JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 02
2024
Research Article
https://doi.org/10.56979/702/2024

Anomaly Detection using Clustering (K-Means with DBSCAN) and SMO

Umair Rashid[1], Muhammad Faheem Saleem[1], Saad Rasool[1], Ahmad Abdullah[1*], Hira Mustafa[1], and Aiman Iqbal[1]

[1]Department of Computer Science, Bahauddin Zakariya University, Multan, 60000, Pakistan.
[*]Corresponding Author: Ahmad Abdullah. Email: ahmadabdullah6@gmail.com

Received: May 23, 2024 **Accepted:** August 13, 2024 **Published:** September 01, 2024

Abstract: In recent times, AI has become a useful tool for describing the properties of information because it can support the Data Mining (DM) procedure by analysing data for identifying patterns or routines. Anomaly detection is of vital importance in DM that helps in the discovery of hidden behaviour within the most vulnerable data. It also aids in the detection of network intrusion. This research proposed a model for detecting anomalies using machine learning (ML) techniques. By leveraging ML, the model can achieve higher detection rates and reduce the number of false positives, resulting in an overall improvement in intrusion classification. This study evaluated a proposed hybrid ML technology using dataset of Network Security Knowledge and Data Discovery. In this study, we used K-means and Density-Based Clustering Algorithm for clustering and Sequential Minimal Optimization for classification purposes. By putting the suggested method for detecting anomalies to test, it is demonstrated by the findings that this hybrid model can increase positive detection rate and anomaly detection accuracy while decreasing rate of false-positives. The proposed algorithm showed superior performance compared with recent closely related studies using similar variables and environments. This algorithm achieved lower false alarm probability (FAP) and high accuracy. This is due to the hybrid nature of producing an optimal detectors quantity that exhibit high accuracy and low FAP. The required time will decrease if the given false alarm probability is small for pre-processing and processing when compared to other algorithms.

Keywords: Network Security; Anomalies Detection; Sequential Minimal Optimization; Data Mining; Hybrid Model; K-Mean; Density-Based Clustering Algorithm.

1. Introduction

As a result of the Internet's complete accessibility to users, computer networks are more susceptible to intrusion and information-disclosing exploits. Recently, network attacks have grown increasingly sophisticated and difficult to detect. According to the report of statistics cited in the Symantec Global Internet Security Threat report, the number of attacks is at an all-time high and is rapidly rising. Given that our reliance on data is growing at an exponential rate, we need algorithms to protect data for keeping confidentiality, security, and accessibility. So, detecting approaches have been limited to identifying previous attacks, whereas algorithms for detecting anomalies have the capability to identify unknown attacks based on how user acts. Speed and efficiency are two of the most important things to think about with anomaly detection [1]. A rapid, complex ID algorithm prior to the attack can be nearly intolerable if there is a lot of data on the network. A lot of recent methods improve IDS rates, although at the expense of extensive time and energy expenditures (in the form of communication, memory, or some other system requisite). If the flow is changed in real time, these problems could become even harder to fix.

A defense-in-depth approach is a fundamental aspect of the safe structure, implementing various security strategies to prevent, identify, contain, and resolve potential attacks. Access control, multi-factor authentication, or data encryption are some of the security mechanisms deployed as a first line of defense to possible threats [2]. Implementing protective measures and technologies such as Intrusion detection systems, firewalls, and antivirus software to facilitate tracking of network systems in order to identify,

mitigate, and respond to potentially malicious activity [3]. In the field of networks security, intrusion detection systems are utilized to keep track of traffic on the network and detect abnormal behavior within the system. This type of approach is widely recognized as important in maintaining network security [4]. Using ML or statistical techniques, it is possible to develop a competent intrusion detection system for network protection [5].

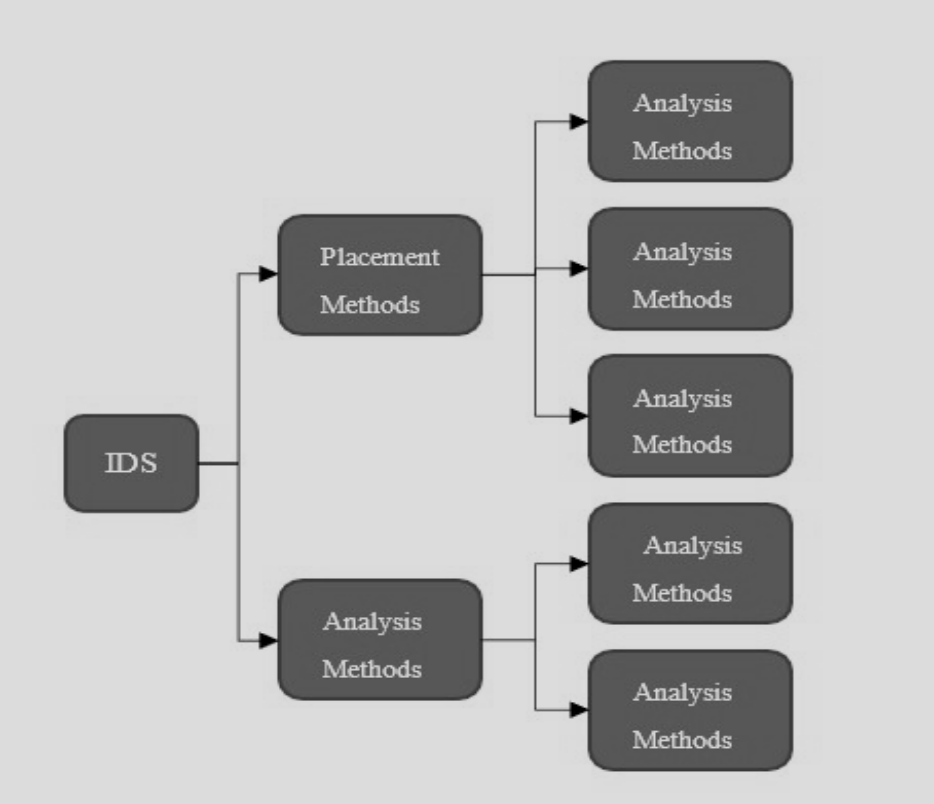

Figure 1. Categorization of Intrusion Detection System.

As a result of advances in network technology, individuals are increasingly relying on network resources to collect data over the Internet. Computer networks and data are subject to several risks. As a result, computer network and data protection defenses are constantly being created to ensure data integrity. The main function of IDS is protection. An efficient IDS protects computers and information systems from potential attacks and helps detect unauthorized access. System anomaly detection protects against potential new and zero-day attacks by detecting anomalies according to usage statistics.

For classifying user's behavior as normal or abnormal, numerous approaches of DM are used. Our study is focused on mitigating constraints that DM approaches have and enhancing the detection rate. Previous studies have shown a number of data mining (DM) strategies based on user behavior classification, but there are significant limitations that this study addresses for decreasing them as well as increasing detection rate. Limitations of the detection technique are:

The main challenges for anomaly detection systems are efficiency and speed. One problem with complex detection algorithms is that they cannot be used fast enough when there is a lot of network traffic. Many theories have been put forward for advanced algorithms based on high detection rates, even though it's highly challenging.

If important data for feature selection during classification is lacking, precise or valid results won't be formed. Real-time traffic analysis is among the major weaknesses of IDS. Flaws in the detection of traffic can leave computer networks vulnerable to attacks.

In this study, a new hybrid method of Intrusion Detection System is presented. This hybrid model uses SMO and K-Mean with DBSCAN to acquire clustering and classification [7]. We combine Sequential Minimal Optimization and K-Mean with DBSCAN methods for improving Intrusion Detection System efficiency, properly identifying novel intrusions, as well as enhancing accuracy of detecting anomalies [6]. It also uses attack patterns to reduce the number of real-time false positives.

2. Literature Review

Previous research offered numerous DM strategies based on user behavior classification, however they have certain limitations. This study proposes strategies to overcome these challenges and enhance mining accuracy. The following sections highlight some of the primary limits and issues associated with detection approaches.

Real-time traffic monitoring is significant drawbacks of algorithms of Intrusion Detection. The computer system can become vulnerable to attack if detection pf real time traffic is faulty.

The key difficulties in anomaly detection systems are efficiency and speed. One of the issues is connected to network traffic volume, because complicated detecting algorithms are utilized at a sufficient speed when there is a large traffic.

Without critical data for feature selection, precise or valid results won't be formed. This portion analyses an evaluation for several ID models and procedures in light of the above limits and obstacles. It also discusses the methodology used to create the IDS as well as the most recent updated models. C. Taylor et al. [8] presented a NATE technique, powered by clustering and multiple factor evaluation. It is easier for IDS to deal with monitoring limits and traffic from big data due to NATE [9]. Furthermore, NATE facilitates limited attack scope performance characteristics and identification of anomalies, while also reducing the amount of network traffic monitoring [10]. The NATE process is divided into two phases: data gathering or observation of potential threats, and identification of attacks [11]. The suggested research demonstrates clustered data collection method which makes it easy to update the database in real time with new attack features [12].

Amen, B. et al. [10] proposed a P-BEST framework for detecting exploit attacks and developing new signing techniques. Production-Based Expert System is able to give effective Intrusion Detection System evaluation [11]. The suggested approach integrates with C programming to provide flexibility and usability. Nonetheless, its ability to detect intrusions and attacks is limited when presented with incomplete or uncertain data, or when operating in an unknown environment [13].

Chen, Z. et al. [14] proposed an approach to develop Intrusion Detection System algorithms using DM. Suggested approach allows for programmed application for Intrusion Detection System algorithms [15]. Inductively learned computations regarding applicable systematic aspects, unprocessed audited information manipulation, and system-dropped information get combined to connections record or features [16], which are important for the functioning of the DM framework tiers. Associations rule as well as frequents-episode [17] are used in this strategy.

M. Saeed et al. [18] proposed an approach of decision tree to combine various hosts-specific detecting devices. Identification metrics and a decision tree support the proposal. The metrics form the backbone of IDS models [19]. The statistical rule-based method [20] is used to carry out the modelling techniques.

T. Minegishi et al. [21] proposed a data mining (DM) approach for developing IDS models. The suggested approach allows for programmed application for Intrusion Detection System algorithms [18]. Inductively learned computations regarding applicable systematic aspects, unprocessed audited information manipulation, and system-dropped information get summed up in connections record or features [18].

Daniel, B. et al. [22] proposed how a decision tree can be used to combine different hosts-specific detecting devices. The presented concept is based over metrics alongside decision tree to find intrusions. These metrics form the basis for Intrusion Detection System modelling. Modelling measurements are performed according to statistical rule-based methods.

Zhang, J. et al. [24] presented anomaly detection method for NIDS in another research. This study used a hybrid approach for detecting anomalies in an effort to reduce the limitations of each technique when regarded separately. The suggested hybrid detection method is tested as a way to find intrusions in the random dataset using data mining.

P. Yuhuai et al. [24] proposed an approach for making decision tree working better in their study. Zhang, P. et al. [25] demonstrated how performance of anomalies identification may improve through AI. The authors evaluated the C4.5 algorithm [26] using the K-mean as well as Euclidean distance for training instances. It is demonstrated by the findings that semi-supervised algorithm outperforms the supervised algorithm and unsupervised training algorithm.

Vynokurova, O. et al. [27] developed a highly efficient intrusions detection method using fuzzy logic. They presented a setup that permits the recognition of network intrusion behavior. It employs a mechanical process to generate fuzzy rules from repeating elements in specified rules [28].

Raymond T. et al. [29] presented back-propagation along with C4.5 algorithms. Apart from addressing known threats, these algorithms are primarily employed for identifying instances of misuse and assessing the degree of anomalies in typical profiles. One potential avenue for exploration is the utilization of supervised machine learning algorithms [12]. Neural Networks (NN) is a good way to find known attacks according to findings, but that using decision tree is a better and more interesting way to find novel intrusions.

Mutanov, G. et al. [31] presented machine learning approach for network identification that combines K-mean with SVM. The proposed study reduces false positive and false negative alarms while also improving detection rates. This study uses NSL-KDD dataset and SVM [32]. It obtains a high accuracy of detecting anomalies and low rate of false alarms.

We proposed a hybrid ML method for the recognition of anomalies in this paper. Our primary goal is to enhance the efficiency and accuracy for detecting anomalies and decrease the rate of false alarms. We use a dataset named NSL-KDD to evaluate K-Mean with DBSCAN and Sequential Minimal Optimization.

3. Anomalies Detection Method

To find an appropriate approach for detection of anomalies in the network, the concept of normality is first adopted. The normality concept is related to the formally defined framework which explains connection amongst system variables. It quantifies the extent to which the identification of anomalies causes system's performance to deviate from a model of normal conditions. The approach we used is based on anomalies identification approach centered K-mean with DBSCAN and SMO. In a computer network, identification approach has been assessed in order to create a sufficient quantity of detection devices.

The presented approach ambitions to lessen the wide variety of capabilities with the aid of using the usage of function choice algorithms with inside the preprocessing phase. Specific trends are decided on from the dataset with the aid of using making use of a steady subset degree and genetic seek algorithm. This choice approach eliminates beside the point capabilities earlier than the cluster system and type technique, observed with the aid of using the K-means with DBSCAN clustering system technique. It reduces training, processing time, and dataset complexity. The type technique improves reputation excellent primarily based totally on SMO.

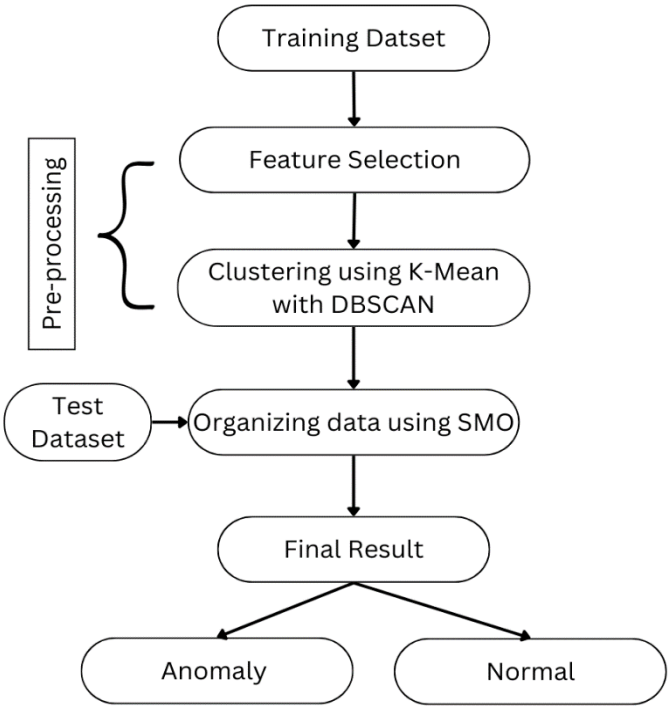

Figure 2. Anomalies Detection Process.

Figure 2 shows the trained model that takes into account both typical and unusual data, assuming that the latter does not occur as often as the former. The data used for training can be either normal or abnormal. Attack data is probably not as prevalent as ordinary data. Given that intrusion-based attackers frequently use aggregate samples to monitor traffic, and that larger samples allow for more effective incursions, this assumption is valid. Consequently, the percentage of data that is considered abnormal is less than x, where x represents the probability of assaults and incursions. The training data is prepared for classification and the data paradox is reduced through preprocessing on the original NSL-KDD dataset. The identification engine below also receives accurate data from it. The feature selection and cluster construction-based K-means algorithm determines the degree of preprocessing. This step contributes to network data cleaning by capturing and processing missing or inadequate properties, which are subsequently preprocessed in the next phase.

3.1. Feature Selection

Among the most important pre-processing procedures in hybrid models is features selection. This is due to the importance of the selection outcome in Machine Learning procedure as well as K-means and DBSCAN clustering formulations. It therefore has an impact on the overall efficiency of the Data Mining technique for the Intrusion Detection approach. As a whole, the high dimensional features' variables of input are categorized.

Certain characteristics are not evaluated in categorization and are therefore irrelevant to the mixed approach. If bad, duplicated, or messy information is not filtered out earlier, it might disrupt your Machine Learning workflow. Useless information within dataset might raise model complexity and training time, lowering the training algorithm's effectiveness. As a result, deleting these attributes enhances the effectiveness and improves overall Intrusion Detection System functioning. It also aids in the detecting procedure, as well as the accuracy of findings or safety as a whole. As a result, these disparate datasets must be discovered and processed. Several approaches (e.g., pairwise attribute approaches, PANDA) have been named for identifying features that are not relevant. PANDA's ability to function with no specialized expertise makes it helpful [33]. Real-world data flow is frequently affected by a variety of circumstances.

Certain crucial elements include noise, flawed and redundant information etc. The non-negotiable problem is a data corruption which occurs when data is considerably harmed throughout the process of collecting and preparing data. Errors come in two different forms. Both implicit noise and random error are provided by acquisition services such as sensors. Actual data that has been mishandled or erroneously assigned at the incorrect hour is referred to as improper data. The algorithm's performance and effectiveness are strongly influenced by the reliability of training data and potency in relation to categorization mistake. Figure 3 demonstrates this procedure.

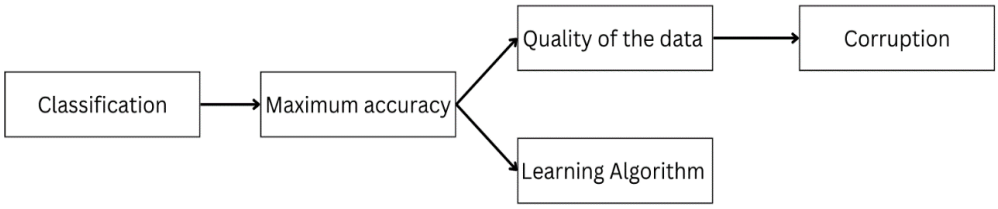

Figure 3. Anomalies Detection Approach using hybrid ML Algorithm.

Your data might contain irrelevant, redundant, or disruptive information that can be filtered away. Duplicate and unnecessary attributes might enter the Machine Learning method as noisy or damaged data and cause it to fail in such a situation. The number of features has been reduced by features selection and eliminate any feature that is ineffective, duplicate or has noise. it improves the overall performance of system and speed up the data mining approaches and enhance accuracy.

Figure 4 depicts the features selection process. Given the feature set (F0 to FN) from the original dataset obtained by traffic across networks, the feature gathered by using a tool (i.e., MATLAB's Feature Selection Toolbox (FST)) is (F0 to FM), as shown in Figure 4. The numbers of features obtained is determined by the selection tool along with correlation. The following level covers the fundamental philosophy of feature analyses.

Figure 4. FST Tool for Features Selection.

Following above phase, the identification engine is given the data set, which now comprises of useful attributes. Feature Selection increase the rate of anomalies identification as well as decrease the false alarms in Intrusion detection System. The Machine Learning tool utilized in this work to compute features subset determination for mixed Data Mining schemas is the Waikato Environment of Knowledge Analyses. By categorizing the test yield for every function subset, this has been accomplished. Prior to clusters creation and classification, some attributes are chosen and non-relevant features are eliminated using genetic searching strategies.

3.2. Clustering Phase

The k-mean clustering formation method is used during the cluster formation step. The 3 clusters have been developed plus evaluated. The strategy iteratively trains the data. Cluster formations are passed down through another. Cluster changes alter center values and have an impact on current clustering elements. Once values cease modifying, clustering formation will establish. Following that, cluster model gets constructed. Following this, DBSCAN algorithms has been applied to the dataset, taking clusters obtained from K-Means as input. DBSCAN discovers densely connected core points and expands clusters based on their density. We either group every data point into a specific cluster or flag it as an outlier. We use clustering results from DBSCAN and mark outliers as potential anomalies.

During the test phase, networking traffic created by collecting data has been examined. The last level belongs to classification, which is wherein it's controlled. In other words, the Sequential Minimal Optimization method has been used to determine if the data is normal or abnormal.

High-traffic IoT along with WSN domains are two excellent implementations for the suggested technique. The strength of suggested model is enhanced precision of detecting intrusions using ML. The suggested approach is intended to process and handle massive quantities of traffic while maintaining functionality.

4. Results

The Waikato environment has been used to create the WEKA (Experiment for Knowledge Analysis). WEKA is useful tool in machine learning alongside data mining. It was designed in 1997. It [41] serves as graphical user interface (GUI) that combines Data Mining and Machine Learning methods.

A total of five methods are there for locating regulations of associations. There are six graphical user interfaces of WEKA. The attribute relation file format is used to save data. There is also an installation utility supplied. WEKA offers a variety of boards for running certain processes. It may also modify and integrated with novel Machine Learning methods. The below metrics are used to detect attacks: (A) the number of false alarms equals the identified intrusions; however, such intrusions are the norm. (B) The number of false negatives reflects the typical cases discovered; however, such attacks represent real intrusions and are victim of Intrusion Detection System. (C) The total identified intrusions that are actually attacks is referred to as True Positives (TP). (D) The total true negatives (TN) discovered is related to the total of cases that are norm.

The following measuring metrics have been adopted: rate of detection, FPR's, and accuracy. DTR has been described by the proportion of identified to total intrusions and can obtained by following equation [34]:

$$DTR = \frac{TP}{TP+FN} \times 100 \quad (1)$$

FPR is the ratio of false alarms to the total number of attacks. FPR can be determined by the following equation:

$$FPR = \frac{FP}{TN+FP} \times 100 \quad (2)$$

Accuracy can be determined by the equation given below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3)$$

Table 1. Confusion Matrix.

Clustering Techniques	Predicted: Abnormal	Predicted: Normal
Actual: Abnormal	True Positive	False Negative
entry 2	False Positives	True Negative

4.1. Hybrid Model Clustering

The hybrid model that is the combination of SMO and K-Means with DBSCAN, with 22 selected attributes applied on a dataset in WEKA. Table 2 depicts the accuracy, and Table 3 shows the confusion matrix.

Table 2. Accuracy of Hybrid Model.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.988	0.052	0.972	0.988	0.98	0.968	Normal
0.948	0.012	0.978	0.948	0.963	0.968	Anomaly
0.974	0.037	0.974	0.974	0.974	0.968	Weighted Avg

Table 3. Confusion Matrix of Hybrid Model

a	b	Classified as
TN = 14285	FP = 177	a = normal
FN = 418	TP = 7664	b = abnormal

Table 4. Measurement Metrics of Hybrid Model.

Algorithm	DTR	FPR	Accuracy
Hybrid K-Mean and Sequence Mining Optimization	94.47	1.2	97.3695

Firstly, we applied the K-Mean algorithm on dataset to divide the dataset into K clusters. It assigns each data point to its closest centroid. Afterwards, we applied the DBSCAN algorithm to the data using the K-means clusters as inputs. DBSCAN discovers densely connected core points and expands clusters based on their density. We either group every data point into a specific cluster or flag it as an outlier. We used the clustering results from DBSCAN to identify anomalies. Potential anomalies are identified from the data points that DBSCAN identifies as outliers. Based on the clustering results, these data points either do not belong to any cluster or are outside of dense regions. Using the SMO approach, we train an SVM classifier on the labelled data (normal and potential anomaly points) generated in the preceding phases. During training, the SVM was taught to differentiate between normal and anomalous data points. Use the trained SVM classifier to determine if the remaining unlabeled data points are normal or abnormal. Each data point is assigned a class label by the SVM based on its learned decision boundary and proximity to support vectors.

The presented method enhances the detection rate while reduces the false alarms. K-mean makes categorization better and DBSCAN identifies potential anomalies whereas Sequential Minimal Optimization enhances the identification accuracy through dropping attributes. The hybrid model outperforms the individual models in terms of false alarms.

It is clear from the above results that Sequential Minimal Optimization combined with the K-Means with DBSCAN algorithms obtained successful outcomes. In this way, likelihood of detecting anomalies also increased. In this way, likelihood of detecting anomalies also increased. The major objective in the majority of cases was to maximize the likelihood of anomalies identification, although false alarm probability also got greatly decreased, maximizing detecting performance.

4.2. Discussion

Both K-means and DBSCAN algorithms rely on parameter values, such as the minimum number of points and neighborhood radius for DBSCAN and the number of clusters (K) for K-means. It can be difficult

to determine the best parameter values, and different parameter settings may provide varied clustering outcomes, which would then alter the performance of anomaly detection.

The combined technique makes the assumption that clusters contain anomalies or can be used to identify outliers. However, some anomalies might not follow standard clustering patterns and might be challenging to find using clustering algorithms. The clustering-based approach might not be able to fully catch anomalies that are located in dense areas or have intricate patterns.

The fundamental distribution of the data can have an impact on how well clustering methods work and how well anomaly identification works after that. Both K-means and DBSCAN may encounter difficulties if the data distribution is extremely skewed, has overlapping clusters, or has changing cluster densities. In such situations, the combination of these algorithms can have trouble correctly identifying anomalies. The initial centroids chosen for the cluster are important for K-means clustering. The final cluster assignment resulting from differing initializations may affect the detection of abnormalities. Poor initial centroids or random initialization could lead to subpar clustering outcomes, which would then affect the efficacy of anomaly detection.

5. Conclusion and Future Work

The hybrid technique based on K-mean with DBSCAN clustering and Sequential Minimal Optimization categorization is suggested in this paper. The technique focuses on problems which occur within setting related to huge datasets. SMO improves a data set by using features selection during the pre-processing step. The consistent subset level approach alongside genetics search approach has been implemented to reduce irrelevant attributes from the dataset.

To boost detection rate, a supervised classification algorithm SMO has used. Sequential Minimal Optimization with K-means and DSBSCAN and other similar methods were used to benchmark contribution techniques. It has been demonstrated by the findings that the presented model performs better than recent and close work (i.e., ADAM, NATE, P-BEST) by utilizing same variables. The suggested technique is capable of being used to detect anomalies for future Data Mining networks wherein network computational speed will probably be significantly lowered. The general technique generates a sufficient number of detection devices alongside adequate precision and negligible false alarms. Because of the reduced false alarms, it's has been strongly anticipated that the time for pre-processing and processing would be reduced.

Yet there are several issues that must be addressed, such as detecting patterns as well as abnormalities in big dataset in actual time while having a nearly infinite numbers of factors and computing capacity. This will need substantial study as well as advancement. The use of K-Means clustering with SMO classification for detecting anomalies results in substantial effectiveness alongside fastness. Meanwhile, as the quantity of live traffic over networks grow, high-speed computation becomes necessary to maintain satisfactory efficiency, particularly for continuous monitoring.

As data becomes more important in businesses, a single disruption to corporate data can result in major failures and excessive expenditures. Further studies on machine learning data anomaly detection will focus on preemptive scheme instead of reacting approaches. Anomalies can be identified almost instantly. As a result, Machine Learning approaches offer the enormous promise by near future.

In order to guarantee its excellent results, the presented technique can be further tested in subsequent research. Furthermore, numerous different features selection methods may be employed to choose better essential features in order to improve the system's efficiency.

References

1. Joseph, M.V. Significance of data warehousing and data mining in business applications. *Int. J. Soft Comput. Eng.* 2013, *1*, 329–333.
2. Tellis, V.M.; Souza, D.J.D. Detecting anomalies in data stream using efficient techniques: A review. In Proceedings of the 2018 International Conference on Control, Power, Communication and Computing Technologies (ICCPCCT), Kannur, India, 23–24 March 2018; pp. 296–298.
3. Zhang, L.; Chen, Y.; Liao, S. Algorithm optimization of anomaly detection based on data mining. In Proceedings of the 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, China, 10–11 February 2018; pp. 402–404.
4. Xie, J.; Wu, D.; Liao, T. Method of anomaly detection of temperature data in vacuum thermal test based on data mining. In Proceedings of the Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC), Harbin, China, 19–21 July 2018; pp. 1040–1045.
5. Cai, S.; Sun, R.; Hao, S.; Li, S.; Yuan, G. An efficient outlier detection approach on weighted data stream based on minimal rare pattern mining. *China Commun.* 2019, *16*, 83–99.
6. Ali, E.S.; Hasan, M.K.; Hassan, R.; Saeed, R.A.; Hassan, M.B.; Islam, S.; Nafi, N.S.; Bevinakoppa, S. Machine Learning Technologies for Secure Vehicular Communication in Internet of Vehicles: Recent Advances and Applications. *Secur. Commun. Netw.* 2021, *2021*, 8868355.
7. Yang, Z.; Ding, W.; Zhang, Z.; Li, H.; Zhang, M.; Liu, C. A Service selection framework for anomaly detection in IoT stream data. In Proceedings of the International Conference on Service Science (ICSS), Xining, China, 24–26 August 2020; pp. 155–161.
8. Sun, W.; Zhang, G.; Zhang, X.; Zhang, X.; Ge, N. Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy. *Multimed. Tools Appl.* 2021, *80*, 30803– 30816.
9. Nurelmadina, N.; Hasan, M.K.; Memon, I.; Saeed, R.A.; Zainol Ariffin, K.A.; Ali, E.S.; Mokhtar, R.A.; Islam, S.; Hossain, E.; Hassan, M.A. A Systematic Review on Cognitive Radio in Low Power Wide Area Network for Industrial IoT Applications. *Sustainability* 2021, *13*, 338.
10. Amen, B.; Grigoris, A. A Theoretical study of anomaly detection in big data distributed static and stream analytics. In Proceedings of the IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; pp. 1177–1182.
11. Cao, N.; Lin, C.; Zhu, Q.; Lin, Y.R.; Teng, X.; Wen, X. Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data. *IEEE Trans. Vis. Comput. Graph.* 2018, *24*, 23–33.
12. Guezzaz, A.; Asimi, Y.; Azrour, M.; Asimi, A. Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection. *Big Data Min. Anal.* 2021, *4*, 18–24.
13. Zhao, Z.; Zhang, Y.; Zhu, X.; Zuo, J. Research on time series anomaly detection algorithm and application. In Proceedings of the IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 20–22 December 2019; pp. 16–20.
14. Iqbal, Z., Imran, A., Yasin, A., & Alvi, A. (2022). Denial of service (DoS) defenses against adversarial attacks in IoT smart home networks using machine learning methods. NUST Journal of Engineering Sciences, 15(1), 18-25.
15. Chen, Z.; Yu, X.; Ling, Y.; Song, B.; Quan, W.; Hu, X.; Yan, E. Correlated anomaly detection from large streaming data. In Proceedings of the IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 982–992.
16. Ergen, T.; Kerpiççi, M. A novel anomaly detection approach based on neural networks. In Proceedings of the 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; pp. 1–4.
17. Lee, J.; Park, S. Mobile memory management system based on user's application usage patterns. *Comput. Mater. Contin.* 2021, *68*, 4031–4050.
18. Mei, L.; Zhang, F. A Novel distributed anomaly detection algorithm for lowdensity data. In Proceedings of the IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dalian, China, 25–27 August 2020; pp. 197–201.
19. Saeed, M.M.; Saeed, R.A.; Saeid, E. Identity division multiplexing based location preserve in 5G. In Proceedings of the International Conference of Technology, Science and Administration (ICTSA), Taiz, Yemen, 22–24 March 2021; pp. 1–6.

20. Elfahal, M.O.; Mustafa, M.; Mustafa, M.E.; Saeed, R.A. A framework for Sudanese Arabic—English mixed speech processing. In Proceedings of the International Conference on Computing and Information Technology (ICCIT1441), Tabuk, Saudi Arabia, 9–10 September 2020; pp. 1–6.
21. Provotar, O.I.; Linder, Y.M.; Veres, M.M. Unsupervised Anomaly detection in time series using LSTM-based autoencoders. In Proceedings of the IEEE International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 18–20 December 2019; pp. 513–517.
22. Minegishi, T.; Niimi, A. Detection of fraud use of credit card by extended VFDT. In Proceedings of the World Congress on Internet Security (WorldCIS-2011), London, UK, 21–23 February 2011; pp. 152–159.
23. Barbará, D.; Couto, J.; Jajodia, S.; Wu, N. ADAM: A testbed for exploring the use of data mining in intrusion detection. *ACM Sigmod Rec.* 2001, *30*, 15–24. 23.
24. Zhang, J.; Zulkernine, M. A hybrid network intrusion detection technique using random forests. In Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06), Vienna, Austria, 20–22 April 2006.
25. Peng, Y.; Tan, A.; Wu, J.; Bi, Y. Hierarchical Edge Computing: A Novel MultiSource Multi-Dimensional Data Anomaly Detection Scheme for Industrial Internet of Things. *IEEE Access* 2019, *7*, 111257–111270.
26. Zhan, P.; Xu, H.; Luo, W.; Li, X. A novel network traffic anomaly detection approach using the optimal ϕ-DTW. In Pro- ceedings of the IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 16–18 October 2020; pp. 1–4.
27. Saeed, R.A.; Saeed, M.M.; Mokhtar, R.A.; Alhumyani, H.; Abdel-Khalek, S. Pseudonym mutable based privacy for 5G user identity. *Comput. Syst. Sci. Eng.* 2021, *39*, 1–14.
28. Vynokurova, O.; Peleshko, D.; Bondarenko, O.; Ilyasov, V.; Serzhantov, V.; Peleshko, M. Hybrid machine learning system for solving fraud detection tasks. In Proceedings of the IEEE Third International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2020; pp. 1–5.
29. Jwo, D.-J.; Wu, J.-C.; Ho, K.-L. Support Vector Machine Assisted GPS Navigation in Limited Satellite Visibility. *CMC-Comput. Mater. Contin.* 2021, *69*, 555–574.
30. Ng, R.T.; Han, J. Efficient and Effective clustering methods for spatial data mining. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94), San Francisco, CA, USA, 12–15 September 1994; pp. 144–155.
31. Ahmed, Z.E.; Hasan, M.K.; Saeed, R.A.; Hassan, R.; Islam, S.; Mokhtar, R.A.; Khan, S.; Akhtaruzzaman, M. Optimizing Energy Consumption for Cloud Internet of Things. *Front. Phys.* 2020, *8*, 358.
32. Mutanov, G.; Karyukin, V.; Mamykova, Z. Multi-class sentiment analysis of social media data with machine learning algorithms. *Comput. Mater. Contin.* 2021, *69*, 913–930.
33. Dridi, A.; Boucetta, C.; Hammami, S.E.; Afifi, H.; Moungla, H. STAD: SpatioTemporal Anomaly Detection Mechanism for Mobile Network Management. *IEEE Trans. Netw. Serv. Manag.* 2021, *18*, 894–906.
34. Chang, H.; Feng, J.; Duan, C. HADIoT: A Hierarchical Anomaly Detection Framework for IoT. *IEEE Access* 2020, *8*, 154530–154539.
35. Mansour, R.F.; Alfar, N.M.; Abdel-Khalek, S.; Abdelhaq, M.; Saeed, R.A.; Alsaqour, R. Optimal deep learning based fusion model for biomedical image classification. *Expert Syst.* 2022, *39*, e12764.