

Time Series Analysis for Understanding the Vaccination Rate using ARIMA

Amulya Maitre
Department of Computer
Engineering
Pimpri Chinchwad College of
Engineering, Akurdi
Pune, India

Dr. K. Rajeswari
Department of Computer
Engineering
Pimpri Chinchwad College of
Engineering, Akurdi
Pune, India

Prof. Sushma Vispute
Department of Computer
Engineering
Pimpri Chinchwad College of
Engineering, Akurdi
Pune, India

Abstract - The pandemic of the novel Coronavirus has led to a devastating situation all around the globe. With the introduction of vaccines by various pharmaceutical companies, there is a hope of recovery from this dreadful condition. In this work, we are using a Machine Learning (ML) approach, based on time series analysis called ARIMA to forecast vaccination rates in various countries. With experimentation, we are able to conclude the forecast of probable increase or decrease in vaccination rates for developed, developing, and underdeveloped countries. This approach will be helpful to apprehend which country needs more attention in terms of vaccine supply or creating awareness amongst its citizens.

Keywords - ARIMA, Covid-19, Machine Learning, Time Series Analysis.

I. INTRODUCTION

The COVID-19 pandemic originated in China at the end of 2019. By January 2020, the virus affected almost a billion people due to irresponsibility and gradually gained the status of COVID-19 as a world-upending pandemic.

Since the COVID-19 pandemic was termed unfavorable for humankind, the only ray of hope was to manufacture a cure for this wildly spreading viral disease at the earliest. Scientific and medical communities quickly began their research and development processes and successfully came up with vaccines in a record-breaking short period.

But not until early 2021, the vaccines were made available to all. With the rapid development, many pharmaceutical companies have come up with their vaccine brands, approved by the World Health Organization (WHO). Although vaccine production is moving in great swing, we cannot help but notice the lack of awareness in people around the globe about the benefits of these vaccines, looking at the rate of vaccinations done. Another possibility could be that some countries haven't been supplied with the vaccines yet.

In this study, we intend to understand the trends of vaccinations done in countries around the world. We categorize these countries as developed, developing and under-developed, to summarize our observations. Using the time series analysis approach, we can predict the rate of

vaccinations for countries in the upcoming period. This will help us analyze the citizens' response to vaccination and indicate the awareness amongst the citizens of that particular country. The analysis could also be an indicator for vaccine supplies in the countries.

The paper further walks through various sections. Section I introduces the enigma, and section II illustrates the prior research done in this field. Section III provides the data set description and methodology, followed by the results and discussion in Section IV. The conclusion of this study is in Section V.

II. LITERATURE REVIEW

Time series analysis is a specialized area of statistics that focuses on time-series data and analysis of trends over time. This kind of data could be in intervals or a series of particular periods.

Ample research is done using time series analysis in various aspects such as economic changes, geographical changes, and many more. Many applications use time series analysis to predict the near future possibilities based on defined parameters. In this section, we will summarize a few applications based on time series analysis. Also, we will discuss some prior studies based on COVID-19.

S. Julian et al. [3] have used the ARIMA model for calculating the prediction of immunization in Indonesia. Their study focused on predicting the vaccine stock for immunization which will help the government determine vaccine stock requirements for clinics to avoid shortage or excess vaccines in stock.

V. Varsharani et al. [4] worked on forecasting Measles vaccine requirements using time series analysis. Their study used the Expert Modeler as it showed a simple seasonal model as the best fit model. The study helped in the vaccine requirement calculation as it gave the upper and lower confidence interval, assuring the optimum supply of vaccines in Medical colleges.

S. Saswat et al. [5] used the ARIMA model to predict the effect of the lockdown-unlock periods on the COVID-19

dataset. The model results were compared with other models like TBATS, N-Beats, Prophet, Single Exponential Method, Double Exponential Method, and Moving Average, using the Root Mean Square Error (RMSE).

H. Tandon et al. [11] contrasted the accuracy of the ARIMA model with other time series analysis models like Linear Trend, Moving Average, Quadratic Linear, Single Exponential, S-Curve Trend, and Double Exponential. The conclusion was that ARIMA(2,2,2) is the most accurate amongst the other models for forecasting the confirmed COVID-19 infection cases, based on MAPE, MAD, and MSD values. They suggest employing this model for foretelling the number of COVID-19 infection cases in India.

Benvenuto, Domenico et al. [14] performed predictions using the ARIMA model on the Johns Hopkins epidemiological data. This study predicted the epidemiological trend of the prevalence and frequency of COVID-19.

Lai D. et al. [15] applied Box-Jenkins, random walk on the daily reported SARS cases, and ARMA model to monitor the effects of SARS in China based on the new cases reported daily by the Ministry of Health of China.

These works mention the effectiveness of the ARIMA model in various use cases, and the research undertaken helps to understand the behavior of COVID-19 infection.

III. METHODOLOGY

A. Dataset

This study uses the vaccination data provided by Hannah Ritchie on “Our World in Data”[1], referenced by [6]. The site has data based on various metrics like Vaccinations, Tests & positivity, Hospital & ICU, Confirmed cases, Confirmed deaths, Reproduction rate, Policy responses, and Other variables of interest. The data is updated daily based on verifiable public official sources. The dataset can be accessed online at: <https://ourworldindata.org/coronavirus-source-data>.

The vaccination data consists of country-wise data on COVID-19 vaccinations for 209 countries [2]. This dataset includes some subnational locations (England, Northern Ireland, Scotland, Wales, Northern Cyprus, etc.) and international aggregates (World, continents, European Union, etc.). The dataset has various attributes which are tabulated in Table 1.

TABLE I. ATTRIBUTES IN VACCINATION DATASET

| location | iso_code |
|-------------------------|-------------------------------------|
| date | total_vaccinations |
| people_fully_vaccinated | total_vaccinations_per_hundred |
| people_vaccinated | people_vaccinated_per_hundred |
| daily_vaccinations_raw | daily_vaccinations_per_million |
| daily_vaccinations | people_fully_vaccinated_per_hundred |

But for this analysis, we used only a few attributes like location, date, and people_vaccinated_per_hundred. The aim is to forecast how many people will be vaccinated per hundred in the next 30 days for selected countries based on their overall development status.

As mentioned in the introduction, the approach is to forecast the vaccination rate for some countries based on their development status. In time series forecasting, we collect past observations and analyze them to develop a suitable mathematical model. The model then captures the underlying pattern of data generation for the series. This trained model can predict future events [8]. As prior studies showcase, ARIMA i.e., Autoregressive Integrated Moving Average performs effectively in forecasting. Hence, we will be incorporating the ARIMA model for forecasting vaccination rates in this study. Fig. 1. shows the flowchart of implementation for this work.

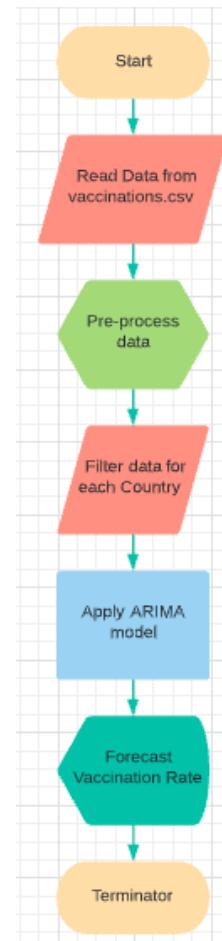


Fig. 1. Flowchart of Implementation

B. ARIMA

The classical regression is insufficient for explaining the dynamics of a time series as there is often additional information in the data that regression does not capture. This may lead to the generation of correlation introduced due to lagged linear relations. Hence, the *autoregressive model*, *AR*, and *autoregressive moving average model*, *ARMA* came

into the picture. For non-stationary scenarios, the *ARIMA* model was introduced [10].

In terms of ARIMA, "autoregressive" terms are the lags in the forecasting equation of a stationary series, and the "moving average" terms are the lags of the forecast errors. A time series that needs to be differentiated to make it stationary is an "integrated" version of the stationary series [7]. ARIMA is a model that fits time-series data to better understand or predict future data points in the series.

As mentioned in [7][11], an ARIMA model is expressed in terms of *ARIMA*(p, d, q), Here,

- p is the number of autoregressive (AR) terms,
- d is the degree of trend differences needed for stationarity (I), and
- q is the number of moving average (MA) terms i.e., the lagged forecast errors in the prediction equation.

We use the following representation of ARIMA for forecasting the vaccination rates in various countries [11].

ARIMA(p, d, q):

$$X_t = (\alpha_1 X_{t-1}) + (\alpha_2 X_{t-2}) + (\beta_1 Z_{t-1}) + (\beta_2 Z_{t-2}) + Z_t \quad (1)$$

$$\text{where, } Z_t = X_t - (X_{t-1}) \quad (2)$$

Here,

- X_t = forecasted values for vaccinations done on t^{th} day,
- $\alpha_1, \alpha_2, \beta_1$, and β_2 = parameters
- Z_t = residual term from the t^{th} day.

For implementation, we are using the *auto_arima* from the *pmdarima* library available in python. The *auto-ARIMA* method attempts to identify the most optimal parameters for an ARIMA model to achieve a single best-fit ARIMA model [9].

C. AIC

AIC, i.e., Akaike's Information Criterion, is used to identify the best ARIMA model. The model with lower AIC is the "better" one when comparing multiple models [12].

In *auto-ARIMA*, the best fit model derives from a given information_criterion. This criterion could be one of Akaike Information Criterion, Bayesian Information Criterion, Corrected Akaike Information Criterion, etc. The *auto-ARIMA* returns the model that minimizes the value.

D. ACF & PACF

The ACF, autocorrelation function graph, and PACF, partial autocorrelation graph help analyze differencing order. The model is tested for stationary characteristics and variance in normality via the Augmented Dickey-Fuller

(ADF) test [11]. *Auto-ARIMA* conducts the differencing tests and determines the order of differencing, d [9]. The test can be mentioned explicitly among the options of Augmented Dickey-Fuller, Phillips-Perro, or Kwiatkowski-Phillips-Schmidt-Shin.

The implementation of this case study is inspired by the work of [11] and [13].

IV. RESULTS & DISCUSSION

As mentioned earlier, we aim to compare the vaccination rates for countries based on their overall development statuses like developed, developing, and under-developed.

A. Developed countries

When searched for developed countries, the top countries that populate the list are Norway, Switzerland, Ireland, Germany, and many more.

The experimentation graphs for Norway and Switzerland data are shown below. Fig. 2. and Fig. 5. show the ACF and PACF graphs for both countries. Fig. 3. and Fig. 6. illustrate the *auto-ARIMA* model summary. Fig. 4. and Fig. 7. depict the forecast of vaccination rate (red line) in Norway and Switzerland, respectively, for the upcoming 30 days.

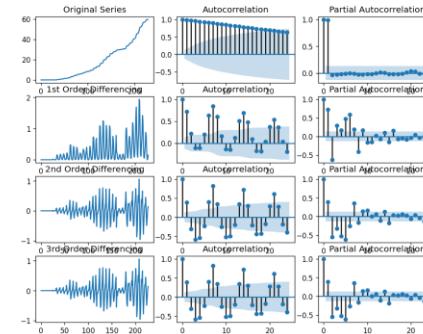


Fig. 2. ACF and PACF graphs for Norway vaccination data

```
Performing stepwise search to minimize aic
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=0.727, Time=0.25 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=70.331, Time=0.06 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=34.579, Time=0.07 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=-1.193, Time=0.12 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=68.333, Time=0.02 sec
ARIMA(0,2,2)(0,0,0)[0] intercept : AIC=inf, Time=0.32 sec
ARIMA(1,2,2)(0,0,0)[0] intercept : AIC=inf, Time=0.65 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=3.02, Time=0.09 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=1.21, Time=0.08 sec
ARIMA(0,2,2)(0,0,0)[0] : AIC=-27.889, Time=0.27 sec
ARIMA(1,2,2)(0,0,0)[0] : AIC=-59.566, Time=0.27 sec
ARIMA(2,2,2)(0,0,0)[0] : AIC=-187.020, Time=0.30 sec
ARIMA(2,2,1)(0,0,0)[0] : AIC=137.712, Time=0.15 sec
ARIMA(3,2,2)(0,0,0)[0] : AIC=185.052, Time=0.56 sec
ARIMA(3,2,1)(0,0,0)[0] : AIC=144.943, Time=0.22 sec
ARIMA(2,2,2)(0,0,0)[0] intercept : AIC=-185.110, Time=0.66 sec

Best model: ARIMA(2,2,2)(0,0,0)[0]
Total fit time: 4.115 seconds
SARIMAX Results
=====
Dep. Variable: y No. Observations: 230
Model: SARIMAX(2, 2, 2) Log Likelihood: 98.510
Date: Wed, 21 Jul 2021 AIC: -187.020
Time: 06:29:22 BIC: -169.874
Sample: - 230 HQIC: -188.102
Covariance Type: opg
```

Fig. 3. ARIMA model summary for Norway vaccination data

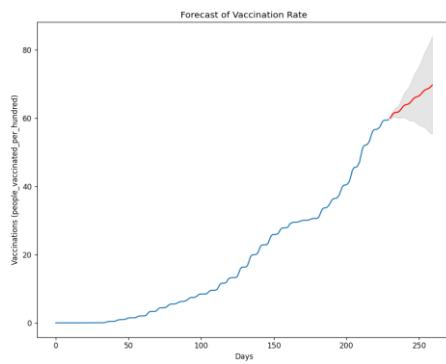


Fig. 4. Forecast of Vaccination Rate for Norway

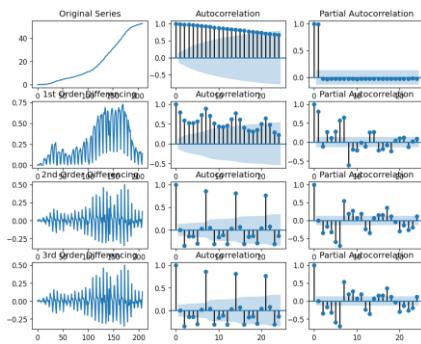


Fig. 5. ACF and PACF graphs for Switzerland vaccination data

```
Performing stepwise search to minimize aic
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=-272.302, Time=0.39 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=-224.737, Time=0.09 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=-232.738, Time=0.04 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=-232.739, Time=0.11 sec
ARIMA(0,2,0)(0,0,0)[0] : AIC=-236.732, Time=0.02 sec
ARIMA(2,2,1)(0,0,0)[0] intercept : AIC=-304.372, Time=0.27 sec
ARIMA(2,2,0)(0,0,0)[0] intercept : AIC=-257.049, Time=0.01 sec
ARIMA(3,2,0)(0,0,0)[0] intercept : AIC=-205.941, Time=0.49 sec
ARIMA(4,2,0)(0,0,0)[0] intercept : AIC=-260.359, Time=0.20 sec
ARIMA(4,2,1)(0,0,0)[0] intercept : AIC=-334.273, Time=0.79 sec
ARIMA(4,2,2)(0,0,0)[0] intercept : AIC=-279.187, Time=0.29 sec
ARIMA(5,2,1)(0,0,0)[0] intercept : AIC=-390.728, Time=0.77 sec
ARIMA(5,2,2)(0,0,0)[0] intercept : AIC=-401.631, Time=1.11 sec
ARIMA(6,2,1)(0,0,0)[0] intercept : AIC=-463.682, Time=0.94 sec
ARIMA(6,2,2)(0,0,0)[0] intercept : AIC=-535.185, Time=1.45 sec
ARIMA(7,2,0)(0,0,0)[0] intercept : AIC=-530.482, Time=1.16 sec
ARIMA(8,2,1)(0,0,0)[0] intercept : AIC=-534.311, Time=1.60 sec
ARIMA(8,2,2)(0,0,0)[0] intercept : AIC=-534.311, Time=1.16 sec
ARIMA(6,2,0)(0,0,0)[0] intercept : AIC=-507.860, Time=1.18 sec
ARIMA(6,2,2)(0,0,0)[0] intercept : AIC=-534.688, Time=1.59 sec
ARIMA(7,2,2)(0,0,0)[0] intercept : AIC=-537.476, Time=0.91 sec
ARIMA(7,2,1)(0,0,0)[0] : AIC=-513.222, Time=0.73 sec
ARIMA(7,2,0)(0,0,0)[0] : AIC=-537.280, Time=0.88 sec
ARIMA(6,2,1)(0,0,0)[0] : AIC=-537.475, Time=0.95 sec
ARIMA(6,2,2)(0,0,0)[0] : AIC=-510.060, Time=0.68 sec
ARIMA(6,2,3)(0,0,0)[0] : AIC=-536.318, Time=0.89 sec

Best model: ARIMA(7,2,2)(0,0,0)[0]
Total fit time: 19.764 seconds
```

SARIMAX Results

| Dep. Variable: | y | No. Observations: | 210 |
|------------------|------------------|-------------------|----------|
| Model: | SARIMAX(7, 2, 2) | Log Likelihood: | 278.738 |
| Date: | Wed, 21 Jul 2021 | AIC: | -537.476 |
| Time: | 06:41:02 | BIC: | -504.101 |
| Sample: | 0 - 210 | HQIC: | -523.981 |
| Covariance Type: | opg | | |

Fig. 6. ARIMA model summary for Switzerland vaccination data

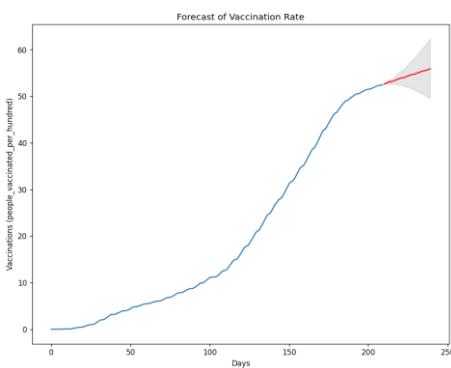


Fig. 7. Forecast of Vaccination Rate for Switzerland

B. Developing countries

In the list of developing countries, we find countries like India, Argentina, Brazil, China, and many more.

The experimentation graphs for India and Argentina data are shown below. Fig. 8. and Fig. 11. show the ACF and PACF graphs for both countries. Fig. 9. and Fig. 12. illustrate the auto-ARIMA model summary. Fig. 10. and Fig. 13. depict the forecast of vaccination rate (red line) in India and Argentina, respectively, for the upcoming 30 days.

C. Under-developed countries

Some of the underdeveloped countries include Niger, Chad, South Sudan, Mali, etc. In the dataset, it is observed that there are fewer tuples for these countries. This also indicates the lack of awareness among the citizens of these countries.

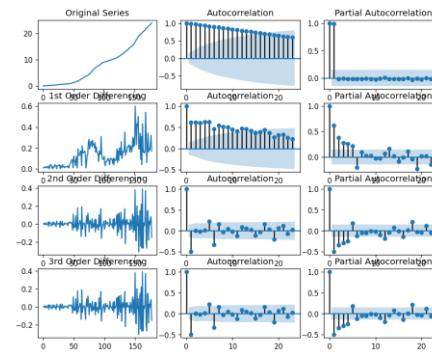


Fig. 8. ACF and PACF graphs for India vaccination data

```
Performing stepwise search to minimize aic
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=-405.821, Time=0.39 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=-318.026, Time=0.96 sec
ARIMA(1,2,2)(0,0,0)[0] intercept : AIC=-407.113, Time=0.10 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=-407.113, Time=0.14 sec
ARIMA(0,2,0)(0,0,0)[0] : AIC=-320.011, Time=0.07 sec
ARIMA(0,2,2)(0,0,0)[0] : AIC=-405.951, Time=0.26 sec
ARIMA(1,2,2)(0,0,0)[0] intercept : AIC=-403.115, Time=0.48 sec
ARIMA(0,2,2)(0,0,0)[0] : AIC=-408.245, Time=0.05 sec
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=-407.057, Time=0.12 sec
ARIMA(0,2,2)(0,0,0)[0] : AIC=-407.225, Time=0.12 sec
ARIMA(1,2,0)(0,0,0)[0] : AIC=-369.929, Time=0.07 sec
ARIMA(1,2,2)(0,0,0)[0] : AIC=-405.561, Time=0.24 sec

Best model: ARIMA(0,2,1)(0,0,0)[0]
Total fit time: 2.106 seconds
```

SARIMAX Results

| Dep. Variable: | y | No. Observations: | 180 |
|------------------|------------------|-------------------|----------|
| Model: | SARIMAX(0, 2, 1) | Log Likelihood: | 206.123 |
| Date: | Wed, 21 Jul 2021 | AIC: | -408.245 |
| Time: | 15:49:48 | BIC: | -401.882 |
| Sample: | 0 - 180 | HQIC: | -405.665 |
| Covariance Type: | opg | | |

Fig. 9. ARIMA model summary for India vaccination data

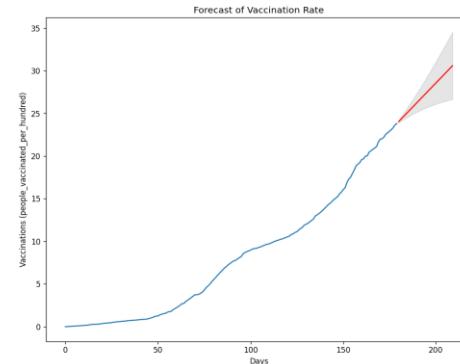


Fig. 10. Forecast of Vaccination Rate for India

The experimentation graphs for Niger and South Sudan data are shown below. Fig. 14. and Fig. 16. illustrate the auto-ARIMA model summary. Fig. 15. and Fig. 17. depict the forecast of vaccination rate (red line) in Niger and South Sudan, respectively, for the upcoming 30 days.

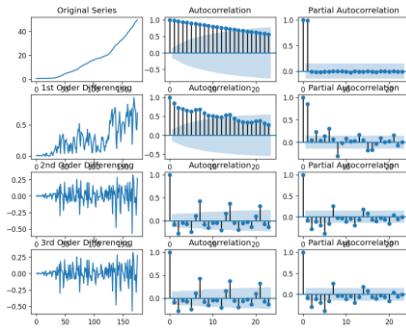


Fig. 11. ACF and PACF graphs for Argentina vaccination data

```
Performing stepwise search to minimize aic
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=-261.694, Time=0.38 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=-233.421, Time=0.03 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=-233.817, Time=0.04 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=-236.232, Time=0.13 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=-236.248, Time=0.02 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=-236.252, Time=0.01 sec
ARIMA(2,1,0)(0,0,0)[0] intercept : AIC=-263.703, Time=0.23 sec
ARIMA(2,2,0)(0,0,0)[0] intercept : AIC=-263.461, Time=0.55 sec
ARIMA(2,2,2)(0,0,0)[0] intercept : AIC=-261.888, Time=0.45 sec
ARIMA(1,2,2)(0,0,0)[0] intercept : AIC=-265.813, Time=0.52 sec
ARIMA(0,2,2)(0,0,0)[0] intercept : AIC=-265.512, Time=0.20 sec
ARIMA(1,2,2)(0,0,0)[0] intercept : AIC=-264.597, Time=0.18 sec

Best model: ARIMA(1,2,2)(0,0,0)[0] intercept
Total fit time: 3.132 seconds
SARIMAX Results
=====
Dep. Variable: SARIMAX(1, 2, 2) No. Observations: 178
Model: SARIMAX(1, 2, 2) Log Likelihood: -261.907
Date: Wed, 21 Jul 2021 AIC: -265.813
Time: 16:58:53 BIC: -249.961
Sample: 0 HQIC: -259.384
Covariance Type: opg
```

Fig. 12. ARIMA model summary for Argentina vaccination data

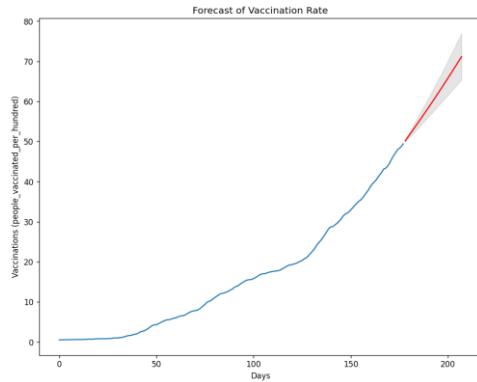


Fig. 13. Forecast of Vaccination Rate for Argentina

```
Performing stepwise search to minimize aic
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=1.068, Time=0.16 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=-0.601, Time=0.01 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=0.583, Time=0.09 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=-0.900, Time=0.11 sec
ARIMA(0,2,0)(0,0,0)[0] : AIC=-1.262, Time=0.01 sec

Best model: ARIMA(0,2,0)(0,0,0)[0]
Total fit time: 0.404 seconds
SARIMAX Results
=====
Dep. Variable: y No. Observations: 9
Model: SARIMAX(0, 2, 0) Log Likelihood: 1.631
Date: Wed, 21 Jul 2021 AIC: -1.262
Time: 17:14:36 BIC: -1.316
Sample: 0 HQIC: -1.930
Covariance Type: opg
```

Fig. 14. ARIMA model summary for Niger vaccination data

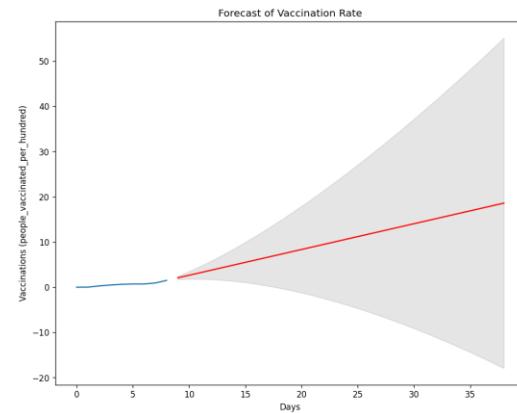


Fig. 15. Forecast of Vaccination Rate for Niger

```
Performing stepwise search to minimize aic
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=-55.633, Time=0.32 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=-46.907, Time=0.03 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=-56.564, Time=0.09 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=-52.082, Time=0.12 sec
ARIMA(0,2,0)(0,0,0)[0] : AIC=-48.907, Time=0.02 sec
ARIMA(2,2,0)(0,0,0)[0] intercept : AIC=-56.030, Time=0.17 sec
ARIMA(2,2,1)(0,0,0)[0] intercept : AIC=-54.051, Time=0.24 sec
ARIMA(1,2,0)(0,0,0)[0] : AIC=-58.535, Time=0.07 sec
ARIMA(2,2,0)(0,0,0)[0] : AIC=-58.023, Time=0.10 sec
ARIMA(1,2,1)(0,0,0)[0] : AIC=-57.605, Time=0.29 sec
ARIMA(0,2,1)(0,0,0)[0] : AIC=-52.929, Time=0.07 sec
ARIMA(2,2,1)(0,0,0)[0] : AIC=-56.047, Time=0.21 sec
```

```
Best model: ARIMA(1,2,0)(0,0,0)[0]
Total fit time: 1.773 seconds
SARIMAX Results
=====
Dep. Variable: y No. Observations: 16
Model: SARIMAX(1, 2, 0) Log Likelihood: 31.267
Date: Wed, 21 Jul 2021 AIC: -58.535
Time: 17:23:51 BIC: -57.257
Sample: 0 HQIC: -58.653
- 16
Covariance Type: opg
```

Fig. 16. ARIMA model summary for South Sudan vaccination data

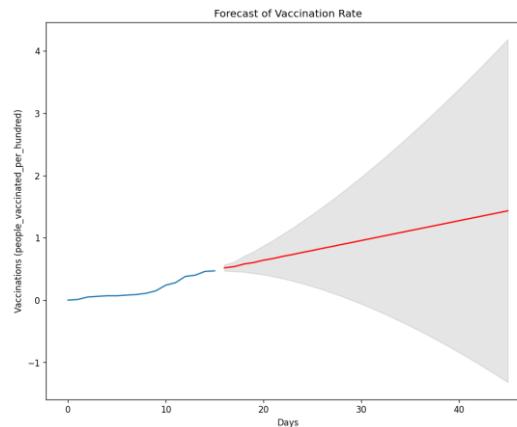


Fig. 17. Forecast of Vaccination Rate for South Sudan

From the above experimentation results, we observe that for the vaccinations dataset, the auto-ARIMA function determines d as a second-order differencing.

Table 1. shows the forecasts of vaccination rate for people per hundred for the upcoming 30 days.

TABLE I. SUMMARIZATION OF VACCINATION RATES IN DIFFERENT COUNTRIES

| Development Status | Countries | Forecast Vaccination Rate (people vaccinated per hundred) | Remark |
|--------------------|---------------------|---|---|
| Developed | Norway, Switzerland | 50 to 80 50 to 65 | Increasing moderately for both |
| Developing | India, Argentina | 27 to 35 65 to 75 | Slow increase for India, Rapid increase for Argentina |
| Underdeveloped | Niger, South Sudan | -20 to 55 -2 to 4.5 | Very slow for both |

V. CONCLUSION

In this work, we forecasted the vaccination rates for some of the developed, developing, and underdeveloped countries for the upcoming 30 days.

From the experimentations, we can conclude that the vaccination rate for developing countries is increasing at quite a pace. The average rate observed is 50 - 80 people per hundred getting vaccinated in the next 30 days.

For developing countries, the average forecasted rate observed is 30 - 60 people per hundred getting vaccinated in the upcoming month. With this, we can understand the level of social awareness amongst the citizens.

However, for underdeveloped countries, we have less data on vaccinations. Due to this, the vaccination rate can't be accurately forecasted, thus providing us a wide range of possibilities ranging from -10 to 50 people per hundred getting vaccinated in the next 30 days on an average. This may also imply that the governments of these countries need to spread awareness amongst their citizens and also supply ample quantity of vaccines to their citizens.

VI. FUTURE WORK

As future research, the model can extend to forecast vaccination rates using other attributes provided in the dataset. The sources [1][2] also have an explicit US vaccinations dataset, which could be used for similar purposes. Different models can be used, and their performance can be compared to the chosen problem statement.

REFERENCES

- [1] Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C., Appel, C., Rodés-Guirao, L. and Roser, M., 2021. *Coronavirus Pandemic (COVID-19)*. [online] Our World in Data. Available at: <<https://ourworldindata.org/coronavirus-source-data>> [Accessed 20 July 2021].
- [2] Mathieu, E., Ritchie, H., Ortiz-Ospina, E. et al. A global database of COVID-19 vaccinations. *Nat Hum Behav* (2021). <https://doi.org/10.1038/s41562-021-01122-8>.
- [3] Sahisnu, Julian Satya, et al. Vaccine Prediction System Using ARIMA Method. ICIC International, 2020. DOI.org (CSL JSON), <https://doi.org/10.24507/icicelb.11.06.567>.
- [4] Vithalrao Kendre, Varsharani, et al. "Forecasting Measles Vaccine Requirement By Using Time Series Analysis." *Journal of Evolution of Medical and Dental Sciences*, vol. 6, no. 28, Apr. 2017, pp. 2329–33. *DOL.org (Crossref)*, doi:10.14260/Jemds/2017/501.
- [5] Singh, Saswat, et al. "Time Series Analysis of COVID-19 Data to Study the Effect of Lockdown and Unlock in India." *Journal of The Institution of Engineers (India): Series B*, Apr. 2021. Springer Link, doi:10.1007/s40031-021-00585-7.
- [6] Tuli, Shreshth, et al. "Predicting the Growth and Trend of COVID-19 Pandemic Using Machine Learning and Cloud Computing." *Internet of Things*, vol. 11, Sept. 2020, p. 100222. *DOL.org (Crossref)*, doi:10.1016/j.iot.2020.100222.
- [7] People.duke.edu. 2021. *Introduction to ARIMA models*. [online] Available at: <<https://people.duke.edu/~rnau/411arim.htm>> [Accessed 20 July 2021].
- [8] R. Adhikari and R. K. Agrawal, An introductory study on time series modeling and forecasting, arXiv Prepr. arXiv1302.6613, 2013.
- [9] Alkaline-ml.com. 2021. *pmdarima.arima.AutoARIMA — pmdarima 1.8.2 documentation*. [online] Available at: <<https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.AutoARIMA.html>> [Accessed 20 July 2021].
- [10] R. H. Shumway and D. S. Stoffer, Time Series Analysis and Its Applications: With R Examples, Springer, 2017.
- [11] H. Tandon, P. Ranjan, T. Chakraborty, V. Suhag, Coronavirus (covid-19): Arima based time-series analysis to forecast near future. arXiv:2004.07859 (2020)
- [12] Keshvani, A. and Keshvani, V., 2021. *Using AIC to Test ARIMA Models*. [online] CoolStatsBlog. Available at: <<https://coolstatsblog.com/2013/08/14/using-aic-to-test-arima-models-2/>> [Accessed 20 July 2021].
- [13] Benny, D., 2021. *COVID-19 Vaccination Rate / Analysis of COVID-19 Vaccination Rate*. [online] Analytics Vidhya. Available at: <<https://www.analyticsvidhya.com/blog/2021/04/time-series-analysis-forecast-covid-19-vaccination-rate/>> [Accessed 20 July 2021].
- [14] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief*, 29, 105340. <https://doi.org/10.1016/j.dib.2020.105340>
- [15] Lai, D. (2021). Monitoring the SARS Epidemic in China: A Time Series Analysis. *Journal of Data Science*, 3(3), 279-293. doi:10.6339/JDS.2005.03(3).229