

# 5<sup>th</sup> ICITB

---

## Forecasting The Number Of Customers Using The Box-Jenkins (Arima) And Linear Regressionmethod (Case Study: PT. AIA Financial LPG Sunrise Agency)

Amalyanda Azhari<sup>1</sup> dan Sri Lestari<sup>2</sup>

<sup>12</sup>Magister Teknik Informatika Postgraduate Programs

Institut Informatika & Bisnis Darmajaya, Bandar Lampung, Indonesia

amalyanda.1721211003@mail.darmajaya.ac.id , srilestari@darmajaya.ac.id

### ABSTRACT

Forecasting is one of the important inputs for a company in the investment decision-making process. We realize that in the forecasting process there are often inaccuracies in the results of forecasting, but forecasting must be done because all organizations operate in an environment that contains an element of uncertainty. In addition, decision-making must still be taken which will affect the future of the company.

The method used in this research is Linear Regression and Arima. Based on these studies it was found that this method is suitable for predicting the number of customers with a case study of the number of customers at PT. AIA Financial LPG Sunrise Agency.

This study provides the results of forecasting the number of customers at PT. AIA Financial LPG Sunrise Agency and produce data that can be used as a reference for what needs to be done in the future so that the policies taken by the company can be better for the company in the future.

Keywords: Forecasting, PT.AIA Financial LPG Sunrise Agency, Linear Regression, Arima

### 1. INTRODUCTION

Forecasting is one of the important inputs for a company in the investment decision making process. We realize that in the forecasting process there are often inaccuracies in the results of forecasting, but why do forecasting still need to be done? The answer is that all organizations operate in an environment that contains an element of uncertainty, but decisions must still be taken which will affect the company's future. A scientific prediction of the future will be far more meaningful than a prediction only relies on intuition.

# 5<sup>th</sup> ICITB

PT. AIA FINANCIAL also requires the forecasting process, this is because PT. AIA FINANCIAL is one of the leading life insurance companies in Indonesia and is a life insurance company registered with and supervised by the Financial Services Authority. Insurance is an attempt to help among a number of people through investments in the form of assets that provide a pattern of return to face certain risks through an agreement that is agreed upon (Ridwan, 2015). Insurance is a way to collect funds from the public in the form of premiums and in return each participant is entitled to receive payment of a certain amount of funds in the event of a particular event or disaster. In principle, insurance is engaged in services, which sell its products to the public. Insurance is also a solution to family problems in meeting their needs (Mulyana, 2015).

PT. AIA FINANCIAL as one of the leading companies has an increasing number of customers every week. This is because it is supported by more than 40,000 network of sales force and serves more than 350,000 customers spread across 6 marketing offices and 110 agency offices throughout Indonesia. (Holmogren, 2015). So far, customer data is only left as a document. While changes in the number of customers can be taken as forecasting data (forecasting) the number of customers for the future. Peramalan customer data can be useful for the company and also makes it easier for the company PT. AIA FINANCIAL in the decision making process and making future plans.

Good forecasting is certainly supported by a variety of methods, such as conducted by Ikrima Nailul Sari, 2016 which uses the LINEAR REGRESSION and ARIMA methods to find out the factors that influence customers choosing Muamalat bank. Based on these studies it was found that this method is suitable for predicting the number of customers. For this reason, this study tries to use a similar method for forecasting the number of customers at PT. AIA FINANCIAL LPG Sunrise Agency which will occur in 2020.

## 2. LITERATURE REVIEW

| No | Researcher                    | Method                                                             | Data period                       | Research and Results<br>(Strength AND weakness)                                                                                                                                        |
|----|-------------------------------|--------------------------------------------------------------------|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | AgusSartono,<br>Firdaus(2019) | <i>Simple Moving<br/>Average<br/>danWeighted<br/>MovingAverage</i> | Periods of 10, 15<br>and 20 weeks | Testing the difference in average<br>stock returns with technical analysis<br>methods. The test results do not give<br>statistically significantly different<br>average stock returns. |

# 5<sup>th</sup> ICITB

|   |                           |                                                       |                                                                |                                                                                                                                                                                                                                                                                                                                             |
|---|---------------------------|-------------------------------------------------------|----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2 | AgusSabardi<br>(2010)     | MACD ( <i>moving average convergence divergence</i> ) | Daily data for 6 months                                        | Research on trading strategies that utilize buy and sell signals. The results showed that the intersection of the MACD lines on 49 shares on the JSE under study were all accurate. Although MACD is not the only most appropriate indicator, but the use of MACD with a combination of other indicators will guarantee investors a profit. |
| 3 | Sri Mulyono<br>(2015)     | MetodeBox-Jenkins<br>(ARIMA)                          | Daily data for January 3-March 31 period 2015                  | Research on JCI short-term forecasting on the JSE. The results showed that the Box-Jenkins method is simple, fast, and inexpensive because it only requires data on previous variables to forecast. This model is also suitable for short-term forecasting.                                                                                 |
| 4 | RewatWongkaroon<br>(2012) | ARIMA,Random walk theory                              | Data for the period 2010 - 2011                                | Test the efficiency of the ARIMA model in predicting the movement of the SET50 index prices in Thailand. The results showed the ARIMA model was only more accurate in predicting the SET50 index in years 2011                                                                                                                              |
| 5 | JokoSangaji<br>(2018)     | <i>Autoregressive Moving Average</i> (ARMA)           | Daily data for the 14 December 2016 - 5 September 2017 period. | Research on the forecasting of PT Telkom's stock prices. The results showed that the ARMA model was fit fit for forecasting.                                                                                                                                                                                                                |

# 5<sup>th</sup> ICITB

---

|   |                                           |                                                         |                                                                    |                                                                                                                                                                                           |
|---|-------------------------------------------|---------------------------------------------------------|--------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6 | AchmadYani<br>(2019)                      | MetodeBox-<br>Jenkins<br>(ARIMA)                        | Period 2 daily<br>data<br>January 2018 -<br>December 30,<br>2003   | IHSG Forecasting on the Jakarta<br>Stock Exchange. The results showed<br>the ARIMA model was suitable for<br>short-term forecasting                                                       |
| 7 | NachrowiDjal<br>al,HardiusUs<br>man(2017) | MetodeBox-<br>Jenkins<br>ARIMA) and<br>GARCH<br>methods | Daily data for<br>the period 3<br>January 2015 - 2<br>January 2006 | Predict the JCI movement on the<br>JSX with several approaches and<br>then compare the predictive power.<br>The results showed that ARIMA<br>had smaller errors than the GARCH<br>method. |

In the literature review above, forecasting using ARIMA has good accuracy because the level of accuracy is above 80%.

However, we will use one more method, namely LINEAR REGRESSION to compare more accurately where ARIMA is with LINEAR REGRESSION so that research will get a really accurate forecasting between 2 methods.

### 3. RESEARCH METHODS

#### 3.1 ARIMA (Autoregressive Integrated Moving average)

ARIMA method is a forecasting method that does not use theory or influence between variables as in the regression model; thus the ARIMA method does not require an explanation of which variables are dependent and independent [penden. This method does not require breaking patterns into trend, seasonal, cyclical or irregular components as in time series data in general. This method purely predicts based only on historical data. It is almost impossible to implement ARIMA manually. Besides known by the name of ARIMA, this method is popularly known as

# 5<sup>th</sup> ICITB

---

the Box-Jenkins method, because it was developed by two US statisticians, namely G.E.P Box and G. Jenkins in 1970. (Santoso, 2009, p. 152)

This chapter introduces a class of models that can produce accurate forecasts based on a description of historical patterns in the data. Autoregressive integrated moving average (ARIMA) models are a class of linear models that is capable of representing stationary as well as non-stationary time series. Recall that stationary process, vary about fixed level and non-stationary process have no natural constant mean level. ARIMA models do not involve independent variable in their construction. Rather, they make use of the information in the series itself to generate forecasts. For example an ARIMA model for monthly sales would project the historical sales pattern to produce a forecast of next month's sales. ARIMA models rely heavily on autocorrelation patterns in the data.

This chapter introduces models that can produce accurate forecasting based on descriptions of past data patterns in the data. The autoregressive integrated moving average (ARIMA) model is a class of linear models that are capable of processing stationary and non-stationary time series data. Keep in mind that stationary processes depend on a fixed level and non-stationary processes do not have a constant natural average level. The ARIMA model does not involve independent variables in its processing. Instead, they use information in the series itself to produce estimates. For example the ARIMA model for monthly sales will project historical sales patterns to produce estimated sales next month. The ARIMA model relies heavily on autocorrelation patterns in the data. (Hanke&Wichern, 2003, p. 381)

Autoregressive Integrated Moving Average (ARIMA) is this method of making explicit the use of autocorrelation in time series, namely the correlation between a variable, which has a period of more than one period, with the variable itself. (Kazmier, 2005)

ARIMA is a method that produces predictions based on the synthesis of historical data patterns (Arsyad, 1995). This ARIMA completely ignores the independent variables because this model uses the present values and past values of the dependent variable to produce accurate short-term forecasting. For long-term forecasting the accuracy of forecasting will usually tend to be flat (horizontal / constant) for a fairly long period.

In making forecasting this model completely ignores the independent variables because this model uses the present value and past values of the dependent variable to produce accurate short-term forecasting. The Box-Jenkins method can only be applied, explained, or represented a series that is stationary or has been made stationary through a differencing process. Because the stationary series does not have a trend element, what we want to explain with this method is the remaining element, namely error. The linear time series model groups included in this method include: autoregressive, moving average, autoregressive-moving average, and autoregressive integrated moving average. (Administrator, 2009).

# 5<sup>th</sup> ICITB

ARIMA models have been widely used in the tourism literature, ARIMA models have been widely used in tourism literature (Claveria&Datzira, 2010).

## 3.2 Simple Linear Regression

Simple linear regression is a regression equation that describes the relationship between an independent variable (X) and an independent variable (Y), where the relationship between the two can be described as a straight line. The relationship between the two variables can be written in the form of an equation:

$$Y_i = \beta_0 + \beta_1 + \varepsilon_i \dots\dots\dots (P1)$$

Y = dependent variable, X = independent variable,  $\beta_0$  = intercept / intersection with upright axis,  $\beta_1$  = slope / gradient, errori independent and spread normal error  $N(0, \sigma^2)$   $i = 1, 2, \dots, n$ .

In reality we often cannot observe all members of the population, so we only take a sample, for example, the sample is size n and is written as  $\{(x_i, y_i), i = 1, 2, 3, \dots, n\}$ . The equation obtained is a conjecture from equation (12.1) and can be written as:

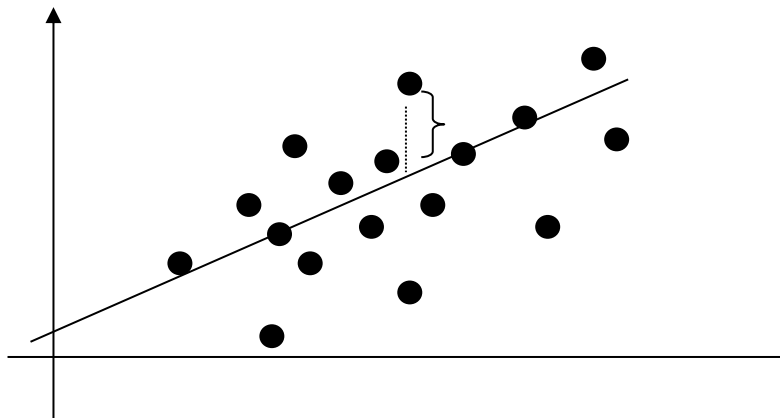
$$= b_0 + b_1 X_i \text{ (P2)}$$

$b_0$  is the estimator for  $\beta_0$ , and  $b_1$  is the estimator for  $\beta_1$ .

For the independent variable  $x_i$  the observation value  $y_i$  is not always right at the line  $= \beta_0 + \beta_1$  (population regression line) or  $= b_0 + b_1 X_i$  (sample regression line)

$$y_i \hat{Y}_i = b_0 + b_1 X_i$$

$e_i$



**picture 1. Estimating relationship line between variables X and Y**

# 5<sup>th</sup> ICITB

---

There is a deviation of  $e_i$  (for samples) or (for populations), thus

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

or

$$Y_i = b_0 + b_1 X_i + e_i \text{ (sample regression model)}$$

$$Y_i = \beta_0 + \beta_1 + e_i \text{ (population regression model)}$$

Assumptions / assumptions in simple linear regression analysis with the model

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

1)  $e_i$  is a random error that spreads normally with  $E(e_i) = 0$  and

$$\text{Var}(e_i) = \sigma^2 \text{ for all } i$$

2)  $Y_i$  spreads normally with  $E(Y_i) = \beta_0 + \beta_1 X_i$  and  $\text{Var}(Y_i) = \sigma^2$  for all  $i$

To estimate the parameter values  $\beta_0$  and  $\beta_1$  there are various methods, for example the least square method, the maximum likelihood method, the weighted least square method, etc.

Here the method used is the least squares method, because it is easy to do manually. The basic principle of the least squares method is to minimize the number of squares of deviations or Number of Squares of Errors

$$(JKG) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Using the help of calculus lessons, the estimated values of the regression parameters are obtained as follows:

$$b_0 = \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i^2 - \sum_{i=1}^n X_i \sum_{i=1}^n X_i Y_i}{n \sum_{i=1}^n X_i^2 - \left( \sum_{i=1}^n X_i \right)^2} \quad b_1 = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - \left( \sum_{i=1}^n X_i \right)^2}$$

Thus relationships can be obtained;

$$b_0 = \frac{1}{n} \left( \sum Y_i - b_1 \sum X_i \right) = \bar{Y} - b_1 \bar{X}$$

# 5<sup>th</sup> ICITB

---

The research method is a set of rules, activities, and procedures used to prepare this research. There are several steps which are carried out, which are as follows:

a. Research Place and Time

1. Place

This research will be conducted at PT. AIA FINANCIAL LPG Sunrise Agency at JalanJendralSudirman No. 57A RawaLaut Bandar Lampung.

2. Time

In carrying out the stages of research, researchers plan research time from March to January 2020.

### 3.4 Data Collection Techniques

In this study, researchers used several methods that will be used to conduct research related to data collection. Here are some of the methods used;

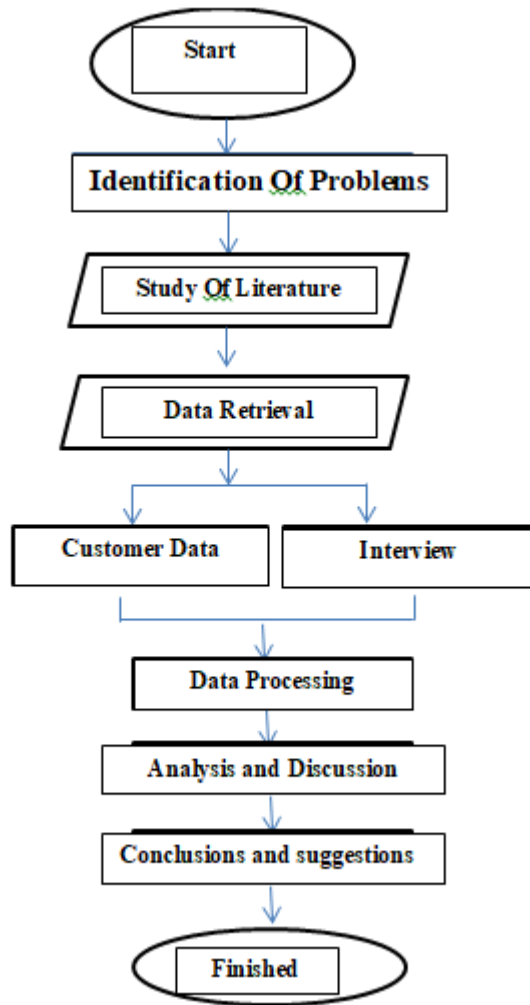
1. Field Research (Field Research)

Field studies are a method of collecting data to obtain data and information by making direct observations. The data collection and information techniques carried out at the time of field study at the point of land to be analyzed is Direct Observation, namely the collection of data by the author when observing directly on Landsat data through various media providers of Landsat data services.

2. Literature Review (Research Library)

Literature review is done by reading, quoting and making notes sourced from library materials that support and relate to research in this case about box-Jenkins.

## 3.5 Research Framework



Picture 2. Research methods

## 4. DISCOVERY AND DISCUSSION

In our research, things that need to be considered before making a forecast are forming a forecasting model for the number of insurance customers, namely as follows:

# 5<sup>th</sup> ICITB

---

## 4.1 ARIMA

The formation of forecasting models for the number of insurance customers using the Box-Jenkins method involves 4 stages, namely model identification, model parameter estimation, model verification and forecasting.

### Stage 1. Model Identification

The first step in the Box-Jenkins procedure is model identification. This stage is carried out to determine the temporary model in accordance with the data, namely by seeing the actual data plot in plain view and ACF and PACF data graphs to be more convincing that the data is not stationary, then the ACF and PACF pair test is performed

### Stage 2. Parameter Estimation

After the provisional model is obtained, then a parameter estimation is performed which aims to determine the parameters of the ARIMA model (0,1,1). Parameter estimation uses the least squares method, but to simplify the calculation the Minitab program is used. The parameter estimation results are presented in the following table:

**Table 1. Model Parameter Estimation**

| Model        | Parameter | Koefisien | <i>P</i> |
|--------------|-----------|-----------|----------|
| ARIMA(0,1,1) | MA(1)     | 0,6161    | 0,000    |
|              | konstanta | 0,9294    | 0,101    |

Based on table 3.1, it can be seen that the MA parameter (1) is significant in the ARIMA model (0,1,1). This is because these parameters have a P value that is smaller than the tolerance level (5%). While the model constants are not significant in the ARIMA model (0,1,1) because they have a P value greater than the tolerance level (5%).

### Stage 3. Model Verification

The ARIMA model (0.1, 1) that has been estimated for its parameters, will then be tested for eligibility to be used for forecasting.

### Stage 4. Forecasting

After the model is obtained, forecasting is then performed. Forecasting phase consists of training, testing and forecasting periods for the year 2018. The amount of data used for the training period is 91 data, data from the first week of January 2008 to the third week of November 2019 and for

# 5<sup>th</sup> ICITB

---

the testing period is 5 data, namely data in the fourth week of November 2019 to the fourth week of December 2019

## a. Training data

Training data forecasting is forecasting that uses actual data. Furthermore, the results of forecasting of training data will be searched using Equation 4.1 by taking the example of predicting data at times  $t = 2, 3, \dots, 91$ .

$$z_t = z_t - e_t - 0,6161 e_t - 1$$

## b. Testing Data

Forecasting data testing is forecasting without using actual data.

## 4.2 Linear Regression Model

To identify the regression model and predict the 2018 data, the data is used to analyze linear regression to determine the value of a constant and the regression coefficient  $b$  in the linear equation with  $a$  and  $b$  can be calculated by the equation

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$y$  = effect variable (Dependent)

$x$  = cause variable (Independent)

$a$  = constant

$b$  = the magnitude of the response caused by the predictor.

## 5. CONCLUSIONS

Based on the results of the problem analyst and discussion of the results of this study are:

1. Provide the results of forecasting the number of customers at PT. AIA FINANCIAL Lpg Sunrise Agency.
2. Provide a data that can be used as a reference for what to do in the future so that the policies taken by the company can be better for the company in the future.

# 5<sup>th</sup> ICITB

---

## REFERENCES

- [1] Astuti, Dwi Puji. “Volume Penjualan pada PT. Industri Sandang Nusantara Analisis Runtun Waktu”. *Tugas Akhir Mahasiswa UNNES* 2016.
- [2] Astuti, Yan. “Peramalan (*Forecasting*) Volume Penjualan Teh Hitam dengan Metode *Exponential Smoothing* pada PT.Perkebunan Tambi Wonosobo”. *Tugas Akhir Mahasiswa UNNES*. Semarang. 2015.
- [3] Arom,Dahlia.“Peramalan Komposisi Penduduk Kota Semarang menurut Jenis Kelamin Pemulusan Eksponential Ganda dari Hold”.*Tugas Akhir Mahasiswa UNNES*.2015
- [4] Djojosoedarso, Seisno. *Prinsip-Prinsip Manajemen Risiko dan Asuransi*. Salemba Empat. Jakarta. 2015.
- [5] Efendi, Riswan. *Analisa Runtun Waktu*. Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. 2015.
- [6] Hanke, John E, dan Dean W. Wichern. *Business Forecasting*. Pearson Education International, USA. 2009.
- [7] Kurniartha, Alvernia. “Keputusan Seseorang dalam Berasuransi dan Peluangnya untuk Memilih Asuransi Syariah Perbandingan Kedua Metode Time Series dan ARIMA”. *Tugas Akhir Mahasiswa UNNES*.2017.
- [8] Mulyana, Sri. “Hubungan Antara Komitmen Organisasi dan Komunikasi Interpersonal dengan Produktifitas Agen Asuransi Bumiputera 1912”. *Pascasarjana UNRI*. 2015.
- [9] RidwanS.M. *Asuransi Syariah*.PT.RichaoxIndonesia.Jakarta.2010
- [10] Sari, Ikrima Nailul. “Faktor-Faktor yang Mempengaruhi Nasabah Memilih Bank Muamalat Metode *Time Series* dan ARIMA”.*Tugas Akhir Mahasiswa UNNES*.2019.
- [11] Fitria -, S. L. (2018). METODE CASE BASED REASONING (CBR) PADA SISTEM DIAGNOSA PENYAKIT KULIT.  
<https://jurnal.darmajaya.ac.id/index.php/JurnalInformatika/article/view/1149>, 21-34.