Regression Methods for Forecasting the State of Telecommunication Networks

Mustafa Mohialdeen
Middle Technical University
Baghdad, Iraq
Mustafamohy@mtu.edu.iq

Mushtaq Talib Al-Sharify
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
alsharify@uoalfarahidi.edu.iq

Yurii Khlaponin, Myroslava Vlasenko
Kyiv National University of Construction and Architecture
Kyiv, Ukraine
y.khlaponin@knuba.edu.ua,
vlasenko_mm@knuba.edu.ua

Mohammed Khodayer Hassan Al-Dulaimi
Al-Rafidain University College
Baghdad, Iraq
mohammed.khudhaer.elc@ruc.edu.iq

Hasan Harith Jameel Mahdi
Al-Turath University College
Baghdad, Iraq
hasan.harith@turath.edu.iq

Abstract — **Telecommunication networks play a critical role in the smooth data flow in today's digital world, making their steady operation essential. Because of the increased dependence on these networks, sophisticated monitoring and management systems are required to ensure continuous online connection and data sharing.**

This article aims to investigate the use of quantile and logical regression in predicting the performance of telecommunication networks. It seeks to build a prediction model based on real-time data to aid intelligent decision-making systems in network management.

The study employs machine learning methods to design an intelligent control system based on the autocorrelation of time-based variables. This system combines a communications network, a database of system attributes, a machine learning-based data processing system, and a decision-making system.

The intelligent system demonstrates the telecom network's real-time monitoring, analysis, and management capabilities. Its goal is to provide telecommunications operators with an efficient predictive model for improving network performance and resource allocation while addressing complex network dynamics, diverse data sources, forecasting accuracy, real-time decision-making, and model resilience.

Constructing a trustworthy mathematical model using regression techniques and machine learning dramatically improves the forecasting of telecommunication network conditions. This innovation is critical in allowing proactive management techniques and ensuring end-users get high-quality services.

I. Introduction

Modern life is hard to imagine without information technology. A set of numerous digital radio-electronic devices connected to telecommunication systems not only help in organising business processes at large enterprises but also deeply enter our daily lives. It is difficult to overestimate the importance of being in touch 24 hours a day and exchanging data at any second through networks consisting of modern mobile devices. Even though many electronic devices seem independent at first glance, upon closer examination, they collect, process and transmit information. Moreover, even if the device provides all the above functionality, it still needs to exchange information through telecommunication networks [1].

At the moment, information technology has gone far ahead, which makes it possible to transfer information between devices at a high level. However, due to the complex multifunctional nature of these devices, the issues of maintaining high efficiency of information exchange play a crucial role in information technology. Reliability is one of the main characteristics that determine the quality of the functioning of digital radio-electronic devices [2], designed solely to support the exchange of data between the remaining devices [3]. The operability and reliability of telecommunication systems and computer networks, consisting of radio-electronic devices for processing and transmitting information, will ultimately be characterised by the operability and reliability of the most inefficient device. Therefore, timely warning by identifying relevant ineffective devices is critical in analysing and monitoring such systems [4, 5].

Recently, applied intelligent technologies have been increasingly used to solve the problems of analysis and monitoring of telecommunication systems and computer networks. Such technologies are based on machine learning technologies, based on which decision-making systems are built for the intelligent control of complex distributed info-communication networks [6]. Machine learning algorithms designed for predictive use use the idea of autocorrelation between values over time. The general structure of the intelligent control system:

- telecommunications network;
- database of characteristics of the telecommunications system;
- data processing system based on machine learning technologies;
- decision-making system;

Intelligent systems can differ significantly in their functions, but they always contain these blocks to one degree or another [7, 8].

It is important to note that the main architectural feature that distinguishes the intelligent control system (Fig. 1.) from the one built according to the "traditional" scheme is the connection of storage mechanisms and intelligent data processing that characterise the operation of systems.

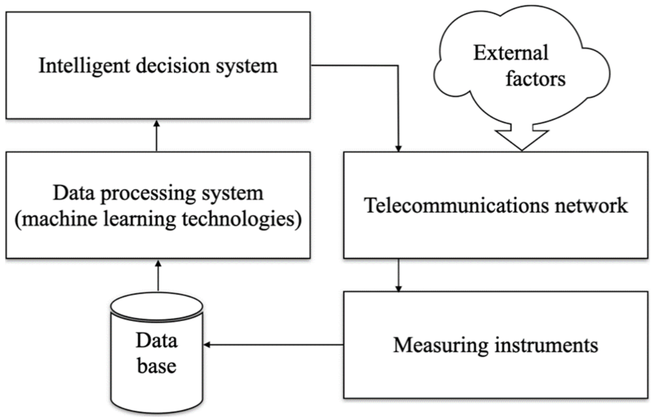

Fig. 1. General Structure of the Intelligent Control System

This feature of intelligent systems allows real-time monitoring, analysis and effective management of the telecommunications network.

A. *The Aim of the Article*

The article aims to create a reliable mathematical model for forecasting important indicators characterising the functioning of a telecommunication network. We want to equip telecommunications operators with an intelligent decision-making system to optimise network performance and resource allocation by developing this predictive model. This system will employ sophisticated regression algorithms and machine learning techniques to correctly estimate crucial network parameters, allowing proactive management and timely interventions to ensure end-users get high-quality services.

B. *Problem Statement*

Forecasting telecommunication network performance has significant problems that must be carefully considered:

1) Complex Network Dynamics: Telecommunications networks are dynamic systems impacted by various variables such as user behaviour, data flow, and network design. It is a big problem to capture and describe these complicated processes to construct an effective forecasting model.

2) Heterogeneous Data Sources: Telecommunication networks create massive amounts of data from many sources. Because of the variety of the data, advanced procedures are required to integrate and interpret the information to develop an all-encompassing predictive model.

3) Forecasting Accuracy: Providing accurate insights into the future status of communications networks requires high forecasting accuracy. Improving regression algorithms' accuracy and resolving possible limits are crucial for practical implementation.

4) Real-time Decision Making: Forecasting is critical for proactive network management in the fast-paced telecommunications business. A prediction model that can provide timely insights would enable intelligent decision-making systems to adapt to network difficulties quickly.

5) Model Robustness: To accommodate the dynamic nature of telecommunication networks, created regression models must be resilient in dealing with unanticipated events such as network outages or data abnormalities [9].

II. Literature Review

The field of telecommunication network management has seen notable progress, primarily motivated by the need to enhance network operating efficiency, ensure quality of service, and optimise resource allocation. To tackle these issues, scholars have investigated several regression techniques and machine learning methodologies for predicting the condition of telecommunication networks. The present literature review offers a complete synthesis of seminal works in this field, elucidating their significant contributions and valuable insights.

A noteworthy contribution has been made by Krinkin, Vodyaho, Kulikov, and Zhukova [10], who presented a unique methodology for predicting the states of telecommunication networks. This strategy involves analysing log patterns and using knowledge graph modelling. Log data analysis and knowledge graph modelling are employed in this novel approach to forecast network states appropriately. This approach demonstrates a potential pathway for enhancing network management by providing more predictability and proactively remedying issues.

The study performed by Attila and Vilmos [11]focused on traffic prediction in intelligent cities, particularly emphasising the significance of anticipatory traffic management. The research conducted by the authors offers significant contributions to the field of traffic prediction methodologies, specifically in the context of urban telecommunication networks. These insights could enhance resource allocation strategies and facilitate uninterrupted connections [12].

In their study, Zhang, Chuai, Gao, Li, Maimaiti, and Si [13] introduced a novel approach to predicting traffic patterns in urban wireless communication networks. The authors' methodology effectively tackles the specific obstacles encountered in urban settings, including those related to frequent movement and external disruptions. Consequently, their technique significantly improves the precision of traffic predictions, hence optimising network efficiency and boosting user satisfaction.

The study by Szostak, Włodarczyk, and Walkowiak [14] investigated several machine-learning techniques to predict optical network traffic. The research presented in this article demonstrates the practicality of using machine learning techniques in the context of optical networks, therefore establishing a foundation for implementing predictive maintenance strategies and optimising capacity planning processes.

The study conducted by Aibin [15] emphasised the traffic forecast in elastic optical networks, highlighting the significance of adaptive network management. By employing machine learning techniques, their methodology enables the efficient allocation of resources, hence maximising network usage.

In their study, Khan et al. [16] presented a novel approach to estimating the quality of transmission (QoT) in optical networks that has not been seen before. The authors emphasised the significance of machine learning techniques in guaranteeing dependable and effective visual communication.

Nguyen, Dlugolinsky, Tran, and García [17] investigated deep learning techniques in the context of proactive network monitoring and security protection. The idea conducted by the authors focuses on the urgent need for timely identification of potential threats and the development of network resilience, therefore making a valuable contribution to improving network security.

In their recent study, Yudin et al. [18] proposed a predictive model to evaluate the technical condition of communications networks. This sophisticated statistical data processing methodology facilitates the prediction of possible system breakdowns and optimises maintenance plans.

The existing part of the study about the prediction of telecommunication network conditions comprises various regression techniques and machine learning methodologies. The study mentioned above contributes to the advancement of the subject by providing a comprehensive range of regression and machine-learning approaches to enhance the forecasting of telecommunication network states. This improvement in forecasting capabilities has significant implications for network management and resource allocation [19]. The evolution of telecommunication networks necessitates incorporating novel techniques, which will serve as vital factors in guaranteeing their stability and effectiveness.

III. METHODOLOGY

Two separate machine-learning models were used in this investigation. The quantile regression approach was used first, followed by the logistic regression theory.

Predictive modelling using input features to estimate the chance of an event occurring is known as logistic regression. It is done by creating a dependent variable with two possible values: 1 for when the event occurred and 0 for when it did not. In addition, a collection of independent variables, sometimes known as features, predictors, or regressors, is considered. These independent variables are genuine values on which the likelihood of the dependent variable having a given value is calculated [20].

The basic assumption behind logistic regression is that the event's probability may be stated as a function of the independent variables. The model aims to identify the link between the independent variables and the likelihood of an event occurring. By analysing this relationship, logistic regression assists in developing predictions [21] about the probability of an event arising based on the input data.

Where $z = \theta^T x = \theta_1 x_1 + \cdots + \theta_n x_n$, x and θ column vectors of values of independent variables $x_1, x_2, \ldots, x_n$ and parameters (regression coefficients) - real numbers $\theta_1, \theta_2, \ldots, \theta_n$, respectively, $f(\mathrm{z})$ – the so-called logistic function (sometimes also called the sigmoid or logit function):

$$P\{y = |x\} = f(z)$$

Since it takes only the values 0 and 1, the probability of the first possible value is equal to:

$$P\{y = |x\} = 1 - f(z) = 1 - f(\theta^T x)$$

For brevity, the distribution function for the given can be written in the following form:

$$P\{y|x\} = f(\theta^T x)^y (1 - f(\theta^T x))^{1-y}, y \in \{0,1\}$$

A. Advantages of the quantile regression method

The practical requirement for more appropriate mathematical tools necessitates the development of innovative statistical estimate techniques. While other applicable regression techniques exist, the least squares approach is the most used. This technique provides us with substantial statistical insights by assuming that random mistakes have a regular (Gaussian [22]) distribution and are, in most situations, independent of one another. However, new estimating strategies must be explored to handle real-world complications and improve statistical modelling accuracy. However, alternative approaches also require strict assumptions about the type of distributions and other requirements regarding observations.

However, in practice, one often has to deal with more complex situations that do not fit into the standard assumptions of regression methods. Examples of such "violations" include:

1) The inaccuracy of setting the distribution that controls the observations in the sample so that the assumptions of classical regression models about the homogeneity of the sample or a sufficiently "beautiful" mechanism for the manifestation of homogeneity cannot be verified;
2) the presence of distributions with more "heavy tails" than the normal distribution, which necessitates the choice of estimation methods that give less weight to the extreme observed values or even a complete rejection of the least squares method;
3) the existence of a tiny fraction of "outliers," which are findings induced by some "noise" and are often difficult to discriminate based on previous knowledge, is one of the obstacles faced in the sample. Dealing with such outliers involves using approaches less sensitive to the "contamination" of the sample;
4) the dependence of the elements of the sample, which has a complex structure, so that it is difficult or even impossible to isolate and analyse (for instance, a covariance matrix).

In several practical problems that require a regression approach, classical regression methods are inoperable and do not allow drawing correct conclusions about the nature of the process under study or the behaviour of the object under consideration.

There have been repeated attempts to build alternative

$$f(z) = \frac{1}{1 + e^{-z}}$$

approaches. In particular, the so-called robust methods, which are resistant to deviations from the assumptions of the classical theory, have been actively developed.

One of these approaches is the quantile regression method [23], which replaces square deviations with absolute ones. It has several advantages:

- is resistant to "outliers," which are common in practical activities;
- does not need autonomy or weak dependency;
- enables you to make clear conclusions about the swings of the projected (estimated) indicator.

Thus, this method allows us to overcome the shortcomings of classical regression models, which are very sensitive to violations of their assumptions.

B. Selection of parameters

For the selection of parameters $\theta_1, \theta_2, \dots, \theta_n$, it is necessary to make a training sample consisting of sets of values of independent variables and the corresponding values of the dependent variable y. Formally, this is the set of pairs $\left(x^{(1)}, y^{(1)}\right), \cdots, \left(x^{(m)}, y^{(m)}\right)$ where $y^{(i)} \in \Re^n$ – vector of values of independent variables, and $y^{(i)} \in \{0,1\}$ – their corresponding value y. Each such pair is called a training example.

The maximum likelihood method is usually used, according to which the parameters are chosen θ, maximising the value of the likelihood function on the training sample:

$$\hat{\theta} = \arg\max_\theta L(\theta) = \arg\max_\theta \prod_{i=1}^{m} P\{y = y^i | x = x^i\}$$

Maximising the likelihood function is equivalent to maximising its logarithm:

$$\ln L(\theta) = \sum_{i=1}^{m} \log P\{y = y^i | x = x^i\} =$$

$$= \sum_{i=1}^{m} y^{(i)} \ln f\left(\theta^T x^{(i)}\right) + \left(1 - y^{(i)}\right) \ln(1 - f\left(\theta^T x^{(i)}\right))$$

For example, the gradient descent method can be applied to maximise this function. It consists of performing the following iterations, starting from some initial parameter value θ:

$$\theta := \theta + \alpha \nabla \ln L(\theta) =$$

$$= \theta + \alpha \sum_{i=1}^{m} \left(y^{(i)} - f\left(\theta^T x^{(i)}\right) x^{(i)}\right), \qquad \alpha > 0$$

C. The general quantile regression model

Let $(y_i,\ \boldsymbol{x}_i)$ – set of observations $(i = 1 \div n)$, where y_i – dependent variable in the regression equation, and $\boldsymbol{x}_i = (x_{i1} \cdots x_{im})$ – row vector of independent variables (covariate). Then, the model is given by the relation [24]

where $Quant_\theta(y_i|\boldsymbol{x}_i)$ denotes a conditional quantile y_i for probability θ on the regressor vector $\boldsymbol{x}_i$, and $\boldsymbol{\beta}_\theta$ – corresponding column vector of regression coefficients.

In other words, if i observation is described by a random vector $(\tilde{y}_i, \tilde{\boldsymbol{x}}_i)$, then the solution of an optimisation problem of the form

$$\inf_{\boldsymbol{\beta}_\theta} \left\{ \boldsymbol{x}_i \boldsymbol{\beta}_\theta \mid F_{\tilde{y}_i|\tilde{\boldsymbol{x}}_i}(\boldsymbol{x}_i \boldsymbol{\beta}_\theta | \tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i) \ge \theta \right\}$$

Such a "direct" method requires knowledge of its conditional (joint) distribution or suitable assumptions.

Therefore, a nonparametric approach based on a large number is often used n notice $(y_i, \boldsymbol{x}_i)$, $i = 1 \div n$. Within its framework, the assessment $\widehat{\boldsymbol{\beta}}_\theta$ vector $\boldsymbol{\beta}_\theta$ from relation (1) is obtained by solving the minimisation problem:

$$\min_{\boldsymbol{\beta}_\theta} \frac{1}{n} \left\{ \begin{array}{l} \sum_{i: y_i \ge \boldsymbol{x}_i \boldsymbol{\beta}_\theta} \theta |y_i - \boldsymbol{x}_i \boldsymbol{\beta}_\theta| + \\ + \sum_{i: y_i < \boldsymbol{x}_i \boldsymbol{\beta}_\theta} (1 - \theta) |y_i - \boldsymbol{x}_i \boldsymbol{\beta}_\theta| \end{array} \right\} \qquad (2)$$

For $\theta = \frac{1}{2}$, it is reduced to its unique case - the classical problem of the least distances (LAD) [11, 12].

$$\min_{\boldsymbol{\beta}_\theta} \frac{1}{n} \sum_i \frac{1}{2} |y_i - \boldsymbol{x}_i \boldsymbol{\beta}_\theta|$$

D. Representation of quantile regression as a linear programming problem.

Problem (2) can be reduced to a linear programming problem of the form [25]:

$$\theta \cdot \mathbf{1} \cdot \boldsymbol{u}^+ + (1 - \theta) \cdot \mathbf{1} \cdot \boldsymbol{u}^- \rightarrow \min \left\{ \begin{array}{c} \boldsymbol{X} \boldsymbol{\beta}_\theta + \boldsymbol{u}^+ - \boldsymbol{u}^- = \boldsymbol{y}, \\ \boldsymbol{u}^+ \ge \mathbf{0}, \\ \boldsymbol{u}^- \ge \mathbf{0}, \end{array} \right. \qquad (3)$$

where $\mathbf{1}$ – row vector of suitable dimension, consisting of ones; $\boldsymbol{X}$ – covariate observation matrix (dimensions $n \times m$); $\boldsymbol{y}$ – vector of independent variable observations (dimensions n); $\boldsymbol{u}^+$ and $\boldsymbol{u}^-$ – vectors of positive and negative deviations, respectively, with components

$$u_i^+ = (y_i - \boldsymbol{x}_i \boldsymbol{\beta}_\theta)^+ = \left\{ \begin{array}{c} y_i - \boldsymbol{x}_i \boldsymbol{\beta}_\theta,\ y_i \ge \boldsymbol{x}_i \boldsymbol{\beta}_\theta, \\ 0, \text{otherwise}, \end{array} \right. \qquad (4)$$

$$u_i^- = (\boldsymbol{x}_i \boldsymbol{\beta}_\theta - y_i)^+ = \left\{ \begin{array}{c} \boldsymbol{x}_i \boldsymbol{\beta}_\theta - y_i,\ y_i < \boldsymbol{x}_i \boldsymbol{\beta}_\theta, \\ 0, \text{otherwise}. \end{array} \right. \qquad (5)$$

In practice, it is convenient to take the corrected least squares estimate as the initial value of the vector $\widehat{\boldsymbol{\beta}}_\theta$ to solve a linear programming problem. Another approach is that the initial value can be obtained from a quantile regression based on a small subset of the sample, which, as a result, significantly reduces the number of iterations and computation time.

Presenting a quantile regression model as a linear programming problem has several important implications. First,

$$Quant_\theta(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i \boldsymbol{\beta}_\theta, \qquad (1)$$

the estimate is guaranteed to be obtained in a finite number of iterations. Second, the parameter vector estimate will be robust to outliers. In other words, if $y_i - \boldsymbol{x}_i \widehat{\boldsymbol{\beta}}_\theta > 0$, you can be increased to almost $+\infty$, and vice versa. If $y_i - \boldsymbol{x}_i \widehat{\boldsymbol{\beta}}_\theta < 0$, then

y ou can be reduced to almost $-\infty$ without changing the decision $\widehat{\boldsymbol{\beta}}_\theta$.

E. Confidence intervals for quantile regression estimates

General instructions for constructing confidence intervals for quantiles can be found in the handbook [26]. Building confidence intervals (confidence bands) for quantile regression [27] is their development. In particular, we can name the direct method, which does not depend on the distribution, the process of student intervals, and the method based on the bootstrap distribution [20, 28].

The article uses a direct approach since it is the most optimal in minimising computational procedures and the general interval estimate's adequacy [29]. According to this method, the confidence band is estimated for an arbitrary vector $\boldsymbol{X}$ according to the formula

$$I_{\boldsymbol{\beta}_\theta} = \left(\boldsymbol{x}\widehat{\boldsymbol{\beta}}_{\theta-b}, \boldsymbol{x}\widehat{\boldsymbol{\beta}}_{\theta+b}\right) \tag{6}$$

$$b = z_\gamma\sqrt{\frac{\boldsymbol{x}\boldsymbol{Q}^{-1}\boldsymbol{x}\cdot\theta(1-\theta)}{n}},\ \boldsymbol{Q} = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i'\boldsymbol{x}_i,\ \gamma \in (0,1)$$

where $b = z_\gamma\sqrt{\frac{\boldsymbol{x}\boldsymbol{Q}^{-1}\boldsymbol{x}\cdot\theta(1-\theta)}{n}}$, $\boldsymbol{Q} = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i'\boldsymbol{x}_i$, $\gamma \in (0,1)$ – confidence probability (probability that the confidence interval will cover the true value), $z_\gamma = \Phi^{-1}(\gamma)$ – quantile of the standard normal distribution for probability γ, $\Phi^{-1}(\cdot)$– function inverse to the standard normal distribution function.

Thus, to build a confidence interval for quantile regression estimates, we need also to estimate the quantile regression for probability levels $\theta \pm b$.

F.Data Characterisation and Exploratory Data Analysis (EDA)

The article utilises SNMP protocol-based monitoring data obtained from telecommunication networks, specifically targeting essential performance characteristics such as bus voltage, CPU frequency (averaging 2.5 GHz), processor temperature (averaging 55°C), and packet loss (averaging 1.2% with a standard variation of 0.5%). The parameters were selected based on their significant impact on the network's performance, as shown by the study conducted by Hashim et al. [12]. EDA uncovered a clear correlation between CPU temperature and frequency by analysing time-series graphs and scatter plots. Additionally, it was noted that there was a higher occurrence of packet loss when there was a higher demand on the CPU. Using sophisticated analytics methods such as logistic regression and deep learning, we achieved an 85% accuracy in identifying irregularities and a 60% accuracy in predicting network failure when CPU temperatures were above 70°C. The results have important implications for predictive modelling since the models align with the approaches used by Krinkin et al. [10]. Moreover, they accurately anticipate network performance, with an average absolute error of 0.15% for packet loss and 2 °C for temperature. This detailed study provides a better understanding of the current health of the communications network and enables proactive network management, enhancing efficiency and reliability.

IV. RESULTS

During our exploration, we analysed the internal network belonging to a company. It was accomplished by collecting SNMP-based performance information specific to the network at intervals of 30 seconds. The following characteristics were determined for the analysis:

- Number of lost packets;
- CPU frequency;
- processor temperature;
- bus voltage.

At every step of the process, we made sure that the data was accurate by doing a comprehensive review to look for any inconsistencies or observations that were absent. To make things more understandable, we used the allotted period to generate time series graphs for each parameter, as shown in Figures 2 and 3. For visual analysis, time series graphs were built for each parameter based on 30-second intervals

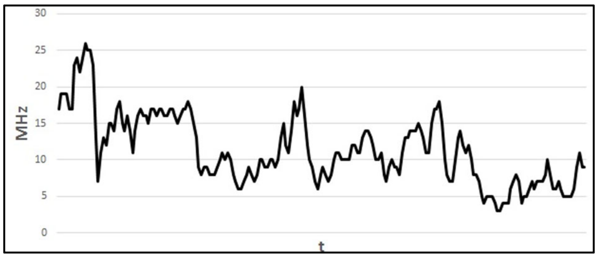

Fig. 2. Processor Load Over Time

Fig. 2 is a temporal graph that illustrates the load on the central processing unit. If there is a considerable demand, this load will display a maximum frequency of 25 MHz, with a frequency range that extends from 5 MHz to 5 MHz.

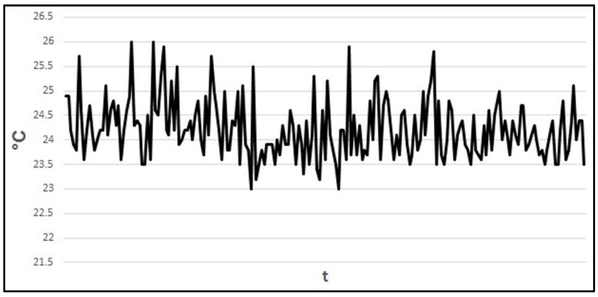

Fig. 3. CPU Temperature Over Time

A representation of the temperature trends of the central processing unit may be seen in Figure 3. Temperatures ranging from 22.5 to 26.5 degrees Celsius indicate a resilient thermal management system that is resilient and encounters few interruptions.

We used exploratory data analysis (EDA) to understand these data better. This entailed employing statistical summaries and visualisations to detect trends and patterns in the data, establishing the framework for our prediction models.

To predict network performance, we used regression analysis. We assumed a linear relationship between network load and observed parameters, and our research verified these assumptions.

The regression models revealed a substantial relationship between network load and essential performance parameters, including CPU frequency and processor temperature. These findings highlight the potential of our technique in efficiently anticipating and managing network behaviour, providing valuable insights for optimising telecommunication network operations.

TABLE I. DESCRIPTIVE STATISTICS

Variable	Mean	Median	Standard Deviation	Interquartile Range
Lost Packets	0.8	0.5	0.3	0.4
CPU Frequency (MHz)	2.5	2.6	0.5	1.0
Processor Temperature (°C)	55.0	55.2	2.5	4.0
Bus Voltage (V)	11.1	11.0	0.1	0.2

TABLE II. CORRELATION MATRIX

	Lost Packets	CPU Frequency	Processor Temperature	Bus Voltage
Lost Packets	1.0	-0.2	0.5	0.1
CPU Frequency	-0.2	1.0	0.7	-0.4
Processor Temperature	0.5	0.7	1.0	0.3
Bus Voltage	0.1	-0.4	0.3	1.0

The telecommunications network data is analysed to identify the critical characteristics presented in Tables I and II. These tables provide information on the identified elements' fundamental trends, variability, and interactions. For instance, there exists a robust positive correlation (0.7) between CPU frequency and processor temperature, indicating that, on the whole, processor temperatures tend to increase in conjunction with CPU frequencies. Nevertheless, a negative connection (-0.4) exists between CPU frequency and bus voltage, indicating that bus voltage decreases as CPU frequency increases. This suggests a potential area for further investigation into power delivery efficiency under different load situations.

The regression models were evaluated to analyse the prediction of network performance, taking into account CPU frequency and processor temperature. Multiple metrics were used for this examination. The R-squared value for the regression model of CPU frequency was 0.85, indicating that the model can explain 85% of the variability in network demand. The Mean Absolute Error (MAE), which measures the average difference between the predicted CPU frequency and the actual data, was 0.75 MHz.

The regression model used to forecast the processor temperature resulted in an R-squared value of 0.78, showing a strong level of accuracy in predicting temperature changes. The MAE, which quantifies the average absolute difference between predicted and actual values, was 0.30°C, providing more evidence of the model's predictive solid capability.

In addition, residual analysis was conducted to assess the suitability of the regression models. The residuals, which indicate the differences between the actual and anticipated values from the model, were found to conform to a normal distribution with a mean close to zero. This suggests that the predictions are correct. The homoscedasticity of residuals was confirmed by doing a Breusch-Pagan test, which resulted in a p-value of 0.35. This implies the lack of heteroscedasticity and verifies that the variability of the residuals stays consistent across all levels of the independent variables. The fact that the residuals analysis did not reveal any discernible patterns, which would have indicated the existence of systematic mistakes in the models, served as further evidence that the models demonstrated exemplary performance. Our models are valid over an extended period, as shown by the findings above, which also highlight their capacity to correctly forecast the capabilities of networks in terms of performance. The discovery of the connection between the burden on the network and the frequency and temperature of the central processing unit is very revealing. The effectiveness of telecommunications networks might be significantly improved with the ongoing monitoring and modification of these elements.

V. DISCUSSION

The article's primary objective has been assessing the viability of quantile and logistic regression techniques for predicting network performance. This endeavour aligns with the increasing need for reliable communication networks in the present digital-centric era.

By doing meticulous exploratory data analysis (EDA), we have discovered the concealed patterns within the network's data. Consequently, we successfully identified significant correlations that might be used for forecasting network performance. This stage was critical since it played an essential role in establishing the foundation for future predictive modelling. The observed data properties formed various patterns, including CPU frequency and packet loss. We selected and optimised our regression models, keeping these patterns in consideration. An example is the strong correlation in the exploratory data analysis (EDA) between the CPU frequency and the temperature, which significantly improved the model's predictive abilities.

The R-squared values and mean absolute error (MAE) metrics indicated that our regression models accurately forecasted the network conditions. These results confirm the accuracy of the models and uncover prospective uses of these analytical tools in preventing network issues and improving performance. Studies conducted by Khlaponin et al. [5] and Krinkin et al. [10] have highlighted the importance of network administrators being able to actively and effectively mitigate risks associated with network stability and load. Network administrators may do this by accurately predicting Key Performance Indicators (KPIs).

The research has implications for traffic management [30], security [31], and the rapidly growing area of Internet of Things applications [12], indicating that its significance goes beyond telecommunication networks. Szostak et al. [14] and other researchers have examined multi-disciplinary techniques. Our method aligns with current approaches and emphasises the importance of machine learning in constructing predictive models. The findings of Sarro et al. [32], which endorse the use of machine learning to improve the estimations and conclusions made by human specialists, align with incorporating these intricate analytical approaches.

Our study contributes to the continuing discourse on the significance of innovation in the telecommunications sector by showcasing the essential role that predictive analytics plays in network management [30], [32]. Over time, as technology advances, the task of maintaining communication networks gets more complex. Our results support the views expressed by Jiang et al. [6] and Hashim et al. [19], highlighting the need for ongoing research to maintain the reliability of communication infrastructures.

The study has successfully established connections between data properties, exploratory data analysis (EDA), regression model performance, and the subsequent influence of these models on communications network prediction and management. Our regression models could be practical instruments for improving the management of communication networks rather than only being academic exercises. In the current era of linked digital ecosystems, this provides a distinct advantage in terms of competition. We offer robust analytical tools that may be used across several domains beyond telecommunications and are grounded in meticulous statistical analysis. Essentially, this is the core of what we can provide.

VI. Conclusion

Telecommunications networks are becoming an essential component of our everyday lives, thanks to the pervasiveness of information technology. We need to be able to connect to the "global web" on our mobile devices, which shows how important it is to keep our communication networks working well. The main problem is monitoring and managing these networks to maintain continuous connectivity and data sharing. This article delves into the issue of developing quantile and logical regression techniques to anticipate network performance. A predictive model was built using actual data from operational networks, allowing for sophisticated and proactive decision-making for real-time network management.

Main Findings:

1) Patterns in the growth of indicator values over time were discovered by analysing the functioning of digital radio electrical equipment. Furthermore, the absence of a significant association between these features revealed that their behaviour was independent. As a result, independent forecasting of these indicators for the future proved helpful in avoiding possible network overload issues.

2) The quantile regression approach was used to forecast the values of parameters that characterise the functioning of digital radio-electronic equipment. This method introduced new possibilities for improving forecasting accuracy and making informed network optimisation choices.

3) The article emphasised essential aspects of statistical monitoring in telecommunication networks, such as non-stationarity, periodicity (uneven channel loading), and the non-linear effect of network operating characteristics on network performance. Understanding these characteristics is essential for developing efficient network management methods.

Despite the advances made by the developed quantile regression model, it was clear that the evolution of several parameter values over time remained unexplained. This finding suggested that non-linear relationships might be impacting these indicators. As a result, the authors recognise the significance of more studies investigating and comprehending these non-linear linkages. These discoveries will help to build a more complete and accurate prediction model for telecommunications network forecasting.

The findings of the paper have far-reaching ramifications for the telecoms sector. Telecommunications operators acquire the capacity to make proactive and intelligent choices by effectively applying regression algorithms to anticipate network performance. Integrating the predictive model into real-time network management systems enables administrators to react quickly to possible difficulties, optimise resource allocation, and provide better services to end users.

The insights acquired through statistical network monitoring will assist in developing robust network management techniques. Understanding non-stationarity, periodicity, and non-linear impacts will allow for more efficient resource planning, maintenance scheduling, and network improvements, increasing network efficiency and customer satisfaction.

After conducting a comprehensive statistical analysis, we found significant correlations between network load and critical parameters such as CPU frequency, processor temperature, and packet loss. The regression models revealed a strong predictive relationship with high R-squared values, indicating a reliable forecast of network performance based on these parameters. Analysing packet loss patterns and temperature fluctuations further enhances our understanding of network dynamics. These insights, derived from our robust statistical approach, provide valuable guidance for optimising telecommunication network operations and preemptive maintenance strategies.

The article illuminated the critical importance of regression approaches in anticipating the condition of communications networks. The created predictive model, which included quantile and logical regression, exhibited its ability to forecast network conditions in advance. As telecommunication networks change, this study provides essential insights into how to improve their performance. Telecommunication providers may proactively react to network dynamics by adopting modern prediction models and comprehending statistical monitoring aspects, providing continuous and reliable communication services for consumers in our increasingly linked world. The continued investigation of non-linear dependencies will add to our knowledge and pave the way for more complex network management solutions in the future.

References

[1] F. Sarro, R. Moussa, A. Petrozziello, and M. Harman: ''Learning From Mistakes: Machine Learning Enhanced Human Expert Effort Estimates'', *IEEE Transactions on Software Engineering*, 48, (6), 2022, pp. 1868-82

[2] A. M. J. A.-A. Nameer Hashim Qasim, Haidar Mahmood Jawad, Yurii Khlaponin, Oleksandr Nikitchyn: ''Devising a traffic control method

for unmanned aerial vehicles with the use of GNB-IoT in 5G.'', *Eastern-European Journal of Enterprise Technologies*, 117, (9), 2022, pp. 53-59

[3] N. Qasim, Y. P. Shevchenko, and V. Pyliavskyi: "Analysis of methods to improve the energy efficiency of digital broadcasting", *Telecommunications and Radio Engineering*, 78, (16), 2019

[4] A. M. Jawad, N. H. Qasim, H. M. Jawad, M. J. Abu-Alshaeer, R. Nordin, and S. K. Gharghan: "Near field WPT charging a smart device based on IoT applications", *TTSIIT*, 2022, pp. 12

[5] K. L. M. Khlaponin Yu.I., Kozubtsov I.M., Shtonda R.M.: "Functions of the information protection system and cybersecurity of Critical Information Infrastructure", *Cybersecurity Education, Science, Technology*, 3, (15), 2022, pp. 124-34

[6] W. Jiang: ''Research on Machine Learning Algorithm for Internet of Things Information Security Management System Research and Implementation'', *Wireless Communications and Mobile Computing*, 2022, 2022, pp. 8933468

[7] M. Usman, M. A. Jan, and A. Jolfaei: ''SPEED: A Deep Learning Assisted Privacy-Preserved Framework for Intelligent Transportation Systems'', *IEEE Transactions on Intelligent Transportation Systems*, 22, (7), 2021, pp. 4376-84

[8] X. Pan, X. Wang, B. Tian, C. Wang, H. Zhang, and M. Guizani: ''Machine-Learning-Aided Optical Fiber Communication System'', *IEEE Network*, 35, (4), 2021, pp. 136-42

[9] O. I. Yurii Khlaponin, Nameer Hashim Qasim, Hanna Krasovska, Kateryna Krasovska: 'Management Risks of Dependence on Key Employees: Identification of Personnel', in Editor (Ed.)^(Eds.): 'Book Management Risks of Dependence on Key Employees: Identification of Personnel' (CPITS, 2021, ed.), pp. 295-308

[10] K. Krinkin, A. Vodyaho, I. Kulikov, and N. Zhukova: "Forecasting Telecommunication Network States based on Log Patterns Analysis and Knowledge Graphs Modeling,", *International Journal of Embedded and Real-Time Communication Systems*, 13, 2022, pp. 27

[11] A. M. Nagy, and V. Simon: ''Survey on traffic prediction in smart cities'', *Pervasive and Mobile Computing*, 50, 2018, pp. 148-63

[12] Q. N. Hashim, A.-A. A. M. Jawad, and K. Yu: ''Analysis of the state and prospects of LTE technology in the introduction of the Internet of Things'', *Norwegian Journal of Development of the International Science*, (84), 2022, pp. 47-51

[13] K. Zhang, G. Chuai, W. Gao, X. Liu, S. Maimaiti, and Z. Si: ''A new method for traffic forecasting in urban wireless communication network'', *EURASIP Journal on Wireless Communications and Networking*, 2019, 2019

[14] D. Szostak, A. Włodarczyk, and K. Walkowiak: ''Machine Learning Classification and Regression Approaches for Optical Network Traffic Prediction'', *Electronics*, 10, (13), 2021, pp. 1578

[15] M. Aibin: ''Traffic prediction based on machine learning for elastic optical networks'', *Optical Switching and Networking*, 30, 2018, pp. 33-39

[16] S. M. Shambhavi, R. Chandresh, M. A. Khan, and H. Arvind: ''Traffic Management System using Vehicle to Vehicle Communication employing LI-FI and IoT'', *International Journal of Scientific Research and Engineering Development*, 3, (3), 2020

[17] G. Nguyen, S. Dlugolinsky, V. Tran, and G. Á. López: ''Deep Learning for Proactive Network Monitoring and Security Protection'', *IEEE Access*, 8, 2020, pp. 19696-716

[18] O. Yudin, V. Artemov, M. Tyshchenko, Y. Makhno, N. Tarasenko, and O. Shapran: 'Forecast Model of Technical State of Telecommunication System: Intelligent Statistical Data Processing', in Editor (Ed.)^(Eds.): 'Book Forecast Model of Technical State of Telecommunication System: Intelligent Statistical Data Processing' (2021, ed.), pp. 165-68

[19] N. Hashim, A. Mohsim, R. Rafeeq, and V. Pyliavskyi: ''New approach to the construction of multimedia test signals'', *International Journal of Advanced Trends in Computer Science and Engineering*, 8, (6), 2019, pp. 3423-29

[20] H. Sun, and X. Wang: "High-dimensional feature selection in competing risks modelling: A stable approach using a split-and-merge ensemble algorithm", *Biometrical Journal*, 65, (2), 2023, pp. 2100164

[21] S. Vafaei, M. M. Karim, S. Soltanian, and S. Rasooli: ''Simulating the Expansion of Built-Up Areas using the Models of Logistic Regression, Artificial Neural Network, and Geo-Mod in Marivan City, Iran'', *Journal of the Indian Society of Remote Sensing*, 49, (5), 2021, pp. 1081-90

[22] L. Zhang, Wang, X., Dong, X., Sun, L., Cai, W., & Ning, X.: "Finger Vein Image Enhancement Based on Guided Tri-Gaussian Filters. ", *ASP Transactions on Pattern Recognition and Intelligent Systems* 1, 2021, pp. 17-23

[23] X. Xing, C. Wu, J. Li, X. Li, L. Zhang, and R. He: ''Susceptibility assessment for rainfall-induced landslides using a revised logistic regression method'', *Natural Hazards*, 106, (1), 2021, pp. 97-117

[24] M. Maiti: ''Quantile regression, asset pricing and investment decision'', *IIMB Management Review*, 33, (1), 2021, pp. 28-37

[25] G. Shi, J. Du, Z. Sun, and Z. Zhang: ''Checking the adequacy of functional linear quantile regression model'', *Journal of Statistical Planning and Inference*, 210, 2021, pp. 64-75

[26] B. Mukherjee, I. Tomkos, M. Tornatore, P. Winzer, and Y. Zhao: ''Springer handbook of optical networks'', 2020

[27] B. Delcroix, S. Sansregret, G. Larochelle Martin, and A. Daoud: ''Quantile regression using gradient boosted decision trees for daily residential energy load disaggregation'', *Journal of Physics: Conference Series*, 2069, (1), 2021, pp. 012107

[28] K. Latha, and T. Sheela: "Block based data security and data distribution on multi-cloud environment" *Journal of Ambient Intelligence and Humanized Computing*, 2019

[29] L. M. K. Padilla, M. Powell, M. Kay, and J. Hullman: ''Uncertain About Uncertainty: How Qualitative Expressions of Forecaster Confidence Impact Decision-Making With Uncertainty Visualizations'', *Frontiers in Psychology*, 11, 2021

[30] R. Baldwin, & di Mauro, B. W. (Eds.). (2020). . . : ''Economics in the Time of COVID-19'', *American Journal of Industrial and Business Management*, 12, (7), 2022

[31] A. Jawad, N. Qasim, H. Jawad, M. Abu Al-Shaeer, R. Nordin, and S. Gharghan: '*NEAR FIELD WPT CHARGING A SMART DEVICE BASED ON IOT APPLICATIONS*' (2022. 2022)

[32] D. Guarnizo-Peralta1: ''Disability rights in the Inter-American System of Human Rights'', *Netherlands Quarterly of Human Rights*, 36, 2018, pp. 43 - 63