Prediction of Regional Revenue and Expenditure Budget using Autoregressive Integrated Moving Average (ARIMA)

R F Rahmat¹, A H Syaputra¹, S Faza², and D Arisandi¹

¹ Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

² Politeknik Negeri Medan, Medan, Indonesia

Email: romi.fadillah@usu.ac.id

Abstract. The Regional Revenue and Expenditure Budget (APBD in Indonesia) is a financial plan stipulated by Regional Regulations. The APBD of the North Sumatra Provincial Government is the reflection of development direction in the Province of North Sumatra since it would not be possible to execute the regional development process without being supported by adequate cash. Therefore, this study aims to predict the APBD allocation in 5 types of accounts, namely Regional Original Revenues, Balancing Funds, Other Legal Regional Revenues, Indirect Expenditures, and Direct Expenditures using time series data from 2002 to 2019. The proposed method in this study is Autoregressive Integrated Moving Average, which is a forecasting method using the time series data in the form of current and past data of the APBD. The results showed that the forecasting of the 2020 APBD value for the Regional Original Revenue account is IDR 6,989,217 Million, for the Balance Fund account is worth IDR 8045.259 Million, for Other Legitimate Income is at IDR 317,546 Million, for Indirect Shopping is worth IDR 9,308,607 Million and for Direct Shopping at IDR 5,495,597 Million.

1. Introduction

The Regional Revenue and Expenditure Budget (APBD in Indonesia) is the annual financial planning of the regional area. APBD consists of three groups, namely: Regional Revenue, Regional Expenditure, and Regional Financing. Regional Revenue is the regional government's right which is recognized as an income to the net worth, Regional Expenditure is a regional government obligation which is recognized as an expense in net worth, and Regional Financing is all revenue that needs to be paid back and/or expenditure to be received again, both on the ongoing fiscal year and in subsequent fiscal years [1].

The North Sumatra Provincial Government's APBD is a real picture of the development directions in North Sumatra. If Regional Revenue and Regional Expenditure are to increase proportionally each year, then it is a sign that the region is in the process of developing for the better. On the contrary, if the expenditure is greater than income and the numbers are worsening every year, it can be ascertained that the region is in bad condition. The reason is that it would be impossible to have a developed area with insufficient balance for development. Therefore, it is necessary to have prediction system to forecast the APBD in the coming year so that decision-makers can make more sensible policies related to the management of the upcoming APBD.
The Regional Revenue and Expenditure Budget (APBD) is the main determinant of regional development, the higher the income of an area, the more the amount of expenditure that can be spent for the infrastructure and other regional development activities. Therefore we need an approach to predict the value of regional budgets so that policymakers can have more precise objectives in planning the financial budgeting. The proposed method for this issue is the ARIMA method.

Numerous studies related to data prediction have been performed. One of which is Hafilah Hamimi (2014), by conducting research on data analysis of regional expenditure income using k-means and its predictions. In the study, clustering data was performed using the k-means algorithm, and predictions were executed using multiple linear regression. The data used in this study were the Regional Budget data of Padang City in the period 2009-2013 (5 years) and was able to produce 3 clusters of 4 types of data used. The result of the predictions was that the Padang city area should be spending APBD worth of IDR 2,352,908.8 Million in the following year.

Subsequent research was conducted by Rahayu Mayang Sari (2015). The study discussed the prediction of regional budget revenue using the k-means algorithm. It has not much different from the previous study since it was also implemented k-means for the data clustering. The clustering process in this study resulted in 2 clusters each in the regional income and expenditure data. For the regional income, the first cluster of regional revenue obtained one substantial income from 13 data; the second one obtained 12 small income. As for the expenditure, the first clusters obtained six large expenditures from 10 types of expense, and the remaining 689 clusters obtained four small expenditures from 10 types of shopping. In addition, this study was also able to predict data in the following year, using only the previous year of data.

Another research was conducted by Liu Qingjie & Wang Huachun (2015). They performed a research to predict the development of income and financial expenditure in China using ARIMA. The ARIMA model processed 64 years of data (1950-2013), and the model (2,1,6) was obtained as the best model. From the generated model, the data prediction was created based on the income date in the last five years. The results indicated that there should be an increase in allocation from 9.93% to 9.97% on China's state finances in the following year.

CJ Ketaren performed a study to predict fuel needs using ARIMA. The data were the monthly data of PT. PERTAMINA UPMS -1 Medan from January 2011 to September 2015 with three types of data consisting Premium, Kerosene, and Solar. From these data, the result showed that the best type of premium model ARIMA was (0,1,1), for kerosene data was (2,1,0), and for solar was (0,1,1). Standard errors of the data were 0.10106 for premium, 0.1283 for kerosene, and 0.1109 for solar.

Another research was carried out by Almasarweh Muhammad & Wadi S. AL (2018). The authors conducted research to predict banking stock data in Jordan using ARIMA. The data were acquired from the Amman stock market in the period 1993 to 2017. From this study, it was concluded that the best ARIMA model was (1,1,2) with a standard error of 1.4. Besides ARIMA there is other methods also suitable for this kind of problem such as DENFIS, or ANFIS.

2. Methodology

The proposed method in predicting the APBD consists of several processes. The processes carried out in this study were as follows: Data collection was carried out from the dashboard on the Ministry of Finance website and from the North Sumatra Province Financial and Asset Management Agency, then the data will go through the pre-processing using ARIMA, after which the ARIMA model was formed. Based on the new model, a prediction was made for the following year's fund allocation. The system produced output in the form of a predicted value of the following year's financial planning. The general architecture of this study can be seen in Figure 1.
2.1. Data Acquisition

The data were in the form of North Sumatra Province's APBD (Regional Revenue and Expenditure Budget) data from the period of 2002 to 2019. They were acquired from the Ministry of Finance data recap via their website and from the Regional Financial and Asset Management Agency of North Sumatra Province. The data used were in .csv format.

2.2. Pre-processing

Pre-processing is a stage in image processing to validate the data by performing data cleaning and selection.

Data cleaning is a process of checking data and deleting irrelevant and inconsistent data to increase the accuracy in the formation of data prediction models. The raw data can be seen in Figure 2.
In this research, due to the process of forming the model and prediction performed at account level 2, then the field—field, which is a derivative of account level 2, will be deleted, such as field _749 regional income, regional retribution, and many more. The final result of this process in the form of recapitulation of North Sumatra Province APBD data is shown in Table 1.

| Year  | Local Original Revenue | Balancing Funds | Others Legitimate Income | Indirect Shopping | Direct Shopping |
|-------|------------------------|-----------------|--------------------------|------------------|----------------|
| 2002  | 440591                 | 366644          | 17760                    | 616423           | 355852         |
| 2003  | 621017                 | 424955          | 0                        | 517237           | 644795         |
| 2004  | 874268                 | 485666          | 0                        | 577225           | 863013         |
| 2005  | 1026891                | 475583          | 0                        | 726859           | 919417         |
| Etc   | ...                    |                 |                          |                  |                |

Data selection is the process of selecting data from the processed dataset. This process aims to ease the prediction process since the data have been sorted. The results of the data selection can be seen in Table 2.

| Year  | Local Original Revenue |
|-------|------------------------|
| 2002  | 440591                 |
| 2003  | 621017                 |
| 2004  | 874268                 |
| 2005  | 1026891                |
| 2006  | 1377138                |
| 2007  | 1502955                |
| 2008  | 1897497                |
| 2009  | 2104203                |
| 2010  | 2226498                |
| 2011  | 3181900                |
| 2012  | 4026427                |
| 2013  | 4809369                |
| 2014  | 4944502                |
| 2015  | 5257837                |
| 2016  | 4630468                |
| 2017  | 4925627                |
| 2018  | 5732425                |
| 2019  | 7583849                |

2.3. ARIMA model
ARIMA has parameters in the form of 3 integers, where these parameters will then become a model of ARIMA itself. Therefore the ARIMA model is denoted by the ARIMA notation of (p, d, q).

p is part of ARIMA's autoregressive, which allows the model to include past data patterns to predict future values. For example, in this study, PAD data from 2002 to 2009 continued to increase, then it was more likely the value of PAD in 2010 to also increasing. The general form of the AR(p) model is as follows:

\[ X_t = \theta_1 X_{t-1} + \cdots + \theta_p X_{t-p} + \epsilon_t \]

Where:
- \( X_t \): variable value at the t-time
- \( X_{t-1}, \ldots, X_{t-p} \): past values of the determined time series at time of t, t-1, t-2, ..., tp
- \( \theta_1, \ldots, \theta_p \): regression coefficient, i: 1, 2, 3, ..., p
- \( \epsilon_t \): error value at t-time

\[ d \] is the integrated part which will determine the number of differences. For example, if the number of values in the previous three years has a small difference, then d will assume the value in the following year will be the same as the previous year.

q is part of Moving Average, which will calculate the average movement in the data. The model of q can be declared as follows:

\[ X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \]

Where:
- \( X_t \): variable value at the t-time
- \( \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q} \): error value at time t, t-1, t-2, .., tq
- \( \theta_1, \ldots, \theta_q \): regression coefficient, i: 1, 2, 3, ..., p
- \( q \): order of MA

One good indicator of a model ARIMA is when the model has a small value of AIC, so in this study, authors excluded all possible combinations of p, d, q and chose the combination with the smallest value of AIC. The sample of determining the value of AIC can be seen in Figure 3.

![Figure 3. List of p, d, q along with its AIC values](image-url)
2.4. Model
The obtained model was then displayed in the form of a chart that is juxtaposed with the actual data graph to see the comparison of the generated data.

2.5. VAR Model Testing
In this study, the model testing was performed in 2 ways, one of which is to use the library calculation of python to get the value of standard error against the resulting prediction, as shown in Figure 4.

![Figure 4. Result of standard error](image)

2.6. Prediction Calculation
At this stage, a prediction was calculated to determine the following year data for each type of account, as per the generated model. The prediction calculation used the following equation.

\[ X_{t+1} = X_t + X_{t-11}X_{t-12} + e_{t+1} - \Theta_1 e_t - \Theta_1 e_{t-11} + \Theta_1 \Theta_1 e_{t-12} \]

3. Result and Discussion
The model creation was performed in 1616 looping of the 16679 model of ARIMA 06689 (p, d, q) to find the smallest value of 06999 AIC. Table 3 shows the model for each account type along with the AIC value.

| Table 3. Models and AIC |
|-------------------------|
| **Local Original Revenue** | **Balancing Funds** | **Others Legitimate Income** | **Indirect Shopping** | **Direct Shopping** |
| ARIMA Model | 1, 3, 0 | 3, 3, 1 | 0, 2, 1 | 3, 3, 2 | 1, 3, 1 |
| AIC value | 442,278 | 474,676 | 491,46 | 459,625 | 444,639 |

The tests were carried out based on the obtained model for each account type to predict the future value of each account. The test also calculated the standard error of each prediction result, where the smaller the standard error, the better the prediction will be. The generated prediction result, along with its standard error values for each account is as follows:

3.1. Test result of local revenue
From the model ARIMA (1, 3, 0) obtained a predictive value of IDR 10,102,217 Million, with a standard error _666 of 6.61e -15, or around 0.00000000000000661, as shown in Figure 5.
3.2. Test result of balanced fund

Based on the ARIMA model of (3, 3, 1), a prediction value of balance fund worth IDR 8,860,877 Million, with standard error 666 of 8.49e-12, or around 0.0000000000000849 as shown in Figure 6.

3.3. Test result of other legitimate revenue

From the model ARIMA (3, 3, 1) obtained, a predictive value of IDR 10,102,217 Million with a standard error 666 of 6.61e-15, or around 0.00000000000000661, as shown in Figure 7.
3.4. Test result of indirect shopping
   As per the ARIMA model of (3, 3, 2), the indirect shopping prediction was worth IDR 10,102,217 Million with the standard error \_666 of 6.61e -15, or around 0.0000000000000000661 as shown in Figure 8.

![Figure 8. Standard error of indirect shopping](image)

3.5. Test result of direct shopping
   According to the ARIMA model of (3, 3, 1), the amount of direct shopping allocation should be at IDR 10,102,597 Million with the standard error \_666 of 6.61e -12 or around 0.0000000000000000661 as shown in Figure 9.

![Figure 9. Standard error of direct shopping](image)

4. Conclusion
   Based on the test results, The ARIMA method was proven to be able to predict the future value of income and expenditure accounts in the North Sumatra Province Regional Budget. The prediction results were based on the ARIMA model with the smallest value AIC and could produce good predictive values with small standard errors.

References
[1] Government of the Republic of Indonesia. 2006 Minister of Home Affairs Regulation No. 13 of 2006 concerning Guidelines for Regional Financial Management.
[2] Hamimi H, 2014 Analysis of Regional Expenditure Budget Data Using Clustering K-Means and Forecasting. Padang: Universitas Negeri Padang.
[3] Sari R M 2015 Local Budget Revenue Prediction Data Using K-Means Algorithm. Riau: STMIK AMIK Riau.
[4] Liu Q and Huachun W, 2015 Research on the Forecast and Development of China's Public Fiscal Revenue Based on the ARIMA Model. *Theoretical Economics Letters*, Vol 5:482-493.

[5] Ketaren C J 2017 Prediction of Fuel Needs Using the ARIMA Method in Pertamina UPMS-1 Medan. Medan: Universitas Sumatera Utara.

[6] Almasarweh M and Alwadi S 2018 ARIMA Model in Predicting Banking Stock Market Data. *Modern Applied Science*.

[7] Rahmat R F, Nurmawan S, Sembiring S, Syahputra M F, Fadli 2018 Adaptive neuro-fuzzy inference system for forecasting rubber milk production *IOP Conference Series: Materials Science and Engineering* 308(1),012014