RESEARCH ARTICLE OPEN ACCESS

Forecasting Model of Student Admission in XYZ University with Arima Forecasting Technique

Nugroho Budiarto, ST.,M.Kom.*, Dr. Arief Wibowo, S.Kom., M.Kom**,Sigit Wijanarko, S.Kom., M.Kom***, Eryk Budi Pratama, S.,Kom., M.Kom, M.M****,Alga Yusri, S.,Kom., M.Kom*****

*(Faculty of Information Technology, Universitas Budi Luhur, and Jakarta Indonesia
Email: nugroho.budiarto2014@gmail.com)

** (Faculty of Information Technology, Universitas Budi Luhur, and Jakarta Indonesia
Email: arief.wibowo@budiluhur.ac.id)

*** (Faculty of Information Technology, Universitas Budi Luhur, and Jakarta Indonesia
Email: sgtwijanarko23@gmail.com)

**** (Faculty of Information Technology, Universitas Budi Luhur, and Jakarta Indonesia
Email: eryk.pratama@gmail.com)

***** (Faculty of Information Technology, Universitas Budi Luhur, and Jakarta Indonesia
Email: Junbie13@gmail.com@gmail.com)

--************************----------------------------------

Abstract:

Forecasting is an important method for effective and efficient planning. Forecasting is a prediction, projection or estimated level of uncertainty in the future. During the last seven years, the number of new student admissions at the XYZ University experienced a graph of ups and downs.Starting from 2012 to 2017, the number has always increased but starting in 2018, it has decreased. In this research, we will present XYZ University's New Admissions Model Forecasting by using the ARIMA Method (Autoregressive Integrated Moving Average) Forecasting Technique. This research uses the academic period of registration as a causal variable and the number of prospective new students as a result variable. It is expected that the results of this study can help the XYZ Universityto predict the number of new students, also improve and optimize lecture operational services in the following period.

*Keyword -**Forecasting, Data Mining, Autoregressive, Moving Average, ARIMA***

--************************----------------------------------

I. INTRODUCTION

The growth in the number of new students at XYZ University - Jakarta has increased from 2012 to 2016, while in 2017 and 2018 there has been a declining trend. With the trend of the number of student admissions that are still unstable, a system is needed to predict the acceptance level of new students for the next period, so that the existence of prediction can support the decision making for operational activities of lectures.

To solve the problem, we need to find the best Forecasting Model to estimate thenumber of new student admissions of XYZ University by using Forecasting Techniques. The objective is to be able to estimate the number of prospective new students which has implications for the readiness of resources to improve education operational services at XYZ University.

To forecast the number of new students at XYZ University, we will use Arima forecasting techniques. The Arima method is an integration of Autoregression with Moving Average.

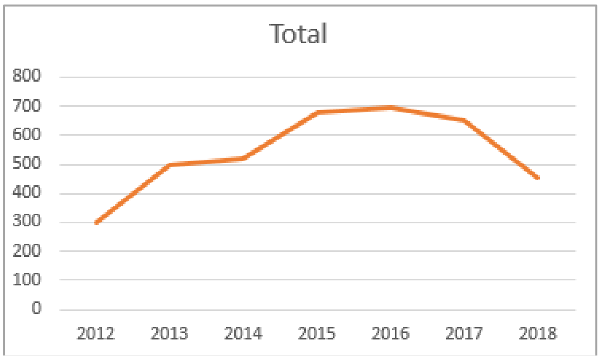

Fig. 1Number of New Student Admissions for XYZ University from 2012 to 2018

A. Problem Identification

Based on the problems outlined in the background, the following problems can be identified in this research:

1. There is less accurate forecasting model for accepting new students, so the current available resources are not utilized.
2. The operational activitiesare disrupted especially in the process of lecture scheduling, room allocation, and lecturer allocation.

There will be amount of excess burden on the resources that have been prepared especially if there is a class cancellation or lecture.

B. Problem Scope

From the problems encountered by XYZ University in determining the prediction of the number of new students, it limits the research as follows:

1. This research focuses on the results of the admission of new undergraduate students from 2012 to 2018 for study programs other than Mathematics (MA), Architecture (AT) and Visual Communication Design (DV) which will be processed into a dataset.
2. This research is to build a Forecasting Model of the Number of New Student Admissions at XYZ University with Forecasting Techniques.

C. Problem Formulation

Based on the description of the problem identification, the authors determined the problem formulation is how to make the Forecasting Number of New Student Admission Model XYZ University with Forecasting Techniques.

D. Research Purposes

The purpose of this research is to build a Model Forecasting the Number of New Student Admissions at XYZ University with Arima Forecasting Techniques.

E. Research Benefits

The benefits of this research are expected to be able to:

1. Help to predict the number of new students, so the prediction results can help improve the operational activities of XYZ university.
2. Prevent the loss of operational activities that are not on the target.
3. Improve the performance of the marketing department in terms of innovation to increase the number of new students in the next period.

II. THEORETICAL BASIS

Forecasting is a procedure for making factual information about future social situations on the basis of information that has been customary about policy issues. Forecasts have three main forms: projections, predictions, and forecasts [1]:

1. A projection is a prediction based on extrapolation of past and present trends to the future. Projections make explicit questions based on arguments obtained from certain methods and parallel cases.
2. A prediction is a prediction based on firm theoretical assumptions. These assumptions can take the form of theoretical law (for example the law of a decrease in the value of money), theoretical propositions (for example the proposition that the breakdown of civil society is caused by a gap between expectations and abilities), or analogies (eg analogies between the growth of government organizations and the growth of biological organisms).

3. A conjecture is a prediction based on informative or expert judgment about the future situation of society.

A. Forecasting Category

Forecasting is usually based on the future time horizon that it covers. The time horizon is divided into several categories [1] :

1. Short-term forecasting
   This forecast includes a period of up to 1 year but is generally less than 3 months.This forecast is usually used to plan purchases, work scheduling, sales, the number of workers, job assignments, and production levels.
2. Medium-term forecasting
   This forecast generally includes a monthly count of up to 3 years. This forecast is useful for planning sales, planning and production budgets, cash budgets, and analyzing various operating plans.
3. Long-term forecasting
   Generally for planning periods of 3 years or more. Long term permalan is used to plan new products, capital expenditure, location or facility development, and research and development (R&D).

B. Forecasting Steps

Good forecasting is forecasting that is done by following the steps or procedures for good preparation. Basically there are three important forecasting steps, namely [2]:

1. Analyzing past data. This analysis is done by making tabulations from past data. With this tabulation you can find out the pattern of the data.
2. Determine the method used. Project past data using the method used, and consider several factors for change.

C. Forecasting Technique

Based on the technique used to predict the prediction can be divided into two parts, qualitative prediction and quantitative prediction [3].

1. Qualitative predictions
   Qualitative predictions are based on qualitative data from the past. Qualitative methods are used if past data from the variable to be predicted are absent, insufficient or less reliable. The results of predictions made are very dependent on the individuals who compose them. This is important because the prediction results are determined based on judgmental thoughts or opinions, knowledge and experiences of the constituents. Therefore this qualitative method is also called judgmental, sudjective, intuitive.
2. Quantitative predictions
   Quantitative predictions are based on quantitative data from the past. The results of predictions made depend on the method used in the prediction. With different methods different predictions will be obtained. The thing to note from the use of these methods is whether the method used and is determined from the deviation between the predicted results with the reality that occurs. A good method is a method that gives the values of differences or possible deviations. Quantitative forecasting can be applied if there are 3 of the following conditions [3]:
   a. Information about other conditions.
   b. This information can be quantified in the form of data..
   c. It can be assumed that past patterns will be sustainable in the future.

   The general procedure used in quantitative forecasting is [3]:
   a. Determine the purpose of forecasting
   b. Determine what elements to predict
   c. Determine the time horizon of forecasting (short, medium or long term)
   d. Determine forecasting models
   e. Collect data needed to make forecasting
   f. Validate forecasting models and apply forecasting results

D. Types and Forecasting Patterns

An important step in choosing an appropriate time series method is to consider the type of data pattern, so that the most appropriate method with the pattern can be tested. Data patterns can be divided into four types, namely [1] :

1. Horizontal Data Pattern
   This data pattern occurs when the data fluctuates around the average value. A product

whose sales have not increased or decreased over a period of time including this type.

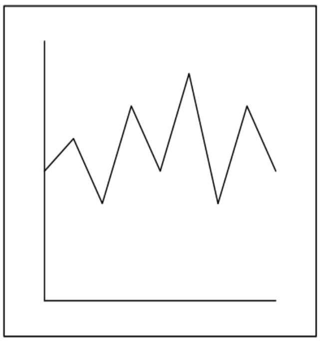

Fig. 2 Horizontal Data Pattern

2. Trend Data Pattern
   This data pattern occurs whenever there is a long-term secular increase or decrease in the data.

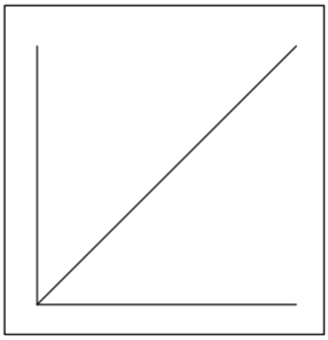

Fig. 3Tren Data Pattern

3. Seasonal Data Pattern
   This data pattern occurs when a series is influenced by seasonal factors (for example a quarter of a particular year, month or days of a particular week).

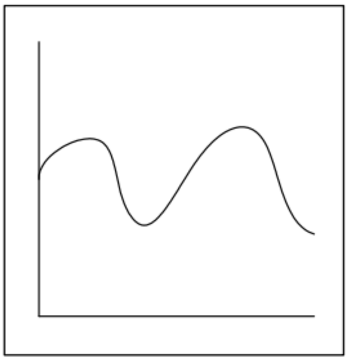

Fig. 4Seasonal Data Pattern

4. Cyclied Data Pattern
   This data pattern occurs when the data is affected by long-term economic fluctuations such as those related to the business cycle.

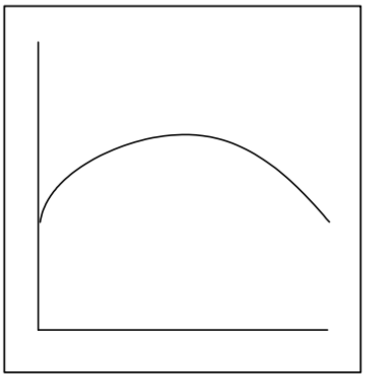

Fig. 5 Cyclied Data Pattern

E. Forecasting Accuracy Measurement

Forecasting models that are carried out are then validated using added indicators. The indicators commonly used are [4] :

1. Mean Absolute Deviation,
2. Mean Square Error,
3. Mean Absolute Percentage Error,
4. Tracking Signal,
5. Moving Range

F. Mean Absolute Deviation (MAD)

Themethod for forecasting methods evaluation uses the sum of absolute errors. The Mean Absolute Deviation (MAD) measures the accuracy of the forecast by averaging the alleged error (the absolute value of each error). MAD is useful when measuring forecast errors in the same unit as the original series [4]

G. Mean Square Eror (MSE)

Mean Squared Error (MSE) is another method for evaluating forecasting methods. Each error or remainder is squared. Then added up and added to the number of observations. This approach manages large forecasting errors because they are squared. The method produces moderate errors which are probably better for small mistakes, but sometimes make a big difference [4].

H. Mean Absolute Percentage Eror (MAPE)

Mean Absolute Percentage Error (MAPE) is calculated using absolute errors in each period divided by the real observed value for that period. Then, average out the absolute percentage error. This approach is useful when the size or size of the forecast variable is important in evaluating the accuracy of the forecast. MAPE indicates how much error in predicting compared to the real value [4].

I. Model Autoregressive

If the stationary series is a linear function of the sequential past values or the present value of the series is the weighted average of the past values along with the current error, then the equation is called the autoregressive model [5].

The general form of this model is:

$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + e_t$...(I)

Where:

Y_t= the predicted AR value

Y_{t-1}, Y_{t-2}, Y_{t-n} = past value of the relevant series; lag value of time series.

A_p = coefficient

e_t = residuals; error that explains the effect of variables that are not explained by the model, forecasting errors with features as before.

The number of past values used (p) in the AR model shows the level of this model. If only a past value is used, it is called the first-level autoregressive model and is denoted by AR. For this model to be stationary, the number of coefficients of the autoregressive model $(\sum_{i=1}^{n} bi)$ must always be less than 1. This is a necessary condition, not sufficient, because there are still other conditions needed to guarantee stationarity.

J. ModelMoving Average

If the stationary series is a linear function of sequential forecasting and past forecast errors, the equation is called the moving average model [5].

The general form of this model is:

$Y_t = e_t - W_1 e_{t-1} - W_2 e_{t-2} - \cdots - W_q e_{t-q}$ ..(2.11)

Where:

Y_t = the predicted MA value

$W_{1,2,q}$ = constants; coefficient or weight

e_t = residuals; error which explains the effect of variables not explained by the model.

It appears that Y_t is a weighted average of errors of n periods backwards. The number of errors used in this equation (q) marks the level of the moving average model. If the model used two past mistakes, it is called the level 2 average model and is denoted as MA.

Almost every exponential smoothing model is in principle equivalent to this model. For this model to be stationary, a necessary (not sufficient) condition, called invertibility condition, is that the number of model coefficients $(\sum_{i=1}^{n} wi)$ is always less than 1. This means that if it goes back to the role of the error getting smaller. If this condition is not fulfilled, mistakes that go back even more play a role.

The MA model predicts the value of Y_t based on a combination of past linear errors (lag), while the AR model shows Y_t as a linear function of a number of previous actual Y_t values.

K. Arima Model

The ARIMA (Autregressive Integrated Moving Average) method is a forecasting method that does not use theories or influences between variables as in the regression model, thus the ARIMA method does not require an explanation of which variables are dependent and independent. 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 the Box-Jenkins method, because it was developed by two US statisticians, namely G.E.P Box and G. Jenkins in 1970.

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 data [5].

The time series model used is based on the assumption that the time series data is stationary, meaning that the average variant (σ2) of a time series data is constant. But as we know that many time series data in economics are not stationary, but rather integrated. If the integrated time series data with order 1 is called I (1) it means the first differencing. If the series goes through the differencing process as many times as d can be made stationary, then the series is said to be non-stationary homogeneous level d. Often stationary random processes cannot be properly explained by the moving average model alone or autoregressive, because they contain both. Therefore, the combination of the two models, called the Autregressive Integrated Moving Average (ARIMA) model can more effectively explain the process [5].

In this combined model the stationary series is a function of the past value and the present value and past error.

The general form of this model is:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + A_p Y_{t-p}$$
$$-W_1 e_{t-1} - W_2 e_{t-2} - \cdots - W_q e_{t-q}$$

Where:

Y_t = stationary series value

Y_{t-1}, Y_{t-2} = past value of the relevant sereies

e_{t-1}, e_{t-2} = independent variables which are lags of residuals

W_1, W_q, A_1, A_p = model coefficient

L. Data Mining

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases. The term data mining has a nature as a scientific discipline whose main purpose is to find, explore, or mine knowledge from the data or information that we have. Data mining, often also referred to as Knowledge Discovery in Database (KDD). KDD is an activity that includes the collection, use of data, historical to find regularities, patterns or relationships in large data sets [6].

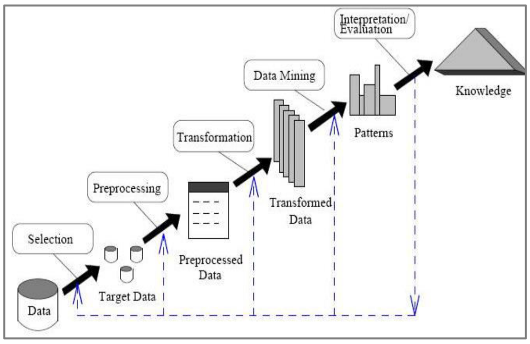

Fig. 6 Data Mining Steps

III. METHODOLOGY AND RESEARCH DESIGN

A. Research Methods

In this study will apply forecasting techniques with the Autoregressive Integrated Moving Average (ARIMA) method to determine the number of students in the next period. The dataset used in this study is the data of new student admissions in the period of 2012 to 2017 period equipped with registration date parameters, this dataset will be used as a database for forecasting the number of new students and then for test data is the data of new students in the 2018 period.

Based on the intent and scope of this study, this study was conducted using forecasting techniques with the Integrated Moving Average Autoregressive Method (ARIMA). This method is carried out by researchers by testing the dataset which later the results of this study will produce a prediction of the number of new students in the next period.

B. Data Collection Methods

In this study the data is taken using primary data by restoring data from the server and can directly query to form a dataset that will be used as a source or research database.

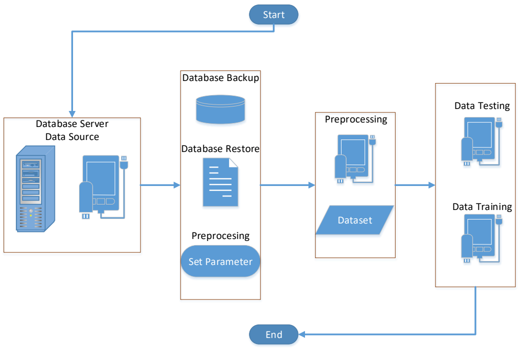

Fig. 7 Data Collection

C. Proposed Model

In this study the authors propose using the Autoregressive Integrated Moving Average (ARIMA) method.

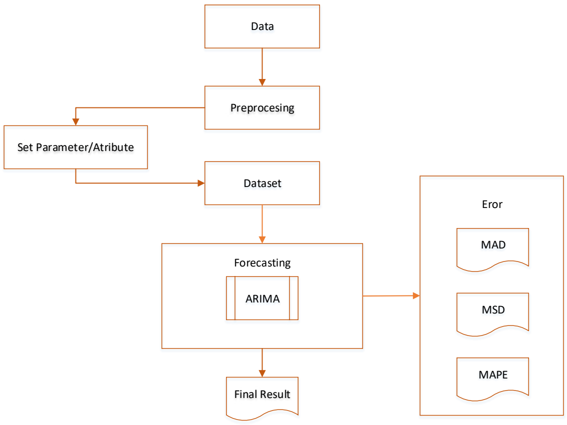

Fig. 8 Proposed Model

D. Data Processing

Authors do a direct query on the source database marketing and then do the preparation.

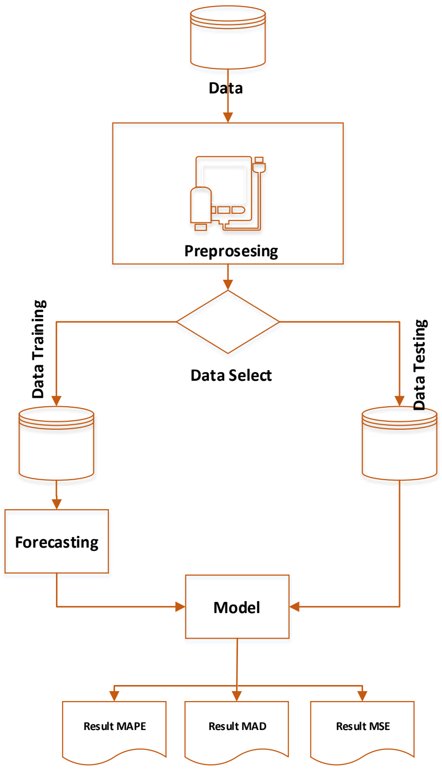

Fig. 9 Data Processing

IV. DISCUSSION AND RESEARCH RESULTS

The following are the stages of implementation carried out in the study.

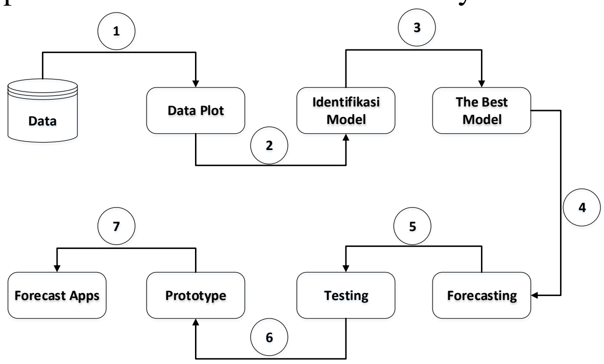

Fig. 10 Research Implementation

The stages of research implementation can be described as follows:

1. The process of Knowledge Discovery Database (KDD)
   In this step is carried out a query process to prepare time series data which is used as the basis for forecasting the number of new student admissions.

2. Data Plot
   The first step that must be done is to plot the original data, from the plot can be seen whether the data is stationary. If the data is not stationary in the mean it is necessary to do a differencing process. In this stage, data distribution is also processed, whether distributed normally or not.
3. Model Identification
   After stationary data in the mean and variance the next step is to look at the ACF and PACF plots. From the ACF (autocorrelation function) and PACF (partial autocorrelation function) plots, several possible models that are suitable to be modeled can be identified.
4. Estimation and Model Selection
   After successfully establishing several possible models that are suitable and estimating the parameters. Then the significance coefficient is tested. If the coefficient of the model is not significant, then the model is not suitable for forecasting.
   The things that need to be considered in taking the model are as follows:
   a. The model should be as simple as possible. In the sense of containing as few parameters as possible, so that the model is more stable.
   b. In comparing models, always choose the model with the highest accuracy, which is the one that gives the smallest error.
5. Forecasting
   At this stage a prediction or forecasting process is done from the model that is considered the best and can be predicted in the next few periods.
6. Model Testing
   From the results of the forecasting will be calculated the value of Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD) and Mean Squared Deviation (MSD).
7. Application Prototype
   The design of the prototype application as a basis for the application development process Forecasting the Number of New Students of XYZ University with Forecasting Techniques.
8. Application Development
   The final process in this implementation is the development of the XYZ UniversityForecasting Number of New Student Application applications with Forecasting Techniques.

The steps in applying the ARIMA method are as follows identification of the model, estimating the parameters of the model, checking the diagnosis and applying the model for forecasting[7].

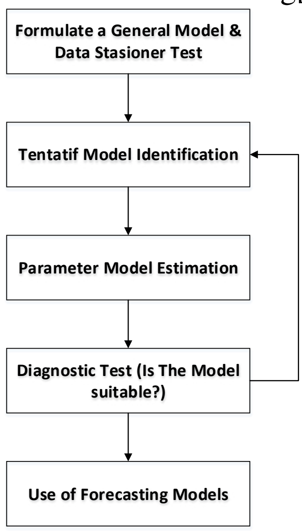

Fig. 11Arima Steps

A. Knowledge Discovery Database

In this study the process of determining the dataset is done by database query.

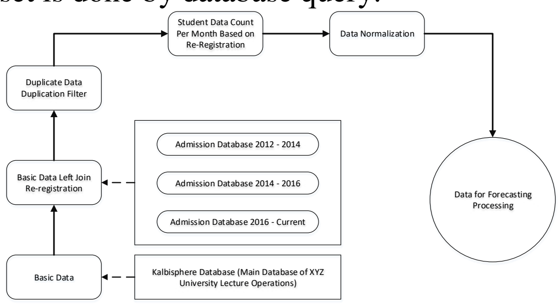

Fig. 12 Data Processing Stages

After obtaining a time series dataset the number of new admissions for the 2013 to 2018 period. Then the time series data will be tested according to the steps below.

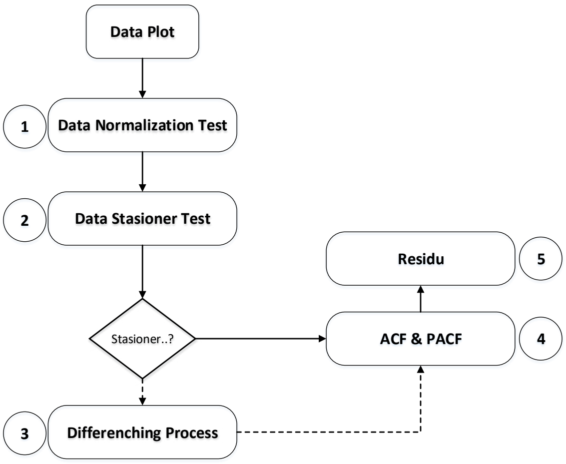

Fig. 13 Dataset Tests

B. Data Normalization Test

The data normalization test in this study was first carried out through a distribution test with the Kolmigorov-smirnov test to find out whether the data distribution was normal or not..

Basic provisions for decision making are based on probability if the probability value> 0.05 then the population is normally distributed and if the probability value <= 0.05 then the population is not normally distributed. With the time series data on the admission of new students in the period 2013 to 2018 shows that the value of Asymp. Sig. (2-tailed) is greater than 0.05. Because the Asymp value. Sig. (2-tailed) is greater than 0.05, it can be concluded that the time series data for new student admissions in the period 2013 to 2018 are normally distributed [8].

One-Sample Kolmogorov-Smirnov Test

		Jumlah
N		90
Normal Parameters[a,b]	Mean	38,87
	Std. Deviation	37,748
Most Extreme Differences	Absolute	,152
	Positive	,118
	Negative	-,152
Kolmogorov-Smirnov Z		1,438
Asymp. Sig. (2-tailed)		,032

a. Test distribution is Normal.

b. Calculated from data.

Fig. 14 Kolmogorov

To test the stationarity of various tests using Box-Cox Transformation, the data is said to be stationary if the value of the Rounded Value is 1.00.

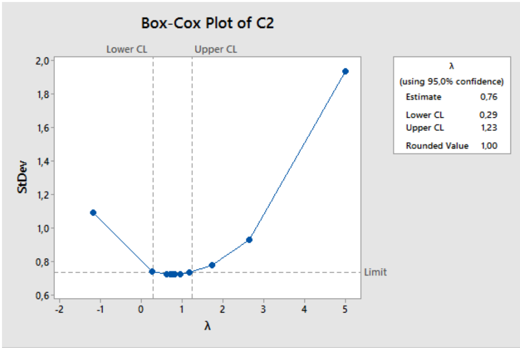

Fig. 15 Box-Cox

Furthermore, the autocorrelation test is performed to test whether in the linear regression model there is a correlation between the error of interruption in the t period with the error of the previous t period. After concluding stationary data the correlation test was performed using Pearson Correlation and Durbin Watson.

1. *Pearson Correlation*

Basis of Decision Making:

- If the significance value <0.05, then it is correlated.
- If the significance value> 0.05, then it does not correlate.

Degree of Relationship:

- Pearson Correlation value 0,00 to 0.20 = no correlation
- Pearson Correlation value 0.21 to 0.40 = weak correlation
- Pearson Correlation value 0.41 to 0.60 = moderate correlation
- Pearson Correlation value 0.61 to 0.80 = strong correlation
- Pearson Correlation value 0.81 to 1.00 = perfect correlation

Correlations

		Waktu	Jumlah
Waktu	Pearson Correlation	1	,018
	Sig. (2-tailed)		,866
	N	90	90
Jumlah	Pearson Correlation	,018	1
	Sig. (2-tailed)	,866	
	N	90	90

Fig. 16 Pearson Correlation

2. *Durbin Watson*

Basis of Decision Making:

- If du <DW <4 - du means there is no autocorrelation problem

- If du <= DW <= du or 4-du> = DW> = 4-dl, no conclusions can be drawn.
- If DW <dl or DW> 4-dl then there is autocorrelation

The following results of the correlation test with Durbin Watson with the SPSS application The amount of N data (amount of data) is 90 with a DW value of 1,184, according to the Durbin Watson table reference the du value is 1.61 and the dl value is 1.70. then from the results above it can be concluded that there is no autocorrelation problem and the data is autocorrelation.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,018[a]	,000	-,011	37,956	1,184

a. Predictors: (Constant), Waktu

b. Dependent Variable: Jumlah

Fig. 17 ACF Durbin Watson

C. Defferenching

At the differencing stage to perform the stationarity of the data against the average range because in the ACF pattern there are still some lags coming out of the confidence interval, so the defferenching process needs to be done so that the data can be stationary against the average range. For the results of differencing plot ACF and PACF can be seen as shown below. [7]

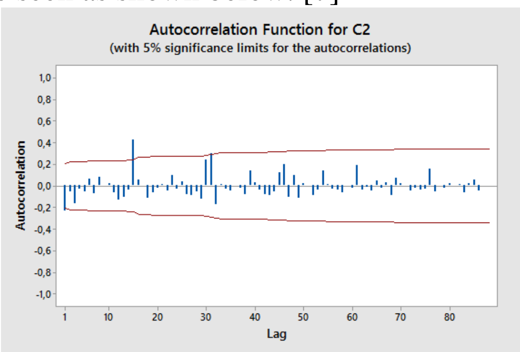

Fig. 18 ACF Diff-1

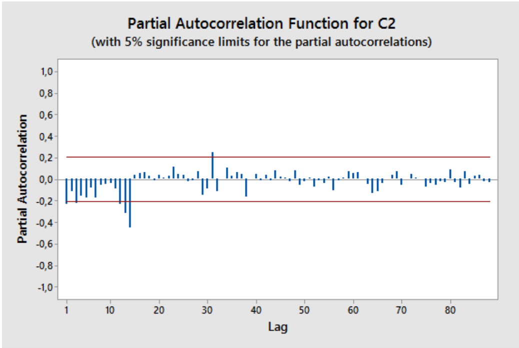

Fig. 19 PACF Diff-1

D. Residual Test

At this residual test stage, it is used to determine whether time series data can be used to determine the process of forecasting the number of new students. Residual is the difference between the predicted value and the actual observed value if the data used is sample data. From the residual test it can be concluded that the time series data in table 4.9 has a deviation value of 2.37%, with this value the time series data can be used for forecasting modeling with the ARIMA Method.

E. Arima Model Selection

The selection of the Arima model is based on time series data on the number of new student enrollments for the period 2013 to 2018. The process of identifying the Arima model used is Model (1,1,0), Model (0,1,1) and Model (1,1,1) , with this model, processing with Minitab application is done by finding the P value that is close to 0 (zero). Following are the time series data test results for the number of new students enrolled for the period 2013 to 2018 by using the Minitab application.

Model (1,1,0), namely the Autoregressive value: 1, Difference: 1, Moving Avarage: 0 based on time series data on the number of new student enrollments for the period 2013 to 2018.

```
Final Estimates of Parameters

Type          Coef  SE Coef      T      P
AR   1     -0,2319   0,1043  -2,22  0,029
Constant    -0,202    4,283  -0,05  0,962


Differencing: 1 regular difference
Number of observations:  Original series 90, after differencing 89
Residuals:     SS =  142024 (backforecasts excluded)
               MS =  1632  DF = 87
```

Fig. 20 Arima (1,1,0)

The P value in the Autoregressive order has a value of 0.029, which means that the value is not significant for any alpha. So the Arima model (1,1,0) cannot be used for forecasting.

Model (0,1,1), namely with Autoregressive value: 0, Difference: 1, Moving Avarage: 1 based on time series data on the number of new students enrolled for the period 2013 to 2018.

```
Final Estimates of Parameters

Type        Coef  SE Coef      T      P
MA   1    0,4163   0,0987   4,22  0,000
Constant  -0,107    2,468  -0,04  0,966


Differencing: 1 regular difference
Number of observations:  Original series 90, after differencing 89
Residuals:    SS =  137957 (backforecasts excluded)
              MS =  1586  DF = 87
```

Fig. 21 Arima (0,0,1)

The P value on Moving Average has a value of 0,000 which is already significant in any alpha, so this model can be used for forecasting.

Model (1,1,1), namely the Autoregressive value: 1, Difference: 1, Moving Avarage: 1 based on time series data on the number of new students enrolled for the period 2013 to 2018.

```
Final Estimates of Parameters

Type          Coef  SE Coef      T      P
AR   1      0,4096   0,1022   4,01  0,000
MA   1      0,9794   0,0410  23,90  0,000
Constant   -0,0175   0,2223  -0,08  0,937


Differencing: 1 regular difference
Number of observations:  Original series 90, after differencing 89
Residuals:    SS =  108112 (backforecasts excluded)
              MS =  1257  DF = 86
```

Fig. 22 Arima (1,1,1)

The P value in Autoregressive has a value of 0,000 and the Moving Avarage has a value of 0,000, both parameters are significant at any alpha, which means that this model is the most suitable for forecasting time series data for the number of new students enrolled in the period 2013 to 2018.

F. Modeling

The model that will be used in forecasting calculations is Model (1,1,1) because it has a P value with Autoregressive = 0,000 and MA = 0,000 and both Autoregressive and Moving Avarage values are significant.

```
Forecasts from period 75

                      95% Limits
Period  Forecast    Lower     Upper  Actual
    76    21,608  -41,902    85,118
    77    31,532  -38,975   102,039
    78    36,090  -36,104   108,284
    79    38,183  -34,519   110,885
    80    39,145  -33,757   112,047
    81    39,586  -33,422   112,594
    82    39,789  -33,291   112,870
    83    39,882  -33,258   113,022
    84    39,925  -33,269   113,119
    85    39,945  -33,301   113,190
    86    39,954  -33,342   113,249
    87    39,958  -33,387   113,303
    88    39,960  -33,435   113,354
    89    39,961  -33,483   113,404
    90    39,961  -33,532   113,454
```

Fig. 23 Forecast Result

For forecasting results for the next 15 periods or 2018 new student admissions are predicted to be 565 new students when compared to the original data of new student admissions are 454 new students. 2018 forecasting results with data on the admission of new students for the period 2013 to 2017 experienced a difference of 24.44%.

G. Application Prototype

In this study the Arima application was developed using the PHP programming language and with a MySQL database.

- Application Login:
  Login access to the Forecasting Application Number of New Students is still limited to Admin Login, not yet equipped with access rights for each menu. To login the application has remainedder if the password or username is incorrect.

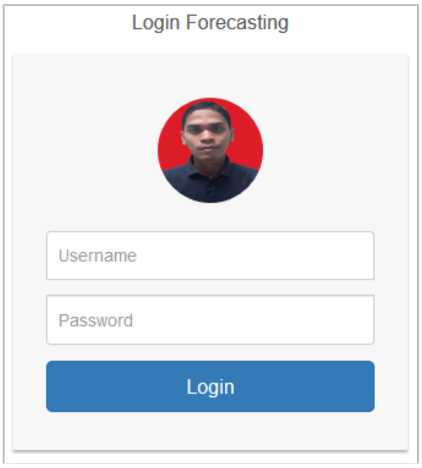

Fig. 24Application Login

- Dashboard Page:
  On the dashboard page this application displays data on the number of new students per period in the form of tables and graphs.

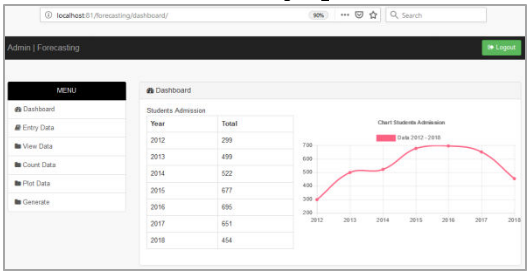

Fig. 25 Dashboard Page

- Generate Forecasting :
  This menu is used to generate Forecasting the Number of New Students by Generating it will display the Results of the Application and Analysis of MSD, MAD and MAPE.

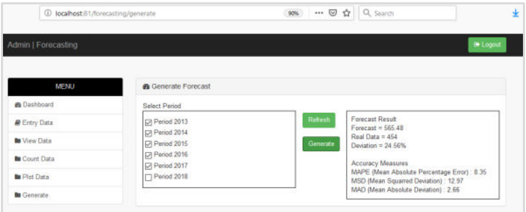

Fig. 26 Generate Forecasting

V. CONCLUSIONS

Some conclusions that can be obtained from the research that has been done are as follows:

1. The forecast deviation rate for the number of new students for the 2018 period based on time series data for new student admissions in 2013 to 2017 reaches 24.4% with actual data.
2. The Arima method is suitable for forecasting time series data with an accuracy rate of 75.6%, this accuracy rate is better than the Linear Regression method 62%, SMOreg 43%, Gausian Processes 37%, and Random Forest 58%.
3. There is data that has not been normally distributed in the 2012 time series data.

Suggestions

In this study there are limitations to the problem and there are still many shortcomings in the system that has been developed in this study. Some things that can be used as a reference or consideration in further research, namely:

1. The research is still focused on the admission data for new undergraduate students, so that further research can also be done for master and associate's degree levels.
2. In future research, it can be done by adding parameters in the time series data which can increase the level of accuracy.
3. There is still a lot of data entry that is not in accordance with the applicable procedures so that the data is not normally distributed, so in the future there needs to be an improvement in terms of data entry according to the procedure, and a validated system should be built to avoid this..

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES

[1] S. Makridakis, S. C. Wheelwright, dan V. E. McGEE, *Forecasting: Methods and Applications*. 1999.

[2] S. Wardah dan Iskandar, "Analisis Peramalan Penjualan Produk Keripik Pisang Kemasan Bungkus (Studi Kasus : Home Industry Arwana Food Tembilahan)," *J. Tek. Ind.*, vol. XI, no. 3, 2016.

[3] A. Purba, "Perancangan Aplikasi Peramalan Jumlah Calon Mahasiswa Baru yang mendaftar menggunakan Metode Single Exponential Smoothing (Studi Kasus: Fakultas Agama Islam UISU)," *J. Ris. Komput.*, vol. 2, no. 6, hal. 8–12, 2015.

[4] A. M. Dharmesta dan N. Susanto, "Peramalan Perencanaan Produksi Terak Dengan Metode Exponential Smoothing With Trend Pada Pt . Semen Indonesia (Persero) Tbk .," hal. 1–10, 2017.

[5] M. As'ad, S. S. Wibowo, dan E. Sophia, "Peramalan Jumlah Mahasiswa Baru Dengan Model Autoregressive Integrated Moving Average (Arima)," *J. Inform. Merdeka Pasuruan*, vol. 2, no. 3, hal. 20–33, 2017..

[6] M. Ridwan, H. Suyono, dan M. Sarosa, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes

Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, hal. 151, 2018..
[7] B. Hendrawan, "Penerapan Model ARIMA Dalam Memprediksi IHSG," *J. Integr.*, vol. 4, no. 2, hal. 205–211, 2012.
[8] M. A. Oktaviani dan H. B. Notobroto, "Perbandingan Tingkat Konsistensi Normalitas Distribusi Metode Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, dan Skewness-Kurtosis," *J. Biometrika dan Kependud.*, vol. 3, no. 2, hal. 127–135, 2014.
[9] A. A. Adebiyi, A. O. Adewumi, dan C. K. Ayo, "Stock Price Prediction Using The ARIMA Model," *Proc. - UKSim-AMSS 16th Int. Conf. Comput. Model. Simulation, UKSim 2014*, hal. 106–112, 2014

Mail your Manuscript toeditorijctjournal@gmail.com
editor@ijctjournal.org