

**APPLICATION TO PREDICT THE NUMBER OF APPLICANTS FOR  
NEW STUDENTS WITH A TIME SERIES MODEL**



**Submitted as partial Fulfillment of the Requirements for Getting Bachelor Degree of  
Informatics Departement Faculty of Communication and Informatics**

**By:**

**NU'MAN NORMAS MUHAMAD**

**L200164003**

**DEPARTEMEN OF INFORMATICS  
FACULTY OF COMMUNICATION AND INFORMATICS  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**2019**

**APPROVAL PAGE**

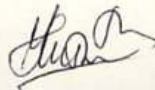
**APPLICATION TO PREDICT THE NUMBER OF APPLICANTS FOR  
NEW STUDENTS WITH A TIME SERIES MODEL**

**SCIENTIFIC PUBLICATION**

Submitted by:

**Nu'man Normas Muhamad**  
**L200164003**

Have been reviewed and approved to be tested by:  
Supervisor



**Husni Thamrin, M.T, Ph.D**  
**NIK. 706**

VALIDATION PAGE

APPLICATION TO PREDICT THE NUMBER OF APPLICANTS FOR  
NEW STUDENTS WITH A TIME SERIES MODEL

BY

Nu'man Normas Muhamad

L200164003

It has been defended in front of Examiners  
Faculty of Communication and Informatics  
Muhammadiyah University of Surakarta  
At ~~Surabaya~~ Feb 8<sup>th</sup> 2020  
And declared to qualify

Examiners:

1. Husni Thamrin, M.T, Ph.D.  
(Chief of Examiners Board)
2. Fajar Suryawan, Ph.D.  
(Member 1 of Examiners Board)
3. Heru Supriyono, Ms.C, Ph.D.  
(Member 1 of Examiners Board)

  
(.....)

  
(.....)



Dean of  
Communication and Informatics Faculty

  
Nurgiyatna, S.T., M.Sc., Ph.D.  
NIK. 881



Head of  
Informatics Departement

  
Heru Supriyono, S.T., M.Sc., Ph.D.  
NIK. 970

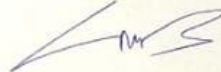
## STATEMENT

I hereby declare that in this publication article there is no work ever submitted for a degree at a college and to the best of my knowledge there is no work or opinion ever written or published by any other person, except in writing referred to in the manuscript and mentioned in the list library.

If later there is proved untruth in my statement above, then I will be fully accountable.

Surakarta, Feb 18<sup>th</sup> 2020

Writer



NU'MAN NORMAS MUHAMAD

L200164003



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

**SURAT KETERANGAN LULUS PLAGIASI**

No Surat 42/A-4-11.3/Inf-FK1/11/2020

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : Nu'man Normas Muhamad  
NIM : L200164003  
Judul : *Application to predict the number of applicants for new students with a time series model*  
Program Studi : Informatika  
Status : Lulus

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 18 Februari 2020

Biro Skripsi Informatika

Ihsah Cahyo Utomo, S.Kom., M.Kom.



UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id> Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

Feedback Studio - Google Chrome  
#xctumzts.com/ooip/cantak/ku/7n...idku-10f7759080/ang-en\_usda-1259274675

feedback studio APPLICATION TO PREDICT THE NUMBER OF APPLICANTS FOR NEW STUDENTS WITH A TIME SERIES MODEL

### APPLICATION TO PREDICT THE NUMBER OF APPLICANTS FOR NEW STUDENTS WITH A TIME SERIES MODEL

**Abstrak**

Permasalahan-permasalahan yang akan dihadapi oleh lembaga pendidikan tinggi khususnya pada fase penerimaan mahasiswa baru. Perencanaan yang matang serta strategi dibutuhkan dalam mengatasi proses penerimaan mahasiswa baru. Data-data untuk melakukan perencanaan dapat diperoleh dengan menggunakan metode peramalan. Model peramalan runtun waktu digambarkan untuk mendapatkan data peramalan. Data-data peramalan digunakan untuk proses pengambilan keputusan. Data penerimaan mahasiswa baru yang diperoleh merupakan data 3 periode (2017 - 2019). Data yang didapat merupakan data yang stasioner. Dikatakan data yang memiliki sudah stasioner, data tidak perlu diferensiasi. Data yang didapat juga memiliki nilai korelasi yang cukup, serta memiliki penalaran pada lag ko-7. Sebelum membuat aplikasi dilakukan pengujian untuk mencari model peramalan time series yang cocok untuk data PMB. Model yang diuji adalah model ARIMA dan model Auto Regressive. Pada pengujian realitas waktu peramalan, model ARIMA mendapat nilai error yang lebih kecil daripada RMSECV maupun MAECV yang lebih kecil dibandingkan model AR. Model ARIMA dipilih untuk diimplementasikan ke dalam aplikasi. Algoritma auto\_arima digunakan agar aplikasi dapat beradaptasi dengan data-data yang berbeda. Model ARIMA diimplementasikan ke dalam aplikasi prediksi dengan menggunakan bahasa pemrograman python. Pengembangan aplikasi menggunakan Django sebagai web-framework aplikasi berbasis web. Bootstrap digunakan untuk membuat interface aplikasi.

Kata Kunci: PMB, Dision, Peramalan, ARIMA, Auto Regressive

Page 2 of 15 Word Count: 5185

Bottom Flavour

Text only Output High Resolution On

12:01 PM 10/10/2024

Match Overview

17%

1	link.springer.com	1%
2	Submitted to Universh...	1%
3	matrimakewangmash...	1%
4	Wahid Sir Rahayu, Reg...	1%
5	ejournal.ums.ac.id	1%
6	www.proceedingsof-dia...	1%
7	Submitted to University...	1%

# APPLICATION TO PREDICT THE NUMBER OF APPLICANTS FOR NEW STUDENTS WITH A TIME SERIES MODEL

## Abstrak

Permasalahan-permasalahan yang akan dihadapi oleh lembaga pendidikan tinggi khususnya pada fase penerimaan mahasiswa baru. Perencanaan yang matang serta strategi dibutuhkan dalam menghadapi proses penerimaan mahasiswa baru. Data-data untuk melakukan perencanaan dapat diperoleh dengan menggunakan metode peramalan. Model peramalan runtun waktu digunakan untuk mendapatkan data peramalan. Data-data peramalan digunakan untuk proses pengambilan keputusan. Data penerimaan mahasiswa baru yang diperoleh merupakan data 3 periode (2017 – 2019). Data yang didapat merupakan data yang stasioner. Dikarenakan data yang dimiliki sudah stasioner, data tidak perlu diferensiasi. Data yang didapat juga memiliki nilai korelasi yang cukup, serta memiliki perulangan pada lag ke-7. Sebelum membuat aplikasi, dilakukan pengujian untuk mencari model peramalan time series yang cocok untuk data PMB. Model yang diuji adalah model ARIMA dan model Auto Regressive. Pada pengujian rentang waktu peramalan, model ARIMA mendapat nilai error yang lebih kecil di hampir semua tes. Dalam metode Cross-validation Model ARIMA juga mendapat nilai RMSECV maupun MAECV yang lebih kecil dibandingkan model AR. Model ARIMA dipilih untuk diimplementasikan ke dalam aplikasi. Algoritma `auto_arma` digunakan agar aplikasi dapat beradaptasi dengan data-data yang berbeda. Model ARIMA diimplementasikan ke dalam aplikasi prediksi dengan menggunakan Bahasa pemrograman python. Pengembangan aplikasi menggunakan Django sebagai web-framework aplikasi berbasis web. Bootstrap digunakan untuk membuat interface aplikasi.

**Kata Kunci:** PMB, Django, Peramalan, ARIMA, Auto Regressive.

## Abstract

Problems that will be faced by higher education institutions, especially in the phase of new student admissions. Careful planning and strategies are needed in dealing with the process of admission of new students. The data for planning can be obtained using the forecasting method. The time series forecasting model is used to get forecasting data. Forecasting data is used for the decision making process. The data of new student admissions obtained is 3-period data (2017 - 2019). The data obtained is stationary. Because the data is stationary, the data does not need differentiation. The data obtained also has a sufficient correlation value, and has a loop on the 7th lag. Before making an application, a test is performed to find a time series model that is suitable for admission data. The tested models are the ARIMA model and the AutoRegression model. In testing the forecast timespan, the ARIMA model gets a smaller error value in almost all tests. In the Cross-validation method, the ARIMA Model also gets a smaller RMSECV or MAECV value than the AR model. The ARIMA model was chosen to be implemented into the application. The `auto_arma` algorithm is used so that applications can adapt to different data. The ARIMA model is implemented into a prediction application using the Python programming language. Application development uses Django as a web-based web application framework. Bootstrap is used to create application interfaces.

**Keywords:** admission, Django, forecasting, ARIMA, Auto Regression.

## 1. INTRODUCTION

Higher education is the final level of formal academic education. Higher education institutions can include Colleges to Universities that cover a variety of levels starting from diploma to doctoral level. Higher education which is a formal educational institution has various problems. One of the most crucial is quality problems. Quality problems for an educational institution especially higher educations are almost always oriented to the achievement and prestige of an educational institution. making educational institutions highly dependent on the quality of learning, supporting facilities, and lecturers. these problems also indirectly give birth to a competition. The current competition depends on the quality of graduates from these higher educational institutions. Graduates become the benchmark for the quality of an educational institution. Besides graduates, a variety of good supporting facilities can support a good learning process. as well as if supporting facilities are lacking or inappropriate. The learning process also becomes disrupted and less optimal. To avoid things like disruption to the learning process, quality improvement is needed. improving the quality of a higher educational institution can begin with future planning. With great academic future planning, the problems that will be faced by educational institutions can be prevented (Umar & Ismail, 2018; Wahyudi, 2018).

The making of academic planning requires estimating data that will be used to determine plans so that educational institutions can be better prepared to deal with future problems. These problems that will be faced, one of them is in the process of admitting new students. Admission of new students is a crucial moment for a higher educational institution In this phase. If the number of new students accepted cannot meet the admission target. They will find it difficult to develop and compete with others (Karmita, Putra, Gaffar, & Wiguna, 2019). The number of applicants that changes each year requires different strategies in dealing with the process of admission of new students. Admission strategies also depend on how universities accept their students. There are many ways in the process of admission of new students. ranging from national joint selection that is intended to accept students at State Universities in Indonesia. And there is the admission process is widely used by private universities which all processes can be done in one day, known as One Day Service (ODS) (Surakarta, 2016). Registrant data in the ODS database form in daily data. It can be used to predict the number of registrants in the coming period with the forecasting method. Forecasting data is shown for decision making and application of priorities for the number of new students who will be accepted later (Anggoro & Supriyanti, 2019; Purba, 2015; Sinaga, Sembiring, & Limbong, 2019) .

Based on the estimation of data in the past to predict a variable in the future that is called forecasting. Forecasting does not provide the same as data that will occur in the future. But forecasting looking for data as close as possible to the data that will occur. Forecasting itself is the science and art of predicting future events by using past data (Pravilovic, Appice, & Malerba, 2014). There are



two kinds of forecasting, qualitative method, and the quantitative method. The qualitative method creates forecast data based on intuition or opinion. The quantitative method obtains forecast data using statistics and mathematical rules. In this research, we use a quantitative method. There are two models in quantitative method, Casual Model and time series model. Because admission data is a sequence data. We selected time series model. The time series model is based on a series of sequential data with the same sequence. This forecast model is a set of observational data measured based on time with the same interval. And measured over a certain period (Thamrin & Murray-Smith, n.d.). To obtain accurate prediction data using time series model. It takes a lot of reference of new student candidate data in previous years (Sinaga et al., 2019). forecasting with time series using a quantitative approach with past data which is a reference to get forecast data. Many time series models can be used such as Artificial Neural Network (ANN) (Azid et al., 2014), Single Exponential Smoothing (Purba, 2015), Auto Regression, Auto Regression Integrated Moving Average (ARIMA) (Janah, Sulandari, & Wiyono, 2014; PRASETYO, 2018; Sinaga et al., 2019), and many others.

In previous studies on the comparison of stock prices using the ARIMA model and the Artificial Neural Network (ANN) model, the level of Root mean square error (RMSE) for ARIMA model was smaller than using the ANN model with a difference of more than double (Pandji, Indwiarti, & Rohmawati, 2019). Next references is refer to other research by Sismi and Darsyah. From their research ARIMA Model is more optimal in producing forecasting compared to MA models (Sismi & Darsyah, 2018). Forecasting results with the ARIMA model are better than the two other forecasting methods as evidenced by the results of the two studies above. The use of ARIMA to predict the number of new students has been used in a study from As'ad. in this study the level of differences in the number of registrants in 2017 and 2018 on actual data and the predicted data only around 5%, so the predicted results can be used for planning references (As'ad, Wibowo, & Sophia, 2017). Instead of predicting admission data yearly like other research in predicting number of new students, we will predicting the data daily in which contain ODS system.

## 2. METHOD

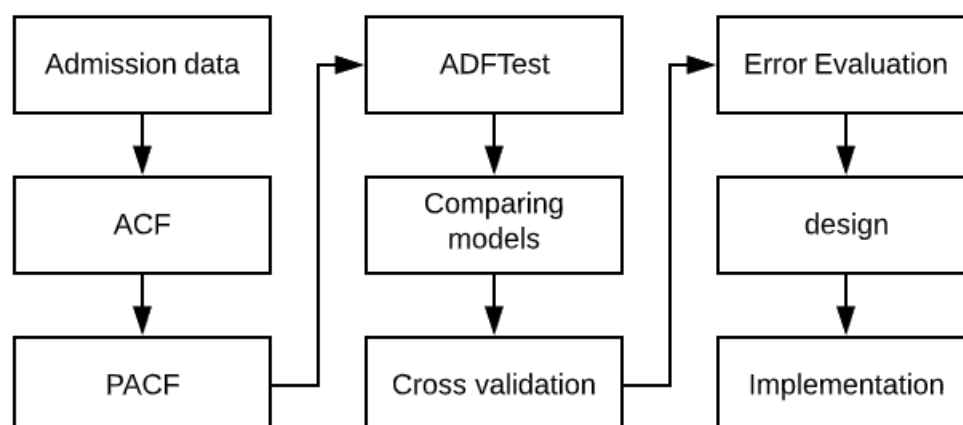


Figure 1. Research Method

In this study, there are several steps that we did these steps can be seen in Figure 1. the initial stage of this research is to collect data obtained from related institution who process ODS data in UMS, the Data obtained is 3 period data (2017 - 2019) in the form of Excel, this data will be converted into CSV format to facilitate the processing of data using the Jupyter Notebook software which will then be processed by the pandas module in python so that it can be further analyzed using the Python programming language. To further analyze the data, the writer uses one department data as a sample, namely Mechanical Engineering department data.

The next stage is the stage to test whether the data to be analyzed has entered the time series criteria or not. This process includes ACF and PACF to see the correlation between the data. ACF is a function to calculate the correlation between data. And PACF is a function to calculate the partial correlation between data. These function also work to define the white noise of the data. To define the data stationary we will use the Augmented Dickey-Fuller (ADF) test. The ADF test is a type of statistical test called a unit root test. it interprets this result using the p-value from the test. A p-value below a threshold (such as 5% or 1%) suggests we reject the null hypothesis (stationary). Otherwise, a p-value above the threshold suggests we fail to reject the null hypothesis (non-stationary).

Third stage is the analysis phase for testing between 2 models. At this stage the researcher used the Auto Regression and ARIMA models to test the suitability of the model for Admission data in UMS. The Auto Regression Model is a model that looks for future predictive values based on a linear combination of the past values of a variable, or it can also be called a regression variable against itself. An autoregressive model can be stated as follows:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t \quad (1)$$

Where y is the value or in this research y is the number of student, t is the time. The  $\mu$  is the population mean that we get from the data. And this symbol  $\phi$ , refer to the AR parameter, which we can find in PACF. The last variable is the error value. To find the forecast value from formula above, we didn't include  $e_t$  variable or error variable. Because actually we didn't know the error at that time yet. And the second model is ARIMA model. ARIMA has 3 main elements, namely: the Auto regression (AR), Moving Average (MA), and Integrated (I) models. The three elements can be modified so that they can form new models. For example Auto regression and Moving Average (ARMA) if the data held is already stationary. ARIMA is formed by 3 orders (p, d, q). p is an order from AR, d is an order from I, and q is an order from MA. Further explanation for each element in ARIMA. Auto Regression (AR), the explanation of AR can be seen in the previous explanation. Integrated (I), here I state the differentiation of data. To make the ARIMA model required to have stationary data, and if the data

to be used by the ARIMA model is stationary, the value of I will be 0. Moving Average (MA), the value of the MA variable here is determined by the error of the variable itself. The ARIMA model is stated as follows:

$$y'_t = \mu + \phi_1 y'_{t-1} + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t \quad (2)$$

Where  $y'$  is the Differenced data, differenced data calculated using simple term by reducing data by data before. Same as AR formula,  $\mu$  is the population mean,  $e$  is the error value. And for  $\theta_q$  is the Moving Average parameter, which reduced over time.

The last stage is the stage to determine which model is having a better performance. To determine this, we use Cross validation method. In time series data the cross-validation method is slightly different from cross-validation in general. We uses the Nested cross-validation method or commonly known as day forward chaining, which has been proven to be good for time series data (Alexander & Rogelio, 2019). To calculating the error measurement of each model we used two error measurement method. The first one is Root Mean Square Error (RMSE) is an error measurement method that measures the average value of a model's error by squaring the error value and finding its root value. Error values in the RMSE method have different weights. The RMSE method stated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (3)$$

Where the  $n$  variable is the number of data, and  $e$  is the error that we got from reducing forecast value by real value. Afterwards the error value is squared, and divided by the number of data. The last step is finding the root from mean squared error. The second error measurement method is Mean Absolute Error (MAE), this method of measuring error that measures the average absolute error value of a model by reducing the predicted value to the original value of a variable and making it an absolute number. In the MAE error measurement method the value of each error variable has the same weight. The MAE method stated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (4)$$

The formula is similar with the RMSE, but instead of erasing negative value by squaring the error. In MAE to erasing negative value is by converting to absolute value. This method also used to find the percentage of error, usually called Mean absolute percentage error (MAPE). After finding the model with best performance between those two methods. The last stage is the stage for designing the

application and implementing the model. The model which obtained in previous stage. This final stage is the process of making forecast applications that will be made using the Python programming language.

The application will be developed using Django web framework for creating the web-based app. Django is a python based web framework. We choose python because python is remarkably easy to understand programming language. And python itself has many library that support machine learning and statistics analysis. These library can be used to analyze the data, and obtain forecast data. After obtain a suitable model. The model will be implement to the application using python programming language. If we got ARIMA model we'll use auto\_arima function in python to determine PDQ order (Jain & Mallick, 2017). Because bootstrap is integrated with Django in visual studio the user interface will be create using bootstrap framework.

This application will only have 2 main UI. In the first UI, the user will have to input the chosen department, and number of days. Number of days is an input variable which is how many days will be forecast. The second UI will show the forecast result. To show the forecast result data, we will use Highchart for displaying the data as chart. For the database we use MySQL database to store the data. The database is used to store training data for the application.

### 3. RESULTS AND DISCUSSION

The entire data that researcher got from the UMS database is 3 period data (2017 - 2019) of the Faculty of Engineering in UMS. In this study the researchers used one of the majors as sample data, the chosen major was the Mechanical Engineering major. This data gathered using pandas library and saved as data frame in python.

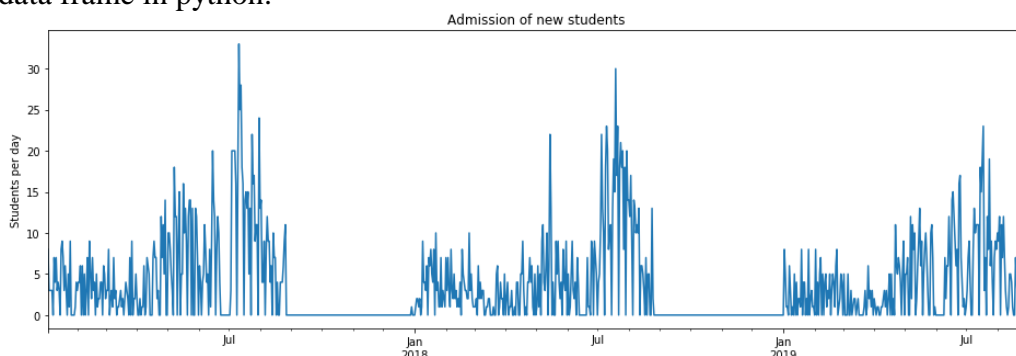


Figure 2. Mechanical Engineering Admission data

In Figure 2, it can be seen that from September to the end of December the data for the number of registrants remained at 0, this is because in previous periods the UMS opened new student admissions starting from January to late August. Before making a forecasting application, researcher took 2 time series forecasting models that will be tested to find which model is more suitable for Admissions data which then will be implemented into forecasting applications.

#### 3.1 ACF and PACF

This step is to determine the correlation between the data. ACF and PACF plot can be useful to see

whether the data is a stationary data or not. The horizontal axis indicate the lag, and vertical axis indicate correlation. Otherwise as we can see in figure 3 and 4. The first value (0<sup>th</sup> lag) is always 1. A sharp drop from lag 0 to lag 1 indicate the data is stationary. And from the 7<sup>th</sup>, 14<sup>th</sup> lag and so on, the value is rising. It indicate that the data has seasonality.

ACF also can be used to test whether the data is white noise or not. The white noise appear when the data in time series is random data and have no correlation. If the 95% of ACF is less than ((+/-) 2/sqrt(data length)) or (0.064/-0.064) it means that the data is white noise. In figure 3, it appears that more than 5% of the data is higher than (0.064/-0.064). So the data is not white noise (Williams, 2003).

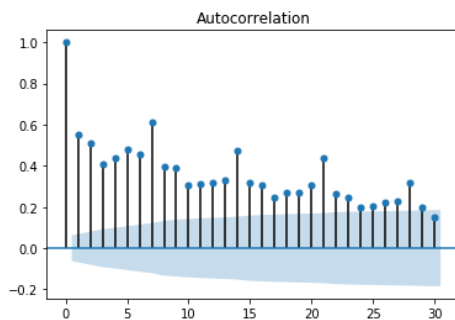


Figure 3. ACF Plot

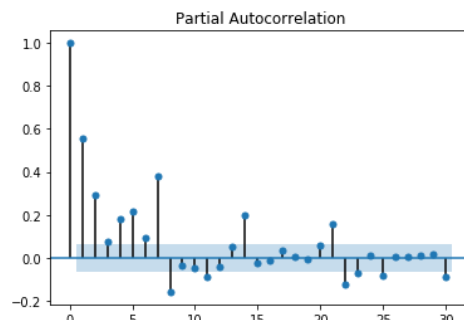


Figure 4. PACF Plot

### 3.2 ADF Test

The results of the ADF Test can be seen in Figure 5, showing the data is stationary. To see the stationary data from the ADF test results can be seen in the p-value column, the results show that the p-value is less than 5% (1.1443%) it is shown that the data has no unit root and will reject the null hypothesis.

```
Augmented Dickey-Fuller Test:
ADF test statistic      -3.386082
p-value                0.011443
```

Figure 5. ADF Test result

From the ADF test results above, it can be concluded that the data is not influenced by trends (Stationary).

### 3.2 Auto\_arima

To determine the order of the ARIMA model the researcher uses the auto\_arima function. Auto\_arima can also be used to predict using the order contained in the auto\_arima results. In figure 6, the model obtained from the auto\_arima algorithm is SARIMA (5,0,5). It is because the data have seasonality. And also the data is stationary, variable I (d) on ARIMA is not needed.

Dep. Variable: y  
 Model: SARIMAX(5, 0, 5)

Figure 6. Auto\_arima result

To determine the order that we got from auto\_arima is acceptable or it's not white noise. We check the ACF plot from the predictions using auto\_arima (SARIMA (5,0,5)). If 95% of the data is less than 0.365. The prediction data is white noise. As we can see in figure 7, more than 5% of the autocorrelation data is higher than (0.365/-0.365). So prediction data which generated by auto\_arima is not white noise, and acceptable.

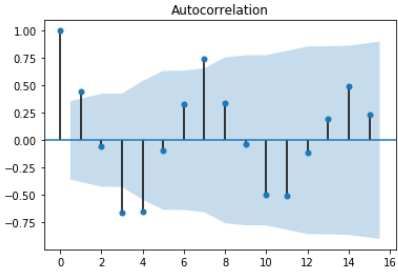


Figure 7. auto\_arima ACF Plot

**3.2 1-Month Testing**

The graph that is formed from the prediction results using the Auto Regression model and the ARIMA model can be seen in Figure 7. In figure 7 there are 3 lines representing the real value of the new students, predicted value using AR, and predicted value using ARIMA model. Using 938 data as training data and 30 data are used as testing data. And for the error measurement it can be seen in table 2 which is the result of error testing using RMSE and MAE on both models. The data table above shows the error rate from the prediction data and the original data for the last 1 month (End of July - August). The results of the error test show that the error value from the ARIMA model is smaller than the error value from the Auto Regression model. The smaller error value shown better predicted value. Same as error measurement using the MAE method which gets the results that the ARIMA model is smaller than the AR model.

Table 2. 1-Month error result

Model	RMSE	MAE
AutoRegression	4.392	3.849
SARIMA	3.747	2.894

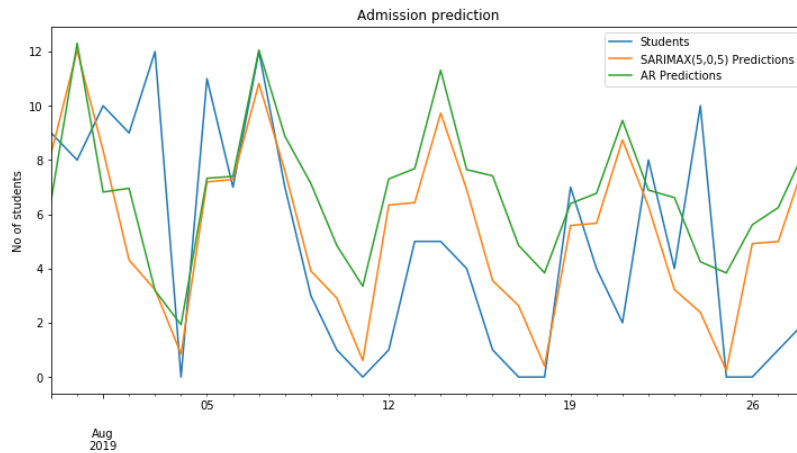


Figure 7. 1-Month test result

In Figure 7, it can also be seen the pattern of prediction results, both Auto Regression and SARIMA have repeated the same pattern and it shrunk, where every 7 days the prediction value of the two models will approach the value of 0, which every once a week the UMS also dismiss the ODS committee which falls on Sunday. And for both model the value that approach the value of 0 appear also in Sunday. The graph that is formed from the SARIMA model is also more integrated with the testing data graph than the graph that is formed from the Auto Regression model.

To see the distribution of error values, the researcher makes a graph that displays the error values of each model displayed in one graph, which can be seen in Figure 8.

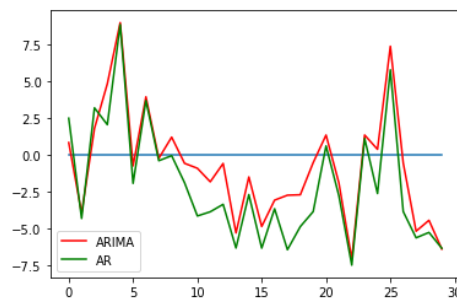


Figure 8. Error Distribution

The distribution of error values of each model has the same value ARMA (-7.5 - 9) and AR (-7.5 - 9). However, the spread of error from the SARIMA model is closer to zero than the spread of error in the AR model which is more away from the value of 0.

### 3.3 2-Weeks Testing

The total of training same as the previous experiment (938 data) and for the testing data it reduced to 14 data (2 weeks). The prediction of the SARIMA and AR model produce a graph that can be seen in Figure 9. In Figure 9 the predictive data pattern of AR still has the same low point as the testing data on Sunday, but the resulting low point pattern of AR model is still far from zero value. The graphs formed by the SARIMA model have low point deviations that still fall on Sunday. The pattern formed by each model has similar pattern.

Table 3. 2-Week error result

<b>Model</b>	<b>RMSE</b>	<b>MAE</b>
AutoRegression	3.967	3.313
SARIMA	3.513	2.542

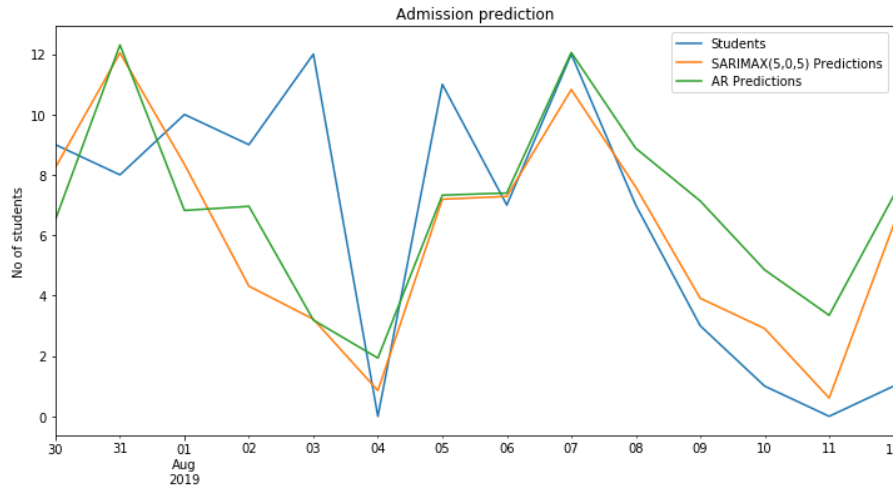


Figure 9. 2-Week test result

In table 3 the value of the measurement error generated by the AR model to predict data 2 weeks is greater than the error value of the SARIMA model, with a large difference. Error value using a span of 2 weeks has a smaller error value than the predicted results with a data span of 1 month.

### 3.4 1-Week Testing

In testing with testing data for 1 week, the error value in both models is higher than the error value in the previous test. The patterns that are formed from the SARIMA and AR models are almost the same to each other. And for the resulting graph, it can be seen in Figure 10. The error measurement value generated from the SARIMA model with the MAE error measurement method is also smaller than the error measurement value in the AR model. But for error measurement using the RMSE error measurement method the error rate of the AR model is smaller than the SARIMA model.

Table 4. 1-Week error result

<b>Model</b>	<b>RMSE</b>	<b>MAE</b>
AutoRegression	4.372	3.773
SARIMA	4.45	3.57



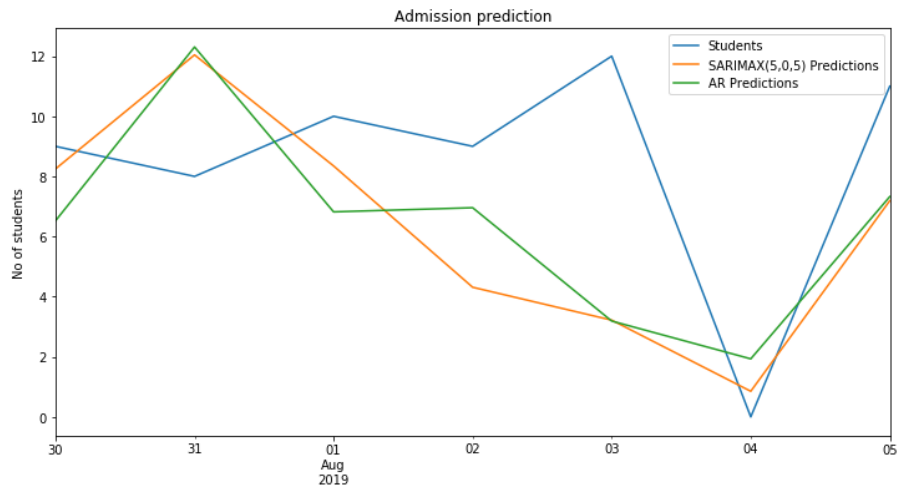


Figure 10. 1-Week test results

In table 4, the difference in error result happened because MAE measure the absolute error for each value. Absolute error give each error same weight and RMSE measure the error value by squaring the error value. squaring the error value have impact to different weight to each error value in RMSE (Chai, Chai, & Draxler, 2014). From the 3 tests (1 month, 2 weeks, 1 week). Commonly ARIMA or AR can't forecast with longer period. Because the more data ARIMA or AR will forecast, it'll only will getting from the forecast value. Which is the forecast value itself has an error. Therefore the more data need to be forecast the more error it'll be get.

### 3.5 Cross Validation

Here we used grouping data for day forward chaining of 10 groups, each group contained 14 data (2 weeks). Researchers used data as much as 14 data because in the previous experiment the smallest error value was generated in the 2-week trial. This day forward chaining method aims to compare between AR and ARIMA which are more suitable to be applied in forecasting applications. In table 5 can be seen is the value of the measurement error of each group of data.

Table 5. Cross validation result

Data	Auto Regression		ARIMA	
	RMSE	MAE	RMSE	MAE
945 – 959	5.66	5.03	4.22	3.63
946 – 960	5.34	4.68	4.01	3.41
947 – 961	5.55	5.03	4.45	3.88
948 – 962	4.71	4.09	3.85	3.22
949 – 963	3.9	3.23	3.36	2.62
950 – 964	3.2	2.54	3.3	2.47
951 – 965	3.29	2.74	3.31	2.49
952 – 966	2.96	2.34	3.1	2.45
953 - 967	3.19	2.67	3.21	2.58
954 - 968	3.32	2.73	3.22	2.63

To see more clearly the error rate of each model in day forward chaining. Values from each model are grouped and calculate the average error value. Mean error in the cross validation of each model can be seen in Table 6. It can be seen that the ARIMA model has smaller Root Mean Square Error Cross Validation (RMSECV) and Mean Absolute Error Cross Validation (MAECV) values than the AR model.

Table 6. Cross validation mean

Model	RMSECV	MAECV
Auto Regression	4.11	3.51
ARIMA	3.6	2.94

### 3.6 User Interface

After testing the error values of the two models, the ARIMA model was chosen because it produced an error value smaller than the error value obtained from the AR model on almost all tests. Although the error value of the AR model in the 1 week test is smaller in the RMSE test using 938 training data. Overall the ARIMA model gets a smaller error value than the AR model. As well as for grouping data, researchers use daily data grouping.

After getting a suitable model to be implemented into forecasting applications for admission in UMS. We implements the ARIMA model by using the auto\_arma algorithm as an order finder. Auto\_arma also used so the application can adapt to different data. The system model in this application is quite simple, the database is used as a container for training data. And the input received by the system is only the department that will be predicted, and the number of days that will be predicted. The output of the application in the form of a graph that displays predictive data. This website based prediction application was developed using the Python and Django programming

languages as web framework.

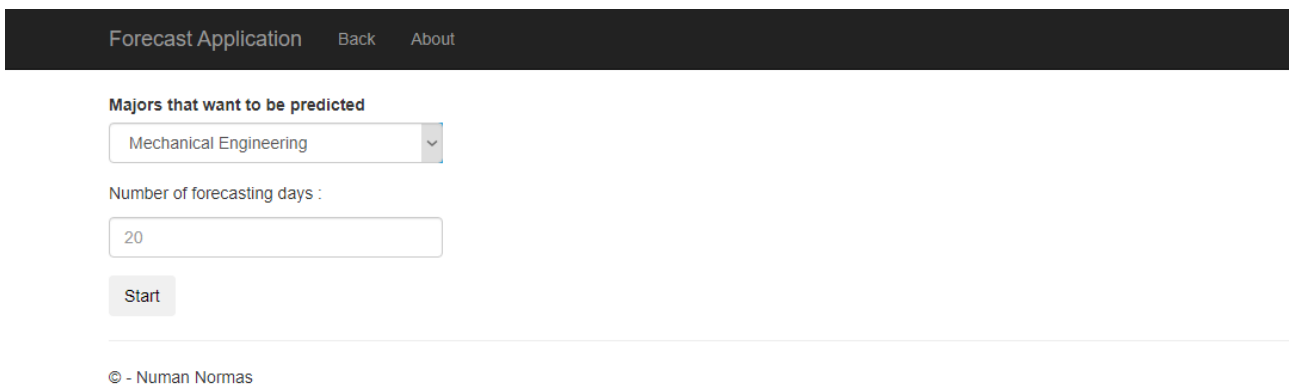


Figure 11. First user interface

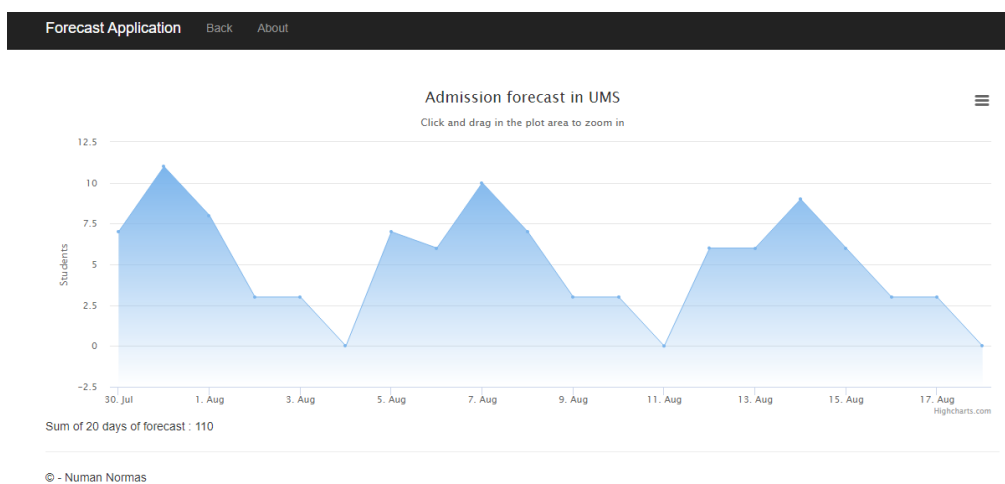


Figure 12. Second user interface

Figure 11, is the initial display of the application that will receive input in the form of majors and the number of days to be predicted, for the display of the research application using Bootstrap as a framework. And in Figure 12, the display of forecast data using Highchart framework as a graphical chart viewer.

#### 4. CONCLUSION

Based on the research. From the ACF and PACF plot we can conclude that the data can be forecast using Time series model. Time series data has seasonality. Forecast with longer period is not suitable with ARIMA nor AR, it's because the ARIMA take value for forecasting is getting from the forecast value. The more data ARIMA need to forecast, more error will be generate. ARIMA model have higher accuracy to be applied into forecasting applications compared to AR models. Because the level of error measurements performed on almost all error tests show ARIMA has a smaller error measurement value than the AR model. Either AR or ARIMA is bad at long-term forecasting, it will just repeating same pattern. Forecasting application development using Django web framework and python as a programming language. Auto\_arima algorithm is used so the application can adapt to different data.

## REFERENCES

- Alexander, J. B., & Rogelio, E. C.-R. (2019). *SCOPE: Selective Cross-Validation over Parameters for Elo*. Salt Lake City: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-19).
- Anggoro, D. A., & Supriyanti, W. (2019). Aplikasi Sistem Pendukung Keputusan dengan Metode AHP untuk Pemilihan Siswa Berprestasi di SMAN Kebakkramat. *Jurnal Penelitian Dan Pengabdian Kepada Masyarakat UNSIQ*, 6(3), 163–171. <https://doi.org/10.32699/ppkm.v6i3.777>
- As'ad, M., Wibowo, S. S., & Sophia, E. (2017). Peramalan Jumlah Mahasiswa Baru dengan Model Autoregressive Integrated Moving Average (Arima). *Jurnal Informatika Merdeka Pasuruan*, 2(3). Retrieved from <https://www.neliti.com/publications/264535/peramalan-jumlah-mahasiswa-baru-dengan-model-autoregressive-integrated-moving-av>
- Azid, A., Juahir, H., Toriman, M. E., Kamarudin, M. K. A., Saudi, A. S. M., Hasnam, C. N. C., ... Yamin, M. (2014). Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques: a Case Study in Malaysia. *Water, Air, & Soil Pollution*, 225(8), 2063. <https://doi.org/10.1007/s11270-014-2063-1>
- Chai, T., Chai, T., & Draxler, R. R. (2014). Ozone health and ecosystem impacts View project Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature. *Geosci. Model Dev*, 7, 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Jain, G., & Mallick, B. (2017). A Study of Time Series Models ARIMA and ETS. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2898968>
- Janah, S. N., Sulandari, W., & Wiyono, S. B. (2014). PENERAPAN MODEL HYBRID ARIMA BACKPROPAGATION UNTUK PERAMALAN HARGA GABAH INDONESIA. *MEDIA STATISTIKA*, 7(2), 63–69. <https://doi.org/10.14710/medstat.7.2.63-69>
- Karmita, S., Putra, A. B. W., Gaffar, A. F. O., & Wiguna, A. S. (2019). Prediksi Jumlah Calon Mahasiswa Baru Menggunakan Fuzzy Time Series-Time Invariant. *Prosiding SAKTI (Seminar Ilmu Komputer Dan Teknologi Informasi)*, 3(1), 208–214. Retrieved from <http://e-journals.unmul.ac.id/index.php/SAKTI/article/view/2093>
- Pandji, B. Y., Indwiarti, I., & Rohmawati, A. A. (2019). Perbandingan Prediksi Harga Saham dengan model ARIMA dan Artificial Neural Network. *Indonesian Journal on Computing (Indo-JC)*, 4(2), 189–198. <https://doi.org/10.21108/INDOJC.2019.4.2.344>
- PRASETYO, G. (2018). Prediksi Harga Saham Sektor Keuangan dan Sektor Infrastruktur di Indonesia dengan Model ARIMA. *Jurnal Ilmiah Mahasiswa FEB*, 6(2). Retrieved from <https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/5035>
- Pravilovic, S., Appice, A., & Malerba, D. (2014). Integrating cluster analysis to the ARIMA model for forecasting geosensor data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8502 LNAI, 234–243. [https://doi.org/10.1007/978-3-319-08326-1\\_24](https://doi.org/10.1007/978-3-319-08326-1_24)
- Purba, A. (2015). PERANCANGAN APLIKASI PERAMALAN JUMLAH CALON MAHASISWA BARU YANG MENDAFTAR MENGGUNAKAN METODE SINGLE EXPONENTIAL SMOOTHING (Studi Kasus : Fakultas Agama Islam UISU). *JURIKOM (Jurnal Riset Komputer)*, 2(6). Retrieved from <https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom/article/view/347>
- Sinaga, E., Sembiring, A. S., & Limbong, R. (2019). PERANCANGAN APLIKASI PREDIKSI JUMLAH KELULUSAN MAHASISWA DENGAN METODE AUTO REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) (STUDI KASUS : PRODI TI STMIK BUDIDARMA MEDAN). *Informasi Dan Teknologi Ilmiah (INTI)*, 13(3), 308–313. Retrieved from <https://ejurnal.stmik-budidarma.ac.id/index.php/inti/article/view/1023>
- Sismi, S., & Darsyah, M. Y. (2018). Perbandingan Prediksi Harga Saham PT.BRI, Tbk dengan METODE ARIMA dan MOVING AVERAGE. *Prosiding Seminar Nasional Mahasiswa Unimus*, 1(0). Retrieved from <http://prosiding.unimus.ac.id/index.php/mahasiswa/article/view/170>
- Surakarta, U. M. (2016). One Day Service – PMB UMS. Retrieved October 3, 2019, from <http://pmb.ums.ac.id/ods/>
- Thamrin, H., & Murray-Smith, D. J. (n.d.). A Mathematical Model of the Human Respiratory Control System during Exercise. *Modelling, Simulation, and Identification / 658: Power and Energy Systems / 660, 661, 662*. ACTA Press.
- Umar, M., & Ismail, F. (2018). Peningkatan Mutu Lembaga Pendidikan Islam (Tinjauan Konsep Mutu Edward Deming dan Joseph Juran). *Jurnal Ilmiah Iqra'*, 11(2). <https://doi.org/10.30984/jii.v11i2.581>

Williams, S. D. P. (2003). The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. *Journal of Geodesy*, 76(9–10), 483–494. <https://doi.org/10.1007/s00190-002-0283-4>