Visualization of Forecasting Number of Patients Visiting in Pratama Clinic Using ARIMA Method

Mardiani Trisno Putri1*, Dedy Sugiarto2, Teddy Siswanto3

Information System Study123, Faculty of Industrial Technology, University of Trisakti

Abstract-The Primary Clinic of the Ministry of Communication and Information of the Republic of Indonesia had a problem that is difficulty in predicting the number of patient visits resulting in a buildup of patients in the clinic. For this reason, it is necessary to forecast patient visits in order to be the basis for planning for the next period and become a material consideration for decisions in clinical development. The research method was conducted using ARIMA by utilizing historical data on clinic visits from 2014-2019. The processed results in RStudio in the form of prediction data for patient visits in 2020-2021 are visualized on a website that uses the PHP Codeigniter framework and is integrated with RStudio. The results of visit forecasting based on services using the ARIMA method have good forecasting abilities, because the MAPE values are in the range of <10% and 10-20%, namely 19.58% and 9.62%.

Intex Terms- Forecasting, ARIMA, MAPE

I. INTRODUCTION

The Primary Clinic, which is supervised by the Ministry of Communication and Information of the Republic of Indonesia, provides medical services including general polyclinic and dental polyclinic for the state civil apparatus which is run by several doctors. Along with the increase in the number of state civil servants and their dependents and increased awareness about health, there is usually aggregate data analysis in a database context. In an OLAP database, there is usually aggregated, historical data, and is usually stored in a multi-dimensional scheme. OLAP is a technology for analytic purposes. OLAP has the ability to do data analysis in a database context. In an OLAP database, there is usually aggregated, historical data, and is usually stored in a multi-dimensional scheme. OLAP is a database used to maximize reading speed (select query). The difference seen between OLAP and OLTP is in the database schema. If OLTP is normalized, while the OLAP database scheme uses a star or snowflake schema.

Visualization of the number of patients visiting in the Primary Clinic is ARIMA. The ARIMA method has several advantages including being able to be applied to predict simple data and the application of methods which are not difficult in reviewing data that have seasonal or trend patterns, has a character that follows data patterns (flexible), has a fairly high degree of accuracy and is suitable when used to predict several variables quickly and cheaply because it only requires historical data to make predictions. So that the method is expected to make accurate prediction data so that it can meet the demand of the Primary Clinic of the Ministry of Communication and Information of the Republic of Indonesia in expanding the clinic.

Previous studies related to forecasting the number of patient visits have been conducted, namely the Forecasting of the Number of General Poly Patients in RSUD Pantura MA Sentot Using the Arima Method. Then the implementation of the Fuzzy Neural Network is to Estimate the Number of Surgical Poly Patient Visits in Surabaya Oncology Hospital. (Anggraeni, 2012) As well as the comparison of the Single Exponential Smoothing and ARIMA methods in predicting drug needs in Indramayu Regional Hospital, the prediction results using the ARIMA method are more accurate than the Single Exponential Smoothing method. (Zahra) Based on the problems and needs mentioned above, visualization is needed to predict the number of visitors in the Primary Clinic of the Ministry of Communication and Information of the Republic of Indonesia next period using the ARIMA method.

II LITERATURE STUDY

A. OLAP

OLAP or commonly known as Online Analytical Processing is a technology for analytic purposes. OLAP has the ability to do data analysis in a database context. In an OLAP database, there is usually aggregated, historical data, and is usually stored in a multi-dimensional scheme. OLAP is a database used to maximize reading speed (select query). The difference seen between OLAP and OLTP is in the database schema. If OLTP is normalized, while the OLAP database scheme uses a star or snowflake schema.
B. FORECASTING

Forecasting or forecasting is the activity of predicting, projecting, and or making an estimate of the various possibilities that will occur before more accurate planning can be done. Forecasting capture historical data and make predictions in the next period by utilizing a number of mathematical models. Forecasting methods consist of two types namely qualitative methods used if they do not have historical data so mathematical predictions cannot be made. While quantitative methods are used for forecasting if they have historical data. (Subagyo, 2008)

C. ARIMA METHOD

ARIMA or commonly known as the Box-Jenkins Periodic Series is a method that can be used to predict time series data (Time Guide Analysis) in the form of quantitative calculating time, by collecting data periodically based on an arrangement to ensure past data patterns are collected regularly, then the ARIMA method will predict based on a statistical understanding that develops to get data series patterns which are then exploited in the future. The implementation of AR, MA, and ARMA models shows that the data already have stationary properties. But sometimes there is often no stationary data. For the data to be stationary a differencing process must be run. Models that have stationary data after going through the differencing stages are called ARIMA models. So if the data is stationary at differencing d times and implements ARMA (p, q), then the model that is realized is ARIMA (p, d, q). (Amira Herwindyani Hutausuhut, 2014)

D. Data Visualization

Data visualization is the latest communication way to convey information by displaying it in graphic form. Visualization in the form of images is easier to understand than in the form of writing. The purpose of data visualization is to use graphical tools to convey information clearly, effectively, and easily. Effective data visualization can support the user to interpret and analyze evidence and data. (Kirk, 2012)

E. CODEIGNITER FRAMEWORK

Framework is a set of instructions or basic functions that build certain provisions and are interconnected so that when building a website application must follow the provisions of the framework. The advantage of using a framework is that the time to build website applications is faster; the website code is easier to understand because it is not many and is essential; no need to build supporting code such as form validation, database interface, security, and GUI; can focus more on the website application problem code regarding the appearance and services that will be displayed on the website. (Wardana)

Codeigniter (CI) is a type of website application framework that can be accessed for free, easy, and lightweight. The main feature of Codeigniter is that it provides convenience and shortens the time to build programs in PHP and manages the layout of the site and limits structural problems. But some things can't be done with Codeigniter, that is, CI doesn't have all the functions available to handle all the problems encountered because of the size of the light and small Codeigniter.

F. MEASUREMENT OF PREDICTION ACCURACY

I. 1. RMSE

RMSE is the result of rooting from MSE. This measurement model is used to get accurate forecasting results based on historical data. The smaller the value of the results, the better the results of the forecasting run.

II. 2. MAPE (MEAN ABSOLUTE PERCENTAGE ERROR)

MAPE is a way of measuring by showing the accuracy of forecasting in a percentage format. Calculations with MAPE are often used because they are easy to read. The smaller the MAPE value, it is concluded that the forecasting model used has good ability. Below is Table 2.2, which shows the range value for MAPE. (P.C. Chang, 2007)

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\%
\]

Note: n = amount of data, xt = actual result value, and yt = predictive value yet

<table>
<thead>
<tr>
<th>MAPE</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10%</td>
<td>Very good predictive ability</td>
</tr>
<tr>
<td>10-20%</td>
<td>Good predictive ability</td>
</tr>
<tr>
<td>20-50%</td>
<td>Decent / adequate predictive ability</td>
</tr>
<tr>
<td>&gt; 50%</td>
<td>Bad predictive ability</td>
</tr>
</tbody>
</table>

III RESEARCH METHODOLOGY

The method in predicting patient visit data starts from preparing raw data in the form of a spreadsheet file. Then the data cleansing stage is carried out which serves to determine the data needed by those not, and to tidy up the structured data format. Next, designing the table layout and its attributes, the fact tables that have been successfully created are then given a connection with the dimension tables so that a schema can be formed. These tables include: service dimensions, doctor dimensions, status dimensions, time dimensions, visit facts, and visit facts based on status. In this design also must identify the dimensions by designing dimensional models. The chosen dimensional model is the galaxy schema or fact constellation schema because the fact tables in this study are twofold and share dimension tables. Furthermore, the data is categorized and inputted to the database by the Extract, Transform, Load (ETL) process using the Spoon Pentaho Data Integration tools to form an OLAP database. The R Programming Language and RStudio tools are used to make predictions of patient visit data at the Primary Clinic from 2020 to 2021 using the time series method. Forecasting technique that is used in the time series method is the ARIMA method. Furthermore, if you have succeeded in forecasting with the ARIMA method in RStudio, then the next step is the visualization of graphical data from forecasting by utilizing the PHP Codeigniter Web
Framework that has been integrated with the RStudio application. The last step is to analyze the results of forecasting data using the ARIMA method by looking at the results of the error rate figures or Root Mean Squared Error (RMSE) and MAPE on the forecasting model used to obtain the lowest possible error results.

IV RESULTS AND DISCUSSION

The prediction data using the ARIMA method can be visualized in the form of a chart of forecasting patient visit data based on service, type of patient status, doctor, and period of visit. The design begins by designing an OLAP database schema model using ETL process which in this processing will take data from the results of cleansing and categorizing raw data with Pentaho Data Integration tools to the database until the OLAP database is formed.

Then do a histogram check of the ARIMA residual model obtained. From the figure below it can be concluded that the ARIMA Model (1,1,1) (1,0,0) can be said to be a significant ARIMA model to be used in predictions because the residual histogram is normally distributed and there is no lag coming out of the significant line boundary in the plot ACF.

Tools Rstudio functions to run prediction scripts in the R language using the ARIMA method. Then the plot results of the patient visit forecast are visualized with PHP Codeigniter which has been integrated with RStudio. Then analyze the accuracy of forecasting by observing the RMSE and MAPE error measure numbers from the ARIMA forecasting method.
Visualization of Forecasting Number of Patients Visiting in Pratama Clinic Using ARIMA Method

Table 1. Analysis of Results of Data Forecasting Visit by Service and Patient Status

<table>
<thead>
<tr>
<th>DATA FORECASTING RESULTS</th>
<th>RMSE (ARIMA)</th>
<th>MAPE (ARIMA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dental Poly Visit</td>
<td>19.71812</td>
<td>19.58%</td>
</tr>
<tr>
<td>General Poly Visit</td>
<td>104.1736</td>
<td>9.62%</td>
</tr>
<tr>
<td>Visits Based on Status of PNS Patients</td>
<td>86.06643</td>
<td>9.47%</td>
</tr>
<tr>
<td>Visits Based on Non-PNS Patient Status</td>
<td>28.92071</td>
<td>13.69%</td>
</tr>
<tr>
<td>Visits Based on Dependent Patient Status</td>
<td>15.87885</td>
<td>11.23%</td>
</tr>
</tbody>
</table>

V. CONCLUSION

The conclusions obtained from this thesis research are:

1. Based on the analysis of historical data on visits to the clinic, the type of dimension model that is suitable to be applied is the galaxy schema caused by two fact tables and the sharing of dimension tables. These tables include: (1) service dimensions, (2) doctor dimensions, (3) status dimensions, (4) time dimensions, (5) visiting facts, and (6) visiting facts based on status.

2. Based on the forecasting results of the dental clinic service visit the highest point forecast in August 2020 and January 2021 was 121 and 122 visits. Whereas in general poly services get the highest point forecast in August 2020 and August 2021 as many as 797 and 770 visits.

3. Based on the prediction results of patient visits that have been made, RMSE and MAPE values obtained by the ARIMA method which when compared with the significance value of MAPE are already good, for example in the prediction of patient visit data in the Dental Poly service gets RMSE of 19.71812 and MAPE of 19.58%, because MAPE is in the range of 10-20%, which means good forecasting ability and prediction results of General Poly service patient visits get RMSE as much as 104.1736 and MAPE as much as 9.62%, because MAPE <10% which means that forecasting ability is very good.

4. Based on the visualization of patient visit forecasting in this study, it can be seen that the number of Public Poly and Dental Poly service visits in 2020 to 2021 has fluctuated and the number of patient visits with PNS status in 2020 to 2021 has decreased, the number of patient visits with NON-PNS status increased in number. As for the number of patient visits with dependent status, it has fluctuated.

REFERENCES


