Artificial Intelligence Powered Revenue Forecasting: Investigate the Use of Ai to Predict Local Government Revenue

Laode Buzyali Fajman Sanbigs

S. IP M.P.P, The Regional Revenue Agency

Abstract
This paper explores the application of Artificial Intelligence (AI) in revenue forecasting for local governments. Revenue forecasting plays a crucial role in effective financial planning and resource allocation, enabling governments to make informed decisions regarding budgeting, expenditure, and policy formulation. Traditional revenue forecasting methods often rely on historical data and statistical models, which may have limitations in capturing complex patterns and dynamic factors that influence the revenue generation. This study aims to investigate the feasibility and effectiveness of using AI techniques to predict local government revenue. By leveraging Machine Learning (ML) algorithms and advanced data analytics, AI has the potential to uncover hidden patterns and correlations within vast and diverse datasets. The research methodology involves collecting historical revenue data from multiple local government entities and also implementing various AI models, including regression-based algorithms and neural networks. The models are trained and validated using a comprehensive dataset that includes relevant economic variables, such as Population, GDP, The Unemployment rate, and Hotel Occupancy. The results of the study demonstrate the potential of AI-powered revenue forecasting in local government contexts with high-accuracy prediction. The findings also highlight the significance of incorporating dynamic external factors into the forecasting process, as AI models can capture complex relationships and identify key drivers of revenue generation.

Keywords: Artificial Intelligence (AI), Revenue forecasting, Local government, and Machine learning (ML).

1. Introduction

Revenue forecasting is an important aspect of financial management not only for private but also for public entities. Revenue prediction can help both sectors to make more informed decisions, stay ahead of competitors, and identify new opportunities for revenue growth (Hasin, et al., 2018). In this regard, artificial intelligence (AI) has emerged as a promising tool for estimating revenue (Bughin, et al., 2017). AI is a highly multidisciplinary system that mimics and enhances human intelligence through computing. It combines concepts and methods from various disciplines, including linguistics, computer science, mathematics, psychology, and philosophy (Russel & Norvig, 2010). This advantage has encouraged public sector institutions such as local governments, to explore more about the use of AI in forecasting revenue (Papadimitriou & Ntatsopoulou, 2021).

Then, how can AI predict local government revenue? On one hand, local government revenues are the money that local governments raise to pay for their operations and services. It comes from taxes, fees, levies, fines, and intergovernmental grants (Smith, et al., 2013). Although the sources vary, more than half (57.7%) of their income comes from taxes such as land and building tax, sales tax, income tax, and consumption tax (The Ministry of Finance, 2020). In this case, several local governments in different countries have used AI to predict local government revenues through their tax sector revenues. For example, a case study from New York City has shown a positive relationship between the use of artificial intelligence and predicting property tax income. It explains how machine learning algorithms can analyze data about property values, property characteristics, and other factors to predict property tax revenue in New York City (Datta & Datta, 2021). The authors used a dataset containing information about related objects and other factors that can affect property tax revenue and found which features were most significant in predicting property tax revenue, such as property price, number of bedrooms and bathrooms in the property, and the age of the building. Additionally, Kim, et al (2018) also wrote a paper that supported it. The authors used machine learning algorithms to predict local tax revenue in Korea. They used a dataset of local tax revenue, economic and demographic indicators, and other relevant data. They then experimented with various machine learning algorithms, such as decision trees, random forests, and regression models. The results show that machine learning algorithms can perform more effectively than traditional statistical methods in predicting tax revenue. In China, recent studies also found a similar result: various machine learning techniques of AI could analyze large datasets related to housing prices, population, and other factors and positively impact property tax revenue generation (Zhang & Lu, 2019). The paper found that machine learning algorithms can perform to identify areas of potential revenue growth and make more targeted investments in infrastructure and other projects to support economic development. These papers show the apparent capabilities of using AI for predicting local government revenue.

Furthermore, AI-powered revenue forecasting can improve the accuracy of revenue predictions, generate revenue forecasts more quickly and identify patterns and trends in data that might not be apparent through traditional methods (Gorecki & O'Mahony, 2021). This high technology can help local governments to better anticipate their future revenue and plan accordingly. For example, if a department has a more accurate forecast of its future revenue, it can make more informed decisions about investments, resource allocation, and pricing strategies that have bigger opportunities for increasing revenue. Every organization, including the government, must consider the importance of analyzing historical trends and anticipating future developments. When planning, an organization is more effective, especially when it involves financial considerations. Analyzing the company's historical data will help with this (Muskaan & Sarangi, 2020).

Many experts posit that AI could make predictions about local government revenue. However, there are also concerns that the use of AI in this field might be difficult to implement as some proponents suggest. In previous studies, AI has not been widely considered a factor driving local government revenues. A report from the National League of Cities stated that while AI has the potential to help local governments manage resources better, it has not been commonly used to generate revenue (National League of Cities, 2020). It is mainly because most local governments still utilize handwritten forecasts and analysis of historical data as their primary methodology for estimating revenue (National State Association, 2020). Additionally, AI also requires large amounts of data to be effective, and local governments may not have access to the large data sets needed to take full advantage of AI (Center for

Digital Government, 2020). The quality and availability of data on local government revenue can vary widely, which may limit the usefulness of AI in making predictions. The results are sensitive to the choice of input variables and data preprocessing and feature selection are critical for improving prediction accuracy.

Nonetheless, there are studies that have defied these barriers. In 2018, a study in the Indian state of Andhra Pradesh showed that machine learning models of AI could accurately predict local government revenue, with a Mean Absolute Percentage Error (MAPE) of less than 10% (Reddy, et al., 2018). MAPE is a commonly used measure of the accuracy of a forecasting model, which expresses the average magnitude of the forecast errors as a percentage of the actual values (Hyndman & Koehler, 2006). A lower number of MAPELs indicates a more precise forecasting model which leads to a better result. In this paper, the authors used a dataset of revenue and expenditure data from the state's urban local bodies and trained predictive models using various machine learning algorithms, including decision trees, neural networks, and support vector machines. Then, another study by Huang, Zhang, and Ye (2020), also reports a similar notion. Their hybrid AI model for revenue forecasting in municipal governments produces more accurate results than traditional methods. They contrasted their method's accuracy with that of a number of other approaches and found that their model had lower mean absolute percentage error (MAPE) and higher correlation coefficients between predicted and actual values. Despite all the limitations, both papers provide evidence that AI may still be an effective tool for predicting local government revenue.

In addition, like other jurisdictions, the Indonesian central government just started to use this AI system in public and government services. There are several initiatives and developments that indicate a growing interest in the use of AI. In 2020, the Ministry of Communication and Information Technology (2020) formulated the National AI Roadmap Strategy 2020-2045, which outlines the government's plan to develop AI in Indonesia, including the goals, objectives, and key strategies that will be implemented over the next 25 years. This document also highlights the various sectors that are prioritized, such as healthcare, education, agriculture, and finance including local government revenue.

This paper aims to do further research in this field by providing a mixed-method analysis of how AI predicts local government revenue. While local government revenue forecasting has been extensively studied, the use of AI as a factor that drives these precepts has not been explored at large. This policy paper analyzes AI efforts and revenue administration policies and how this technology works to provide revenue predictions. Then, it offers policy advice for local government revenue administration, such as Indonesia, which has not maximized the use of AI to predict local government revenue.

2. Research question

How can artificial intelligence help the local government predict its own revenue?

3. Methodology

This paper uses a mixed method to describe how AI works to predict local government revenue. Firstly, for the qualitative portion of this study, I examined the existing research to support the argument that Artificial Intelligence can forecast local government revenues. Then, I reviewed existing practices of AI-powered revenue forecasting in specific revenue areas including government policies, reports, forum transcripts, and academic literature related to this topic. The interview session was conducted with AI experts and representatives of relevant agencies in Indonesia as well as other jurisdictions. Finally, the quantitative analysis through a neural network model illustrates how the basic principle of AI works in this field.

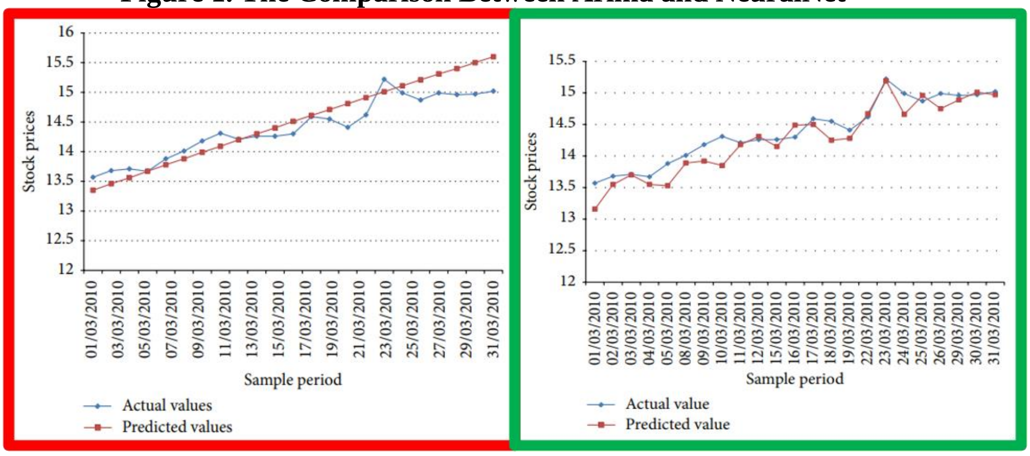
4. AI Model for Predicting Local Government Revenue

1.2 4.1 Choosing The Machine Learning Model

Artificial Intelligence (AI) encompasses the development of intelligent machines capable of performing tasks that usually rely on human intelligence. Within this field, machine learning emerges as a specific subset, concentrating on the creation of algorithms and models that learn from data to make predictions or take actions without direct programming. A machine learning model is a mathematical representation of patterns and relationships in the data, which is used to make predictions or decisions. According to Sarker (2021), there are some types of machine learning models such as supervised, unsupervised, semi-supervised, and reinforcement learning exist in the area. Firstly, supervised learning involves training a machine learning model to understand the relationship between inputs and outputs using example pairs of input-output data. Secondly, unsupervised learning involves analyzing datasets without labeled information, allowing the model to discover patterns on its own, without human intervention. Thirdly, semi-supervised learning combines elements of supervised and unsupervised methods by utilizing both labeled and unlabeled data. Lastly, reinforcement learning is a machine learning approach where software agents or machines autonomously learn the best actions to take in a specific context or environment, aiming to enhance efficiency. It focuses on learning from the consequences of actions rather than explicit instructions.

In this context, AI can predict local government revenue precisely using one of the reinforcement learning model types which is called neural networks. The neural network has been chosen because it has better results in terms of predictions. According to a study by Xu, L., et al. (2020), two of the best machine learning models for revenue prediction are Artificial Neural Networks (ANNs) and Gradient Boosting Machines (GBMs). ANNs are highly adaptable and can model nonlinear relationships between input and output variables in large, complex datasets. Based on these patterns and relationships, the neural net model can predict future revenue trends with a high degree of accuracy. Another research from Adebiyi, et al., (2014), also recommends a neural networks model than a conventional statistical analysis which is commonly used such as the Arima Model in terms of prediction (See Figure 1).

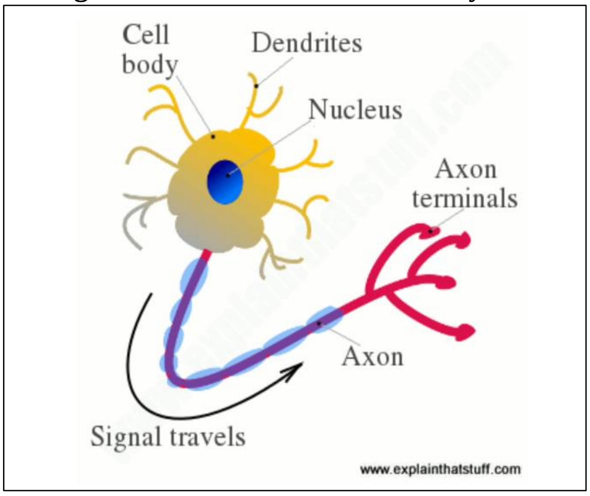
Figure 1. The Comparison Between Arima and NeuralNet



Source*:* (Adebiyi, A. A et al., 2014)

The graphs compellingly demonstrate the superiority of neural networks over the Arima model. Unlike the Arima model, which merely performs linear predictions, the neural network intelligently captures trends, whether they are upward or downward. This adaptability enables the neural network to outperform the Arima model consistently.

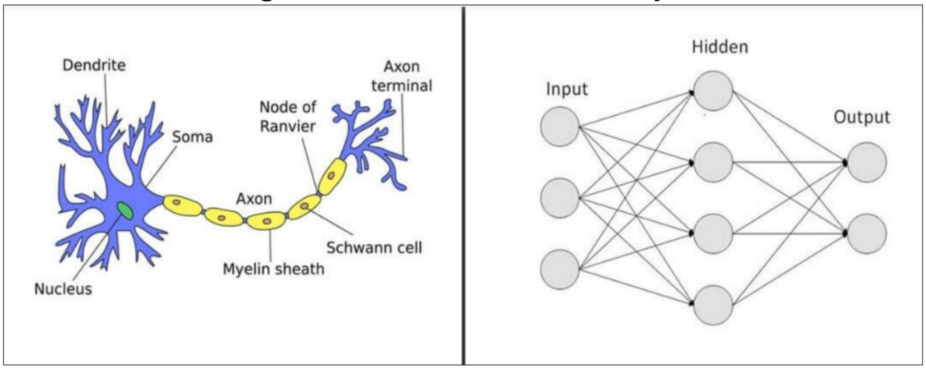
Figure 2. Nueron Structure and System



The high performance of the neural net is attributable to the fact that it draws inspiration from the human brain in its structure and function (Nielsen, 2015). In line with that notion, Woodford (2023) explained that a typical brain comprises approximately 100 billion small cells known as neurons, which is each neuron consists of a cell body, a central mass, and multiple connections, including dendrites (which carry information toward the cell body) and a single axon (which carries information away from the cell) (see figures 2). In an attempt to replicate this biological system in a computer, humans use tiny endoscopic switching devices called transistors as the equivalent of a brain cell. This model of computer consists of interconnected nodes, referred to as neurons, which are responsible for processing and transmitting information. Each neuron receives input from other neurons and applies mathematical functions to produce output. The fundamental concept of a neural network involves creating a simulation of numerous densely interconnected brain cells within a computer. This simulation enables the computer to learn, recognize patterns, and make decisions in a manner that resembles human thinking (Davis, 2019). The fundamental concept is to simulate interconnected brain cells, called neurons, in a computer system. Afzal et al. (2023) explain that a neural network typically comprises three primary types of layers: the input layer, the hidden layer(s), and the output layer (see Figure 3). Here are the explanations:

2. The Input Layer serves as the initial data entry point, where each neuron corresponds to a distinct feature or attribute related to the given problem.
3. The Hidden Layer(s), positioned between the input and output layers, are responsible for processing and converting the input data. Neurons in these layers receive input from the previous layer, perform mathematical operations, and generate intermediate representations within the network, hence their name "hidden."

Figure 3. The Neural Network Layer



4. The Output Layer produces the ultimate result of the neural network. It delivers predictions or decisions based on the processed information from the input and hidden layers. The number of neurons in the output layer varies according to the specific problem, such as two neurons for binary classification or multiple neurons for multi-class classification.
5. Weights represent the connections between neurons in different layers. They play a crucial role in determining the strength and impact of each neuron's contribution to the overall computation. In a neural network, calculating the output of a neuron involves taking the weighted sum of its inputs, adding a bias term, and applying an activation function.

Mathematically, the formula for a single neuron in a layer can be expressed as:

$$z = (wx) + b$$

Where *z* is a matrix representing the weighted sums for each neuron in the current layer. *w* is a matrix of weights connecting the previous layer's neurons to the current layer's neurons. For example *w* = [0.2, 0.3, 0.4, 0.5]. This number can be changed during the process. *b* is a vector representing the biases for each neuron in the current layer. During training, the neural network adjusts these weights to optimize its performance and improve its ability to make accurate predictions. This optimization is achieved through a technique called backpropagation. Backpropagation involves feeding the network with training examples, comparing the predicted outputs with the desired outputs, and calculating the error (Mujtaba & Sowgath, 2022). The network then propagates the error backward, layer by layer, and adjusts the weights in the opposite direction of the gradient of the error with respect to the weights. This iterative process continues until the network achieves a satisfactory level of accuracy. By leveraging the interconnectedness of neurons and the ability to adjust their weights, neural networks can capture complex relationships and extract meaningful features from data. This flexibility allows them to excel in various tasks, including image recognition, natural language processing, speech recognition, and many others. It is important to note that the explanation provided here is a simplified overview of neural networks. The field of neural networks encompasses various architectures, activation functions, optimization techniques, and other advanced concepts that contribute to their effectiveness in solving complex problems.

It is essential to recognize that neural networks are essentially software simulations. Programming standard computers create them to mimic the behavior of billions of interconnected brain cells, though they operate in a conventional, serially connected manner using ordinary transistors and logic gates. It is worth noting that no attempts have been made to construct a computer with transistors wired in a densely parallel structure identical to the human brain. Essentially, a neural network can be likened to a computer model of the weather, while a human brain is comparable to the actual clouds, snowflakes, or sunshine. This analogy highlights the distinction between the two, emphasizing that neural networks are simulations designed to mimic the brain's functioning, rather than being the real thing (Nagyfi, 2018). Computer simulations consist of algebraic variables and mathematical equations connecting them, which are essentially numbers stored in changing values within boxes. These values are insignificant to the computers themselves but meaningful to the programmers who design them. Consequently, when building a model, it is crucial to tailor it to our specific interests, such as predicting revenue, to achieve the best possible outcomes.

5.2 4.2 Building AI Model

4.2.1 Data Collection

The Neuralnet model can be used to predict local government revenue by analyzing historical data and making use of machine learning algorithms. This process involves gathering relevant variables such as population and GDP (Lumy, et al., 2018), GDP per capita (Tahwin, 2013), Hotel Occupancy, and Unemployment Rate (Rafsanjani, 2015). This paper collected data from 20 provinces during 2010-2021, and then the neural net model will analyze this data to identify patterns and relationships within it. The huge amount and complex data are extremely important for machine learning before running the analysis.

4.2.2 Cleaning Data

Cleaning the dataset before inputting it into RStudio or any other data analysis tool is an important step to ensure the accuracy and reliability of your analysis. It provides a solid foundation for accurate and meaningful analysis, leading to more robust insights and reliable results. Dataset cleaning involves handling missing values, dealing with outliers, handling inconsistent data formats, and addressing other data quality issues. This is a step-by-step explanation of the dataset-cleaning process:

a. Handling missing values. Identify and handle missing values in your dataset. Missing values can occur due to various reasons such as data collection errors, equipment malfunction, or human error. You can handle missing values by either removing the rows or columns containing missing values or imputing the missing values with appropriate techniques such as mean imputation, median imputation, or regression imputation.
b. Dealing with outliers. Outliers are extreme values that deviate significantly from the normal pattern of the data. Outliers can distort statistical measures and affect the performance of models. You can identify outliers using statistical techniques like box plots, scatter plots, or z-scores. Once identified, you can handle outliers by either removing them if they are due to data entry errors or transforming them using techniques like winsorization or logarithmic transformation.
c. Handling inconsistent data formats. In real-world datasets, data can often have inconsistent formats, such as inconsistent date formats, categorical variables with different levels, or numerical variables with different units. Standardize the data formats to ensure consistency. For example, convert dates into a common format, ensure consistent labeling of categorical variables, and convert numerical variables to a consistent unit of measurement if required.
d. Addressing data quality issues. Analyze the dataset for any data quality issues, such as duplicate records, erroneous data, or inconsistent naming conventions. Remove duplicates if they exist and correct any erroneous or inconsistent data entries.
e. Feature selection and transformation. Evaluate the relevance and importance of the variables in your dataset. Remove any irrelevant or redundant variables that do not contribute to the analysis. Additionally, you may need to transform variables if they violate the assumptions of the analysis technique being used. For example, transforming skewed variables using logarithmic or power

transformations.

f. Data normalization. Normalize the data if necessary. Normalization scales the data to a common range to avoid the dominance of certain variables due to their scale. Common normalization techniques include min-max scaling or z-score standardization.

g. Data validation. Finally, perform a thorough validation of the cleaned dataset to ensure its integrity and correctness. Check for any remaining anomalies, inconsistencies, or errors that might have been missed during the cleaning process.

By following these steps, this paper collects and cleans the raw data from the National Statistical Agency of Indonesia. After ensuring that the dataset is clean, reliable, and suitable for analysis in RStudio or any other data analysis environment (see Table 1).

Table 1. The Panel Data

1	Province	Year	Population	GDP	GDP Per Kapita	Hotel Occupancy	Unemployment Rate	Local Government Revenue
2	Jakarta	2010	9610000	1075183480000000	111500000	51.76	11.05	17825987294431
3	Jakarta	2011	9891943	1224218480000000	125530000	56.05	10.80	22040801447924
4	Jakarta	2012	9862138	1369432640000000	138860000	56.37	9.87	26852192452636
5	Jakarta	2013	9969948	1546876490000000	155150000	55.68	8.63	31274215885720
6	Jakarta	2014	10075300	1762316400000000	174914360	58.27	8.47	33686176815708
7	Jakarta	2015	10177921	1989088750000000	195460000	62.47	7.23	36888017587716
8	Jakarta	2016	10277628	2159073600000000	210075000	57.89	6.12	43901488807742
9	Jakarta	2017	10374235	2365353850000000	228003000	67.66	7.14	43327136602811
10	Jakarta	2018	10467629	2592606570000000	247678000	66.65	6.24	45707136602811
11	Jakarta	2019	10557810	2815636160000000	268052000	59.71	6.22	57561136602811
12	Jakarta	2020	10562088	2767273490000000	262702000	41.22	10.95	41617136602811
13	East.Java	2010	37565706	990648840000000	26371100	46.05	4.25	8898616683297
14	East.Java	2011	37840657	1120577160000000	29613050	47.81	5.33	9584081971227
15	East.Java	2012	38106590	1248767290000000	32770380	47.44	4.09	11579340719022
16	East.Java	2013	38363195	1382501500000000	36037180	46.22	4.30	14442216534959
17	East.Java	2014	38610202	1537947630000000	39832680	50.81	4.19	15402647674503
18	East.Java	2015	38847561	1691477060000000	43541400	55.56	4.47	15817795024797

Source: The National Statistical Agency of Indonesia, 2023

4.2.3 Determinant Variable Selection.

After the data has been prepared, the next step is to select the most relevant variables (local government revenues) that will be used to train the AI model such as decision trees, neural networks, genetic algorithms, and probabilistic models, among others (Russell, S. J., & Norvig, P., 2010). In the case of Indonesia, however, before we used this dataset we have to determine whether all the variables of the dataset have significance to the targeted variables or not. The output print (see Tabel 2) displays the results of the variable significance analysis based on the p-values obtained from t-tests. Here is the interpretation of the results:

Tabel 2. The Significance of Variables

```
                                         Variable       P_Value
GDP                                           GDP 2.430409e-24
Unemployment.Rate               Unemployment.Rate 5.131832e-18
Hotel.Occupancy                   Hotel.Occupancy 5.131832e-18
Population                             Population 5.132465e-18
GDP.Per.Kapita                     GDP.Per.Kapita 5.134402e-18
Local.Government.Revenue Local.Government.Revenue 1.000000e+00
```

1. GDP with P-Value: 2.430409e-24. This indicates strong evidence against the null hypothesis. Therefore, it is highly likely that the variable "GDP" is statistically significant in its association with the target variable. Other than this significance this variable is chosen because it represents the economic situation of any region. As we know that GDP contains consumptions, investment, government spending,

and also export minus imports. In this regard, consumption reflects people's purchasing power, which directly impacts tax revenue generated from the sale of goods and services. It serves as an essential indicator of consumer demand and economic activity, influencing the overall fiscal health of a region or country. On the other hand, government spending refers to the expenditures made by the government on various public goods, services, and infrastructure projects. This includes investments in education, healthcare, transportation, defense, and social welfare programs. Government spending plays a pivotal role in stimulating economic growth, creating job opportunities, and supporting various sectors of the economy. Additionally, the net exports component of GDP, which is the difference between a country's exports and imports, signifies the international trade balance. A positive net export value contributes to a trade surplus, indicating that the country is exporting more than it imports, while a negative value suggests a trade deficit. This trade balance influences a nation's overall economic competitiveness and can have implications on its exchange rates and foreign investments. By considering these essential economic components, GDP becomes a powerful indicator of an economy's overall performance and its potential impact on the target variable being studied. By involving this variable, this paper tries to include any condition as much as possible, to increase the accuration of the prediction.

2. The Unemployment Rate with P-Value: 5.131832e-18 presents compelling evidence against the null hypothesis, signifying that the variable holds significant statistical importance in its correlation with the target variable. This indicates that the Unemployment Rate plays a crucial role in influencing the target variable's outcomes. Furthermore, the Unemployment Rate serves as a more focused indicator compared to the broad measure of GDP. While GDP encapsulates various economic factors, the Unemployment Rate provides specific insights into the purchasing power of people by examining the employment situation. High unemployment rates can lead to decreased consumer spending, as unemployed individuals often face financial constraints, affecting overall consumption patterns and demand. By incorporating the Unemployment Rate as an independent variable, this study aims to gain a deeper understanding of its impact on the target variable, considering its potential influence on people's purchasing behavior and, subsequently, the broader economic conditions. This finer granularity allows for a more nuanced analysis, contributing to the accuracy and effectiveness of the predictive model in capturing the complexities of the relationship between the Unemployment Rate and the target variable.

3. The Hotel Occupancy with P-Value: 5.131832e-18. This indicates strong evidence against the null hypothesis, suggesting that the variable is likely to be statistically significant in its association with the target variable. Apart from its statistical significance, the inclusion of the Hotel Occupancy variable is justified by its relevance to the study. Hotel Occupancy serves as a valuable proxy for gauging tourism and hospitality activity within a region. Higher occupancy rates suggest increased tourism and business travel, which can stimulate economic growth through various channels such as hospitality sector revenues, local businesses, and job creation. By incorporating the Hotel Occupancy variable, this research aims to capture the influence of the tourism industry on the target variable. Understanding its impact is vital, as tourism can significantly contribute to local economies and affect government revenues, particularly through taxes and other related income streams. As a result, the consideration of Hotel Occupancy in the analysis enhances the model's ability to discern and predict the effects of tourism-related dynamics on the target variable, contributing to a more comprehensive and accurate predictive framework.

4. Population with P-Value: 5.132465e-18. This suggests strong evidence against the null hypothesis, indicating that the variable is likely to be statistically significant in its association with the target variable. Other than this significance value, in general, this variable has a strong relationship with tax revenue. The

greater population will lead to greater tax revenue which significantly affects the local government revenue (Yusuf, 2016). As the population grows, it can lead to increased tax revenue, which plays a vital role in shaping the financial condition of the local government.

5. GDP Per Capita. P-Value: 5.134402e-18. This suggests strong evidence against the null hypothesis, indicating that the variable is likely to be statistically significant in its association with the target variable and can contribute meaningfully to the predictive model's accuracy. Other than this significance this variable is chosen because it provides insights into the economic well-being of individuals in a region, taking into account the average income or output per person. By considering GDP Per Capita in the analysis, we can better understand the distribution of economic prosperity among the population and its potential influence on the target variable. Moreover, GDP Per Capita serves as a valuable indicator of the overall standard of living, reflecting the purchasing power and consumption patterns of the individuals in the region. It complements the broader GDP measure by accounting for population differences, allowing us to gain a more nuanced understanding of the economic dynamics at the individual level.

Overall, all the variables from the dataset are likely to be statistically significant in their associations with the target variable which is local government revenue. In this regard, we can apply this data use that to predict local government revenue.

4.2.4 Modify the Model

After establishing the statistical significance of the variables, the next step is to adjust the Neural Network Architecture accordingly. Since all the independent variables have significance to the targeted variable, then all of them will include into the calculation, resulting in the following formula:

$$z = (w_1x_1) + (w_2x_2) + (w_3x_3) + (w_4x_4) + (w_5x_5) + B$$

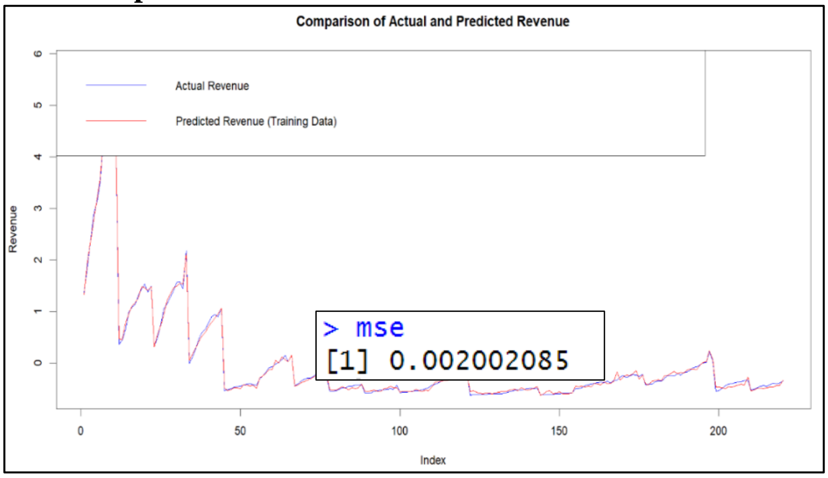
Where:

Independent variables: x_1 represents GDP, x_2 represents Population, x_3 represents Unemployment Rate, x_4 represents Hotel Occupancy, and x_5 represents GDP Per capita. w_1, w_2, w_3, w_4, and w_5 are the respective weights assigned to each independent variable. B represents the bias term. The dependent variable z represents local government revenue. These adjustments aim to enhance the model's accuracy and capture the relationships between the independent variables and the target variable more effectively.

4.2.5 Training the Model

In this paper, we used historical data from 2010-2020 to build the model, then, after feeding the data into the model and adjusting its parameters to minimize the error between the model's predictions and the actual values, the results obtained from the graphs (see Figure 4) indicate a promising level of accuracy and effectiveness in the model's predictions. The close alignment between the predicted values and the actual observed values suggests that the neural network has successfully learned the underlying patterns and relationships present in the data. The close proximity of the prediction and actual values on the graphs demonstrates the model's ability to generalize well to unseen data, indicating that it has not overfit the training dataset. This means that the model is not merely memorizing the training data but has acquired the capability to make accurate predictions on new, unseen data points. Moreover, the successful reduction in prediction errors validates the choice of the selected variables and the neural network architecture, showcasing the model's capacity to effectively capture the complexities and interdependencies among the independent variables in relation to the target variable. However, it is essential to conduct further evaluations and statistical tests to assess the model's robustness and generalizability rigorously. Techniques such as cross-validation, out-of-sample testing, and additional performance metrics can be employed to ensure the model's reliability and validate its predictive capabilities across diverse datasets

Figure 4. The Comparison Between Predicted and Actual Value For 2010-2020



Overall, the close match between the predicted and actual values, as observed from the graphs, instills confidence in the predictive power of the developed model. This outcome strengthens the credibility of the findings and supports its potential applications in real-world scenarios, where accurate predictions based on the chosen variables can be valuable for decision-making, policy planning, or resource allocation.

5.2.6 Model Validation and Testing.

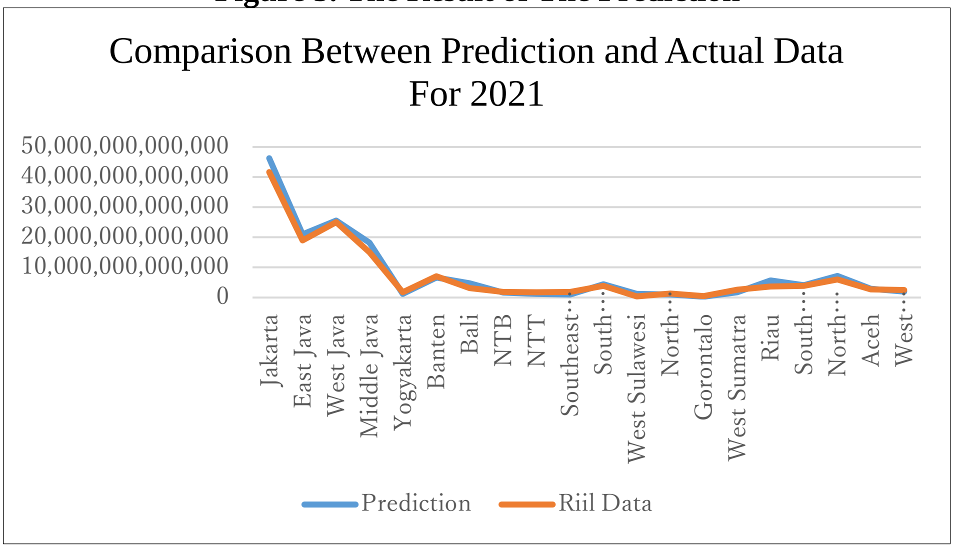
Once the model has been trained, it needs to be validated and tested to ensure that it is accurate and reliable. This involves comparing the predicted revenue figures to the actual revenue figures for a given period. In this neural network model, the output Mean Squared Error (MSE) is 0.002002085 (see Figure 4). MSE measures the average squared difference between the predicted values and the actual values. In this case, A smaller MSE value suggests a better fit of the model to the data, as it indicates that the predicted values are closer to the actual values on average. So that, it can be used to predict the next year's revenue.

4.3 Implementation of The Model

4.3.1 Deployment

After the model has been validated and tested, it can be deployed to predict future revenue figures for the local government in 2021. Here is the result (see Figure 5):

Figure 5. The Result of The Prediction



The results demonstrate a remarkable alignment between the predictions generated by the neural network and the actual data trends. With the successful deployment of this model, decision-makers and local government authorities stand to benefit significantly by gaining invaluable insights into the projected revenue figures for the upcoming year. Armed with this crucial information, they can make well-informed financial decisions and conduct effective budget planning for 2021, relying on the model's reliable projections. The model's rigorous validation and testing inspire confidence in its predictive capabilities, positioning it as an invaluable tool for revenue forecasting in the context of the local government.

4.3.2 Monitoring and Evaluation

After deploying the model, a crucial step is to continually monitor and evaluate its performance to ensure consistent and dependable predictions. This ongoing process may entail regular retraining of the model or updating it with the latest data and relevant features. Effective monitoring allows us to identify potential drifts or shifts in the data distribution, ensuring the model remains aligned with real-world changes. By regularly evaluating its output against ground truth data, we can assess its accuracy and reliability. Additionally, as new data becomes available, incorporating it into the model can improve its predictive capabilities, ensuring it remains relevant and up-to-date. This iterative approach to monitoring and evaluation helps us maintain a high-performing and trustworthy model throughout its operational life.

6. Conclusion

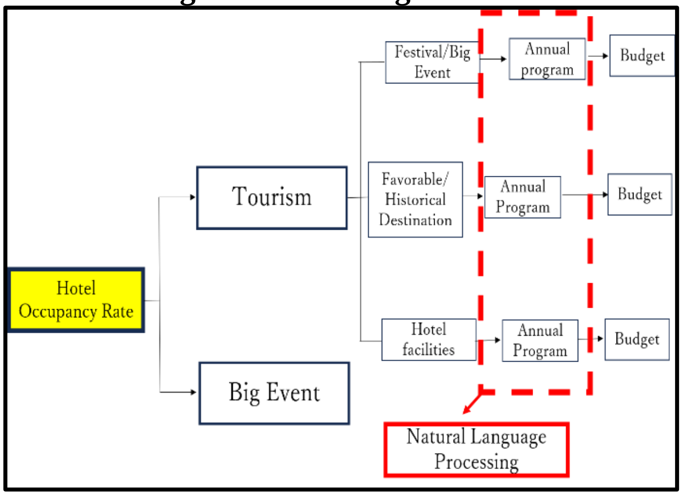
6.2 5.1 The Usefulness of Predicting Revenue

The AI model described here has the potential to revolutionize revenue forecasting and financial planning for local governments. It can be seamlessly integrated into existing systems or utilized as a standalone tool. Its primary objective is to assist local governments in making well-informed budgetary decisions. One of the significant advantages of this AI model is its flexibility in handling different scenarios. For example, the model can take inputs such as the adjustment value of the hotel occupancy variable and use it to predict the expected revenue (see Figure 6). This information is invaluable for financial planning as it allows the government to set realistic revenue targets and allocate resources accordingly. Furthermore, the AI model considers the impact of various policies and external factors on revenue forecasts. By incorporating economic trends, demographic changes, and policy shifts, the model can adapt its predictions to reflect the changing landscape more accurately. For instance, if a new policy is introduced that influences local tax revenues, the AI system can swiftly adjust its predictions to accommodate these changes. To enhance the accuracy of its forecasts, the AI model leverages references from other local governments with similar demographics and economic conditions. By analyzing data from these comparable regions, the model gains valuable insights into revenue trends and patterns. This benchmarking approach enables the local government to identify strategic programs that have been successful elsewhere and apply them to their own region, potentially leading to an increase in revenue.

Overall, the AI model outperforms basic forecasting models commonly used in previous studies. Its ability to consider a wide range of variables, adapt to policy changes, and learn from comparable regions sets it apart as a sophisticated and highly valuable tool for revenue forecasting and financial planning.

6.3 The Limitations of my model and Future Projections

Figure 6. The Usage of NLP



The main limitation of this research is mainly focused on numerical data while AI also can analyze language or characteristical data. There is a machine learning methods such as Natural Language Processing (NLP) that can be used to identify any policies related to all variables that can significantly increase local government revenue (projectpro, 2023). In the future, It would be better to investigate the model that covered not only predicting local government revenue but also a model that can identify a specific program that has the highest contributions not only to the local government revenue but also to regional development as well (see Figure 6). This is an important step because sometimes a program has less or poor economic value but a greater social impact. In this term, a model with more hidden layers and complex computations is significantly needed. Secondly, data limitations. Due of time limitations, this paper only uses 20 out of 34 Provinces' data. It means that information from the remaining provinces is not included in the training dataset. Whereas, larger datasets provide more diverse examples and patterns, which help the neural network to learn better representations and make more accurate predictions. By expanding the dataset to include information from all 34 provinces, the neural network can potentially improve its performance. More data means more examples and more variation in the data, enabling the neural network to learn more comprehensive patterns and relationships.

To conclude, AI is a powerful tool that can augment human capabilities and streamline various processes. By leveraging AI technologies, humans can enhance their productivity, efficiency, and problem-solving abilities. Those who comprehend the intricacies of AI algorithms, machine learning, and data analysis will have a competitive advantage, especially in terms of revenue generation by predicting its potential sources ahead. As AI continues to evolve and permeate various industries, individuals who possess the ability to harness its potential will be sought after. The key lies in recognizing that AI is a tool to be mastered and utilized by humans, rather than a replacement for human intelligence and creativity. Therefore, while AI may not entirely replace human roles, individuals who embrace and understand AI will become indispensable in driving innovation and shaping the future of work. As a reminder, individuals who possess a deep understanding of AI will gradually replace those who lack such knowledge. In a global scale, countries who has advance AI technology will always one step further than the other who don't.

References

7. Adebiyi, A. A et al. (2014). Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, 614342.
8. Bughin, et al. (2017). *Artificial Intelligence: The Next Digital Frontier?* McKinsey Global Institute.
9. Datta, A., & Datta, A. (2021). Using artificial intelligence to predict property tax revenue: A case study of New York City. *International Journal of Public Administration, 44(1)*, 46-56.
10. Davis, T. (2019, December 12). *A beginners guide to neural networks*. Retrieved from towardsdatascience.com: /a-beginners-guide-to-neural-networks-d5cf7e369a13
11. Gorecki, T., & O'Mahony, M. (2021). Artificial Intelligence in Public Services: Opportunities, Risk, and Best Practices. *OECD Working Papers on Public Governance*.
12. Hasin, M. A. A., et al. (2018). The impact of forecast accuracy on business performance: A study of Australia's mining and construction industry. *Journal of Construction Engineering and Management*, 144(10).
13. Hyndman, R.J., & Koehler, A.B. . (2006). Another Look At Measures of Forecast Accuracy. *International Journal of Forecasting*, 679-688.
14. Iqbal M. Mujtaba, Md Tanvir Sowgath. (2022). Application of artificial intelligence in desalination processes. In M. T. Iqbal M. Mujtaba, *Desalination Technologies* (pp. 541-593). Elsevier Inc.
15. Lumy, Denny George, dkk. (2018). Analisis Faktor-Faktor Yang Mempengaruhi Penerimaan Pajak Pemerintah Daerah Provinsi Sulawesi Utara [Analysis Of Factors Affecting Tax Revenue Regional Government Of North Sulawesi Province]. *Pembangunan Ekonomi dan Keuangan Daerah*, Vol.19 No.2.
16. Ministry of Finance. (2020). *Local Government Finance Statistics Yearbook.* Government of Japan.
17. Muskaan & Sarangi, Pradeepta. K . (2020). A Literature Review on Machine Learning Applications in Financial Forecasting. *Journal of Technology Management for Growing Economies*.
18. Nagyfi, R. (2018, September 4). *The differences between Artificial and Biological Neural Networks*. Retrieved from towardsdatascience.com: https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7
19. Nielsen, M. (2015). *Neural Networks and Deep Learning.* Determination Press.
20. Papadimitriou, I., & Ntatsopoulou, G. . (2021). A review of artificial intelligence applications in revenue management and pricing. *Revenue and Pricing Management*, 4-22.
21. Projectpro. (2023, 07 15). *Machine Learning (ML) vs NLP - What's the Difference?* Retrieved from www.projectpro.io: https://www.projectpro.io/article/machine-learning-vs-nlp/
22. Rafsanjani, F. A. (2015). Pengaruh Inflasi, Pengangguran, & Pendapatan Per Kapita Daerah Terhadap Penerimaan Pajak Dan Retribusi Daerah [The Effect Of Inflation, Unemployment, & Regional Per Capita Income On Regional Tax Revenue And Retributions]. *Ilmu Administrasi Universitas Brawijaya*, Vol.6 No.2.
23. Reddy, K. Sreenivas, et al. (2018). Predictive Analysis of Local Government Revenue: A Machine Learning Approach. *International Journal of Advanced Research in COmputer Science and Software Engineering*, 83-89.
24. Russel, S.J., & Norvig, P. (2010). *Artificial Intelligence : A Modern Approach.* Pretince Hall.
25. Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: a modern approach (3rd ed.).* Prentice Hall.
26. Sadegh Afzal, Behrooz M. Ziapour, Afshar Shokri, Hamid Shakibi, Behnam Sobhani. (2023). Building energy consumption prediction using multilayer perceptron neural network-assisted models;

comparison of different optimization algorithms. *Science Direct*, 282.
27. Sarker, I. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI. 2*, 160 .
28. Shakil, M.H, & Tasnia, Mashiyat. (2022). Artificial Intelligence and Tax Administration in Asia and The Pasific. *Research Gate*.
29. Smith, Robert. W. et al. . (2013). *Public Budgeting in America.* CQ Press.
30. Tahwin, M. (2013). Identifikasi Faktor-Faktor Yang Mempengaruhi Penerimaan Pajak Daerah [Identification Of Influence Factors Regional Tax Revenue]. *Buletin Studi Ekonomi*.
31. Woodford., C. (2023, May 12). *Neural networks*. Retrieved from www.explainthatstuff.com: https://www.explainthatstuff.com/introduction-to-neural-networks.html
32. Xu, L., et al. . (2020). An Improved Artificial Intelligence Prediction Model Based on Elastic Net and Gradient Boosting Machine for Financial Time Series. *IEEE Access*, 165692-165707. doi: 10.1109/access.2020.3028802.
33. Yusuf, A. (2016). Pengaruh Jumlah Penduduk, Produk Domestik Regional Bruto dan Inflasi Terhadap Pendapatan Pajak Daerah di Malang Raya [The Influence of Total Population, Gross Regional Domestic Product and Inflation on Local Tax Revenue in Malang Raya]. *Jurnal Ilmiah Mahasiswa Fakultas Ekonomi dan Bisnis*.
34. Zhang & Lu. (2019). Property Tax Forecasting Using Machine Learning Techniques: A Case Study of Beijing. *Journal of Urban Planning and Development*.