JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2021, VOL 4, No S4

Machine Learning Models for Predictive Maintenance and Performance Optimization in Telecom Infrastructure

Venkata Bhardwaj Komaragiri,

Lead Data Engineer*, *bhardwajkommaragiri@gmail.com*, *ORCID ID : 0009-0002-4530-3075

Abstract

Internet Protocol (IP) networks transform entire global communication systems. The growing demand for IP networks with immense capacity and ultra-low latency has substantially changed telecommunications operators' infrastructures. As the number of elements within these infrastructures has multiplied, telcos must shift from today's profitability-driven network maintenance to a proactive approach. Reactive maintenance using simple indicators and counter levels will lead to network failure, affecting customer experience and business continuity. Hence, there is a growing need for Predictive Maintenance. Telco predictive maintenance and performance optimization efforts focus on Machine Learning models that leverage historical data within data warehouses. Such databases comprise both structured and unstructured data related to network design, operation, and performance. The use of temp-data with new technologies represents a further breakthrough in regulatory telecom management. Different analysis and predictive analytics models have been developed, from basic statistical models to complex algorithms.

Imperceptibly, many solutions have been implemented and used successfully. However, a significant challenge associated with these models, such as extreme events or unplanned IP element maintenance, hampers their use. Relying solely on historical data creates inherent limitations, as data patterns will change or disappear risk being out of service. Despite model advances, telecom data storage and technology create additional predictive maintenance challenges. As new data generation levels rise, current data warehouses will become too expensive. Even with cost-efficient storage, questions arise regarding data relevance. Additionally, as network elements become homogenized, clearly defined settings lead to standard behavior. Moreover, storage issues exist due to network evolution over time, data system merge, and tele-operational sequence changes, where the absence of historical data creates knowledge gaps despite passing internal and third-party regulations.

Keywords:Predictive Maintenance,Telecom Infrastructure Monitoring,Machine Learning in Telecommunications,Anomaly Detection Models,Network Fault Prediction,Telecom Equipment Failure Forecasting,Time Series Forecasting Telecom,Performance Optimization Algorithms,Condition-Based Maintenance (CBM),AI-Driven Network Management,Reinforcement Learning Telecom Optimization,Edge Analytics for Telecom,Telecom Network Health Scoring,Big Data in Telecom Maintenance,Root Cause Analysis ML Models.

1. Introduction

The telecommunications industry is a crucial part of the global economy because it lays the groundwork for virtually all businesses. In the last several years, the communications network has become increasingly sophisticated and broad, resulting in a significant rise in the quantity of data

and information held in this field. Telecom infrastructure must function 24 hours a day and offer customers a high-quality experience . Preventing downtime and ensuring maximum efficiency while considering growth rate and service demand are also crucial for critical customer satisfaction and protecting income sources. Machine learning (ML) has undergone enormous growth in recent years and has been utilized in various applications across several areas as the price of technology has dropped and processing capabilities and data have grown. ML applications in telecom networks or properties enable predicting failures of hardware/planned maintenance, predicting customer outages and providing solutions, predicting performance collapse or drop-offs, among many others. The industry has recognized ML's potential benefits and consequences; as a result, there is a booming need for more efficient telecom networks and penalized service providers. All this pushes the need to analyze, model, and optimize performance from a telecom infrastructure perspective. Telecom infrastructure that provides a general understanding of telecom infrastructure and its properties is heavily reliant on hardware. Different equipment requires maintenance times and ways, as well as information about how often they fail or run out of operational resources. Modeling and building ML applications based on these properties are crucial for understanding how much weight will play in general performance. An ML-based ecosystem that optimizes the operation of telecom infrastructure from hardware perspective is prototyped, explored, and demonstrated in this article. The novel concept of ML Model Framework is proposed for communication of services/sensor data and modeling resulting data. With this framework, models for predicting hardware utilization metrics (such as CPU/Core/Memory load) on OSS-level, and for failure prediction at DWDM components are created from, in total, more than 70 M time-dependent data points from a well-known telecom network. The models assess kPI/SLAs or forecast the necessity of planned maintenance, pointing out optimization directions.

Fig 1: Machine Learning for Predictive Maintenance in Telecom Infrastructure.

1.1. Background and Significance

In telecommunications, the rapid evolution of technologies has created excitement, opportunity, and uncertainty about the future of the domain. Equipment vendors and service providers are investing significant resources to keep up with the trends and remain competitive. To adapt, Telecommunications Operators need to develop their understanding of data, use cases, and system engineering expertise. Many use cases in predictive maintenance (PM) and performance optimization and enhancement are still addressed in a manual fashion. There is a lack of intelligent or data-driven end-to-end holistic solutions with the right automation level. Currently, these use cases are either not addressed, poorly addressed, or ad-hoc tasks. With the increasing need for automated processes and optimization, there is an immense opportunity to proactively enhance the services of the entire operation process.

Given the Customer Experience and Service priority, the importance of the operation processes is critical. Oftentimes, these processes depend on telecommunication network data and Critical Product Quality monitoring. Needed improvements include precise product quality monitoring in order to reduce the manual effort, focus on KPI dependencies rather than individual KPIs, seamless visualization of results, focus on discovering genuine problems rather than outlier detection, actionable Big Data analysis, and integration with other operation processes. These enhancements will enable intelligent and automated decision making or alerting, thus boosting the overall efficiency. It is hoped that these intelligent solutions can boost numerous operation processes with rapid business returns. There is much confidence in the win-win situation these intelligent solutions will create, and that the company will become a data miner as well as a Business Expert in Telecom domain. As the first step toward this goal, initial experiences, lessons learned, and research on the telecom Telco Network Data Intel for predictive maintenance will be described.

There are multiple reasons to involve machine learning (ML) technologies to enhance the telecom network with intelligent solutions. Telecom networks and their infrastructure are immensely complex, which makes it more difficult to get insights from the data and find the root cause of problems. Additionally, these networks generate and store massive amounts of data every day. Telecom infrastructure products are extremely expensive, where an edge node product costs over an order higher than enterprise switches from other business domains [3]. Merely monitoring alarms to significant events is not enough, since the level of insight is low and the information extracted is too late.

2. Overview of Telecom Infrastructure

Telecom infrastructure supports broadband communication across antennas, base stations, control and monitoring equipment, transmission equipment, switching equipment, etc., spanning core networks, backbone networks and access networks. Intelligent fault prediction for delivering telecom infrastructure is still a challenging problem. Deep learning models are shown to be promising for intelligently predicting faults over a network but can hardly be applied to telecom infrastructure due to the lack of contextual data. On the other hand, knowledge graphs have been widely adopted in various intelligent applications, including fault prediction. They can assist in reasoning and logic and output explainable knowledge to reflect the described knowledge. As two promising techniques, they are independently used but inspiring scenarios may be raised on the synergy of both domains. The event observation sequence can be used to compare different telecom infrastructure nodes. Clustering algorithms can be applied to the extracted sequence, and node groups can be formed with the same types of similar historical events using model-based approaches. Further, deep learning classification models are used to predict the occurrence probabilities of events. Meanwhile, knowledge graphs can be constructed to model telecom infrastructure nodes, events, observations, etc. On the one hand, a general knowledge graph can be constructed to automatically enrich the original event-interpretation pairs based on their embeddings or node features. On the other hand, a telecom knowledge graph can be constructed based on the pre-designed logical schema and commercial-internal knowledge.

Tx Equipment types including DWDM Mux, Optical Amplifier, Wavelength Router, Mux/Demux, etc., representing different manufacturers, models, and numbers of subcomponents can be involved in generating dummy tx equipment records. For example, based on the dummy base stations,

telecom data including equipment monitoring thresholds, historical pm despike and forecasted pm records, event and geo information can be generated. Telemetry pm records are collected and transferred to online data engines along with other event data. In addition, the generated telecom data can be saved as records for testing and benchmarking purposes. Meanwhile, additional real-world pm data are collected from telecom pm data servers for training and prediction purposes. A knowledge graph that describes telecom equipment, events, and monitoring records can be constructed based on detailed telecom domain knowledge.

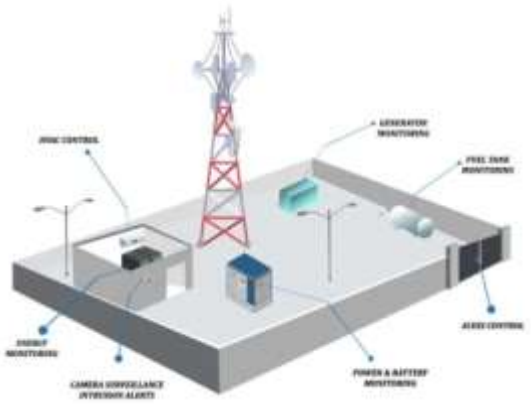

Fig 2: Telecom Infrastructure Management

2.1. Evolution and Modern Trends in Telecom Infrastructure

Telecommunication Networks (Telecom) Countries across the world are investing heavily on remodeling their national network infrastructure. Telecom has been one of the fastest growing domains across the globe. End user demand for faster connectivity, installation of Wireless Local Area Network (Wi-Fi) in offices, homes, metros, public zones and malls has increased the burden on Telecom infrastructure providers. The primary concern of Telecom Infrastructure Providers (TIP) is to provide and maintain quality uptime of the network so that quality of service to end users is not compromised. Telecom Infrastructure consists primarily of mobile towers or BTS (Base Transmitter Station). BTS infrastructure consists of first introduced 2G, which uses Frequency Division Multiple Access (FDMA) for voice and GSM 2G technology. The next 3G upgrade uses Code Division Multiple Access (CDMA) which is spread spectrum technology based. The latest 4G technology upgrade is Time Division Multiple Access (TDMA) based Long Term Evolution (LTE). The latest upgrade by service provider will be 5G which is a duty based millimeter wave technology fighting for more spectrum. Each version upgrade, addition of different make and added with various feeders, amplifiers, additional routers and switches have increased the complexity of the Telecom systems.

The number of alarms from various devices has increased exponentially with the deployment. TIPs employ a large number of engineers for managing the alarms and faults bubble and troubleshooting. Processing of alarm and faults has thus turned into a tedious, manual and resource intensive activity. As mentioned above the Telecom field is evolving with the upward trend in growth where data network of a country is likely to contribute more than 5–7 percent of national GDP in the coming years. Having spent a long decade acquiring expertise in this field, it was intended to innovate a change in how alarms, faults, and device health are managed. Introducing a completely automated data intelligence alert and executive management that will track past

alarms and propose knowledge-based future troubleshooting concerns with graphical representation and correlation with historical device health. The demand-driven growth of data networks has significantly increased the need for management of Service Provider Network infrastructure. Service Providers network is geographically distributed over a very large area with varying topology and no. of devices which cannot be entirely manned by Technical Engineers. Presently Satellite and Surface earth observations are extensively used for passive and active fight's, assets and activities monitoring of the Telecom Sector Infrastructure providing huge amounts of data in geo-referenced format which is being processed and analyzed off-line with images being interpreted for change detection.

Equ 1: Linear Regression for Failure Prediction.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

$\hat{y}$: Predicted output (e.g., equipment health or time-to-failure)

x_i: Input features (e.g., temperature, voltage, usage time)

β_i: Learned model weights

2.2. Core Components of Telecom Infrastructure

The telecommunications infrastructure has a long history, resulting in large and complex networks over time. Performance evaluation and monitoring require sophisticated data architecture and engineering knowledge, especially in cellular networks. Optimizing networks to enhance customer experience and reduce costs comes with many complexities, but it is necessary for effective service execution. Quality of service in cellular networks is highly variable, leading to the idea that service should be tailored for each customer through customization in the context of data mining and predictive analytics . That entails proactive network management to identify when the service should be improved or problems handled before the customer realizes it. Real-time processing and expert knowledge are used to evaluate these KPIs and a well-defined approach for online processing. Traditional processing requires data cloning in a huge working architecture that is not manageable. Statically stored Hadoop partitioning and processing result in a processing time that is not real-time. Integration of expert knowledge is difficult, time-consuming, and prone to error. Solutions using a unified storage for systematic and centralized data storage are required. Big Data results from the revolution in radio technologies and an upsurge in the number of radio devices providing telecommunication services. Many telecommunications key components have advanced in parallel for a long time with the learning and knowledge area of experts spanning over decades. Seeking for performance evaluation, enhancement, or monitoring require profound analytical competence and knowledge of current operating conditions. Nevertheless, performance numbers and metrics, KPIs, logging-selected data, vary considerably in style and dimensionality from one component vendor to the other. In addition, implementation is missing for advanced machine learning techniques. This leads to suboptimal design and tuning decisions based on half-digested raw numbers. Important business questions therefore remain unanswered. The organizational focus in telecommunications shifted during the past decades from network toward business aspects. As a result, some learning capabilities were discarded in favor of quickly applicable off-the-shelf approaches. However, the racing tempo from one Innovation Day to the next left many questions regarding the industrial, economic and basic functioning of a few devices and systems ignored. Outside influences on performance are feared by technicians and decision makers alike, but not deeply enough studied to be ironically and systematically combated. Such a recommendation was quaint, subsequently instantly laughed at. It is recognized that phenomena

on a different scale impact the performance of the systems and must necessarily be included in the performance evaluation and tuning procedure.

3. Importance of Predictive Maintenance

Communication is a crucial technology that is evolving rapidly over the years with the exponential growth of devices that utilize wireless communication technology. Due to this unprecedented growth of telecommunication devices in the 5G era, biological frameworks are heavily relied on. Adoptive and purposefully designed APIs into Small Cells are treated as cell organs for the success of cellular infrastructure. Telecommunication networks are a major component of 5G, which is usually composed of Macrocells, Microcells and Small Cells. Each of the elements has recently experienced growth in scale and heterogeneous usage scenarios. Each telecom element has its own motivation, expectation, objective, and idea for its design. A portion of Low-Power Wide-Area cellular devices is designed to be ultra-low-cost, ultra-low-power, extremely small size, enormous scale, long coverage, and solely-connected mechanisms. Besides, Automatic Number Plate Detector applications are considered. Most micro-cells are co-api with the existing Macro-cell infrastructure; any upgrades would usually need to roll-out new Macro-cells. Modern upgrades on the existing Macro-cells will impact a small portion of affected devices only for a short period of time. Such significantly different designs need and thus should be coordinated differently.

As such, unique and distinct Minimum Performance Targets are formulated for each of the telecommunication scenarios. A TL is adopted to uncover the past and current TL for each telecom scenario under an exponential smoothing mechanism. The predictability of the future TL is considered. The risk is also quantified and ranked for each major decision. Finally, a policy mix is proposed for the solution. It is suggested to keep the current state for Macrocells and Microcells, while a strict training schedule is recommended for Small Cells and Automatic Number Plate Detectors. This thorough performance evaluation of the telecom infrastructure across multiple telecom scenarios is unprecedented, and this work is expected to shed further light on the performance evaluation of other complex public infrastructures.

3.1. Enhancing Equipment Reliability and Lifespan

With the expansion in the telecom industry with a focus on edge data centers, spending on site equipment maintenance has also significantly increased. Telecom site equipment comprises different electronic and electromechanical equipment that is critical for generating revenue for telecom operators. Failure of such mission-critical equipment would interrupt production, hence revenue, and lead to monetary loss from several hundred dollars up to millions. Repair and service contracts with third-party vendors are expensive, and contracted repair time is often too long and beyond acceptable levels. Preventive maintenance can cause high manual scheduling costs with no guarantee on equipment failure mitigation. Given that bottom-line costs are involved, a more proactive approach with high return on investment is needed, to enhance equipment reliability and lifespan, historical failure/root-cause analysis and condition monitoring technologies would be of great importance.

With attributes such as installation and maintenance dates, resulting costs and contracted service details for equipment, huge data is accumulated within telecom operators on equipment

maintenance and failures. However, little effort has been made in data mining and exploring such maintenance data for actionable business intelligence in telecom operators. Because of the massive amount of data, a testbed of a particular telecom operator with base station, microwave, data center, IP core and core network equipment is chosen as a research target. The 7 W's of such maintenance data, on the evolution of telecom infrastructure sites/equipment in the radio access network, transport network and core network from procurement, installation, service, upgrades and decommissioning is studied.

The tedious data preparation, processing and visualization tasks enable the building of a comprehensive big data analytic framework for unfilled research avenues in a telecom operator. Grouping of equipment for maintenance data extraction, based on attributes such as vendor, country of origin, maintenance contracts, series and metrics, is investigated. An EM group-based point-trajectory formulation for telecom maintenance expenses forecasting is proposed, along with a K-Means clustering approach. Holes and wastage in preventative maintenance are studied with Decision Tree clustering process, and actionable exploratory knowledge is derived to drive next-year schedules and resource allocation. Root-cause analysis on equipment failures with complex interdependencies and Bayesian network learning and inference is proposed.

Fig 3: Equipment Reliability and Lifespan

3.2. Reducing Operational Costs and Downtime

Telecom service providers capture user voice and data and need to store and post-process or pre-process them. Infrastructure providers build massive data centers full of servers and routers to control the distributed systems, and designers provide user equipment. Traffic compression, voice scraping, and other tasks related to the call need processing. The information is then stored in a format that allows fast access and low latency in the result. Furthermore, the Appearance Controller creates a DTMF signal for each request route-dependent upon the destination number.

Small-scale complaints are mapped to different sources with multiple transformer models; event contexts are created as input features, and the model returns the root cause in a user-friendly way. Labeling the complaint may take time. The system can draw the most probable root cause first and give definitive candidates to real-world engineers. Human-in-the-loop promotes people retraining the model better. The above methods, using neural network models, improve many industry NLP

engineering tasks and are coping with growth energetically. Nevertheless, experiments with telecom data have not been made public in a preferred shareable format. The development of architecture will stop for now except for some efficiency improvements (on sharing embeddings and cached potential parts of models) to accommodate all classes. Furthermore, a study on model training tasks should come next with some good benchmark baseline models.

Asset management in telecom points to querying the asset system. Using available sources to create good initial assets and filtering inflow bad ones is required, while classical methods usually rely on a lot of query rules. Instead, a set of feasibility rules are designed, among which most are no duplicates. Attention with hard negative mining or spanning graphs might produce better filtering results, as each node is a representation of a certain aspect .

4. Machine Learning Fundamentals

AI and ML are the driving forces behind new-age applications in every domain and play a crucial role in the transformation of technologies. A dataset or information obtained from ML surroundings can help organizations identify information, patterns, and scenarios needed in decision-making, business planning, area analysis, and telecommunications engineering. For classification, ML is generally perceived as a toolbox. However, it offers a much larger and generative way of working with the data, such as semi-supervised clustering and performance optimization. It encompasses all the skimpy gadgets and offers scopes. Understanding this working framework can enhance applications for potential solutions, which is often neglected in tutorials or classes.

ML is generally thought of as a universal toolbox, able to assist with tools for classification problems and regression tasks. ML necessitates a software ecosystem, including data monitoring and transformation, model selection and optimization, performance evaluation, and model integration. The first step concerns the construction of the models and focuses on the training phase. The data have to be pre-processed to remove or normalize outliers, as well as handle missing data. The next step concerns data transformation, where relevant data features are selected, redundancies are minimized, and data formats adapted. Model selection is done online. The most important steps of model selection and behavior learning are carried out in tandem. Initial guess models are selected and then trained on a small proportion of the data. Thereafter, more data should be sampled in order to decide on model decisions (all models allocated the same data). After training, some data is sampled for validation; a proportion of the validation data is retained to serve as a test set, and the rest is used to independently verify the rules constructed in trained models. In this regard, it is crucial to decide when enough experimental data has been collected, ensuring that the ML models produced are representative of the nature of the data. The goal of validation is to ensure that ML models are not over- or underfitting the observed data.

ML offers enumerable approaches depending on the objectives of the learning task. Of three broad categories, learning to think tackles the problem of truly understanding pattern recognition and identification from the data samples (clustering). More often than not, this process of learning is invoked and interpreted only after raw data comes in. While very few datasets are tagged, the need to estimate the properties of the data randomly becomes of higher priority. Therefore, this category of ML is needed to devote meaningful time to the understanding of the universe from the

anomalous data collected. Targeting induction learning, diverse types of approaches generate information or rules supported by data. Some of these models can flexibly switch between rules, thus requiring complex internal structures in learning.

Equ 2: Logistic Regression for Binary Failure Classification.

$$P(y=1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}$$

- $P(y=1|x)$: Probability of failure
- Useful for threshold-based maintenance alerts

4.1. Key Components of a Machine Learning Model Predictive models are mathematical models designed to provide a notion of the state of a system in the future, whether it is in a few seconds or days. Statistical models, signal processing, and machine learning (ML) models fall under this category. The output of the models is a score or a decision that feeds into actions or an event that drives the ML models. Actions are usually taken by trained operators. Whereas anomaly scores exist for automated or operator-initiated alerts to prevent further degradations of measured variables. This section focuses on machine learning models because of their effectiveness in dealing with massive datasets with numerous measurable variables. Generally, predictive models accept and process past data, i.e., measured variables, behavior variables, as well as contextual information, and predict the future state of the relevant variable.

When deploying ML models in production systems, the focus is mainly on the signal acquisition and model training aspects. Once the model is trained and deployed, it continuously operates on the data stream. However, new data can be collected, which is most likely different from the training data (data drift). Also, new data increments to the datasets can degrade the performance of the model over time and require to be retrained and related query changes (concept drift). Collecting incorrect datasets, such as operation mode changes, input space changes, or unanticipated anomalies, can lead to unexpected model outputs. Robustness and performance tracking are crucial to ensure the viability of ML models in production. These mechanisms frequently check the behavior of ML models deployed in the real world. A performance monitoring system automatically tracks the changes in model performance. By measuring the model performance using metrics, it can automatically alert operators regarding the degradation of model performance in the production environment. Here are the six KPIs frequently used to monitor telecom workloads. These KPIs can be used for the performance tracking of other ML models.

Lose of message metric: This metric provides the total number of lost messages, i.e., predictions that could not be calculated due to the absence of data regarding the input relevant variables in the streaming data. Monitored variables values comparison metric: This metric compares the static monitored variable values obtained online and the reference values (monitored variable values at the time of model training) offline. Hence it allows recognizing possible abrupt changes in the operating conditions of the monitored systems.

5. Types of Machine Learning Models

Different types of machine learning models are used to predict telecommunication indicators and to detect and predict maintenance tickets in an automated way. A big part of the models has been

used specifically in different industrial branches or sectors with different datasets. In order to adapt them to the telecommunications sector and infrastructure, models have been selected based on their features, applicability and outcomes.

Regression Models and Approaches: Ticket count models belong to the regression models, as the expected value of the output variable is predicted in this case. ML regression approaches are the following: autoregressive and Auto Regressive Integrated Moving Average, Exponential Smoothing, Monte Carlo method, LSTM, and Random Forest. Autoregression is a linear regression model which models the time series yt as a function of its own previous values yt-1, yt-2…yt-p. An ARIMA model differs from an autoregression model in that it can include differencing which can make a non-stationary time series more stationary, as well as variables that model short-term deviations from average behaviour. Exponential Smoothing models time series as a function of its own previous values, using exponential weights on previous values, with the "forgetting factor" in this case leading to the "smoothing factor". The Monte Carlo method includes random sampling and simulation of a stochastic model. It can be employed to generate a sequence of "worst case" scenarios for planning purposes. An LSTM model is based on a recurrent neural network that can learn long-term dependencies and produce output values with a temporal delay. Random Forest is an ensemble learning method for classification and regression. It constructs a multitude of decision trees and merges the outcomes.

Classification Models and Approaches: Maintenance or troubleshooting ticket models belong to the classification models as one or more categories are predicted for the output variable (i.e. fault type, fault group). Classification approaches model the probability of a class assignment of a data sample based on textual features. Naive Bayes and Multinomial Naïve Bayes models belong to these group of models as well. Additionally, tree-based classification algorithms produce a model of decisions based on input features to classify a data sample. Decision Tree, Random Forest, and XGBoost are among the state-of-the-art tree-based classification algorithms.

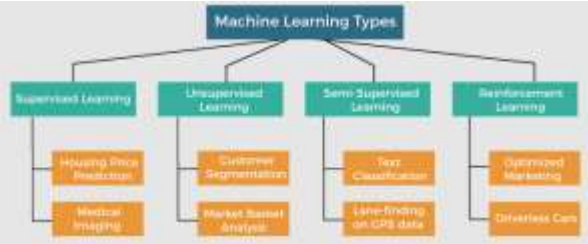

Fig 4: Types Of Machine Learning Techniques

5.1. Supervised Learning

The predictive maintenance models were built on a subset of data containing the time to breakdown of equipment along with other associated parameters such as age, scores, and so on. After the 80/20 split of the training and testing data, several supervised models were applied. After analyzing the metrics, logistic regression resulted in the most accurate prediction rate in the testing phase, and thus it was chosen as the primary model with which further analysis was carried out. All results regarding the logistic regression model were found to be very encouraging. It was found that with the right choice of thresholds, it is possible to catch 96% of breakdowns while impacting only 7% of non-breakdowns. It was found possible to analyze why machines were rated with high

breakdown probability scores and provide the owners with a more realistic set of enforceable actions. Summarily except for Logistic Regression, all other tested models resulted in something with relatively similar accuracy with the Random Forest Classifier exhibiting a much faster rate of execution . The choices of both the input features and model hyperparameters impact the final results immensely, and a much more precise tuning strategy is necessary for a more accurate prediction rate. With proper tuning, perhaps another model such as the Random Forest Classifier has a chance of being more accurate than the Logistic Regression. Hence it would be interesting to study the optimization of the input features and hyperparameters to configure and train better predictive maintenance models. The metric used to analyze the quality of predictions also extensively impacts the results. There are different ways to approach the metric score optimization, and perhaps using the F1 score, the best results would not drop by as much. It would also be interesting to study if different methods of mislabeling time to breakdowns would yield a better result.

5.2. Unsupervised Learning

In recent years, the telecommunication industry has become increasingly reliant on machine learning (ML) in many applications. This technology is expected to play an increasingly central role in future telecommunications networks. In particular, unsupervised machine learning (UML) will enable received performance self-diagnosis by identifying anomalies in a real-time manner. In this work, UML approaches are explored that can perform telecom infrastructure performance analysis while keeping the desired properties of low computational complexity and high online reactiveness. In particular, the approach can deduce performance anomalies in terms of time (i.e., fault, configuration, or network procedure changes), identification of affected areas, and a performance indicator root-cause determination . It can also find relationships in terms of logical dependencies between multiple performance indicators.

Unsupervised learning applications can enable the performance self-diagnosis of telecommunication infrastructure. Since telecom infrastructure generates performance monitoring data that can vary with orders of magnitudes in size and number of statistics, the complexity of processing the telemetry data grows. Proposed applications of UML aim at diagnosing performance anomalies and discovering performance relationships in real-time. First, an approach is developed to detect anomalous performance and find its root cause in terms of whether it is due to faults, configuration, or network procedure failures. The approach processes performance deformation multi-dimensional time-series data using dimensionality reduction. Anomalies are detected in near-constant time complexity using well-known techniques. Then, the root cause identification is cast as an anomaly explanation, where a diverse set of causative temporal events is deduced by examining whether a performance indicator’s change-time coincides with other indicators’ shift times. A robust ranking technique is proposed that captures the casualties of raw performance time series data while keeping execution complexity low. Second, a clustering-based approach is proposed to learn the relationship between two performance indicators in terms of whether they have some logical function between them.

5.3. Reinforcement Learning

In the search of new solutions to handle the increasing complexity and growing size of deployed networks, it is necessary to explore cost-effective strategies to optimize continuous improvements of their performance, while guaranteeing their availability. These strategies require modelling, monitoring, and control steps. Given that faults manifestation and corrective actions happen over time, the dynamical properties of the associated information streams are among the major challenges when processing the huge amounts of data produced and stored along these streams. Telecommunication equipment events, alarms and performance management monitoring, found in various Network Fault Management systems, represent good candidates to be investigated. A Reinforcement Learning (RL) framework to the more generic Network Maintenance Problem is proposed. This framework allows for a sequence of optimized actions over time, as well as the definition of a horizon for them to take effect. Deep reinforcement learning, which combines classic RL reinforcement learning with deep learning, has been a recent trend in machine learning, achieving very good performance in training intelligent agents to play games. Together with the necessity of computationally feasible solutions to the Network Performance Optimization Problem, such technology seems promising to also tackle the long sequences of discrete actions that need to be chosen for telecommunication equipment maintenance.

6. Data Collection and Preprocessing

As a key step for most machine learning models, data collection and data preprocessing will determine the quality of these models. Thus, it is important to have objective criteria to evaluate incoming data to avoid unexpected data quality issues. To this end, the data collection and preprocessing tasks are described below in detail.

Two types of data are acquired for this study. The first type includes telemetry data that are generated at a base station and/or a cell level every five minutes. It contains infrastructure performance metrics that are collected every five minutes. In the preprocessing stage, it is important to have a placeholder for each collected data file to avoid future data quality issues. For this task, a feedback mechanism is needed to plan an action in case the questions are not answered with a specified time period, i.e., unavailability of the data within expected time frames. In this framework, a historical file containing the same data is also retained to compare test run results with historical ones for expected data values. Furthermore, their associated metadata, e.g., for indexation purposes and time metrics, are also collected each time.

The second type of data, alarm data, indicates when a problem occurred in the infrastructure. Together with telemetry data, it provides a valuable use case for predictive maintenance. Alarms are logged into a database and are indexed by their identification codes (IDs). They are collected daily and, in each run, the corresponding new data are processed. This is done by taking the earliest time of each relevant alarms and retrieving the last 24 hours of telemetry data. Each telemetry data reflects the performance of some infrastructure elements. Therefore, as the locating ID of an alarm is unknown, the appropriate IDs must be found using the alarm names. This is done by looking at the latest alarm collected previously and examining a subset of elements in the predisposed cells.

Data preprocessing steps consist of cleaning, filtering, and merging the collected telemetry and alarm data to train predictive models. Missing values account for the most data quality issues in this data. Each data recording and value is associated with a clock-time stamped time on a base 10 min granularity. In a day, deleting a recording means losing values for each of its monitored metrics for 288 10 minute blocks. However, when telemetry data recording is missed, an alarm can also be deemed invalid. Thus, there are too many recordings and pruning invalid ones is a desired preprocessing step. Another anticipated data quality issue that needs to be handled is the outliers, specifically in the form of negative numbers in telemetry metrics. This is a domain-specific issue, since models are trained based on what insights were gained from the data. The telemetry data to build the baseline models must be decided before any change occurs in the preprocessing pipeline.

Equ 3: LSTM for Sequence Modeling.

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$$

h_t: Hidden state at time t

x_t: Input at time t

σ: Activation function

Captures patterns like degradation over time

6.1. Data Sources

Telecommunications service providers collect and analyze huge volumes of customer and network data every day from static fixed and dynamic machine-generated sources. As part of the proposed analysis, in preparation for machine learning models, there is a requirement to bring data from multiple sources together, in a manner that a new target variable could be defined and calculated independently from any past calculations. Therefore, Dataframe is designed as a stacked transaction structure that holds transformation rules. Data definition is defined as business readable data structure and information system defining unified analytics. As a consequence, DataNode is a Transformation object holding some data related constraints. A chain of transformation rules based on Type and Implementation could increment potentially infinite/set transformation. Internal edge weight defines recursive and sequential transition types. The clear direct connection to DataNode Target and Source is distinguished by the DataType.

In Telecom, business Key Performance indicators versus time models can often be made repeated time series augmentation based. Such functional models are restricted by present and past customer stream behavior or their time length. Therefore, ExternalType transformation limits its possible newly created grinding irrespective of affecting any historical Transformed commercial variables. Machine learning models outperform existing statistical and numerical model performance on real target variables, and on modelling at least their past error is in deviation in Telecom.

In telecommunication domains generate huge volumes of concise, clean PCI, IP, and SIP protocol state logs storing the source end and the integrity check of message flows. Those multi-sourced logs hold information about all existing Machine to Machine Action Upon Events and the corresponding analytic base table preparation. A universal model creation can be shown stacking protocol-aware trees with indefinite tree depth or complex model depth. Recent Machine Learning

model performance surpasses existing statistical models as the volume, complexity, and dimensions of input data increases.

6.2. Data Cleaning Techniques

The data is cleaned according to the following procedures as shown in Table. Entries which are not Feature engineerable: Such data are irrelevant data for the goal of predictive maintenance. Such irrelevant data is deleted. Useless or Neutral data: There existed some entries which possessed a uniform value for all the prior timestamps up to the present. Such entries guess the same value for all of the former timestamps without prediction. Hence entries produced uniform predictions in this respect are deleted. Entries with missing values: If a time step(s) data is missing for an entry, it will not be possible for that entry to predict it using regression and hence the entry will be deleted. Noise Removal: Some entries are showing abnormal behavior. It is either due to communication issues or device self-damage or a possible cyber-attack. Some entries in such cases tend to chromatographically spread be at unjustifiable ranges. Hence such entries are eliminated after examining its values intelligently. Deleting redundant data: There exist some entries of data per hour (60 minute). Generally speaking, the devices pinging continuously generate such data. It is either useless or of negligible value for predictive maintenance. Such data is hence deleted after examination. Aggregation: Data aggregation is most often desired as the devices ping hourly. But a data-aggregation function of hourly mean (or median) for each hour of the day (24 bin classification) is provided to run early experiments. Changing in the inspection time range every 14 days keeps varying seasonally. Therefore this is to avoid possible overfitting on the data captured within only one season (even though telecom infrastructures follow a regular standard). Each day of 24 hours' data is kept in bagged format and employed in a daily round-robin sampling manner .

Fig 5: Data Cleaning Techniques

6.3. Feature Engineering

As mentioned, features are of crucial importance for the models and on how they perform regarding QoS estimation . The general approach is to strive for the feature set that performs best. However, even at this early stage, one can recognize some early indications for the assessment of how well low-cost features can perform. Therefore, this list is thought to be very extensive and can be focused throughout the process. Studies have shown that part of this set of features carries most of the predictive power. While this was mostly based on knowledge of the area, it is desired to show this behavior regarding short PNG series. Another aspect of the feature set is its individual characteristics. Several factors, among others, are relevant. If the feature itself contains enough

information, what type of noise is it subjected to, can it be recreated, and how costly is it? A detailed exploration of the question of consistency is yet to be done. The information displayed within the features was analyzed to assure the building of a sound feature set. Hereby, the full feature set already displayed a lot of potential, outperforming even full-size bands of DX. Nonetheless, again relying on the knowledge of the area, it was shown that this feature set can be distilled even further. Part of this work was built on literature showing which features would emphasize more non-linearity and thus provide better predictive power. As of today, the selected features based on this literature displayed the best accuracy scores, emphasizing that even decreased costs across the board can lead to superior predictive capabilities.

7. Training and Validation of Models

To evaluate the trained models, three per-case datasets were defined: The training set with the first 1200 time series, the in-the-sample dataset with the last 60 time series, and the out-of-sample dataset with the one-hour allocation of the in-the-sample. The labels were prepared in advance for all the data given training data with calculated accuracy. The evaluation got MME values for all the models on these datasets. As there were cases with very long time series fragments to be analyzed, it was decided to test how the proposed models performed on converted velocity distribution to fraction format. Testing is focused on evaluating the per-case evaluation for the out-of-sample dataset. The whole validation pipeline should be tested at first. During the training, labels were assigned based on the behavior on the in-order sequence of the incoming time series.

Quantification of the prediction capability of a trained model is performed by imitating the system-wise allocation of the future time series for which performance should be quantified while waiting for many-labeled cases. Every day, the last hour per-cases were predicted. ML and DL estimation were performed in a sliding-window fashion, with the window size of 60 min. It was decided to predict malfunctions 30 min with 30 min. Before predictions of malfunctions, the training of the Zero-Algorithm and all the proposed models should end. To analyze the algorithm's prediction capability, 20 out of 265 system days were selected as test sets. Detailed inspection of the oracles and per-case predictions for these cases should be provided. To select the cases to be analyzed and graphically illustrate the differences between label and prediction ground truth, the selected test set should be reviewed. The selected false cases should illustrate the worst performance of each α-Algorithm. It was decided to predict malfunctions at t + 30 min from September 24. To evaluate the forecasting horizon of the trained models, tests were performed to quantify the prediction performance with different prediction horizons of t + 30 min, t + 60 min, and t + 90 min.

It was decided to compare the per-case evaluation measure MME and σ metric on both cases with uniform and non-uniform PPLs. The main focus was focused on examining the cases with the mZ-vote and t + 30 min prediction horizon. Alternative models for predicting the injection molding machinery faults using data from IoT sensors were compared. The results of the competing models were analyzed to characterize the differences between them regarding their training workload and performance in terms of accuracy and effort on monitoring the predicted system status with models of different types, complexities, and performances.

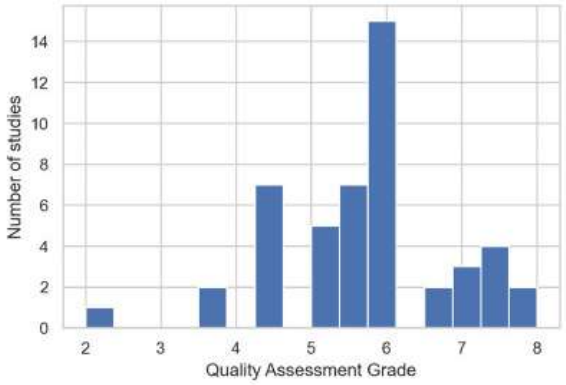

Fig 6: Predictive Maintenance in Industry

7.1. Training Strategies

The strategies discussed may be implemented to enhance the generalizability and robustness of machine learning models with a variety of architectures for different components of the countrywide telecom infrastructure. More broadly these strategies can be adopted for any distributed network of complex systems running in heterogeneous and dynamic environments while creating large streams of data. It is hoped that holistic failure prevention can be achieved for the world's telecom infrastructures running in a highly competitive yet fast-growing communication market if a greatly flexible, adaptive, and generalizable predictive approach can be achieved. Such an approach can then be extended to many other application domains characterized by distributed systems, dynamic environments, and wide ranges of input features producing data streams, such as transportation system, power grid, etc.

Adaptive learning strategies may be proposed to help build an efficient prediction model for an infrastructure having no fault history or limited fault data. Complex learning architectures may be deployed to provide accurate prediction for the small subset of predictors as compared to the very large set of background signals during normal operation. Techniques like attention or multi-scale learning may be incorporated to handle temporal variation of the input data. This can also be deemed as a temporal convolutional strategy to deal with very long input signals leading to excessive model parameters if all data are summed into a very high-dimensional input vector. It can be straightforwardly extended to a sequence of predictive maintenance as compared to one-shot prediction of faults.

Transfer Learning methods may be investigated to handle the scenario when a new infrastructure comes into operational state. The architecture and hyperparameters being the same, Partial Transfer may be achieved to learn about the new infrastructure from the data and model obtained through training on a different but similar infrastructure. For instance, considering the two machine architectures having a deeper learning architecture for model L2 and a shallow architecture for model L1, a L1 > L2 pre-trained model, can be executed on by level-lift prediction to a L2 transfer model. To learn the predictive part from L1's training via techniques is to abandon higher weights of L1's model and retrain parameters related to L2's learning outputs. Here, knowledge obtained from a general architecture can be transferred to a more complex specific model through matching the input-output formula of the two architectures. Unmatched parameters are set to be ignored via

appropriately moved weight initializing or layer size computation. Transfer learning procedures can be initially tested on a simple toy problem based on timer estimation regards with learning data of incremental hardware. It can also be explained as a small-scale exhaustive hyperparameter optimization process applied on a limited compute resource in designing a whole chain beforehand on a large setup.

7.2. Cross-Validation Techniques

The statistical performance evaluation of quantitative models is a critical step in the modeling process, providing insight into the model's generalization power. Several models that involve a set of model parameters or weights must be tuned in advance to achieve adequate performance. The most common approach for performance estimation and tuning parameters is cross-validation (CV) with a training set and a test set. In this way, one can obtain an unbiased estimate of model performance and also select the best hyper-parameters Θ. Apart from CV, there are other variations to choose model hyper-parameters. These include using multiple, fixed partitions of the training set to estimate performance and tune parameters and using the test set to simultaneously estimate performance and tune parameters.

Researchers emphasize the differences between these methods in terms of how unbiased or informative performance estimates are, how well model parameters can be interpreted, and how tunability of hyper-parameters is affected. CV and testing estimates of performance are unbiased, which is particularly important if the performance estimate is to be used as a quality measure for model comparison. In contrast, CV and testing will be predicted to generate the worst performance. This holds for both unbiased estimates of performance and lower average performance when those estimates are biased. The other two methods, cross-validation with a fixed test set and cross-testing, are predicted to yield estimates of performance between the two extremes. CV and the training set are predicted to yield estimates of performance that are positively biased, while withhold testing will be negative. Common to both will be methods yielding less interpretable choices of weights and parameters.

Unlike balance testing with datasets that are taken from an identical ensemble, CV is sensitive to how the data were derived. In the extreme case, prediction performance is expected to rise with training set size for a fixed, correctly selected, model. Since models with many fewer parameters can be fit to training data more closely, performance is predicted to exceed testing performance for that model. Fixing the assigned training set, and generating larger samples for testing, is expected to yield greater differences in performance. Cross-validation with data sets that are of different training/test sizes, without different data generation mechanisms, is predicted to yield greater results than with data that differ in this regard. Given multiple models, even optimal tuning can fail to yield genuinely comparable performance. In such cases, fixed partitions of sets are predicted to yield better performance comparison. Evaluation methods are usually considered to be meant solely for one of two purposes: estimating how well a given model generalizes to withhold data drawn from the same process (testing), or assessing what model to choose from a space of candidate models on the basis of with-hold dataset performance (model selection).

8. Performance Metrics for Evaluation

Data quality, loading, and sampling impact the performance of data-driven Machine Learning (ML) models. Therefore, it is imperative to perform a detailed analysis of the data quality, treatment, and its effect on the resulting ML model. In predictive maintenance, a number of metrics are already defined, including false positive rate, false negative rate, precision, recall, and f1-score. These metrics are, however, generic and not optimized for the telecom infrastructure. Existing performance metrics for the telecom space mostly comprise simple average KPIs like average throughput, jitter, or average latency. In this study a new set of metrics is being proposed that addresses the telecom space and also takes the data treatment and insights gained into account.

Overall Model Performance Use case independent requirements include model performance metrics (True Positive Rate (TPR), False Positive Rate (FPR), precision, recall, and f1-score) . Model performance data points estimate the fraction of true observations each of the states (faulty, watchful, or healthy) that successfully trigger each class of indicators, i.e. TPR is defined as R×R/(R+F), where R and F are the counts of truly positive and falsely identified positive observations respectively. The implementation of these performance metrics visualizes a water-filling shape diagram where each of the states of the telecom equipment (faulty, watchful, or healthy) forms a corner. Further, the telecom infrastructure is large, serves hundreds of cities and millions of connections. Therefore before assessing a new model or data, data from a system level is aggregated to the core layer of telecom architecture (RAN and core independently). Core nodal MGs returned by a selected model are highlighted, which should lead to a more careful investigation.

8.1. Accuracy and Precision

Current telecommunications networks are based on a complex set of machines, each capable of rendering tens or hundreds of disparate services. Over the last twenty years, network availability has been pushed beyond 99.99%. Each service is the output of a discrete set of machines, many of which are independent of one another. Altering the service as ordinarily provided means considerable changes to the machine network, particularly as machines age, fail, and are replaced with different vendors. The result is a multitude of distinct machine states, each with unique combinations of operating and potential failure conditions. After time service is restored, with considerable human intervention. These inherent rat-holes to be exploited by artificial intelligence.

Telecommunications infrastructure consists of both critical machines that must be supervised and highly redundant machines that can be discarded. Historical data consists of observations at reduced time resolution that are largely nominal, with wakes of alarms having noted fierceness and idea. The infrastructure consists of a mix of machines: some are quite new and just come on line, some are very old and not well understood, and some are stored drums. It is crucial to collect and analyze the relevant data to completely understand state and path. Also, the ranking of importance and alignment with failure records should be established.

Suspicious observations typically introduce pre-processor features that highlight signals to be trained on. New machine states and previously unobserved but possible behaviors should be allowed within the system. With regard to state of the art network topologies, using hierarchical

feature selectors should accomplish it. By building up seeds of failures and tracking over time, classes of possible faults can be defined. Simple timed based metrics across machines can prune the ensemble and bring them to a common base.

8.2. Recall and F1 Score

Recall is a measure of sensitivity, which for a two-class classification scheme is defined as the ratio of correctly predicted failures to the sum of actual failures. It is computed as (TP / (TP + FN)), where TP refers to the number of true positives, the number of actual state changes that are predicted as such, and FN refers to false negatives, the number of missed failures. The recall metric allows the algorithm to emphasize sensitivity and to predict a class more than is strictly required. The in-sample recall rates as a function of class thresholds, ensembles, and history length is shown. The strong decline in recall rates as the decision thresholds increase fits with the development and application of more general ensembles of models. Through careful feature group selection and hyperparameter tuning, fixed and shifting time windows can be used, with thresholds giving equal improvement across all class sizes. In general, as the number of classes grows, the sensitivity to failure events is lost. This result relates to the metrics that discriminate more finely by constructing metrics that include the total and class-specific F1 scores. Precision is a measure of specificity, which for a two-class classification scheme is defined as the ratio of correctly predicted failures to the total number of predicted failures. It is computed as (TP / (TP + FP)), where FP refers to false positives, the number of normal states that are predicted as failures. Precision indicates how conservative the algorithm is in labeling states as a class. A low precision rate indicates that there is high acceptance of likely true positives together with many false-positives. Precision rates are of less concern in the telecom domain than recall rates, as it is better to automatically check hundreds of false failures rather than miss a few predicted faults.

8.3. ROC-AUC

To draw the ROC curve for a binary classifier, two different classes of examples (i.e., negative and positive) are necessary, for each of which samples have to be observed as either, and for which truth labels are known in order to calculate predictions. If the classifier under study returns a probability or score for each of the predictions, the ROC curve can be easily plotted by the Receiver Operating Characteristic methodology. Generally, when plots of the ROC curves are examined, the performance of several models can be compared using the areas under the curves (AUC) . The AUC is a widely used evaluation measure in the ROC analysis which is equal to the expected probability that the classifier scores higher on a random positive than on a random negative.

With the traditional methods, that assumes the word occurrence is conditionally independent given any class label, the AUC computation may have a time complexity of O(n2) which is not feasible with large datasets having large numbers of documents. To therefore avoid this unimproved calculation burden, a random sampling method was proposed to explain the AUC statistics which, however, cannot be incorporated into the standard optimization framework for model training. To bridge this gap, an unbiased estimation of the AUC and Random Sampling Gradient (RSG) approaches are presented for reduced complexity estimation and efficient computation with O(m+ n log(n)) time and O(m+n) space, where m is number of positive documents and n is size of

negative document set. Moreover, by adjusting collection and processing of negative documents, an essentially exact greedy sampling method can be employed in reducing the complexity an order of magnitude.

Experimental results have shown that the proposed methods can achieve great speed-ups with high estimation accuracy, much better fitting performance than traditional methods, and do favor the convergence of model training. Finally, it is noteworthy that the AUC estimator with a sampling size as low as 104 achieves better performance than existing AUC optimizers.

9. Case Studies in Predictive Maintenance

Telecom networks are subject to the effects of prolonged overload. Shared, multi-purpose equipment or a single equipment malfunction can lead to the whole network being impacted. Not always inoperable (or slowly executable), equipment performance can degrade in time, leading to one or several performance metrics being violated, which will affect customer QoE. Service providers need tools to identify early on the performance degradation of equipment in a network before customer QoE gets impacted . This study proposes building a full causal inference framework for telecom performance analysis.

Each telecom equipment is submitted to a set of performance statistics per monitored metric for a given period of time (typically, the past 96 measuring minutes monitor one-day performance). Each measuring comes from a monitored equipment thesis or a set of performance statistics from a controlled equipment’s thesises. The latter determines the threshold value as well as the number of missing measuring(s) needed to provide valid analyses concerning operational types of equipment. Because all statistics influence performance a priori, a graphical model is built and subsequent analyses are applied on each species’ model. The proposed solution, Vision, is a Swiss Knife for telecom performance statistical analyses, conveying the graphical model, GUI-based performance contribution graph construction, and the relevant user-defined scope of interest.

The solution is validated on real-life telecom performance statistics in the FTTH equipment of an international service provider. A voice-over-DSL equipment in a telecom service provider is analyzed as a context aware case study to evaluate the tool usability and track performance propagation paths. Lessons learned and the impact of the proposed techniques in telecom are discussed.

9.1. Case Study 1: Network Equipment

This chapter presents two representative case studies for applying the proposed schemes for intelligent predictive insights to predictive maintenance and performance optimization in Integrated Fixed and Mobile Network (IFMN) applications. These schemes can be generalized to wider applications in other telecom infrastructure elements, thus expediting use-case development. The first case study is related to the telecom infrastructure maintenance in predicting equipment faults, while the second case study focuses on quality of experience (QoE) in video streaming service over telecom infrastructure optimization.

The stable operation of telecom network equipment is a prerequisite for ensuring telecom service continuity and consequently customer satisfaction. In the context of the rapid deployment of various telecom network equipment and services, including traditional and Cloud RAN with increasing complexity, telecom network operations have become an increasingly sophisticated business endeavor. Cost efficiency, quality control, and customer satisfaction have increased in challenges due to the de-monopolization and the diversification of telecom infrastructure ownership, which adds an additional layer of complexity. Therefore, intelligent maintenance of such network equipment is highly sought after and represents a dominant proportion in the cost of telecom operation. Preventive maintenance, corrective maintenance, and predictive maintenance are three types of common maintenance.

With the rapid advancement of sensors in telecom network equipment tracking operation status such as input voltage, air conditioning unit rectification current, rectification equipment work mode, ventilation monitoring, etc., a fully sensorized equipment status observation infrastructure capable of generating granular real-time equipment operation data is formed. However, through large equipment data repositories and predictive modelling approaches based on them, feature extraction through temporal and spectral filtering and subsequent machine learning model based fault class regression, they only work on resolution time scale orders from minutes to hours due to their reliance on batch mode predictive model processing, which is incapable of real-time processing on the order of milliseconds to seconds to prevent faults.

9.2. Case Study 2: Base Stations

In cellular infrastructures, it is hard to have the right allocation of crews, because the failures of a particular site are not easy to predict. With historical failure data, it is conceivable to train machine learning models that can output failure probabilities for sites in the network. This can boost resource allocation efficiency and increase the network quality. The prediction capability of Random Forest and XGBoost were evaluated on a dataset, which contains more than 3 years of historical data from 11,000 cells. The 5- and 10-days-ahead prediction accuracies were measured, by which Random Forest obtained satisfying results. The combination of an ARIMA prediction model and a Random Forest classification model is projected to not only pinpoint the major risks but also provide proper encapsulating measures.

The case study is conducted on the Target Variable 1, which refers to a predefined condition that the Key Performance Indicators of a particular site may deviate significantly (thresholds are chosen based on knowledge from domain experts). In the telecommunications infrastructure, the cellular base station is a fundamental component of the radio access network, which ensures the quality of service that the subscribers access. An essential observation is that unplanned outages can happen randomly, but after an outage is detected by the network engineers (or automatically), it may take time to pinpoint the causes and escalate the proper resources needed for corrective measures. Hence, the effectiveness of these resources cannot be guaranteed, especially when bursts of failures occur and they may affect a whole area in terms of QoS degradation. Failed cases which are resolved without resource usage after a certain time usually carry valuable information but are often neglected. With more than 3 years of accumulated operational data, one essential question is raised: Can data analytics techniques help diagnose the outage and point out the major alarms that are very likely to have corresponding outage causes? With this question, it can help increase

resource allocation efficiency and boost the network's health state, especially when new Advanced Features are deployed.

With more than 45,000 sites, the network of a Mobile Operator involves many human-reliant processes and is exposed to various failure causes. These failures can be divided into two types: service deletions (or planned outages) and not-service deletions (or unplanned outages). The former is often planned ahead in advance, while the latter can happen non-randomly and are often protracted. After this study was discussed with high-level engineers, both were considered as follows:

10. Challenges in Implementation

One of the most challenging issues in implementing machine learning models is handling the exponential growth in data. In telecom networks, base stations (eNBs) may produce several thousand KPIs, which can be further sliced across dimensions, generating over a million readings every minute. This scalability issue is exacerbated by the presence of legacy mechanisms, where models written in a different programming language will require additional overhead for preprocessing. The continuous monitoring of a performance metric to detect abrupt changes in behavior is also difficult, due to the need for updating production models, and the time taken to collect enough data for meaningful analysis.

Another significant hurdle during implementation is the lack of validation mechanisms to determine the accuracy of existing models. Telecom models may be constructed and evaluated in a contained test environment, but they can be exposed to anomalous behavior in the production system, leading to a loss in prediction accuracy. Producing additional performance metrics and feature engineering are basic steps that can help improve this, but it is vital to know when models need maintenance and are no longer fit for purpose. This highlights the importance of robust and easy-to-interpret performance metrics to evaluate credibility, adjust reactively to network changes, and help maintain high availability telecom systems that run 24/7 .

Another issue is the need for more explainable models that provide a clear understanding of their working mechanisms. Complex DNN architectures sometimes produce amazing results, but they tend to behave like black boxes and do not communicate how predictions are formed. This is an issue for automated tasks such as root cause analysis, as the models need to guide analysts through the reasoning process. To ensure that the maintenance procedure succeeds, an explanation model is needed that provides clear insights into how the system logically reached its conclusions. Models with precursor finding algorithms can be customized with configurations that enhance their interpretability. Using models with built-in explanations is also advisable. Despite ongoing research, appropriate interpretable models are often needed in a telecom setting.

10.1. Data Quality Issues

The advent of deep learning techniques has had a profound impact on the internet of things ecosystem, most notably in network performance management, where mobile broadband infrastructure must handle a growing volume of increasingly complex traffic types. Tools based on machine learning concepts can augment necessary human analysis and experience by providing

predictions, alerts, and suggestions that will enhance operations and service quality while handling many times the current infrastructure with fewer overheads. While algorithm frameworks and simulations resulting in good analysis and predictions can be devised, they do not guarantee success unless specific requirements are understood and met. Such requirements include inputs, derived metrics, model selection, training, and evaluation.

The telecom operator must assess its data foundations and ML algorithms. There is also a need to carefully assess algorithms for appropriate performance measurement to produce forecasting or labeling results that can assist any operational activities. This includes appropriate evaluation performance indicators and metrics banks, observing protocols, labeling segmentation definition and composition, and tagging preparation for supervised learning tasks.

Telecommunications operators must implement several processes as their first stage to involve in all basics of data preparation before ML applications can happen on the data they already have. Performance Labelling Segmentation Processes aggregate raw performance measurements into derived performance metrics at different levels of granularity, which are useful inputs for performance-related supervised learning tasks targeting network performance forecasting and anomaly detection. All initial post-processing steps such as filtering labels to a desired time horizon can be done in-house. Performance Segmentation Labelling is aggregate and non-generic ML applications where a pre-definition of the label variable is mandatory. Such system processes delineate an ML-based paradigm for both hydrocarbon production forecasting and control.

10.2. Integration with Existing Systems

As discussed in previous sections, the approach for predictive maintenance clearly focuses on maintenance and performance optimization based on the efficient utilization of predictive models. If the resulting models do not translate into either model or outcome, then the model does not serve its purpose regardless of how accurate it is. On the other hand, it is vital that wherever such a system is implemented, it either expands into or integrates with the existing information technology stack. This includes but is not limited to cloud-based service enablement platforms, data pipelines, temporary storage, and business-related systems. This section details the approach and considerations for such an implementation .

To evaluate the performance of the deployed approach, telemetry data that was previously part of the raw datasets must again be processed and once again filtered in terms of specific models. The telemetry filters then get applied against the production data obtained from the database-as-a-service. Once again this filtered data serves for the extraction of statistics, features, and target variables for the ML models and the state variables for the expert systems. The concerns of this section regard processing telemetry data gathered in an environment external to the previous locally run stack.

Telemetries provide state indicators reflecting physical or logical changes in the network elements which enable integer predictions of service-affecting failures. The goal is not to forecast a single numerical value but score these telemetries against ranges of predictions and assign a condition as being OK or NOT OK. The resulting states lower the chance of false negatives and create alarms for further disclosures for the network operations teams. The same concept can group the states

into an overall score that takes network segment arrangements, priorities, and criticalities into consideration. In order to base the state analysis on clear and interpretable criteria, expertise knowledge descriptions must be obtained from experts and specifically agreed upon.

11. Conclusion

The telecom Information and Communication Technology (ICT) sector is growing aggressively, but the expansion of this sector in terms of customer base burden is increasing steeply. Leading Telecom companies are looking at making certain improvements in their telecom infrastructures to efficiently cater to their growing customer base. Major telecom companies' infrastructures are made up of elements such as Chassis, Routers, Patching Frames, Optical Units, etc. In order to operate these elements accurately and efficiently, it is of utmost importance to look into their performance and maintenance actions. These telecom elements undergo normal wear and tear with time and need to be taken up for periodic maintenance to avoid step-mistakes which can lead to complete failure. Furthermore, there can be a sudden operational failure of some of the elements due to external reasons like natural calamities or man-made accidents. These unexpected failures account for a very huge loss not only financially but also a degradation in network quality which leads to loss of revenue.

Machine Learning (ML) models are built to satisfy the need of telecommunication companies by keeping a tab on infrastructure performance, which enhances the expected performance of the telecom elements, and predicting maintenance schedules of infrastructure elements, which enhances the reliability of the telecom elements thereby avoiding breakdown due to unusual behavior. These ML models address the two-fold objective of this research. Poor performing telecom elements are filtered using a Multi-Linear Regression model by parsing the Routing and Sonic log data of network use. These telecom elements are again analyzed using ‘SURF & K-Means Clustering’ models to provide a better understanding of the infrastructure so that corrective actions can be taken. Predicting maintenance tasks of these telecom elements is achieved using Extreme Gradient Boosting, Random Forest, Support Vector Classifier, and Light Gradient Boosting models.

Ultimately, a predictive maintenance approach is presented which depicts the importance of work history parameters for robustness and accuracy in predicting maintenance schedules. Six ML models are trained that give a very high F1 score which is very much appreciable for any prediction task. This predictive maintenance model enables the stakeholders to focus on probable candidates and educate and extract knowledge from them on what may have gone wrong and what corrective efforts can be made. Predictive maintenance is an important dimension under asset management as it detects issues with equipment and machinery before failure occurs. Many industries want to adopt predictive maintenance for their systems to save on maintenance costs and avoid failures. Deciding what predictive model to use and how to implement it are important questions every business should ask before jumping into their predictive deployment projects. This project will help identify the obstacles and trade-offs, which provide clarity of decision-making trail to enable higher success rates in predictive maintenance implementations.

11.1. Future Trends

With the penetration of 5G, the boom of the IoT, and the appreciation of virtual solutions, a new round of telecommunication technological revolution is around the corner. Establishing better performance networks at a lower cost, improving customer experience, increasing automation, and directing services intelligently are some of the challenges set by the telecommunication technological upgrade. Intelligent systems leveraging advanced analytics are critical for Telecommunications Service Providers (TSPs) to thrive in this new environment. This paper discusses the state of the art of machine learning applications for performance optimization and predictive maintenance of telecommunication networks. As the number of installed data sources increases, the leadership of advanced analytics could be deemed as a top competitive advantage. Time-related events in the telecommunication infrastructure such as alarms, counters, performance monitoring, and customer complaint tickets represent a vast array of data. Most of the knowledge assumed is either quantitative in nature and not proactively taken advantage of, or qualitative and stored in text blobs spread in multiple documents. On top of that, how to obtain the highest value from it is a current hot-topic of research with an important impact on decision-making.

The reduced model of the cellular-radio network used is denoted by a homogeneous Poisson point process of users and base stations, with independent micro-mobility. Model parameters are the arrival rate, contact rate, cell residence time, and transmission range. The traffic of educational establishments is shown to exhibit distinct patterns considered relevant to educational planning. Using empirical data, an analytical model providing insights into the impact of school traffic on the network is derived in a mean sense. By exploiting model parameters, the model can be used to explore the impact of cellular network design and traffic demand on educational traffic. Consequently, best practices for planning cellular networks to account for educational zones are suggested. The plans include the upgrade of base station equipment, increase in tower height, and/or reduction in cell-size, jointly pursued with feasible ways of economically deploying new cells, in order for the initial peak-to-average of generated and offered traffic to remain invariant. Alternative plans should also be pursued when the latter condition does not hold, for regulating offered traffic, i.e., pricing schemes, transmission power requirements for users, and basic traffic control mechanisms.

12. References

[1] Chava, K., Chakilam, C., Suura, S. R., & Recharla, M. (2021). Advancing Healthcare Innovation in 2021: Integrating AI, Digital Health Technologies, and Precision Medicine for Improved Patient Outcomes. Global Journal of Medical Case Reports, 1(1), 29–41. Retrieved from https://www.scipublications.com/journal/index.php/gjmcr/article/view/1294

[2] Nuka, S. T., Annapareddy, V. N., Koppolu, H. K. R., & Kannan, S. (2021). Advancements in Smart Medical and Industrial Devices: Enhancing Efficiency and Connectivity with High-Speed Telecom Networks. Open Journal of Medical Sciences, 1(1), 55–72. Retrieved from https://www.scipublications.com/journal/index.php/ojms/article/view/1295

[3] Avinash Pamisetty. (2021). A comparative study of cloud platforms for scalable infrastructure in food distribution supply chains. Journal of International Crisis and Risk

Communication Research , 68–86. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/2980

[4] Anil Lokesh Gadi. (2021). The Future of Automotive Mobility: Integrating Cloud-Based Connected Services for Sustainable and Autonomous Transportation. International Journal on Recent and Innovation Trends in Computing and Communication, 9(12), 179–187. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11557

[5] Balaji Adusupalli. (2021). Multi-Agent Advisory Networks: Redefining Insurance Consulting with Collaborative Agentic AI Systems. Journal of International Crisis and Risk Communication Research , 45–67. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/2969

[6] Singireddy, J., Dodda, A., Burugulla, J. K. R., Paleti, S., & Challa, K. (2021). Innovative Financial Technologies: Strengthening Compliance, Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures. Universal Journal of Finance and Economics, 1(1), 123–143. Retrieved from https://www.scipublications.com/journal/index.php/ujfe/article/view/1298

[7] Adusupalli, B., Singireddy, S., Sriram, H. K., Kaulwar, P. K., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks. Universal Journal of Finance and Economics, 1(1), 101–122. Retrieved from https://www.scipublications.com/journal/index.php/ujfe/article/view/1297

[8] Gadi, A. L., Kannan, S., Nandan, B. P., Komaragiri, V. B., & Singireddy, S. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization. Universal Journal of Finance and Economics, 1(1), 87–100. Retrieved from https://www.scipublications.com/journal/index.php/ujfe/article/view/1296

[9] Cloud Native Architecture for Scalable Fintech Applications with Real Time Payments. (2021). International Journal of Engineering and Computer Science, 10(12), 25501-25515. https://doi.org/10.18535/ijecs.v10i12.4654

[10] Pallav Kumar Kaulwar. (2021). From Code to Counsel: Deep Learning and Data Engineering Synergy for Intelligent Tax Strategy Generation. Journal of International Crisis and Risk Communication Research , 1–20. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/2967

[11] Chinta, P. C. R., & Katnapally, N. (2021). Neural Network-Based Risk Assessment for Cybersecurity in Big Data-Oriented ERP Infrastructures. Neural Network-Based Risk Assessment for Cybersecurity in Big Data-Oriented ERP Infrastructures.

[12] Katnapally, N., Chinta, P. C. R., Routhu, K. K., Velaga, V., Bodepudi, V., & Karaka, L. M. (2021). Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights. American Journal of Computing and Engineering, 4(2), 35-51.

[13] Routhu, K., Bodepudi, V., Jha, K. M., & Chinta, P. C. R. (2020). A Deep Learning Architectures for Enhancing Cyber Security Protocols in Big Data Integrated ERP Systems. Available at SSRN 5102662.

[14] Chinta, P. C. R., & Karaka, L. M.(2020). AGENTIC AI AND REINFORCEMENT LEARNING: TOWARDS MORE AUTONOMOUS AND ADAPTIVE AI SYSTEMS.