Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)
19/2 (2026), 225-239. DOI: http://dx.doi.org/10.21609/jiki.v19i2.1779

Forecasting BPHTB, PBB, and Non-Property Tax Local Tax Revenues Using ARIMA, Feed-Forward Neural Networks, and LSTM: A Case Study of Banjarmasin City Government

Muhamad Fajar, Denny

Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

muhamad.fajar32@ui.ac.id, denny@ui.ac.id

Abstract

Local tax targets in Banjarmasin City, Indonesia are often set through negotiation rather than accountable calculation, which can contribute to recurring gaps between targets and revenue achievement and, consequently, to disruptions in government programs financed by local taxes. This study evaluates forecasting approaches for three major local tax streams, among them: PBB (property tax), BPHTB (land and building acquisition tax), and Non-Property Tax (non-property local taxes, including PBJT). This study using monthly payment records from January 2020 to December 2024 extracted from the city’s core tax administration system. The analysis compares ARIMA-family baselines with neural models (Feed-Forward Neural Network/FFNN and Long Short-Term Memory/LSTM) under a consistent train–test protocol, while additionally testing a multivariate setting for Non-Property Tax using sectoral components (restaurant, hotel, and entertainment tax aggregates) as explanatory inputs and ARIMAX as a statistical benchmark for the multivariate case. Model selection is based on quarterly and annual RMSE, compliance with user-acceptable error thresholds, and paired statistical testing against the closest competing model (runner-up). The results indicate that a multivariate FFNN provides the strongest performance for Non-Property Tax, an LSTM configuration is most suitable for PBB, and an FFNN configuration is preferred for BPHTB. The findings support tax-type-specific model selection as a practical and defensible basis for improving revenue target-setting and short-term fiscal planning.

Keywords: *ARIMA, FFNN Backpropagation, LSTM, Prediction on Local Tax, time-series forecasting, Banjarmasin, BPHTB, PBB, Non-Property Tax*

1. Introduction

Local governments in Indonesia are mandated to strengthen their fiscal autonomy by optimizing Regional Own-Source Revenue (Pendapatan Asli Daerah/PAD), in line with the national regulation on fiscal relations between the central and local governments (Undang-Undang Hubungan Keuangan Antara Pemerintah Pusat dan Pemerintah Daerah/UU HKPD) [1]. Within PAD, local taxes are primary revenue components, directly linked to the public service provision and development programs at the municipal and district level. The ability to forecast and manage local tax revenues in a systematic and accountable manner therefore become critical for the credibility of regional budgets and the continuity of public services.

However, in the reality of the process, local tax revenue targets are still set solely based on negotiation-driven approaches between the local agency of tax and the members of the regional representative council. As the “common” practice negotiations will be based on previous achievements and an assumption of possible percentages of increase in the next year rather than based on more accountable methods of calculation based on historical data [2]. This method will certainly increase uncertainty factor and lower confidence level of the agencies to achieve the target.

In the case of Banjarmasin City, “Satu Data” platform of the city government exposes publicly in 2022 the achievement of the local tax revenue reached approximately IDR 258.15 billion, or 86.63% of the initial target of IDR 298.00 billion [3]. The target–achievement gap persisted in subsequent years, as stated by The Head of the

Tax Data Collection and Determination Division at the Regional Financial, Revenue, and Asset Management Agency (BPKPAD) of Banjarmasin City [4].

BPKPAD at the same time already operates core local tax administration system. This system records taxpayer's registration, tax objects, tax reports, billing and payment of all regulated local tax in Banjarmasin City. However, the data available on the system has not yet been utilized to support decision making, including calculating or forecasting revenue of the local tax. The system is mainly used only for data management, daily operation and services to the taxpayer, not as the basis for quantitative forecast or predictive analysis.

Table 1. Published achievement of 2022 tax revenue at Banjarmasin city's satu data (one data) website.

No	Regional Own-Source Income	Amount (IDR)	Achievement	
			IDR	%
1	**Local Tax**	**298,000,000,000**	**258,154,500,148**	**86.63**
2	Local Retribution	49,774,666,000	33,344,874,546	66.99
3	Results of Management of Separated Regional Assets	30,184,032,799	19,909,453,001	65.96
4	Other Legitimate Regional Own-Source Income	114,699,670,175	87,585,445,472	76.36
	Total	492,658,368,974	398,994,273,167	80.99

The structure of Banjarmasin's local tax revenue consists of three important groups. First, Urban and Rural Land and Building Tax (PBB), hereinafter we will refer to it as PBB. PBB is levied on land and buildings and billed annually with official statement from the government regarding the valuation of the tax object and the tax amount at the current year need to be paid by the building owner. Second, Land and Building Acquisition Tax (BPHTB) is collected based on property transactions and tends to be spontaneously billed at the time property transfer or acquisition happened. These two tax types use official-assessment principle [5], thus the amount of the tax officially published by the local government. Third, composite group of tax other than PBB and BPHTB, unrelated tax to the land or buildings, these include all the rest of the tax such as goods and services tax (food and beverages, road light, hotel services, parking services, entertainment or hospitality services), groundwater tax, non-metallic minerals and rocks tax, swallow's nests tax, and advertisements tax.

Most non-property local taxes in this study are administered under a self-assessment rule[5], except for specific taxes that are assessed differently (e.g., groundwater and advertisement taxes). Nevertheless, these non-property taxes are billed and paid on a monthly cycle. Therefore, for forecasting and target-setting purposes, they are aggregated into a single category, hereinafter referred to as non-property tax, to ensure a consistent monthly time index and to improve the stability of the series by reducing irregular variation across individual tax components.

Table 2. Tax revenue based on billing year of 2022 from core local tax administration system.

No	Tax Type	Tax Payment	Contribution
1	BPHTB	50,896,706,687	25.50%
2	Non-Property Tax	118,784,048,270	59.52%
3	PBB	29,897,134,815	14.98%
	Total	199,577,889,773	100.00%

Machine learning approach or statistical approach commonly used to forecast any kind of time series data [6], [7]. The central research question formulated in this research is: What kind of predictive models could provide practical acceptable accuracy for quarterly and annual forecasts of PBB, BPHTB, and Non-Property Tax revenues, based on available data from the core local tax administration system? Answering this question requires not only evaluating technical performance like MSE, RMSE, and MAE [8], [9], [10], but also interpreting errors in monetary terms and in relation to the scale of Banjarmasin City's local revenues, so that BPKPAD can judge whether the proposed models are operationally acceptable.

To address this question, SLR Kitchenham [11] conducted to systematically select relevant articles and give several relevant models and relevant evaluation metrics to be used in this research. Cross-Industry Standard Process for Data Mining (CRISP-DM) adopted as the overarching process model [12].

2. Literature Review

2.1 Systematic Review on Revenue Prediction in Machine Learning

This research adopts Systematic Literature Review [11], [13] to map the of the art on tax and revenue prediction using machine learning techniques. Follow Kitchenham et al., the SLR is structured as an evidence-based process that systematically identifies, screens, and synthesizes relevant studies using explicit inclusion and exclusion criteria and documented search engine strategies. The SLR is driven by two research questions:

1. Which techniques/algorithms and explanatory variables (features) have been used to predict tax or public revenue?
2. What research designs and evaluation methods have been employed to assess machine-learning or prediction models in tax/financial data-related applications?

The search process initially identified total of 469 publications, consisting of 82 articles from SCOPUS, 331 from ACM Digital Library, and 83 IEEE Xplore. These studies were filtered through multi-stage screening process. In the first stage, titles and abstracts were reviewed to assess their relevance to the research topic (time series problem in financial data or revenue data), which reduced the candidate studies to eight papers. Subsequently, a full text assessment was conducted to evaluate methodological relevance and alignment with the research objectives. After this final screening stage, four studies were retained as the most relevant references for understanding the application of forecasting models in revenue predictions [9], [10], [14], [15].

A SLR findings shows that statistical approaches such as ARIMA and machine learning approaches such as FFNN Backpropagation, GRU, SVM, and LSTM can be used to develop predictive models and solve time series problems. However, from these various sources, in the context of predicting tax revenue, it appears that FFNN Backpropagation is the best machine learning model for predicting such data [9], [10]. Other than tax related research topics, but still at the topic of time series problem, SVM, GRU, LSTM and ARIMA are also used to predict stock prices [14][15]. This study will test the statistical approach of ARIMA, the machine learning approach of FFNN Backpropagation, and Long Short-Term Memory to solve the problem of predicting regional tax revenue. The selection of these three approaches is subjective but still based on considerations that all three approaches have been used to solve time series problems in financial / tax data in the previous research.

For tax streams where additional explanatory signals are available, the study extends the statistical baseline to an ARIMAX specification (i.e., ARIMA with exogenous regressors) and extends the neural models to multivariate inputs. This design allows a fair comparison between univariate and multivariate forecasting settings and clarifies whether performance gains are attributable to model architecture, additional information content, or both.

2.2 Forecast Methodology in Tax Revenue Prediction

2.2.1 Process Framework

The CRISP-DM (Cross-Industry Standard Process for Data Mining) framework provides a widely used, domain-independent structure for organizing data-mining and predictive analytics projects [12]. It comprises six iterative phases:

1. Business Understanding: Translation organizational needs into analytical objectives; defining the need to forecast PBB, BPHTB, and Non-Property Tax revenue at quarterly and annual horizon to support reachable target setting.
2. Data Understanding: Exploring and profiling tax payment data from the core local tax administration system, including completeness, seasonality, and volatility.
3. Data Preparation: Extracting, transforming, and aggregating transactional records into modelling-ready time series.
4. Modelling: Specifying and estimating ARIMA, FFNN Backpropagation, LSTM models, including selection of hyperparameters and model structures
5. Evaluation: Comparing model performance using error metrics, interpreting errors in monetary terms and relative to revenue scales.
6. Deployment: Designing how selected models can be integrated into BPKPAD’s system and workflow (e.g. as a forecasting module that is periodically retrained)

In this study we only focus on phases one to five to get the best model for every tax type mentioned above.

2.2.2 Time-Series Approach: ARIMA

Autoregressive Integrated Moving Average (ARIMA) models are widely used for univariate time-series forecasting. An ARIMA(p, d, q) specification combines three components: (i) an autoregressive (AR) term of order p, which relates the current observation to its past values; (ii) an integration term d, implemented through differencing to remove trend and achieve stationarity; and (iii) a moving-average (MA) term of order q, which models the effect of past forecast errors on the current value.

Model development typically follows the Box–Jenkins procedure, which iterates through identification, parameter estimation, diagnostic checking, and refinement [16]. In this study, candidate ARIMA models for each tax series are generated using a grid search over bounded ranges of *(p, d, q)*, with selection primarily based on the

Akaike Information Criterion (AIC) and secondarily on the Bayesian Information Criterion (BIC) [17], [18], [19], [20], [21]. ARIMA identification is automated through a grid search over non-seasonal orders *(p, d, q)* with the following ranges:

1. p ∈ {0, 1, 2, 3}
2. d ∈ {0, 1}
3. q ∈ {0, 1, 2, 3}

From this full grid, a shortlist of models with the smallest AIC values is constructed for each series ("best_models" table). To assess the adequacy of the fitted ARIMA models, several diagnostic tests are employed. These diagnostics help ensure that ARIMA models are not only good in-sample fitters but also structurally adequate for forecasting. These diagnostics are:

1. The Jarque–Bera test is used to examine whether residuals are approximately normally distributed, based on sample skewness and kurtosis [22].
2. The Durbin–Watson statistic evaluates the presence of first-order autocorrelation in residuals, with values near 2 indicating little serial correlation [23], [24].
3. The Ljung–Box test assesses the joint hypothesis that residual autocorrelations up to a specified lag are zero; non-rejection suggests that the model has adequately captured the serial dependence structure [25], [26].

To support multivariate forecasting when explanatory inputs are available, the ARIMA baseline is extended to ARIMAX, i.e., ARIMA with exogenous regressors. In this study, ARIMAX is used as the statistical benchmark for the Non-Property Tax multivariate setting, where sectoral component aggregates (restaurant, hotel, and entertainment) are included as exogenous variables. This allows a fair comparison between multivariate statistical modelling and multivariate neural approaches under the same information set.

2.2.3 Neural Networks Approach: LSTM and FFNN Backpropagation

In contrast to linear time-series models, neural networks can approximate complex nonlinear relationships between inputs and outputs. Two architectures are particularly relevant: Feed-Forward Neural Networks (FFNN) and Long Short-Term Memory (LSTM) networks.

A FFNN consists of an input layer, one or more hidden layers, and an output layer. Data flow from input to output without recurrent connections. In forecasting settings, sliding windows of past observations (and possibly additional features) serve as inputs, and the next value (or a sequence of future values) serves as the target. During training, backpropagation combined with gradient-based optimization (e.g., stochastic gradient descent or Adam) iteratively updates weights to minimize a loss function such as MSE [27].

FFNNs have been applied to various financial and revenue-prediction problems because of their ability to capture nonlinear patterns and interactions. Noor et al. demonstrate that FFNN models can achieve competitive or superior accuracy in predicting government revenues compared to traditional models, provided that the network architecture and hyperparameters are carefully tuned [9], [10].

LSTM networks are a specialized form of recurrent neural networks designed to handle sequential data with long-range dependencies. LSTM units contain internal "cell states" and gating mechanisms (input, forget, and output gates) that control the flow of information across time steps, allowing the network to retain or discard information as needed [28]. This architecture mitigates the vanishing/exploding gradient problem found in standard RNNs and is thus well-suited for modelling longer historical dependencies in time series.

Studies in stock price and financial time-series forecasting report that LSTM models often outperform both ARIMA and shallow neural networks when series exhibit strong nonlinear and temporal structures [14], [29]. In the context of local tax revenues, LSTM offers the potential to model patterns such as seasonal cycles, although the relatively short historical window may limit these advantages.

Neural models are evaluated in both univariate and multivariate input settings. In the univariate setting, lagged values of the target tax series are used as predictors. In the multivariate setting (applied to Non-Property Tax), the input vector at each time step is expanded to include lagged values of the sectoral components (restaurant, hotel, and entertainment) alongside the target series history. This design tests whether additional explanatory signals improve forecast accuracy and whether the performance gain is consistent across quarterly and annual reporting horizons.

3. Research Methodology

3.1 Research Design

One of the important parts of this research is at the business-understanding stage. At this stage, the focus is on reducing the gap between revenue targets and achievement for three key local tax categories (BPHTB, PBB, and Non-Property Tax) by providing more accountable forecasts to support annual and quarterly target setting. The subsequent stages operationalize this objective into measurable variables, implement data-engineering and modelling workflows, and evaluate model performance using error metrics that are interpretable in rupiah and aligned with the needs of local revenue administrators.

3.2 Computing Environment

The development of the training, hyperparameter tuning, grid search and testing code done using laptop. Meanwhile, the real hyperparameter tuning, grid search and the test run on virtual machine server in the cloud infrastructure. Table 3 lists all the specifications of these two environments.

Table 3. Computing environment specifications.

Component	Laptop	Cloud VM
Operating System	Microsoft Windows 10 Pro	CentOS Linux 7 (Core)
Version / Build	Version 10.0.19045Build 19045	Kernel 3.10.0-1160.119.1.el7.x86_64
OS Architecture	64-bit	x86_64 (32-bit and 64-bit modes supported)
Environment Type	Physical Machine (Bare Metal)	Virtual Machine (KVM Hypervisor)
CPU Model	Intel® Core™ i5-8365U @ 1.60 GHz	AMD EPYC 9754 128-Core Processor
CPU Vendor	Intel	AMD
Physical Cores	4 Cores	2 Cores (allocated to VM)
Logical Processors / Threads	8 Threads	4 vCPU (2 Threads per Core)
Clock Speed	Max 1.896 GHz	~2.25 GHz
CPU Architecture	x64	x86_64
Virtualization	No Virtualization	Full Virtualization (KVM)
Total RAM	16 GB (16,917,159,936 bytes)	15 GB (MemTotal: ~16,007,912 kB)
Available RAM	~3.3 GB Available	~9.4 GB Available
GPU Model	Intel UHD Graphics 620	Cirrus Logic GD 5446 (Virtual VGA Adapter)
GPU Memory	1 GB	Not significant (virtual display adapter)
GPU Type	Integrated Graphics	Virtualized Graphics
Storage Model	Samsung MZVLB512HBJQ-000L7	vda (Virtual Disk)
Storage Interface	SCSI	VirtIO (Virtual Block Device)
Filesystem	NTFS (default Windows filesystem)	ext4

3.3 Study Context and Data

The empirical context is the Government of Banjarmasin City, specifically the local revenue authority responsible for managing regional taxes. The data originates from the city's core tax administration system, which records transactional payment information for all local taxes, including identifiers for taxpayers, tax objects, tax types, and payment timestamps.

For this study, the raw transactional data are aggregated into monthly revenue time series for the period January 2020 to December 2024, yielding 60 monthly observations for each tax category:

1. BPHTB (land and building/property acquisition tax),
2. PBB-P2 (urban and rural property tax; further it will be mentioned as PBB), and
3. Non-Property Tax (a composite of other local taxes such as PBJT and other tax that don't relate to property tax objects).

The choice of the 2020–2024 period is driven by the availability and stability of the core system; earlier years are excluded due to system changes that could introduce structural breaks and reporting inconsistencies.

3.3.1 Non-Property Tax Multivariate Variant

Most experiments in this study use univariate time series, i.e., monthly achieved payments for each tax stream (PBB, BPHTB, and Non-Property Tax). This constraint reflects the limited availability of other primary variables in monthly or quarterly time-series form within the operational system; many potentially relevant attributes are annual assessments, static tax-object attributes, or receivables that do not match the required temporal granularity. However, for Non-Property Tax, this paper additionally evaluates a multivariate dataset derived from the composition of Non-Property Tax itself. Specifically, the monthly aggregate of total Non-Property Tax is treated as the prediction target, while three sectoral component aggregates including restaurant tax, hotel tax, and entertainment tax, these data are used as explanatory inputs. Accordingly, the multivariate dataset is constructed to capture both temporal dynamics and the main revenue-generating components of Non-Property Tax, comprising the following fields: year, month, total_restaurant, total_hotel, total_entertainment, and total_nonpbb. This specification provides sufficient information for forecasting because the component series represent the primary drivers of variation in total_nonpbb, while the year–month index preserves seasonality and trend.

3.4 Data Preparation and Model Development

This subsection will describe the processes of the data preparation including handling missing values and outliers. It also outlines development processes of the models, including the specifying architecture, hyperparameter tuning, functions, and parameters that used.

3.4.1 ETL and Aggregation

Data preparation is implemented through an Extract–Transform–Load (ETL) pipeline. In the extract phase, relevant transactional tables from the core tax database are queried to retrieve payment records for BPHTB, PBB, and Non-Property Tax, including payment date and amount achieved.

In the transform phase, records are filtered to the 2020–2024 period and grouped by year, month, and tax category to compute monthly achievement. Additional cleansing steps handle inconsistent records, such as duplicated entries or technical adjustments, while preserving the financial meaning of the series. For Non-Property Tax, negative entries representing reversal or correction transactions are retained because they are part of the real revenue dynamics; for PBB, a small set of clearly anomalous negative totals (the negative data exists before 2020) is excluded to avoid distorting the series.

Regarding FFNN and LSTM models are sensitive to the scale of input features, each data series is normalized using the MinMaxScaler from scikit-learn, mapping training values into the [0, 1] range. The scaler is fit only on the training portion (2020–2023), and then the same transformation is applied to the test period. After prediction, forecasted values are transformed back to rupiah using the inverse transformation, so errors and interpretations remain in the original monetary scale. Lastly, in the load phase, the aggregated monthly data are exported to comma-separated values (CSV) files and then ingested into the Python-based modelling environment as time-indexed series.

3.4.2 Handling Missing Values and Outliers

The ETL process includes checks for missing months and out-of-range values. For the 2020–2024 horizon, all three-tax series are complete at the monthly level, so no synthetic imputation is needed for missing timestamps. Nevertheless, the scripts implement a forward-fill and backward-fill mechanism as a safeguard for potential deployment or replication in other regions where gaps might appear. Extreme outliers are examined in consultation with domain experts; where an extreme value corresponds to a real policy event (e.g., an intensification program or rate change), it is retained, while clearly erroneous records are excluded at the database level.

3.4.3 Model Development, Train-Test Split and Supervised Learning

Three model families are developed and compared: ARIMA as a classical time-series benchmark, Feedforward Neural Network (FFNN) with backpropagation, and Long Short-Term Memory (LSTM) network. All models are implemented in Python using a common library stack including NumPy, pandas, scikit-learn, statsmodels, and TensorFlow/Keras.

To evaluate models on unseen data that are operationally relevant, the time series for each tax category are partitioned into:

1. Training set: January 2020 – December 2023 (48 months),
2. Test set: January 2024 – December 2024 (12 months).

For the neural network models (FFNN and LSTM), the univariate series are reframed into a supervised learning structure. Each training example uses the most recent 12 months as input features to predict the next 3 months as outputs (a 12-step input window with a 3-step forecast horizon). With 48 training months per series and a stride of one month, this yields 34 overlapping training windows per tax category. The same window specification (12 inputs, 3 outputs) is used for the test period, ensuring that all models are evaluated on a consistent forecasting task aligned with quarterly planning.

3.4.4 FFNN Architecture and Hyperparameter Tuning

The FFNN models are implemented with TensorFlow/Keras as fully connected networks operating on the 12-step input vectors. The baseline architecture consists of:

1. Input layer of size 12 (one feature per past month),
2. Two hidden Dense layers with ReLU or tanh activation,
3. Output layer with 3 units (one for each forecasted month), using linear activation after inverse scaling.

A hyperparameter tuning is performed over a predefined hyperparameter space that reflects practical constraints on model complexity:

1. Hidden layer 1 units: {50, 100, 150, 200, 250, 300, 350, 400, 450, 500}
2. Hidden layer 2 units: {50, 100, 150, 200, 250, 300, 350, 400, 450, 500}

3. Activation functions: {ReLU, tanh}
4. Learning rates: {0.1, 0.01, 0.001}

In total, this yields 600 candidate FFNN configurations per tax series. Each configuration is trained with the Adam optimizer and a maximum of 200 epochs, using an 80/20 split of the training windows into sub-training and validation sets in time chronological order. The first 80% became sub-training data, while the last 20% became validation data. Early stopping monitors validation loss and halts training when no further improvement is observed, reducing overfitting and unnecessary computation.

For each tax category, the configuration with the lowest validation loss is selected as the final FFNN model. Its weights are then retrained on the full 2020–2023 training windows (using the same early-stopping criterion) before evaluation on the 2024 test period.

3.4.5 LSTM Architecture and Hyperparameter Tuning

The LSTM models are also implemented using TensorFlow/Keras and operate on the same 12-step input windows but preserve temporal ordering within each window. Each input window is reshaped into a 12×1 sequence (timesteps × features). The base architecture includes:

1. One LSTM layer with a tuneable number of units,
2. One Dense layer (or second Dense layer) to map the LSTM output to the 3 forecasted months,
3. Linear output activation after inverse scaling.

Hyperparameters are tuned via grid search over:

1. LSTM units (First Layer): {16, 32, 48, 64, 80, 96},
2. Second Dense layer units: {16, 32, 48, 64, 80, 96},
3. Activation functions: {ReLU, tanh},
4. Learning rates: {0.1, 0.01, 0.001}.

The same 80/20 train–validation split and early-stopping strategy as in FFNN are applied, with a maximum of 200 epochs per configuration. The tuning process typically yields 216 candidate configurations per series, from which the best-performing model is selected based on minimum validation loss. As with FFNN, the best LSTM configuration is subsequently retrained on the full set of training windows before being evaluated on the 2024 test period.

3.5 Model Evaluation

3.5.1 Rolling-Origin Evaluation

Because the data are chronological and intended for operational forecasting, model evaluation uses a rolling-origin strategy on the combined (scaled) training and test series. An initial origin is placed after the end of 2023. For each origin, the most recent 12 months are used to generate a 3-month forecast, mimicking quarterly planning. The origin is then shifted forward by three months (one quarter) and the process is repeated until the end of 2024.

For ARIMA, this is implemented by re-estimating or updating the model at each origin and forecasting the next three months iteratively. For FFNN and LSTM, the already trained networks take the 12 most recent scaled observed data as input and output predictions for the next three months. The forecast segments within 2024 are compared against actual achievements for final model comparison.

3.5.2 Per Horizon Evaluation and User Threshold

For each tax category and each model, MAE, MSE, and RMSE are computed at two levels:

1. Quarterly horizon: grouping predictions into four quarters (Q1–Q4) based on the 3-month forecast blocks. To see whether any model in every tax type always have minimum RMSE in every quarter.
2. Annual horizon: RMSE and absolute error of the aggregated 12 forecasted months of 2024. It also be used as main indicator to calculate compliance model to user error thresholds and paired t-test.

RMSE, expressed in rupiah, is used as the primary criterion for ranking models because it penalizes larger errors more heavily while retaining interpretability in the original unit of analysis. MAE complements RMSE by providing a robust measure of average deviation, whereas MSE is useful for analytical purposes and consistency with related studies.

Beyond statistical accuracy, model selection incorporates user-acceptable error thresholds to reflect operational decision needs in local revenue planning. A model is considered practically eligible only if its error level falls within tolerance levels validated with end users, acknowledging that acceptable deviation may differ by tax type and planning horizon. This criterion prevents the selection of models that are statistically competitive but operationally unusable for budgeting and cash-flow control. Thus, all models selected in the per horizon evaluation and pass the error threshold, will be evaluated using paired T-Test and Tie Breaker Principle.

3.5.3 Final Decision: Paired T-Test and Tie-Breaker Principle

Every model that has the lowest RMSE value and passes user error threshold in annual horizon metrics, will be chosen as primary model. Thus, to strengthen the robustness of the final choice, the study applies paired statistical testing between the selected primary model and the closest competing model (runner-up of the lowest RMSE), only if the competing model also passes the user error threshold, else the lowest RMSE will be chosen without t-test assessment.

A one-sided paired test is used to assess whether the candidate achieves systematically lower errors than the runner-up at a conventional risk level (e.g., 5%). When the paired test does not provide sufficient evidence of superiority, the final decision follows a tie-breaker principle prioritizing simpler models (lower computational cost and easier operational deployment), provided that the candidate has already passed the practical eligibility thresholds. All the evaluation process will select final model to conduct the quarterly and annual forecasting horizon, but specific for every tax stream, thus one model will be selected for only one tax stream.

4. Result And Discussion

4.1 Exploratory Analysis of Local Tax Revenue Series

The exploratory analysis focuses on the three main local tax groups administered by BPKPAD Banjarmasin (Non-Property Tax, PBB, and BPHTB) using monthly aggregates for 2020–2024.

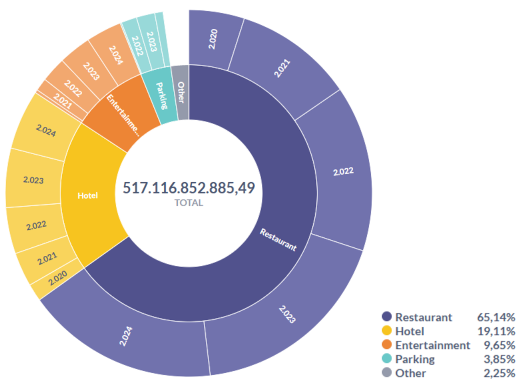

Figure 1. Contribution of tax types in Non-Property Tax at 2020 to 2024.

4.1.1 Revenue composition.

Descriptive statistics on Fig 1. show that Restaurant, Hotel, and Entertainment contribute the largest share of total local tax revenue over the 2020–2024 period in Non-Property Tax.

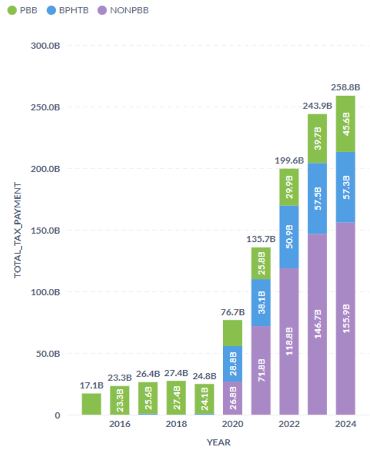

Figure 2. Contribution of tax types from 2015 until 2024.

At Fig. 2 Non-Property Tax is also biggest revenue contributor followed by PBB and then BPHTB. This confirms that an accurate forecast of Non-Property Tax is particularly important for PAD planning.

4.1.2 Seasonality and trend decomposition

Seasonal decomposition (additive) is applied to the monthly versions of each series (Figure 3-5). The results can be summarized as follows:

1. Non-Property Tax shows a clear upward trend with noticeable seasonal and calendar-related spikes (e.g., during holiday seasons or local events), but the residual component remains relatively jagged, especially at weekly granularity.
2. PBB exhibits pronounced seasonal peaks around typical billing and payment periods (e.g., when annual bills fall due), with a smoother trend component and comparatively strong seasonality in the monthly decomposition.
3. BPHTB displays the highest volatility. The trend is upward but with intermittent surges and drops tied to large individual property transactions. Seasonality is present but less regular than in PBB.

Overall, all three series show increasing trends but retain substantial short-term fluctuation, particularly Non-Property Tax and BPHTB. This suggests that purely linear models may struggle to capture all dynamics, and that models with nonlinear capacity (FFNN, LSTM) could exploit complex temporal patterns.

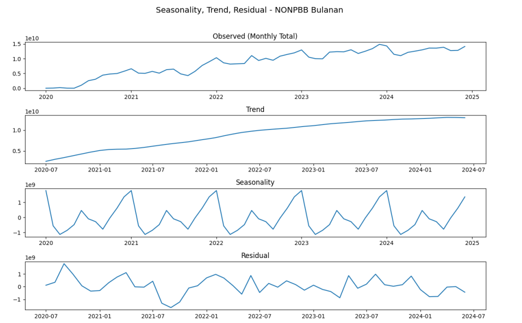

Figure 3. Non-Property Tax monthly sum data seasonal decomposition result.

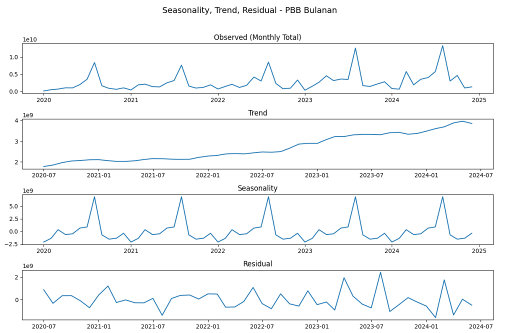

Figure 4. PBB monthly sum data seasonal decomposition result.

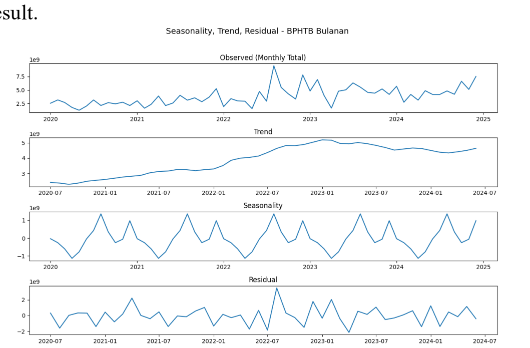

Figure 5. BPHTB monthly sum data seasonal decomposition result.

4.2 GRID Search and Hyperparameter Tuning Result

4.2.1 ARIMA models.

For each tax series, the ARIMA grid search over orders ($p \in \{0,1,2,3\}$), ($d \in \{0,1\}$), and ($q \in \{0,1,2,3\}$) produces a shortlist of candidate models ranked by AIC. The final selected models are:

1. PBB: ARIMA(2,0,3)
2. Non-Property Tax: ARIMA(2,1,3); ARIMAX(0,0,3)
3. BPHTB: ARIMA(2,1,3)

4.2.2 FFNN and LSTM models.

The hyperparameter tuning for FFNN and LSTM explore a wide range of architectures:

1. FFNN: 2 hidden layers, each with 16–450 units; ReLU or tanh activation; learning rate in ({0.1, 0.01, 0.001}).
2. LSTM: one LSTM layer and one Dense layer; 16–96 units; ReLU or tanh activation; same learning-rate grid.

For all three tax types, the models selected as "best" according to validation loss share some patterns:

1. Tend to use moderate learning rates (0.01 or 0.001) rather than 0.1 for stable convergence.
2. Employ non-trivial hidden sizes (e.g., HL1/HL2 between 50–400 units for FFNN; 64–96 units for LSTM) to capture nonlinearity but not so large as to overfit the relatively short history.
3. For Non-Property Tax, LSTM configurations such as LL = 80, DL = 96, LR = 0.1, ReLU and FFNN with HL1 = 450, HL2 = 200, LR = 0.01, ReLU (Fig. 15) emerge as top performers in different evaluation setups.

4.3 Quarterly Forecast Performance (2024)

Quarterly performance is evaluated using a rolling-origin scheme with a 12-month input window and a 3-month forecast horizon, and models are compared using quarterly RMSE. This analysis is intended to characterize short horizon forecasting behaviour in 2024 and to provide supporting evidence for the final annual selection rules presented in the subsequent sections. In other words, this section is used to describe quarter-level accuracy patterns, while the final recommendation is determined under the annual evaluation and eligibility criteria.

As in Table 4, although nonlinear models capture certain intra-year dynamics, empirical results indicate that their superiority is tax and horizon dependent. ARIMA-family baseline remains competitive in several quarters/tax streams. For Non-Property Tax, the quarterly results indicate that neural models are consistently competitive and frequently achieve the lowest or near-lowest RMSE relative to ARIMA-family baselines. In particular, the multivariate FFNN emerges as a strong candidate at the quarter level, which is consistent with the intuition that aggregate Non-Property Tax movements are driven by co-movements among its major components. The best-performing multivariate FFNN configuration highlighted in this study uses HL1 = 250, HL2 = 350, activation = tanh, and learning rate = 0.001, which provides a practical and reproducible specification for subsequent annual evaluation.

Table 4. RMSE of each model in quarterly result

Tax Stream	Row Labels – Best RMSE Label	Min of RMSE
Non-Propoerty Tax	**multivariate-ARIMAX**	**958,730,081.98**
	Q1 (2024)	1,683,065,084.29
	Q2 (2024)	1,418,935,276.53
	Q3 (2024)	958,730,081.98
	Q4 (2024)	1,482,610,622.15
	multivariate-FFNN	**283,567,352.83**
	Q1 (2024) – Best RMSE in Q1	1,175,535,178.37
	Q2 (2024) – Best RMSE in Q2	283,567,352.83
	Q3 (2024)	570,237,038.67
	Q4 (2024)	581,342,461.16
	multivariate-LSTM	**390,727,184.16**
	Q1 (2024)	1,774,951,640.30
	Q2 (2024)	390,727,184.16
	Q3 (2024)	501,825,942.14
	Q4 (2024) – Best RMSE in Q4	475,396,317.07
	univariate-ARIMA	**1,174,049,345.36**
	Q1 (2024)	2,264,897,981.47
	Q2 (2024)	1,174,049,345.36
	Q3 (2024)	2,025,060,029.24
	Q4 (2024)	1,624,390,553.70
	univariate-FFNN	**303,336,200.57**
	Q1 (2024)	1,506,017,318.68
	Q2 (2024)	586,270,445.02
	Q3 (2024)	303,336,200.57
	Q4 (2024)	600,995,766.11
	univariate-LSTM	**94,232,229.61**
	Q1 (2024)	1,723,216,677.92
	Q2 (2024)	346,876,763.37
	Q3 (2024) – Best RMSE in Q3	94,232,229.61
	Q4 (2024)	735,295,740.44
PBB	**pbb-ARIMA**	**1,441,703,192.77**
	Q1 (2024)	2,932,572,835.11
	Q2 (2024)	2,028,725,912.50
	Q3 (2024)	6,585,208,079.71
	Q4 (2024) – Best RMSE in Q4	1,441,703,192.77
	pbb-FFNN	**1,770,571,703.03**
	Q1 (2024) – Best RMSE in Q1	2,144,424,955.71
	Q2 (2024)	1,770,571,703.03
	Q3 (2024) – Best RMSE in Q3	1,885,795,015.13
	Q4 (2024)	2,355,840,413.10
	pbb-LSTM	**1,255,679,422.48**
	Q1 (2024)	2,170,825,756.92
	Q2 (2024) – Best RMSE in Q2	1,255,679,422.48
	Q3 (2024)	4,918,770,298.98
	Q4 (2024)	2,107,628,174.87
BPHTB	**bphtb-ARIMA**	**445,744,226.80**
	Q1 (2024) – Best RSME in Q1	1,326,931,155.26
	Q2 (2024)	798,274,461.71
	Q3 (2024) – Best RSME in Q3	445,744,226.80
	Q4 (2024)	2,255,752,201.03
	bphtb-FFNN	**472,545,833.66**
	Q1 (2024)	1,368,752,500.60
	Q2 (2024) – Best RSME in Q2	472,545,833.66
	Q3 (2024)	697,865,122.92
	Q4 (2024) – Best RSME in Q4	1,636,936,501.18
	bphtb-LSTM	**594,169,116.74**
	Q1 (2024)	1,501,620,760.00
	Q2 (2024)	1,113,211,805.21
	Q3 (2024)	594,169,116.74
	Q4 (2024)	1,644,795,816.30

For PBB and BPHTB, quarter-level RMSE patterns also suggest that non-linear models can match or exceed classical baselines, but quarterly minima should be interpreted as descriptive rather than decisive. For PBB, the LSTM candidate is retained as the primary model for final evaluation due to its overall robustness under the annual criteria; the reported configuration uses LSTM units (LL) = 80, Dense units (DL) = 80, activation = ReLU, and learning rate = 0.001. For BPHTB, a compact FFNN remains a strong candidate under the same principle of reproducible specification, with HL1 = 100, HL2 = 50, activation = ReLU, and learning rate = 0.010. The model configurations with lowest RMSE recapped in Table 5.

Table 5. Quarterly summary (2024): best-RMSE highlight and candidate configurations

Tax stream	Quarterly-level observation (Count of lowest RMSE in 2024)	Candidate model carried forward (reported configuration)
Non-Property Tax	Multivariate FFNN have two lowest RMSE in Q1=1,175,535,178.37 and Q2=283,567,352.83	Multivariate FFNN: HL1=250; HL2=350; AF=tanh; LR=0.001
PBB	Univariate LSTM have lowest RMSE in Q2=1,255,679,422.48	Univariate LSTM: LL=80; DL=80; AF=ReLU; LR=0.001
BPHTB	Univariate FFNN is ranked have lowest RMSE in Q2=472,545,833.66 and Q4=1,636,936,501.18	Univariate FFNN: HL1=100; HL2=50; AF=ReLU; LR=0.010

4.4 Annual Forecast Performance (2024)

Quarterly evaluation reveals that model performance is not stable across all quarters. No single model consistently achieves the lowest quarterly RMSE in every quarter, which means an additional selection step is required to support a defensible single-model recommendation for annual target-setting while still acknowledging quarter-to-quarter variability.

4.4.1 Univariate Non-Property Tax

As in Table 6, with the actual annual total ≈ IDR 155.17 billion,

1. The FFNN configuration with the best RMSE for Non-Property Tax achieves:
   - RMSE ≈ IDR 938.4 million,
   - Predicted annual total ≈ IDR 160.44 billion,
   - Absolute annual error ≈ IDR 5.27 billion, or about 3.4% of Non-Property Tax revenue achievement.
2. The ARIMA(2,1,3) model yields:
   - RMSE ≈ IDR 1.82 billion,
   - Predicted annual total ≈ IDR 156.79 billion,
   - Absolute annual error ≈ IDR 1.62 billion, or about 1.04% of revenue achievement.

These results show a trade-off: FFNN attains a lower RMSE (better month-by-month fit), whereas ARIMA yields a smaller bias in the annual aggregate. From a planning perspective, both are within a ±5% error band, but ARIMA is slightly more conservative in matching the annual total.

Table 6. Non-property tax annual forecast result

Model	Model Code	RMSE	Absolute Error	Best?
ARIMA	Order(2. 1. 3)	1,819,867,270.22	1,626,922,933.56	Best AbsErr
LSTM	LR=0.100 HL1=80 HL2=96	1,043,386,430.22	3,167,153,657.00	
FFNN	LR=0.001 HL1=450 HL2=200	962,875,402.56	3,340,363,271.00	
FFNN	LR=0.100 HL1=50 HL2=100	1,011,883,579.17	4,330,185,223.00	
FFNN	LR=0.010 HL1=50 HL2=400	938,400,706.45	5,270,637,063.00	Best RMSE
LSTM	LR=0.001 HL1=96 HL2=16	1,113,583,720.40	7,280,678,407.00	
LSTM	LR=0.010 HL1=80 HL2=96	1,108,086,805.37	7,309,126,151.00	

4.4.2 Multivariate Inclusion in Non-Property Tax

For Non-Property Tax, the multivariate setting is evaluated by comparing ARIMAX, multivariate FFNN, and multivariate LSTM under the same training–testing split and reporting the annual performance over the 2024 hold-out period. Table 7 summarizes the annual RMSE and absolute error (in currency and percentage) for the multivariate candidates. The results show that the preferred model at the annual horizon is not determined solely by architecture, but by the combined decision rules applied in this study (accuracy, operational tolerances, and robustness checks).

Table 7. Pivot of model RMSE in non-property tax

RMSE (million IDR)	Model Code	Absolute Error on Year (million IDR)	BEST?	Data Variate-Model
728.94	LR=0,001 HL1=250 HL2=350	**299.76**	**Best AbsErr & Best RMSE**	**M-FFNN**
938.40	LR=0,01 HL1=50 HL2=400	5,270.64		U-FFNN
972.23	LR=0,001 HL1=96 HL2=64	1,619.02		M-LSTM
1,043.39	LR=0,100 HL1=80 HL2=96	3,167.15		U-LSTM
1.,410.97	Order(0,0,3)	12,601.67		M-ARIMAX
1,819.87	Order(2,1,3)	1,626.92		U-ARIMA

When contrasted with the univariate Non-Property Tax results, the multivariate experiment indicates that incorporating component aggregates can materially change the ranking of candidate models at the annual horizon. This motivates treating "univariate vs multivariate" as a substantive modelling decision rather than a mere feature-engineering detail, particularly for composite tax streams such as Non-Property Tax.

4.4.3 PBB

Showed in Table 8, with actual annual total ≈ IDR 45.56 billion,

1. The FFNN model with the smallest RMSE for PBB gives:
   - RMSE ≈ IDR 2.15 billion,
   - Predicted annual total ≈ IDR 30.30 billion,
   - Absolute annual error ≈ IDR 15.26 billion, or roughly 33.5% underestimation.
2. Several LSTM configurations for PBB produce slightly higher RMSE values (around 3.37 billion) but much smaller annual biases, with absolute annual errors on the order of IDR 3.64 billion, corresponding to about 7,99% of realized PBB revenue.

This contrast is important: an FFNN tuned purely for minimum RMSE can lead to substantial annual underestimation, while an LSTM with somewhat higher RMSE delivers a far more realistic year-end figure. For PBB, therefore, LSTM is preferable from the standpoint of budget realism, even if FFNN looks better on RMSE alone.

Table 8. PBB annual forecast result

Model	Model Code	RMSE	Absolute Error	Best?
LSTM	LR=0.010 HL1=80 HL2=80	3,370,168,686.18	3,643,068,678.00	Best AbsErr
LSTM	LR=0.001 HL1=16 HL2=16	3,398,287,274.66	3,880,444,166.00	
LSTM	LR=0.100 HL1=80 HL2=48	2,954,894,797.47	8,214,588,794.00	
FFNN	LR=0.100 HL1=50 HL2=50	2,863,174,539.94	14,427,853,386.00	
FFNN	LR=0.010 HL1=350 HL2=350	2,151,345,506.27	15,263,671,098.00	Best RMSE
ARIMA	Order(2. 0. 3)	3,813,108,502.60	20,412,905,790.10	
FFNN	LR=0.001 HL1=450 HL2=500	2,586,766,868.69	20,842,147,290.00	

4.4.4 BPHTB

Showed in Table 9, with the actual 2024 realization approximately IDR 57.30 billion, the best BPHTB configuration (FFNN-based) with the smallest annual bias yields:

- RMSE ≈ IDR 1.21 billion,

- Predicted annual total ≈ IDR 56.83 billion,
- Absolute annual error ≈ IDR 466.38 million, or around 0.8% of realized BPHTB revenue.

Table 9. BPHTB annual forecast result

Model	Model Code	RMSE	Absolute Error	Best?
FFNN	LR=0,010 HL1=100 HL2=50	1.212.746.327,50	466.376.160,00	Best RMSE & Best AbsErr
FFNN	LR=0,100 HL1=50 HL2=150	1.387.020.740,20	1.229.212.192,00	
LSTM	LR=0,100 HL1=80 HL2=48	1.356.985.947,21	2.396.781.088,00	
ARIMA	Order(2, 1, 3)	1.386.100.133,04	3.090.485.587,35	
LSTM	LR=0,001 HL1=64 HL2=16	1.332.894.174,60	3.909.506.080,00	
LSTM	LR=0,010 HL1=16 HL2=96	1.349.656.725,49	4.180.856.352,00	
FFNN	LR=0,001 HL1=100 HL2=150	1.343.071.476,01	4.615.056.928,00	

For BPHTB, all three model families perform reasonably well, but FFNN achieves both low RMSE and very small annual bias, making it the best model at the annual horizon.

4.5 Synthesis

As explained in all the subsection under section 3.5, this sub section will follow explained rules to select the most appropriate model for every tax stream. The maximum annual error tolerated by users is Rp 4.0 billion for each tax type and expressed as a percentage of the 2024 actual value for each tax stream (Table 10).

Table 10. User error threshold

Tax stream	Max annual error (Billion IDR)	Equivalent threshold (% of 2024 actual)	Eligibility rule
BPHTB	4.0	6.98%	Annual AbsErr% ≤ threshold
Non-Property Tax	4.0	2.58%	Annual AbsErr% ≤ threshold
PBB	4.0	8.78%	Annual AbsErr% ≤ threshold

Models that exceed the user tolerance are excluded from recommendation, even if they achieved a strong validation loss ranking during hyperparameter tuning. This pre-screening step prevents the selection of statistically competitive models that are operationally unacceptable for budgeting and cash-flow planning. Screening outcome: BPHTB has 5/7 eligible candidates, Non-Property Tax has 5/10, and PBB has 2/7. This confirms that "best validation loss rank" does not necessarily imply meeting user tolerance.

After pre-screening, the study proceeds as follows: for each tax stream, the two lowest-RMSE models that pass eligibility are carried forward to significance testing (paired t-test), as summarized in Table 11. An exception occurs when the runner-up under the chosen comparison set does not pass pre-screening; in such cases, the remaining eligible best model can be recommended directly based on the combined criteria (lowest RMSE among eligible candidates and smallest annual absolute error).

Table 11. Models that pass user error threshold.

Tax	Candidate	Model code (key params)	RMSE (Billion IDR)	Annual AbsErr (Billion IDR; %)	Rank
BPHTB	Univ. FFNN	LR=0.010; HL1=100; HL2=50; AF=ReLU	1.213	0.466; 0.81%	#1 (lowest RMSE)
BPHTB	Univ. LSTM	LR=0.001; LL=64; DL=16; AF=ReLU	1.333	3.910; 6.82%	#2 (runner-up)
Non-Property Tax	Multiv. FFNN	LR=0.001; HL1=250; HL2=350; AF=tanh	0.729	0.300; 0.19%	#1 (recommended)
PBB	Univ. LSTM	LR=0.010; LL=80; DL=80; AF=ReLU	3.370	3.643; 8.00%	#1 (lowest RMSE)
PBB	Univ. LSTM	LR=0.001; LL=16; DL=16; AF=tanh	3.398	3.880; 8.52%	#2 (runner-up)

Following the selection protocol, the Non-Property Tax multivariate FFNN becomes the recommended model at the pre-screening stage when the runner-up (univariate FFNN LR=0,01 HL1=50, HL2=400 at Table 6) under the selected comparison set does not satisfy the user tolerance criterion; therefore, it does not proceed to paired testing in the same manner as tax streams where two eligible top-RMSE candidates are available.

To confirm whether the selected candidate model (Model A) provides a statistically reliable improvement over the closest competing model (Model B), a paired one-sided t-test was applied to the 12 monthly observations of 2024. The test was conducted on the paired monthly error differences $d_t = a_{t,A} - a_{t,B}$ where $a_{t,\cdot}$ denotes the monthly absolute error for month t. The hypotheses follow the selection rule used: H_0: $\mu_d = 0$ (equal performance) versus H_1: $\mu_d < 0$ (Model A has smaller error). A significance level of $\alpha = 0{,}05$ was used.

Table 12 reports the mean difference, t-statistic, and one-sided p-value for each tax stream where two eligible candidates were available. For BPHTB, the mean difference is negative (about Rp. -168.82 million), indicating that Model A (FFNN) has a smaller average error than Model B (LSTM) in descriptive terms; however, the result is not statistically significant (*p*=0.31 ≥ 0.05). For PBB, the mean difference is positive (about Rp. 289.85 million), implying Model A (LSTM 16/16) is not better than Model B (LSTM 80/80) on average; the result is also not statistically significant (*p*=0.74≥0.05). Consistent with the selection rules, when the paired test is not significant, the final recommendation follows the stated tie-breaker logic (e.g., operational simplicity for BPHTB; lower bias/mean signed error for PBB). For Non-Property Tax, the paired test is not applicable because the runner-up candidate does not pass the error-threshold pre-screening; therefore, the multivariate FFNN is directly recommended under the pre-selection rule.

Table 12. Paired t-test result.

Tax stream	Model A (candidate)	Model B (runner-up)	Mean diff (A–B)	t-stat	p-value (one-sided)	Decision at α=0.05
BPHTB	FFNN (H1=100, H2=50, ReLU, LR=0.01)	LSTM (64, 16, ReLU, LR=0.001)	about Rp. -168.82 Million	-0.52	0.31	Not significant
PBB	LSTM (16, 16, tanh, LR=0.001)	LSTM (80, 80, ReLU, LR=0.01)	about Rp. 289.85 Million	0.65	0.74	Not significant
Non-Property Tax	Multiv. FFNN (250, 350, tanh, LR=0.001)	N/A	N/A	N/A	N/A	Not applicable (runner-up fails threshold)

4.5.1 Non-Property Tax

The multivariate FFNN becomes decisive for Non-Property Tax: it achieves the best annual performance while satisfying the user tolerance threshold, whereas the runner-up candidate with the second-smallest annual RMSE does not pass the pre-screening threshold, so the paired test is not executed for this tax type and the multivariate FFNN is directly recommended as the deployed model.

4.5.2 PBB

For PBB, the paired t-test between the top candidates does not indicate a statistically significant difference, so the final decision is made using an operationally meaningful tie-breaker: preference is given to the LSTM configuration that yields lower bias, operationalized as mean signed error closer to zero (i.e., a smaller systematic tendency to over-forecast or under-forecast). Under this criterion, the selected PBB forecasting model is LSTM (LSTM layer = 16, Dense layer = 16, LR = 0.001, activation = tanh) as the primary model for deployment.

4.5.3 BPHTB

For BPHTB, after applying the user tolerance threshold (Rp. 4 billion converted into a tax-specific percentage bound) and comparing the leading candidates, the paired t-test does not provide evidence of a statistically significant difference between the two best-performing candidates. Consequently, the final recommendation prioritizes a model that is operationally efficient and sufficiently accurate under the acceptance criteria. The chosen model for BPHTB is therefore FFNN Backpropagation (HL1 = 100, HL2 = 50, LR = 0.001, activation = ReLU), selected as the primary deployed model because it satisfies the evaluation criteria while remaining computationally simpler for routine use in an institutional forecasting pipeline.

5. Conclusion

Overall, the results demonstrates that a forecasting model that appears optimal under hyperparameter tuning alone is not automatically acceptable for operational decision-making, because institutional adoption requires conformance to user tolerance constraints and governance checks. By integrating (i) annual RMSE, (ii) an explicit error-threshold rule derived from a Rp. 4 billion maximum tolerated annual deviation (expressed as a tax-specific percentage of realized revenue), and (iii) paired t-tests against runner-up models when eligible, the final recommendations are: Non-Property Tax using multivariate FFNN (HL1 = 250, HL2 = 350, LR = 0.001, tanh); PBB using LSTM (16, 16, LR = 0.001, tanh); and BPHTB using FFNN (100, 50, LR = 0.001, ReLU). Importantly, quarterly evaluation highlights that no single model consistently wins across all quarters, so the paper distinguishes quarterly monitoring insights from the annual model choice required for target-setting and policy planning.

In practical terms, the selected models produce annual errors that are materially smaller than the historical “target vs. achievement” gaps observed in 2024, meaning that the forecasting module can serve as a credible analytical input for budgeting and revenue management. For example, the final comparison shows that the selected models’ annual absolute errors (in percent) are substantially lower than the corresponding 2024 gap between target and achievement for each tax type, supporting their suitability for institutional use where explainability, auditability, and bounded risk are required.

Further research should extend the paper’s evidence-driven pipeline by (i) using the residual diagnostics results more directly to guide ARIMA grid search when diagnostics indicate potential misfit, (ii) testing multivariable forecasting (when explanatory variables are available) to assess whether additional covariates improve stability across tax types other than Non-Property Tax, and (iii) evaluating broader architectures and strategies such as GRU, TCN, hybrid ARIMA-LSTM, ensembles, and hierarchical forecasting to enforce coherence across monthly–quarterly–annual levels, while maintaining the same out-of-sample discipline used in this study. Further research may also explore emerging foundation models as state-of-the-art models by the time this research conducted, such as TimesFM, Prophet Model, Temporal Fusion Transformer (TFT), and N-Beats.

Acknowledgement

The author gratefully acknowledges the Banjarmasin City Revenue, Financial Management, and Asset Agency (BPKPAD Kota Banjarmasin) for providing access to the core local tax administration data used in this study. The author also expresses sincere appreciation to the faculty, the lectures and staff of the Master of Information Technology Program, Faculty of Computer Science, Universitas Indonesia, for their guidance and support throughout the research process. Any errors or omissions in this paper remain the sole responsibility of the author.

References

[1] Pemerintah Indonesia, "Undang-Undang Nomor 1 Tahun 2022 tentang Hubungan Keuangan Antara Pemerintah Pusat dan Pemerintah Daerah," 2022.

[2] Syafrul and L. Farida, "Penetapan target pajak dan retribusi daerah," *Jurnal Kebijakan Publik*, vol. 4, no. 1, pp. 1–118, 2013.

[3] Satu Data Banjarmasin, "Realisasi Pendapatan Asli Daerah (PAD) Kota Banjarmasin Tahun 2022," 2022. [Online]. Available: https://satudata.banjarmasinkota.go.id/data-statistik/7537ac2a-d6c2-46a3-81d9-c1b730cf4079

[4] M. Syahid, "Personal interview (Zoom), Head of Tax Data Collection and Determination Division, BPKPAD Kota Banjarmasin," Sep. 2024.

[5] D. J. P. K. Direktorat Kapasitas dan Pelaksanaan Transfer, *Pedoman Umum Pajak Daerah dan Retribusi Daerah Kabupaten / Kota*. Jakarta Pusat: Direktorat Kapasitas dan Pelaksanaan Transfer, Direktorat Jenderal Perimbangan Keuangan, 2023. Accessed: Oct. 27, 2025. [Online]. Available: https://djpk.kemenkeu.go.id/wp-content/uploads/2023/07/Konsep-Buku-Pedum-PDRD-Kabupaten-atau-Kota-Revisi.pdf

[6] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principle and Practices (3rd Edition)*. Melbourne, Australia: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/

[7] Y. H. Lee and E. Kim, "Deep learning-based delinquent taxpayer prediction: A scientific administrative approach," *KSII Transactions on Internet and Information Systems*, vol. 18, no. 1, pp. 30–45, 2024, doi: 10.3837/tiis.2024.01.003.

[8] A. A. Khalil, Z. Liu, A. Salah, A. Fathalla, and A. Ali, "Predicting Insolvency of Insurance Companies in Egyptian Market Using Bagging and Boosting Ensemble Techniques," *IEEE Access*, vol. 10, pp. 117304–117314, 2022, doi: 10.1109/ACCESS.2022.3210032.

[9] N. Noor, A. Sarlan, and N. Aziz, "GOVERNMENT REVENUE PREDICTION USING FEED FORWARD NEURAL NETWORK," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 6, pp. 2459–2473, 2023, Accessed: Jun. 29, 2026. [Online]. Available: https://www.jatit.org/volumes/Vol101No6/34Vol101No6.pdf

[10] N. Noor, A. Sarlan, and N. Aziz, "Revenue Prediction for Malaysian Federal Government Using Machine Learning Technique," *ACM International Conference Proceeding Series*, pp. 143–148, Feb. 2022, doi: 10.1145/3524304.3524337.

[11] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Inf. Softw. Technol.*, vol. 55, no. 12, pp. 2049–2075, 2013, doi: 10.1016/j.infsof.2013.07.010.

[12] P. Chapman *et al.*, "CRISP-DM 1.0: Step-by-step data mining guide," 2000. [Online]. Available: https://web.archive.org/web/20220401041957/https://www.the-modeling-agency.com/crisp-dm.pdf

[13] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/J.INFSOF.2008.09.009.

[14] A. Yadav, V. Kumar, S. Singh, and A. K. Mishra, "A Survey on Stock Price Prediction using Machine Learning Techniques," *Winter Summit on Smart Computing and Networks, WiSSCoN 2023*, 2023, doi: 10.1109/WISSCON56857.2023.10133865.

[15] C. Anuradha, D. P. S. G. Babu, and A. A. Reddy, "Prediction of Long-Term Stock returns using Long Short Term Memory Neural Network," in *2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC)*, 2023, pp. 1–5. doi: 10.1109/ICACIC59454.2023.10435336.

[16] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed. New York, NY: Springer, 2002. doi: 10.1007/b97391.

[17] R. J. Hyndman and Y. Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *J. Stat. Softw.*, vol. 27, no. 3, pp. 1–22, 2008, doi: 10.18637/jss.v027.i03.

[18] R. Shibata, "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion," *Biometrika*, vol. 63, no. 1, pp. 117–126, 1976, doi: 10.1093/biomet/63.1.117.

[19] M. Rostami, M. Bahaghighat, and M. M. Zanjireh, "Bitcoin daily close price prediction using optimized grid search method," *Acta Universitatis Sapientiae, Informatica*, vol. 13, no. 2, pp. 265–287, 2021, doi: 10.2478/ausi-2021-0012.

[20] L. Zhao, Z. Li, and L. Qu, "Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition," *Heliyon*, 2022, doi: 10.1016/j.heliyon.2022.e12239.

[21] T. Ozaki, "On the Order Determination of ARIMA Models," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 26, no. 3, pp. 290–301, 1977, doi: 10.2307/2346970.

[22] W. H. Greene, *Econometric Analysis*, 8th ed. Harlow: Pearson, 2018.

[23] J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression. II," *Biometrika*, vol. 38, no. 1–2, pp. 159–178, 1951, doi: 10.1093/biomet/38.1-2.159.

[24] J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression. I," *Biometrika*, vol. 37, no. 3–4, pp. 409–428, 1950, doi: 10.1093/biomet/37.3-4.409.

[25] G. M. Ljung and G. E. P. Box, "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978, doi: 10.1093/biomet/65.2.297.

[26] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ: John Wiley & Sons, 2015.

[27] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011.

[28] E. Hossain, *Machine Learning Crash Course for Engineers*. Springer, 2024.

[29] W. Bao, Y. Cao, H. Yang, and H. Che, "Data-driven stock forecasting models based on neural networks: A review," *Information Fusion*, vol. 113, p. 102616, 2025, doi: 10.1016/j.inffus.2024.102616.