35th CIRP Design 2025

An Indirect Measurement Method for Accurate Quality of Service Assessment

Faadhil Assayfullooh*, Kaito Uchiyama, Kahoru Furuya, Yoshiki Shimomura

Faculty of System Design, Tokyo Metropolitan University, Asahigaoka 6-6, Hino-shi, Tokyo, Japan

* Corresponding author. Tel.: +81-70-6949-1010. *E-mail address:* assayfullooh-Faadhil@ed.tmu.ac.jp

Abstract

Existing methods for assessing service quality rely on customer feedback, which can introduce bias and fail to capture true feelings and causes. This research proposes a novel approach that models service quality by analyzing changes in customer behavior rather than direct feedback. We construct a predictive model using gradient boosting regression techniques to analyze big data and identify how various aspects of service provider performance (the "5W1H"—Who, What, When, Where, Why, and How) impact customer behavior metrics. This model provides an objective and detailed assessment of service quality, revealing complex patterns and interactions often missed by traditional feedback methods.

Keywords: Service quality assessment; Customer behavior; Predictive model; Regression technique; Gradient boosting.

1. Introduction

In today's competitive market, the imperative to consistently deliver high-quality service has become a vital factor for the success of companies in all industries. Companies must guarantee that their service quality meets or exceeds customer expectations, as this is the primary differentiator that distinguishes them from their competitors [1]. Consequently, companies must conduct quality assessments to gather pertinent information or feedback for their improvement initiatives, which serve as countermeasure steps. Bias or erroneous information can lead to actions that adversely affect a company's competitiveness and sustainability.

The conventional approach of evaluating service quality predominantly depends on direct consumer feedback, usually collected via surveys or interviews [2]. Although these methodologies yield significant insights, they are frequently time-intensive, not easily replicable, prone to bias, and may not comprehensively reflect customer satisfaction.

This research provides a unique, indirect approach to assessing service quality by utilizing machine learning techniques to overcome existing restrictions. This strategy seeks to forecast service quality by evaluating previous service activity data and customer behavior patterns, eliminating the necessity for direct customer input. This method facilitates a more objective, efficient, and scalable assessment of service performance, providing profound insights into areas requiring enhancement.

2. Literature Review

2.1. Service Quality

Researchers have extensively developed service quality models, including notable frameworks such as technical and functional quality [3], SERVQUAL gap model [2], and SERPERF [3]. Service quality models can be categorized into two groups based on their measurement methods: direct and indirect.

To assess service quality, the direct model relies on customer or internal personnel feedback. The three models mentioned earlier [1][2][3] and other models, including the

synthesized model of service quality [4], the attribute service quality model [5], and the online banking model [6], all of which utilize direct feedback to assess service attributes and overall quality. The PCP attribute model [7], as well as the retail service quality and perceived value model [8], emphasize the collection of direct feedback to improve service quality across various situations.

On the other hand, indirect model assesses service quality using behavioral data analysis, operational performance indicators, or proxy variables, such as customer retention rates, sales growth, or complaint levels, without soliciting direct feedback from customers. This approach is implicitly suggested by Grönroos [9]. For example, the internal service quality DEA model [10] uses operational indicators, such as resource utilization, to rate the quality of internal service. However, although the IT alignment model [11] uses system data and IT performance to improve quality, it cannot be categorized as indirect method because its objective is to implement improvement actions for service delivery rather than to measure service quality.

2.2. Service Engineering

According to service engineering theory [13], which is shown in figure 1, a service is an activity carried out by the provider that changes the receiver's state. This change can manifest in various forms, including behaviors, thoughts, and actions. If we reverse the logic, then every change of state in the customer can be associated with the service performance factor that was received previously.

In relation to this concept, the direct method attempts to measure the manifestation of state changes in customers' thoughts by asking direct questions, which may be influenced by bias or subjectivity responses. On the other hand, another form of state change, e.g., behavior, such as increased engagement, reduced churn, or altered purchasing patterns, might serve as more objective proxies for assessing service performance, providing a more nuanced and real-time measure of how well the service succeeded in transforming the customer's state.
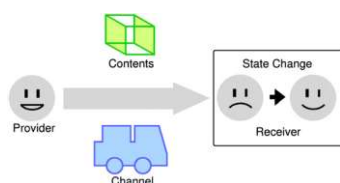

Fig. 1. Service framework [13].

The service framework depicted in figure 1 illustrates that the elements of a service activity comprise a "provider," denoting the individual executing an action (who), and a "receiver," referring to the customer who gets the service and undergoes a change in state as a consequence (who). The "contents" refers to something which is received by the service recipient or object of service (what). And the "channel" refers to the medium used to provide the service, i.e., the delivery method (how). Additionally, when analyzing the more intangible of factors related to service activities, we can

consider the temporal factor (when), the locational factor (where), or the contextual factor (why) of the service activity. Thus, the 5W1H concept can be used to analyze factors related to service activities.

2.3. Research Gap and Objective

Current indirect service quality approaches primarily emphasize operational metrics and internal resource utilization, such as response time, system uptime, or resource efficiency, to assess service quality without soliciting customer feedback. While this method provides useful insights for assessing service delivery efficiency, they fall short in addressing a critical gap: the use of customer behavior changes post-service as a conceptual means to evaluate service quality. Furthermore, direct methods, such as surveys or questionnaires, fail to examine the alterations in customer behavior post-service. This behavioral change may better reflect the true quality of the service, as it can provide a more accurate representation of customers' genuine perceptions than direct feedback. Unlike questionnaires or surveys—which are prone to bias and cannot be administered continuously—behavioral data offers a more objective and consistent measure of service effectiveness.

3. Methodology

This section explains the development of a method that combines the 5W1H framework with customer behavior analysis as an indicator of state change post-service. We describe how this approach connects with the KPI database and how machine learning techniques are applied to model these relationships.

3.1. Framework for Indirect Service Quality Model

In this research, we measure service quality by analyzing changes in the customer's state that reflected in their behavior after receiving service. The 5W1H framework is used to examine factors related to service delivery that may influence these behavioral outcomes. The resulting measurement is represented by a Service Quality Index (SQI).

Using regression analysis, this study indirectly assesses service quality by predicting changes in customer behavior, which may better reflect their true perceptions of the service. This approach leverages historical service performance data and employs machine learning to develop an accurate service delivery model.
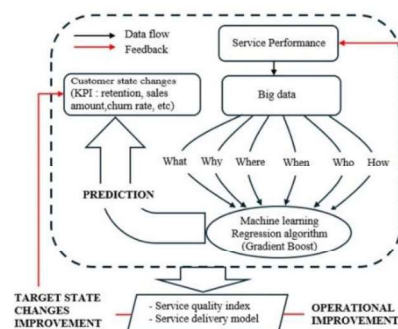

Fig. 2. Indirect service quality framework.

In addition, machine learning algorithms, particularly regression methods, have the capability to extract influential coefficients that impact the predicted model, providing service providers with valuable insights into their service delivery performance's strengths and weaknesses. This allows them to identify unbiased targets for service quality improvement. Figure 2 shows the framework for this research's indirect service quality method.

3.2. Service Quality Index

The SQI in this research reflects the cumulative impact of multiple customer state changes, in form of behaviors after service such as purchase amount, visit time, variation in purchases, and churn rate. Rather than focusing on a single metric, the index aggregates how well the service aligns with multiple expected behavior changes set by the service provider, offering a comprehensive view of service quality. In other words, this approach measures the gap between the provider's expectations and the actual observed customer behaviors, thereby offering a new perspective on gap theory to measure service quality.

The SQI is calculated by determining the percentage gap between the target outcome (provider's expectations) and the actual outcome (observed customer behavior) for several predefined key performance indicators (KPIs). The service quality gap for each KPI is then combined into a weighted average to derive the overall SQI.

3.2.1. Gap Calculation for Each KPI

For each KPI, the percentage gap is computed using the following formula:

$$Gap\ (\%) = \left(\frac{Target\ outcome - Actual\ outcome}{Target\ Outcome}\right) \times 100 \qquad (1)$$

This process is repeated for all customer behaviors under consideration and with minimum gap value is 0%.

3.2.2. KPI Weight

Given that each KPI may contribute differently to the overall success of the company, weights are assigned to each target behavior. The formula for the Total Weighted Gap is:

$$Total\ Weighted\ Gap = \sum_{i=1}^{n} \omega_i \times |KPI_i| \qquad (2)$$

ω_i: The weight assigned to the i-th KPI
KPI_i: The gap for the i-th KPI
n: The total number of KPIs

3.2.3. Interpretation of Service Quality

The SQI provides a normalized measure of how well the service aligns with the provider's expectations.

- **Low SQI (closer to 0%)**: The provider's targets for customer behavior are largely being met, indicating effective service delivery.

- **High SQI (closer to 100%)**: The actual customer behaviors deviate significantly from expectations, highlighting underperformance in certain service areas.

3.3. Machine Learning Algorithm – Gradient Boosting

To anticipate a non-linearity in the actual regression scenario, we employ the gradient-boosting approach in this research. This is a machine learning technique for constructing a series of models that rectify the errors of previous iterations. Gradient boosting is chosen for this methodology because of its robust prediction powers, particularly in intricate and non-linear situations, rendering it ideal for examining consumer behaviors and enhancing service quality [14].

3.4. Application Step

To implement the methodology, we follow these steps:
- Step 1: Collect service delivery data
- Step 2: Identify and Define KPIs for Customer Behavior Change.
- Step 3: Categorize Variables using the 5W1H Framework.
- Step 4: Apply Gradient Boosting to Calculate Service Quality Index (SQI) and Develop Predictive Model.

4. Application case

The methodology in this research is applied to the service repair and maintenance sector, with a specific focus on measuring the service quality of an automotive repair shop in Indonesia. The workshop's daily transaction database from March 2020 to February 2022 provides the data for analysis, with the assumption that the database analysis begins in March 2022 and focuses solely on transactions from personal customers. Organizational customers are excluded due to their contractual relationship with the shop, which may affect their behavior beyond the service performance factors. The process involved in this study follows these key steps:

4.1.1. Step 1: Collect service delivery data

Service delivery data was collected from daily service transactions recorded at an authorized automotive repair shop. This data was extracted in CSV format from the shop's operational software. The data is automatically transmitted to the company's principal for KPI analysis, providing a consistent and structured dataset for evaluating service quality.

4.1.2. Step 1: Identify and Define KPIs for Customer Behavior Change

Four KPIs or dependent variables will be used in this study are: *purchase amount per customer*, *number of visits*, *variety of menus purchased*, and *churn rate*. The target values for each KPI and its weight can be found in Table 1.

Table 1. An example of a table.

KPI Item per customer (average)	Target	Weight
Purchase amount per visit	IDR 500,000	0.3

Number of visits / 6-month period	1		0.2
Variety of menu purchased	3		0.1
Churn rate	20%		0.4

4.1.3. Step 2: Categorize Variables using the 5W1H Framework

We categorize the independent variables data or features in machine learning terms, related to service performance using the 5W1H framework. There which are extracted from the available transaction database, as shown in Table 2.

Table 2. List of variables data using 5W1H domain.

Domain	Data	Content
When	Day	Monday / Tuesday… etc.
	Time	12:39
	Period	Morning / Afternoon
Who (Staff)	Name (Reception)	A / B / C
	Name (Technician)	D / E / F
What (Object)	Car Brand	TOYOTA / BMW … etc.
	Car Age	1 / 2 / 3 … etc.
	Service Menu	Engine / Transmission / rake / Underbody / lectrical / A/C / Other
What (Customer)	Status	New / Repeat Customer
	Sex	Man / Woman
What (Previous Result)	Purchase amount	IDR 500,000
	Number of visits	1 / 2 / 3 … etc.
	Variety of menu	1 / 2 / 3 … etc.
	Losing customer	Yes / No
How (long)	Service Duration	1 / 2 / 3 … minutes
Where	Unused (same place)	
Why	**Not available**	

4.1.4. Step 3: Apply Gradient Boosting to Calculate Service Quality Index (SQI) and Develop Predictive Model

In this step, we prepare the data for machine learning by applying essential preprocessing techniques, including feature engineering and data cleaning. Following data preparation, we conduct hyperparameter tuning to optimize model accuracy.

To evaluate the model's performance, we use different metrics tailored to the nature of the data, as continuous and classification data have distinct characteristics and evaluation needs. For continuous data, where the goal is to predict a numerical value, we use metrics such as Root Mean Square Error (RMSE) and R-squared (R²). For classification data, where the objective is to categorize data points into distinct classes, we rely on a classification report that includes accuracy, precision, recall, and F1-score.

4.2. Result - Service Quality Index

Using formula (1), we can get the quality gap from the data as shown in table 3.

Table 3. KPI result.

KPI Item per customer (average)	Target	Result	Gap (%)
Purchase amount per visit	500,000	434,605	13.08
Number of visits / 6-month period	1	0.2	78.24
Variety of menu purchased	3	2	33.61

Churn rate	20%	40%	101.29

Then, with using formula (1), finally we can calculate the SQI as follows:

$$SQI = (\omega_1 \times |Gap_1|) + (\omega_2 \times |Gap_2|) + (\omega_3 \times |Gap_3|) + (\omega_4 \times |Gap_4|)$$

$$SQI = (0.3 \times 13.08) + (0.2 \times 78.24) + (0.3 \times 33.61) + (0.4 \times 101.29)$$

Service Quality Index $= 63.45\,\%$

This result indicates that there is still a 63.45% gap between the actual behavior of customers and the expectations of the service provider.

4.3. Result - Service Delivery Model

We categorize machine learning models based on the type of output—either continuous or categorical—and use the appropriate performance metrics for each type. In this case, three KPIs(1, 2, and 3) are continuous, while one (4) will be categorical. This is because the churn rate will be represented indirectly through the prediction of the "losing customer" category, rather than as a direct percentage number. Therefore, the performance report outputs will be displayed separately for each type.

4.3.1. Evaluation Performance Result – Continuous Data

To evaluate the performance of predictions for continuous data, we use RMSE and R². RMSE measures the average difference between the predicted values and the actual values, expressed in the same units as the target variable. It quantifies the amount of error in a regression model; the lower the RMSE, the better the model fits the data. R^2 on the other hand, indicates how well the model explains the variance in the target variable. It ranges from 0 to 1, where a value closer to 1 means the model explains a higher proportion of the variability in the target data. An R² of 1 would mean the model perfectly explains all the variability in the target data, while an R² of 0 would mean it explains none.

Table 4. Performance evaluation of the service delivery model.

Prediction Model	RMSE	R²
Purchase amount	256,361	41%
Number of visits	0.37	93%
Variety of menu purchased	0.82	5%

Table 4 shows the performance of the predictive models. It indicates that the current model is only well-suited for predicting visit frequency, as evidenced by low RMSE and high R² values. However, the model's performance for predicting purchase amount is weaker, and even worse for variety of menu purchased, it is suggesting that additional features or an alternative approach may be needed to improve its accuracy.

To understand which features influence the model's predictions, we can refer to the top 5 important features identified in each model, as shown in Figures 3 to 5.
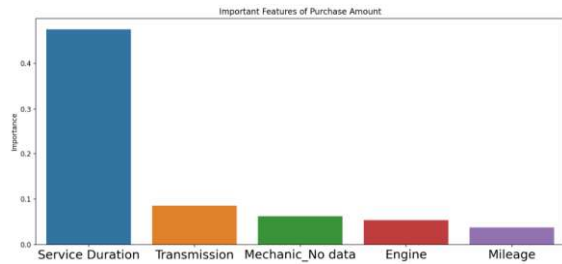

Fig. 3. Importance features of purchase amount

Fig. 3 shows that service duration (How) is the most influential feature for predicting the purchase amount, which is reasonable since longer service durations typically result in higher charges. However, service duration alone only explains 41% of the variance in the data. This indicates that there may be other influential factors affecting the purchase amount. The low impact of the other existing features suggests that more relevant features might be missing, or there could be inaccuracies in the service duration data itself.
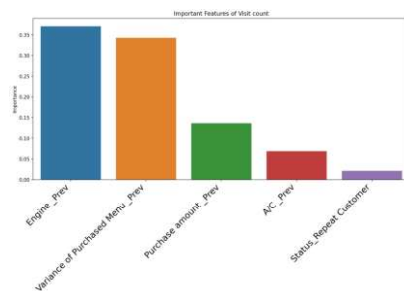

Fig .4. Importance features of number of visits

Fig. 4 shows that information related to previous results (What) is highly dominant compared to other features for predicting visit count. This suggests that past service types and customer behavior, particularly previous menu choices, are crucial factors influencing the prediction of visit count.
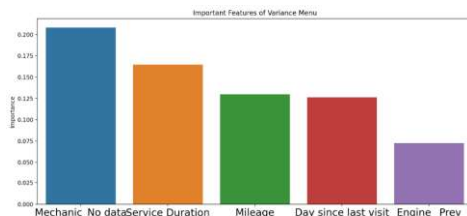

Fig .5. Importance features of variety of menu purchased

Fig. 5 shows that the contribution of features appears to be evenly distributed with mechanic (Who) is most dominant, suggesting that no single feature is significantly influencing the prediction.

4.3.2. Evaluation Performance Result – Categorical Data

To evaluate the performance of predictions for categorical data we will use accuracy, precision, recall, and F1 Score.

Accuracy is the ratio of correctly predicted observations to the total number of observations, providing an overall measure of the model's performance. Precision is the ratio of correctly predicted positive observations to the total predicted positives, indicating the quality of positive predictions. Recall is the ratio of correctly predicted positive observations to all actual positives, measuring the model's ability to identify positive cases. F1 Score is the harmonic mean of precision and recall, balancing the two metrics to provide a single score that considers both false positives and false negatives. These metrics help ensure a comprehensive evaluation of the model's performance, especially in imbalanced datasets. The classification performance report can be seen in Fig 6.

Fig .6. Classification report for the 'Losing Customer' prediction model

In Fig. 6, the classification report evaluates the model's performance in predicting the 'Losing Customer' outcome, with "No" representing customers who did not leave, and "Yes" indicating customers who did leave. The performance indicators include precision, recall, and F1-score for both classes ("No" and "Yes").

The model performs better for the "No" class, as seen in its higher precision, recall, and F1-score compared to the "Yes" class. This suggests that the model is more accurate in identifying customers who are likely to stay (predicted as "No") but struggles to identify those who may leave (predicted as "Yes"). The lower precision and recall for the "Yes" class imply that the model has difficulty detecting leaving customers accurately and may have a tendency to predict "No" more often than "Yes".

The "support" column represents the number of actual instances in each class within the dataset: there are 172 instances of "No" (customers who stayed) and 102 instances of "Yes" (customers who left). This imbalance in class distribution might be affecting the model's ability to predict the minority class ("Yes") accurately. To address this, additional techniques, such as rebalancing the dataset, may help improve the model's performance for predicting leaving customers.
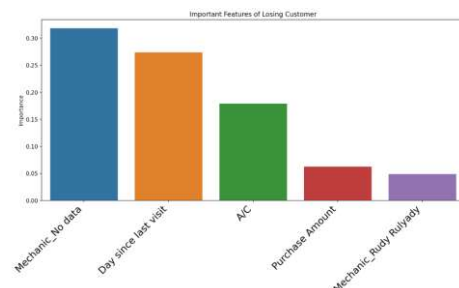

Fig .7. Importance features of churn rate

Fig. 7 shows the top 5 important features for predicting customer loss. It is evident that the mechanic factor (Who) is the top contributor. However, due to incomplete source data, it

is not possible to trace back which mechanic was involved. Despite this, another mechanic, named "Rudy Rulyady" appears as the lowest-ranked feature among the top contributors. This suggests that, while he may have a lesser direct influence on the model compared to other factors, his involvement still significantly impacts customer retention, making him the most notable individual mechanic associated with customer loss.
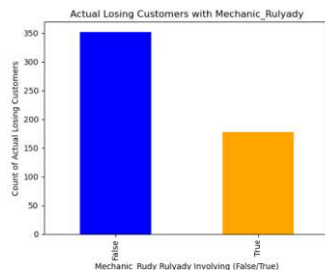

Fig .8. Relation of Actual losing customer data with Rudy Rulyady

To validate this assumption, we can confirm the actual relationship between mechanic Rudy Rulyady and the losing customer phenomenon, as shown in Fig. 8. The data indicates that Rudy Rulyady has minimal involvement in the losing customer factor. This supports the prediction results, even though the model's accuracy is moderate.

5. Discussion

Although the accuracy of the models in this study was not optimal for predicting service performance, valuable insights can still be drawn regarding the most influential features using the 5W1H framework. The influential features identified from initial predictions can inform further analysis to explore additional related factors. For instance, when predicting losing customers, mechanics emerged as a significant factor. This insight allows for subsequent predictions to identify other features that may impact mechanic's performance, fostering a continuous improvement loop. This iterative process can help pinpoint key factors that require attention or uncover hidden patterns in service performance.

Suboptimal accuracy results may be due to two main reasons: non-optimized parameter settings or incomplete data. Addressing data completeness can provide crucial feedback for service companies, emphasizing the need to identify and collect relevant data for effective performance monitoring. The 5W1H framework can serve as a valuable guide in this process.

While this method shows promise for application in the service sector, it has certain limitations, particularly concerning data quality. The accuracy of machine learning predictions relies on the validity and relevance of the data, the more robust the data, the better the prediction outcomes. Additionally, this method has not been extensively tested in service sectors beyond maintenance and repair, which may have unique characteristics and patterns that could affect its adaptability and performance.

6. Conclusion

This study introduces a new perspective for assessing service quality, shifting from the traditional customer viewpoint to the provider's perspective. This approach aims to minimize bias and subjectivity, which can lead to inaccurate information for service improvement. Additionally, we propose a framework and measurement method that not only yield a more objective service quality index but also provide actionable insights for implementing improvements. In future work, we plan to apply this method to other types of service activities while continuing to refine machine learning performance.

Acknowledgements

References

[1] Boshoff C, Gray B. The relationships between service quality, customer satisfaction and buying intentions in the private hospital industry. South African Journal of Business Management. 2004;35(4):27-38.

[2] Parasuraman A, Zeithaml VA, Berry LL. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing. 1988;64(1):12-40.

[3] Grönroos C. A service quality model and its marketing implications. European Journal of Marketing. 1984;18(4):36-44.

[4] Cronin JJ, Taylor SA. SERVPERF versus SERVQUAL: Reconciling performance-based and perception-minus-expectations measurement of service quality. Journal of Marketing. 1994;58(1):125-131.

[5] Brogowicz AA, Delene LM, Lyth DM. A synthesised service quality model with managerial implications. International Journal of Service Industry Management. 1990;1(1):27-44.

[6] Haywood-Farmer J. A conceptual model of service quality. International Journal of Operations & Production Management. 1988;8(6):19-29.

[7] Broderick AJ, Vachirapornpuk S. Service quality in internet banking: The importance of customer role. Marketing Intelligence & Planning. 2002;20(6):327-335.

[8] Philip G, Hazlett SA. The measurement of service quality: A new P-C-P attributes model. International Journal of Quality & Reliability Management. 1997;14(3):260-286.

[9] Sweeney JC, Soutar GN, Johnson LW. Retail service quality and perceived value. Journal of Consumer Service. 1997;4(1):39-48.

[10] Grönroos C. Service Management and Marketing: Managing the Moments of Truth in Service Competition. Lexington Books; 1990.

[11] Soteriou AC, Stavrinides Y. An internal customer service quality data envelope analysis model for bank branches. International Journal of Bank Marketing. 2000;18(5):246-252.

[12] Berkley BJ, Gupta A. Improving service quality with information technology. International Journal of Information Management. 1994;14:109-121.

[13] Luo X, Homburg C. Neglected outcomes of customer satisfaction. Journal of Marketing. 2007;71(2):133-149.

[14] Shimomura Y., Tomiyama T. Service Modeling for Service Engineering. Proceedings of The 5th Intl. Conf. on Design of Information Infrastructure Systems for Manufacturing 2002;309-316.

[15] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of Statistics. 2001;29(5):1189-1232.