SISFOTEK GLOBAL

Journal of Sisfotek Global

ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

Forecasting Model Number Production of Car Spare Parts at PT. Showa Katou Indonesia with Arima Method

Cici Emilia Sukmawati[1] & Ayu Ratna Juwita[2]

[1,2]University of Buana Perjuangan Karawang, Karawang, Indonesia, 41361
E-mail: [1]cici.emilia@ubpkarawang.ac.id , [2]ayurj@ubpkarawang.ac.id

ARTICLE HISTORY

Received : 9 March, 2022
Revised : 21 March, 2022
Accepted : 22 Marcj, 2022

KEYWORDS

Forecasting
PPIC
ARIMA
Data Mining.

ABSTRACT

In the case of single part production planning at PT Showa Katou Indonesia The problem is the production plan is only a production schedule, the schedule is made only two times (morning and evening) in a day. The schedule is created after the Production Planning Inventory Control (PPIC) contacts the customer to ascertain what the customer needs. After knowing what the customer needs, a production schedule and planning are made. The impact of this erratic production schedule causes loss of production time because if there is no demand then nothing is done by workers and the machine stops production because they have to wait for an erratic production schedule. Another impact is the absence of stock in the warehouse and delays in delivery because they are only waiting for the production schedule from PPIC and waiting for finished goods to be produced. To reduce the bad impact, it is necessary to forecast production planning with data mining methods to help these problems. The method used is the ARIMA method with the model (p,d,q) (1,1,1). The results of testing using tools and manual testing showed significant values with MAD = 52.45, MSE = 3917.84, MAPE = 0.05.

1. INTRODUCTION

Karawang is one of the largest industrial cities in Indonesia. Karawang has even become a city that is trusted by foreign companies to establish a business in Karawang. PT Showa Katou Indonesia is one of the Japanese companies established in this city. PT Showa Katou is a company that produces single parts. As one of the big companies, PT Showa Katou Indonesia should have a structured production schedule. But, in the Production Planning and Inventory Control (PPIC) section, there is no good arrangement for the single part production planning.

Single part production planning is only a schedule for production, where the schedule is made two times (morning and evening) in a day. The schedule is made after the PPIC contacts the customer to ensure what the customer needs. After knowing what the customer needs, then a schedule and planning for production are made. If there is a request or pre-order from a customer, a production schedule is made. But if there is no demand, then there is no production schedule.

The impact of this erratic production schedule causes lost production time because if there is no demand then nothing is done by workers and the machine stops production because they must wait for an uncertain production schedule. In addition, other impacts are the absence of stock in the warehouse and delivery delays, because they are only waiting for the production schedule from PPIC and waiting for the goods that have been produced to be finished.

The impact of this erratic production schedule causes lost production time because if there is no demand then nothing is done by workers and the machine stops production because they must wait for an uncertain production schedule. In addition, other impacts are the absence of stock in the warehouse and delivery delays, because they are only waiting for the production schedule from PPIC and waiting for the goods that have been produced to be finished. To overcome the uncertain production plan, it is necessary to forecast production planning as an alternative to reduce the adverse impacts that have been mentioned previously. One of the data mining methods that can be used to help with this problem is the forecasting method, where the purpose of forecasting is to find patterns involving variables to make predictions and classify the future behavior of an entity.

SISFOTEK GLOBAL
Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

2. THEORITICAL BASIS

2.1. Data Mining

Data Mining is a process that uses statistics, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various databases [1].

According to the Gartner Group, data mining is a process of finding meaningful relationships, patterns, and trends by examining large sets of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques [2].

The relationship sought in data mining can be a relationship between two or more in one dimension. For example, in the product dimension, you can see the link between the purchase of a product and another product. In addition, the relationship can also be seen between two or more attributes and two or more objects [3].

Figure 2.1 shows that data mining has long roots in scientific fields such as artificial intelligence, machine learning, statistics, databases, and also information retrieval.

2.2. Forecasting

Forecasting in Prasetya [4] is a form of effort to predict future conditions through testing past circumstances. Argues that forecasting is a business function that seeks to estimate the sales and use of products so that these products can be produced in a fixed quantity [5]. According to him, there are nine steps to ensure the effectiveness and efficiency of the forecasting system in demand management, there is:

1. Determine the purpose of forecasting
2. Choose the independent demand item to be forecast
3. Determine the time horizon of the forecast (short, medium, or long term)
4. Choose forecasting models
5. Obtaining the data needed for forecasting
6. Validation of forecasting models
7. Make a forecast
8. Implementation of forecasting results
9. Monitoring the reliability of forecasting results.

2.3. ARIMA

ARIMA (Autoregressive Integrated Moving Average [6] is often referred to as the Box Jenkins time series method. ARIMA is well known for time series forecasting. ARIMA is very accurate for short-term forecasting and for non-stationary time series data when linear. Meanwhile, for forecasting data that is long enough, the accuracy is not good because it usually tends to be flat. The shorter the forecasting data, the better the accuracy. In addition, ARIMA will experience a decrease in accuracy if there is a non-linear time-series component in the observation data.

The group of time series models included in this method includes Auto-Regressive, Moving Average, and Autoregressive moving average, and ARIMA:

1. Auto-regressive Method

This model was first introduced by Yule in 1926 and developed by Walker in 1931 [7], this model has the assumption that the current period data is influenced by the data in the previous period. The general form of the autoregressive model with the order p or ARIMA model (p,0,0) is expressed in equation 2.1:

$$X_t = \emptyset_1 X_{t-1} + \cdots + \emptyset_p X_{t-p} + \varepsilon_t \qquad (2.1)$$

Where:

X_t : variable value at time to- t

$X_t, X_{t-1}, X_{t-2,\ldots,} X_{t-p}$:
past value of the time series concerned at time t, t-1, t-2,..., t-p

$\emptyset_i$: regression coefficient, i: 1, 2, 3,...,p

ε_t : error value at time t

2. Moving average model

The process of moving the average of order q states the dependency relationship between the observed value Yt and the consecutive error values from the t-q period. The moving average model was first introduced by Slutzky in 1973 [8] with the order of q written MA (q) or ARIMA (0,0,q) the general form can be expressed in equation 2.2:

$$X_t = \mu' + \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-k} \qquad (2.2)$$

Where:

$\mu' = a\ constant$

$\theta_1\ until\ \theta_q\ are\ the\ moving\ average\ parameters$

$e_{t-k} = error\ value\ at\ time\ t_k$

3. ARMA models

The Ar and MA models can be combined into a model known as the Autoregressive Moving Average so that it has the assumption that the current period's data is influenced by the previous period's data and the residual value of the previous value. The ARMA model with periods p and q is written

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

ARMA (p,q) or ARIMA (p,0,q) The general model for a mixture of pure AR(1) and pure MA(1) processes, for example, ARIMA(1,0,1) is expressed in equation 2.3:

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2.3)$$

With:

X_t : variable value at time t
$\emptyset_i$: regression coefficient, i: 1,2,3,..,p
ρ : order AR
q : oerder MA
θ_i : parameter model MA ke-i, i=1,2,3,...q
ε_t : error value at time t
$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$: error at time t-1, t-2,..., t-q and
ε_t *Assuming white noise and normal.*

4. ARIMA models

The AR, MA and ARMA models use the assumption that the resulting time series data is stationary. In fact, time series data are mostly non-stationary. If the data is not stationary, then the method used to create stationary data is differencing for non-stationary data in the mean and the transformation process for non-stationary data in variance. The equations for several ARIMA models are as follows:

a. Model ARIMA (1,1,1)

$$X_t = (1 + \varphi_1)X_{t-1} - \varphi_1 X_{t-2} + \mu' + e_t - \theta_1 e_{t-1} \quad (2.4)$$

b. Model ARIMA (1,1,2)

$$X_t = \mu' + Y_{t-1} + \varphi_1 (Y_{t-1} - Y_{t-2)} - \varphi_1 e_{t-1} - \varphi_2 e_{t-2} \quad (2.5)$$

c. Model ARIMA (0,1,1)

$$X_t = \mu + X_{t-1} - \varphi_1 e_{t-1} \quad (2.6)$$

The best model selection is done by using the ARIMA model with the smallest mse resulting from the following equation:

$$MSE = \frac{\sum_{i=1}^{j} (X_t - X_t')^2}{n} \quad (2.7)$$

Where n = number of data
$X_t = data\ in\ period\ t$
$X_t' = forecasting\ data\ in\ period\ t$

The steps in time series analysis using the ARIMA model (p,d,q) or better known as the Box-Jenkins method are as follows:

1) Plot Data
The first step that must be done is to plot the original data, from the data plot it can be seen whether the data is stationary. If the data is not stationary in the mean, then the differencing process is carried out.
2) Model Identification
After the data is stationary in the mean, the next step is to look at the ACF and PACF plots. From the two data plots , several possible models can be identified that are suitable for use in time series forecasting.
3) Model Estimation
After successfully determining several possible suitable models and estimating the parameters, then a significance test is carried out on the coefficients. If the coefficient of the model is not significant then the model is not suitable for forecasting.
4) Residual Assumption Test
From several significant models, residual assumptions were tested.
5) Best Model Selection
The things that must be considered in taking the model are as follows:
   a) The principle of parsimony is that the model should be as simple as possible. In the sense that it contains as few parameters as possible so that the model is more stable.
   b) The model fulfills (at least approaches) the underlying assumptions.
   c) In the comparison of models, the model with the highest level of accuracy is always chosen, that is, the one that gives the smallest error.
6) Forecasting
The last step of the time series process is the prediction or forecasting of the model that is considered the best, and the value can be predicted for several periods in the future.

2.4. CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) was developed in 1996 [9] by analysis from several industries such as Daimler Chrysler, SPSS, and NCR. CRISP-DM provides

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

standardized data mining processes as a general problem-solving strategy of a business or research unit.

In CRISP-DM a data mining project has a life cycle that is divided into six phases. Figure 2.4. All of these sequential phases are adaptive. The next phase in the sequence depends on the output of the previous phase. Important relationships between phases are depicted by arrows. For example, if the process is in the modeling phase. Based on the behavior and characteristics of the model, the process may return to the data preparation phase for further refinement of the data or move forward to the evaluation phase.

Figure 1. Cycle CRISP-DM

DM is a standardization of data mining compiled by three data mining market initiators. From Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR. Then it was developed in various workshops (between 1997-1999) [10]. More than 300 organizations contributed to this modeling process and finally, CRISP-DM 1.0 was published in 1999 [11].

The data mining process based on CRISP-DM consists of 6 phases. That is:

1. Business Understanding or understanding the domain (research). This phase requires an understanding of the substance of the data mining activities that will be carried out, the needs from a business perspective. Its activities includes: determining business goals or objectives, understanding business situations, determining data mining objectives, and making strategic plans and research schedules.
2. Data understanding is the phase of collecting initial data, studying the data to be able to recognize the data that will be used. This phase tries to identify problems related to data quality, detects an interesting subset of data to make initial hypotheses.
3. Data preparation or data preparation. This phase is often referred to as the labor-intensive phase. Activities carried out include selecting tables and fields to be transformed into a new database for data mining materials (raw data sets).
4. Modeling is the phase of determining the data mining techniques used, determining data mining tools, data mining techniques, data mining algorithms, determining parameters with optimal values.
5. Evaluation is the interpretation phase of the data mining results shown in the modeling process in the previous phase. Evaluation is carried out in-depth to adjust the model obtained to suit the objectives to be achieved in the first phase.

Deployment or deployment is the phase of preparing reports or presentations of the knowledge gained from the evaluation of the data mining process.

3. RESEARCH METHOD

3.1. Problem Solving

The method used for problem solving is by utilizing the method of standard data mining processes as a general problem-solving strategy of the business or research unit, namely the CRISP-DM method.

The following is a step-by-step method for troubleshooting the CRISP-DM method:

1. *Business Understanding*
   By observing the place that will be used as the object of research, namely the PPIC unit of PT. Showa Katou Indonesia will then get the information needed to complete this research. The information obtained later is expected to provide information to achieve business understanding which refers to production planning at PT. Showa Katou Indonesia.
2. *Data Understanding*
   In this stage, 1,000 data were collected by taking Stratified Random Sampling samples that had been obtained from PT. Showa Katou Indonesia, which then the data will be understood further. From the data obtained, there are 11 attributes. These attributes include:

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

Customer : The name of the customer or order of a product
Delivery Date : This is the product delivery time
Invoice Date: This is the time the invoice was issued
No. Invoice : Invoice id
Part No. : Is the product id produced
Part Name : The name of the product produced
Qty : The number of products ordered
Price Idr : The unit price of goods produced
Amount IDR: The total price between the unit price multiplied by the number of quantities ordered

3. *Data* Preparation
In this stage the writer chooses the case or attribute that he wants to analyze, makes changes to several attributes. To make it easier for later modeling, the authors change the original data order into a customized order to make modeling easier. The following data arrangement of the original data is presented in table III-1 and the data arrangement that has been changed is in table III-2.

Table 1. Data with Riil Attribute

CUSTOME R	DELIVER Y DATE	INV. DAT	No Invoice	Part No.	PART NAME	Qty	PRIC E IDR	Amount IDR	Total Qty
PT FTI	01-Nov-16	01-Nov-16	982/SKI/11/201 6	50250-TSA-K010-20	UPR HALF R	360	12,059	4,341,240	**79,07 3**
PT FTI	01-Nov-16	01-Nov-16	982/SKI/11/201 6	50255-TSA-K010-20	UPR HALF L	360	12.059	4.341.240	**65,40 0**
PT FTI	01-Nov-16	01-Nov-16	982/SKI/11/201 6	50250-TSA-K010-21	LWR HALF R	360	9.820	3.535.200	**7,600**
PT FTI	01-Nov-16	01-Nov-16	982/SKI/11/201 6	50255-TSA-K010-21	LWR HALF L	360	9,820	3,535,200	**7,600**
PT YMI	01-Nov-16	01-Nov-16	1035/SKI/11/20 16	18310-K255G-6000-21	Cap B Catalyst K25G	400 0	5.683	22,732,00 0	**10,75 0**
PT YMI	01-Nov-16	01-Nov-16	1035/SKI/11/20 16	18310-K25A-9000-35	CAP A, CATALYS T K25A	360 0	4.235	15,246,00 0	**10,75 0**
PT SDI	02-Nov-16	02-Nov-16	974/SKI/11/201 6	17441-BZ130-SG00	Flange Exhaust	725	27,300	19,792,50 0	**13,80 0**
PT FTI	02-Nov-16	02-Nov-16	1028/SKI/11/20 16	50237-TXRA-P010-YI	BRKT, STIFF CTR STRG G/BOX	830	4,211	3,495,512	**7,957**
PT FTI	02-Nov-16	02-Nov-16	1030/SKI/11/20 16	50237-TXRA-P010-YI	BRKT, STIFF CTR STRG G/BOX	70	4.211	294.802	**450**

Table 2. *Preprocessing Data*
November 1, 2016

No	Product ID	Supplier	Qty
1	F001R	PT. FTI	1.000
2	F001L	PT. FTI	1.000
3	ST001	PT. YMI	800
4	Cat021	PT. SDI	960
5	Con011	PT. GTIM	890

4. Modeling
Choose a forecasting modeling technique using the ARIMA method to solve problems with the case of spare part production planning. The selection of the model is based on the temporary test results by comparing several time series methods that are often used. The methods compared include ARIMA, Exponential Smoothing, and Trend Analysis.

5. Evaluation
The following are the results of the temporary comparison of the three methods, namely ARIMA, Exponential Smoothing, and Trend Analysis using the Minitab 17 tools:

Table 3. Comparison of Time Series Results

Method	MAD	MSE	MAPE
ARIMA best model ARIMA (1,1,1)	215,25895	1353169319	8,86198%
Exponential	215,56976	1436553590	9,40981%

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

Smoothing			
Trend Analysis	218,69856	142565566	9,52354%

From the temporary test results, it was found that the ARIMA method was the best method, because the MAD, MSE, and MAPE values were smaller than the other 2 methods. These results are the reference for using the ARIMA method.

6. Deployment

The last stage of the CRISP-DM [12] [13] methodology is the Deployment stage. At this stage, a forecasting or prediction system prototype will be built using the python programming language that is able to provide reports for system users.

3.2. Testing

The evaluation results will be presented in tabular form by comparing the predicted value with the actual value. With MSE, MAD and MAPE.

4. Finding and Discussion

4.1. Business Understanding

1. Determine Business Objectives

PT Showa Katou Indonesia is a Japanese company established in the city of Karawang. PT Showa Katou is a company that produces single parts. As one of the big companies, PT Showa Katou Indonesia should have a structured production schedule. But in fact, in the Production Planning and Inventory Control (PPIC) section there is no good arrangement for planning the production of the single part.

Single part production planning is only a schedule for production, where the schedule is made 2 times (morning and evening) in a day. The schedule is made after the ppic party contacts the customer to ensure what the customer needs. After knowing what the customer needs then a schedule and planning for production is made. If there is a request or pre-order from the customer, a production schedule is made, and vice versa if there is no demand then there is no production schedule.

The impact of this erratic production schedule causes lost production time, because if there is no demand then nothing is done by workers and the machine stops production, because they must wait for an uncertain production schedule. In addition, other impacts are the absence of stock in the warehouse and delivery delays, because they are only waiting for the production schedule from PPIC and waiting for finished goods to be produced.

If the production of the goods ordered by the customer is completed, a payment invoice is made and then it is notified to the customer to make the payment as stated on the invoice. Then, after payment is made by the customer, the ordered goods are shipped.

2. Asses the Situation

a. To confirm the order, PPIC calls the customer to make sure what spare parts will be ordered and then produced.
b. Customers inform what products will be ordered.
c. The PPIC records the details of each order then makes a production schedule for the order and submits it to the sales department.
d. Sales department creates sales invoices for customers.
e. The customer pays for the order stated on the invoice.
f. Delivery is made after the order production process and payment by the customer have been completed.
g. Single part production planning is only a schedule for production.

3. Determine the Data Mining Goals

The purpose of data mining or the purpose of this research is to create a forecasting model for spare part production planning at PT. Showa Katou Indonesia to find out the results of the forecast some time in the future.

4.2. Data Understanding

Data understanding is carried out by collecting initial data and studying the data to be able to recognize and understand what can be done with the data. Understanding the data refers to the Excel file containing the spare part sales table at PT. Showa Katou Indonesia. The stage of understanding the surface data format (form and report format) and in more depth (physical data form).

1. Collect the Initial Data

The stage of collecting data from 2016 to 2018. In 2016 there were 2,438 records from 1,203 invoices, in 2017 there were 2,718 data from ± 1,598 invoices, and in 2018 there were 2,992 records from ± 2,341 invoices with a total of 63 products. Format

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

the data obtained in the form of excel documents as much as > 5,000 data with the data collection method Stratified Random Sampling sample that has been obtained from PT. Showa Katou Indonesia, so that the data will be understood further. From the data obtained there are 9 attributes. These attributes are: No, Customer, Delivery Date, Invoice Date, No. Invoice, Part No., QTY. Price IDR, and Amount IDR. Here are some data presented to find out what spare parts are produced (the rest is attached):

The spare parts shown in table 4 are the names of any spare parts produced.

Table 4. Names of Spare Parts produced

No	Part No.	Part Name
1	PRBDI01002	Body Inner LB
2	PRP1P10001	Exh Pipe End
3	17441-BZ130-SG00	Flange Exhaust
4	17510-BZ150-H-02	BRKT, FR R
5	17510-BZ150-H-03	BRKT, FR L
6	17531-BZ020-SG00	Cone Catalytic Lower
7	17532-BZ020-SG00	Cone catalytic upper
8	18220-T7EF-K000-W1-41	Protector B Exhaust 2CF (S/C)
9	18224-TE7A-K000-200	Inner Chamber 2SJ (S/C)
10	18308-TE7F-K000-20	Pipe A inner 2MD (S/C)
11	18308-TE7F-K000-21	REINF TORSION BEAM RH
12	18310-K255G-6000-21	Cap B Catalyst K25G
13	18310-K25A-9000-35	Cap Catalyst A Ø 25.4
14	18310-K25A-9000-35	Cap A Catalyst K25A

2. Describe the Data

The data obtained is contained in the appendix. The attributes in the data obtained are as follows:

Table 5: Attributes of the data obtained

Atribut	Type Data	Description
Customer	Char	Show customer name
Delivery Date	Date	Show delivery time
Invoice Date	Date	Shows invoice creation time
No. Invoice	Char	Show invoice number
Part. No	Char	Show spare part number or spare part id
QTY	Integer	Shows the number of items ordered
Price Idr	Integer	Shows the unit price of goods
Amount IDR	Integer	Indicates the amount of the price to be paid

These attributes are obtained from spare part sales data at PT. Showa Katou Indonesia.

3. Explore the Data

From the thousands of existing data, the 5 most ordered data were obtained by the customer, the average order value, the minimum order value, and the maximum order value were obtained.

Table 6: Average Calculation of the 5 Most Spare Parts

Part Name	Componen	Value
Half Elbow	Average	8.659,38
	Max	70.250
	Min	84
Cap Catalyst A Ø 25.4	Average	10.010,71
	Max	74.1750
	Min	100
CAP B CATALYST COMP K60A (S/C)	Average	1.086,21
	Max	4.500
	Min	63
BODY INNER A K59A (S/C)	Average	1.104,13
	Max	6.225
	Min	100
Jasa Perforating Body Inner K15A	Average	4.427,19
	Max	19.000
	Min	10

From the table above, it can be explained that the 5 most spare parts were ordered with average sales, maximum and minimum values obtained from the period 2016 to 2018 or for 36 months.

4.3. *Data Preparation*

Activities to build data sets to be incorporated into the modeling from the initial raw data. Its main function is specifically for modeling to predict a condition.

1. Select Data

The stage of selecting the data used for analysis is based on several criteria, including its relevance to data mining objectives, as well as quality and technical constraints such as limits on data volume or data types.

SISFOTEK GLOBAL

Journal of Sisfotek Global

ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

The data selected for testing on sales data or spare part production is data from January 2016 to December 2018 with a total data of ± 5,000. While the attributes in the understanding data are not all used, there are several attributes that are not influential attributes in the test, which are removed in order to maximize the test results.

2. Clean Data

The verification results at the data understanding stage still show poor quality data. The data needs to be cleaned, repaired, changed even if it needs to be deleted. Some of the data that need to be corrected are due to typing errors. While the data that needs to be deleted because the data is incomplete and produces ambiguous, so it is necessary to delete it. In this cleaning stage, all attributes in the sales data are carried out.

3. Integrate Data

The stage of combining two or more tables that have different information about the same object into a new data set that has been prepared in the early stages of data preparation. At this stage the tables that are integrated are the monthly sales report tables for the period January 2016 to December 2018, where these tables were originally separate tables in chronological order. Then the table is then used as a combined table of the data used. So that it can facilitate the process of testing and modeling by involving data mining techniques.

4.4. *Modeling*

The phase that directly involves data mining techniques. Selection of data mining techniques, algorithms and determining parameters with optimal values. The steps in modeling are as follows:

1. Select Modeling Techniques

Modeling selection is done by doing trial-error with several ARIMA orders which are presented in the following table:

Table 7 : Parameter Estimation with Trial

Model Parameter	Variable	Coef	P-Value	Decs.	MSE	MAD	MAPE
0,0,1							
0,1,0							
0,1,1	MA (1)	0.96	0.01	Significant	3871.85	56.75	0.09
1,0,0							
1,0,1							
1,1,0							
1,1,1	AR (1)	-0.44	0.01	Significant	3917.84	52.45	0.05
	MA (1)	0.98	0.00	Significant			

The ARIMA model used is the ARIMA model (0,1,1) or (1,1,1), because in several random tests and as a comparison using other models or orders, the results show an error compared to the ARIMA model [15] (0,1, 1) or (1,1,1) from forecasting tests for spare parts Cap Catalyst A 25.4 in 2018.

2. Generate Test Design Testing Data for Forecasting 2019

A. Using Minitab Program

Forecasting using time series analysis for 36 months or 3 years so that to forecast the production of spare parts Cap Catalyst A 25.4 at PT. Showa Katou Indonesia in 2019 required spare part sales data for Cap Catalyst A 25.4 PT. Showa Katou Indonesia for the period January 2016 to December 2018. The total data is 36 months. The data can be seen in table 4.8:

Table 8 : Real Sales Data 2016 to 2018

Year	Month	Sell Amount
2016	January	45.400
	February	60.000
	March	71.400
	April	45.200
	May	54.800
	June	53.041
	July	26.800
	Agustus	68.752
	September	72.200
	Oktober	78.033
	November	81.793
	December	56.770
2017	January	62.200
	February	73.500
	March	51.700
	April	73.700
	May	111.400
	June	63.800
	July	108.600
	Agustus	79.000
	September	74.200
	Oktober	106.800
	November	91.000
	December	57.600
2018	January	61.000
	February	94.000
	March	81.600
	April	89.200
	May	99.200
	June	88.200
	July	61.800
	Agustus	54.000
	September	68.000
	Oktober	75.900
	November	87.000
	December	95.600

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

Year	Month	Sell Amount
Average		72.866,36

From Table 8 it can be seen that the data obtained is monthly sales data for 36 months, in the period January 2016 to December 2018 with an average sales of 72,866.36.

1. Model Identification (Forecasting 2019)

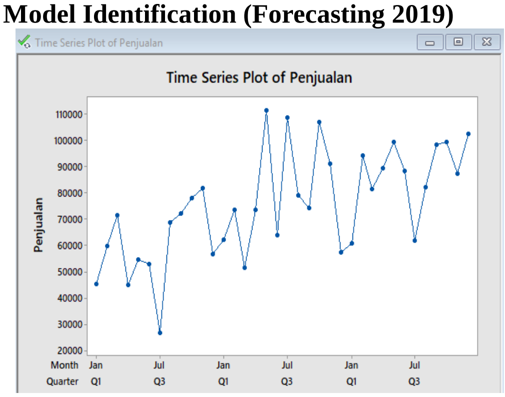

Figure 2. Graph of Spare Part Sales for Cap Catalyst A Ø 25.4 January 2016-December 2018 Period.Information

Description
Sales: Total Sales
Month: Period in months
Figure 4.2 shows the pattern of sales sales data for the period January 2016 to December 2018. From the graph, the data is not stationary with respect to the average or variance. In several months there has been a significant increase and decrease, this is influenced by customer demand.

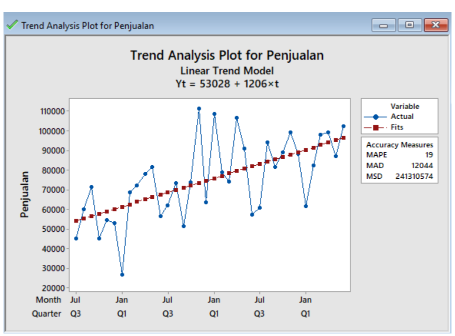

Figure 3. Original Data Trend Graph

Figure 4.3 shows a trend analysis graph of sales data for the period January 2016 to December 2017. The fits variable (red line) is trend data, while the actual variable (blue line) is fact data or real data. In Figure 4.3 between the actual variables are not in accordance with the fits variable which is trend data.

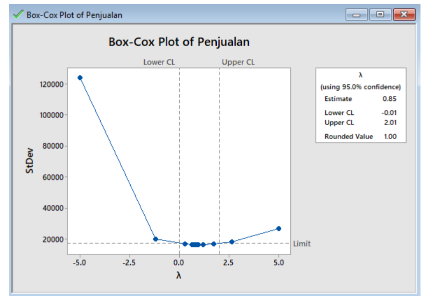

Figure 4. Stationarity Graph Against Variety

A data can be said to be stationary with respect to variance if the value of in the rounded value is 1. Figure 4.4 shows that the rounded value is already worth 1. Thus, it can be ascertained that the data is stationary with respect to variance.

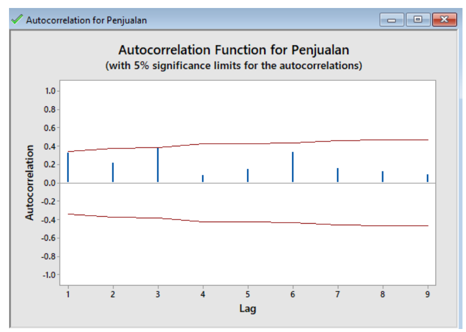

Figure 5. Graph of the Original Data Autocorrelation Function

From the graph of the autocorrelation function above, it can still be seen that the data is not stationary because there is still linearity in some lags, so it is necessary to do differencing for the first data.

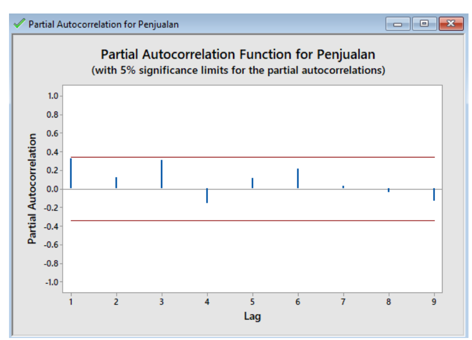

Figure 6. Graph of Original Data Partial Autocorrelation Function (2)

SISFOTEK GLOBAL
Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

Similarly, in the graph of the autocorrelation function in Figure 4.6, the graph of the partial autocorrelation function (FAKP) above still shows that the data is not stationary and needs to be differencing for the first data.

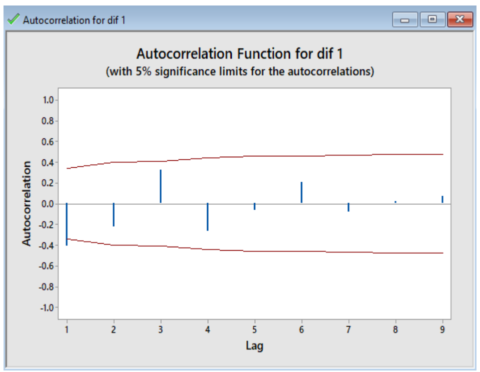

Figure 7. Graph of Autocorrelation Function After Differencing

From the FAK Graph (Autocorrelation Function) it can be seen that the data is stationary because the graph does not decrease slowly and is linear so that the model can be estimated directly.

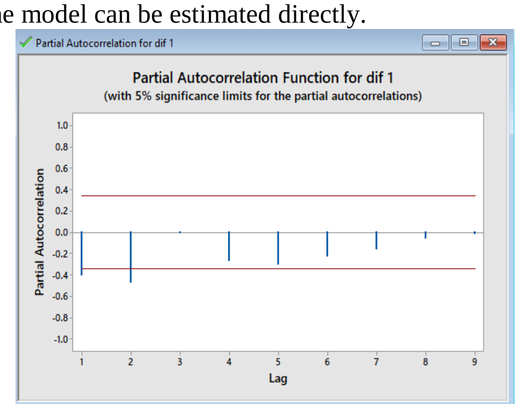

Figure 8. Graph of Partial Autocorrelation Function After Differencing

The FAKP graph looks like a graph following a sine graph. By looking at the state of the FAK and FAKP graphs, it can be seen that the value of r_k cut the white noise line on lag-1 and lag-2, whereas φ_k because the graph follows a sine graph it is not taken into account.

2. Estimation of Parameter Values in the Model (Forecasting 2019)

Based on several experiments and the proposed research proposal, the ARIMA model used is ARIMA (1,1,1).

3. Verification (Forecasting 2019)

To verify the initial model, a more parsimony model must be chosen (simple in terms of parameters). The model selected for verification of the initial model is ARIMA (1,1,1), with reference to the results of the following table:

Table 10: Parameter Estimation with Trial

Model Parameter	Variable	Coef	P-Value	Desc.	MSE	MAD	MAPE
0,0,1							
0,1,0							
0,1,1	MA (1)	0.96	0.01	Signifikan	3871.85	56.75	0.09
1,0,0							
1,0,1							
1,1,0							
1,1,1	AR (1)	-0.44	0.01	Signifikan	3917.84	52.45	0.05
	MA (1)	0.98	0.00	Signifikan			

```
ARIMA Model: dif 1

Estimates at each iteration

Iteration         SSE          Parameters
        0  49370850269    0.100  0.100     13.325
        1  35720994221   -0.050  0.249   -248.349
        2  33591772588    0.047  0.399   -230.707
        3  31305598636    0.132  0.549   -203.134
        4  28828375754    0.203  0.699   -167.018
        5  26215925119    0.259  0.849   -125.750
        6  23494801317    0.226  0.924   -113.779
        7  20310863241    0.076  0.982    -62.305
        8  17596680121   -0.074  0.980    -81.996
        9  15863404151   -0.224  0.977    -97.894
       10  15090229234   -0.374  0.972   -106.846
       11  14998113390   -0.432  0.978    -98.985
       12  14996938760   -0.441  0.977   -123.810
       13  14987642037   -0.442  0.981   -125.255
       14  14987581293   -0.442  0.980   -128.921

Unable to reduce sum of squares any further


Final Estimates of Parameters

Type          Coef   SE Coef       T       P
AR   1     -0.4424    0.1719   -2.57   0.015
MA   1      0.9799    0.1827    5.36   0.000
Constant    -128.9     254.0   -0.51   0.615
```

Figure 9. ARIMA results (1,1,1) in Minitab 2017

The results of the verification can be seen from the final estimates of parameters that the AR and

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

MA orders produce P-values of 0.015 and 0.00, respectively. While the value of = 5% with parameter P-value <. So, it means that Arima's model (1,1,1) can be accepted. Thus, the equation that can be written as Arima (1,1,1) modeling is as follows:

$$Z_t = -128.9 + (-0.4424Z_{t-1}) - 0.9799e_{t-1}$$

4. *Forecasting* 2019

```
Forecasts from period 36

                        95% Limits
Period  Forecast  Lower    Upper  Actual
    37     95903  63059   128746
    38     97289  64446   130133
    39     98446  65585   131307
    40     99609  66733   132486
    41    100773  67880   133665
    42    101936  69027   134845
```

Figure 10: Forecasting Results for the Next 6 Months 2019

The results of forecasting the production of spare parts for Cap Catalyst A 25.4 PT. Showa Katou Indonesia 2019 for the months of January to June can be seen in Figure 4.18.

Table 11: Forecasting Results for the next 6 months in 2019

Period	Production
Januari	95.903
February	97.289
March	98.446
April	99.609
May	100.773
June	101.936
Average	98.992,67

Table 4.11 explains that the results of forecasting in 2019 each month have increased, although not too significantly, with an average monthly production of Rp.

B. Results Using Manual Method

Forecasting testing using manual methods with the help of Microsoft Excel the results are as follows:

Table 12. Manual Forecasting Results for the next 6 months in 2019

Period	Forecasting Result
January	42.199,82
February	80.722,91
March	114.344,39
April	144.499,93
May	144.512,64
June	144.511,66
Average	**111.788,56**

Table 4.12 shows the results of manual calculations for forecasting for 2019 for 6 months. Forecasting results show that every month there is an increase in sales.

4.5. *Evaluation*

The interpretation phase of the data mining results. The evaluation is carried out in depth with the aim that the results at the modeling stage are in accordance with the objectives to be achieved in the business understanding stage.

Evaluate Result

Table 13. Prediction Test Results for 2019

Period	Real Production 2019	Production With Minitab	Production With Manual	Difference With Program Minitab	Different with Manual
January	62.200	95.903	42.199,82	33.703	20.000,18
February	80.100	97.289	80.722,91	17.189	-622,91
March	85.600	98.446	114.344,39	12.846	-28.734,39
April	78.900	99.609	144.499,93	20.709	-65.549,93
May	88.000	100.773	144.512,64	12.773	-56.512,64
June	90.100	101.936	144.511,66	11.836	-54.411,66
Average	**80.817**	**98.992,67**	**111.788,56**	**18.176**	**-30.971,89**

From the table above, it can be explained that the spare part tested is Cap Catalyst A 25.4 in 2017. The data used is monthly sales data for the period January 2016 to December 2018, then the data is tested in such a way using minitab 2017 tools and testing manually, so that generate predictions for sales for the next 6 months of 2019.

In the real production column, the sales results in 2019 are in fact, then compared with the test results with the 2017 Minitab tools and with manual testing.

From the comparison above, the average difference in the minitab program results is 18,176 with manuals, while the average difference in manual results is 29,338.25. The difference obtained with the minitab program is smaller than the manual method. So, the forecasting results that are quite

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

close to the real data in the field are the results of forecasting using the minitab program.

From the forecasting results of spare part sales for the period January to June 2019 the results are different from the previous year. This means that there is a significant increase in sales. This shows that the demand for spare parts production for Cap Catalyst A 25.4 is quite high every month. So there need to be anticipatory things that must be done by PT. Showa Katou Indonesia seems to provide more raw materials so that the need to produce each spare part is not lacking.

4.6. *Implementation*

1. The architectural design

The architectural design is made according to the needs that exist in the Arima application. To model Arima, it takes several main packages and other supporting packages that must be included in the coding.

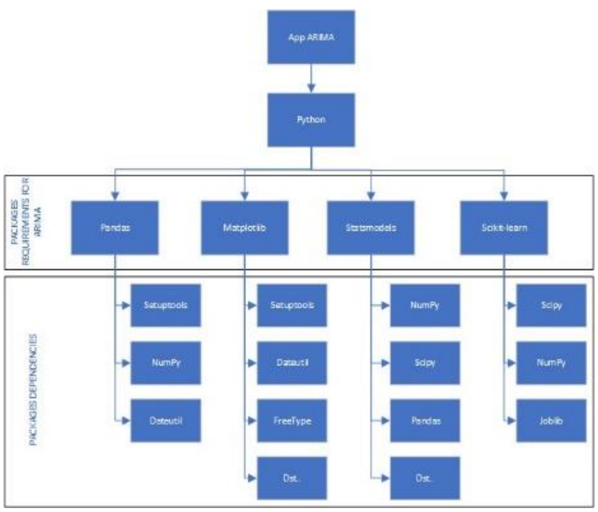

Figure 11. Arima Architectural Design with Python

From the architectural design in Figure 4.11 the packages used for Arima modeling have the following main functions:

1. Pandas, to read csv files. which will be used as a test
2. Matplotlib, to display a graph of test results
3. Statsmodel, for scipy with statistical calculation functions. This includes descriptive statistics and estimates and inferences for statistical models.
4. Scikit-learn, to do some work in data science.

2. Flowchart

Arima's modeling steps made with the python programming language when poured into a flowchart can be seen as follows:

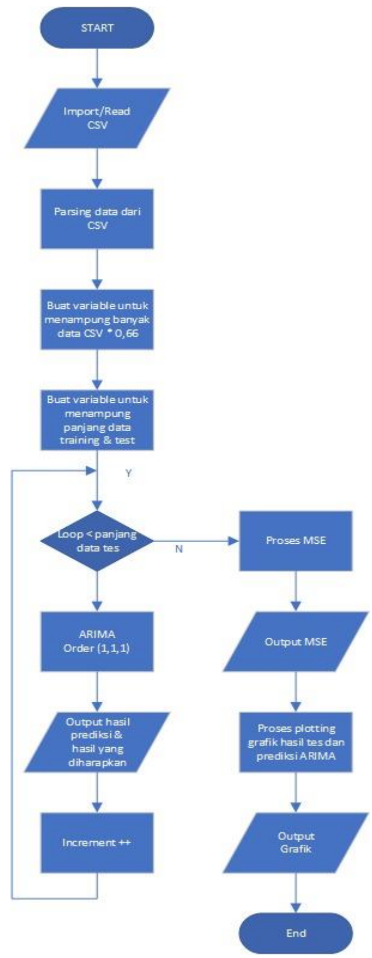

Figure 12. Arima Flowchart in Python

The workflow of an Arima application built in python is described in Figure 4.12.

After importing the .csv file, then the .csv file is parsed and a new variable is created to accommodate the large amount of csv data. Then a loop is carried out along the test data using the Arima model (p, d, q), so that the forecasting results come out. After the forecasting results come out, the MSE process is carried out, and the MSE results along with the graph come out.

Journal of Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

5. CONCLUSIONS AND SUGGESTIONS

5.1. Conclusions

In writing a thesis entitled "Forecasting Model of the Number of Production of Car Spare Parts at PT. Showa Katou Indonesia Using the ARIMA Method" the author concludes :

1. Forecasting the amount of spare part production based on sales data from 2016 to 2019 using the ARIMA model (1,1,1) method and producing a forecast with a significant amount.
2. After doing several experiments by doing trial-error, the MAD results are 52.45, MSE is 3917.84 and MAPE is 0.05% using the ARIMA (1,1,1) model.
3. From the comparison results in manual calculations and using minitab 2017, the average difference shows results that are significantly different from real production data. The difference between using the minitab program is greater than the difference using the manual method. So the forecasting results that are quite close to the factual data in the field are the results of forecasting using the manual method because the results of the difference are smaller.

   In manual calculations, the compiler takes $\theta = 0.5$, while θ lies between –1 and 1, so there is a possibility that if is different, the θ results will also be different. Therefore, the Arima model (1,1,1) when referring to the results of the minitab program is not suitable for use in this research data.

5.2. Suggestions

From the results of the research conducted, the suggestions given for further research references are :

1. From testing using tools and manuals, it can be seen that there are significant differences in results, so for further research, other better models can be used which allow for better forecasting abilities.
2. For better test results, it is recommended to use more test data.
3. Applying ARIMA modeling into a program based on a GUI (Graphical User Interface) that is user friendly, and easy to use.

REFERENCES

[1] Alwadi, S., Almasarweh, M. and Alsarairah, A. (2018) ‘Predicting Closed Price Time Series Data Using ARIMA Model’, (November). doi: 10.5539/mas.v12n11p181.

[2] Berry, M.J.A., L. (1997) *Data MiningTechniques: For Marketing, Sales, and Customer Support*. New York: John Wiley & Sons.

[3] Fayyad, U. (1996) *Advances in Knowledge Discovery and Data Mining*. MIT Press.

[4] Ilić, I., Jovanović, S. and Milić, V. J. (2016) ‘Forecasting Corn Production In Serbia Using Arima’, 564(11), pp. 1141–1156.

[5] Larose, D. T. (2005) *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.

[6] Muhammad, A. D. and Susanto, N. (2017) ‘Peramalan Perencanaan Produksi Terak Dengan Metode Exponential Smoothing With Trend Pada Pt . Semen Indonesia (Persero) Tbk.’, pp. 1–10.

[7] Nofiyanto, A., Adi Nogroho, R. and Kartini, D. (2015) ‘Peramalan Permintaan Paving Blok dengan Metode ARIMA’, pp. 9–10.

[8] Oktarina, T. (2018) ‘Peramalan Produksi Crude Palm Oil (Cpo) Menggunakan Metode Arima Pada Pt . Sampoerna’, (November).

[9] Ponniah, P. (2001) *Data Warehouse Fundamentals : a Comprehensive Guide for IT Professional*. New York: John Wiley & Sons.

[10] Turban, Efraim, et al (2005) *Decision Support Systems and Intelligent Systems*. 7th Ed. New Jersey: Pearson Education.

[12] S. Wardah, and I. Iskandar, "ANALISIS PERAMALAN PENJUALAN PRODUK KERIPIK PISANG KEMASAN BUNGKUS (Studi Kasus : Home Industry Arwana Food Tembilahan)," J@ti Undip: Jurnal Teknik Industri, vol. 11, no. 3, pp. 135-142, Jan. 2017. https://doi.org/10.14710/jati.11.3.135-142.

[13] C. Schröer. 2021. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science Volume 181, Pages 526-534.

ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v12i1.478
Vol. 12, No. 1, March 2022, pp. 65-78

[14] J. Fattah, et.all. 2018. *Forecasting of demand using ARIMA model.* International Journal of Engineering, Business Management, Volume 10: 1–9