Sciforce

Journal of Artificial Intelligence and Machine Learning

Journal homepage: www.sciforce.org
ISSN : 2995-2336

Open Access Review Article

Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression

Sudhakara Reddy Peram*

Engineering Leader, Illumio Inc., United States

Abstract

As organizations increasingly rely on computer networks for their daily operations, network traffic visualization has become increasingly important. The complexity and volume of raw network traffic data pose significant challenges to human interpretation, forcing the need to transform numerical data into understandable visual formats. This study addresses these challenges by introducing an innovative approach that combines multiple visualization techniques and machine learning methods to improve network traffic analysis and anomaly detection. The research explores four primary visualization approaches: data filtering and transformation, pixel-based techniques, graph-based representations, and integrated multi-view systems. A new method is proposed that incorporates Principal component analysis (PCA) and multidimensional scaling (MDS) serve as efficient techniques for dimensionality reduction, and colour mapping techniques. The system analyses key network performance metrics, including traffic load (MBps), latency (ms), packet loss percentage, and jitter (ms), to provide comprehensive network monitoring capabilities. To improve the prediction accuracy, histogram gradient boosting regression is used, providing excellent performance in handling large-scale datasets with missing values and categorical features. The combined approach demonstrates significant advantages over traditional methods such as t-SNE, especially in preserving multidimensional properties and managing extensive data volumes. This research contributes to more efficient network management through improved network security, improved anomaly detection capabilities, and advanced visualization and machine learning integration.

Keywords: Network traffic visualization, anomaly detection, principal component analysis, multidimensional scaling, histogram gradient boosting regression, dimensionality reduction, network security, packet loss, machine learning, data visualization

Introduction

Network traffic visualization plays an important role as we increasingly rely on computer networks, including the Internet, in our daily operations. Irregular or unstable network behaviour disrupts operations and highlights the need for continuous monitoring of network performance. Since raw network traffic data is generally challenging for humans to understand, visualization methods transform this data into more understandable formats, such as images, so that it can be easily interpreted and analysed using image processing techniques. This method helps in detecting unusual activities, including activities such as Distributed Denial of Service (DDoS) attacks or network monitoring provides valuable insights into connectivity and performance, thereby supporting administrators in protecting the integrity and reliability of the network [1]. To protect digital environments and sensitive information, it is essential to understand and evaluate the vast amount of daily network traffic. However, the dynamic nature and complexity of network traffic patterns make this task difficult and time-consuming. To overcome these challenges, an interactive, web-based visualization system has been introduced. The system uses a combination of integrated visualizations and rich user interactions to enhance users' ability to understand and analyse network traffic data. In addition, it incorporates feature extraction and uncertainty estimation techniques to increase analysis accuracy and effectively detect unusual network traffic patterns [2]. Effective visualization is a key tool for understanding network operations, allowing users to discover insights into network flows and recognize communication patterns.

However, the primarily numerical nature of network traffic data, which contains elements such as timestamps, packet sizes, and inter-packet intervals, makes it difficult to perceive relationships and underlying structures. Many existing visualization techniques face challenges in managing the complexity and volume of extensive network traffic data. This study addresses these existing challenges by introducing a new and efficient A technique for visualizing complex network traffic data enhanced with statistical attributes, combining principal component analysis (PCA) and multidimensional scaling (MDS) perform effective dimensionality reduction, improving visual clarity through the use of colour maps. It is capable of producing high-quality visualizations on real-world datasets, providing a more adaptive and scalable solution compared to widely used techniques such as t-SNE, especially in preserving multidimensional properties and managing large data volumes [3]. Identifying intrusive activities within a network is essential to ensure security. Despite the availability of numerous computational approaches, confirming suspicious events and interpreting their specific characteristics remains a challenging task. To overcome this problem, various visualization frameworks have been designed improve data understanding. However, the full potential of these Graphical methods used for network traffic analysis has yet to be fully explored. This paper bridges the gap by reviewing the current literature and outlining four main approaches to efficient visualization of network traffic includes data filtering and transformation, pixel-based methods, graph-based models, and detailed multi-view architectures.

Received date: October 17, 2023; **Accepted date:** October 28, 2023; **Published date:** November 11, 2023

***Corresponding Author:** Peram, S. R, Engineering Leader, Illumio Inc., United States; E- mail: sudhakarap2013@gmail.com

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281

Evaluate these methods, Initial visualization models were developed and evaluated considering aspects such as implementation difficulty and the amount of data pre-processing, pattern interpretation, and the ability to detect unusual events [4]. Like other systems designed to help users interpret large datasets, Flow Scan focuses on collecting, storing, and visualizing network traffic data. Like many software tools, Flow Scan has predecessors and related systems that have influenced its development and features. Various previous tools have established methods for collecting network traffic data through passive measurement techniques [5]. Its goal is to visualize network traffic between workloads using multiple formats such as graphs, tables, and web-based views. These diverse visualization formats help users observe network relationships between workloads, labels, IP addresses, and ports or protocols. In addition, the system supports enhanced data centre security by improving network visibility and analysis. Beyond facilitating anomaly detection, a key advantage of this approach is its ability to provide an overall a representation of network traffic activity is provided. Since the primary focus of this study is to visualize network traffic data rather than analyse network architecture or topology, it limits its review to previous research focused on network data visualization [6]. A wide range of various techniques are currently used to analyse network traffic. Some visualization focuses on specific machines or depicts the connection between a single host and external entities.

However, although many administrators we interviewed manage more than 100 systems, existing methods generally do not support very large home or external networks. We believe that our approach offers the highest scalability among the concrete network visualization techniques available [7]. This allows the use from image processing and video compression methods, including techniques such as scene change analysis and motion estimation, in packet header data, which uncovers significant characteristics of network traffic. In this study, we combine methods from image processing and video analysis to improve traffic description [8]. Recently, there has been a significant increase in network traffic across the Internet, local area networks, enterprise private networks, and data centre environments. To effectively manage this traffic, it is often necessary to visualize overall connection patterns, such as end-to-end connection diagrams. For example, in an enterprise in a virtual private network (VPN), administrators need access up-to-date Monitoring traffic behaviour is crucial to maintaining situational awareness. In the event of safety alerts, administrators should conduct interactive traffic analysis to implement immediate mitigation measures and appropriate remediation measures [9]. Based on the effects of the architecture, Grid View effectively presents network traffic in grids categorized by application layer protocols. In addition, Plotter View visualizes wireless network traffic across the entire campus on a single dynamic screen that automatically scales to the size of the network. These interactive features greatly improve the ability to detect and monitor compromised devices, while also reducing response time [10].

Materials and Methods

Traffic Load MBps: Traffic load it refers to the amount of data exchanged across a network within a given time frame. It an important parameter that indicates how much demand is placed on a network's resources, including bandwidth and processing capacity. Traffic load varies depending on the number of active users, the types of applications in use, and the time of day. Monitoring traffic load helps improve network performance, avoid congestion, and ensure efficient data delivery across the system.

Latency ms: Latency is the time delay between the initiation of a request and the corresponding response in a network or computer system. It is usually measured in milliseconds and is a key factor in determining the performance of time-sensitive applications such as video conferencing, online gaming, and cloud computing. Latency can be affected by many factors such as transmission distance, routing paths, and hardware processing times. Low latency leads to faster communication and an improved user experience, which is essential for efficient network operations.

Packet Loss %: Packet loss describes a situation where data packets within a network fail to reach their designated destination. Occurs when one or more data packets travelling across a network are rejected or dropped Network congestion, downtime caused by hardware, software glitches, or weak signal strength. Packet loss can reduce network performance and lead to outages, delays, or poor quality in Services such as VoIP, multimedia streaming, and interactive online gaming. Monitoring and minimizing packet loss is essential to maintaining reliable and efficient data communications.

Jitter ms: Jitter refers to the variation in packet arrival times during data transmission over a network. It is caused by network congestion, path changes, or time drift, which leads to inconsistent delays between data packets. Jitter is particularly problematic for Time-sensitive applications such as VoIP, video meetings, and interactive online gaming, where a consistent data stream is critical for quality performance. High jitter can lead to choppy audio, video lag, or dropped calls. Reducing jitter is important to ensure consistent and reliable communication.

Instructions for machine learning:

Hist Gradient Boosting Regression: Hist Gradient Boosting Regression is an advanced machine learning algorithm used to predict continuous outcomes. It is a variant of gradient boosting that uses histogram-based techniques to improve both training speed and memory capacity. Instead of evaluating all possible split points for continuous features, the algorithm bins the values into discrete intervals (histograms), which significantly reduces computational time and allows it to scale efficiently to large datasets. This method builds a set the decision is based on a set of trees, with each successive tree aiming to correct the mistakes made by its predecessor. one by minimizing a specific loss function, typically the mean square error for regression tasks. The trees are grown continuously, and the overall prediction is obtained by aggregating the outputs of all individual trees. Hist Gradient Boosting Regression is particularly effective in dealing with missing values and categorical features, and it often outperforms traditional gradient boosting methods in both speed and accuracy. It is implemented in libraries such as Scikit-learn and is well suited for large-scale machine learning tasks.

Results and Discussions

The datasets provide network traffic visualization data that shows how various performance metrics vary with increasing traffic load (in MBps) from 10 to 80. As traffic load increases, latency (ms), packet loss (%) and jitter (ms) typically increase, indicating network congestion. Initially, all metrics show gradual changes, but after about 40 MBps, latency and jitter increase significantly, and packet loss becomes more erratic, peaking near 4%. This indicates that the network becomes less efficient under heavy loads. The table is useful for analysing quality of service (QoS) under varying traffic conditions and identifying thresholds for optimal performance.

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281

Table 1. Provides descriptive statistics for network performance measurements across 100 data points				
	Traffic Load MBps	Latency ms	Packet Loss %	Jitter ms
count	100.00000	100.00000	100.00000	100.00000
mean	45.00000	77.29231	2.25446	4.53245
std	20.51318	30.90392	1.04440	2.07430
min	10.00000	25.78408	0.21693	0.59399
25%	27.50000	50.04677	1.40546	2.72256
50%	45.00000	77.64174	2.29654	4.47465
75%	62.50000	104.99794	3.18598	6.28564
max	80.00000	129.53083	3.97629	8.33572

Table 1 provides descriptive statistics for network performance measurements across 100 data points. The average traffic load is 45MBps, with corresponding averages of 77.29ms latency, 2.25% packet loss, and 4.53ms jitter. The standard deviation indicates moderate variability, especially in latency. The minimum values indicate optimal conditions (e.g., 25.78ms latency and 0.22% packet loss), while the maximum values reflect congestion (e.g., 129.53ms latency and 3.98% packet loss). The median range is 50% of the data falls between 50.05 and 105ms latency and 2.72 to 6.29ms jitter, indicating performance degradation as traffic increases.

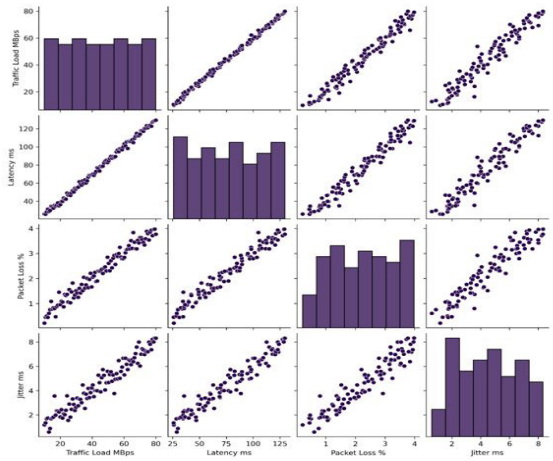

Figure 1: Scatter plot of the variousNetwork Traffic Visualization

Figure 1. Scatter plot matrix illustrating the relationships between key network traffic parameters: traffic load (Mbps), latency (ms), packet loss (%) and jitter (ms). The graphs reveal strong positive correlations between all metrics, with increased traffic load leading to higher latency, packet loss and jitter, indicating potential network congestion

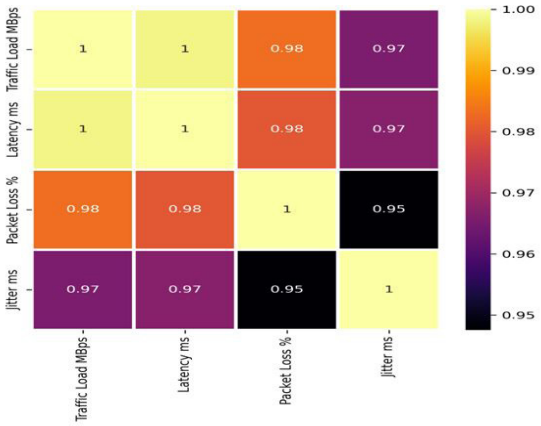

Figure 2: Heat map of the connection between process variables and outcomes

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281

Figure 2. Heat map illustrating the relationship between network process variables and performance outcomes. The values indicate strong positive correlations between traffic load, latency, packet loss, and jitter, with coefficients ranging from 0.95 to 1.00. This indicates that an increase in traffic load significantly affects the overall network performance degradation.

Hist Gradient Boosting Regression (Traffic Load MBps):

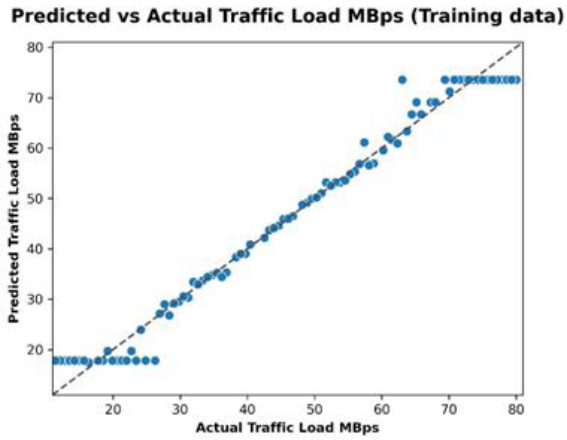

Figure 3: Hist Gradient Boosting Regression on Traffic Load MBps: training data

Figure 3. Predicted vs. actual traffic load (Mbps) using histogram-based gradient boosting regression on the training data. The scatterplot shows predictions closely aligned with the diagonal reference line, indicating high model accuracy and effective learning. Most of the data points cluster around the line, reflecting strong agreement between predicted and actual values.

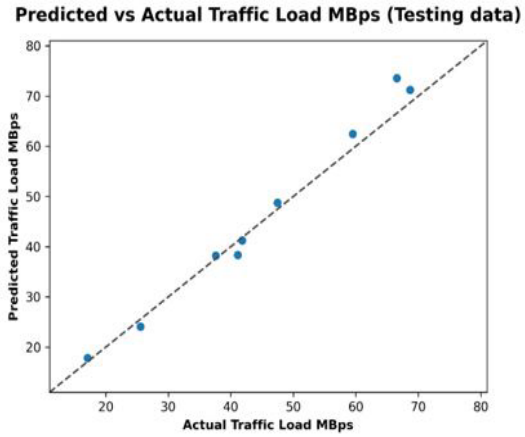

Figure 4: Hist Gradient Boosting Regression on Traffic Load MBps: testing data

Figure 4. Predicted vs. actual traffic load (Mbps) using histogram-based gradient boosting regression on test data. The data points closely align with the diagonal reference line, indicating the strong generalization performance of the model. The tight clustering of the points indicates reliable prediction accuracy and minimal deviation between the actual and predicted traffic loads.

Table 2. Performance Metrics of Hist Gradient Boosting Regression on Traffic Load MBps (Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Traffic Load MBps	Train	HGBR	Hist Gradient Boosting Regression	0.98000	0.98000	8.43276	2.90392	1.91380	10.52296	0.01434	1.11462
	Test	HGBR	Hist Gradient Boosting Regression	0.96141	0.97071	13.83801	3.71995	2.78138	7.85079	0.03141	1.99473

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281

Table 2 summarizes the performance metrics of the Hist Gradient Boosting Regression (HGBR) model for predicting traffic load (MBps) using both Datasets used for model training and evaluation. For the training set, model performs exceptionally well, achieving R^2 and EVS of 0.98 with low errors (MSE: 8.43, RMSE: 2.90, MAE: 1.91). On the test data, the performance is strong (R^2: 0.961, EVS: 0.971), although the errors increase slightly (MSE: 13.84, RMSE: 3.72, MAE: 2.78), indicating good generalization ability. The low MSLE and Median Absolute Error (Med AE) on both sets confirm that HGBR reliably models traffic load patterns.

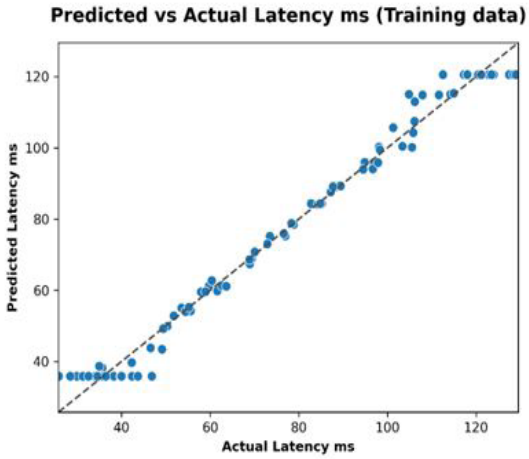

Figure 5: Hist Gradient Boosting Regression on Latency ms: training data

Figure 5. Predicted vs. actual latency (ms) using histogram-based gradient boosting regression on the training data. The data points closely follow the diagonal line, indicating strong prediction accuracy. The model effectively captures latency patterns with minimal deviation between predicted and actual values, demonstrating its robustness and relevance on the training dataset.

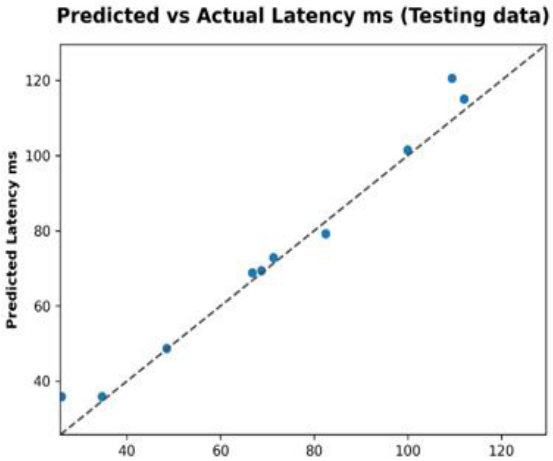

Figure 6: Hist Gradient Boosting Regression on Latency ms: testing data

Figure 6. Predicted vs. actual latency (ms) using histogram-based gradient boosting regression on test data. The scatterplot shows a strong alignment with the reference line, reflecting accurate model predictions. The close clustering of points indicates effective generalization, maintaining high latency prediction accuracy in data where the model is not observed.

Table 3. Performance Metrics of Hist Gradient Boosting Regression on Latency ms (Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Latency ms	Train	HGBR	Hist Gradient Boosting Regression	0.98465	0.98465	14.70970	3.83532	2.69249	10.91003	0.00572	1.68710
	Test	HGBR	Hist Gradient Boosting Regression	0.96816	0.97819	25.44815	5.04462	3.47578	11.18833	0.01124	1.84285

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281

Table 3 presents the performance evaluation of the Hist Gradient Boosting Regression (HGBR) model for predicting latency (ms) using both training and test data. This model demonstrates excellent accuracy on the training set with an R^2 and EVS of 0.98465 and low error metrics (MSE: 14.71, RMSE: 3.84, MAE: 2.69). On the test data, the model maintains strong predictive ability with an R^2 of 0.96816 and EVS of 0.97819, although the error values increase slightly (MSE: 25.45, RMSE: 5.04). Overall, the low MSLE and Med AE indicate strong model performance and minimal deviation from the true latency values.

Hist Gradient Boosting Regression (Packet Loss %):

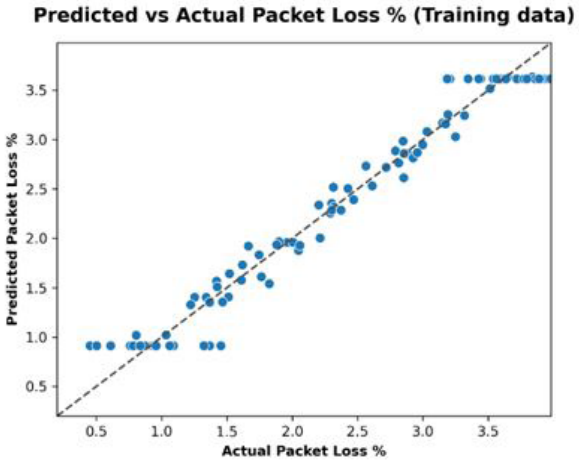

Figure 7: Hist Gradient Boosting Regression on Packet Loss %: training data

Figure 7. Predicted vs. actual packet loss (%) using histogram-based gradient boosting regression on training data. The scatterplot shows a strong correlation, with most points closely following the diagonal line. This indicates that the model accurately captures the relationship in the training data and effectively predicts the packet loss percentages.

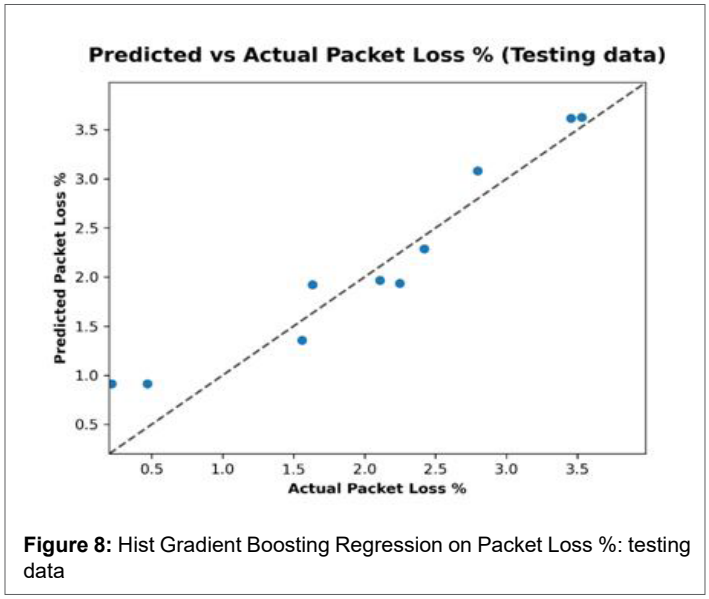

Figure 8: Hist Gradient Boosting Regression on Packet Loss %: testing data

Figure 8. Predicted vs. actual packet loss (%) using histogram-based gradient boosting regression on experimental data. The scatterplot shows a strong linear relationship, with most data points closely aligned to the diagonal reference line. This indicates that the model maintains high prediction accuracy and generalizes well to unobserved packet loss values.

Table 4. Performance Metrics of Hist Gradient Boosting Regression on Packet Loss %(Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Packet Loss %	Train	HGBR	Hist Gradient Boosting Regression	0.96342	0.96342	0.03915	0.19787	0.14871	0.53758	0.00644	0.10939
	Test	HGBR	Hist Gradient Boosting Regression	0.90507	0.91780	0.10605	0.32566	0.27626	0.69774	0.03138	0.24340

Table 4 outlines the performance metrics of the Hist Gradient Boosting Regression (HGBR) model for predicting packet loss (%) using the training and testing datasets. On the training set, the model performs well with an R^2 and EVS of 0.96342 and low error metrics (MSE: 0.03915, RMSE: 0.19787, MAE: 0.14871). The testing results show slightly lower performance with increased errors (MSE: 0.10605, RMSE: 0.32566) (R^2: 0.90507, EVS: 0.91780), but are still within acceptable limits. The low MSLE and Median Absolute Error (Med AE) for both sets indicate accurate modelling, indicating that HGBR effectively captures the packet loss behaviour under varying network conditions.

Table 5. Hist Gradient Boosting Regression on Jitter ms(Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Jitter ms	Train	HGBR	Hist Gradient Boosting Regression	0.94492	0.94492	0.23948	0.48936	0.35801	1.67351	0.01463	0.24147
	Test	HGBR	Hist Gradient Boosting Regression	0.84946	0.85970	0.51046	0.71446	0.62887	1.18019	0.01965	0.69081

Table 5 presents the performance metrics of the Hist Gradient Boosting Regression (HGBR) model for predicting jitter (ms) on both the training and testing datasets. On the training data, this model achieves robust performance with an R^2 and EVS of 0.94492 and low error values (MSE: 0.23948, RMSE: 0.48936, MAE: 0.35801). On the testing data, the performance is reliable with an R^2 of 0.84946 and EVS of 0.85970, although the errors increase moderately (MSE: 0.51046, RMSE: 0.71446). The low MSLE and Med AE indicate that the model generalizes well and accurately captures jitter trends across different network conditions.

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281

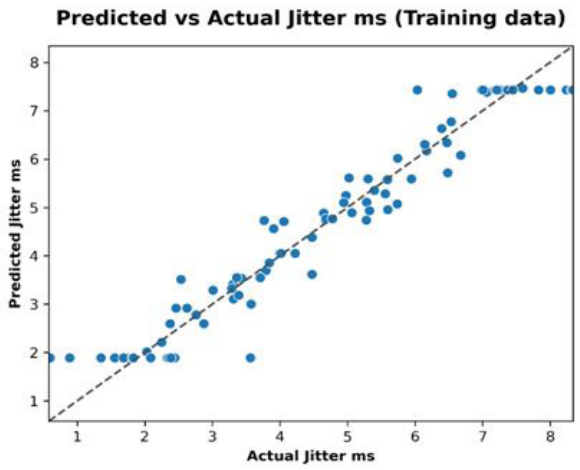

Figure 4: Hist Gradient Boosting Regression on Jitter ms: training data

Figure 9. Predicted vs. actual jitter (ms) using histogram-based gradient boosting regression on the training data. The scatterplot illustrates a strong linear fit along the diagonal line, indicating accurate model predictions. The close clustering of points confirms the model's effectiveness in learning jitter patterns from the training dataset with minimal error.

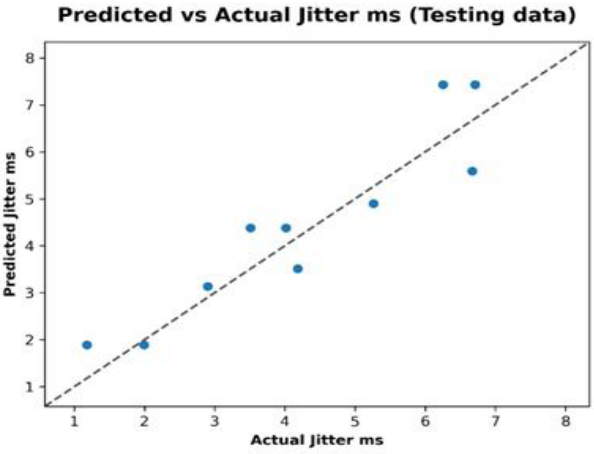

Figure 10: Hist Gradient Boosting Regression on Jitter ms: testing data

Figure 10. Predicted vs. actual jitter (ms) using histogram-based slope boosting regression on experimental data. The scatterplot reveals a strong correlation, with most points closely following the diagonal line. This indicates that the model generalizes well, accurately predicting jitter values in unobserved data with minimal deviation from actual observations.

Conclusion

Based on the detailed analysis presented in this study, the research successfully demonstrates the effectiveness of combining advanced visualization techniques with machine learning methods for network traffic analysis and anomaly detection. The proposed approach, which integrates Dimensionality reduction techniques such as PCA and MDS with colour mapping techniques, provides significant improvements over traditional methods such as t-SNE, especially in preserving multidimensional properties and managing extensive data volumes. The implementation of Histogram Gradient Boosting Regression (HGBR) has been shown to be highly effective in all network performance metrics. The model demonstrates exceptional accuracy in predicting traffic load (R^2 = 0.96-0.98), latency (R^2 = 0.97-0.98), packet loss (R^2 = 0.91-0.96), and jitter (R^2 = 0.85-0.94). These results indicate strong predictive capabilities with minimal deviation between actual and predicted values, confirming the model's reliability for real-world network monitoring applications. Correlation analysis reveals strong positive relationships between all network parameters, with coefficients ranging from 0.95 to 1.00. As traffic load increases, especially evident beyond 40 MBps, latency and jitter show a significant increase, and packet loss becomes more erratic.

The research contributes significantly to network management by providing administrators with advanced tools for proactive monitoring and anomaly detection. The system's ability to handle large-scale datasets with missing values and categorical features makes it particularly valuable for enterprise environments managing extensive network infrastructures. Future work should focus on expanding visualization techniques to incorporate real-time streaming capabilities and exploring additional machine learning algorithms for comparative analysis. The integration of deep learning approaches could further improve anomaly detection accuracy, while creating automated alert systems based on established performance thresholds could provide immediate operational benefits for network administrators looking to maintain optimal network performance and security.

References

1. Plonka, Dave. "{FlowScan}: A Network Traffic Flow Reporting and Visualization Tool." In 14th Systems Administration Conference (LISA 2000). 2000.

2. Corchado, Emilio, and Álvaro Herrero. "Neural visualization of network traffic data for intrusion detection." Applied Soft Computing 11, no. 2 (2011): 2042-2056.

3. Ji, Soo-Yeon, Bong-KeunJeong, and Dong Hyun Jeong. "Evaluating visualization approaches to detect abnormal activities in network traffic data." International Journal of Information Security 20, no. 3 (2021): 331-345.

4. Ball, Robert, Glenn A. Fink, and Chris North. "Home-centric visualization of network traffic for security administration." In Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, pp. 55-64. 2004.

5. Sridhar Kakulavaram. (2022). Life Insurance Customer Prediction and Sustainbility Analysis Using Machine Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 390 –. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7649

6. Ruan, Zichan, Yuantian Miao, Lei Pan, Yang Xiang, and Jun Zhang. "Big network traffic data visualization." Multimedia Tools and Applications 77 (2018): 11459-11487.

7. Kim, Seong Soo, and AL Narasimha Reddy. "NetViewer: A Network Traffic Visualization and Analysis Tool." In LISA, vol. 5, pp. 18-18. 2005.

8. Adari, V. K., Chunduru, V. K., Gonepally, S., Amuda, K. K., & Kumbum, P. K. (2020). Explainability and interpretability in machine learning models. Journal of Computer Science Applications and Information Technology, 5(1), 1–7. https://doi.org/10.15226/2474-9257/5/1/00148

9. Shi, Lei, Qi Liac, Xiaohua Sun, Yarui Chen, and Chuang Lin. "Scalable network traffic visualization using compressed graphs." In 2013 IEEE International Conference on Big Data, pp. 606-612. IEEE, 2013.

10. Bhardwaj, Amit Kumar, and Maninder Singh. "Data mining-based

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281

integrated network traffic visualization framework for threat detection." Neural Computing and Applications 26 (2015): 117-130.

11. Patwari, Neal, Alfred O. Hero III, and Adam Pacholski. "Manifold learning visualization of network traffic data." In Proceedings of the 2005 ACM SIGCOMM workshop on Mining network data, pp. 191-196. 2005.

12. Landge, Aaditya G., Joshua A. Levine, AbhinavBhatele, Katherine E. Isaacs, Todd Gamblin, Martin Schulz, Steve H. Langer, Peer-Timo Bremer, and Valerio Pascucci. "Visualizing network traffic to understand the performance of massively parallel simulations." IEEE Transactions on Visualization and Computer Graphics 18, no. 12 (2012): 2467-2476.

13. Popa, F. (2009). Network Traffic Visualization. https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2009-01-1/NET-2009-01-1_12.pdf

14. Jeong, D. H., Cho, J.-H., Chen, F., Kaplan, L. M., Jøsang, A., & Ji, S.-Y. (2022). Interactive Web-Based Visual Analysis on Network Traffic Data. Information, 14(1), 16. https://doi.org/10.3390/info14010016

15. Ruan, Z., Miao, Y., Pan, L., Xiang, Y., & Zhang, J. (2018). Big network traffic data visualization. Multimedia Tools and Applications, 77(9), 11459–11487. https://doi.org/10.1007/S11042-017-5495-Y

16. Ji, S.-Y., Jeong, B.-K., &Jeong, D. H. (2021). Evaluating visualization approaches to detect abnormal activities in network traffic data. International Journal of Information Security, 20(3), 331–345. https://doi.org/10.1007/S10207-020-00504-9

17. Nicholls, J., Peters, D., Slawinski, A., Spoor, T., Vicol, S., Happa, J., Goldsmith, M., &Creese, S. (2013). NetVis: a Visualization Tool Enabling Multiple Perspectives of Network Traffic Data. 9–16. https://doi.org/10.2312/LOCALCHAPTEREVENTS.TPCG.TPCG13.009-016

18. Bethel, E. Wes, Scott Campbell, Eli Dart, Kurt Stockinger, and Kesheng Wu. "Accelerating network traffic analytics using query-driven visualization." In 2006 IEEE Symposium on Visual Analytics Science and Technology, pp. 115-122. IEEE, 2006.

19. Chunduru, V. K., Gonepally, S., Amuda, K. K., Kumbum, P. K., & Adari, V. K. (2021). Real-time optical wireless mobile communication with high physical layer reliability using GRA method. Journal of Computer Science Applications and Information Technology, 6(1), 1–7. https://doi.org/10.15226/2474-9257/6/1/00149

20. Chen, Wei, FangzhouGuo, and Fei-Yue Wang. "A survey of traffic data visualization." IEEE Transactions on intelligent transportation systems 16, no. 6 (2015): 2970-2984.

21. Swing, Edward. "Flodar: Flow visualization of network traffic." IEEE Computer Graphics and Applications 18, no. 5 (2002): 6-8.

22. Shiravi, Hadi, Ali Shiravi, and Ali A. Ghorbani. "A survey of visualization systems for network security." IEEE Transactions on visualization and computer graphics 18, no. 8 (2011): 1313-1329.

23. Le Malecot, Erwan, Masayoshi Kohara, Yoshiaki Hori, and Kouichi Sakurai. "Interactively combining 2D and 3D visualization for network traffic monitoring." In Proceedings of the 3rd international workshop on Visualization for computer security, pp. 123-127. 2006.

24. Estrin, Deborah, Mark Handley, John Heidemann, Steven McCanne, Ya Xu, and Haobo Yu. "Network visualization with nam, the vint network animator." Computer 33, no. 11 (2000): 63-68.

25. Arnold, David, Mikhail Gromov, and JafarSaniie. "Network traffic visualization coupled with convolutional neural networks for enhanced iot botnet detection." IEEE Access (2024).

26. Goodall, John R., Wayne G. Lutters, Penny Rheingans, and Anita Komlodi. "Preserving the big picture: Visual network traffic analysis with tnv." In IEEE Workshop on Visualization for Computer Security, 2005. (VizSEC 05). pp. 47-54. IEEE, 2005.

27. Promrit, Nuttachot, and AnirachMingkhwan. "Traffic flow classification and visualization for network forensic analysis." In 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, pp. 358-364. IEEE, 2015.

28. Shi, Ronghua, Mengjie Yang, Ying Zhao, Fangfang Zhou, Wei Huang, and Sheng Zhang. "A matrix-based visualization system for network traffic forensics." IEEE Systems Journal 10, no. 4 (2015): 1350-1360.

Citation: Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. https://doi.org/10.55124/jaim.v1i3.281