Mesopotamian journal of Cybersecurity
Vol.2023, **pp**. 64–71
DOI: https://doi.org/10.58496/MJCS/2023/011; ISSN: 2958-6542
https://mesopotamian.press/journals/index.php/CyberSecurity

Research Article

Cyber Security System Based on Machine Learning Using Logistic Decision Support Vector

M.Sahaya Sheela[1], D.Hemanand[2], Vallem Ranadheer Reddy[3]

[1]Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu-600062, India.
[2]Department of Computer Science and Engineering, S.A. Engineering College (Autonomous), Thiruverkadu, Chennai-600077, Tamil Nadu, India.
[3]Chaitanya Deemed To Be University, Kishanpura, Hanamkonda, Warangal -506001, Telangana, India.

ARTICLE INFO

Article History

Received 9 Jan 2023
Accepted 19 March 2023
Published 24 March 2023

Keywords

Cyber security
KDD cup99

Feature selection
malicious activity

Classification

ABSTRACT

Nowadays, we are moving towards cybersecurity against digital attacks to protect systems, networks, and data in developing areas. A collection of technologies and processes is at the core of cybersecurity. A network security system is a feature of network and computer (host) security. Cybercrime leads to billion-dollar losses. Given these crimes, the security of computer systems has become essential to reduce and avoid the impact of cybercrime. We propose the Logistics Decision Support Vector (LDSV) algorithm dealing with this problem. Initially, we collected the KDD Cup 99 dataset to create a network intrusion detection, such as penetrations or attacks, a prognosis model that varies between the "Non Malicious" and "Malicious" standard links. These method finds the cyber-attack category based on the behavior features. In the second step, data preprocessing should be cleaned from errors, and raw data should be converted into a prepared dataset. The third step is Feature Selection (FS) techniques often improve the feature selection process in an Intrusion Detection System (IDS) that is more convenient for using the mean of the Chi-square test (MAC) method. Finally, a classification is done to classify and detect the network intrusion detection based on LDSV for Cyber security. The proposed LDSV simulation is based on the Precision F-Measure, Recall, and Accuracy for the best result.

1. INTRODUCTION

Cyber security protects digital information from unauthorized access, damage, or theft throughout its lifetime. Cyber security protects networks, devices, and data from unauthorized access and misuse and maintains information confidentiality, integrity, and availability. Advancements in computer networks, servers, and mobile devices have significantly increased Internet usage. The digital age offers many social and public business opportunities. Cyber security protects systems such as hardware, software, and data connected to the Internet from cyber threats. Individuals and businesses use the procedure to prevent unauthorized access to data centers and other computerized systems [1].

However, the amount of text information to be processed is enormous, and separating threats from the noise requires much routine work. They are more challenging to recognize in the spatial domain. However, with the proliferation of information on the Internet, the traditional signature and rule-based network security mechanisms are facing challenges. Cybersecurity issues are no longer just Information Technology (IT) issues. Compared to conventional security practices, the ability to monitor and secure assets is beyond manual controls. This is now a business risk that needs to be handled carefully at the highest level of the organization [2]. However, these opportunities come with risks such as cyber-attacks, data breaches, intellectual property loss, and financial fraud. Ransomware attacks, distributed denial-of-service attacks, phishing, web and mobile applications, exploits, and many other cyber-attacks can bankrupt financially, productivity and legal costs.

This study first collects a dataset called KDD Cup-99 to detect network penetration in predictive model intrusions or attacks. Second, data pre-processing can remove errors in the network and transform them into a prepared raw dataset. Thirdly, an intrusion detection system can improve the process using feature selection techniques. Finally, LDSV is considered a suitable classifier for cyber security that can be used to classify and detect cyber security.

*Corresponding author. Email: hisheelu@gmail.com

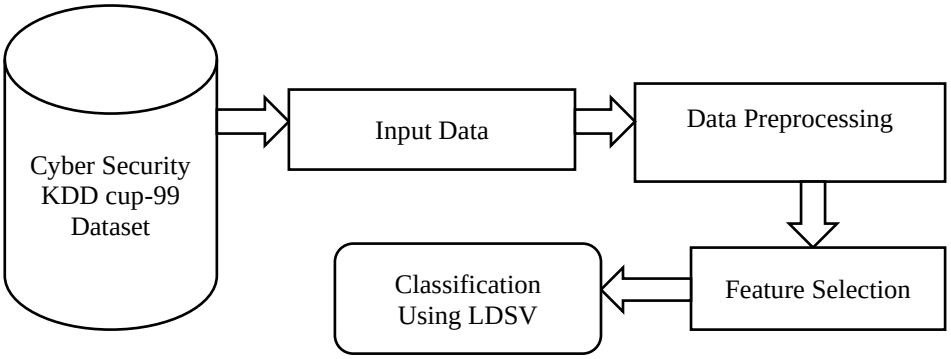

Fig 1. Basic Structure of Cyber Security

Fig 1 describes the basic structure of cyber security. First collect the KDD CUP99 Cyber dataset. The dataset can then be used to perform preprocessing on the input data. After that, the malicious link can be detected in the feature selection. Finally, detection for cyber security can be categorized.

2. LITERATURE SURVEY

Anna L et al, (2016):[1] The author proposed that the Machine Learning (ML) and data mining (DM) method is cyber analysis to support focused literature review intrusion detection. Each method's papers are identified and summarized based on emerging method relevance or citation count. A Lakshmanarao et al, (2020):[2] The author proposes that the potential for cyber security can be leveraged with sophisticated ML applications. ML algorithms are analyzed on the most common cyber security threats. Cybersecurity is considered one of the most challenging problems due to the effectiveness of ML and Deep Learning. G Apruzzese et al, (2018):[3] The author proposed that ML techniques can be used to detect and analyze intrusions, malware, and spam. These goals can be divided into two categories: assessing the current maturity in solutions and identifying fundamental limitations preventing the immediate adoption of ML web detection programs. H I Sarker et al, (2020):[4] The author proposed to build a security model using an intrusion detection tree ("IntruDTree") based on ML. The selected essential features can be used to construct a tree-based general intrusion detection model and verify the ranking according to the importance of the security features. D Dasgupta et al, (2020): [5]The author proposes that the most comprehensive methodology in ML is a survey of works such as basics of cyber-security and related defenses and basics of commonly used ML algorithms. Features, dimensionality reduction, and classification/detection techniques are essential in cyber security. Sarker IH et al, (2021): [6] The authors present a collection of relevant aspects of "cyberlearning," ML-based cyber security modeling, and detailed empirical analysis of the performance of various ML-based security models. Binary classification for anomaly detection and multi-class classification for multiple cyber-attack types should be observed models in CL modeling. Sarker IH et al, (2020): [7] The author briefly discusses the cyber security of data collection and data science in communications. Effective, secure solutions can be analyzed and meet data-driven patterns. Cybersecurity allows the computing process to become more efficient and more innovative than traditional ones. Al-Omari et al, (2021): [8] The author proposed an intelligent intrusion detection model that can be used to predict and detect attacks in cyberspace. Should consider the ranking of security features to design the models based on the decision tree. However, this system increases computational complexity and affects the data detection accuracy. Also, the dimensions and safety aspects pose a considerable problem.\Landauer M et al, (2020): [9] The author proposed the existing studies in grouping categories through clustering techniques using approaches, reviewing their applicability and limitations, discussing trends, and identifying gaps. However, log data consists of unstructured and heterogeneous messages collected from various sources, which makes automated analysis of log data difficult. Xi B et al, (2020):[10] The author proposes ML techniques against three main types of attacks. They are poisoning attacks, evasion attacks, and privacy attacks. Weaknesses and limitations in introducing security approaches are critical to existing interactions. Sarker IH et al, (2021):[11] The author proposed a comprehensive vision of "AI-Driven Cybersecurity" that is important in cybersecurity services and management. Security intelligence modeling is an automated system for calculating cyber security that can be intelligently developed based on conventional security and AI methods. Also, it helps to highlight and research several research directions within the scope of the study. Z. Zhang et al, (2022): [12] The author proposed an Explainable Artificial Intelligence (XAI) approach to cyber security applications in current literature. The rapid development of AI has led to widespread use in cyber security fields, including intrusion and malware detection and spam. Also, these are internet-connected systems, including ML and DL. A. Halbouni et al, (2022): [13] The author proposed Intrusion Detection Systems (IDS) using ML and DL to protect data from malicious behavior. Functional intrusion detection systems can be developed with network implementations, applications, algorithms, learning approaches, and datasets. M. A. Ferrag et al, (2022): [14] The author proposed to analyze IDS for Agriculture 4.0 cyber security. In particular, it offers cybersecurity threats and metrics for performance evaluation of agriculture 4.0 IDS. It provides a detailed classification of intrusion detection value for each

emerging technology. D. Gumusbas et al, (2021): [15] The authors proposed to provide a road map for readers who wish to understand the potential of DL methods in network security and intrusion detection systems. Also, analysis of benchmark datasets is used in the literature to train DL models.

2.1 Problem of Statement

- ❖ Detecting network collisions and attacks is a challenging task in modern network security.
- ❖ However, cyber security is increasing everywhere and exploiting vulnerabilities according to the computing environment.
- ❖ This further complicates the development of robust learning techniques because strong can withstand multiple attacks.
- ❖ The increasing use of the Internet has improved the volume and complexity of data, leading to the occurrence of Big Data.

3. PROPOSED METHODOLOGY

This section analyzes malicious or non-malicious cyber security attacks based on Logistic Decision Support Vector (LDSV). As per this analysis, the dataset is first collected, then pre-processed and analyzed in feature selection to analyze and classify it to get the right accuracy in detecting malicious or benign intrusions for cyber security.

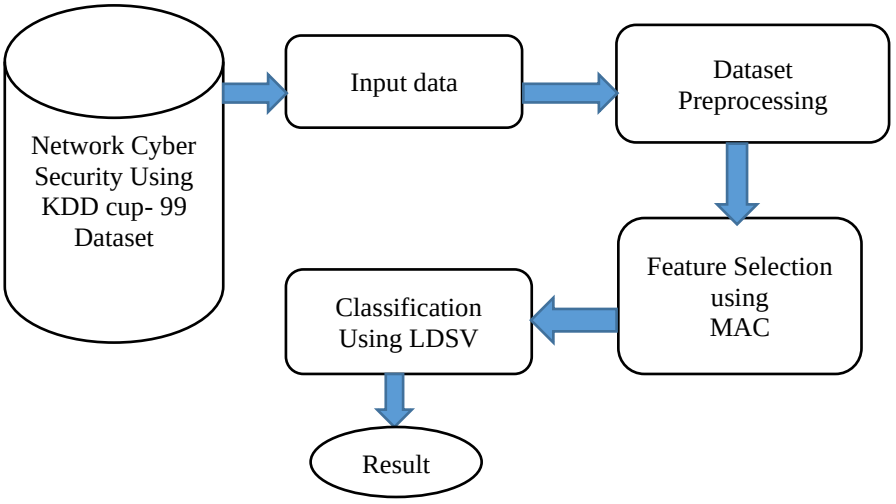

Fig.2 Cyber security Classify using the LDSV

Fig 2. First, collect the dataset from KDD CUP-99 Cyber Security. The dataset can then be used to perform preprocessing on the input data. Afterward, feature selection can be detected according to malicious and non-malicious intrusion attacks. Finally, the assessment for cyber security can be categorized.

3.1 Data collection

Cybersecurity datasets can be collected using the KDD Cup-99. Collections are classified according to the source of relevant information, such as unique parts of a data set that cannot classify according to a particular source. Use a dataset from cyber security as a predictive model to differentiate malicious links from malicious ones based on network intrusion attacks.

3.2 Preprocessing in cyber security

Data preprocessing techniques can be used to process data with ML algorithms to build a predictive model. Also, these techniques are generally used to increase the model's accuracy. The raw data should be transformed into a predefined format, and the data should be cleaned of errors. Anomalies of missing values are detected to indicate important information in the field. A better approach is to use min-max scaling to overcome this problem. However, the scale will be used to normalize data in contrasts. Equation 1 finds the minimum and maximum values,

$$a^* = \frac{a-min}{max-min}\,(new\_max - new\_min) + new\_min \qquad (1)$$

Let's assume a^* -normalized value, a- input data set from the dataset, and min and max –variable; the standardization technique involves making communal data value.

$$F^* = \frac{a^* - Mean}{c} \quad (2)$$

Equation 2 Standardize the value from the attribute. Where F^* - standardized value, a^*- Input dataset, Mean - represents the Mean attribute, c- standard deviation of the attribute. A standardization technique can create a dataset with zero mean and unit variance.

3.3 Feature selection using Mean Absolute on Chi-square test

Cyber security is developing predictive models to act in a limited number of ways against malicious intruders. Leads to cyber security by detecting malicious issues in feature datasets. This method is similar to the variance threshold method, but the mean is not a perfect square. From these, the absolute deviation from the mean value can be calculated. The chi-square method tests the relationship between categorical variables. The feature set can be used to detect malicious attacks on the cyber security.

Algorithm

Input: Standardize the data value F^*

Output: observed values.

Start

Step 1: Find the distance between two features

$$\frac{1}{x} = \sum_{j=1}^{x} |n_j - y| = \frac{|n_1 - y|\ |n_2 - y| + \cdots +\ |n_3 - y|}{x} \quad (3)$$

Step 2: Find the average data value.

$$M = \frac{\sum_{i=1}^{n} (a_i - \bar{a})}{n} \quad (4)$$

Step 3: First, the mean or center measure must be specified in m.

Step 4: Each data value is deviated from the data set by m and takes the difference between the data value.

Step 5: Calculate the feature data set. Where equation5 estimate the optimal features of cyber security.

$$a^2 = \sum_{j=1}^{x} \frac{(C_j - D_j)^2}{D_j} \quad (5)$$

Step 6: Return Observed value.

Stop

Let's assume, initial measure the data value in a feature and calculate the better result in an observation value in a class. Where a^2 – Calculate the between quality and best data set, c- Observation value, D- Expected value, j- class. n- The total number of data, m- Derivate the data values, J- Period of value. Let, a_i-performance value, $\bar{a}$ – Average value, x - number of data.

3.4 Classification using Logistic Decision Support Vector (LDSV)

Data classification is the process followed in cyber security to identify vulnerabilities, prevent intrusions and determine the effort and cost required to protect critical data properly. Classification is used to get more accuracy in cyber security by detecting malicious or benign problems. LDSV is a classification algorithm used to predict the probability of some classes based on the predicted dependent variable. This model can analyze the malicious or non-malicious network intrusion attacks for cyber security. The LDSV model calculates the output by short the input features. This method uses the LDSV model as an essential feature to detect cyber-attacks as they occur. With this classification, this type can be used to solve intruder's problems and improve quick accuracy.

Algorithm:

Input: Observed value a

Output: Optimized result

Start

Step 1: The logistic model also predicts the probability that a particular data value is present.

$$f(a) = \frac{1}{1+x^{(OF)}} \quad (7)$$

Step 2: Select the optimal value for the support vector.

$$x_{ab(u)} = \frac{\partial}{\partial u^{fa}} \frac{\partial}{\partial u^{fb}} \quad M(u, u_b) \quad (8)$$

Step 3: The quantitative is used to the data value.

$$R = [a_i, f] = \frac{G(a_i, f)}{sinfo(a_i, f)} \quad (9)$$

Step 4: Increase the distance between each data point

Step 5: Classifying test data points.

Step 6: Return optimize result

Stop

Let assume, for determining the training and test data, predicting the accuracy of the specified value presents the optimal data value, and each data point performs the classification of the increase distance between each data points and detect the quantitative data set. f (a) – Classify the value, a- number of value, x- represents the exponential function. Classify the optimal value. R- Ratio, G-Grade, $a_i - KDD\ dataset$, sinfo- split value, f- feature information.

4. RESULT AND DISCUSSION

The proposed implementation is tested in Python language using the publically available KDDcup99 dataset. Malicious or non-malicious network cyber security can be effectively detected by comparing the parameters like classification accuracy. Cyber security uses the KDD Cup99 dataset to detect network intrusions. Cyber security trained data 1000 and testing data 500.

TABLE I SIMULATION PARAMETER

Value	Parameter
Dataset Name	KDD cup-99
Number of records	494021
Training Data	1000
Testing Data	500
Language	Python
Simulation Tool	jupyter

Table 1 describes the simulation parameters tested on windows 10 OS and we use anaconda environment python language.

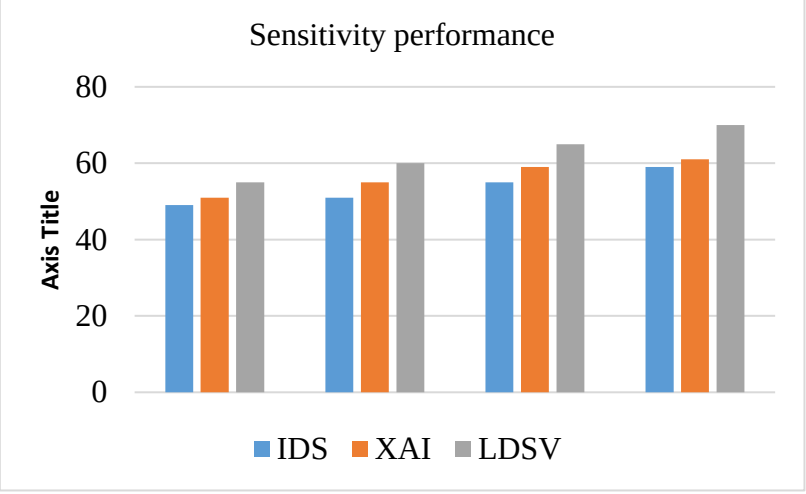

Fig 3. Model performance for Sensitivity

Fig 3 is a comparative result of the proposed sensitivity performance analysis approaches presented. Among them, the IDS has the lowest rating of 51%; XAI has a low rating of 62%. LDSV has a higher rating of 70% when comparing the two.

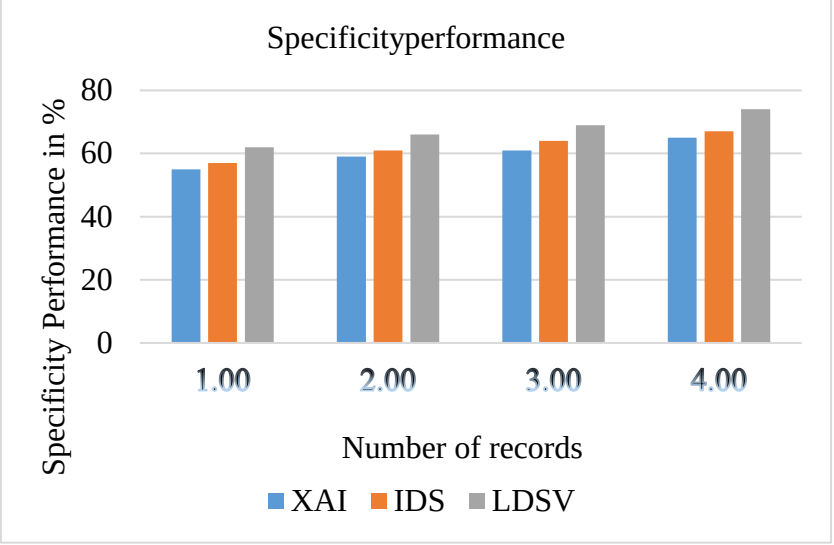

Fig 4 Performance model of Specificity

Fig. 4 is a comparative result of the proposed Specificity performance analysis approaches presented. Among them, the XAI has the lowest rating of 65%; IDS has a low rating of 67%. LDSV has a higher rating of 74% when comparing the two.

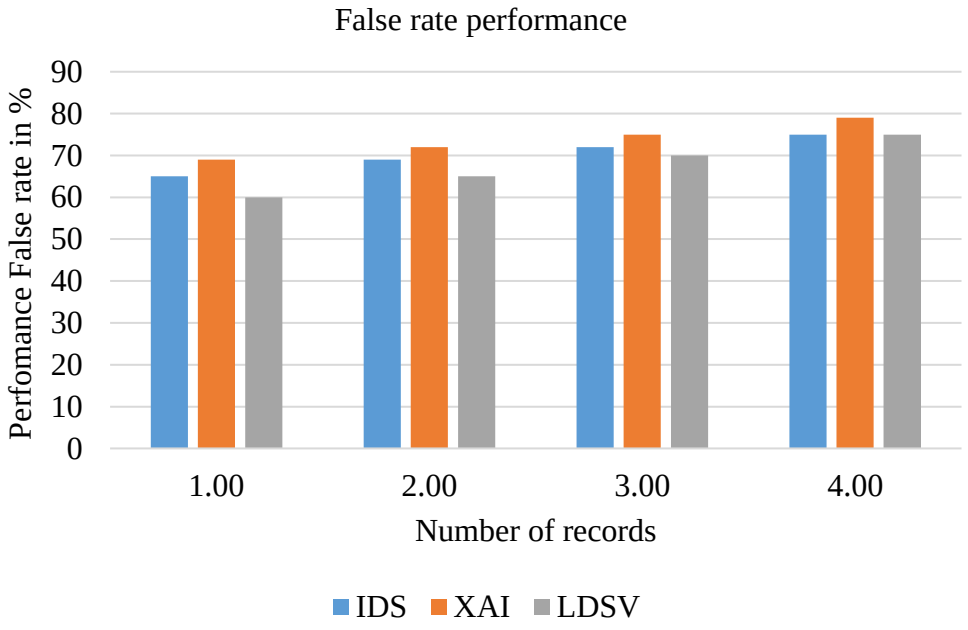

Fig 5. False Rate Performance

Fig. 5 is a comparative result of the proposed false rate performance and an analysis of the presented approaches. Among them, XAI and IDS have higher accuracy. LDSV has lower accuracy when compared to both.

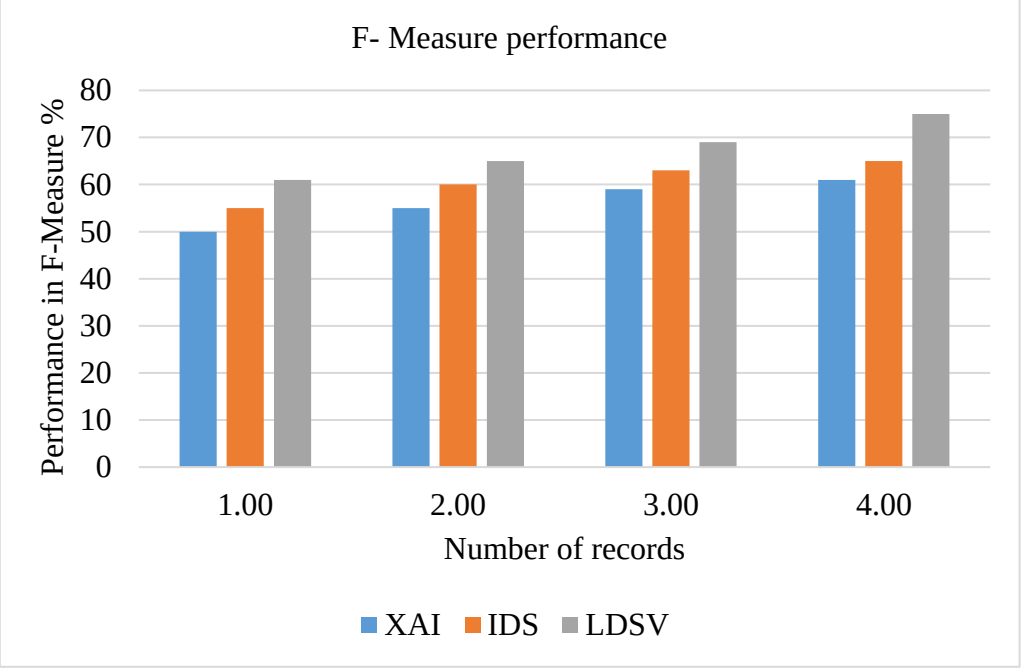

Fig. 6 F-Measure Performance

Fig. 6 the comparative results of the proposed F-Measure performance are an analysis of the presented approaches. Among them, XAI and IDS are the least accurate. LDSV has higher accuracy when compared to both.

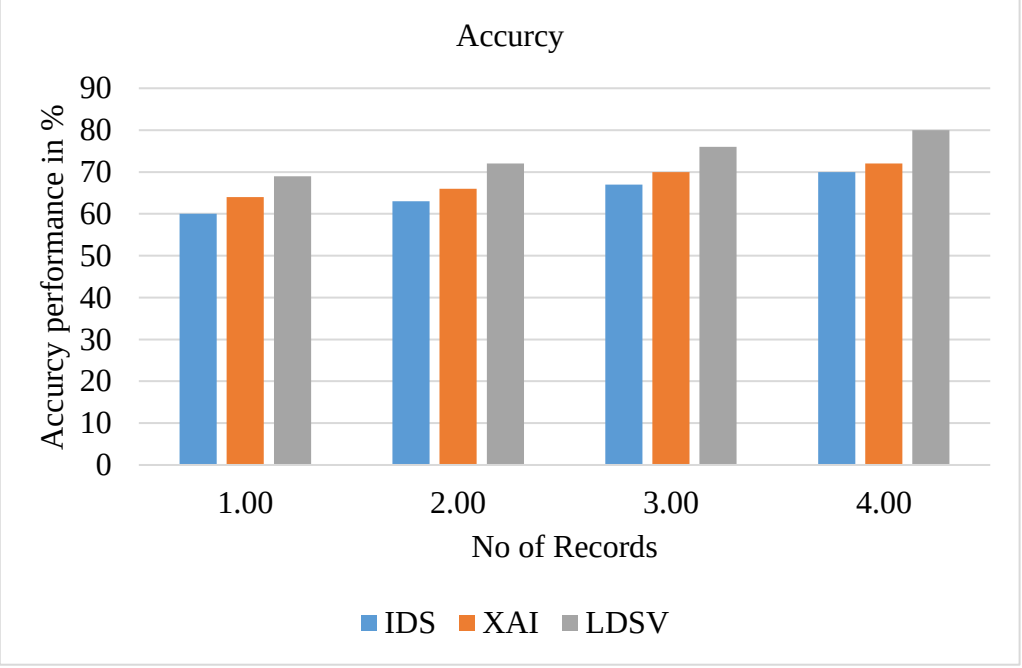

Fig 7. Performance of accuracy model

Fig. 7 compares the accuracy of the proposed analysis approaches, and the computational results are presented. XAI and IDS have lower accuracy. The proposed LDSV has 80% higher accuracy compared to previous methods.

5. CONCLUSION

Cyber security requires gathering complete and accurate knowledge of malicious intent, where relevant information is challenging. So to know whether malicious or malicious intrusion detection in cyber security, firstly, we need to use the KDD cup99 dataset. Cyber-security can be detected at an early stage. Next, check Malware Detection in Cyber Security to detect malicious attacks in the feature selection. Furthermore, proposed a new classification called LDSV to eliminate cyber security network intrusion attacks. Used to detect and protect against malicious or benign network intrusions and cyber-attacks. A high accuracy 70% was obtained when analyzing cyber security-related sensitivity, and a high of 74% when analyzing specific performance. A new classifier called LDSV achieves a high accuracy rate of up to 80% and prevents cyber-security from occurring.

REFERENCE

[1]. L. Buczak and Erhan Guven. “A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection.” In: IEEE Communications Surveys Tutorials 18.2 (2016), pp. 1153– 1176. doi: 10.1109/COMST.2015.2494502.

[2]. A Lakshmanarao, M Shashi, A survey on machine learning for cyber security international Journal of Scientific & Technology Research, volume 9, issue 1, p. 499 – 502 Posted: 2020

[3]. G Apruzzese, L Ferretti, M Marchetti, M Colajanni, A Guido, On the Effectiveness of Machine and Deep Learning for Cyber Security, p. 1 – 17 Posted: 2018

[4]. H I Sarker, B Y Abushark, F Alsolami, I. A Khan IntruDTree: A machine learning based cyber security intrusion detection model Symmetry, volume 12, issue 754, p. 1 – 15 Posted: 2020

[5]. D Dasgupta, Z Akhtar, S Sen Machine learning in cybersecurity: a comprehensive survey The Journal of Defense Modeling & Simulation, p. 1 – 50 Posted: 2020

[6]. Sarker IH (2021) Cyberlearning: effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. Internet Things 14:100393

[7]. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A (2020) Cybersecurity data science: an overview from a machine learning perspective. J Big Data 7(1):1–29

[8]. Al-Omari M, Rawashdeh M, Qutaishat F, Alshira’H M, Ababneh N (2021) An intelligent tree-based intrusion detection model for cyber security. J Netw Syst Manag 29(2):1–18

[9]. Landauer M, Skopik F, Wurzenberger M, Rauber A (2020) System log clustering approaches for cyber security applications: a survey. Comput Secur 92:101739

[10]. Xi B (2020) Adversarial machine learning for cybersecurity and computer vision: current developments and challenges. Wiley Interdiscip Rev Comput Stat 12(5):e1511

[11]. Sarker IH, Hasan Furhad M, Nowrozy Ra (2021) AI-driven cybersecurity: an overview, security intelligence modeling, and research directions. SN Comput Sci 2(3):1–18

[12]. Z. Zhang, H. A. Hamadi, E. Damiani, C. Y. Yeun and F. Taher, "Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research," in IEEE Access, vol. 10, pp. 93104-93139, 2022, doi: 10.1109/ACCESS.2022.3204051.

[13]. A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi and R. Ahmad, "Machine Learning and Deep Learning Approaches for CyberSecurity: A Review," in IEEE Access, vol. 10, pp. 19572-19585, 2022, doi: 10.1109/ACCESS.2022.3151248.

[14]. M. A. Ferrag, L. Shu, O. Friha and X. Yang, "Cyber Security Intrusion Detection for Agriculture 4.0: Machine Learning-Based Solutions, Datasets, and Future Directions," in IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 3, pp. 407-436, March 2022, doi: 10.1109/JAS.2021.1004344.

[15]. D. Gumusbas, T. Yldrm, A. Genovese and F. Scotti, "A comprehensive survey of databases and deep learning methods for cybersecurity and intrusion detection systems", *IEEE Syst. J.*, vol. 15, no. 2, pp. 1717-1731, Jun. 2021.