Predictive Framework for Enhanced Wireline Network Reliability

Unveiling Anomalies and Streamlining Maintenance

A technical paper prepared for presentation at SCTE TechExpo24

Madiha Sahar
Sr. Data Scientist
Rogers Telecommunications
Madiha.sahar@rci.rogers.ca

Ray Stevens
Manager Service Reliability Engineering
Rogers Telecommunications
ray.stevens@rci.rogers.com

Jenny Panman
Manager Wireline Data Science
Rogers Telecommunications
evgenia.panman@rci.rogers.com

Vikram Karwal
Sr. Data Scientist
Rogers Telecommunications
vikram.karwal@rci.rogers.com

Anna Korchatov
Sr. Data Scientist
Rogers Telecommunications
anna.korchatov@rci.rogers.com

Peter Theodorakidis
Sr. Cable Access Network Specialist
Rogers Telecommunications
panagiotis.theodorakidis@rci.rogers.com

Davy Ma
Sr. Data Scientist
Rogers Telecommunications
davy.ma@rci.rogers.com

Courtney Lovell
Sr. Network Quality Specialist
Rogers Telecommunications
courtney.lovell@rci.rogers.com

Mahmood Mohiuddin
Network Quality Specialist
Rogers Telecommunications
mahmood.mohiuddin@rci.rogers.com

Table of Contents

List of Figures

List of Tables

1. Introduction

In today's sophisticated wireline network infrastructures, the Cable Modem Termination System (CMTS) interfaces with customer homes through an intricate array of branching connections, establishing distinct pathways for each subscriber. This setup not only facilitates individualized connections but also constitutes a comprehensive map of the access network. This map delineates key network elements including ports, media access control (MAC) domains, and critical components such as amplifiers and splitters, all of which play essential roles in managing the link between the gateway modem and the CMTS.

At the heart of this connectivity is the node, a pivotal aggregation point strategically located based on geographic considerations. Each node typically services a varying number of devices, making it an optimal point to measure quality of service (QoS) of the access network.

QoS is a critical measure used to evaluate and ensure the reliability and efficiency of these network connections. It encompasses various metrics such as network latency, availability, bandwidth, jitter, and packet loss, which collectively determine the network's ability to deliver a consistent and high-quality user experience. Measuring QoS at the node level is crucial for identifying and addressing area-specific issues within the access network. By focusing on nodes, network operators can differentiate between widespread, location-specific problems and individual customer issues.

This targeted measurement and analysis not only help maintain a responsive, reliable, and accessible network for all users but also assist service technicians in pinpointing and isolating root causes of network problems. By identifying issues at the node level, technicians can more effectively plan and execute targeted interventions, addressing network-wide problems more efficiently and minimizing disruptions for individual customers.

To effectively manage and troubleshoot network performance, it's essential to understand the roles of key physical network components and the potential issues that can impact QoS. Additionally, environmental factors can significantly influence network performance.

Table 1 below lists some of the physical components of network that can directly impact QoS.

Table 1 - Physical Components of Network

Distribution	Access	Transmission	Customer Premises
Digital Subscriber Line Access Multiplexer (DSLAM), fibre to the node (FTTN), fibre to the curb (FTTC)	Remote Terminal	Amplifiers	Modems
Optical Line Terminal (OLT)	Street Cabinets	Repeaters	Routers
Uplink & Downlink Cards, Chasis	Media Access Point Controllers	Transceivers	Drops

Each one of these components may have different issues that could impact the customer, with each outline below:

- Digital Subscriber Line Access Multiplexer (DSLAM)**: A**ggregates multiple DSL connections.

  Issues: Failures or misconfigurations can cause slow speeds and connectivity problems by inefficiently managing bandwidth.

- Fiber to the Node (FTTN) / Fiber to the Curb (FTTC)**:** Brings fiber closer to subscribers for better performance.

  Issues: Fiber breakage or degradation can reduce bandwidth and increase latency, impacting overall speed.

- Remote Terminal: Extends network reach.

  Issues: Failures can disrupt connectivity and cause inconsistent service for users connected through that terminal.

- Optical Line Terminal (OLT)**:** Manages the interface between fiber networks and local networks.

  Issues: Hardware or software problems can lead to connectivity issues and decreased data throughput.

- Amplifiers: Boost signal strength to extend coverage.

  Issues: Faulty amplifiers can cause signal degradation and noise, leading to poor connectivity and reduced speeds.

- Repeaters: Regenerate and boost signals.

  Issues: Malfunctions can result in signal loss or attenuation, affecting connectivity over long distances.

- Transceivers: Facilitate data transmission and reception.

  Issues: Problems with transceivers can lead to packet loss and increased latency.

- Street Cabinets: House essential network equipment.

  Issues: Power failures, overheating, or physical damage can disrupt equipment functionality and cause localized service interruptions.

- Modems: Provide connectivity between the network and customer devices.

Issues: Faulty modems can lead to slow speeds, frequent disconnections, and poor service quality.

- Routers: Manage data flow and network traffic.

  Issues: Configuration errors or hardware failures can result in routing problems and network congestion.

- Drops: Connect the network to customer premises.

  Issues: Physical damage or poor connections can lead to unreliable service and connectivity issues.

- Uplink & Downlink Cards, Chassis: Support data transmission between network components.

  Issues: Failures or issues with these cards can disrupt data flow and network efficiency.

- Media Access Point Controllers**:** Manage and coordinate network traffic.

  Issues: Problems can lead to traffic congestion and reduced network performance.

Additionally environmental factors may impact network performance, each of these is listed in Table 2 below.

- Radio Frequency Interference (RFI)

Issues: External radio frequency signals can interfere with network equipment, causing data transmission errors and reduced performance.

- Electromagnetic Interference (EMI)

  Issues: Electromagnetic fields from nearby electronic devices can disrupt network signals, leading to connectivity issues and degraded service quality.

- Power Supply Stability

  Issues: Fluctuations or failures in power supply can affect network equipment reliability, leading to outages or degraded performance.

- Infrastructure Accessibility

  Issues: Limited access to network infrastructure for maintenance or repairs can delay issue resolution and impact overall network performance.

- Temperature, Precipitation, Storms

  Issues: Extreme weather conditions can damage physical network components, affect signal quality, and lead to service interruptions.

Table 2 - Environmental Components of Access Network

Radio Frequency Interference (RFI)	Electromagnetic Interference (EMI)	Power Supply Stability	Infrastructure Accessibility	Temperature, Precipitation, Storms

This study aims to advance proactive network management by developing a comprehensive tool for monitoring both upstream and downstream channels. Central to this effort is the creation of a predictive model that defines and forecasts the health of Access Networks. Emphasizing the impact on user experience and connectivity issues, the model assesses the likelihood of node failures, categorized as Red, Yellow, and Green. By integrating these capabilities, telecommunications providers can enhance network reliability and optimize maintenance strategies, ensuring robust service delivery to end-users.

The framework's primary objective is to assess the likelihood of node degradation before failure, empowering proactive network response teams with alerts of potential network disruptions. The framework also equips field technicians with the information derived from access layer metrics to devise actionable strategies toward resolution, facilitating efficient network management.

2. Literature Review

In the evolving landscape of network management, ensuring high QoS and effective prioritization remains critical for optimizing performance and user experience. Recent advancements in technology and methodologies have introduced innovative approaches to address these challenges. This literature review examines key contributions to the field of QoS and prioritization, focusing on studies that propose solutions for managing network performance and resource allocation. By exploring these approaches, we gain insights into how modern techniques can enhance network reliability and address issues related to traffic management.

2.1. Predictive Analytics in Satellite Telecommunications

Ochuba et al. (2024) provide an in-depth review of predictive analytics techniques for satellite telecommunications infrastructure, emphasizing the use of statistical modeling, machine learning algorithms, and big data tools. Their work underscores the importance of integrating predictive analytics into Proactive Network Maintenance (PNM) to forecast equipment failures and optimize maintenance schedules. This approach aligns with the PNM perspective by aiming to pre-emptively address potential disruptions, thus enhancing network reliability and performance. However, the review lacks a comparative analysis of different predictive techniques, practical integration challenges, and scalability considerations for diverse satellite systems. Future research could benefit from exploring real-world implementations and integration strategies, as well as addressing scalability and practical application across various satellite environments.

2.2. Scheduling Policies in Real-Time Systems

Kargahi and Movaghar (2006) analyze the Earliest-Deadline-First (EDF) scheduling policy, focusing on optimizing real-time task management based on deadlines. This policy enhances system responsiveness and reduces missed deadlines, contributing to effective network performance management. While EDF's theoretical advantages are well-documented, the study does not compare EDF with other scheduling policies or address practical implementation challenges. Additionally, scalability in complex or large systems is not discussed. Incorporating comparative analyses with alternative scheduling policies and exploring real-world case studies could offer deeper insights into EDF's practical applicability and performance.

2.3. Outlier Detection Methods

Ren et al. (2004) present Relative Density Factor (RDF), a density-based outlier detection method that utilizes vertical data representation. This method aims to detect anomalies by analyzing the density of data points, which is crucial for maintaining network performance through early detection of irregularities. The paper, however, lacks a comparative analysis with other outlier detection techniques, and there is limited discussion on scalability and real-world applications. Future research should address these gaps to better evaluate RDF's effectiveness and practical implementation in various contexts.

Wang et al. (2009) introduce a distance-based outlier detection method for uncertain data, which is vital for handling anomalies in datasets with inherent uncertainty. While innovative, the study does not compare this method with other techniques for uncertain data, nor does it discuss performance metrics and scalability. Exploring these aspects could enhance the understanding of the method's effectiveness in diverse scenarios.

Radovanović et al. (2015) explore the use of reverse nearest neighbors (RNN) in unsupervised distance-based outlier detection. This method improves anomaly detection accuracy by analyzing neighborhood relationships. The study lacks a comparative analysis with other distance-based methods and does not address scalability or practical applications. Including these elements would provide a more comprehensive evaluation of RNN's effectiveness in various network contexts.

Kriegel et al. (2012) focus on outlier detection in arbitrarily oriented subspaces, dealing with high-dimensional data. The method's contribution to network performance management is significant, yet it lacks detailed comparisons with other subspace-based methods and scalability considerations. Practical validation is also missing. Addressing these gaps could offer a more robust assessment of this approach.

Zimek et al. (2012) survey unsupervised outlier detection techniques for high-dimensional data. While the survey is extensive, it lacks a detailed comparative evaluation and real-world application examples. Discussing emerging trends could provide additional insights into the current state and future directions of outlier detection methods.

2.4. Anomaly Detection in Mobile Networks

Gajic et al. (2015) propose an improved anomaly detection method using incremental time-aware clustering. This approach enhances traditional clustering by incorporating temporal patterns, which is critical for maintaining network performance and addressing anomalies proactively. Despite its innovation, the paper does not compare this method with other techniques and lacks scalability and real-world validation discussions. Future research should address these aspects to provide a more comprehensive evaluation of the method's effectiveness.

Hadj-Kacem et al. (2020) focus on anomaly prediction in mobile networks, employing a data-driven approach to select suitable machine learning algorithms. This aligns with the prediction perspective by aiming to improve prediction accuracy. However, the study lacks a comparative analysis of different algorithms, practical implementation challenges, and detailed performance metrics. Addressing these elements could enhance the applicability and effectiveness of the proposed approach.

2.5. Conclusion

The reviewed literature provides significant insights into predictive analytics, scheduling policies, and anomaly detection methods, each contributing to PNM and predictive strategies. However, common gaps such as the need for comparative analyses, scalability considerations, and real-world validation are evident. Addressing these gaps in future research can enhance the practical applicability and effectiveness of these methods, leading to more robust and reliable network management solutions.

3. Methodology

In the dynamic realm of telecommunications, effectively managing network performance demands advanced tools capable of navigating the complexity and rapid evolution of modern networks. Our methodology responds to this need by leveraging a tool designed for near real-time data analysis, enabling precise insights and enhanced network efficiency through sophisticated analytics. This tool is adeptly engineered to handle the substantial volume and velocity of data, generating actionable results every three hours.

To address gaps identified in existing research, our approach integrates a comparative analysis of various unsupervised machine learning algorithms, focusing on density-based and decision-based methods for network classification. While literature highlights theoretical frameworks and algorithmic innovations, practical implementations often face limitations, including discrepancies between theoretical benchmarks and actual operational thresholds. Our methodology bridges this gap by establishing operational key performance indicators (KPI) benchmarks tailored to real-world conditions, taking into account the physical and environmental factors that influence network performance.

The framework we propose is structured around four key components as presented in Figure 1:

a. **Threshold Analysis**: Develops operational benchmarks based on observed performance trends, moving beyond theoretical models to accommodate the practical nuances of network segments.
b. **Anomaly Classification**: Employs advanced techniques to detect and categorize anomalies, providing insight into deviations from expected performance and their potential impacts.
c. **Node Health Prediction**: Utilizes predictive modeling to forecast the future state of network nodes, prioritizing maintenance efforts based on anticipated needs.
d. **Priority Assessment**: Optimizes resource allocation and maintenance scheduling by evaluating predictive insights and ensuring timely responses to potential disruptions.

By capturing both temporal and spatial data, the framework addresses the continuous evolution of networks, treating them as dynamic systems requiring ongoing monitoring and maintenance. This approach not only enhances network reliability but also ensures that maintenance strategies are informed by real-time data and predictive insights, thereby overcoming limitations highlighted in previous research and advancing the state of network management.

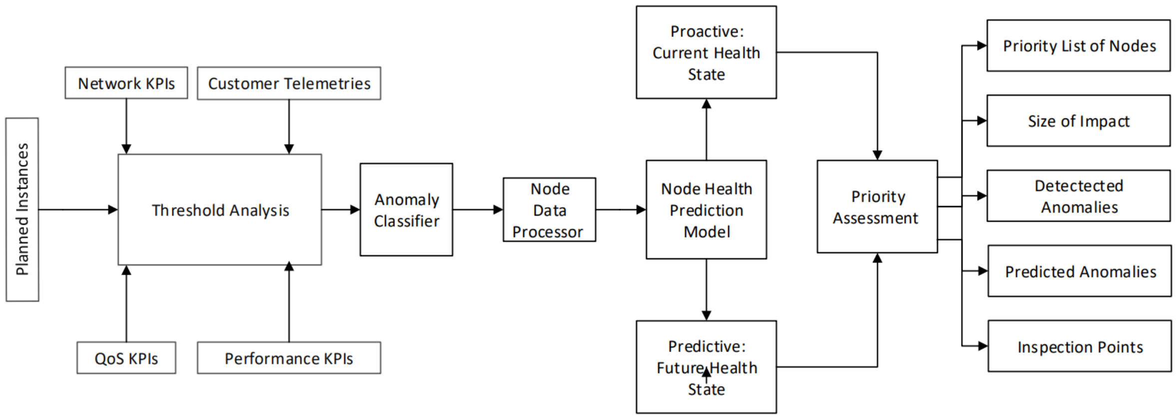

Figure 1- Proposed Framework for Enhanced Network Reliability

3.1. Threshold Analysis

Access network connectivity is best assessed through key telemetries such as Signal-to-Noise Ratio (SNR), Receive Power (Rx), Transmit Power (Tx), Modulation Error Rate (MER), and Packet Error Rate (PER). However, the dynamic, nonlinear, and non-stationary nature of this data, compounded by its dependence on physical, seasonal, and environmental factors, poses significant challenges to prediction accuracy. Traditional approaches often rely on aggregated data, which can create blind spots, particularly in extreme scenarios or when observed over brief periods. To overcome these limitations, our framework employs a benchmark solution that enhances the accuracy and reliability of network performance assessment.

Existing research, including studies on PNM and QoS, frequently highlights discrepancies between theoretical KPI criteria and actual network performance. For instance, while Cable Labs provide specific criteria for KPIs such as pass, marginal pass, and fail, real-world data often reveals that network segments may function satisfactorily even when individual KPIs fail to meet these documented thresholds. This indicates a critical gap where theoretical models do not fully capture the practical operational state of the network. The physical components of the network significantly influence performance, with observed values frequently deviating from prescribed ranges

To address these challenges, our framework incorporates a robust benchmark solution by evaluating network performance through a detailed analysis of pass, marginal pass, and fail criteria over extended periods. This approach ensures the integrity of the information relayed by the data and mitigates the limitations associated with traditional threshold-based evaluations.

a. Network Segmentation: The framework begins by performing a similarity analysis of the network to identify distinct segments using unsupervised clustering algorithms such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Spectral Clustering. The similarity analysis is performed using Euclidean distance represented in Eq 1 which measures absolute distance and sensitive to scale and magnitude. Other approach used is Cosine Similarity measures as shown in Eq 2 which captures directional similarity. It was found that cosine similarity resulted in better separation in clusters due to the high variance and dimensionality of data. This segmentation helps in managing and analyzing network data more effectively as presented in Figure 2 and 3.

Euclidean Distance Calculation:

$$d(v_i, v_j) = \sum k = \ln(v_{ik} - v_{jk})^2 \quad (1)$$

Cosine Similarity Calculation:

$$(v_i, v_j) = \|v_i\|\|v_j\|v_i v_j \quad (2)$$

b. Threshold Analysis: It involves both inter-cluster and intra-cluster examinations to determine the optimal number of network segments with significant distinctions. This analysis employs extreme value analysis and advanced outlier detection algorithms to refine the thresholds for each KPI. Techniques such as Isolation Forest, Local Outlier Factor (LOF), and One-Class Support Vector Machines (OC-SVM) are utilized to identify and handle anomalies effectively.

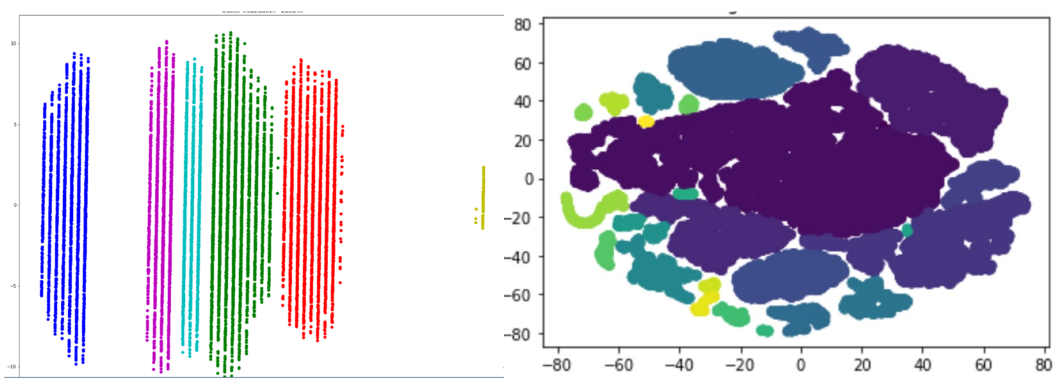

Figure 2 & 3- Network Segments DBSCAN (left) Network Segments Spectral Clustering (right)

In the initial phase of threshold analysis, a modified density-based unsupervised clustering algorithm is employed to account for the temporal and spatial dynamics of network data. This method reveals five distinct clusters, each large enough to be considered as individual network segments. Among these, one cluster is notably denser and is designated as the "standard cluster." This standard cluster can either be further subdivided into more specific sub-clusters using hierarchical clustering techniques or be treated as a representative model of the network's "healthy" state, indicating optimal performance.

To enhance the precision of performance assessment, the framework establishes two key benchmarks for defining a 'healthy' state. The primary benchmark involves identifying the standard cluster as a proxy for optimal network conditions. The secondary benchmark is a representative vector for each segment, calculated as a sixty-day rolling median of each KPI. This approach allows for a comparative analysis of segment performance over time, helping to identify and address chronic underperformance by highlighting deviations from the historical performance norms.

Additionally, cluster profiling techniques and machine learning algorithms, such as Random Forest and Support Vector Machines (SVMs), are used to compute precise thresholds for each cluster. These thresholds are crucial for evaluating network performance against established criteria, ensuring a robust and dynamic assessment process.

Receive power (Rx) which is one of the QoS KPIs is one of the key components of the analysis. Figure 4 displays the passing thresholds for Rx computed for above mentioned 5 network segments in green and red represents theoretical thresholds provided.

Figure 4 - Threshold for Rx

3.2. Anomaly Classification

In the anomaly classification phase, the framework applies previously computed thresholds to evaluate network performance at a granular level. Instead of aggregating KPI data, this model assesses how frequently each KPI measurement falls outside the predefined passing range. This approach offers a precise view of network health, quantifying the number of instances where each KPI is classified as a pass or fail.

a. Each node is evaluated based on whether its KPI measurements meet the pass, marginal pass, or fail criteria. This model tracks the frequency of deviations from the acceptable range for each KPI, ensuring that performance assessment reflects true operational conditions.
b. The resulting data presents a detailed performance profile of the network, showcasing the frequency of pass versus fail occurrences for each KPI. This profiling provides insight into the distribution and severity of performance issues across different network segments.
c. To further analyze the KPI data, distance-based clustering algorithms such as K-Nearest Neighbors (KNN), DBSCAN, and ordering points to identify the clustering structure (OPTICS) are utilized. These algorithms identify patterns and correlations among KPIs by evaluating the proximity of KPI values.

   KNN: Classifies or clusters data based on the distance between points, highlighting network segments with similar performance characteristics.

   DBSCAN: Detects clusters based on density, identifying areas of similar KPI performance and potential outliers.

   OPTICS: Handles varying densities to uncover clusters with different characteristics, providing a detailed view of KPI performance.

d. Dimensionality Reduction and Key KPI Identification: Techniques such as Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) are employed to identify the key contributing KPIs. PCA reduces dimensionality and reveals the most significant KPIs that explain variance in the data, while t-SNE visualizes high-dimensional KPI data in lower dimensions, facilitating the identification of patterns and important features.

e. Feature Selection with Least Absolute Shrinkage and Selection Operator (LASSO) Regression: LASSO regression is used to calculate the coefficients of each KPI as shown in Table 3. This regression technique applies a penalty to the size of coefficients, effectively selecting the most relevant KPIs by shrinking less important ones to zero. The resulting coefficients are used as weights in further analysis, ensuring that the most influential KPIs are prioritized.

f. Root Cause Analysis and Troubleshooting: The insights from clustering, dimensionality reduction, and feature selection are integrated to enhance root cause analysis. By understanding KPI correlations and identifying key performance indicators, onsite technicians can more accurately diagnose performance issues. This targeted approach streamlines troubleshooting, improves diagnostic accuracy, and enhances overall network maintenance efficiency.

This comprehensive methodology not only provides a detailed assessment of network performance but also ensures that the analysis is based on the most relevant and impactful KPIs. By combining threshold evaluation, clustering, dimensionality reduction, and regression analysis, the framework delivers a robust solution for proactive network management and optimization.

Table 3 - KPI Importance

KPI Name	Feature Weight
User Experience (Video)	3.5
Accessibility	2.6
Customer Interaction	5.2
QoS KPI	3.1

3.3. Node Health Prediction

To achieve accurate and actionable node health predictions in dynamic telecommunications networks, this framework integrates detailed methodologies and advanced algorithms. Here’s a technical overview:

The functional state of network components is assessed using two key references:

a. Reference 1: Anomaly classification results, which categorize each network component's performance based on detected deviations from normal operation.
b. Reference 2: A 60-day rolling median of each KPI, providing a historical baseline for normal performance under the assumption that the network is predominantly functional.

The system performs a comparative analysis by evaluating the current KPI measurements against those from previous hours. This involves calculating the deviation of current measurements from historical baselines using statistical methods such as z-scores or Mahalanobis distance.

Short-Term Forecasting: Predictive models generate forecasts for the next three hours based on observed trends and deviations. This is achieved through time-series forecasting techniques, such as Autoregressive Integrated Moving Average (ARIMA) or exponential smoothing, tailored to capture the network's rapid fluctuations.

In predictive modeling, Gradient Boosting Machines (XGBoost) is utilized for anomaly classification. XGBoost processes high-dimensional KPI vectors to classify anomalies. It employs decision trees with gradient boosting, optimizing the loss function to improve prediction accuracy. XGBoost evaluates feature importance through gain metrics, assessing each KPI's contribution to anomaly detection.

Graph Neural Networks (GNNs) were implemented to model the spatial relationships between nodes. GNNs aggregate information from neighboring nodes to predict future states, using node embeddings and message-passing techniques to capture complex dependencies and network topology. This approach enables the prediction of node degradation and potential failures based on graph-based analysis.

The framework includes mechanisms for periodic re-evaluation and retraining of the models to adapt to changing network conditions. This involves recalibrating XGBoost models and updating GNN parameters using recent data batches, ensuring the models reflect current network dynamics and maintain prediction accuracy.

For adaptive learning, techniques such as online learning or incremental training are employed to continuously integrate new data, allowing the models to learn from recent trends and anomalies without requiring complete retraining from scratch.

By incorporating these detailed methodologies, the framework provides a robust solution for predicting node health. It effectively captures the dynamic nature of network performance through advanced statistical analysis, predictive modeling, and continual model refinement. This approach ensures accurate, short-term forecasts and facilitates proactive network management. In this framework determining the priority of network segments for intervention involves a sophisticated analysis of the predicted state of nodes. This section outlines the approach and technical details used to prioritize maintenance tasks effectively.

The predicted state of nodes, derived from the Gradient Boosting Machines (XGBoost) and Graph Neural Networks (GNNs), provides a forecast of potential degradation and future anomalies. This prediction is critical for evaluating which segments are at risk and require immediate attention. To rank network segments based on their predicted state, a One-Class SVM algorithm is employed. This algorithm is particularly effective for anomaly detection in high-dimensional spaces. It defines a boundary around normal data points and identifies deviations as outliers. For our application:

The SVM is trained on historical KPI data to establish a boundary of normal operational states for each network segment.

Using the predicted states, the One-Class SVM computes the anomaly scores for each segment. These scores reflect how much a segment deviates from normal behavior, thus determining its priority for maintenance.

The assessment process incorporates various factors influencing network development and growth. This includes:

a. Active Factors: Current network conditions, recent changes, and ongoing issues.
b. Planned Factors: Upcoming network expansions, scheduled upgrades, and anticipated changes in traffic patterns.

c. Impact Size and Resolution: The framework evaluates the size of potential impacts and the complexity of required resolutions. Segments with higher impact potential and complex resolution needs are prioritized higher.

By leveraging the One-Class SVM for anomaly-based ranking and considering both current and predicted states of the network, the framework provides a robust method for prioritizing maintenance tasks. This approach ensures that network segments most in need of intervention are addressed efficiently, enhancing overall network reliability and performance.

a. Dynamic Allocation: Based on the SVM-derived rankings, maintenance service technicians receive prioritized lists of network segments. This ensures that critical issues are addressed promptly, optimizing resource allocation and minimizing downtime.
b. Continuous Adjustment: The priority list is updated regularly, reflecting the latest predictions and network conditions. This allows for adaptive maintenance strategies that align with the network's evolving state and operational demands.

4. Implementation

Implementation Key Success Points:

a. Near real time data collection and processing
b. High granularity of the data used for analysis (15-minute interval)
c. End-to-end cloud-based solution allows scalability, powerful compute, data accessibility and use of innovative technologies.
d. Multiple updates during a day with most recent outcomes.
e. The model provides a practical base to cross-business functionality improvement.

Effective documentation and adherence to industry best practices are critical for managing complex data processing and machine learning systems. The framework's documentation encompasses detailed records of ETL workflows, including data extraction, transformation, and loading procedures. Comprehensive logs of machine learning model parameters, including hyperparameters, training epochs, and learning rates are maintained. Implemented data lineage track to document the flow and transformation of data across ETL processes.

k-fold cross-validation, where $k = 10$, is used within the process to assess model performance. This involves splitting the dataset into 10 subsets, training the model 10 times with different training sets, and evaluating its performance on each subset. Metrics such as precision, recall, and F1-score should be monitored. For instance, aim for a precision of 0.90 and a recall of 0.85. Integration of techniques to monitor and detect model drift regularly by comparing recent model predictions against historical performance metrics to adapt models as necessary.

Using these insights and evolving requirements, a feedback loop is established to review the system performance and user feedback regularly. Structured approach to feature updates and model refinements to ensure continuous enhancement of system performance.

Once the model training for anomaly classification and network segment identification for thresholding is performed, the inferencing is run on the Cloud using optimized techniques and dedicated resources. The final outcomes of this model are presented as a dashboard to be utilized across various departments. Constant feedback is collected and integrated in the system to improve its intelligence.

For data quality assurance, scripts to detect data anomalies and inconsistencies are performed at each step. When established thresholds for acceptable data quality are violated it triggers a warming or process fails depending on the nature of the alert. These automated alerts for critical issues like data pipeline failures or high resource utilization ensure immediate action can be taken and to resolve bottle necks.

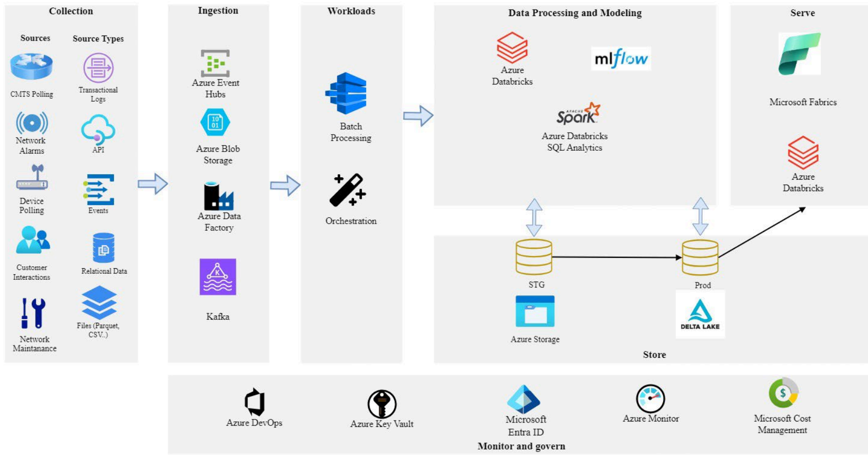

Figure 5: Process Design and Implementation

Figure 5 provides a clear flow of end-to-end implementation and deployment of the framework highlighting some of the technologies and tools utilized.

5. Conclusion

The innovative framework and advanced methodologies introduced in this paper represent a significant leap forward in network management and maintenance. By addressing critical gaps in traditional approaches, this solution empowers network operators to effectively tackle the complexities of modern telecommunications infrastructures. The proof of concept (POC) run with network operators and maintenance proved the model accuracy to be 80%.

One of the major strengths of this framework is its ability to expedite troubleshooting and fault resolution, which is essential during peak periods of network usage. Through informed segmentation and dynamic load balancing, the framework optimizes node performance and ensures that resources are allocated equitably. This leads to enhanced overall network efficiency and effectiveness, addressing a key challenge identified in current research—ensuring consistent performance amidst varying traffic patterns and operational demands.

The framework's emphasis on early prediction of potential network failures enables pre-emptive interventions, effectively mitigating service disruptions and improving reliability. This proactive approach goes beyond conventional reactive strategies, enhancing customer satisfaction by minimizing unexpected outages and maintaining seamless connectivity.

Moreover, the detailed fault classification across different time intervals provides actionable insights that are crucial for informed decision-making. These insights guide strategic network policies and infrastructure investments, promoting operational resilience and ensuring that network assets are utilized optimally. By addressing the limitations of theoretical benchmarks and integrating real-world data, the framework establishes robust operational thresholds and eliminates blind spots that have previously hindered performance assessments.

The integration of advanced techniques, such as Gradient Boosting Machines (XGBoost), Graph Neural Networks (GNNs), and One-Class Support Vector Machines (SVMs), accelerates root cause analysis, swiftly identifying and resolving underlying issues. This reduces downtime and operational costs, setting a new standard for proactive, data-driven network maintenance.

In summary, this groundbreaking methodology redefines traditional reactive management practices by introducing a comprehensive, predictive approach to network maintenance. By transforming how network performance is monitored and managed, the framework not only enhances network resilience and performance but also establishes a new benchmark for proactive maintenance strategies. It provides network operators with the tools needed to continuously improve service delivery, optimize customer experience, and adapt to the dynamic nature of modern network environments.

Abbreviations

AP	access point
bps	bits per second
CMTS	Cable Modem Termination System
DSLAM	Digital Subscriber Line Access Multiplexer
DBSCAN	forward error correction
EDF	earliest-deadline-first
EMI	electromagnetic interference
FEC	forward error correction
FTTC	fiber to the burb
FTTH	fiber to the home
FTTN	fiber to the node
GNN	graph neural networks
HD	high definition
Hz	hertz
K	kelvin
KNN	k-nearest neighbors
KPI	key performance indicators
LASSO	least absolute shrinkage and selection operator
LOF	local outlier factor
MAC	media access control
MacDomain	media access control domain
MER	modulation error rate
OC-SVM	one-class support vector machines
OLT	optical line terminal
OPTICS	ordering points to identify the clustering structure
PER	packet error rate
PNM	proactive network maintenance
PCA	principal component analysis
QoS	quality of service
RFI	radio frequency interference
RNN	reverse nearest neighbors
Rx	receive power
SCTE	Society of Cable Telecommunications Engineers
SNR	signal to noise ratio
SVM	support vector machines
t-SNE	t-distributed stochastic neighbor embedding
Tx	transmit power
XGBOOST	Gradient Boosting Machines
RDF	Relative Density Factor
ARIMA	Autoregressive Integrated Moving Average
ELT	Extract, Load, Transform

Bibliography & References

[1] Ochuba, N. A., Usman, F. O., Okafor, E. S., Akinrinola, O., & Amoo, O. O. (2024). Predictive analytics in the maintenance and reliability of satellite telecommunications infrastructure: A conceptual review of strategies and technological advancements. Engineering Science & Technology Journal -ISSN: 2708-8952, Volume 3 (Issue 5), P.No. 704-715, March 2024

[2] M. Kargahi and A. Movaghar, "A Method for Performance Analysis of Earliest-Deadline-First Scheduling Policy", J. Supercomputing, vol. 37, no. 2, pp. 197-222, 2006.

[3] Ren D, Wang B, Perrizo W (2004) Rdf: a density-based outlier detection method using vertical data representation. In: extitFourth IEEE international conference on data mining (ICDM'04), pp 503–506

[4] Wang B, Xiao G, Yu H, Yang X (2009) Distance-based outlier detection on uncertain data. In: 2009 Ninth IEEE international conference on computer and information technology, vol 1, pp 293–298

[5] Radovanović M, Nanopoulos A, Ivanović M (2015) Reverse nearest neighbors in unsupervised distance-based outlier detection. IEEE Trans Knowl Data Eng 27(5):1369–1382

[6] Kriegel H, Kröger P, Schubert E, Zimek A (2012) Outlier detection in arbitrarily oriented subspaces. In: 2012 IEEE 12th international conference on data mining, pp 379–388

[7] Zimek A, Schubert E, Kriegel H-P (2012) A survey on unsupervised outlier detection in high-dimensional numerical data. Stat Anal Data Min ASA Data Sci J 5(5):363–387

[8] Borislava Gajic, Szabolcs Nováczki and Stephen Mwanje, "An improved anomaly detection in mobile networks by using incremental time-aware clustering", 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), pp. 1286-1291, 2015.

[9] Imed Hadj-Kacem, Sana Ben Jemaa, Sylvain Allio and Yosra Ben Slimen, "Anomaly prediction in mobile networks: A data driven approach for machine learning algorithm selection", NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium, pp. 1-7, 2020.

[10] Pichanun Sukkhawatchani and Wipawee Usaha, "Performance evaluation of anomaly detection in cellular core networks using self-organizing map", 2008 5th International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology, vol. 1, pp. 361-364, 2008.

[11] Jun Wu, Patrick PC Lee, Qi Li, Lujia Pan and Jianfeng Zhang, "Cellpad: Detecting performance anomalies in cellular networks via regression analysis", 2018 IFIP Networking Conference (IFIP Networking) and Workshops, pp. 1-9, 2018

[12] Putina A, Rossi D, Bifet A, Barth S, Pletcher D, Precup C, Nivaggioli P (2018) Telemetry-based stream-learning of BGP anomalies. In: Proceedings of the 2018 workshop on big data analytics and machine learning for data communication networks. Big-DAMA'18, pp 15–20. Association for Computing Machinery, New York.

[13] Cao F, Estert M, Qian W, Zhou A (2006) Density-based clustering over an evolving data stream with noise. In: Proceedings of the 2006 SIAM international conference on data mining, pp 328–339.

[14] Ming-Chang Lee, Jia-Chun Lin, Volker Stolz, "NP-Free: A Real-Time Normalization-free and Parameter-tuning-free Representation Approach for Open-ended Time Series", 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), pp.334-339, 2023.

[15] Zhiwen Tian, Ming Zhuo, Leyuan Liu, Junyi Chen, Shijie Zhou, "Anomaly detection using spatial and temporal information in multivariate time series", Scientific Reports, vol.13, no.1, 2023.

[16] J. Ma and S. Perkins, "Time-series novelty detection using one-class support vector machines", Proceedings of the International Joint Conference on Neural Networks (IEEE), vol. 3, pp. 1741-1745, July 2003.

[17] Zhang K, Hutter M, Jin H (2009) A new local distance-based outlier detection approach for scattered real-world data. In: Theeramunkong T, Kijsirikul B, Cercone N, Ho T-B (eds) Advances in knowledge discovery and data mining. Springer, Berlin, pp 813–822

[18] Juan M Ramírez, Fernando Díez, Pablo Rojo, Vincenzo Mancuso and Antonio Fernández-Anta, "Explainable machine learning for performance anomaly detection and classification in mobile networks", Computer Communications, 2023.

[19] Ming-Chang Lee, Jia-Chun Lin, Ernst Gunnar Gran, "How Far Should We Look Back to Achieve Effective Real-Time Time-Series Anomaly Detection?", Advanced Information Networking and Applications, vol.225, pp.136, 2021.

