MedCrave
Step into the World of Research

International Robotics & Automation Journal

Mini Review

Open Access

CrossMark

Research and analysis machine learning methods in network security in telecommunication system

Abstract

The effectiveness of the use of machine learning in network safety in the telecommunication system is analyzed. This is due to the fact that the use of machine learning technology (ML, Machine Learning) in the field of telecommunication systems, information security, or cybersecurity is extremely in demand for experts. Given a complete and labeled training dataset, it makes sense to build a cluster using Machine Learning and Data Mining technologies, which have proven to be the most effective in network security for implementing systems and tools for detecting various attacks using machine learning methods. In particular, machine learning tools are used to identify threats of network safety in telecommunication systems, respectively, threats with confidential data that are stored, processed, transmitted and accepted in these networks. In this work, the tasks of studying machine learning methods in network safety in the telecommunication system and analysis of the metric assessment of the quality of algorithms clustering algorithms are considered. The aim of the study is to increase the quality indicators of machine learning models in solving problems of clustering the state of telecommunication systems. Analytical expressions are given to assess the quality of trained models that use various metrics to evaluate the advantages and disadvantages of clustering methods.

Keywords: machine learning method, recall, unsupervised learning, precision, clustering

Volume 11 Issue 2 - 2025

Bayram Ibrahimov
Department of Radio Radioelectronic and Aerospace Systems,Azerbaijan Technical University, Azerbaijan

Correspondence: Bayram Ibrahimov, Department of Radio Engineering and Telecommunication, Azerbaijan Technical University, Baku, Azerbaijan, Tel +99470-649-07-79

Received: July 23, 2025 | **Published:** August 15, 2025

Introduction

Currently, the problem of information security, the danger factors and intensity many information security events, traffic classification and clustering used in modern multi-service telecommunication networks using machine learning methods, is very acute.[1,2] To solve this problem in telecommunication systems, it is necessary to detect and resist network attacks, analyze and eliminate vulnerabilities, fill the knowledge base on cyber threats, conduct cyber intelligence, etc. However, the huge volume data transmitted over high-speed communication channels and numerous tasks do not allow for real-time analysis.[3–5]

Mathematical methods in multiservice telecommunication networks are rarely created to solve specific engineering problems - they are quite universal. But, nevertheless, the same approach cannot be applied to different types of forecasting. It is clear that forecasting the load in the communication system and finding the upselling pattern cannot be done in the same way. Therefore, a rather important and interesting task is to select forecasting algorithms for different needs of communication operators.[6,7] Some of the trends that arise with the existing complexity of the multiservice network and the use of Big Data are trends in Artificial Intelligence, machine learning, neural networks. At this stage of the study, it was decided to consider the possibilities of machine learning in network security.

It is worth noting that such problems can be solved using ML and artificial intelligence technologies, which are well suited for studying network traffic in telecommunication systems, helping to identify "normal" traffic - including user actions - and separate it from suspicious and potentially dangerous traffic.[8–10] Recently, ML has been considered one of the key tools for ensuring network security and cybersecurity in telecommunication systems.[2,11]

The most promising and relevant technology in the multiservice telecommunication network at present is automated ML, which is a set of instrumental and methodological means that allow to significantly reduce the share human participation in the creation artificial intelligence systems, including by means of automatic validation modeling results.[12–14] Considering the above, this paper examines the research tasks and analysis of machine learning methods in network security in the telecommunication system

General statement of the problem and solution methods

Machine learning is a method of analyzing data in a telecommunications system that allows the analytical system to learn by solving many similar problems, extract information from the source data, identify patterns, and make decisions with minimal human involvement.[1] It is the minimal human involvement and the ability to effectively process and transmit data that makes machine learning methods so relevant given the exponential growth of data volumes in our time. The conducted research shows that in communication systems and computer networks, machine learning is based on three key elements:[2,7–9]

1. Collecting an experimental data set. This can be Internet traffic, network flows, logs, email messages, user activity, and much more. The larger and more diverse the training data, the more accurate the prediction result will be. The effectiveness of ML depends on the quality of the data set.

2. Selection of attributes - features characterizing the data being processed in communication systems. Depending on the task being solved, there may be hundreds of attributes. This may be metadata associated with the file being analyzed: name, creation date, size, presence of network connections, registry access

3. Selection of existing or development of new algorithms and models ML in communication systems.

The correct choice of an algorithm or model that searches for specific features of the searched dataset is a compromise between the speed of the algorithm and its complexity.

To formalize the problem, a new approach is proposed that will most accurately reflect the telecommunication processes in communication systems occurring in the studied link of multiservice networks when assessing the effectiveness of the metric, and will allow obtaining analytical expressions for cluster analysis of data.

Metrics for assessing the quality of clustering algorithms in a communication system

To solve the stated problem and evaluate the quality of trained models, we will consider various metrics.[2,3,6,8,10–14]

1. Complete – Recall reflects what share of objects found by a cluster belongs to a particular class relative to all objects of this class in a test sample. The value can be calculated by the formula:

$$\mathbb{R}\ call = \frac{A_{c,c}}{\sum_{i=1}^{n} A_{i,c}} \le 1 \quad , \qquad (1)$$

where $A_{c,c}$ - diagonal element of c-th class; $\sum_{i=1}^{n} A_{i,c}$ – the sum of all elements of the column of the c-th class.

2. Precision - reflects the proportion objects that actually belong to a long class, relative to the total number of objects that the system assigned to this class. The value can be calculated using the formula:

$$\mathbb{P}\ ecision = \frac{A_{c,c}}{\sum_{i=1}^{n} A_{i,c}} \le 1, \qquad (2)$$

where $A_{c,c}$ - diagonal element of c-th class ; $\sum_{i=1}^{n} A_{i,c}$ – the sum of all elements of the column of the c-th class.

3. Cleanliness- reflects the degree in which clusters contain one class. The value can be calculated by the formula:

$$Purity = \frac{1}{N} \sum_{m \in M} \underset{d \in D}{m} |m \cap d|, \qquad (3)$$

where M-set of clusters; D-set of classes.

4. Rand Index - reflects the percentage of the correct decisions made by the algorithm.

Takes values from 0 to 1, where 0 means 6t, that the result of the distribution of data on clusters does not fully coincide with the classes, and 1-what contents of the clusters completely coincide with the classes.

The value can be calculated by the formula:

$$RandInd = \frac{P + N}{P + P + N + N} \le 1 \quad , \qquad (4)$$

where TR is the number of true positive results when assessing the quality of clustering algorithms;

TN is the number of true negative results;

FP piece of false positive results;

FN is the number of false negative results.

5. F-measure- calculated on the basis of Precision and Recall and is used to balance the deposit of falsely negative distributions. The value can be calculated by the formula:

$$F - measure = \frac{(B^2 + 1) \cdot \mathbb{P}\ ecision \cdot \mathbb{R}\ call}{B^2 \cdot \mathbb{P}\ ecision + \mathbb{R}\ call} \to \max \quad , \qquad (5)$$

where $B -$ the importance of the impact of completeness on F-measure.

6. Jaccard Index - is a numerical assessment of similarity between two data sets.

Takes values from 0 to 1, where 0 means that there are no general elements, and 1-what sets are identical.

The value can be calculated by the formula:

$$J(A, B) = \frac{P}{P + P + N} \le 1 \qquad (6)$$

7. Index DiceDice index is a numerical assessment of the similarity between two data sets, such as Jaccard Index, a measure with doubleting the number of trials. The value can be calculated by the formula:

$$DSC = \frac{2P}{2P + P + N} \le 1 \qquad (7)$$

8. The Fowlkes-Mallows-reflects the similarity between the returned clusters and the standard - classes.

The value can be calculated by the formula:

$$M = \sqrt{\frac{P}{P + P} \frac{P}{P + N}}, \qquad (8)$$

Thus, the proposed formulas (1), ... (8) based on a new approach allow you to evaluate the effectiveness of the metric of assessing the quality of clustering algorithms in telecommunication systems.[2,3,8,11]

Results analysis numerical calculations and experimental research

To evaluate the efficiency of clustering algorithms, traffic measurements were conducted in the data center of the AzerTelecom operator using the Ubuntu 16.04- tcpdump program on the edge router of the communication network. The main volume of traffic was collected during the daytime. TCP (Transport Control Protocol) and UDP (Uzer Datagram Protocol) were used as transport layer protocols.[2,5,14]

First, we consider clustering and partial learning tasks as experimental studies to assess the quality of trained models, which are the following: clustering tasks, partial learning task, and clustering quality criteria. Next, we consider clustering algorithms, which are: K-means method, DBSCAN algorithm, and hierarchical methods. However, this paper examines the metrics for assessing the quality of clustering algorithms KNN (K-nearest Neighbors Method) and SVM (Support Vector Vachine) for different ratios of training and test samples in network security.[5,12,13,15]

To conduct the study and analyze the effectiveness of the proposed calculation method, it was carried out in the Matlab, R 2019b (9.7; 64 bit) environment using the Communications Toolbox package, designed to calculate and model metrics for assessing the quality of clustering algorithms.[2,5] Figure 1 shows the training time of the KNN and SVM algorithms for each of the training and test sample ratios.[2,5,13]

Analysis from Figure 1 shows that the KNN algorithm is fast to learn, but its testing requires significant time, while the SVM algorithm, on the contrary, is slow to learn, but is tested fairly quickly.[13] Based on numerical calculations, Figure 2 shows a graph showing averaged metrics for assessing the quality of clustering using the KNN and SVM algorithms for the best ratios in terms of accuracy.[5,13]

Citation: Ibrahimov B. Research and analysis machine learning methods in network security in telecommunication system. *Int Rob Auto J.* 2025;11(2):40–42. DOI: 10.15406/iratj.2025.11.00299

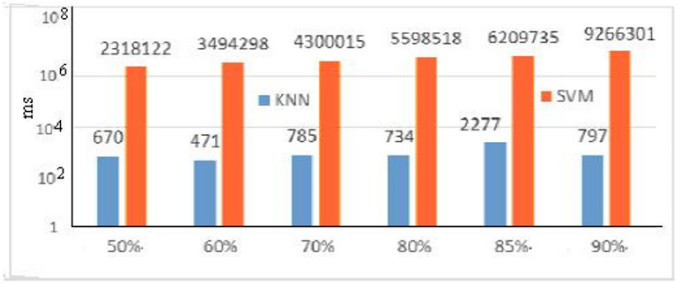

Figure 1 Training time of KNN and SVM algorithms.

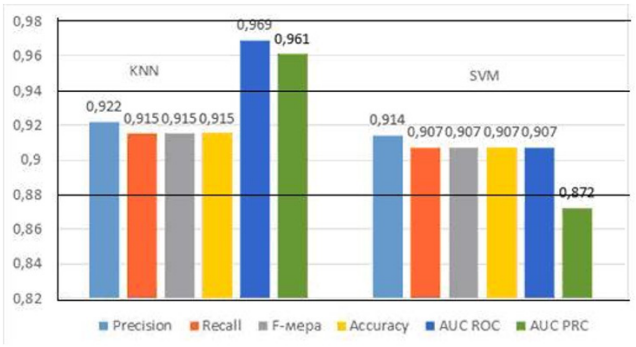

Figure 2 Graph of averaged metrics for assessing the quality of clustering using KNN and SVM algorithms for the best ratios in terms of accuracy.

Analysis of the graphical dependence (Figure 2) shows that the KNN algorithm has greater accuracy in comparison with the SVM algorithm.

Conclusions

As a result of the study of machine learning in network safety in the telecommunication system, they showed that the task of the next stage of research is to verify the possibility of applying the designated methods for previously listed tasks and the choice of the most effective algorithm for each case to evaluate the qualities of clustering algorithms.

It was found that in the presence of a complete and labeled training data set, it is advisable to build a cluster using Machine Learning and Data Mining technologies, which have proven to be the most effective in network security for the implementation of systems and tools for detecting various attacks using machine learning methods. The creation of an "ideal" cluster is impossible until the problems inherent in this area are solved.

Acknowledgements

None.

Conflicts of interest

Authors declares that there is no conflict of interest.

References

1. Ahmad Azab, Mahmoud Khasawneh, Saed Alrabaee, et al. Network traffic classification: Techniques, datasets, and challenges. *Digital communication and Network*. 2024;10(3):676–692.
2. Ahmad Azab, Mahmoud Khasawneh, Saed Alrabaee, et al. Network traffic classification: Techniques, datasets, and challenges. *Digital communication and Network*. 2024;10(3):676–692.
3. Sheloukhin OI, Erokhin SD, Polkovnikov MV. Machine learning technologies in network security. Moscow: Hotline – Telecom. 2021. 360 p.
4. Ibrahimov BG, Mammadov EV. Analysis of some questions on systems for breaking and computer attack detection. Problems of informatization. Proceedings of 12–th International Scientific and Technical Conference November 21–22, 2024; Vol 1: sections 2, 3. Baku–Kharkiv–Bielsko–Biala; 2024. pp. 89– 90.
5. Brink H., Richards J., Feverolf M. Machine learning. St. Petersburg: Piter, 2017. 336 p.
6. Ibrahimov B, Valiyev F. Research useful and service traffic based on machine learning in multiservice software defined network. The proceedings international conference on artificial intelligence and digital development: current realities and future perspectives (AIDD–2024), NASA (https://aidd.anas.az), 17 July, Baku. National Academy Sciences of Azerbaijan. 2024. p. 41–44.
7. Larose DT. Discovering knowledge in data: An introduction to data mining. –Wiley–Interscience. 2014. p. 200–222.
8. Fisher DH. Knowledge acquisition via incremental conceptual clustering. *Machine Learning*. 2007;2:139–172.
9. Ibrahimov B.G., Rafizade U.R. Methods of classification and forecasting network traffic based on machine learning// Problems of informatization. Proceedings of 12–th International Scientific and Technical Conference November 21–22, 2024. Vol 1: sections 1, 2. Baku–Kharkiv–Bielsko–Biala–2024. pp. 48– 49.
10. Lebedev IS, Sikarev IA, Sukhoparov ME. Using uncontrolled clusterization to increase multilevel data processing models quality indicators. *T–Comm*. 2024;18(4)30–35.
11. B Valencia–Vidal, E Ros, I Abadia, et al. Bidirectional recurrent learning of inverse dynamic models for robots with elastic joints: a real–time real–world implementation. *Frontiers in Neurorobotics*. 2023;17:1–15.
12. Este A, Gringoli F, Salgarelli L. Support vector machines for TCP traffic classification. *Computer Networks*. 2009;53(14):2476–2490.
13. Witten Ia H, Frank E Data mining: practical machine learning tools and techniques with java implementations. 2nd ed. San Francisco: Morgan Kaufmann Publ. 2015. 525 p.
14. Bolshakov AS, Gubankova EV. Detection of anomalies in computer networks using machine learning methods. *Telecommunication devices and systems*. 2020;1:37–42.
15. Brett Lantz. Machine learning with R. Pack Publishing. *Birmongham Mumbai*. 2013. 468 p.

Citation: Ibrahimov B. Research and analysis machine learning methods in network security in telecommunication system. *Int Rob Auto J.* 2025;11(2):40–42. DOI: 10.15406/iratj.2025.11.00299