, Nidhi Kushwaha and OP Vyas "Exploiting Anomaly Detections for high Dimensional data using Descriptive Approach of Data mining" IEEE(ICCT) 2013.
15. S. Gnanapriya, R. Adline Freeda, M. Sowmiya
a. "Evaluation of Clustering Capability Using WekaTool" International Journal of Innovations in Engineering and Technology (IJIET) 2017.
16. Bora Bardük "Modelling Time Statistics for Customer Churn Prediction", 2020 28th Signal Processing and Communications Applications Conference (SIU) Year: 2020.
17. ChiakiDoi; Masaji Katagiri; Takashi Araki; DaizoIkeda; Hiroshi Shigeno "Is he Becoming an Excellent Customer for us? A Customer Level Prediction Method for Customer Relationship Management System" 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA) Year: 2018
18. DehuaKong; XingLi; Yongxia Zhao "Research on Product Recommendation Based on Web Space-Time Customer Behavior Trajectory", 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)Year: 2019.
19. Harsh Valecha; Aparna Varma; Ishita Khare; Aakash Sachdeva; Mukta Goyal "Prediction of Consumer Behaviour using Random Forest

Algorithm" 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) Year: 2018.