Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 3, September 2026, pp. 697–706
e-ISSN: 2964-2922, p-ISSN: 2963-6191

Open access

Original Research

Customer Service Time Prediction Using Gradient Boosting Regressor for Sequential Queuing Systems

Amelia Sari Dewi [a*], Eva Rahmawati [b], Sri Diantika [c]

[a,b,c] Department of Information Systems, Faculty of Information Technology, Universitas Bina Sarana Informatika, Jakarta, Indonesia

*Corresponding Author: ameliasarid0@gmail.com

ABSTRACT

Service quality in financial service companies depends on the ability to provide efficient and timely customer services. PT FIFGROUP Cikampek Branch utilizes the Integrated Self-Service Smart Queue Information System (ISS-SIAP), which implements a sequential queueing system where customers may pass through different service stages with varying processing durations. This complexity makes customer service time prediction challenging and may affect resource allocation and queue management. This study aims to develop a customer service time prediction model using the Gradient Boosting Regressor algorithm based on historical ISS-SIAP data. A quantitative approach was applied through data collection, preprocessing, 10-Fold Cross Validation, model development, and performance evaluation using R^2, MAE, MSE, and RMSE metrics. The experimental results show that the proposed model achieved an R^2 value of 0.866, MAE of 2.726 minutes, MSE of 15.308, and RMSE of 3.903 minutes. These results demonstrate that the Gradient Boosting Regressor provides good predictive performance in estimating customer service time. The proposed model provides practical support for improving operational efficiency at PT FIFGROUP Cikampek by enabling better service planning, resource allocation, and queue management decisions.

How to Cite:
A. S. Dewi, E. Rahmawati, and S. Diantika, “Customer Service Time Prediction Using Gradient Boosting Regressor for Sequential Queuing Systems,” *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 5, no. 3, pp. 689–706, 2026.
Doi: https://doi.org/10.69916/jkbti.v5i3.590

KEYWORDS

Gradient Boosting Regressor
ISS-SIAP
K-Fold Cross Validation
Machine Learning
Service Time Prediction

ARTICLE INFO

Received: July 18, 2026
Accepted: August 28, 2026
Available Online: September 18, 2026

1 INTRODUCTION

In everyday life, queuing is a common activity that occurs as a consequence of the need for a service or goods, whether in the buying and selling process, obtaining health services, or various other forms of transactions. This phenomenon occurs when the demand for services exceeds the available service capacity, so that individuals must wait to receive services [1, 2]. In the service sector, the success of a financing service company lies in the customer satisfaction index which is based on several factors including waiting time, service time, and service quality.

As one of the largest financing companies in Indonesia, PT FIFGROUP has implemented the Integrated Self-Service Smart Queue Information System (ISS-SIAP) to manage customer service processes across all branches, including the Cikampek Branch. This system applies the sequential queue concept, meaning each customer receives only one queue number to undergo several service stages depending on the type of service selected. This characteristic causes service durations to vary and be difficult to predict, potentially increasing waiting times, reducing resource utilization efficiency, and lowering customer satisfaction levels if not managed optimally.

Previous studies addressing queuing and service-time prediction fall into two broad streams. Classical queuing research such as work using M/M/S formulations provides analytical tools for capacity planning and server-count adjustments to improve average waiting times and utilization [5]. Separately, a growing body of machine-learning research (including ensemble approaches) has demonstrated strong point-prediction performance across diverse applications; for example, ensemble and Gradient Boosting methods have been applied successfully in service-time or load prediction contexts and many other domains [3, 4, 6–11]. These literatures emphasize different strengths: queuing theory offers closed-form insights under stochastic assumptions, while machine-learning approaches excel in capturing complex, non-linear mappings when rich historical data are available.

For a multi-stage sequential system such as ISS-SIAP, the service process combines path dependence (different combinations of stages per customer) and stage-level heterogeneity in processing times characteristics that limit the direct applicability of simple analytical queuing formulas. In this context, a data-driven ensemble method is appropriate. The Gradient Boosting Regressor (GBR) is an additive, stagewise ensemble that models non-linear interactions and can ingest mixed categorical and numerical operational features. Prior studies report high coefficients of determination and low prediction errors for GBR in varied prediction tasks [6–11], supporting its suitability for mapping ISS-SIAP operational attributes (e.g., counters visited, arrival time, and service type) to total service time.

Recent developments in machine learning have shown significant potential in improving waiting time prediction in complex service systems. Unlike traditional queueing models that rely on probabilistic assumptions, machine learning approaches are capable of capturing non-linear relationships and dynamic patterns from historical operational data [12, 13]. Several studies have demonstrated that ensemble learning methods, particularly tree-based algorithms such as Gradient Boosting, can achieve high predictive accuracy in various domains [14, 15]. Moreover, the integration of predictive models into operational systems enables real-time decision-making, allowing organizations to optimize resource allocation and reduce service bottlenecks [16]. These advancements highlight the growing relevance of machine learning in addressing the limitations of classical queueing approaches, especially in multi-stage and sequential service environments [17].

However, most previous research has focused on research subjects outside the service sector of finance companies. Research on predicting service times in sequential queuing systems with multiple service stages within a single queue number, particularly in the ISS-SIAP system at PT FIFGROUP's Cikampek Branch, is still very limited. Based on the gap, this study proposes the use of the Gradient Boosting Regressor algorithm to build a customer service time prediction model based on historical data obtained from the ISS-SIAP system. The novelty of this study lies in the application of the Gradient Boosting Regressor algorithm to a sequential queuing system in a finance company environment by utilizing real operational data and validating it using 10-Fold Cross Validation to obtain a prediction model that has good generalization capabilities. The resulting model is expected to be able to help companies in estimating service times more accurately so that it can support resource planning and improve service quality [18–20].

Based on the description, this study aims to identify the characteristics of the ISS-SIAP queuing system, build a service time prediction model using the Gradient Boosting Regressor algorithm, evaluate the model performance using the R-Square (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) metrics, and provide recommendations for PT FIFGROUP Cikampek Branch in improving service effectiveness based on the prediction results obtained. This research is limited to the analysis and prediction of service times using historical queue data sourced from the ISS-SIAP application at the FIFGROUP Cikampek Branch. The data used is queue data for a period of 13 working days (Monday - Saturday), namely April 1, 2026 to April 15, 2026. This research focuses on a sequential queue system, where each customer only gets one queue number for various types of services. The variables analyzed are limited to data available in the system, such as arrival time, service time, and other related attributes. The results of this study do not cover the direct implementation of the system, but are limited to the modeling and evaluation stages of the prediction model performance.

2 RESEARCH METHODS

This study uses a quantitative research method with a machine learning-based predictive approach to develop a customer service time prediction model. The research stages were carried out systematically, starting with problem identification, historical data collection, data preprocessing, model validation using 10-Fold Cross Validation, modeling using the Gradient Boosting Regressor algorithm, and model performance evaluation, as depicted in Figure 1.

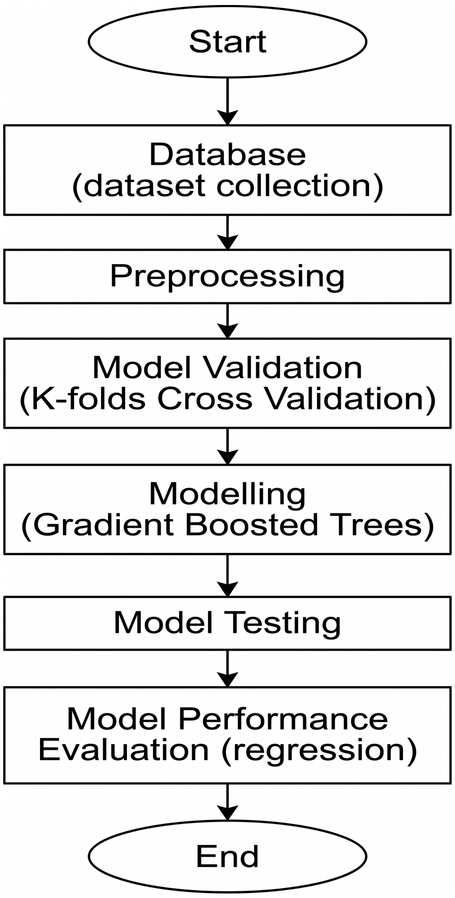

Figure 1. Research Stages

This series of stages is designed to produce a prediction model that is capable of providing accurate prediction results and has good generalization capabilities to new data.

2.1 Dataset and Data Collection

The object of this study is customer service time recorded in the ISS-SIAP queuing system used by FIFGROUP Cikampek Branch. Data collection was conducted using three main techniques:

a. Participatory Observation

Researchers were directly involved in the use of the ISS-SIAP queuing system during the service process to understand the service flow in depth.

b. Literature Study

Reviewing journals, books, and previous scientific publications related to waiting time prediction, queuing systems, and ensemble algorithms.

c. Documentation

Extracting historical customer service data directly from the ISS-SIAP system for the period from April 1, 2026 to April 15, 2026 (13 working days) with an initial total of 1,227 service data records. Therefore, although the representativeness of the dataset across broader time periods remains limited, the model results still provide a valid representation of service patterns within the observed operational context and can serve as a foundational basis for further model development.

Table 1. Research Database Attribute Structure

No	Attribute	Description
1	Day	Customer service day
2	Arrival Time	Customer arrival time
3	Service Type	Selected service category
4	Via Cashier	Cashier service indicator (0 = No, 1 = Yes)
5	Via CCP	CCP service indicator (0 = No, 1 = Yes)
6	Via CRE	CRE service indicator (0 = No, 1 = Yes)
7	Total Service Time	Total service duration in time units (minutes)

2.2 Proposed Method

The method proposed in this study is the Gradient Boosting Regressor algorithm, an ensemble learning method that builds a predictive model through a series of regression trees in stages. The learning process begins by forming an initial prediction based on the average value of the target variable, then calculating the residual as the difference between the actual value and the predicted value. The algorithm then builds a regression tree to learn the error pattern and iteratively updates the prediction using the learning rate. To obtain the best model configuration, hyperparameter optimization was performed using the Grid Search (Optimize Parameters) method, while model validation was performed using 10-Fold Cross Validation to ensure the model has more stable predictive capabilities and does not experience overfitting.

a. Initial Prediction

The first stage is carried out by calculating the average value (mean) of the target variable (Total Service Time) in the training data as $F_0(x)$:

$$F_0(x) = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{1}$$

where N represents the number of training data and y_i is the actual value of customer service time.

b. Residual Calculation (Error)

The algorithm calculates the residual value (r_i) or the difference between the actual value and the previous model prediction at each iteration:

$$r_i = y_i - F_0(x) \tag{2}$$

c. Regression Tree Formation and Boosting Process

Based on the residual values, a regression tree is constructed to learn the error patterns. Instead of directly using the residuals, a weak learner $h_m(x)$ is trained to approximate the pseudo-residuals. The model is then updated iteratively as:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \tag{3}$$

where $h_m(x)$ represents the regression tree trained at iteration m, and η is the learning rate.

2.3 Experimental Setup

The experiment was conducted using Microsoft Excel for data preprocessing and RapidMiner Studio to build and evaluate the predictive model. Parameter optimization was performed using the Grid Search method, while model validation used 10-Fold Cross Validation, which divides the dataset into ten parts, with nine parts used as training data and one part as testing data at each iteration. This process was repeated until all folds were used as testing data. This approach was chosen because it can reduce data division bias and produce a more objective and representative model evaluation of data that has not been used in the training process.

2.4 Evaluation Metrics

Model performance was evaluated using four regression metrics: Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The R^2 value measures the model's ability to explain variations in actual data, while the MAE, MSE, and RMSE measure the model's prediction error. The higher the R^2 value and the lower the MAE, MSE, and RMSE values, the better the model's performance in predicting customer service times.

3 RESULTS AND DISCUSSION

3.1 Data Understanding and Preprocessing

The secondary dataset extracted from the ISS-SIAP web-based application at PT FIFGROUP Cikampek initially consisted of 1,227 records. During the data cleaning stage, customer queue records with missing or invalid queue-type entries were identified, particularly those with a Total Service Time equal to 0 (zero) minutes, which typically indicate cancelled transactions or input errors. In addition, records exhibiting abnormally long total service times (defined in this study as greater than 60 minutes) were flagged as anomalous, as such durations are inconsistent with normal operational conditions and may reflect data recording errors or exceptional cases. A total of 40 anomalous records were removed, resulting in a final dataset of 1,187 valid customer service records used for model development. A sample of the initial dataset is presented in Table 2.

Table 2. Sample ISS-SIAP Historical Service Dataset

TARGET	PREDICTOR ATTRIBUTES						
Total Service Time	**Day**	**Service Type**	**Arrival Time**	**Via Cashier**	**Via CCP**	**Via CRE**	**Cashier Officer**
4	1	0	8	1	0	0	1
5	1	0	8	1	0	0	1
9	1	1	8	1	1	1	1
6	1	1	8	0	1	1	0
5	1	0	8	1	0	0	2
...	...	...	...	...	...	...	...
3	5	0	14	1	0	0	2
4	5	0	15	1	0	0	1
6	5	0	15	1	0	0	2
3	5	0	16	1	0	0	2
18	5	0	16	1	0	0	2

The data transformation process is performed to convert nominal and time data into numeric form to suit the input of the Gradient Boosting algorithm. The Arrival Time attribute is rounded to the nearest hour (30-minute rounding), and the duration at the Cashier, CCP, and CRE counters is not used directly; instead, this information is transformed into binary variables (0 and 1) that represent the sequential flow of customer service, indicating whether a customer passes through specific service stages based on the service type selected at the time of queue registration. Consequently, the model only utilizes features that are available at the customers arrival, such as service type and service flow indicators. Target label encoding is completed using the Nominal to Numerical and Set Role operators in RapidMiner.

3.2 Model Validation and Hyperparameter Configuration

Model validation using 10-Fold Cross Validation randomly divided 1,187 service data sets into 10 evenly sized folds. Folds 1 to 7 had 119 test data sets and 1,068 training data sets, while Folds 8 to 10 had 118 test data sets and 1,069 training data sets. The final parameter configuration of the Gradient Boosting Regressor algorithm was obtained through automatic optimization using the Optimize Parameters (Grid) operator in RapidMiner Studio. The optimal parameter combination that resulted in the lowest error is presented in Table 3.

Table 3. Hyperparameter Optimization Results of Gradient Boosting Trees

Hyperparameter	Value	Function
Number of Trees	30	Determines the number of decision trees built sequentially.
Maximal Depth	2	Determines the maximum depth of each decision tree.
Learning Rate	0.1	Regulates the contribution of each tree in correcting prediction errors at each iteration.

3.3 Model Training Phase

The training process is described through manual calculation simulations on Fold-1 training data ($N = 1,068$ data). The initial step is to determine the basic prediction $F_0(x)$ using the average actual value of total service time:

$$F_0(x) = \frac{15159}{1068} = 14.19382 \approx 14.194 \text{ (minutes)} \tag{4}$$

Based on these initial values, the first residual (r_i) is calculated for each training sample data. The actual sample values and initial residual values are shown in Table 4.

Table 4. Initial Prediction Sample and Initial Residual Values (Fold-1)

No	Actual (y_i)	Initial Prediction ($F_0(x)$)	Residual (r_i)
1	22	14.194	7.806
2	22	14.194	7.806
3	26	14.194	11.806
4	7	14.194	-7.194
5	6	14.194	-8.194
...	...	...	...
1065	8	14.194	-6.194
1066	8	14.194	-6.194
1067	22	14.194	7.806
1068	18	14.194	3.806

Next, the algorithm constructs the first regression tree using the initial residual values as learning targets to find the separator with the greatest variance reduction. Based on the modeling results, the model provides meaningful insights into the key factors influencing customer service duration. One of the most notable findings is the selection of the "Via CCP" attribute as the root node of the regression tree, indicating that this variable has the highest contribution in explaining variance in service time. A representation of the separation structure of the first regression tree with a maximum depth of 2 is shown in Figure 2.

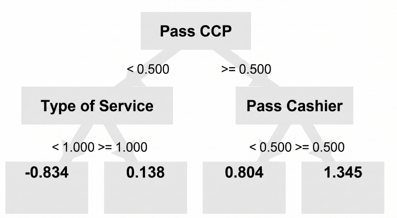

Figure 2. First Regression Tree Structure of Modeling Results

From an operational perspective, the "Via CCP" condition reflects additional service steps or intermediate processes that customers must undergo before reaching the main service point. This suggests the presence of potential bottlenecks within the service flow, which may increase both waiting time and overall service duration. Therefore, this finding has important business implications, highlighting that optimizing the CCP stage could be a strategic priority to improve service efficiency and reduce customer waiting time.

Based on Figure 2, the leaf node value represents the average residual of the data group. This value forms the basis for the correction factor. Through an additive boosting process with a learning rate ($\eta = 0.1$), the model predictions are updated for the next iteration. This iteration and residual update are executed continuously until the maximum limit of 30 decision trees is reached. The combination of all regression tree contributions produces the final model with the following additive formula:

$$F(x) = F_0(x) + \eta h_1(x) + \eta h_2(x) + \eta h_3(x) + \cdots + \eta h_{30}(x) \tag{5}$$

3.4 Performance Testing and Evaluation Phase

The final model was tested using independent testing data on Fold-1 ($n = 119$ data points). Using the Apply Model operator, the predicted total service time was calculated and compared directly with the actual data to obtain the squared error and absolute error values, as detailed in Table 5.

Table 5. Calculation of Prediction Error Values on Fold-1 Test Data

No	Total Service Time	Prediction	Error	Absolute Error	Squared Error
1	15.0	22.1	-7.1	7.1	50.2974
2	8.0	6.3	1.7	1.7	3.0333
3	5.0	6.0	-1.0	1.0	1.0586
4	4.0	6.3	-2.3	2.3	5.1002
5	4.0	6.0	-2.0	2.0	4.1164
...	...	...	...	...	...
115	8.0	6.3	1.7	1.7	3.0333
116	28.0	26.7	1.3	1.3	1.6877
117	19.0	15.5	3.5	3.5	12.4458
118	8.0	15.3	-7.3	7.3	53.6797
119	21.0	15.3	5.7	5.7	32.1870
TOTAL				**327.2463**	**1895.8532**

Based on the accumulated error values in Table 5, the regression performance metrics for the Fold-1 sample were calculated manually:

a. **Mean Absolute Error (MAE):** 2.750 (minutes). This proves that the average absolute deviation of the model prediction results is around 2.75 minutes from the actual service time.

b. **Mean Squared Error (MSE):** 15.9315. This value describes the sensitivity to extreme errors due to error squaring.

c. **Root Mean Squared Error (RMSE):** 3.9914 (minutes). The model's mean squared deviation is 3.99 minutes from the real target.

d. **Coefficient of Determination (R^2):** Using the total variation divisor against the mean of the data $R^2 = 1 - 0.1500 = 0.8500$ (85%). The model was shown to be able to control 85% of the service time variation in the Fold-1 test data.

3.5 Final Result Analysis

The manual testing results above were executed simultaneously on Fold-2 through Fold-10 using RapidMiner Studio. The mean and standard deviation of the overall final performance metrics are presented in Table 6.

Table 6. Final Model Evaluation Results

Evaluation Metric	Value
RMSE	3.903 ± 0.286
MAE	2.726 ± 0.152
MSE	15.308 ± 2.247
R^2 (Squared Correlation)	0.866 ± 0.013

Based on the comprehensive data presented in Table 6, the proposed Gradient Boosting Regressor model achieved an R^2 value of 0.866, indicating strong predictive capability in explaining customer service time variance and effectively capturing complex integrated sequential queue patterns. This performance is further supported by low error values, with MAE of 2.726 minutes and RMSE of 3.903 minutes, as well as relatively small standard deviations across all evaluation metrics, which indicate stable generalization performance on unseen data and a low risk of overfitting.

When compared to previous studies, this result is considered competitive. For instance, Ferinan et al. reported an R^2 value of 0.696 [4], while Rusdianto Hidayat et al. achieved a higher R^2 of 0.992 in a more structured sports analytics context [7]. Similarly, Gradient Boosting has demonstrated strong performance in various domains, including housing price prediction [6], flood prediction [8], student depression analysis [9], and climate modeling [10], all showing relatively low prediction errors. In contrast to traditional queuing models such as M/M/S, which focus on system-level averages and are limited in capturing dynamic and non-linear service patterns [5], machine learning approaches particularly ensemble methods offer greater flexibility in modeling complex sequential processes [3, 11]. These findings are consistent with recent studies which emphasize that machine learning models can provide more flexible and accurate predictions in complex service systems compared to traditional queueing approaches [12, 13].

Therefore, although the R^2 value in this study is not the highest reported, the model remains robust and reliable within the context of a real-world, multi-stage sequential queuing system characterized by high variability and operational uncertainty. Furthermore, the model provides a valid predictive basis for management to optimize the allocation of frontliners across service counters based on estimated service time workloads.

3.6 Discussion

This study successfully demonstrates that the Gradient Boosting Regressor (GBR) is a highly effective algorithm for predicting customer service times within a complex, multi-stage sequential queuing system. The model achieved a strong coefficient of determination (R^2) of 0.866, indicating that it explains 86.6% of the variance in total service time. Furthermore, the low error metrics, Mean Absolute Error (MAE) of 2.726 minutes and Root Mean Squared Error (RMSE) of 3.903 minutes, confirm the model's precision in real-world operational contexts.

A critical and highly actionable insight derived from the models internal structure is the identification of the "Via CCP" (CCP counter) attribute as the root node of the regression tree. This reveals that intermediate processing stages significantly drive service time variance, highlighting the CCP stage as a potential operational bottleneck that disproportionately impacts overall service duration.

The performance of the proposed GBR model aligns with and extends recent advancements in machine learning applications for service systems. Traditional queuing theories (e.g., M/M/S models) provide valuable analytical baselines for capacity planning but often struggle to capture the non-linear, path-dependent complexities and stage-level heterogeneity inherent in sequential queue systems like ISS-SIAP. In contrast, the data-driven GBR approach excels in mapping these complex operational features. The achieved R^2 of 0.866 is highly competitive and surpasses several prior predictive studies in similar domains, such as the work by Ferinan et al. (2026), which reported an R^2 of 0.696. While it remains naturally lower than models applied to highly structured, deterministic environments (e.g., sports analytics with $R^2 = 0.992$), this performance is expected and appropriate for noisy, real-world financial service data, confirming GBR's robust generalization capabilities.

From a managerial perspective, the predictive model offers actionable intelligence for PT FIFGROUP Cikampek Branch. By accurately estimating service durations at the exact moment of customer arrival, branch management can dynamically optimize frontliner allocation across Kasir, CCP, and CRE counters based on predicted workload rather than static schedules. Furthermore, integrating this predictive model directly into the existing ISS-SIAP user interface could provide customers with real-time, transparent waiting time estimates. This transparency directly addresses a primary determinant of customer satisfaction in financial services, transforming the queuing experience from a source of frustration into a managed, predictable process.

Despite its robust performance, this study acknowledges several limitations. First, the dataset spans only 13 working days, which may not fully capture long-term seasonal variations, month-end transaction rushes, or the distinct impact of national holidays on service demand. Second, the study focused exclusively on the GBR algorithm and did not include a comparative baseline analysis against other standard regression models (e.g., Linear Regression, Decision Trees, or Random Forest). This limits the ability to definitively quantify GBR's relative performance advantage in this specific context. Finally, the model relies solely on internal system attributes, omitting dynamic external factors such as real-time lobby congestion, daily staff availability, or specific customer demographic complexities.

Future research should address these limitations by extending the data collection period to encompass broader temporal patterns and seasonal fluctuations. Incorporating external operational variables, such as the number of active tellers per shift, real-time queue density, and holiday indicators, could further refine prediction accuracy and better reflect real-world service dynamics. Additionally, subsequent studies should conduct rigorous benchmark comparisons against other advanced ensemble regression methods, such as XGBoost, CatBoost, and LightGBM, to identify the most computationally efficient and accurate architecture. Ultimately, transitioning this model from a retrospective analytical tool to a real-time, embedded module within the ISS-SIAP system represents a vital next step for practical deployment and continuous operational improvement.

4 CONCLUSION

This study successfully developed a service time prediction model for a sequential queuing system based on the Integrated Self-Service Smart Queue Information System (ISS-SIAP) at PT Federal International Finance (FIF-GROUP) Cikampek Branch using the Gradient Boosting Regressor algorithm. A structured data preprocessing stage, including anomalous data cleaning, resulted in 1,187 valid records for modeling. Through the application of 10-Fold Cross Validation and hyperparameter optimization using grid search, the optimal configuration was obtained, consisting of 30 decision trees, a maximum tree depth of 2, and a learning rate of 0.1. The resulting model demonstrates strong predictive performance, with an RMSE of 3.903 ± 0.286 minutes, MAE of 2.726 ± 0.152 minutes, MSE of 15.308 ± 2.247, and a coefficient of determination (R^2) of 0.866 ± 0.013, indicating that the model explains 86.6% of the variance in customer service duration.

Although the dataset size is relatively modest compared to large-scale machine learning studies, it represents real operational data that capture actual customer service behavior within a sequential queuing system. To mitigate limitations associated with sample size, this study employs 10-Fold Cross Validation to provide a more robust evaluation and reduce the risk of overfitting. The relatively low standard deviation across folds further suggests stable model performance. However, it is important to acknowledge that the limited observation period and dataset size may constrain the models generalizability across different time periods or operational conditions. In addition, this study does not include a direct comparison with baseline regression models (e.g., Linear Regression, Decision Tree, or Random Forest), which may limit the ability to quantitatively demonstrate the relative performance advantage of the proposed model.

For future strategic development and application, further research is recommended to expand the scope of the historical dataset by extending the observation period to better capture seasonal variations in queue patterns. In addition, prediction accuracy may be improved by incorporating relevant external operational variables, such as the number of active frontliners per day, real-time queue density in the waiting area, and national holiday status. These variables directly address the theoretical gap identified in this study, where previous approaches have primarily focused on internal system variables and have not adequately captured dynamic operational factors that influence service time in sequential queuing environments. By integrating these external factors, future models are expected to better reflect real-world service dynamics and enhance predictive performance. Future studies are also encouraged to include benchmark comparisons with baseline models under the same evaluation framework to provide a more comprehensive performance assessment.

Comparative performance evaluation with other advanced ensemble regression algorithms, including XGBoost, CatBoost, and LightGBM, is also recommended to identify the most efficient and accurate model. From a practical perspective, it is recommended that the management of FIFGROUP Cikampek Branch develop this predictive model into a real-time service time estimation module integrated directly into the ISS-SIAP system interface. Such implementation can function as a predictive monitoring tool for counter staff while simultaneously enhancing customer satisfaction through improved transparency and clarity of service waiting time information.

ACKNOWLEDGMENTS

The authors would like to express their deepest gratitude to the management and all operational staff of PT Federal International Finance (FIFGROUP) Cikampek Branch for permission, assistance in providing the historical database of the ISS-SIAP application, and technical support during the observation and data collection process for this research.

DECLARATIONS

Author Contributions:

Amelia Sari Dewi: Conceptualization, Methodology, Software, Writing – Original Draft.
Eva Rahmawati: Formal Analysis, Resources, Supervision.
Sri Diantika: Validation, Supervision, Writing – Review & Editing.

Funding:

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest:

The authors declare no conflict of interest regarding the publication of this paper.

Data Availability:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI Use Statement:

During the preparation of this manuscript, the authors used Gemini solely for language refinement and readability improvement and used ChatGPT as an assistant for programming support in RapidMiner Studio and Python, including workflow suggestions, debugging assistance, and parameter optimization. All scientific content, analysis, interpretations, conclusions, and final model implementations were developed, verified, and validated by the authors. Following the use of these tools, the authors carefully reviewed and edited the manuscript and assume full responsibility for the content of the published article.

Additional Information:

No additional information is available for this paper.

REFERENCES

[1] A. Hariputra, R. S. Wijaya, and M. Hidayat, "Analisis antrean pelayanan jasa pada sektor pembiayaan konsumen," *Jurnal Sistem Informasi*, vol. 9, no. 1, pp. 12–20, 2022.

[2] P. K. Santi and M. Andesta, "Pengaruh waktu tunggu terhadap indeks kepuasan nasabah lembaga keuangan non-bank," *Jurnal Manajemen Pelayanan Jasa*, vol. 4, no. 2, pp. 88–95, 2023.

[3] D. Cendani and A. Wibowo, "Penerapan metode ensemble learning untuk prediksi beban waktu pelayanan," *Jurnal Sains Data Komputasional*, vol. 5, no. 3, pp. 45–53, 2022.

[4] R. Ferinan et al., "Model prediktif indeks kebahagiaan berbasis gradient boosting regressor dengan optimalisasi seleksi fitur dan implementasi web," *Jurnal Komputer Bisnis dan Teknologi Informasi (JKBTI)*, vol. 1, no. 1, pp. 25–34, 2026.

[5] N. Pramesti, M. Fuad, and R. Safrizal, "Analisis sistem antrean dalam meningkatkan pelayanan pada SPPBE Tandem, KSO PT. Pertamina (Persero)," *Jurnal Teknologi Pelayanan Jasa*, vol. 11, no. 2, pp. 102–110, 2023.

[6] A. Fitri, "Analisis perbandingan metode regresi linier, random forest regression dan gradient boosted trees regression untuk prediksi harga rumah," *Jurnal Informatika dan Komputasi*, vol. 14, no. 1, pp. 55–64, 2023.

[7] A. Rusdianto Hidayat et al., "Analitik prediktif sepakbola: model machine learning BRI Liga 1 Indonesia," *Jurnal Kecerdasan Buatan dan Aplikasi*, vol. 3, no. 2, pp. 78–86, 2024.

[8] R. Maulita et al., "Analisis komparatif linear regression, random forest, dan gradient boosting untuk prediksi banjir," *Jurnal Sains Komputasi Terapan*, vol. 8, no. 4, pp. 210–218, 2024.

[9] S. Nur et al., "Perbandingan prediksi depresi mahasiswa dengan linear regression, random forest, dan gradient boosting," *Jurnal Psikologi Komputasional*, vol. 6, no. 1, pp. 40–49, 2024.

[10] M. Saputra et al., "Prediksi dan analisis pola perubahan iklim menggunakan algoritma gradient boosting," *Jurnal Ilmu Lingkungan dan Kebijakan*, vol. 13, no. 2, pp. 134–142, 2025.

[11] R. Ferinan, N. Harani, and S. Pane, "Analisis performa algoritma boosting pada sistem informasi eksekutif perusahaan," *Jurnal Komputasi Intern*, vol. 10, no. 3, pp. 167–175, 2025.

[12] M. Mitzenmacher and R. Shahout, "Queueing, Predictions, and Machine Learning," *Stochastic Systems*, vol. 15, no. 1, pp. 1–22, 2025.

[13] E. Chocron, I. Cohen, and P. Feigin, "Delay prediction for service systems: a machine learning approach," *Operations Research Letters*, vol. 52, no. 3, pp. 210–218, 2024.

[14] S. S. Bhandari and T. G. Deepak, "Machine learning-based waiting time prediction for queueing systems," *Journal of Service Science*, vol. 18, no. 2, pp. 145–158, 2026.

[15] H. Öztürk et al., "Time prediction using gradient boosting methods," *Reliability Engineering & System Safety*, vol. 245, 2025.

[16] Y. Hong et al., "Real-time wait time prediction using machine learning models," *Journal of Data Science Applications*, vol. 9, no. 1, pp. 33–41, 2024.

[17] L. Taton et al., "Queue waiting time prediction using ensemble learning models," *Journal of Big Data Analytics*, vol. 7, no. 2, pp. 101–115, 2025.

[18] H. A. Irawan, "Deteksi Fraud Kartu Kredit Dengan Logistic Regression, Random Forest dan Gradient Boosting," *J. Teknol. Inf.*, vol. 11, no. 2, pp. 92–97, 2025.

[19] W. Wardalisah, "Peramalan Harga Saham menggunakan Metode Light Gradient Boosting Machine dengan Optuna Hyperparameter Optimization (Studi Kasus: Saham PT. Vale Indonesia Tbk.)," Skripsi, Dept. Matematika, Universitas Hasanuddin, 2024.

[20] A. C. Nugraha and I. Irawan, "Komparasi Deteksi Kecurangan pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost)," *J. Comput. Sci.*, vol. 12, no. 1, 2023.