ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 08, August 2026

Forecasting Motor Vehicle Transfer Fee Revenue in East Java Province Using a Hybrid Arima-ANN Model

[1] Arianto Nugroho, [2] Ahmad Saikhu

[1] [2] Department of Informatics, Sepuluh Nopember Institute of Technology, Surabaya, East Java, Indonesia
Corresponding Author's Email: [1] 6025202006@student.its.ac.id, [2] saikhu@if.its.ac.id

Abstract*— Motor Vehicle Transfer Fee (BBNKB) is one of the main sources of local revenue (PAD) in East Java Province. Since the enactment of Law No. 1 of 2022, BBNKB applies only to the first transfer of vehicle ownership, making revenue highly dependent on new vehicle sales. This condition creates uncertainty and fluctuations that require accurate forecasting to support fiscal planning. This study evaluates the application of a hybrid ARIMA–Artificial Neural Network (ANN) model for forecasting BBNKB revenue using weekly data from 2010–2024 obtained from the East Surabaya Vehicle Tax Office (Samsat). The research involves stationarity analysis using the Augmented Dickey–Fuller (ADF) test, ARIMA modeling based on the Akaike Information Criterion (AIC), and residual testing using the BDS test to detect nonlinear components. Results show that ARIMA residuals contain nonlinear patterns, so further an ANN model was employed to capture these components. The hybrid model combines ARIMA forecasts with ANN residual corrections. However, performance evaluation reveals slightly higher RMSE, MAE, and MAPE than the single ARIMA model, indicating limited accuracy improvement but providing insight into the linear and nonlinear fiscal data behavior.

Index Terms*— Fiscal Revenue Forecasting, Hybrid ARIMA–Artificial Neural Network (ANN), Motor Vehicle Transfer Fee (BBNKB), Nonlinear Time Series Analysis.

I. INTRODUCTION

Regional Revenue (PAD) is a key component in supporting regional development and the provision of public services. One of the main sources of PAD in East Java Province is the Motor Vehicle Transfer Fee (BBNKB), which is a tax imposed on the transfer of vehicle ownership resulting from a bilateral agreement, a unilateral action, or circumstances arising from transactions such as sale and purchase, exchange, grants, inheritance, or incorporation into a business entity [1]. BBNKB constitutes a significant revenue source for the East Java Provincial Government. Based on Law Number 1 of 2022 concerning Fiscal Relations between the Central Government and Regional Governments (UU HKPD), starting in 2025 BBNKB will only be applied to the first transfer of motor vehicle ownership. This regulatory change makes BBNKB revenue highly dependent on new vehicle sales, potentially leading to substantial fluctuations in regional revenue.

Uncertainty in regional tax revenue—driven by fluctuations in the automotive market, economic conditions, and regulatory changes—demands accurate, data-driven fiscal planning. In this context, forecasting BBNKB revenue becomes an essential instrument for projecting potential income and supporting more adaptive policymaking. The application of statistical methods and artificial intelligence in forecasting is expected to provide a more realistic depiction of future revenue trends.

The Auto Regressive Integrated Moving Average (ARIMA) model is a classical method widely used in time-series forecasting due to its ability to capture linear and seasonal patterns [2]. However, ARIMA has limitations in modeling nonlinear relationships that may emerge in the data However, ARIMA has limitations in modeling nonlinear relationships that may emerge in the data [3]. Conversely, Artificial Neural Networks (ANN), which excel in capturing nonlinear patterns, tend to be less effective when dealing with predominantly linear structures [4].

Several studies have shown that ARIMA and ANN are two commonly used approaches in time-series forecasting. Zhang 2003 [3] introduced the concept of the hybrid ARIMA–ANN model, which combines the ability of ARIMA to capture linear components with the capability of ANN to model nonlinear patterns. Pai and Lin [5], also demonstrated that integrating statistical models with artificial intelligence can reduce forecasting uncertainty and enhance model stability when dealing with complex datasets.

In 2010, Khashei and Bijari demonstrated that hybrid approaches can significantly improve forecasting accuracy compared to single-model methods [6]. Tarmanini, Sarma, Gezegin, and Ozgonenel [7] conducted a comparative study between ARIMA and ANN for Short-Term Load Forecasting. Their results indicated that ANN achieves higher accuracy than ARIMA when dealing with nonlinear data patterns, whereas ARIMA remains superior in capturing linear and seasonal components.

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 08, August 2026

In Indonesia, Prasetyono and Anggraini applied the ARIMA model to forecast national poverty levels, while Sidik and Ansawarman 2022 [9] employed machine-learning methods based on linear regression to predict the number of motor vehicles.

These findings indicate that statistical and machine-learning methods hold substantial potential for analyzing economic and fiscal data. However, research on the application of hybrid ARIMA–ANN models for forecasting regional tax revenues—particularly BBNKB receipts—remains limited. Therefore, this study applies the hybrid model to weekly BBNKB revenue data, which may contain both linear and nonlinear components, in order to obtain more accurate forecasts. The results are expected to contribute to the development of hybrid-based forecasting methods in regional fiscal analysis and serve as a foundation for formulating BBNKB revenue strategies in the era of the UU HKPD implementation.

II. METHOD

This section presents the dataset and the methods used in the study. Fig. 1 provides a representation of the research methodology.

A. Dataset

The dataset used in this study consists of daily BBN revenue records obtained from the East Surabaya Vehicle Tax Office (Samsat). The data cover the period from 1st January 2010 to 31rd December 2024, with a total of 519,023 entries. The dataset contains 12 attributes—TGBAYAR, NOPOL, NAMA, ALAMAT, GOLKEND, KDASAL, KDMERK, KDTYPE, THBUAT, TGMSLAKU, PKBPOKOK, and BBNPOKOK—each of which is described in Table I.

Table. I Dataset Description

No	Atribut	Description
1	TGBAYAR	BBNKB payment receipt date
2	NOPOL	Police number
3	NAMA	Taxpayer name
4	ALAMAT	Taxpayer Address
5	GOLKEND	Vehicle category classification ‘01’ for Sedans and similar vehicle types ’02’ for Jeeps and similar vehicle types ’03’ for station wagons and similar vehicle types ’04’ for Bus and similar vehicle types ’05’ for Truck and similar vehicle types ’06’ for motorcycles and similar vehicle types
6	KDASAL	Country-of-origin code for motor vehicles
7	KDMERK	Motor vehicle brand code
8	KDTYPE	Motor vehicle type code
9	THBUAT	Motor vehicle manufacturing year
10	TGMSLAKU	Motor vehicle tax due date
11	PKBPOKOK	PKB payment amount
12	BBNPOKOK	BBNKB payment amount

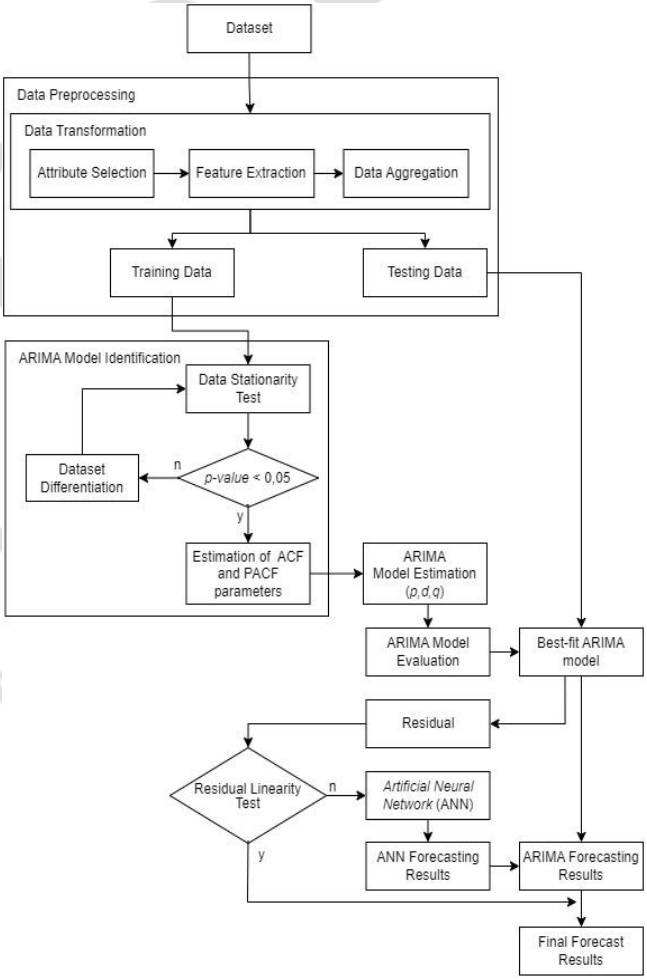


Fig. 1 Model Flowchart

B. ARIMA

ARIMA is a model that combines the Autoregressive (AR) and Moving Average (MA) processes to model time-series data [10]. This method was developed by George E. P. Box and Gwilym M. Jenkins in 1976, and is therefore often referred to as the Box–Jenkins method. In general, ARIMA is a statistical analysis model based on linear time-series behavior [11].

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 08, August 2026

Before applying the ARIMA method, a stationarity test is conducted on the training data using the Augmented Dickey–Fuller (ADF) test. The stationarity criterion is satisfied when the ADF test yields a probability value (p-value) below 0.05. This test determines the value of parameter d to be used in the ARIMA model. The next step is the estimation of the Autocorrelation Function (ACF) for parameter p and the Partial Autocorrelation Function (PACF) for parameter q.

Ketiga parameter (p,d,q) yang dihasilkan dari tahap tersebut digunakan sebagai basis untuk menduga model ARIMA. Pemilihan parameter terbaik ARIMA dilakukan dengan memilih kombinasi yang menghasilkan nilai Akaike Information Criterion (AIC) paling rendah.

C. Residual

From the ARIMA modeling process, residuals are obtained, which represent the difference between the actual values and the predicted values. Let e_t denote the residual at time t from the linear model, then:

$$e_t = y_t + \hat{L}_t \quad (1)$$

where $\hat{L}_t$ is the forecast value at time t obtained from the estimated linear relationship.

D. Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are a type of model capable of approximating various forms of nonlinearity within data. ANN provides a flexible computational framework for modeling a wide range of nonlinear problems. One of the key advantages of ANN over other classes of nonlinear models is its capability as a universal approximator, meaning it can approximate a broad class of functions with high accuracy. This strength comes from its ability to process information in parallel based on the available data. The single hidden layer feedforward network is the most widely used form of ANN in time-series modeling and forecasting. In this study, ANN is employed to capture the nonlinear structure present in the residuals of the ARIMA model.

E. Hybrid ARIMA-ANN

ARIMA and ANN are two of the most popular approaches in time-series forecasting. ARIMA excels at capturing linear patterns, while ANN is effective in learning nonlinear relationships. However, neither model is suitable for all conditions [3]. The basic concept of this hybrid method is that a time-series dataset is assumed to consist of both linear and nonlinear components, as expressed in the following equation (2)

$$y_t = L_t + N_t \quad (2)$$

where L_t represents the linear component and N_t denotes the nonlinear component.

The hybrid method begins by using ARIMA to capture the linear patterns in the data. The residuals produced by the ARIMA model, which are assumed to contain nonlinear structures, are then used as training data for the ANN to learn the nonlinear relationships within the residuals.

The ANN predictions are then combined with the ARIMA results to obtain the final hybrid forecast. The final forecasting output of the ARIMA–ANN hybrid model can be expressed mathematically as follows (3)

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (3)$$

with:

$\hat{L}_t$ = the ARIMA forecast

$\hat{N}_t$ = the ANN prediction of the residuals

The hybrid model is able to capture both linear and nonlinear patterns more accurately and provides better forecasting performance compared to single models [3][5][6].

F. Residual

From the modeling process, residuals are obtained, representing the difference between the actual values and the predicted values. Let e_t denote the residual at time t from the linear model, then:

$$e_t = y_t + \hat{L}_t \quad (4)$$

where L_t is the forecast value at time t obtained from the estimated linear relationship.

A linearity test is conducted on the residuals to determine whether any nonlinear patterns remain. If the residuals are already linear, the modeling process ends with the ARIMA model alone.

G. Artificial Neural Network (ANN)

In this study, ANN is used to learn the nonlinear structures that cannot be captured by the ARIMA model. ANN is applied only when nonlinear patterns are still present in the ARIMA residuals. Accordingly, the linear component is first modeled using ARIMA, while the remaining nonlinear patterns are subsequently modeled using ANN.

H. Final Forecast

At this stage, the final forecasting results are obtained. If the residuals still contain nonlinear patterns and ANN modeling is performed, the final forecast is produced by combining the outputs of the ARIMA model and the ANN. Conversely, if no nonlinear patterns remain in the ARIMA residuals, the ARIMA model alone provides the final forecasting result.

III. EXPERIMENT AND RESULT

A. Experimental Setup

All data processing and modeling procedures were carried out using the Python programming language. The main packages used include pandas and numpy for data processing, statsmodels for ARIMA modeling and diagnostic testing, and scikit-learn for Artificial Neural Network (ANN) modeling and cross-validation. Visualization was performed

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 08, August 2026

using matplotlib.

B. Data Preprocessing

Data preprocessing is an initial stage aimed at transforming raw data into a properly structured dataset that is ready for analysis [12]. Two data preprocessing steps are carried out in this study, namely:

1) Data Transformation

At this stage, the first step is attribute selection, where irrelevant attributes for forecasting are removed from the dataset, including NOPOL, NAMA, ALAMAT, KDASAL, KDMERK, KDTYPE, THBUAT, TGMSLAKU, and PKBPOKOK.

The next step is feature extraction from the TGBAYAR attribute, which contains the date of BBNKB payment. Through this process, two new and more specific attributes are generated, namely TAHUN dan MINGGU. An illustration of this feature extraction process can be seen in Fig. 2.

After the data transformation stage, the number of attributes in the dataset is reduced to five, namely TAHUN, MINGGU, GOLKEND, JUMLAH, dan POTENSI with further descriptions provided in Table II.

2) Dataset Splitting

The transformed dataset is divided into training and testing sets. This division is based on the data period, where data from 2010 to 2023 are used as the training set, and data from 2024 are used as the testing set.

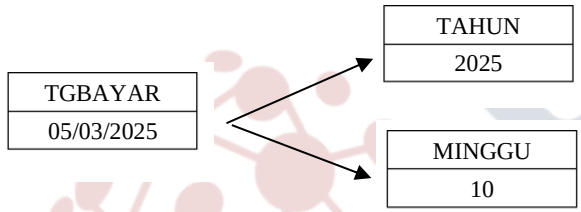


Fig. 2 Illustration of Feature Extraction on the TGBAYAR Attribute

Table. II Database Attributes After Data Transformation

No	Atribut	Deskripsi
1	TAHUN	Year of BBNKB payment
2	MINGGU	Week number within a year
3	GOLKEND	Vehicle category classification '01' for Sedans and similar vehicle types '02' for Jeeps and similar vehicle types '03' for station wagons and similar vehicle types '04' for Bus and similar vehicle types '05' for Truck and similar vehicle types '06' for motorcycles and similar vehicle types
4	JUMLAH	Number of new vehicles
5	POTENSI	Amount of BBNKB revenue from new vehicles

C. ARIMA Modeling

The first step is to test the stationarity of the data using the Augmented Dickey–Fuller (ADF) test. The results show an ADF statistic of –3.0042 and a p-value of 0.0345. Since the p-value 0.0345 < 0.05, the training data can be considered stationary, indicating that ARIMA modeling without differencing (d = 0) is still appropriate.

In addition to the ADF test, the autocorrelation structure of the series is examined through the ACF plot (Fig. 3) and PACF plot (Fig. 4) up to lag 30. These patterns indicate the presence of notable AR and MA components at several lower lags, supporting the use of ARIMA as the baseline model.

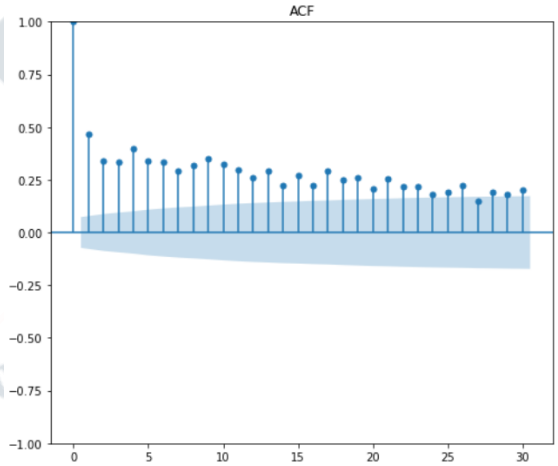


Fig. 2 Grafik ACF

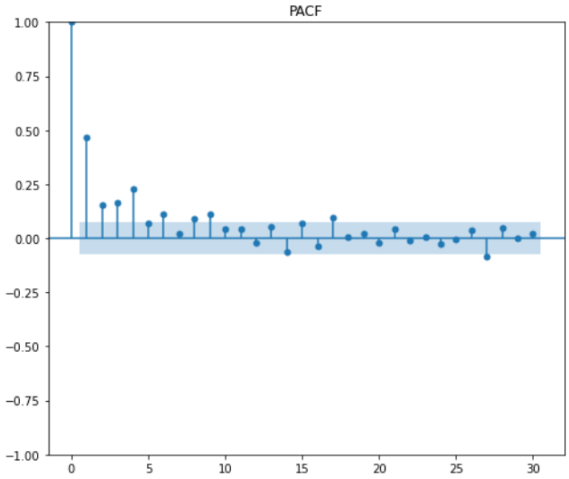


Fig. 3 Grafik PACF

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 08, August 2026

	p	d	q	AIC
0	3	0	3	31606.996191
1	2	0	4	31607.020143
2	1	0	5	31608.107976
3	3	0	4	31608.671853
4	2	0	5	31608.728739
5	1	0	4	31609.967058
6	3	0	5	31610.665622
7	1	0	2	31611.442073
8	2	0	3	31612.354542
9	1	0	3	31613.020253
10	3	0	2	31613.081328
11	2	0	2	31613.153954
12	3	0	1	31613.487643
13	2	0	1	31614.609185
14	1	0	1	31624.647177
15	3	0	0	31666.455913
16	2	0	0	31684.696480
17	0	0	5	31685.655321
18	0	0	4	31691.839328
19	1	0	0	31699.910173
20	0	0	3	31723.897755
21	0	0	2	31726.660948
22	0	0	1	31752.401370
23	0	0	0	35891.851869

Fig. 4 AIC Results

Interpretation of the ACF and PACF plots indicates that suitable values for $p$fall within the range of 0 to 4, while values for $q$lie within the range of 0 to 6 as potential ARIMA model candidates. Based on these results, the parameter combinations for the ARIMA model.

- p ϵ {0,1,2,3}
- d = 0
- p ϵ {0,1,2,3}

For each combination of (p, d, q), the model is estimated and the Akaike Information Criterion (AIC) value is computed. The AIC results for all evaluated models are presented in Table III. The model with the lowest AIC value is selected as the best candidate. The best-performing model obtained is ARIMA (3, 0, 3) with an AIC value of 31,606.996.

After obtaining the ARIMA (3,0,3) model, the residuals are calculated as the difference between the actual data and the fitted values. These residuals are then analyzed using several diagnostic functions, including tests for stationarity, autocorrelation, heteroskedasticity, and nonlinearity. A summary of the residual analysis results is as follows:

- ADF

Based on the Augmented Dickey–Fuller (ADF) test results, the residuals show an ADF statistic of –26.6960 with a p-value of 0.0000. This value is far below the significance threshold, indicating that the residuals are highly stationary.

- Ljung-Box

The Ljung–Box test at the selected lag yields a Q statistic of approximately 22.94 with a p-value of around 0.5235. Since the p-value is greater than 0.05, there is no significant evidence of remaining autocorrelation. Thus, in terms of linear structure, the ARIMA residuals can be considered to approximate white noise.

- ARCH-LM

The test result shows a test statistic of approximately 25.07 with a p-value of around 0.0145. Since the p-value is below 0.05, there is strong evidence of heteroskedasticity (ARCH effects) in the residuals, indicating that the residual variance is not constant.

- BDS

The bootstrap-based BDS test results for embedding dimensions $m = 2$ to $m = 6$ show consistently high BDS statistics, ranging from 7.0510 to 7.5710. For all values of m, the obtained p-values are approximately 0.004975, which is below 0.01. This leads to the rejection of the null hypothesis that the residuals are independent and identically distributed (i.i.d.) at the 1% significance level. Thus, even though the linear autocorrelation components have been addressed by the ARIMA model, these results indicate that the residuals still contain nonlinear structures that cannot be explained by the pure ARIMA model.

m	BDS_stat	p_value
2	7.051032	0.004975
3	7.487528	0.004975
4	7.397230	0.004975
5	7.526234	0.004975
6	7.571009	0.004975

Fig. 5 BDS Bootstrap

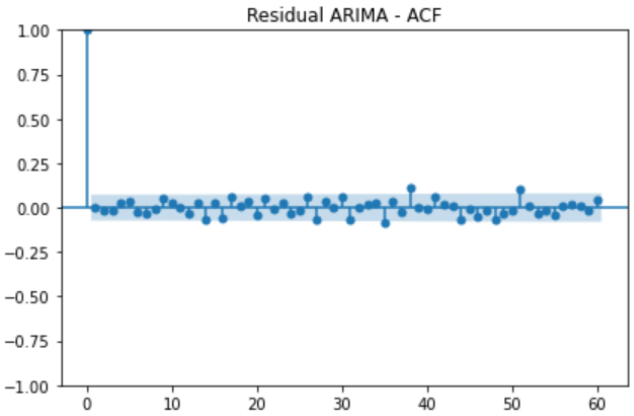


Fig. 6 Residual ARIMA - ACF

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 08, August 2026

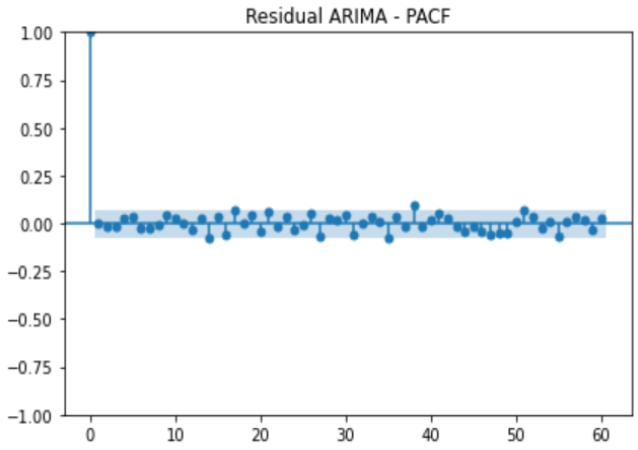


Fig. 7 Residual ARIMA - PACF

These test results further reinforce the conclusion that the residual pattern is not entirely random and that the residual distribution deviates from normality. Therefore, the hybrid ARIMA–ANN approach becomes relevant for capturing the remaining nonlinear components.

The three parameters (p, d, q) obtained from these steps are used as the basis for estimating the ARIMA model. The optimal ARIMA parameters are selected by choosing the combination that yields the lowest Akaike Information Criterion (AIC) value.

D. Hybrid ARIMA-ANN Modeling

As the initial step of the hybrid approach, the residuals from the ARIMA (3,0,3) model are used as the target for the ANN model. The residual lags employed are [1,2,3,4,5,6,7,14,21,28,52]. Lags 1–7 represent short-term memory (up to approximately one week), while lags 14, 21, 28, and 52 are intended to capture weekly cyclical and seasonal patterns (such as monthly and annual cycles) that may not be captured by the ARIMA model. The lagged features are normalized using StandardScaler to accelerate and stabilize the neural network training process.

The ANN model used in this study is a Multilayer Perceptron (MLPRegressor) implemented in scikit-learn. The basic parameters employed include:

- max_iter = 5000
- random_state = 42
- early_stopping = True
- n_iter_no_change = 30

To obtain the most suitable ANN configuration within the ARIMA–ANN hybrid framework, a Grid Search is performed using TimeSeriesSplit (cv = 5) as the cross-validation method. This approach ensures that the temporal order is preserved, allowing the ANN performance evaluation to remain consistent with the characteristics of time-series data. The parameters explored include various combinations of hidden layers, activation functions, alpha values, and learning rates. The recorded Grid Search results are:

- Best RMSE (CV) ≈ 1,01 × 10^9
- Best params:
  - hidden_layer_sizes = (128, 64)
  - activation = 'relu'
  - alpha = 0,001
  - learning_rate_init = 0,0005

The optimal configuration is then retrained using the full set of scaled residuals. After training is completed, the predicted residuals for the training data are transformed back to their original scale using the inverse transform of the target scaler.

For evaluation purposes, this forecasting scheme is directly applied to the 2024 test data. The model's performance is assessed by computing MAE, RMSE, MAPE, and NRMSE on the test set, as summarized in Table III

Table. III Hasil Pengujian Data

s	MAE	RMSE	MAPE(%)	NRMSE
ARIMA	9,899271×10^8	1,271799×10^9	25,3343	0,2257
Hybrid	1,056292×10^9	1,390152×10^9	27,3948	0,2468

As shown in Table III, the hybrid model's performance does not surpass that of the pure ARIMA model. All error metrics for the hybrid model—including MAE, RMSE, MAPE, and NRMSE—are higher than those of ARIMA. The ARIMA (3,0,3) model delivers better results, with a MAPE of approximately 25.33% and an NRMSE of about 0.226. In contrast, the hybrid model produces a MAPE of around 27.39% and an NRMSE of about 0.247, indicating that the addition of the ANN component does not improve forecasting accuracy for this dataset.

These findings indicate that, for the vehicle category examined, adding an ANN component to model the ARIMA residuals does not improve forecasting accuracy. In fact, the performance tends to decrease slightly compared to the pure ARIMA model. This suggests that the nonlinear components captured by the ANN may not be sufficiently significant to yield performance gains for this dataset.

Overall, the experimental results indicate that ARIMA (3,0,3), as a standalone linear model, already provides a reasonably strong performance in forecasting weekly BBNKB revenue potential for the analyzed vehicle category. Meanwhile, the ARIMA–ANN hybrid approach, with the configuration and data used in this notebook, does not demonstrate improved accuracy. These findings highlight an important point for discussion: hybrid models do not automatically outperform single models; achieving their benefits may require adjustments to the ANN architecture, feature selection, or even more specific data segmentation to fully leverage the strengths of the hybrid approach.

IV. CONCLUSION

The ARIMA (3,0,3) model provides reasonably good forecasting performance for the weekly BBNKB revenue

potential data, with residuals that are linearly close to white noise. Although the BDS test indicates the presence of nonlinear patterns, these components appear to be non-dominant, allowing ARIMA to capture most of the data dynamics.

The hybrid ARIMA–ANN approach in this experiment does not improve accuracy; all error metrics show a slight deterioration compared to the standalone ARIMA model. This suggests that the remaining nonlinear components are either too small or insufficiently stable for effective modeling by the ANN. Therefore, for this dataset, ARIMA proves to be more effective than the hybrid approach employed.

REFERENCES

[1] K. Katjong, T. Yanuaria, H. M. Polontoh, and D. Mulyadi, "Penerapan Ipteks Undang-Undang Nomor 1 Tahun 2022 Tentang Hubungan Keuangan Antara Pemerintah Pusat dan Pemerintahan Daerah," *Bakti Hayati J. Pengabdi. Indones.*, vol. 3, no. 1, pp. 8–14, 2024, doi: 10.31957/bhjpi.v3i1.3855.

[2] I. M. Wirawan, T. Widiyaningtyas, and M. M. Hasan, "Short Term Prediction on Bitcoin Price Using ARIMA Method," *2019 Int. Semin. Appl. Technol. Inf. Commun.*, pp. 260–265, 2019, doi: 10.1109/ISEMANTIC.2019.8884257.

[3] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003, doi: 10.1016/S0925-2312(01)00702-0.

[4] G. P. Zhang, B. E. Patuwo, and M. Y. Hu, "A simulation study of artificial neural networks for nonlinear time-series forecasting," *Comput. Oper. Res.*, vol. 28, no. 4, pp. 381–396, 2001, doi: 10.1016/S0305-0548(99)00123-9.

[5] P.-F. Pai and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005, doi: https://doi.org/10.1016/j.omega.2004.07.024.

[6] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 479–489, 2010, doi: 10.1016/j.eswa.2009.05.044.

[7] C. Tarmanini, N. Sarma, C. Gezegin, and O. Ozgonenel, "Short term load forecasting based on ARIMA and ANN approaches," *Energy Reports*, vol. 9, pp. 550–557, 2023, doi: 10.1016/j.egyr.2023.01.060.

[8] R. I. Prasetyono and D. Anggraini, "Analisis Peramalan Tingkat Kemiskinan Di Indonesia Dengan Model Arima," *J. Ilm. Inform. Komput.*, vol. 26, no. 2, pp. 95–110, 2021, doi: 10.35760/ik.2021.v26i2.3699.

[9] A. D. Sidik and A. Ansawarman, "Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning," *Formosa J. Multidiscip. Res.*, vol. 1, no. 3, pp. 559–568, 2022, doi: 10.55927/fjmr.v1i3.745.

[10] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 1394–1401, 2018, doi: 10.1109/ICMLA.2018.00227.

[11] G. P. Ganapathy *et al.*, "Rainfall Forecasting Using Machine Learning Algorithms for Localized Events," *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 6333–6350, 2022, doi: 10.32604/cmc.2022.023254.

[12] R. S. Wahyuwidayat Ahmad; Hidayati, Shintami Chusnul, "Analysis of Taxpayer Behavior to Predict Motor Vehicle Tax Payments Using the Weighted Majority Voting Ensemble Approach," *IPTEK J. Technol. Sci.*, no. Vol 35, No 2 (2024): In Press, pp. 122–134, 2024.