BAREKENG: Journal of Mathematics and Its Applications
September 2025 Volume 19 Issue 3 Page 1625-1636
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss3pp1625-1636

PREDICTION INTERVALS IN MACHINE LEARNING: RESIDUAL BOOTSTRAP AND QUANTILE REGRESSION FOR CASH FLOW ANALYSIS

Wa Ode Rahmalia Safitri [1*], Farit Mochamad Afendi[2], Budi Susetyo[3]

[1,2,3]Statistics and Data Science Department, Faculty of Mathematics and Natural Sciences, IPB University
Jln. Dramaga, Bogor, Jawa Barat, 16680, Indonesia

*Corresponding author's e-mail: * **rahmaliasafitri68@gmail.com***

Article History:

Received: 2nd December 2024
Revised: 11th January 2025
Accepted: 2nd March 2025
Published: 1st July 2025

Keywords:

Bootstrap Method;
CatBoost;
LightGBM;
Prediction Intervals;
Quantile Regression;
XGBoost.

ABSTRACT

Time series forecasting often faces challenges in producing reliable predictions due to inherent uncertainty in dynamic systems. While point predictions are commonly used, they may not adequately capture this uncertainty, especially in financial systems where forecasting accuracy directly impacts decision-making. Prediction intervals offer a solution by providing a range of likely outcomes rather than single-point estimates. This study implements multivariate time series forecasting using gradient boosting algorithms (XGBoost, CatBoost, and LightGBM) to predict cash flow patterns in banking transactions, focusing on constructing reliable prediction intervals. Using transaction data from Bank Rakyat Indonesia (BRI), the research analyzes both office and e-channel transactions with different lag structures based on Granger Causality tests. Model performance was evaluated using RMSLE, MAE, and MAPE metrics, with RMSLE chosen as primary due to its ability to handle skewed distributions. LightGBM achieved best performance for office cash-in transactions (RMSLE: 0.2395), while CatBoost outperformed others for office cash-out (RMSLE: 0.2848), e-channel cash-in (RMSLE: 0.3946), and e-channel cash-out (RMSLE: 0.4221). For prediction intervals, two methods were compared: Residual Bootstrap with 500 samples and Quantile Regression. Residual Bootstrap generally produced coverage probabilities closer to the 80% level (i.e., 10–90% prediction interval), especially for office transactions, while maintaining narrower interval widths. In contrast, Quantile Regression tended to generate wider intervals and often overestimated uncertainty, resulting in overly high coverage in some cases. However, both methods showed clear limitations when applied to e-channel transactions, particularly for cash-in e-channel, where coverage probabilities fell below 50% due to high volatility and irregular transaction patterns. Unlike previous work focused only on point forecasts, this study offers insights into forecast uncertainty by evaluating how well each method quantifies, providing practical guidance for financial institutions aiming to improve risk management through interval-based forecasting.

How to cite this article:

W. O. R. Safitri, F. M. Afendi and B. Susetyo., "PREDICTION INTERVALS IN MACHINE LEARNING: RESIDUAL BOOTSTRAP AND QUANTILE REGRESSION FOR CASH FLOW ANALYSIS," *BAREKENG: J. Math. & App.*, vol. 19, no. 3, pp. 1625-1636, September, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Prediction often comes with a level of uncertainty that needs to be considered. To quantify this uncertainty, prediction intervals are used. A prediction interval is a range between two bounds that indicates where an unknown value is estimated to lie with a certain probability [1]. The upper and lower bounds of this interval provide decision-makers with valuable information about the reliability and potential variability of forecasts.

Statistical methods for generating prediction intervals have evolved significantly over time. In the article "Prediction Intervals in Machine Learning" [2], various approaches to creating prediction intervals are explored, including Analytical Methods, Bootstrap Methods, Bayesian Methods, Monte Carlo Dropout, Deep Quantile Regression, and Heteroskedastic Models. Traditional approaches often relied on parametric assumptions and symmetric intervals based on normal distributions. However, real-world data frequently violates these assumptions, particularly in fields such as finance where asymmetric distributions and heteroscedasticity are common. This has led to the development and application of more robust methods, particularly resampling techniques and quantile regression approaches.

The bootstrap method involves repeatedly sampling the original dataset with replacement to estimate the sampling distribution of statistics of interest [3]. In the context of time series forecasting, this method helps capture both model uncertainty and inherent data variability. Quantile regression, developed by Koenker and Bassett [4], extends beyond traditional mean regression by modeling different quantiles of the response variable, providing a more comprehensive view of the relationship between variables at various points in the conditional distribution.

These methods offer several advantages in financial forecasting. Bootstrap resampling provides non-parametric prediction intervals that don't rely on restrictive distributional assumptions [5], making them particularly suitable for financial data that often exhibits non-normal characteristics. Quantile regression can capture heteroscedasticity and asymmetric responses that are common in financial time series, offering more realistic uncertainty estimates at different probability levels [5].

Previous approaches to cash flow forecasting in banking institutions have often relied on traditional time series methods such as ARIMA or exponential smoothing [6], [7], which typically assumes homoscedastic errors and normal distributions. However, these assumptions are frequently violated in financial data, which can exhibit changing volatility patterns and heavy-tailed distributions. Furthermore, conventional prediction intervals based on standard errors often underestimate uncertainty during periods of market stress or structural changes.

Several researchers have explored similar applications in various contexts. In [8], researchers developed a hybrid ARIMAX-Quantile Regression model for forecasting currency inflow and outflow, while [9] analyzed bootstrap methods for financial risk. However, these studies primarily focused on currency circulation forecasting or different financial applications. In Indonesia's banking sector, most previous research has concentrated on point forecasts rather than prediction intervals.

The novelty of this study lies in its comprehensive application of both residual bootstrap and quantile regression methods to BRI's cash flow data, providing a more complete understanding of forecast uncertainty in an emerging market banking context. By combining these methods with modern machine learning algorithms (XGBoost, LightGBM, and CatBoost), this research offers insights into how advanced forecasting techniques can be effectively applied to Indonesian banking data. Additionally, this study provides a comparative analysis of different prediction interval methods in a real-world banking context, which can serve as a valuable reference for similar applications in other financial institutions, particularly in emerging markets.

This research aims to address challenges in financial forecasting by employing interval prediction methods to analyze cash flow data from Bank Rakyat Indonesia (BRI), a major banking institution in Indonesia. Cash flow forecasting is crucial in banking as it affects liquidity management and financial stability. The primary objective of this research is to construct accurate prediction intervals that can provide clearer insights into the uncertainty of the company's financial conditions. By accurately estimating prediction intervals, this research's findings are expected to provide valuable insights for decision-makers in the financial sector, helping them manage risks and make better-informed decisions.

2. RESEARCH METHODS

2.1 Data Description

The data used in this study is sourced from Kaggle. This dataset focuses on financial information from BRI related to cash flow. It includes changes in cash values and other variables that influence and are processed to achieve cash optimization. The data range covers the period from July 31, 2019, to September 30, 2020. In this study, variables are divided into two main groups: dependent variables and independent variables. Dependent variables include four variables: cash_in_office, cash_out_office, cash_in_echannel, and cash_out_echannel. Meanwhile, independent variables encompass all variables except for those dependent variables. In addition to independent and dependent variables, the study also incorporates dummy variables such as weekend and national holidays.

Table 1. Variables and Descriptions of Cash Flow Parameters in BRI

Variables	Description
Cash_in_office	The total cash inflow at the BRI office cash desk
Cash_out_office	The total cash outflow at the BRI office cash desk
Cash_in_e-channel	Total cash inflow at Automated Teller Machines (ATMs) and Cash Recycle Machines (CRMs)
Cash_out_e-channel	Total cash outflow at ATMs and CRMs
Current_account	Total giro deposits
Deposits	Total deposit deposits
Additional_liabilities	Deposits other than giro, savings, and deposits, including issued securities
Savings	Total savings deposits
IsWeekend	0 for weekends (not operating); 1 for weekdays (operating)
IsHoliday	0 for national holidays (not operating); 1 for working days (operating)

2.2 XGBoost

XGBoost (eXtreme Gradient Boosting), introduced by Chen and Guestrin in 2016 [10]. At its core, XGBoost minimizes a regularized objective function:

$$L(\varphi) = \Sigma\big(l(y_i, \hat{y}_i)\big) + \Sigma\big(\Omega(f_k)\big) \tag{1}$$

where l : the loss function, y_i : the response variable, $\hat{y}_i$: the prediction, and $\Omega(f)$: the regularization term.

XGBoost employs a second-order Taylor expansion of the loss function, allowing for more accurate optimization. It features built-in regularization to prevent overfitting, efficient handling of sparse data, and parallel processing capabilities, making it suitable for a wide range of problems [11].

2.3 LightGBM

LightGBM (Light Gradient Boosting Machine), developed by Microsoft, represents a significant advancement in gradient boosting frameworks. This algorithm enhances the conventional Gradient Boosting Decision Tree approach through two innovative techniques: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). The main idea of GOSS is to focus more on under-trained data by employing a selective approach to data sampling. The technique prioritizes instances with larger gradients while maintaining a balanced distribution through random sampling of instances with smaller gradients. This can be expressed mathematically as

$$\nabla L \approx a\Sigma_{\{i\in A\}} g_i \; + \; b\,\Sigma_{\{i\in B\}} g_i \tag{2}$$

where A: the subset of instances with larger gradients and B: a random sample of the remaining instances.

Meanwhile, EFB aims to bundle exclusive features to reduce the number of features. These complementary techniques are specifically designed to enhance both computational efficiency and scalability while maintaining model accuracy [12].

2.4 CatBoost

CatBoost (Categorical Boosting), created by Yandex, distinguishes itself through its effective handling of categorical variables. It introduces ordered boosting to reduce prediction shift, a common issue in gradient boosting methods [13]. The prediction for each example is calculated using the formula:

$$\hat{y}_i = F_i(x_i) = \sum_{\{j=1\}}^{i} Model_{j(x_i)} \tag{3}$$

where $Model_j$: trained on a subset of the data preceding the i-th example.

This ordered approach helps to mitigate the bias introduced by using modified labels in training. CatBoost also employs a novel categorical feature handling method based on ordering the response variable statistics:

$$\hat{s} = \frac{\sum_{\{j=1\}}^{n}[j < i] * y_j + a * P}{\sum_{\{j=1\}}^{n}[j < i] + a} \tag{4}$$

where $[j < i]$: an indicator function, y_j : the response variable, P : a prior, and a : a weighting parameter.

2.5 Prediction Interval using Residual Bootstrap

The bootstrap method provides a robust solution for prediction interval calculations when the assumption that residuals follow a normal distribution cannot be met. This method is straightforward, assuming only that residuals are uncorrelated and have constant variance. Here is the procedure using naïve forecasting [14].

The error of a one-step forecast is defined as $e_t = y_t - \hat{y}_{t|t-1}$. For naïve forecasting, $\hat{y}_{t|t-1} = y_{t-1}$. Thus, it can be rewritten as:

$$y_t = y_{t-1} + e_t \tag{5}$$

Assuming that future errors will be like past errors, when $t < T$, we can replace e_t by sampling from the set of previously occurred errors (residuals). This way, we can simulate the next observation:

$$y^*_{T+1} = y_T + e^*_{T+1} \tag{6}$$

where e^*_{T+1} : a randomly sampled error from the past and y^*_{T+1} : a possible future value.

The asterisk (*) denotes that this is not the actual value of y_{T+1}, but one potential future value. By adding this simulated observation to the dataset, we can iterate on the process to improve the results [15].

$$y^*_{T+2} = y_T + e^*_{T+2} \tag{7}$$

By repeating these steps, a collection of predictions (potential future values) is formed, representing the expected variance in forecasting.

2.6 Prediction Interval using Quantile Regression

Quantile regression serves as an alternative to classical regression with Ordinary Least Squares (OLS). In classical regression, OLS is used to predict the average response variable across predictor values, whereas quantile regression predicts the median or other quantiles of the response variable. OLS assumes that residuals have constant variance across all values of the independent variables. However, quantile regression offers a superior approach by employing quantile loss, which provides sensible prediction intervals even for residuals with non-constant variance or non-normal distribution [5].

Just as regression minimizes the squared-error loss function to predict a single point estimate, quantile regression minimizes quantile loss when predicting specific quantiles [16]. Quantile loss is defined as:

$$\rho_\tau(u) = \tau \max(u, 0) + (1 - \tau) \max(-u, 0) \tag{8}$$

where τ : the required quantile and has a value in the range (0, 1) and u : the error or difference between the actual observation (y) and the predicted value $(\hat{y})$.

2.7 Evaluation Metrics

Root Mean Squared Logarithmic Error (RMSLE) is an evaluation metric used when the response variables have varying scales, or the data exhibits a skewed distribution [17]. RMSLE measures the difference between the logarithms of predicted and actual values. Mathematically, the formula for RMSLE is:

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\log(\hat{y}_i + 1) - \log(y_i + 1))^2} \tag{9}$$

RMSLE tends to weigh more heavily when the predicted value is less than the actual value, while applying less weight when the predicted value exceeds the actual value.

Mean Absolute Percentage Error (MAPE) is the average absolute difference between predicted and actual values, expressed as a percentage of the actual value [18]. MAPE is used to calculate the percentage error between actual and predicted values, defined by the equation:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{10}$$

Mean Absolute Error (MAE) measures the average absolute difference between predicted and actual values. Unlike other metrics, MAE does not square the errors, thus assigning equal weight to all errors irrespective of their direction [19]. This property makes MAE particularly useful when understanding the magnitude of errors without considering whether they are overestimations or underestimations.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{11}$$

3. RESULTS AND DISCUSSION

In this study, the data used covers the period from July 31, 2019, to September 30, 2020. The data consists of various variables recording daily observations during the study period. Descriptive analysis was conducted to understand the characteristics and basic patterns of this time series data. This information serves as a crucial foundation for further analysis and the development of predictive models in subsequent stages. **Table 2** below presents the descriptive statistics for these variables.

Table 2. Descriptive Statistics of Variables

Variables		Mean (IDR)	Min (IDR)	Max (IDR)	Std (IDR)
$y_{1,t}$	Cash_in_office	89,779,694,500	0	656,925,500,445	92,603,694,470
$y_{2,t}$	Cash_out_office	-62,862,353,978	-344,749,440,186	0	52,247,391,641
$y_{3,t}$	Cash_in_e-channel	703,341,412	0	3,744,400,000	342,185,919
$y_{4,t}$	Cash_out_e-channel	-699,203,294	-2,670,100,000	0	373,507,578
$x_{1,t}$	Current_account	881,283,069,011	382,093,559,531	4,678,342,418,901	386,604,058,907
$x_{2,t}$	Deposits	900,630,117,960	729,321,441,460	3,464,394,920,252	191,594,178,023
$x_{3,t}$	Additional_liabilities	13,765,019,988	10,080,295,596	47,590,591,384	3,401,930,180
$x_{4,t}$	Savings	678,195,351,170	617,056,714,583	2,794,601,471,249	109,363,810,824

Based on the descriptive statistical analysis presented in **Table 2**, several patterns emerge in BRI's transaction and deposit products. Cash_in_office transactions show a mean of IDR 89.77 billion with a standard deviation of IDR 92.60 billion, significantly higher compared to cash_in_e-channel transactions, which only average IDR 703.34 million with a standard deviation of IDR 342.18 million. This substantial difference indicates that customers still predominantly prefer conducting deposit transactions through physical offices rather than digital channels. Meanwhile, cash_out_office transactions recorded an average of IDR 62.86 billion with a standard deviation of IDR 52.24 billion, while cash_out_e-channel transactions averaged IDR 699.20 million with a standard deviation of IDR 373.50 million. The high standard deviation in office transactions indicates greater volatility, implying the need for more stringent liquidity management.

Regarding deposit products, time deposits dominate with an average balance of IDR 900.63 billion, followed by current accounts at IDR 881.28 billion, and savings accounts at IDR 678.19 billion. Although time deposits maintain the highest average balance, they exhibit a lower standard deviation (IDR 191.59 billion) compared to current accounts (IDR 386.60 billion), indicating greater stability. Current accounts show the widest fluctuation range, with minimum values of IDR 382.09 billion and maximum values reaching IDR 4.67 trillion, reflecting the dynamic nature of this product in line with corporate customer transaction behavior. Other liabilities show relatively small values with an average of IDR 13.76 billion and a standard deviation of IDR 3.40 billion, indicating a more stable component within the bank's liability structure.

3.1 Data Preprocessing and Feature Engineering

The preprocessing phase began with an exploratory analysis of transaction patterns in the banking dataset. The analysis revealed an interesting pattern where office transactions consistently showed zero values on Fridays and Saturdays, while significant transactions were recorded on Sundays. After a thorough investigation, it was discovered that these patterns indicated a temporal misalignment in the recording system, where transactions were being recorded a day before their actual execution date. Based on this finding, date shifting (H+1) was implemented as the primary preprocessing step to address the temporal misalignment in the transaction records. This adjustment was necessary after confirming that transaction records on certain dates represented cash flows for the following day. The shift alignment resulted in a more accurate representation of operational patterns, particularly evident in the proper recording of zero values on Saturdays and Sundays, consistent with non-operational branch conditions.

Following the date shift adjustment, a comprehensive analysis of data quality and temporal patterns was conducted. This examination revealed systematic missing values in the dataset, particularly on specific dates such as June 21, 2020, June 25, 2020, and August 28, 2020. These patterns were not random but reflected the operational characteristics of banking transactions, especially during non-business hours and weekends. For handling these missing values in time-dependent variables, a Seasonal Decomposition method **[20]** with a 7-day seasonal period was employed. The 7-day period was selected based on ACF and PACF analyses revealing consistent weekly patterns in banking transactions, with strong correlations. This sophisticated approach began with linear interpolation for initial gap filling, followed by decomposition into trend, seasonal, and residual components. The decomposed components were then carefully reconstructed, maintaining the data's inherent characteristics while providing more accurate representations of missing values. This method proved particularly effective for variables such as cash_in_e-channel, cash_out_e-channel, current_accounts, deposits, additional_liabilities, and savings, as well as office cash transactions during business days.

Further enhancement of the dataset involved sophisticated feature engineering techniques. Holiday flags were introduced to mark major religious and national holidays such as Eid al-Fitr, Christmas, and New Year, while weekend flags were added to differentiate between weekday and weekend patterns. These temporal indicators proved crucial for capturing the distinct behavioral patterns that emerged during holidays and weekends compared to regular business days.

Following these preprocessing steps, the optimal lag structure was determined using Granger Causality testing **[21]**, and the data was appropriately partitioned into training and testing sets. The combination of these preprocessing steps resulted in a more robust and information-rich dataset. The enhanced data quality provided a stronger foundation for subsequent analysis, particularly in improving the accuracy of predictive models and supporting more reliable insights into cash flow patterns. This comprehensive preprocessing approach was essential in addressing the unique characteristics of banking transaction data while preserving the temporal integrity necessary for accurate forecasting.

3.2 Multivariate Time Series Forecasting

This research implements multivariate forecasting using XGBoost, CatBoost, and LightGBM algorithms through the ForecasterAutoregMultiVariate() function from the skforecast library. This function enables the creation of forecasting models that can handle multiple input variables simultaneously, automatically generate lagged features based on specified parameters. To achieve optimal performance, parameter tuning was conducted using the grid_search_forecaster_multivariate() function, which optimizes forecasting parameters using RMSLE, MAE, and MAPE evaluation metrics. The optimization process performs a systematic search through specified parameter combinations, employs cross-validation to prevent overfitting, and utilizes multiple evaluation metrics for comprehensive model assessment. Following the Granger Causality test results, the training data implemented different lag values for each variable: lag 6 for cash_in_office, lag 2 for cash_out_office, lag 4 for cash_in_e-channel, and lag 2 for cash_out_e-channel, ensuring appropriate temporal dependencies are captured in the model.

Table 3. Performance Comparison of Optimal Models for Different Cash Flow Categories

	Best Model	RMSLE	MAE (IDR)	MAPE
Cash_in_office	LightGBM	0.2395	20,686,654,405	0.1947
Cash_out_office	CatBoost	0.2848	14,090,147,801	0.2073
Cash_in_e-channel	CatBoost	0.3946	228,662,201	0.3591
Cash_out_e-channel	CatBoost	0.4221	255,251,432	0.4019

The comparative analysis of model performance reveals distinct patterns across different cash flow categories. LightGBM demonstrated superior performance for cash_in_office transactions, achieving the lowest RMSLE (0.2395) and MAPE (0.1947) among all categories, despite having a relatively high MAE of IDR20.69 billion. CatBoost emerged as the optimal choice for both cash_out_office and cash_in_e-channel predictions, with notably different error magnitudes between these categories. For e-channel transactions, both cash-in and cash-out categories showed higher error metrics (RMSLE > 0.39, MAPE > 0.35) compared to office transactions, suggesting greater prediction complexity in e-channel cash flows. XGBoost, while performing best for cash_out_e-channel predictions, exhibited the highest error metrics across all evaluation criteria, indicating challenges in this category's prediction accuracy.

3.3 Prediction Intervals

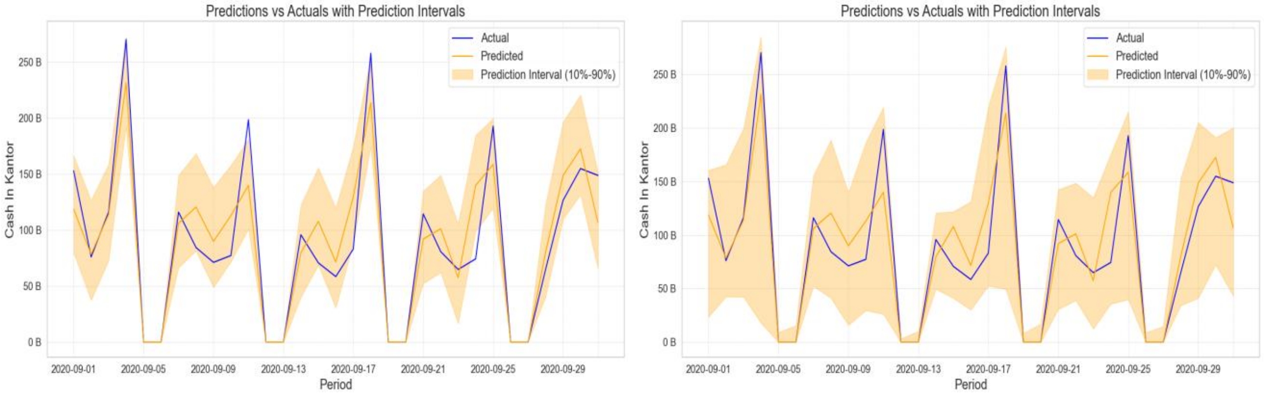
Enhancing the robustness of prediction, prediction intervals were constructed using Residual Bootstrap and Quantile Regression methods. Implementation was performed through the backtesting_forecaster_multivariate () function with 500 bootstrap samples and percentile ranges of 10-90 to accommodate significant variations in data distribution. This function evaluates model performance across multiple time windows while generating prediction intervals through both resampling of model residuals and estimation of conditional quantiles of the target variable. This approach enables the model to provide more comprehensive estimates by considering prediction uncertainties. The performance comparison of both methods for four types of transaction data is presented in Table 4.

Table 4. Performance Comparison of Prediction Interval Methods

Data	Method	Coverage Probability	Interval Range
Cash_in_office	Residual Bootstrap	87.10%	10% - 90%
	Quantile Regression	100.00%	10% - 90%
Cash_out_office	Residual Bootstrap	93.55%	10% - 90%
	Quantile Regression	96.77%	10% - 90%

Data	Method	Coverage Probability	Interval Range
Cash_in_e-channel	Residual Bootstrap	51.61%	10% - 90%
	Quantile Regression	45.16%	10% - 90%
Cash_out_e-channel	Residual Bootstrap	77.42%	10% - 90%
	Quantile Regression	87.10%	10% - 90%

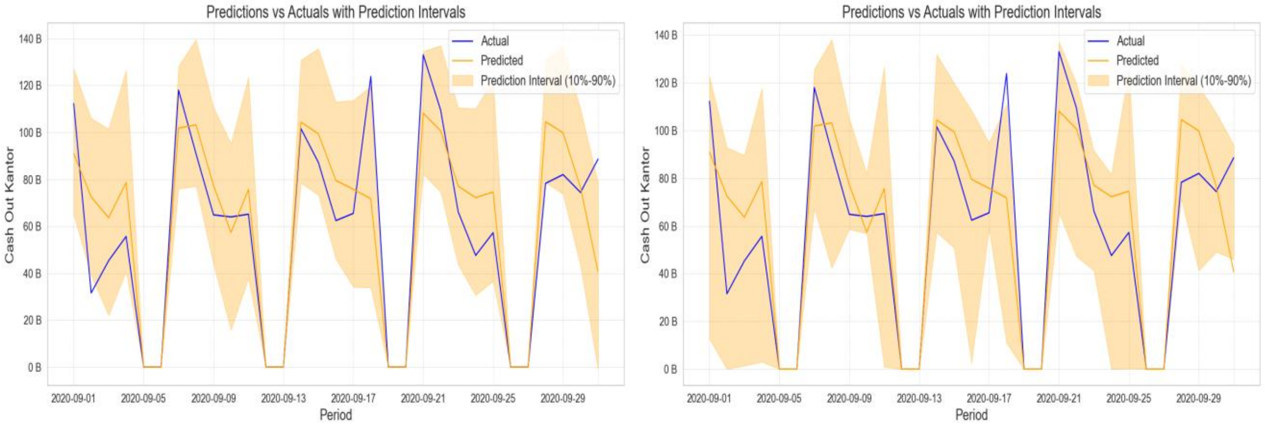
Based on the results obtained in **Table 4**, prediction interval visualizations were created to provide a clearer picture of each method's performance. **Figure 1** shows a visual comparison between the Residual Bootstrap and Quantile Regression methods for one of the analyzed variables, namely Cash In Office.



(a) (b)

Figure 1. Cash In Office Analysis: Actual vs. Predicted Values with Prediction Intervals using (a) Residual Bootstrap and (b) Quantile Regression Approach

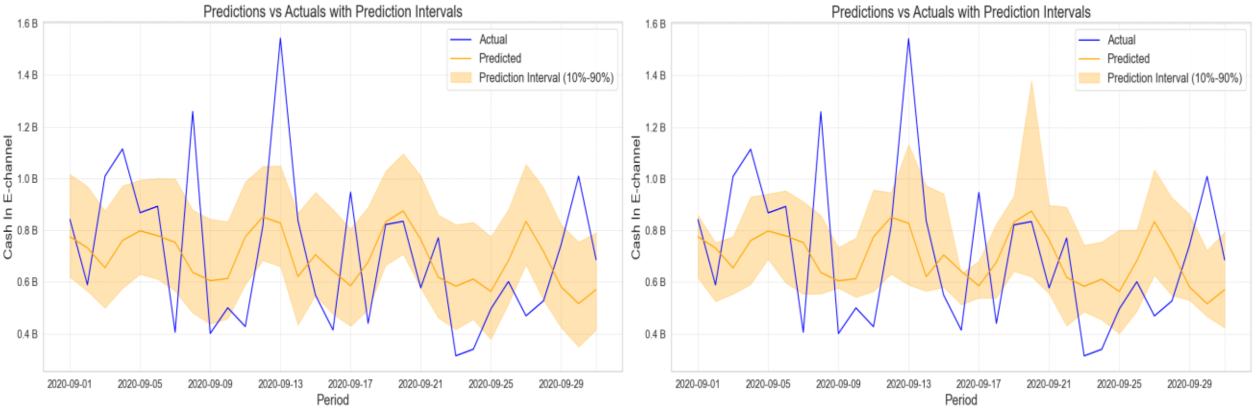
The blue line represents actual values, while the orange line shows prediction values, with the yellow area depicting the prediction interval. The Residual Bootstrap method achieved a coverage probability of 87.10% within the 10–90% prediction interval range. This value is relatively close to the nominal coverage level of 80%, indicating a well-calibrated model with efficient uncertainty estimation. Moreover, the prediction intervals are relatively narrow, providing sharper and more informative guidance while still successfully capturing the majority of actual observations, even during periods of significant fluctuation. In contrast, the Quantile Regression method attained a coverage probability of 100% in the same 10–90% interval range. Although this suggests all actual values were captured within the interval, it also indicates that the model may be overly conservative. The wider intervals produced, especially during periods of normal variation, reduce the precision and informativeness of the predictions. Overall, the Residual Bootstrap method offers a more balanced performance by providing prediction intervals that are both well-calibrated and efficiently narrow, making it more suitable for capturing the dynamic behavior of Cash In Office data without overestimating uncertainty.



(a) (b)

Figure 2. Cash Out Office Analysis: Actual vs. Predicted Values with Prediction Intervals using (a) Residual Bootstrap and (b) Quantile Regression Approach

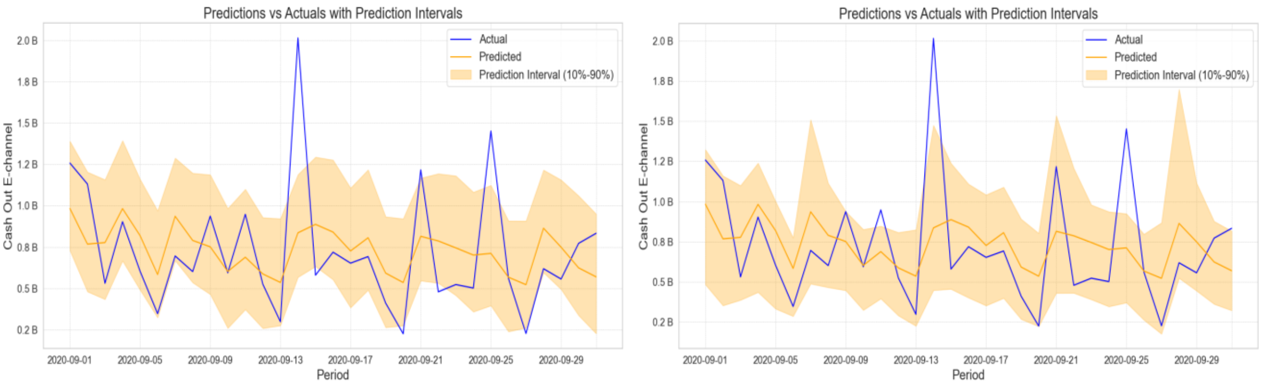
The Residual Bootstrap method achieved a coverage probability of 93.55% within the 10–90% prediction interval range. Although this indicates strong performance and well-calibrated intervals, several actual observations—particularly on September 18 and 21, 2020—fell outside the prediction bounds, likely due to extreme fluctuations. In comparison, the Quantile Regression method yielded a higher coverage probability of 96.77% within the same interval range. This approach produced prediction intervals that more consistently encompassed actual values, especially during periods of sharp variation. The wider intervals generated by Quantile Regression reflect its conservative nature but also its ability to adapt effectively to outlier behavior and sudden spikes in data. Overall, while both methods demonstrate solid performance, the analysis suggests that Residual Bootstrap offers a more balanced trade-off between interval sharpness and reliability. Its intervals are narrower and more efficient, with a coverage probability that remains close to the target level, making it particularly suitable for operational forecasting where clarity and precision are important.



(a) (b)

Figure 3. Cash In E-channel Analysis: Actual vs. Predicted Values with Prediction Intervals using (a) Residual Bootstrap and (b) Quantile Regression Approach

The prediction performance for Cash In E-channel transactions was notably weak using both the Residual Bootstrap and Quantile Regression methods. The Residual Bootstrap approach (a) achieved a coverage probability of only 51.61%, while Quantile Regression (b) performed even lower, at 45.16%, within the 10–90% prediction interval range. These values are far below the nominal 80% target, indicating poor reliability of the prediction intervals. The Cash In E-channel data exhibited extreme volatility, with unpredictable transaction spikes and sharp declines. One significant anomaly occurred around September 13, 2020, where cash inflows surged to approximately 1.6 billion, far exceeding the typical range of 0.6 to 0.8 billion. These abrupt changes were not captured adequately by either method. Overall, the analysis reveals fundamental limitations of traditional predictive models in handling highly volatile and irregular transaction behavior in E-channel systems. This suggests the need for more advanced and adaptive forecasting approaches, possibly involving hybrid models, external drivers, or real-time learning mechanisms to better anticipate unexpected surges and shifts in electronic transaction flows.



(a) (b)

Figure 4. Cash Out E-channel Analysis: Actual vs. Predicted Values with Prediction Intervals using (a) Residual Bootstrap and (b) Quantile Regression Approach

For Cash Out E-channel transactions, the Residual Bootstrap method (a) showed limited effectiveness, achieving a coverage probability of 77.42% within the 10–90% prediction interval. Meanwhile, Quantile Regression (b) offered a slight improvement with a coverage probability of 87.10%. Although Quantile Regression provided higher coverage, both methods struggled to capture the extreme fluctuations characteristic of Cash Out E-channel data. A notable spike occurred in mid-September 2020, where the transaction value surged to approximately 2.0 billion, far exceeding typical daily patterns. These findings highlight the highly dynamic and volatile nature of E-channel financial data, underscoring the need for more sophisticated and adaptive predictive models. Incorporating hybrid methods, external features, or real-time learning could better accommodate irregular transaction behaviors and sudden surges.

This finding aligns with previous research **[22]** by using the same Bank Rakyat Indonesia dataset, where they compared Transfer Function and Artificial Neural Network (ANN) methods. While their study found ANN to be the superior method overall, it still showed higher error rates for e-channel transactions compared to office transactions. This consistent pattern across different studies suggests that e-channel transactions inherently possess greater volatility and are more challenging to predict than office-based transactions. The analysis revealed critical limitations in current predictive modelling approaches for E-channel transactions. The high volatility and irregular patterns suggest the need for more sophisticated, integrated predictive strategies.

4. CONCLUSIONS

This study evaluated machine learning models and prediction interval methods for cash flow forecasting at Bank Rakyat Indonesia (BRI). Among the machine learning models compared (XGBoost, LightGBM, CatBoost), CatBoost and LightGBM demonstrated superior performance across different transaction types. CatBoost excelled in predicting cash_out_office, cash_in_e-channel, and cash_out_e-channel transactions, while LightGBM showed better accuracy for cash_in_office predictions. In terms of prediction interval methods, Residual Bootstrap generally produced coverage probabilities closer to the nominal 80% coverage level (i.e., 10–90% prediction interval), while also maintaining relatively narrower interval widths compared to Quantile Regression. This balance makes its prediction intervals more efficient and better calibrated for most transaction types. However, both methods faced challenges when applied to E-channel transactions, which exhibited extreme volatility. For example, in cash_in_e-channel, low coverage probabilities were primarily due to highly erratic transaction patterns influenced by factors such as early-month salary disbursements, promotional programs, and pandemic-related shifts.

REFERENCES

[1] D. C. Montgomery, C. L. Jennings, dan M. Kulahci, “WILEY SERIES IN PROBABILITY AND STATISTICS,” 2015.

[2] H. Peters, “PREDICTION INTERVALS IN MACHINE LEARNING,” https://medium.com/@heinrichpeters/prediction-intervals-in-machine-learning-a2faa36b320c.

[3] B. Efron, “BOOTSTRAP METHODS: ANOTHER LOOK AT THE JACKKNIFE,” *The Annals of Statistics*, vol. 7, no. 1, hlm. 1–26, Jan 1979, https://doi.org/10.1214/aos/1176344552.

[4] R. Koenker dan G. Bassett, “REGRESSION QUANTILES,” *Econometrica*, vol. 46, no. 1, hlm. 33–50, 1978, https://doi.org/10.2307/1913643.

[5] Rob Hyndman dan G. Athanasopoulos, *FORECASTING: PRINCIPLES AND PRACTICE*, 3rd ed. OTexts, 2021.

[6] A. Susanti, Suhartono, H. J. Setyadi, M. Taruk, Haviluddin, dan P. P. Widagdo, “FORECASTING INFLOW AND OUTFLOW OF MONEY CURRENCY IN EAST JAVA USING A HYBRID EXPONENTIAL SMOOTHING AND CALENDAR VARIATION MODEL,” dalam *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mar 2018. https://doi.org/10.1088/1742-6596/979/1/012096.

[7] Fahmi, “THE ACCURACY FORECASTING OF CASH INFLOW AND CASH OUTFLOW USING DETERMINISTIC, STOCHASTIC AND HYBRIDIZATION MODELS,” *Jurnal Manajemen dan Perbankan (JUMPA)*, vol. 11, no. 2, hlm. 1–11, 2024, https://doi.org/10.55963/jumpa.v11i2.628.

[8] N. A. Salehah, “PENERAPAN MODEL HYBRID ARIMAX-QUANTILE REGRESSION UNTUK PERAMALAN INFLOW DAN OUTFLOW PECAHAN UANG KARTAL DI JAWA TIMUR,” Institut Teknologi Sepuluh Nopember, Surabaya, 2017.

[9] P. Cogneau dan V. Zakamouline, “BOOTSTRAP METHODS FOR FINANCE: REVIEW AND ANALYSIS,” Mei 2010.

[10] T. Chen dan C. Guestrin, "XGBOOST: A SCALABLE TREE BOOSTING SYSTEM," dalam *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, dalam KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, hlm. 785–794. https://doi.org/10.1145/2939672.2939785.

[11] A. Mehdary, A. Chehri, A. Jakimi, dan R. Saadane, "HYPERPARAMETER OPTIMIZATION WITH GENETIC ALGORITHMS AND XGBOOST: A STEP FORWARD IN SMART GRID FRAUD DETECTION," *Sensors*, vol. 24, no. 4, 2024, https://doi.org/10.3390/s24041230.

[12] G. Ke *dkk.*, "LightGBM: A HIGHLY EFFICIENT GRADIENT BOOSTING DECISION TREE," 2017. [Daring]. Tersedia pada: https://github.com/Microsoft/LightGBM.

[13] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, dan A. Gulin, "CatBoost: Unbiased Boosting With Categorical Features," dalam *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, dan R. Garnett, Ed., Curran Associates, Inc., 2018. [Daring]. Tersedia pada: https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf

[14] V. L. Justus, V. B. Rodrigues, dan A. R. dos S. Sousa, "Bootstrap CONFIDENCE INTERVALS: A COMPARATIVE SIMULATION STUDY," Apr 2024, [Daring]. Tersedia pada: http://arxiv.org/abs/2404.12967

[15] E. Paparoditis dan H. L. Shang, "BOOTSTRAP PREDICTION BANDS FOR FUNCTIONAL TIME SERIES," Apr 2020, [Daring]. Tersedia pada: http://arxiv.org/abs/2004.03971.

[16] R. Koenker dan K. F. Hallock, "QUANTILE REGRESSION," *Journal of Economic Perspectives*, vol. 15, no. 4, hlm. 143–156, Des 2001, https://doi.org/10.1257/jep.15.4.143.

[17] R. Y. Assoc, R. Rao Kurada, dan S. Pattem Asst, "AN APPROACH TO IDENTIFY ACCURATE MACHINE LEARNING MODEL TO BUILD HUMAN STRESS LEVEL PREDICTION SYSTEM," 2023. [Daring]. Tersedia pada: https://ssrn.com/abstract=4379061.

[18] A. Botchkarev, "A NEW TYPOLOGY DESIGN OF PERFORMANCE METRICS TO MEASURE ERRORS IN MACHINE LEARNING REGRESSION ALGORITHMS," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 14, hlm. 45–76, 2019, https://doi.org/10.28945/4184.

[19] A. Jadon, A. Patil, dan S. Jadon, "A COMPREHENSIVE SURVEY OF REGRESSION BASED LOSS FUNCTIONS FOR TIME SERIES FORECASTING," Nov 2022, [Daring]. Tersedia pada: http://arxiv.org/abs/2211.02989.

[20] K. Bandara, R. J. Hyndman, dan C. Bergmeir, "MSTL: A SEASONAL-TREND DECOMPOSITION ALGORITHM FOR TIME SERIES WITH MULTIPLE SEASONAL PATTERNS," Jul 2021, [Daring]. Tersedia pada: http://arxiv.org/abs/2107.13462.

[21] C. Amornbunchornvej, E. Zheleva, dan T. Y. Berger-Wolf, "VARIABLE-LAG GRANGER CAUSALITY FOR TIME SERIES ANALYSIS," Des 2019, https://doi.org/10.1109/DSAA.2019.00016.

[22] A. Faricha *dkk.*, "COMPARISON STUDY OF TRANSFER FUNCTION AND ARTIFICIAL NEURAL NETWORK FOR CASH FLOW ANALYSIS AT BANK RAKYAT INDONESIA," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 6, hlm. 6635–6644, Des 2022, https://doi.org/10.11591/ijece.v12i6.pp6635-6644.