Predicting NYSC Mobilization Period Using ARIMA Model

Ajayi, Olusola O., Adegbite, Adewuyi A., Zidah, John I., Akinrolabu, Olatunde D. & Obi, Ebuka D.
Department of Computer Science
Adekunle Ajasin University
Akungba-Akoko, Ondo State, Nigeria
E-mails: olusola.ajayi@aaua.edu.ng; adewuyi.adegbite@aaua.edu.ng ; jzidah@gmail.com; horlamy007@gmail.com; davidobi023@gmail.com
Phones: +2347056433798; +2347030714447; +2348137956463; +2348064518483; +2348130770663

ABSTRACT

This research paper predicts the period of mobilization of graduated student during their NYSC by detecting the frequent sub sequences in a sequence database and observing the values that a variable takes at a different time. This paper explores the application of Machine Learning for predicting of NYSC mobilization period. Numerous factors or factors affect the prediction of the time series such as the pre-mobilization period, the year, batch and registration period served as the variable used for predicting. The data for the study was obtained from various website like www.nairaland.com, www.portal.nysc.org.ng, www.nigeriaschool.com.ng, www.myschoolgist.com, www.myschool.ng, www.hotnigerianjobs.com, www.corpr.com.ng, www.myjobmag.com, www.aaua.edu.ng, www.ngcareers.com, Preprocessing steps were adopted to convert the data into a usable form. After this, the Autoregressive Integrated Moving Average (ARIMA) model was built, the data were split into training and testing data and the model was evaluated giving an accuracy of 86.8%.

Keywords: NYSC, Posting, Mobilization, Registration, ARIMA Model, Time Series, Batch, Graduating Students.

CISDI Journal Reference Format
Ajayi, O.O., Adegbite, A. A., Zidah, J.I., Akinrolabu, O.D. & Obi, E.D. (2022): Predicting NYSC Mobilization Period Using ARIMA Model. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 13 No 4, Pp 13-.28
Available online at www.isteams.net/cisdijournal.
dx.doi.org/10.22624/AIMS/CISDI/V13N4P2

1. BACKGROUND TO THE STUDY

The NYSC was established in 1973 as a post-civil-war strategy in Nigeria, with a view to promote the development of common ties among the youths of Nigeria and the promotion of national unity. Despite the fact that the NYSC was established in Nigeria's military regime as a strategy to ensure reconstruction, rehabilitation, and reconciliation of Nigeria's post-war economy, it remains relevant and an apex youth organization in Nigeria's democratic era, that enforces the policies of national development through the mobilization, deployment, and development of youth capabilities (NYSC, 1999, 2008; Bodley-Bond and Cronin, 2013). As the joint second longest standing national youth programme in sub-Saharan Africa, the relevance of the NSYC scheme cannot be overemphasized as it stands a critical platform that links educated Nigerian youths to places of employment (Bodley-Bond and Cronin, 2013).

The mobilization of eligible youth participants in the NYSC scheme is carried out by the National Directorate while the State Committees under the authority of State Coordinators are responsible for deploying mobilized corps members to places of critical national and community development needs (Olutola, 1979; Marenin, 1989b, 1990). At the grassroots level, the operation of the scheme is supervised by the State Zonal Officers and the Local Government Inspectors (NYSC, 1999). On average, the NYSC programme mobilizes an average of 250,000 graduates for national service and community development (FMYD, 2013a). Mobilization of corps members is guided by two policies: deployment and posting policy. This research is to predict the period of mobilization of graduated student during their NYSC by detecting the frequent sub sequences in a sequence database and observing the values that a variable takes at a different time.

Statement of the Problem
Studies have shown that time forecasting have contiguous observations, such has one observation each hour, day, month or years. A time series where the observations are not uniform over time may be described as discontiguous. The lack of uniformity or the observations may be caused by missing or corrupt values. The demands and yearnings of graduating youths to meet up with service year as at when due, cannot be over-emphasized. However, many of these expectations fall short of fulfillment because of the unpredictability of NYSC posting period and its non-alignment with the academic calendar of higher institutions of learning.

Objectives
The following objectives was achieved:

i. Extensive reviews of related literature;
ii. Extraction of 10years NYSC mobilization period data;
iii. Evaluation of collated data using ARIMA model.

2. THE ARIMA MODEL

The ARIMA model is a conventional time series model that was proposed by Box and Jenkins in 1970 and is referred to as an ARIMA model, a Box-Jenkins model or a B-J model. It consists of three basic model forms: (1) the Moving Average model MA(q), in which q is the order of the moving average; (2) an autoregressive model AR(p), in which p is the order of the autoregression; and (3) the autoregression Moving Average model ARMA(p,q). The first two model forms are special cases of the third model form; that is, MA(q) can be viewed as ARMA(0,q) and AR(p) can be viewed as ARMA(p,0). If the difference in the surveillance series constitutes ARMA(p,q), the series model is referred to as the ARIMA(p,d,q) model.

The ARIMA model has been widely used in econometrics. Over the past several years, it has also attracted more attention for the surveillance and early warning of infectious disease. In practice, the ARIMA model is used to fit surveillance series data {yt}, calculate the corresponding expected value and early warning limit, and then determine whether an alarm should be triggered. The following steps are used to build this model:

Step 1: Determine and handle series stationarity. The B-J model requires that the surveillance series {yt} is a stationary series; that is, the mean and variance for the series do not vary with time. In other words, the mean of the series is constant at all time points, and the covariance of the surveillance data at two time points is only related to the length of time. If the mean of the surveillance series is not a constant value, data should be used to transform the non-stationary series into a stationary series. For a series in which the mean is not stationary, the difference method can be used. For a series in which the variance is not stationary, square root transformations and other methods can be used.

Step 2: Identify and estimate the model. Series characteristics are needed to identify orders p, q, and d for the model. Indicators that are used to identify these orders include the autocorrelation (ACF), partial autocorrelation (PACF) and inverse autocorrelation functions (IACF). The order q for autoregression can be identified through PACF and IACF; and the order p for the Moving Average can be identified through ACF. After the orders are identified, the model parameters can be estimated using the nonlinear least square or maximum likelihood methods.

Step 3: Diagnose the model. After estimating the parameters, the model should be further diagnosed and can only be used with proven rationality. In fact, model diagnosis conducts a series of tests for the residual error. The residual error of a "satisfactory" model should have the following characteristics: (1) The mean of the residual error should not significantly differ from 0. A t-test can be used to test this assumption; (2) The residual error should be subject to a normal distribution, which can be evaluated with a Kolmogorov-Smirnov test; and (3) The residual error series cannot be autocorrelated, which can be determined with a Box-Ljung test.

Once a rational early warning model is constructed according to the above three steps, the model can calculate the expected value and confidence limit. The confidence limit is based on a normal distribution. Early warning is determined by comparing the expected value and upper limit of the confidence interval. Currently, many standard statistical software applications are can be used to calculate the ARIMA model, such as SAS, Stata, Python and Minitab.

Watier et al. (1991) used the Seasonal Autoregressive Integrated Moving Average model to analyze the bio surveillance of salmonellosis . The following expected model was constructed:

(3.28)εεyt−yt−12=αyt−1−yt−13+εt+βεt−12 (1)

In the equation, y_t is the number of cases at time t, and ε_t is the seasonal difference random disturbance term at time t. Before fitting a model, it is important to first identify the historical outbreaks that are included in the surveillance series, and the non-outbreak expected value can be used to replace the monitored value during an outbreak to eliminate the impact of the historical outbreak on the model. Then, seasonal fluctuations can be adjusted, the prediction limit is calculated according to one-step forecasting, and the upper limit is used as the early warning threshold. In 2007, Feng et al. used the ARIMA model to analyze the monthly incidence data for modifiable reports of infectious diseases in Mainland China from 1995 to 2004 (Feng et al., 2007) and constructed an ARIMA time series model. After identifying, fitting, and testing the model, the ARIMA $(0,1,0)(0,1,0)_{12}$ model was selected as the optimal model, in which the variance forfitting the residual error is 2.28 and the mean absolute error of extrapolation and prediction is 0.34.

Although the ARIMA model is highly promising, there are several challenges in applying this method in CIDARS. For example, the steps to build this method are very complex, as there is a need to identify and diagnose the stationarity, order, and residual error of the series. Additionally, the robustness of the model is poor and significantly restricts the application of this method in CIDARS.

Conceptual Model

The conceptual model is presented below

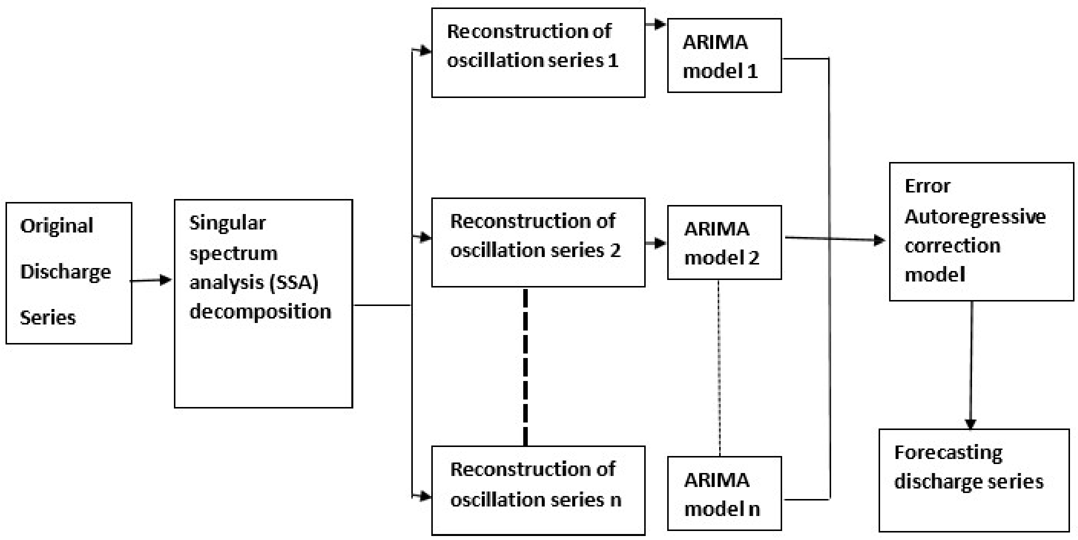


Figure 1: Conceptual model for ARIMA model

3. METHODICAL DETAILS

The model is a conventional time series model that was proposed by Box and Jenkins in 1970 and is referred to as an ARIMA model. Sequential pattern mining is a data mining in which a frequent pattern of sequence is generated from the database. The following procedures are followed to ensure the successful implementation of the work:

i. **Data Collection:** Data collection as part of the methodology for this study entails going online and surfing the internet to get/download the past NYSC Mobilization time table from various website. After getting the time table of the NYSC mobilization period was extracted the predictive attributes are further extracted:
    i. **Pre-Mobilization Period:** this period the NYSC board Collate and store Senate/Academic Board approved result lists from Corps Producing Institutions. Vetting/screening of the uploaded data of the Prospective Corps Members (PCM). Collation of the Signature Specimens from Corps Producing Institutions
    ii. **Registration Period:** this the period where by all prospective corps whose names appear in the Senate/Academic Board Approved Result lists submitted by their Institutions enroll for registration on the NYSC portal.
    iii. **Post Mobilization Period:** this is the period where prospective corps member after undergoing orientation programme are posted to their various Primary Place of Assignment (PPA).

Table 1 below shows the component extracted which are: Pre-Mobilization Period, Registration Period and Post Mobilization Period

Table 1: Extracted Data.

S/N	Year	Batch	Pre-Mobilization	Registration Period	Post Mobilization
1.	2009	B	15th – 16th April, 2009	4th – 15th June, 2009	4th – 6th August 2009
2.	2010	A	19th Nov, 2009	1st Dec 2009 – 8th Jan, 2010	30th March, 2010
3.	2010	B	30th March, 2010	12th April -19th May, 2010	27th July, 2010
4.	2010	C	10th August, 2010	16th Aug -15th Sept, 2010	16th November 2010
5.	2011	A	25th November, 2010	1st Dec 2010 - 25th Feb, 2011	30th March 2011
6.	2011	B	12th April, 2011	1st April -14th June,2011	26th July, 2011
7.	2011	C	9th - 10th August, 2011	15th August- 14th October, 2011	6th December 2011
8.	2012	A	13th December, 2011	12th Dec, 2011- 15th Feb, 2012	27th March 2012
9.	2012	B	17th to 20th April 2012.	2nd to 4th June 2012	30th July 2012
10.	2012	C	4th - 6th Sept, 2012	1st August- 5th October, 2012	27th November 2012
11.	2013	A	16th January, 2013	10th Dec, 2012- 26th Feb, 2013	26th March 2013
12.	2013	B	N/A	25th June, 2013	6th August 2013
13.	2013	C	N/A	5th November, 2013	26th November, 2013
14.	2014	A	N/A	9th Dec, 2013 - 14th Feb, 2014	1st April, 2014
15.	2014	B	6th - 9th June, 2014	14th April, 2014 - 11th July, 2014	26th August, 2014
16.	2014	C	N/A	20th August – 12th October, 2014	25th November, 2014
17.	2015	A	N/A	2nd March - 4th April, 2015	26th May, 2015
18.	2015	B	21st – 24th July 2015	8th September – 9th October 2015	16th Nov.2015 (stream 1) 14th Dec, 2015 (stream II)
19.	2016	A	5th - 9th January, 2016	8th - 28th February, 2016	13th May, 2016 (Stream 1) 6th June, 2016 (Stream II)
20.	2016	B	N/A	17th October – 12th November 2016	14th Dec, 2016 Stream I Jan, 2017 Stream II
21.	2017	A	N/A	17th April - 12th May 2017	14th Jun, 2017. (Stream I) 13th Jul, 2017 (Stream II)
22.	2017	B	N/A	23rd Oct to 12th Nov 2017	N/A
23.	2018	A	5th - 9th March 2018	26th March - 8th April 2018	N/A
24.	2018	B	N/A	2nd July, 2018.	N/A
25.	2018	C	N/A	23rd Oct -12th Nov.2018	N/A
26.	2019	A	N/A	4th – 19th March 2019	16th April, 2019
27.	2019	B	14th – 18th May 2019	1st June – 12th June 2019	9th July 2019 (stream I) 10th Sept (stream II)
28.	2019	C	7th - 12th October, 2019	21st October - 1st November, 2019	26th Nov 2019 (stream I) 19th Dec 2019 (stream II)

ii. **Data Preparation:** Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. The dataset gather from the internet is inconsistent, and/or lacking in certain behaviors or trends, therefore it needs to preprocess in order to carry out an accurate prediction.
The follow steps were followed in the pre-processing of the data
Step1: gather the data
Step2: the data is transformed by generating the weeks differences from the dates, this is done by calculating the amount of it takes before the next pre-mobilization period and registration period.
Step3: check for missing values, then the 'replacing with mean value' method is usedto handle the missing values
Step4: the data is then stored in Microsoft Excel Spreadsheet which will be imported into our python code. The table below shows the days differences between the dates.

Table 2: Weeks Difference.

S/N	Year	Batch	Weeks Difference (PM)	Weeks Difference (RP)
1.	2009	B	0	0
2.	2010	A	31.3	25.8
3.	2010	B	18.9	19
4.	2010	C	19.1	18.1
5.	2011	A	15.4	15.4
6.	2011	B	19.8	17.4
7.	2011	C	17.1	19.5
8.	2012	A	18.1	17.1
9.	2012	B	18.1	24.8
10.	2012	C	20.1	8.7
11.	2013	A	19.2	18.8
12.	2013	B	20.3	28.2
13.	2013	C	21.3	19.1
14.	2014	A	No Data	5
15.	2014	B	18.4	18.1
16.	2014	C	16.5	18.4
17.	2015	A	15.6	27.8
18.	2015	B	15.7	27.2
19.	2016	A	20.3	22
20.	2016	B	18.4	36.1
21.	2017	A	19.3	26.1
22.	2017	B	15.3	27.1
23.	2018	A	20.3	22.1
24.	2018	B	No Data	14.1
25.	2018	C	18.2	20.5
26.	2019	A	No Data	19
27.	2019	B	62.2	12.8
28.	2019	C	21	24.8

jupyter Untitled Last Checkpoint: Last Monday at 2:14 AM (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Run Code

```
In [2]: import pylab
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import sklearn
        from pandas import datetime
        from pandas import Series
        from pandas.plotting import lag_plot
        from pandas import datetime
        from matplotlib import pyplot
        from statsmodels.tsa.arima_model import ARIMA
        from sklearn.metrics import mean_squared_error
        from hmmlearn.hmm import GaussianHMM

In [3]: data = pd.read_csv('mydata.csv')
        b = data.replace(np.NaN, data['PMWD'].mean()).fillna(0)
        b.head()

Out[3]:
```

	S/N	Year	Batch	PMWD	RPWD
0	1	2009	B	0.0	0.0
1	2	2010	A	31.3	25.8
2	3	2010	B	18.9	19.0
3	4	2010	C	19.1	18.1
4	5	2011	A	15.4	15.4

Figure 2: Source Code for Processing the Data

Data Analysis: This process we use sequential pattern mining for analyzing the sequential time, to discover statistically relevant pattern of the pre-mobilization dates and registration dates, and then making the data stationary by applying differencing and using the lags of the data to predict the future of outcomes, thus ARIMA model is applied following the architecture as shown in Figure 5.

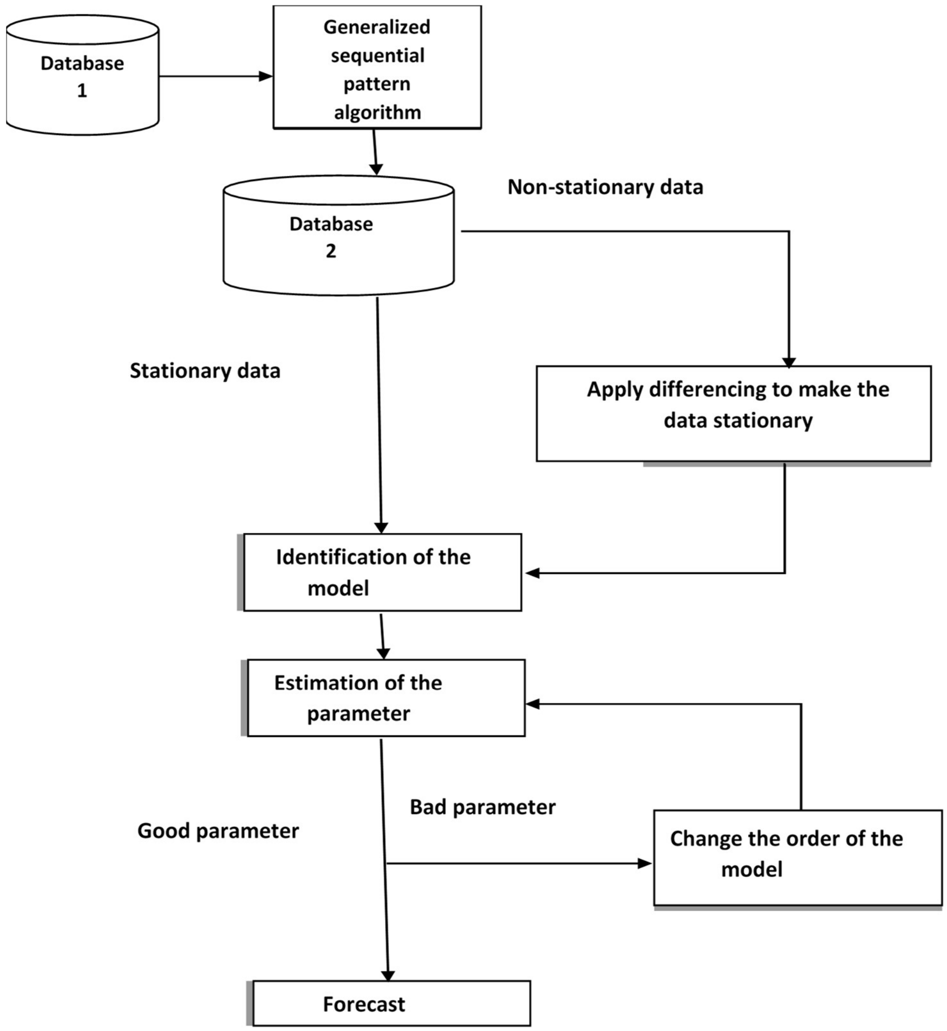


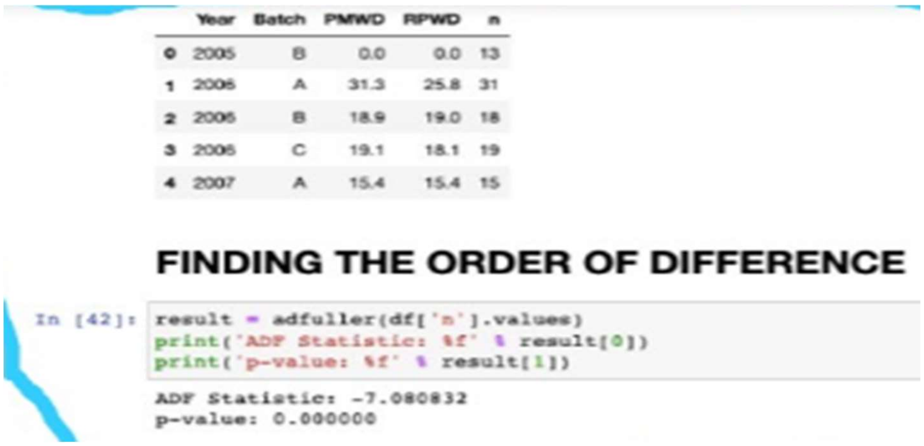
Figure 3: Detailed Architecture

4. IMPLEMENTATION

The model was implemented on Window 10 pro OS, Intel core i5 processor, 2.4Ghz, 4GB RAM using Python programming language on Jupyter Notebook IDE (Integrated Development Environment).

4.1 Implementation Phase Autoregressive Integrated Moving Average

The first step in building the ARIMA model is by checking if the data is not stationary. We do this by running nd ADF test to check for the p-value, if the p-value of the test is less than the significance level (0.05) then you reject the null hypothesis and infer that the time series is indeed stationary. So, in our case, our P-Value < 0.05 therefore the data is indeed stationary. Since P-value is lesser than the significance level, we will need to determine the value of d (Differencing required)



	Year	Batch	PMWD	RPWD	n
0	2005	B	0.0	0.0	13
1	2006	A	31.3	25.8	31
2	2006	B	18.9	19.0	18
3	2006	C	19.1	18.1	19
4	2007	A	15.4	15.4	15

FINDING THE ORDER OF DIFFERENCE

```
In [42]: result = adfuller(df['n'].values)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])

ADF Statistic: -7.080832
p-value: 0.000000
```

Figure 4: Finding Order of Difference

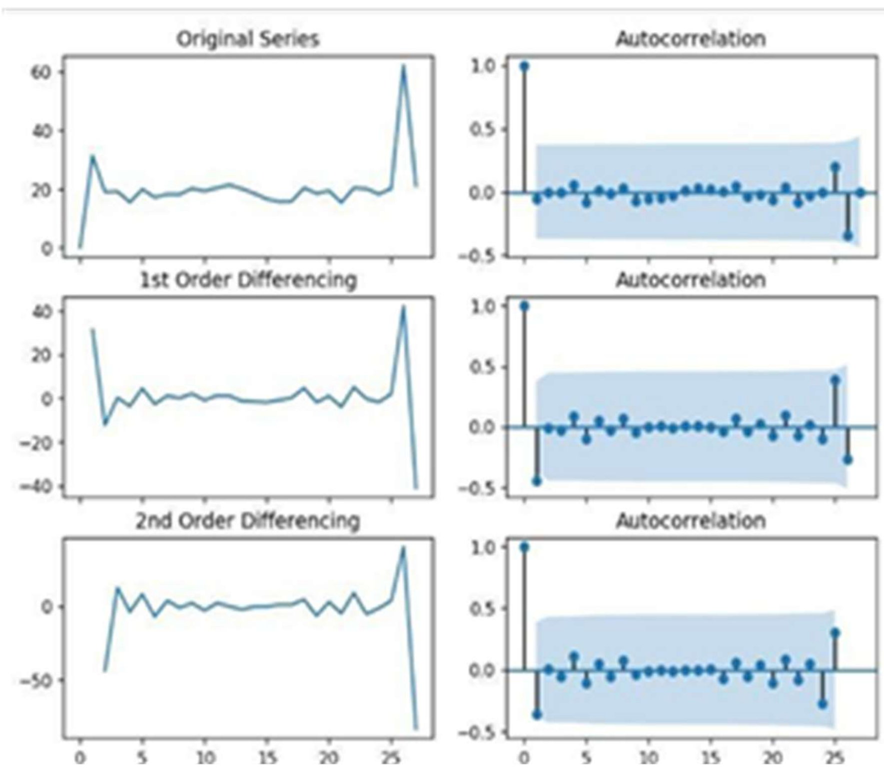


Figure 5 Differencing plot

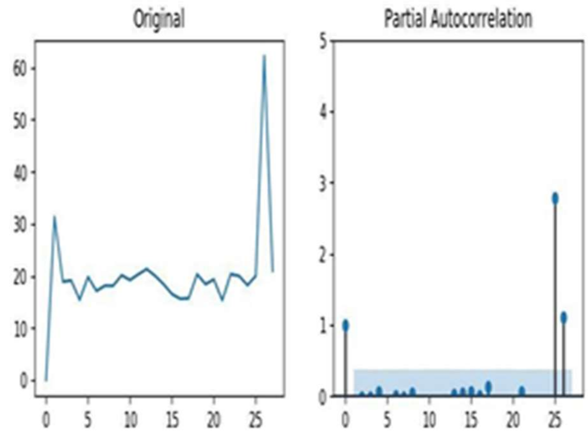


Figure 6 Partial Autocorrelation Graph

For the above series, the time series reaches stationarity with two orders of differencing. But on looking at the autocorrelation plot for the 2nd differencing the lag goes into the far negative zone fairly quick, which indicates, the series might have been over differenced so, we are going to tentatively fix the order of differencing as 1 even though the series is not perfectly stationary (weak stationarity).

4.2 Identification Of Model

We need to the identify the number of AR and MA needed for the model to identify if the model needs any AR terms. We have to plot inspecting the Partial Autocorrelation (PACF) graph to find out the required number of AR terms needed for the model. It is observed that the PACF lag 1 is quite significant since is well above the significance line, slightly managing to cross the significance limit (blue region). So we are going to be conservative and tentatively fix the p as 1. Just like how we looked at the PACF plot for the number of AR terms, we will look at the ACF plot for the number of MA terms. An MA term is technically, the error of the lagged forecast.

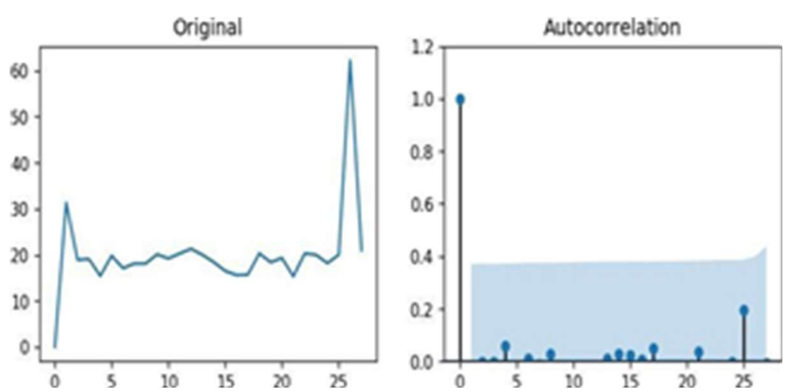


Figure 7: Auto-correlation Graph

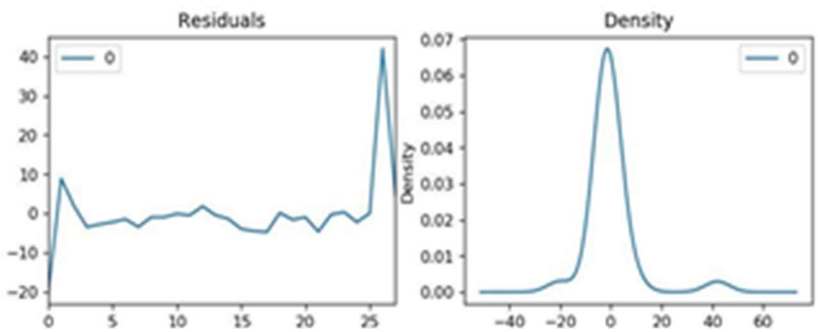


Figure 8: Residuals Graph

Just one lag is well above the significance line. So, we tentatively fix q as 1. When in doubt, go with the simpler model that sufficiently explains the Y.

4.3 Building The ARIMA Model

Since we have determined the values of p, d and q, we have everything needed to fit the ARIMA model. We use the ARIMA implementation in stats models package and the plot the residuals to ensure there are no patterns (that is, look for constant mean and variance). The residual errors seem fine with near zero mean and uniform variance. The next step is to plot the actuals against the fitted values using plot predict.

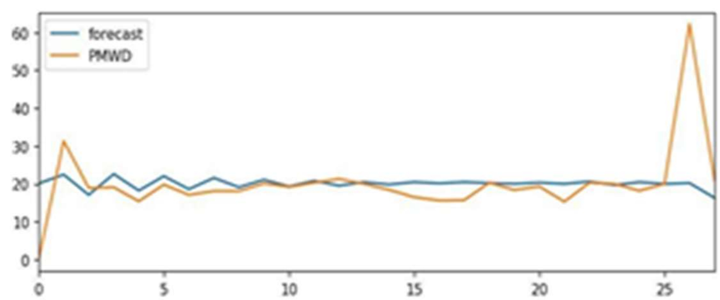


Figure 9: Actual vs Fitted

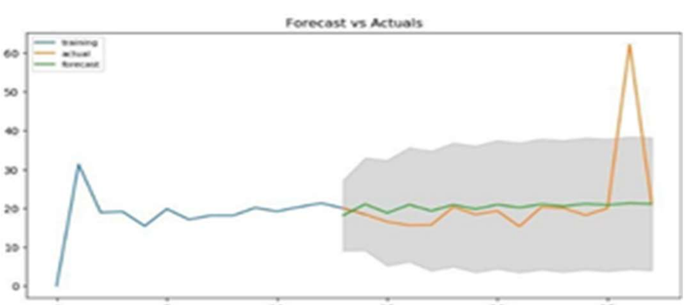


Figure 10: Reverse Future vs. Actual

Since the predicting forecast is above the actual therefore, we have a good model.

4.4 Model Summary

Table 4: Model Summary

ARMA Model Results

Dep. Variable: 28 Model:	**PMWD ARMA(1, 1)**	**No. Observations: Log Likelihood**	**-102.051**
Method:	css-mle	S.D. of innovations	9.244
Date:	Mon, 20 Jan 2020	AIC	212.103
Time:	05:36:23	BIC	217.431
Sample:	0	HQIC	213.732

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Const	20.1047	1.675	12.001	0.000	16.821	23.388
ar.L1.PMWD	-0.8873	0.239	-3.720	0.001	-1.355	-0.420
ma.L1.PMWD	0.7991	0.267	2.989	0.006	0.275	1.323

Roots

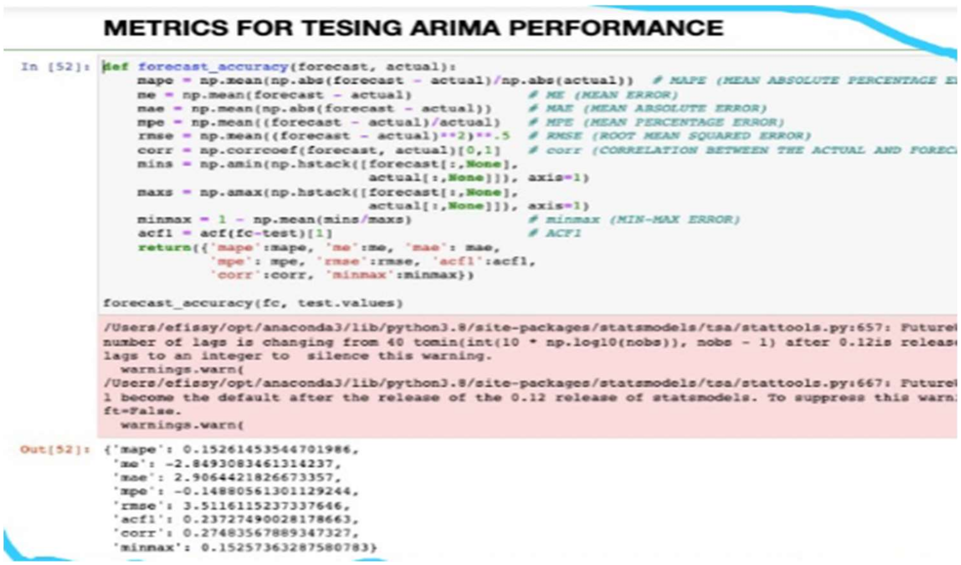
	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	-1.1270	+0.0000j	1.1270	0.5000
MA.1	-1.2515	+0.0000j	1.2515	0.5000

It is seen that the Akaike Information Criterion (AIC) value is a 212.103 and the Bayesian Information Criterion (BIC) is 217.431, this has a minimum. With the AR and MA been less than 0.005, it implies that the model is a good one.

4.5 Model Metrics

The commonly used accuracy metrics to judge forecasts are: Mean Absolute Percentage Error (MAPE), Mean Error (ME), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Root Mean Squared Error (RMSE), Lag 1. Autocorrelation of Error (ACF1), Correlation between the Actual and the Forecast (corr) and Min-Max Error (minmax).

Computing, Information Systems & Development Informatics Journal
Vol. 13 No. 4, December 2022 - www.isteams.net/cisdijournal



METRICS FOR TESING ARIMA PERFORMANCE

```
In [52]: def forecast_accuracy(forecast, actual):
             mape = np.mean(np.abs(forecast - actual)/np.abs(actual))  # MAPE (MEAN ABSOLUTE PERCENTAGE E
             me = np.mean(forecast - actual)             # ME (MEAN ERROR)
             mae = np.mean(np.abs(forecast - actual))    # MAE (MEAN ABSOLUTE ERROR)
             mpe = np.mean((forecast - actual)/actual)   # MPE (MEAN PERCENTAGE ERROR)
             rmse = np.mean((forecast - actual)**2)**.5  # RMSE (ROOT MEAN SQUARED ERROR)
             corr = np.corrcoef(forecast, actual)[0,1]   # corr (CORRELATION BETWEEN THE ACTUAL AND FOREC
             mins = np.amin(np.hstack([forecast[:,None],
                                       actual[:,None]]), axis=1)
             maxs = np.amax(np.hstack([forecast[:,None],
                                       actual[:,None]]), axis=1)
             minmax = 1 - np.mean(mins/maxs)             # minmax (MIN-MAX ERROR)
             acf1 = acf(fc-test)[1]                      # ACF1
             return({'mape':mape, 'me':me, 'mae': mae,
                     'mpe': mpe, 'rmse':rmse, 'acf1':acf1,
                     'corr':corr, 'minmax':minmax})

         forecast_accuracy(fc, test.values)

         /Users/efissy/opt/anaconda3/lib/python3.8/site-packages/statsmodels/tsa/stattools.py:657: Future
         number of lags is changing from 40 tomin(int(10 * np.log10(nobs)), nobs - 1) after 0.12is releas
         lags to an integer to  silence this warning.
           warnings.warn(
         /Users/efissy/opt/anaconda3/lib/python3.8/site-packages/statsmodels/tsa/stattools.py:667: Future
         1 become the default after the release of the 0.12 release of statsmodels. To suppress this warn
         ft=False.
           warnings.warn(

Out[52]: {'mape': 0.15261453544701986,
          'me': -2.8493083461314237,
          'mae': 2.9064421826673357,
          'mpe': -0.14880561301129244,
          'rmse': 3.5116115237337646,
          'acf1': 0.23727490028178663,
          'corr': 0.27483567889347327,
          'minmax': 0.15257363287580783}
```

Figure 11:Reverse Future vs. Actual

The Mean Absolute Percentage Error (MAPE) is around 13.2% which implies the model is 86.8%.

5. INTERPRETATION AND DISCUSSION OF RESULT

In this study, the ARIMA (1,0,1) was the best candidate model selected for making predictions for up to 5 years for NYSC Mobilization period using a 10 years' time series data. ARIMA was used for the reasons of its capabilities to make predictions using a time series data with any kind of pattern and with autocorrelations between the successive values in the time series. The study also statistically tested and validated that the successive residuals (forecast errors) in the fitted ARIMA time series were not correlated, and the residuals seem to be normally distributed with mean zero and constant variance. Hence, we can conclude that the selected ARIMA(1,0,1) seem to provide an adequate predictive model for the NYSC mobilization period. While training and testing dataset by splitting the time series into 2 contiguous parts and Out-of-Time cross-validation which take few steps back in time and forecast into the future to as many steps you took back, all the figures show good results, i.e. the ARIMA model generates an outputs with an error of 13.2% which gives an overall prediction accuracy of 86.8%. The model is therefore valid.

6. SUMMARY

The ability to predict NYSC Mobilization period would be of great importance and benefit to Nigerian Institutions. Knowing the next mobilization period has often been a great deal to educational body, with the Implementation of machine learning using Autoregressive Integrated Moving Average (ARIMA) in predicting the future outcome of the Time series, identifying the number of AR and MA needed for the model and removal of seasonality from Time series is also useful in getting the accurate result needed. The implementation of the work was carried out using PYTHON programming language. The aim of this research was to predict the next NYSC mobilization period based on the past date at which the mobilization period started.

7. CONCLUSION AND RECOMMENDATION

This research work shows how reliable the ARIMA model is, considering the performance (accuracy score) of the model on the series of data gathered. The results of the study also show that, ARIMA model is an optimized approach for time series prediction. The research and output evaluated shows that ARIMA is resource intensive in terms of computational time and memory. This model was developed to assist institutions in Nigeria to meet up and get ready for the next mobilization period of NYSC using sequential pattern mining and time series analysis using ARIMA model with the intention to get the best result.

8. CONTRIBUTION TO KNOWLEDGE

Autoregressive Integrated Moving Average approach for increased efficiency and effectiveness in time series analysis to provide better results in predicting the possible outcomes.

REFERENCES

1. Aileen P. Wright, Adam T. Wright, Allison B. McCoy, Dean F. Sittig Journal of Biomedical Informatics (2015) The use of sequential pattern mining to predict next prescribed medications
2. C. Arden Pope III, Douglas W. Dockery , John D. Spengler , and Mark E. Raizenne vol. 144, No. 3_pt_1 Sep 01, 1991 Respiratory Health and PM10 Pollution: A Daily Time Series Analysis Daniel Pefia “Forecasting Growth with Time Series Models”
3. David E.Rose Journal of Econometrics Volume 5, Issue 3, May 1977, Pages323-345 Forecasting aggregates of independent Arima processes
4. Dr.IlangoVelchamy, Dr.UmaIlango, Nitya Ramesh (2015) “TIME SERIES DATA MINING RESEARCH PROBLEM, ISSUES, MODELS, TRENDS AND TOOLS” Volume IX,
5. Ette Harrison Etuk and Tariq Mahgoub Mohamed (2014) “Full Length Research Paper Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan.”
6. Fallaw Sowell Journal of Econometrics Volume 53, Issues 1–3, July–September 1992, Pages 165- 188 Maximum likelihood estimation of stationary univariate fractionally integrated time series models
7. Fallaw Sowell Journal of Monetary Economics Volume 29, Issue 2, April 1992, Pages 277-302 Modeling long-run behavior with the fractional ARIMA model
8. Florent Masseglia, Pascal Poncelet, Maguelonne Teisseire Data & Knowledge Engineering
9. G TOULOUMI, S J POCOCK, K KATSOUYANNI, D TRICHOPOULOS International Journal of Epidemiology, Volume 23, Issue 5, October 1994, Pages 957–967, Short-Term Effects of Air Pollution on Daily Mortality in Athens: A Time-Series Analysis
10. G. Peter Zhang (2003) “Time series forecasting using a hybrid ARIMA and neural network model”.
11. Guo-Cheng Lan Tzung-Pei Hong Vincent S.Tseng Shyue-Liang Wang Expert Systems with Applications Volume 41, Issue 11, 1 September 2014, Pages 5071-5081 Applying the maximum utility measure in high utility sequential pattern mining
12. Helen Lingard, Valerie Francis, Michelle Turner “Work-life strategies in the Australian construction industry: Implementation issues in a dynamic project-based work environment”
13. Hirotugu Akaike Mathematics in Science and Engineering Volume 126, 1976, Pages 27-96 Canonical Correlation Analysis of Time Series and the Use of an Information Criterion

14. Hui Liu Hong-qi Tian Yan-fei Li Applied Energy Volume 98, October 2012, Pages 415-424 Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction
15. Ihueze Chukwutoo Christopher and Okafor Emeka Christian (July 2010). “Multivariate Time Series Analysis for Optimum Production Forecast: A Case Study of 7up Soft Drink Company in Nigeria.”
16. Joe MARTIN, Theresa Keoughan BURROWS and Ian PEGG, United Kingdom “Predicting Construction Duration of Building Projects”
17. Joong Hyuk Chang Knowledge-Based Systems Volume 24, Issue 1, February 2011, Pages 1-9 Mining weighted sequential patterns in a sequence database with a time-interval weight
18. Judy Kay, Nicolas Maisonneuve, and OsmarZaïane “Mining patterns of events in students’ teamwork data
19. K. A. Al Mamun &H. K. Nath Export-led growth in Bangladesh: a time series analysis
20. Lingling Zhou, Ping Zhoa and Hao Haung (2018) “Time series for forecasting the number of new admission inpatient.”
21. M Zakria and Faqir Muhammad (2009) “FORECASTING THE POPULATION OF PAKISTAN USING ARIMA MODELS
22. M. Mudelsee · K. Stattegger (1997) “Exploring the structure of the mid-Pleistocene revolution with advanced methods of time-series analysis”
23. MARC SAEZ, JORDI SUNYER, JORDI CASTELLSAGUÉ, CARLES MURILLO, JOSEP M ANTÓ International Journal of Epidemiology, Volume 24, Issue 3, June 1995, Pages 576–582, Relationship between Weather Temperature and Mortality: A Time Series Analysis Approach in Barcelona
24. Martin Casdagli Physica D: Nonlinear Phenomena Volume 35, Issue 3, May 1989, Pages 335- 356Nonlinear prediction of chaotic time series
25. Maxim Dunaev, Konstantin Zaytsev, Mikhail Titov (2018) “A study of sequential pattern mining algorithms for use in detection of user activity patterns”
26. Ms. Pooja Agrawal, Mr. Suresh kashyap , Mr.Vikas Chandra Pandey, Mr. Suraj Prasad Keshri (2013) “An Analytical Study on Sequential Pattern Mining With Progressive Database” volume1
27. Nizar R. Mabroukeh, C.I Ezeife ACM Computing Surveys (CSUR) December 2010 article No.: 3 A taxonomy of sequential pattern mining algorithm
28. Philippe Fournier-Viger, Jerry Chun-Wei Lin, Rage UdayKiran, Yun Sing Koh and Rincy Thomas (2017) “A Survey of Sequential Pattern Mining” volume 1
29. Prof. Kirti S. Patil. (2014) “Overview of Sequential PatternMining Algorithms” volume 3.
30. S. M. Pincus, and A. L. Goldberger 01 APR 1994https://doi.org/10.1152/ajpheart.1994.266.4.H1643 Physiological time-series analysis: what does regularity quantify?
31. Sadok Rezig, Zied Achour and Nidhal Rezg.(2018) “Using Data Mining Methods for Predicting Sequential Maintenance Activities”
32. Samuel Erasmus Alnaa and Ferdinand Ahiakpor (May 2011). ‘ARIMA (autoregressive integrated moving average) approach to predicting inflation in Ghana.”
33. Stéphanie Jacquemont, François Jacquenet and Marc Sebban (2009). “Mining probabilistic automata: a statistical view of sequential pattern mining.”
34. STEVEN M. PINCUS AND ARY L. GOLDBERGER “Physiological time-series analysis”
35. Thomas E. MaCurdy Journal of Econometrics Volume 18, Issue 1, January 1982, Pages 83-114 The use of time series processes to model the error structure of earnings in a longitudinal data analysis
36. Vincent S.Tseng Chao-Hui Lee Expert Systems with Applications Volume 36, Issue 5, July 2009, Pages 9524-9532 Effective temporal data classification by integrating sequential pattern mining and probabilistic induction

37. Volume 46, Issue 1, July 2003, Pages 97-121 Incremental mining of sequential patterns in large databases
38. Yen-Liang Chen Mei-Ching Chiang Ming-TatKo Expert Systems with Applications Volume 25, Issue 3, October 2003, Pages 343-354 Discovering time-interval sequential patterns in sequence databases
39. Yen-Liang Chen Ya-Han Hu Decision Support Systems Volume 42, Issue 2, November 2006, Pages 1203-1215Constraint-based sequential pattern mining: The consideration of recency and compactness
40. Yi-Chung Hu Gwo-Hshiung Tzeng Chin-Mi Chen Information Sciences Volume 159, Issues 1– 2, 20 January 2004, Pages 69-86 Deriving two-stage learning sequences from knowledge in fuzzy sequential pattern mining