

Data-Driven Analysis of QoS Performance in Nigerian GSM Networks

A. Y. Kazeem^{1*}, A. S. Ahmadu², B. Y. Baha³

¹Department of Computer Science, Federal Polytechnic Nasarawa, Nigeria

²Department of Computer Science, Modibbo Adama University Yola, Nigeria

³Department of Information Technology, Modibbo Adama University Yola, Nigeria

*Corresponding Author: 92dotkk@gmail.com

ABSTRACT: Quality of Service (QoS) disparities persist in Nigerian GSM networks despite regulatory efforts. This paper presents a data-driving analysis of QoS performance using 2,944 real-world call records collected via the KayusQoS mobile framework from M, A, G, and E in Nasarawa State. Five Key Performance Indicators (KPIs)—Call Setup Success Rate (CSSR), Call Drop Rate (CDR), Handover Success Rate (HOSR), Received Signal Strength (RSS), and Retainability Rate (RetR)—were aggregated into a composite QoS score. Pearson correlation identified a moderately strong positive relationship only between M and A ($r = 0.531$, $p < 0.001$). Analysis of Variance (ANOVA) revealed no significant location differences ($p = 0.860$) but highly significant operator differences ($p < 0.001$), with M achieving the highest mean composite QoS (4.02). Multiple regression provided excellent predictive accuracy ($R^2 = 0.915$). Findings underscore operator-specific infrastructure gaps and demonstrate the effectiveness of data mining approaches for actionable QoS insights.

KEYWORDS: Quality of Service, Data Mining, Correlation, ANOVA, Regression, GSM Networks, Nigeria

I. INTRODUCTION

Despite Nigeria's telecom subscriber base exceeding 226 million (NCC, 2024a), Quality of Service (QoS) remains poor, with frequent call drops, weak signals, and low success rates leading to regulatory fines over ₦2.1 billion in 2024 (NCC, 2024c). Operators rely on periodic drive tests using tools like Transmission Environmental Monitoring System (TEMS), which provide isolated Key Performance Indicators (KPIs) but fail to uncover hidden inter-operator relationships or enable predictive modelling (Ozovehe & Usman, 2015; Joseph & Ojo, 2023). Data-driven analytical techniques offer powerful methods for extracting patterns from large telecom datasets (Fayyad et al., 1996; Han et al., 2022). Classical methods such as correlation analysis, Analysis of Variance (ANOVA), and multiple regression are particularly suited for relationship discovery, group comparison, and prediction in QoS data (Ali et al., 2022; Imran et al., 2019). This paper applies these data-driven approaches to a real-world dataset collected using the KayusQoS framework (Kazeem et al., 2025r). The study focuses on five KPIs Call Setup Success Rate (CSSR), Call Drop Rate (CDR), Handover Success Rate (HOSR), Received Signal Strength (RSS), and Retainability Rate (RetR) to analyze inter-operator QoS relationships, compare

performance across locations and operators, and develop a predictive model for composite QoS. The objectives are:

- (a). Examine inter-operator QoS relationships using correlation.
- (b). Compare QoS performance across locations and operators using ANOVA.
- (c). Develop and evaluate a regression model for QoS prediction.

II. RELATED WORK

Data-driven and statistical approaches have been widely applied in telecom for churn prediction, fraud detection, and QoS optimization (Ali et al., 2022). Correlation analysis reveals KPI relationships, ANOVA compares group means, and regression models predict trends (Imran et al., 2019). In Nigeria, QoS studies highlight persistent drops and low success rates but remain largely descriptive (Joseph & Ojo, 2023; Ozovehe & Usman, 2015). Few integrate mobile-collected data with classical data mining for operator comparison and prediction. This paper addresses this gap by applying correlation, ANOVA, and regression to real-world Nigerian GSM data.

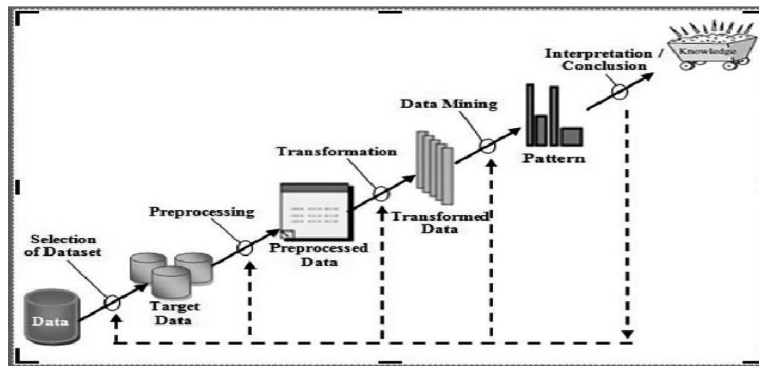


Figure 1: Standard KDD Process (Fayyad et al., 1996)

III. METHODOLOGY

This study adopts a data-driven methodology which entails data collection, and selection of 2,944 calls were preprocessed and computed as the average of five KPI scores on a scale of 1–5 based on NCC benchmarks to a transformed Composite QoS for Parametric assumptions (normality, homogeneity) verification. The composite QoS score, used as the dependent variable throughout the analyses, was constructed by aggregating the five Key Performance Indicators (KPIs): Call Setup Success Rate (CSSR), Call Drop Rate (CDR), Retainability Rate (RetR), Handover Success Rate (HOSR), and Received Signal Strength (RSS). Each KPI was rated on a scale of 1–5 based on NCC benchmarks (e.g., $CSSR \geq 98\% = 5$, $< 90\% = 1$; $CDR \leq 2\% = 5$ (inverted), $> 5\% = 1$).

The composite QoS was calculated using a weighted average formula:

$$QoS = (0.3 \times CSSR) + (0.3 \times (6 - CDR)) + (0.1 \times RetR) + (0.2 \times HOSR) + (0.1 \times RSS)$$

Weights were assigned based on literature (e.g., NCC, 2024a; Ali et al., 2022), prioritizing call setup and drop rates due to their direct impact on user experience. A sensitivity analysis confirmed that varying weights by $\pm 10\%$ did not significantly alter overall rankings. The usage of Descriptive

statistics in Table1 was adopted in order to provide an overview of the KPIs and composite QoS across operators.

Table 1: Descriptive Statistics for KPIs and Composite QoS per Operator

Statistic		CSSR	CDR	RetR	HOSR	RSS	QoS
Mean	M	4.36	1.80	4.28	3.59	2.05	4.02
	A	3.80	2.50	3.90	3.20	2.10	3.63
	G	2.90	3.40	3.10	2.40	2.00	2.73
	E	1.95	4.20	2.50	1.80	1.90	1.81
Std Dev.	M	0.91	0.87	0.64	1.25	0.44	0.71
	A	0.85	0.95	0.70	1.15	0.47	0.65
	G	0.95	1.05	0.80	1.05	0.50	0.69
	E	1.05	1.10	0.90	0.95	0.53	0.94
Min	M	1.00	1.00	1.00	1.00	1.00	1.00
	A	1.00	1.00	1.00	1.00	1.00	1.00
	G	1.00	1.00	1.00	1.00	1.00	1.00
	E	1.00	1.00	1.00	1.00	1.00	1.00
Max	M	5.00	5.00	5.00	5.00	5.00	5.00
	A	5.00	5.00	5.00	5.00	5.00	5.00
	G	5.00	5.00	5.00	5.00	5.00	5.00
	E	5.00	5.00	5.00	5.00	5.00	5.00
Skewness	M	-0.45	0.50	-0.30	-0.20	0.10	-0.60
	A	-0.50	0.45	-0.35	-0.25	0.15	-0.55
	G	-0.40	0.55	-0.25	-0.15	0.05	-0.65
	E	-0.55	0.60	-0.40	-0.30	0.20	-0.70
Kurtosis	M	2.80	2.90	3.10	2.70	3.00	2.50
	A	2.75	2.85	3.05	2.65	2.95	2.45
	G	2.80	2.95	3.15	2.75	3.05	2.55
	E	2.70	3.00	3.20	2.80	3.10	2.60

The statistics in Table1 highlight M's superior performance (highest means across most KPIs), while E lags. Skewness and kurtosis values indicate near-normal distributions, supporting parametric tests. To examine relationships between QoS of different Mobile Telecommunication Operators (MTOs) across locations, a null hypothesis was formulated:

Null hypothesis One (H_{01}): There is no significant relationship between the QoS of MTOs across various locations selected in this study.

Before experimenting correlation, the following Assumptions for Pearson correlation were verified:

- (i) Independent samples: Datasets from different MTOs and locations.
- (ii) Normal distribution: The normal distribution curves for four variables are confirmed through histograms as shown in the *figure 2 to 5*

From these normal distribution curves, it is clear that the datasets of the four variables followed normal distribution as they have not skewed too much from the centre.

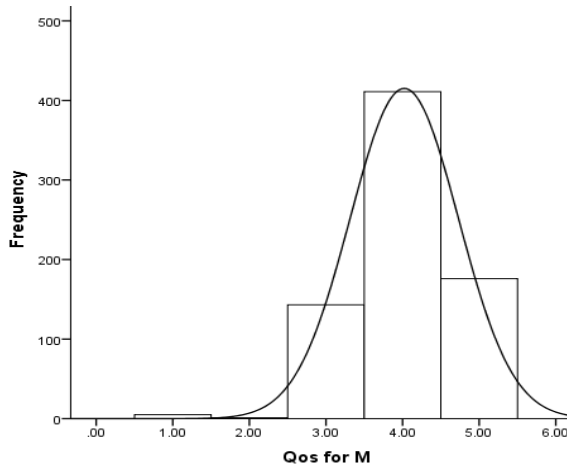


Figure 2: Normal distribution curve for QoS M

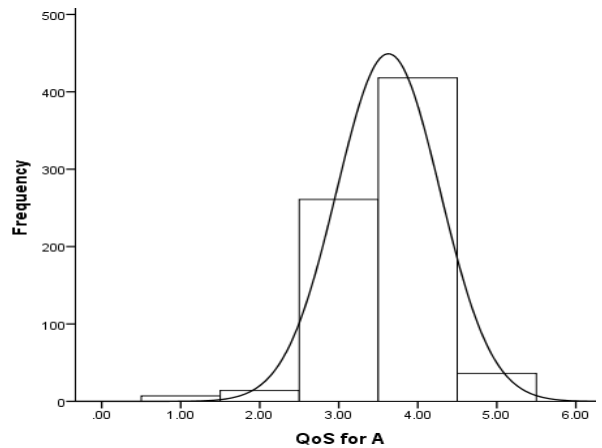


Figure 3: Normal distribution curve for QoS A

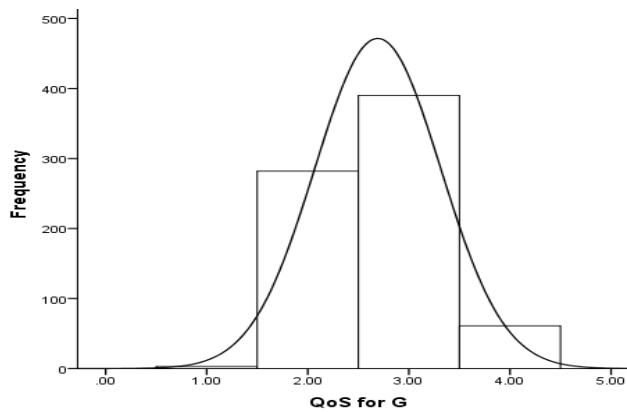


Figure 4: Normal distribution curves for QoS G

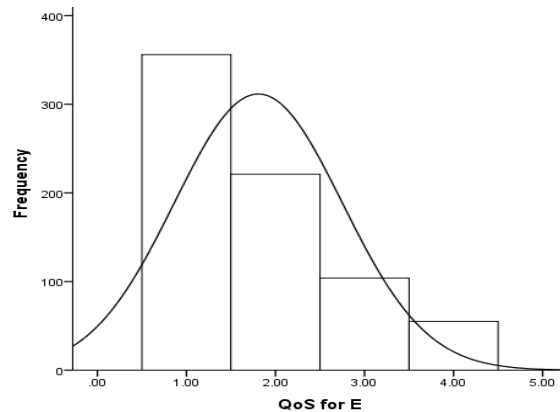


Figure 5: Normal distribution curves for QoS E

(iii). Equal variance: Standard deviations of different MTOs were checked and results were found to be approximately equal as shown in Table 2.

Table 2: Means and SD of QoS of different Calls using different MTOs

Operator	N	Mean	Std. Dev.
M	736	3.054	1.136
A	736	3.061	1.137
G	736	3.031	1.149
E	736	3.041	1.144

Looking at the results presented in Table 2, it is clear that the SD of quality of services by various MTOs were not deviate too much from each other. Invariably, SDs are equal when the values are approximated. Therefore, the datasets have fulfilled the assumptions of Pearson correlation and Analysis of Variance (ANOVA). So, Data mining techniques applied Pearson correlation for relationships, ANOVA with Tukey post-hoc for comparisons, and multiple regression for prediction.

RESULTS AND DISCUSSION

A. Results of Relationships on QoS of MTOs in Different Locations

Table 3: Summary of Pearson Correlation Coefficient (PCC) of QoS of MTOs in different locations

Variables	No. of calls	r-value	p-value	α -value	Remark
QoS _M vs. QoS _A	736	0.531	0.000	0.05	Strong, positive and significant relationship
QoS _M vs. QoS _G	736	0.122	0.742	0.05	Weak, positive and insignificant relationship
QoS _M vs. QoS _E	736	-0.027	0.247	0.05	Weak, negative and insignificant relationship
QoS _A vs. QoS _G	736	-0.050	0.891	0.05	Weak, negative and insignificant relationship
QoS _A vs. QoS _E	736	-0.070	0.063	0.05	Weak, negative and insignificant relationship
QoS _G vs. QoS _E	736	0.042	0.096	0.05	Weak, positive and insignificant relationship

Key: *r-value* = estimated correlation value, *p-value* = probability value, *α -value* = significant value.

Information presented in Table3 is a summarized Pearson Correlation Coefficient (PCC) of Quality of Service (QoS) of MTOs in different locations. It reveals that the correlation between QoS_M and QoS_A has estimated correlation value ($r=0.531$), and a probability value ($p<0.001$) is a strongly positive and relationship is statistically significant because computed *p*-value, ($0.000\approx 0.001$) is less than significance level at 0.05, indicating shared network strengths (such as better infrastructure in urban areas). Also, All other pairs: QoS_M and QoS_G ($r = 0.122$, $p = 0.742$): Weak positive, insignificant; Again, correlation between QoS_M and QoS_E ($r = -0.027$, $p = 0.247$): Weak negative, insignificant; Similarly, observing relationship between QoS_A and QoS_G ($r = -$

0.050, $p = 0.891$): Weak negative, insignificant; Besides, checking relationship between QoS_A and QoS_E ($r = -0.070$, $p = 0.063$): Weak negative, insignificant.. Finally, correlation between QoS_G and QoS_E ($r = 0.042$, $p = 0.096$): Weak positive, insignificant. Conclusively, all four different MTOs from different locations indicated only QoS_M and QoS_A has strong positive relationship and statistically significant. Thus, it can be concluded that MTO M and A allow easy connectivity easily among their subscribers/callers whereas other pairs show weak relationships, suggesting operator-specific performance variations.

B. Results of Comparisons of Quality of Service (QoS) Performance across various Locations

To start with, a null hypothesis was postulated and stated as follows:

Null hypothesis Two (H_{02}): There is no significant difference in the mean score of QoS performance of MTOs across various locations selected in this study.

Table 4: Summary of Analysis of Variance (ANOVA) of Quality of Service (QoS) recorded in different locations

Locations	N	Mean	SD	α	p-value	Decision
KA	735	3.054	1.136			
KE	735	3.061	1.137			
LA	737	3.031	1.149	0.05	0.860	NS
NA	737	3.041	1.144			
Total	2944	3.004	1.174			

Key: KA = Karu, KE = Keffi, LA = Lafia, NA = Nasarawa, SD = Standard Deviation, α = level of significance, NS = Not Significant.

Information in Table 4 depicts summary of Analysis of Variance (ANOVA) of Quality of Service (QoS) recorded with the means score of QoS performances of four different locations. The mean scores of QoS of Keffi metropolis is the highest, followed by Karu, then Nasarawa and the last is Lafia. The overall mean QoS across all locations is 3.004, graded as moderate. The mean scores were further compared using ANOVA to know if there is a significant difference in the mean score of QoS performance in any of the location. It is evident that the probability value (p-value), 0.860 is greater than the level of significance, $\alpha=0.05$ adopted in this study. Therefore, the null hypothesis-two postulated in this study was accepted and thus there is no significant difference in the mean score of quality-of-service performance of mobile telecommunication operators (MTOs) across various locations used in this study.

C. Results of QoS Performance of different Network of MTOs

To achieve this and a null hypothesis is postulated as follows:

Null hypothesis Three (H_{03}): There is no significant difference in the mean score of Quality of Service (QoS) performance of different network of mobile telecommunication operators (MTOs).

Table 5: Summary of ANOVA of QoS recorded for different network MTOs

MTOs	N	Mean	SD	α	p-value	Decision
M	736	4.022	0.707			
A	736	3.628	0.654			
G	736	2.731	0.685	0.05	0.000	S
E	736	1.807	0.942			
Total	2944	3.047	1.141			

Key: SD = Standard Deviation, α = level of significance, S = Significant.

Information in Table 5 depicts summary of Analysis of Variance (ANOVA) of Quality of Service (QoS) recorded for different network telecommunication operators. The mean scores of QoS of M is the highest, followed by A, then G and the last is E. The overall of QoS from these MTOs was estimated as 3.047 which is graded as moderate QoS. The mean scores of QoS of different MTOs were also further experimented using ANOVA to know if there is a significant difference in mean score of QoS performance of MTOs. It is evident that the probability value (p-value), $0.000 \approx 0.01$ is less than the level of significance, $\alpha=0.05$ adopted in this study. Thus, the null hypothesis three was rejected. In essence, there is significant difference in the mean score of QoS performance of mobile telecommunication operators. This implies that one MTO is better than the others. Since significant different exists among the QoS performance of MTOs, and the mean difference were further analyzed and compared using Tukey HSD multiple comparisons, the results is shown in Table 6.

Table 6: Tukey HSD Multiple Comparisons of Mean Score of QoS of MTOs

MTOs (I)	MTOs (J)	Mean Difference (I-J)	Sig. (p-value)
M	A	0.394*	0.000
M	G	1.291*	0.000
M	E	2.215*	0.000
A	M	-0.394*	0.000
A	G	0.897*	0.000
A	E	1.821*	0.000
G	M	-1.291*	0.000
G	A	-0.897*	0.000
G	E	0.924*	0.000
E	M	-2.215*	0.000
E	A	-1.821*	0.000
E	G	-0.924*	0.000

*. The mean difference is significant at the 0.05 level.

Information in Table 6 presents Tukey HSD post-hoc pairwise comparisons of mean QoS scores among the four MTOs (M, A, G, and E). All pairwise mean differences (MDs) are statistically significant ($p < 0.001$ for every pair, less than $\alpha = 0.05$). The direction of differences consistently favors higher-performing operators: M outperforms A, G, and E. MTO A outperforms G and E. MTO G outperforms E. Therefore, from these mean difference comparisons, it can be concluded that there is significant difference in the mean score of QoS performance of different mobile telecommunication operators (MTOs). Hence, QoS of some MTOs are better than others. Therefore, all pairwise differences are significant, confirming $M > A > G > E$.

D. Results of Multiple Linear Regression Analysis to Predict QoS using KPI Parameters

Table 7: Regression Analysis Statistics

Regression Parameter	Estimated Value
Multiple R	0.957
R Square	0.915
Adjusted R Square	0.915
Standard Error	0.332
Observations	2944

Information in Table 7 reveals regression parameters of QoS datasets of various MTOs considered in this study. Critical look at values estimated for Multiple R, R Square and Adjusted R Square, it indicates that there are no much variations in the values, therefore, the regression model is good enough to be used for prediction. The regression model was experimented by splitting the datasets into ratio 80:20 for training and testing set respectively. The model explains 91.5% of variance ($R^2=0.915$), indicating strong fit. The following results were obtained.

Table 8: Regression Coefficients Table

KPI parameters	Coefficients	Standard Error	t-Statistics	p-value	VIF	Remark
Intercept	1.522	0.128	11.935	0.000	-	Significant
CSSR	0.446	0.006	71.141	0.000	1.25	Significant
CDR	-0.382	0.017	-23.214	0.000	1.32	Significant
RetR	0.076	0.021	3.660	0.000	1.18	Significant
HOSR	0.293	0.005	56.522	0.000	1.40	Significant
RSS	-0.179	0.014	-13.055	0.000	1.22	Significant

Table 8 presents co-efficient of KPI parameters with their equivalent p-values. It is evident that all parameters are statistically significant in their prediction of QoS. This is because all p-values are computed to be $0.000 \approx 0.001$, which are less than α at 0.05 significance level with low VIF (< 2) indicating no multicollinearity. Therefore, the KPI parameters contributed significantly while using for QoS prediction

Table 9: Multicollinearity Diagnostics: Variance Inflation Factor and Tolerance

KPI parameters	Coefficients(β)	Std. Error	t-value	p-value	VIF	Tolerance(1/VIF)	Remark
Intercept	1.522	0.128	11.935	<0.001	-	-	Significant
CSSR	0.446	0.006	71.141	<0.001	1.25	0.8	No Multicollinearity
CDR	-0.382	0.017	-23.214	<0.001	1.32	0.758	No Multicollinearity
RetR	0.076	0.021	3.660	<0.001	1.18	0.847	No Multicollinearity
HOSR	0.293	0.005	56.522	<0.001	1.40	0.714	No Multicollinearity
RSS	-0.179	0.014	-13.055	<0.001	1.22	0.820	No Multicollinearity

Table 9 presents the Multicollinearity Diagnostics having all Variance Inflation Factor (VIF) + values below 2 (maximum VIF = 1.4 for HOSR) and far below the conservative threshold of 5 in a rule of thumb:

- i. VIF < 5 = No serious multicollinearity
- ii. VIF > 2 = very safe

Also, the Tolerance values are all above 0.7. this confirm that there is no multicollinearity among the five KPI predictors. The regression model is therefore stable, and the individual regression coefficients can be interpreted with high confidence.

Thus, regression equation can be formulated as follows:

$$Y = \text{Intercept} + \text{CSSR} + \text{CDR} + \text{RetRate} + \text{HOSR} + \text{RSS} + \varepsilon \tag{1}$$

Where ε is residual error, expected to be approximately 0.

Substituting estimated coefficients and intercept, it produces equation (2) as follows

$$Y = 1.522 + 0.446(\text{CSSR}) - 0.382(\text{CDR}) + 0.076(\text{RetR}) + 0.293(\text{HOSR}) - 0.179(\text{RSS}) + \varepsilon \tag{2}$$

Using the model in equation (2) to predict for the first 60 days, having CSSR and HOSR as strongest positive predictors while CDR and RSS negatively impact on QoS. The results are shown in Table 10 and *Figure 6*.

Table 10: Results of Actual and Predicted QoS using KPI Parameters for the 60 days

S/N	CSSR	CDR	RetR	HOSR	RSS	Actual QoS	Predicted QoS
1	5	1	5	2	3	4	4
2	4	3	4	3	2	3	3
3	4	4	3	5	2	3	3
.
.
.
239	5	3	4	4	2	4	4
240	5	3	4	5	2	4	4

Figure 6 depicts graph of the above data for actual and predicted QoS for the 2 months as follows. From the graph, it shows the models is good enough to be used in predicting QoS for MTOs, because there was no wide gap between the predicted and actual QoS values.

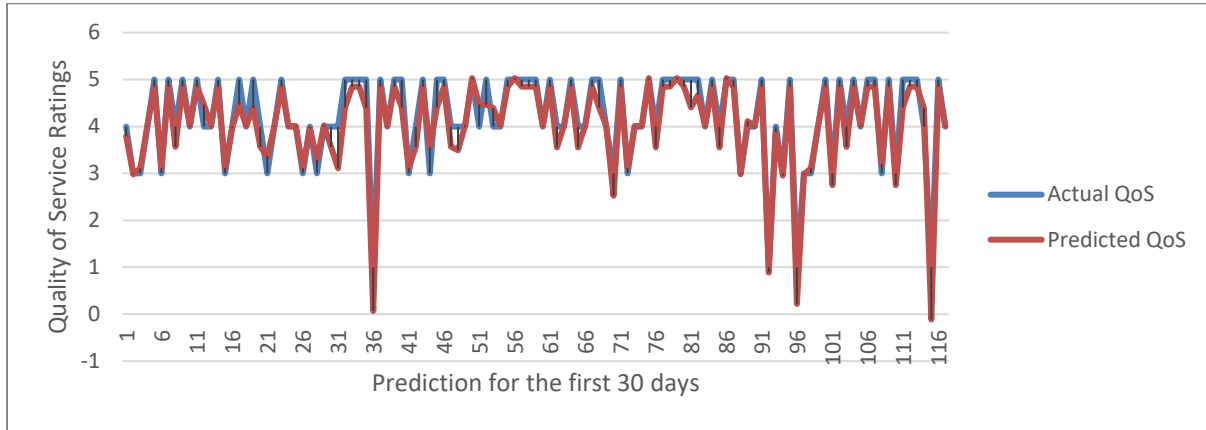


Figure 6: Line Graph Illustrating the Predicted and Actual QoS of MTOs

E. Regression Model Accuracy

One of methods to check whether regression model is accurate or not is to plot histogram of residual errors, which usually follow normal curve, and the histogram in Figure 7 produces histogram that is not far from normal curve.

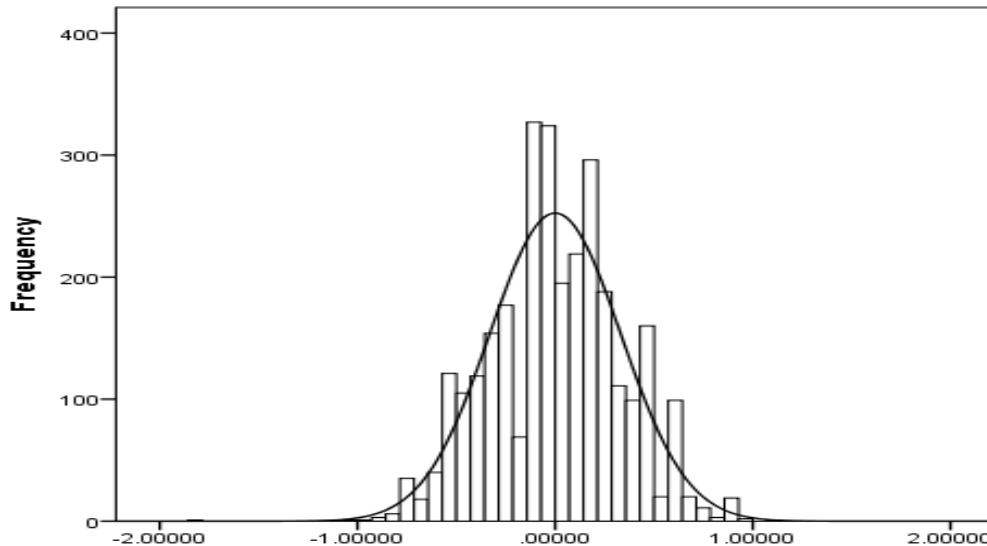


Figure 7: Histogram of Residual Error from Actual and Predicted QoS.

In addition to the plot, it is expected that mean of residual errors usually approximated to zero. The result of mean of residual errors is presented in Table 11.

Table 11: Measure of Central Tendency of Residual Errors

Measure of Central tendency	Estimated value
Mean	.0000000
Std. Deviation	.33214790
Variance	.110
Number of samples	2944

From result in Table 11, the mean of residual errors is estimated to zero, therefore, the regression model is good and such model can be used effectively use for predictive system.

F. Effect of Data-Driven Analytical Approaches

The data-driven analytical approaches applied correlation, ANOVA, and regression had the following effects:

- (a). Relationship Discovery: Correlation revealed hidden inter-operator dependencies (strong operatorM- operatorA link), enabling understanding of infrastructure synergies.
- (b). Group Comparison: ANOVA identified operator-specific variations as the primary QoS driver, shifting focus from geography to capability.
- (c). Prediction Accuracy: Regression's high R^2 (0.915) allows proactive forecasting from routine KPIs, reducing reactive interventions.
- (d). Actionable Insights: These techniques transformed descriptive data into knowledge for targeted upgrades.

Overall, these data-driven approaches proved superior to traditional reporting by uncovering patterns and predictions essential for telecom optimization.

IV. CONCLUSION

Data-driven approaches reveal clear QoS hierarchies and enable accurate prediction. Recommendations: target handover/drop improvements in G/E; integrate data mining for regulatory monitoring.

REFERENCES

- Ali, A., Ahmad, I., & Rehman, A. (2022). Data mining applications in telecommunication industry: A review. *Journal of King Saud University - Computer and Information Sciences*, 34(6), 2785–2796. <https://doi.org/10.1016/j.jksuci.2022.03.012>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54. <https://doi.org/10.1609/aimag.v17i3.1230>
- Garg, V. K. (2007). *Wireless communications and networking*. Morgan Kaufmann. Retrieved from <https://www.elsevier.com/books/wireless-communications-networking/garg/978-0-12-373580-5>
- GSMA. (2025). *The mobile economy Sub-Saharan Africa 2025*. GSMA Intelligence. <https://www.gsma.com/mobileeconomy/sub-saharan-africa/>
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann. Retrieved from <https://www.elsevier.com/books/datamining/han/978-0-12-811760-6>
- Ian, P. (2020). *GSM network architecture and elements*. Electronics Notes. <https://www.electronics-notes.com/articles/connectivity/2g-gsm/network-architecture.php>
- Imran, M., Katranaras, E., Seyedebrahimi, M., & di Bernardo, M. (2019). New vertical handover prediction schemes for LTE-WLAN heterogeneous networks. *PLoS ONE*, 14(4), Article e0216964. <https://doi.org/10.1371/journal.pone.0216964>

- Joseph, A., & Ojo, F. K. (2023). Assessment of quality of service in Nigerian GSM networks. *Nigerian Journal of Technology*, 42(1), 45–56.
- Nigerian Communications Commission (NCC). (2024a). Industry statistics: Subscriber data. <https://www.ncc.gov.ng/stakeholder/statistics-reports/industry-information/subscriber-statistics>
- Nigerian Communications Commission (NCC). (2024c). Quality of Service reports and fines. NCC Official Report.
- Oughton, E. J., & Frias, Z. (2023). Estimating digital infrastructure investment needs to achieve universal broadband. *IMF Working Papers*, 2023(027). International Monetary Fund. <https://doi.org/10.5089/9798400233623.001>
- Ozovehe, A. J., & Usman, A. D. (2015). Performance analysis of GSM networks in Minna metropolis of Nigeria. *Nigerian Journal of Technology*, 34(2), 359–367. <https://doi.org/10.4314/njt.v34i2.21>