Application of the Artificial Neural Network Method to Predict the Realization of the Duty Tax on the name of Motor Vehicles in Lampung Province

by dian kurniasari

Application of the Artificial Neural Network Method to Predict the Realization of The Duty Tax on The Name of Motor Vehicles in Lampung Province

[1*]Dian Kurniasari, [2]Putri Salsabila Ramadhani, [3]Wamiliana, [4]Warsono

[1,2,3,4] Departement of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung, Indonesia

Email: [1]dian.kurniasari@fmipa.unila.ac.id, [2]putrisalsabila21@gmail.com, [3]wamiliana.1963@fmipa.unila.ac.id, [4]warsono.1963@fmipa.unila.ac.id

Article Info

Article history:

Received May 12th, 2019

Revised Jun 20th, 2020

Accepted Jul 26th, 2020

Keyword:

BBNKB Tax

Machine Learning

Artificial Neural Network

Predictions

ABSTRACT

Regional taxes, specifically the Motor Vehicle Name Return Tax (BBNKB), provide the primary source of revenue for regions from the several forms of taxes. The BBNKB tax is crucial in funding government and regional development due to its significant annual growth, encompassing four-wheeled and two-wheeled vehicles. Furthermore, the BBNKB tax catalyzes regional economic expansion and significantly contributes to the government's income. Hence, predicting and forecasting the BBNKB Tax in Lampung Province is necessary to monitor future tax rate fluctuations. That will enable the government to devise innovative tax payment systems and establish tax revenue targets. This study utilizes the Artificial Neural Network (ANN) methodology, using many approaches for distributing training and testing data to forecast. In addition, we utilize hyper-tuning on several factors to obtain the most favourable configurations. The ideal model achieved has a training data allocation of 80% and a testing data allocation of 20%. It was trained for 50 epochs and used a batch size of 16. The model has exceptional predictability, attaining an accuracy rating of 96.51%. Additionally, it showcases a low root mean square error (RMSE) of 0.246 and a minimal mean absolute percentage error (MAPE) score of 3.48%. Therefore, it is appropriate to predict the next two-year term. As a result, the forecast for the amount of tax collected from motor vehicle name returns in Lampung has fluctuated.

Corresponding Author:

Dian Kurniasari

Departement of Mathematics, Faculty of Mathematics and Natural Sciences

University of Lampung

Jl. Sumantri Brojonegoro No 1, Bandar Lampung, Indonesia.

Email: dian.kurniasari @fmipa.unila.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v2i2.xxxx

1. INTRODUCTION

The tax sector is a promising source of domestic revenue with significant potential for further development [1]. Local taxes play a crucial role in local original taxes (PAD) and should be raised due to their substantial contribution to PAD. The Motor Vehicle Name Return Tax (BBNKB) is the primary municipal tax that generates the region's highest income. This phenomenon occurs due to an annual surge in individuals utilizing motorized vehicles, including four- and two-wheeled vehicles. BBNKB tax is levied on the transfer of ownership rights of motor vehicles as a result of a mutual agreement, unilateral actions, or situations arising from purchases, sales, exchanges, grants, inheritances, or involvement in a corporate entity.

The BBNKB tax plays a crucial role in financing the government and promoting regional development, as it stimulates economic growth in the region and serves as the primary source of state revenue. Hence, it is imperative to scrutinize and forecast the BBNKB tax in Lampung Province to comprehend forthcoming patterns of increase or fall and devise advancements in the tax payment system while establishing aims for tax revenue.

Predictions, as defined by Gosavi [2], refer to the techniques employed to identify future patterns based on available facts and limitations. An extensively employed approach for predicting and resolving intricate problems involving several factors is the Artificial Neural Network (ANN) [3]. An ANN is a computational technique designed based on human neurons' structure to model data [4]. The artificial neural network (ANN) comprises three distinct layers: the input, hidden, and output layers [5].

The application of Artificial Neural Networks has been extensive across numerous domains [6–9], encompassing optimization, prediction, modelling, clustering, pattern recognition, simulation, and other disciplines. That can be attributed to its exceptional performance in self-learning, adaptability, and nonlinearity [10]. Furthermore, utilizing Artificial Neural Network (ANN) models can significantly decrease research expenses and efficiently conserve computing resources [11, 12].

The ANN algorithm has been extensively implemented across many fields for prediction purposes. Fidan [13] constructed an ANN model to forecast the thermal properties of concrete based on its mechanical features. In order to optimize model performance, a range of activation functions, including sigmoid, tanh, and triangle functions, are employed. Furthermore, they employ cross-validation to guarantee robust generalization and mitigate the risk of overfitting. The results demonstrated that the suggested ANN model for forecasting the thermal characteristics of concrete can accurately carry out prediction tasks.

In his study, Naser [14] utilizes ANN to model trip creation, the initial phase of transportation planning. This study examined the performance of ANN using data collected from the Central Business Department (CBD) in Nasiriyah City. ANN models outperform statistical models, yielding higher performance. Kareem [15] assesses the efficacy of the Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) in predicting weather forecasts by comparing their levels of accuracy. While both have distinct traits and features, ANN demonstrates higher performance when carrying out prediction tasks.

Yang [11]) employed an ANN model to forecast the remaining durability of carbon fibre-reinforced composites (CFRCs) following low-speed impacts. They constructed an ANN model using the Backpropagation technique and successfully obtained favourable outcomes with an error rate of less than 5%. Baashar [16] constructed an ANN model to forecast student accomplishment and academic performance. The results demonstrate the efficacy and attain a high level of accuracy in predicting academic achievement, hence facilitating the evaluation of student performance in the academic domain.

According to the published studies, it may be inferred that ANN exhibits high accuracy. However, to achieve precise outcomes, it is necessary to undergo appropriate training in ANN [6]. Thus, in this study, we aimed to implement hyperparameter tweaking on the Backpropagation technique, one of the ANN algorithms. The objective is to acquire ideal parameters to enhance the constructed model's performance.

2. RESEARCH METHOD

This study used an ANN-based Backpropagation algorithm. Backpropagation is one of the most popular ANN algorithms because of its efficiency and ease of implementation [17]. The data used is monthly data on the realization of the reverse duty tax on the name of motor vehicles in Lampung Province obtained from the Office of the Regional Revenue Agency of Lampung Province from January 2012 to June 2021. The stages to be carried out in the research process can be seen in the Figure 1:

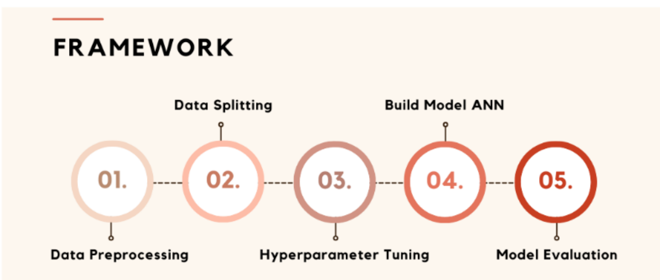

Figure 1. Proposed Model Framework

2.1 Data Preprocessing

Data preprocessing is the preliminary phase in the learning process, wherein data is manipulated or transformed into input that machines can readily process [18]. Furthermore, the utilization of applied data preparation techniques has the potential to enhance the performance of artificial neural networks (ANNs) and decrease their level of complexity [19]. Data preprocessing is predicted to take up 50% to 80% of the time necessary for the learning process. Preprocessing has a crucial role in model construction, as demonstrated by Kadhim [20].

2.1.1 Data Cleaning

Data cleaning refers to the systematic procedure of identifying and rectifying mistakes and nonconformities in data sources to enhance their quality. Data cleaning encompasses the necessary steps to rectify incomplete and missing data, among other issues, to preserve the accuracy of prediction outcomes [21].

2.1.2 Data Transformation

Data scaling converts values from their original format to a different one to facilitate data analysis and processing. Although multiple data scaling approaches exist, identifying the appropriate scaling method is the primary difficulty. Several studies have demonstrated the impact of data scaling on various algorithms [22]. Sinsomboonthong 's research [23] specifically examined the performance of scaling techniques in the ANN model, such as statistical column, decimal scaling, and min-max normalization. Min-max normalization is widely regarded as the optimal method for scaling data due to its ability to yield the most precise outcomes. Consequently, this study will employ min-max normalization as a method of scaling. Min-max normalization, sometimes referred to as min-max scaling, is a straightforward method for scaling data by adjusting the range of feature values to either [0, 1] or [−1, 1] [24].

2.2 Data Splitting

Data splitting, or train-test split, refers to partitioning data into distinct subsets to train and evaluate models independently. That aims to prevent overfitting in the learning method, which might result in subpar model performance during testing [25].

Kumar [26] suggests that larger training sets are preferable to evaluation sets. Moreover, numerous studies propose using data training or testing schemes that partition datasets by ratios of 70:30 or 80:20.

Pham et al. [27] found that enlarging the training data can enhance training performance and increase the stability of the model. Consequently, we employed a modified version of data splitting, allocating 60% of the data for training: 40% for testing, 70% for training: 30% for testing, and 80% for training: 20% for testing. That was done to get the best possible performance of the model.

2.3 Hyperparameter Tuning

Yang and Shami [28] define hyperparameter tuning as constructing an ideal model architecture by systematically exploring different parameter combinations. Hyperparameters are crucial in creating efficient learning models, particularly for tree-based and ANN models with numerous parameters [29]. The variables employed in this investigation were epoch and batch size.

According to Vijayalakshmi [30], an epoch is the number of iterations needed throughout the training process. In this context, batch size denotes the number of training groups employed during a single iteration. These two factors heavily influence the learning process and are considered the most crucial hyperparameters [31].

2.4 Artificial Neural Network

2.4.1 Nodes in the Hidden Layer

The number of hidden layers applied impacts changes in learning outcomes. The criteria typically employed for determining the number of hidden layers are as follows:

a. The number of hidden layers exceeds the number of nodes in the input or output layers.
b. The number of hidden layers is calculated as two-thirds of the sum of the total nodes in the input layer and the number of nodes in the output layer.
c. The maximum number of hidden layers is twice the number of nodes in the input layer.

2.4.2 Activation Function

The activation function is also known as the transfer function. The activation function determines whether a neuron needs to be activated [32]. The activation function used in this study is the Rectified Linear Unit (ReLU).

The Rectified Linear Unit (ReLU) activation function is employed to assign values collected from the layer. The ReLU function sets a threshold at zero so that when x is less than or equal to zero, x is equal to zero, and when x is more than zero, x remains unchanged. Equation (1) represents the mathematical expression that defines the activation function of the ReLU.

$$f(x) = \max(0, x) \tag{1}$$

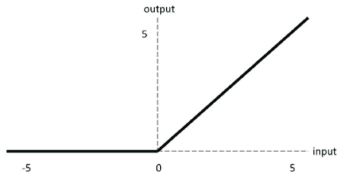

Figure 2. ReLU Activation Function

2.5 Validation Model

Hanke [33] asserts that forecasting approaches utilizing quantitative data with a specific temporal sequence can result in inaccuracies or mistakes. Hence, there is a want for methodologies that may be employed to assess inaccuracies in predicting methodologies. According to Kim [34], there are multiple techniques available for assessing models, such as:

a. Mean Squared Error (MSE) is a metric that measures the average squared difference between the observed target value and the anticipated value.
b. Root Mean Squared Error (RMSE) is a statistic that quantifies the average discrepancy between projected and actual values.
c. Mean Absolute Percentage Error (MAPE) is a metric that quantifies the percentage difference between the anticipated value and the actual value. A model's prediction or forecasting is deemed accurate when the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) numbers approach zero.

The mathematical equations of MSE, RMSE, and MAPE are as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2 \tag{2}$$

$$RMSE = \sqrt{MSE} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - X_i}{Y_i}\right| \tag{4}$$

Where:

X_i = Predicted i-th value
Y_i = Actual value i-th
n = number of attempts

3. RESULTS AND ANALYSIS

3.1 Data Preprocessing

Once the data is entered, it will be visualized by generating plots to display the distribution pattern within the data.

Figure 3. BBNKB Tax Data Plot

In addition, several stages are carried out, including verification of missing data and data normalization. At the verification stage, there is no missing data. So, no data cleaning is needed. Next, the data is normalized using min-max normalization, where the data range is converted into a range of 0 to 1.

3.2 Data Splitting

The methodology for splitting the training and testing data in this investigation is illustrated in Table 1.

Table 1. Splitting Training Data and Testing Data

Data Splitting	Amount of Data
60% data training and 40% data testing	54 data training and 36 data testing
70% data training and 30% data testing	63 data training and 27 data testing
80% data training and 20% data testing	72 data training and 18 data testing

3.3 Build model

The model constructed in this study is the ANN model utilizing the ReLU activation function. The ReLU activation function transforms the data range from 0 to infinity. That results in negative values being replaced with 0, while positive values remain unchanged. Hence, the ReLU activation function is highly appropriate for time series data as it yields superior neuron activation when actual values are employed. The additional parameters employed are the input layer, two hidden layers, output layer, number of neurons or units in the hidden layer, epoch, and batch size. The parameter values utilized for this investigation are as follows.

Table 2. Parameters Used

Parameter	Parameter Value
Epoch	50 dan 100
Batch Size	16 dan 32
Activation Function	ReLU

The ideal parameter values are determined through hyper-tuning, which involves implementing early stopping to halt the model's learning process once the best parameter has been identified. The optimal parameter values are determined based on the findings of hypertuning.

Table 3. Optimal Parameters for Composition of 60% Training Data and 40% Testing Data

Parameter	Parameter Value
Epoch	50
Batch Size	16
Fungsi Aktivasi	ReLU
Layer	Input: 1 layer Hidden: 2 layers consisting of 24 nodes and 17 nodes. Output: 1 layer

Table 4. Optimal Parameters for 70% Data Training and 30% Data Testing

Parameter	Parameter Value
Epoch	100
Batch Size	16
Activation Function	ReLU
Layer	Input: 1 layer Hidden: 2 layers consisting of 24 nodes and 17 nodes. Output: 1 layer

Table 5. Optimal Parameters for 80% Training Data and 20% Testing

Parameter	Parameter Value
Epoch	50
Batch Size	16
Activation Function	ReLU
Layers	Input: 1 layer Hidden: 2 layers consisting of 24 nodes and 17 nodes. Output: 1 layer

The structure of the artificial neural network (ANN) in Figure 4 is an architecture built based on the parameters obtained via the hyperparameter tuning technique in Table 3, Table 4, and Table 5.

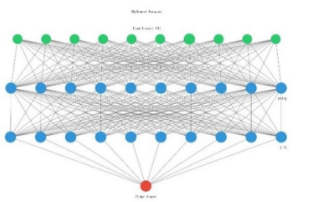

Figure 4. Built Artificial Neural Network Structure

3.4 Model Testing

Model testing is conducted after identifying the optimal parameters for constructing the model. At this stage, we acquire the loss value and validation loss value. A lower loss value indicates a higher-quality model.

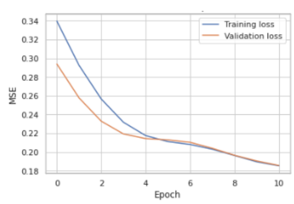

Figure 5. Loss Graph for 60% of Training Data

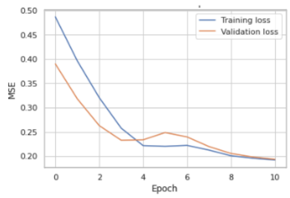

Figure 6. Loss Graph for 70% of Training Data

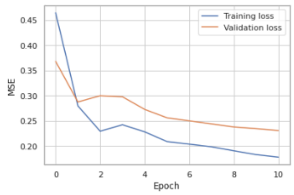

Figure 7. Loss Graph for 80% of Training Data

Figures 5, 6, and 7 indicate that all models using different training testing data-splitting strategies, such as 60:40, 70:30, and 80:20, have demonstrated optimal performance without encountering overfitting. That is evident from where the loss values and validation loss come together.

3.5 Prediction

The model was employed to predict the imposition of inverse duty tax on motor vehicle designations in Lampung Province after the testing phase. An additional analysis is conducted on the derived predictions by comparing them with the actual data. Once the anticipated outcomes demonstrate a strong correspondence with the empirical data, the developed model may be a dependable instrument for predicting the forthcoming tax realization on motor vehicle names.

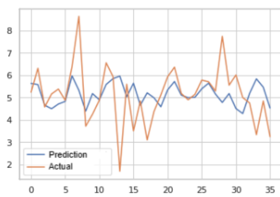

Figure 8. Plot 60% Training Data Prediction

Figure 8 displays two lines: the blue line represents the predicted results, while the orange line shows the actual data. Consequently, the results did not align with the data distribution when 60% of the training data was used for predictions. There were notable disparities between the predicted values and the actual data, indicating that the constructed model is insufficient for accurate forecasting.

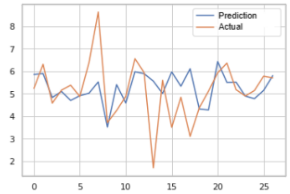

Figure 9. Plot 70% Training Data Prediction

The plot of prediction results utilizing 70% of the training data exhibits a striking resemblance to the actual data distribution pattern, as confirmed by Figure 9. Nevertheless, a disparity persists between the predicted values and the actual data, suggesting that the model's prognostications are reasonably precise.

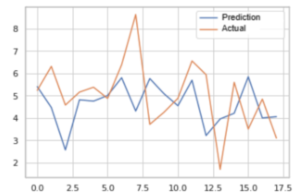

Figure 10. Plot 80% Training Data Prediction

The plot of prediction results utilizing 80% of the training data more closely than other models resembles the distribution pattern of the actual data, as indicated by Figure 10. This finding illustrates that the predicted value corresponds precisely to the actual data, suggesting that the developed model can be relied upon to foresee or predict BBNKB taxes.

3.6 Model Evaluation

RMSE and MAPE values are utilized in this study to evaluate the accuracy of prediction models. A comparison of the accuracy values acquired for each constructed model is presented below:

Table 6. Comparison of Accuracy Values

Data Schema	Model Evaluation		
	RMSE	MAPE	Accuracy
60% data training, 40% data testing	0.189	4.51%	95.48%
70% data training, 30% data testing	0.207	4.16%	95.83%
80% data training, 20% data testing	0.204	3.48%	96.51%

As shown in Table 6, the highest achievable accuracy was 96.51% when a data training scheme of 80% was utilized in conjunction with a 20% data testing scheme. This finding illustrates the model's exceptional predictive capabilities, attaining a 96.51% exact accuracy rate. Thus, this strategy for training data is exceptionally suitable for forecasting.

3.7 Forecasting

Once the accuracy number reaches a reasonable level, the next step is forecasting. Based on the existing findings, the model that yields the highest accuracy in forecasting future events employs an 80% data training and a 20% data testing approach. The following findings provide predictions on the implementation of the reverse duty tax on motor vehicles in Lampung province:

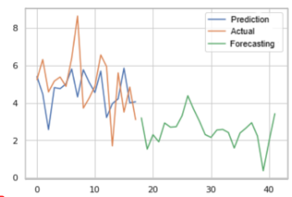

Figure 11. Plot Combination of Forecasting Results

Based on Figure 11, the forecasted results for the Lampung Province motor vehicle reverse tax name for two years have exhibited volatility, as depicted by the green line. The blue line corresponds to the projected outcomes, whereas the orange line corresponds to the factual facts.

4. CONCLUSION

This study utilizes the Artificial Neural Network (ANN) method to predict the future collection of motor vehicle name return duty tax in Lampung for two years. The algorithm is trained using monthly data from the Office of the Regional Revenue Agency of Lampung Province, covering January 2012 to June 2021.

We employ different data splitting schemes for training and testing: 60%: 40%, 70%: 30%, and 80%: 20%. Furthermore, we employ hyperparameter tuning to optimize the epochs and batch sizes to achieve optimal values. The optimal parameters for predicting the occurrence of motor vehicle name reverse duty tax are 50 epochs and 16 batch sizes. The training data scheme consists of 80% of the data, while the remaining 20% is used for testing. The ReLU activation function is employed.

The model exhibits high predictability, attaining an accuracy rate of 96.51%. Furthermore, it exhibits a mean absolute percentage error (MAPE) of 3.48% and a low root mean square error (RMSE) of 0.246. Therefore, it is appropriate to make predictions for the forthcoming two-year duration. Therefore, the forecast for collecting BBNKB taxes in Lampung has fluctuated.

Application of the Artificial Neural Network Method to Predict the Realization of the Duty Tax on the name of Motor Vehicles in Lampung Province

ORIGINALITY REPORT

10%	11%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ejournal.uin-suska.ac.id Internet Source	5%
2	ejournal.unsri.ac.id Internet Source	3%
3	Moon Seong Rang. "APPLICATION OF GREY MODEL AND ARTIFICIAL NEURAL NETWORKS TO FLOOD FORECASTING", Journal of the American Water Resources Association, 4/2006 Publication	1%
4	link.springer.com Internet Source	1%
5	Yuhe Wang, Yu Yang, Rencai Gao, Shiming Li, Yan Zhao. "A Security Situation Prediction Model for Industrial Control Network Based on EP-CMA-ES", IEEE Access, 2023 Publication	1%
6	fastercapital.com Internet Source	1%

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	Off		