Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 179 (2021) 524–532

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on Computer Science and Computational Intelligence 2020

Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET

Christophorus Beneditto Aditya Satrio[a,*], William Darmawan[a], Bellatasya Unrica Nadia[a], Novita Hanafiah[b]

[a]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480
[b]Computer Science Department, BINUS Online Learning, Bina Nusantara University, Jakarta, Indonesia 11480

Abstract

The spread of COVID-19 has caused it to be a pandemic. This has caused massive disruption to our daily lives, both directly and indirectly. We aim to utilize Machine Learning model in attempt to forecast the trend of the disease in Indonesia with finding out the approximation when normality will return. This study uses Facebook's Prophet Forecasting Model and ARIMA Forecasting Model to compare their performance and accuracy on dataset containing the confirmed cases, deaths, and recovered numbers, obtained from the Kaggle website. The forecast models are then compared to the last 2 weeks of the actual data to measure their performance against each other. The result shows that Prophet generally outperforms ARIMA, despite it being further from the actual data the more days it forecasts.

Keywords: COVID; Forecasting; Prophet; ARIMA

1. Introduction

Coronavirus disease (COVID-19) is a new disease caused by the SARS-COV-2 virus. The virus first originated in Wuhan, Hubei province in December 2019. While at first it is just a series of pneumonia with unknown cause in

* Corresponding author. Tel.: +62-817-690-2964.
E-mail address: christophorus.satrio@binus.ac.id

10.1016/j.procs.2021.01.036

Wuhan, it quickly became an international crisis in less than a month. Almost six million people have been infected with over three hundred thousand deaths worldwide. In effect, countries have been locked down, public places have been closed, and various other activity-limiting policies have been implemented to slow down the spread of the disease. The COVID-19 virus spreads primarily through droplets that come out from a person's mouth or nose as they sneeze or cough. It may not sound deadly if people play safe and not coughing or sneezing carelessly, but the fact that it has spread through the globe denies the fact that COVID-19 cannot be treated as deadly.

Machine Learning has expanded its field into forecasting and there have been plentiful of researches conducted around this. In order to find out approximately when the disease will be eradicated and when our life will resume to normality, we aim to forecast the time series of COVID-19 in Indonesia based on existing data recorded from January 20, 2020 until May 21 2020 and exhibit the patterns of the time series of the disease for 4 weeks ahead. We apply two different approaches that have the capability to produce future results. They are Auto Regressive Integrated Moving Average (ARIMA) model and the forecasting procedure, PROPHET. ARIMA models are quite popular and have been applied to many fields, such as banking stock market[1] and tourism demand[2,3]. While PROPHET can be considered new approach as it was released 3 years ago, nevertheless has been renowned for its ease of usability but powerful model. This section of the paper reviews the literature which has been used by people in using several algorithms as the approaches to predict, classify, and forecast various diseases.

1.1. SIR Model (Classification and Forecasting)

Several studies in predicting diseases focus on the Scheme of Susceptible-Infectious-Recovered/Death (SIRD) model. We found studies[4,5] that focus on influenza outbreaks. The epidemic onset can be rapidly tracked, and variations of incidence can be captured. There are studies that focuses on the coronavirus[6,7,8] however the model for[6] has high variation[4,5] and long range of lowerbound and upperbound which depicts uncertainty. Qihui Yang et al[7] bounded Kalman Filter into SIR model using Gilespie algorithm and simulates the spreading of virus based on Markovian and non-Markovian processes. However, it resulted in high inconsistency. Louis Kim et al[8] tackles the inaccuracy of the SIRD model by further comparing it with real-world data which creates the best fit when the actual research data are fed. Cieo Anastassopoulou[9] also used SIRD model to forecast the COVID-19 outbreak. However, it did not take into account of the factors that play an important role of determining the dynamics of the disease.

1.2. Classification

A research[10] uses several classification algorithms to compare their sensitivity, specificity, and balanced-accuracy metrics. While another[11] uses combinations of modified RNN and GRU to classify using SVM. Vijayarani et al[12] uses SVM with Naive Bayes to predict liver disease to compare both algorithm's differences, while Kapoor et al[13] uses J48 Decision Tree instead of SVM with generally the same reason. Predictive Models for MERS-CoV infections[14] are also built with Naive Bayes and J48 Decision Tree. Carlos Sather[15] tested the predictive capability of MultiLayer Perceptron, Gated Recurrent Unit (GRU), LSTM with bayesian ridge regression. GRU showed the best result. Minh Nguyen[16] uses RNN to predict the progression rate of Alzheimer disease which proves to be superior to SVM and LSS.

1.3. Time Series Analysis (Classification and Forecasting)

A research from China uses search engine queries as its dataset for forecasting China's AIDS epidemic[17]. The approach used was surprisingly, an Artificial Neural Network. The multilayer perceptron worked really well with the forecast and the perfect pick for the Pearson Correlation Coefficient (PCC) threshold. Mahmood Akhtar et at[18] successfully showed high accuracy in forecasting the zika virus in a 4-week span using Nonlinear AutoRegression with eXogeneous inputs (NARX) Neural Network. Joceline Lega[19] proposed a self-made model called EpiGro which could estimate the order of magnitude, peak, and ultimate size of an ongoing outbreak. The result was very satisfying as the model could interpolate through widely oscillating reported incidence data, given that the data is very noisy. Qihui Yang[7] also uses the Kalman Filter to produce short-term forecasting, which only lasted for 3 days apparently. Dennis Ndanguza et al[20] analyzed biased and unbiased Extended Kalman Filter model, which did not have major differences. Xianghei Zhu et al[21] used RNN with LSTM and an attention mechanism to forecast influenza epidemic. The attention mechanism could improve the Mean Absolute Percentage Error of the model. Sangwon Chae et al[22] predicted disease using Big Data with LSTM and ARIMA with search trend data. LSTM was proven to yield better result. Carsten Kirkeby et al[23] proposed the evaluation of precision of Poisson regression, SIS and SIScom model. All methods perform equally when the sampling interval is large, and transmission is low. J.Zhang and K. Nawata[24] used Multi-Step prediction LSTM for forecasting influenza outbreak.

There are several studies which uses Facebook's Prophet Forecasting Procedure. Ye[25] mixes Prophet and ARIMA to make his own model to make weather forecasts, focusing on air quality with its pollutant with data from 11 air quality monitoring stations in Shenzhen. Yenidoğan, et al[26] aims to compare the performance of Prophet against ARIMA in forecasting Bitcoin's value. While both performs well in training and validation sets, ARIMA fits considerably worse than Prophet on their test set. While Tyralis, et al[27] uses Prophet for multi-step ahead forecasting of monthly streamflow, where they developed 4 models of prophet, 2 models of Naïve, a SES model and an ARIMA model. Alabi, et al[28] utilizes Prophet to forecast the growth of COVID-19 cases with the aim of emphasizing the importance of integrating easing lockdowns with a careful evaluation to prevent a second wave of the pandemic.

2. Data Preparation

2.1. Dataset and Data preprocessing

COVID-19 Dataset is obtained from the Kaggle website. The dataset consists of 27618 rows and 8 columns. The features of this data are: Serial Number, Observation Date, Province/State, Country/Region, Last Update, Confirmed, Deaths, and Recovered. The latest time stamp for the dataset is at May 21, 2020 and the oldest being at January 20, 2020. Dataset is shown in Table 1.

Table 1. Overview of COVID-19 Dataset

SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
1	1/22/2020	Anhui	Mainland China	1/22/2020 17:00	1	0	0
2	1/22/2020	Beijing	Mainland China	1/22/2020 17:00	14	0	0
3	1/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6	0	0
4	1/22/2020	Fujian	Mainland China	1/22/2020 17:00	1	0	0
5	1/22/2020	Gansu	Mainland China	1/22/2020 17:00	0	0	0
6	1/22/2020	Guangdong	Mainland China	1/22/2020 17:00	26	0	0
7	1/22/2020	Guangxi	Mainland China	1/22/2020 17:00	2	0	0
8	1/22/2020	Guizhou	Mainland China	1/22/2020 17:00	1	0	0
9	1/22/2020	Hainan	Mainland China	1/22/2020 17:00	4	0	0
10	1/22/2020	Hebei	Mainland China	1/22/2020 17:00	1	0	0
11	1/22/2020	Heilongjiang	Mainland China	1/22/2020 17:00	0	0	0
12	1/22/2020	Henan	Mainland China	1/22/2020 17:00	5	0	0
13	1/22/2020	Hong Kong	Hong Kong	1/22/2020 17:00	0	0	0
14	1/22/2020	Hubei	Mainland China	1/22/2020 17:00	444	17	28
15	1/22/2020	Hunan	Mainland China	1/22/2020 17:00	4	0	0
16	1/22/2020	Inner Mongolia	Mainland China	1/22/2020 17:00	0	0	0
17	1/22/2020	Jiangsu	Mainland China	1/22/2020 17:00	1	0	0
18	1/22/2020	Jiangxi	Mainland China	1/22/2020 17:00	2	0	0

Dataset cleaning was done in order to extract the right information to be forecasted: The 'Last Update' column was removed because Observation Date was chosen to be the time stamp. Both PROPHET and ARIMA have their default time stamp, so a conversion into datetime was necessary in order to forecast. Then, a new dataset where the Country name is Indonesia was created by separating from the original COVID-19 data. Indonesia's COVID-19 data has the same number of columns but with 81 total rows, with each row representing data recorded in one day period. As ARIMA only accepts univariate data, a split was required for the confirmed, deaths, and recovered to have their own data frames with time stamp being their indexes.

Table 2. Splitting results to make the dataset univariate

Date	Confirmed	Date	Deaths	Date	Recovered
3/2/2020	2	3/2/2020	0	3/2/2020	0
3/3/2020	2	3/3/2020	0	3/3/2020	0
3/4/2020	2	3/4/2020	0	3/4/2020	0
3/5/2020	2	3/5/2020	0	3/5/2020	0
3/6/2020	4	3/6/2020	0	3/6/2020	0
3/7/2020	4	3/7/2020	0	3/7/2020	0

2.2. Checking for stationarity (ARIMA)

The first step in applying ARIMA model is to check whether the time series is stationary or not. ARIMA works best when our data has a stable or consistent pattern overtime, meaning that the variance and mean of the data have to remain constant overtime. Thus, when the data has a trend of going upwards or downwards and has a particular pattern (seasonality), then the data is not stationary. Fig. 1. Shows that COVID-19 confirmed, death, and recovered cases in Indonesia do not have seasonality, however they do have a trend of going upwards and their variance and mean are not constant, which indicate non-stationary data.

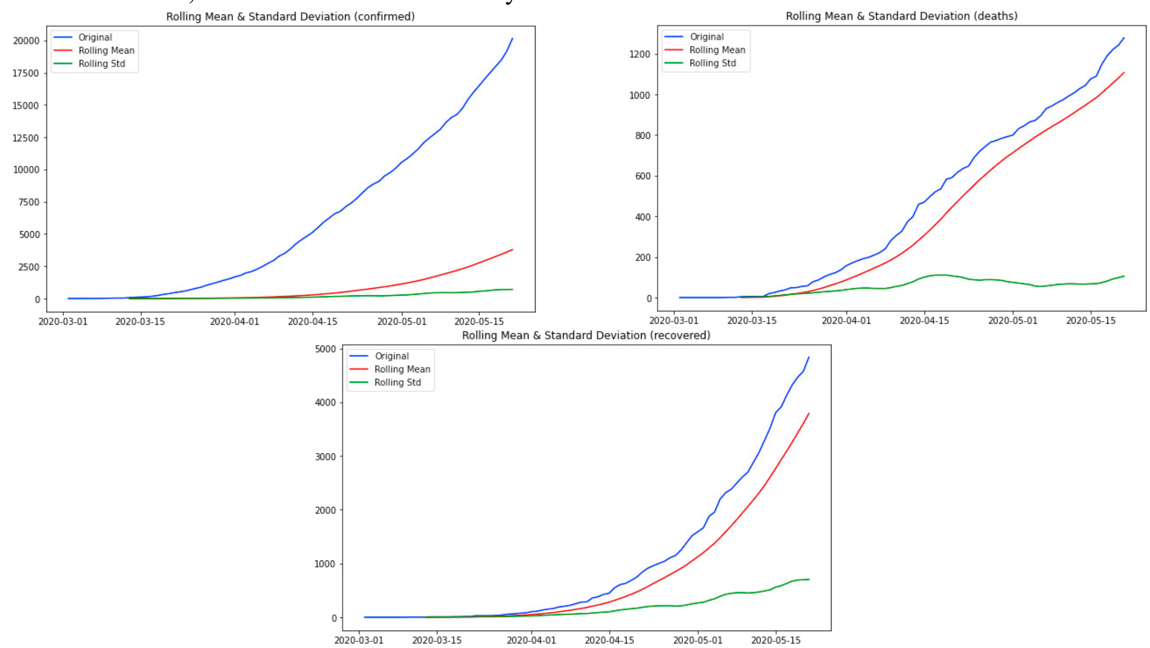


Fig. 1. Variance, Mean, and the trend from COVID-19 confirmed, death, and recovered data

To further confirm the stationarity of data, we used Augmented Dickey Fuller (ADF) Test as statistical test like this makes strong assumptions about our data. ADF informs that the degree to which a null hypothesis can be rejected or not rejected in order to determine if our data is stationary or not. This is interpreted using a threshold (0.05) that will suggest if we reject or accept the null hypothesis. The ADF results of the data are shown in Fig. 2. Below.

```
Results of Dickey Fuller Test (confirmed):
Test Statistic                    1.720431
p-value                           0.998177
#Lags Used                       10.000000
Number of Observations Used      70.000000
Critical Value (1%)              -3.527426
Critical Value (5%)              -2.903811
Critical Value (10%)             -2.589320
dtype: float64
```

```
Results of Dickey Fuller Test (deaths):
Test Statistic                    0.538579
p-value                           0.986000
#Lags Used                       11.000000
Number of Observations Used      69.000000
Critical Value (1%)              -3.528890
Critical Value (5%)              -2.904440
Critical Value (10%)             -2.589656
dtype: float64
```

```
Results of Dickey Fuller Test (recovered):
Test Statistic                    0.257808
p-value                           0.975324
#Lags Used                       12.000000
Number of Observations Used      68.000000
Critical Value (1%)              -3.530399
Critical Value (5%)              -2.905087
Critical Value (10%)             -2.590001
dtype: float64
```

Fig. 2. ADF Tests of COVID-19 confirmed, death, and recovered cases in Indonesia.

From the results it can be inferred that :

- Confirmed, death, and recovered cases data have the p-value higher than the threshold, in which they failed to reject the null hypothesis (the data is not stationary).
- The test statistics for confirmed, death, and recovered cases data are none near or within the region of the critical values, which shows that the data is not stationary.

2.3 Data Transformation (ARIMA)

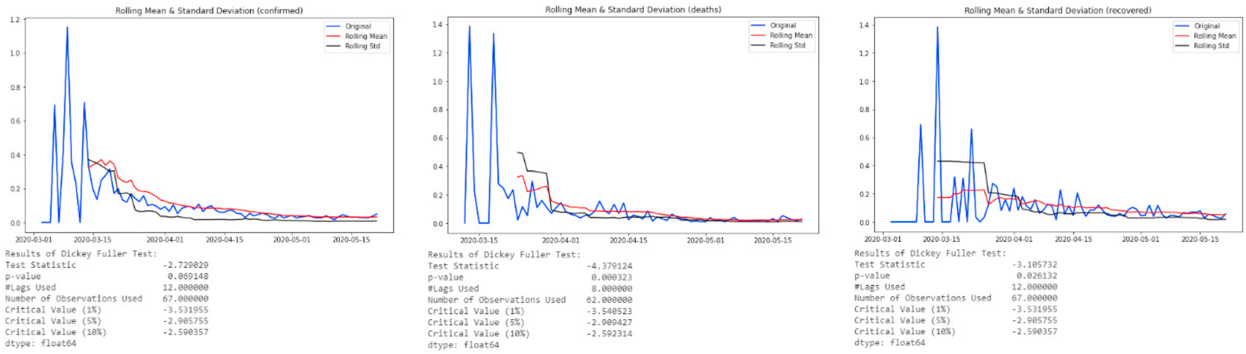


Fig. 3. Variance, Mean, Original values, and ADF Test results from COVID-19 confirmed, death, and recovered data after performing transformation

There are many methods that can be used to transform data into being stationary. In this case, log-scale transformation and time-shifting transformation were used. Log-scale transformation can be used to stabilize series that have non-constant variance so that the series would have more normal distribution. This is done by simply taking the log of the numerical values inside the dataset. After log-scaling we found that inside the recovered and deaths dataset, there are data which have the values of negative infinity. This is normal as death and recovered always start at 0 in the beginning and taking the log of values near 0 would result in this kind of issue. Therefore, we simply replaced values which have negative infinity into 0. The next step is Time-Shifting transformation process.

Given a set of observation on the time series $x_0, x_1, x_2, \ldots, x_n$. The values will be shifted to the right, resulting in *null*, $x_0, x_1, x_2, \ldots, x_n$. Thus, the time series with shifted values is shown in (1).

$$null, (x1 - x0), (x2 - x1), (x3 - x2), \ldots (xn - (xn - 1)) \tag{1}$$

The results after transformation proved to be quite effective. As seen in Fig. 3, even though there are noises, the graphs are constant and show no trend. The ADF Test also prove that the data have changed into being stationary.

3. Forecasting Model

3.1. Forecasting with ARIMA

The AR in ARIMA stands for AutoRegressive which is represented by *p* in the model. It refers to the number of lags of *Y* to be used as the predictor. A pure AR model will be where *Yt* depends only on its own past values (*Yt-1, Yt-2*,…) . A common representation of an autoregressive model of order p can be written as (2).

$$yt=c+\phi 1yt-1+\phi 2yt-2+\cdots+\phi pyt-p+\varepsilon t, \tag{2}$$

where *εt* represents white noise.While the MA in ARIMA stands for Moving Average, represented by *q* in the model. Unlike AR model which uses past values, MA model depends only on past forecast errors. A common representation of a moving average model of order q can be written follow.

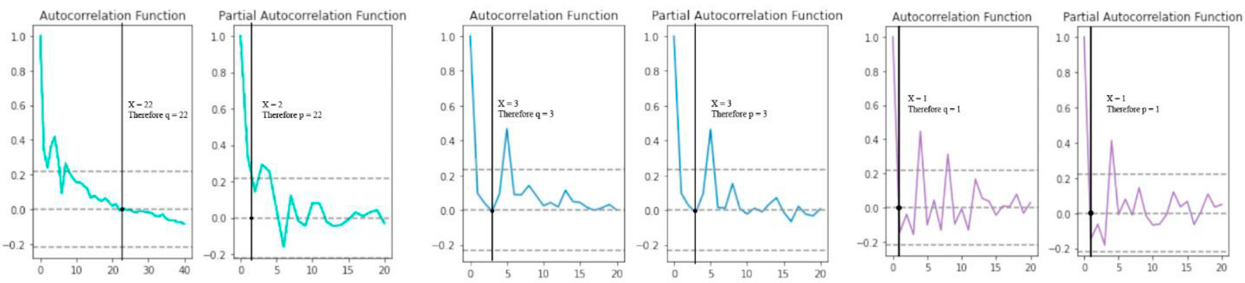
$$yt=c+\varepsilon t+\theta 1\varepsilon t-1+\theta 2\varepsilon t-2+\cdots+\theta q\varepsilon t-q, \tag{3}$$

where *εt* represents white noise, which forecast errors follow on. The I in ARIMA stands for Integrated and is represented by *d* in the model. When time series was differenced at least once to make it stationary and combined with AR and MA models, it will generate a non-seasonal ARIMA model, where in this context, differencing is the reverse of integration. The formula for ARIMA could be represented as

$$y't=c+\phi 1y't-1+\cdots+\phi py't-p+\theta 1\varepsilon t-1+\cdots+\theta q\varepsilon t-q+\varepsilon t, \tag{4}$$

where *y't* represents the series of difference as it may have been differenced multiple times, while the right side of the formula include both lagged values of *yt* from AR model and lagged errors from MA model. As our data have transformed into being stationary, it can be safely said that the order differencing (*d*) in our ARIMA model has the value of 1.

Identifying the AR term (*p*) was done by inspecting the Partial Autocorrelation (PACF) graph. The PACF is the correlation between any two points with a specific time shift, called lag where the linear effects of the points in between is removed[29]. Simply put, PACF gives the correlation between a time series and lag after excluding the contributions of intermediate lags. Finding *p* can be done by looking at the value of *x* which crosses *y* when *y* is 0 (when the correlations end up starting with a negative value). Identifying the MA term (*q*) was done by inspecting the Autocorrelation (ACF) graph. The ACF is the correlation between any two values in a time series with specific lag[29]. Same as finding *p*, finding *q* can be done by looking at the value of *x* which crosses *y* when *y* is 0. Both PACF and ACF graphs which can be seen in Fig. 4.



(a) (b) (c)

Fig. 4. PACF and ACF plot of the COVID-19 a) confirmed data; b) deaths data; c) recovered data

From Fig. 4a it can be seen from the ACF graph that the line touches $y = 0.0$ at $x = 22$. Therefore $q = 22$. It could be seen from the PACF graph that the line touches $y = 0.0$ at $x = 5$. Therefore $p = 5$. While in Fig. 4b, it can be seen from the ACF graph that the line touches $y = 0.0$ at $x = 3$. Therefore $q = 3$. It could be seen from the PACF graph that the line touches $y = 0.0$ at $x = 3$. Therefore $p = 3$. From Fig. 4 it can be seen from the ACF graph that the line touches $y = 0.0$ at $x = 1$. Therefore $q = 1$. It could be seen from the PACF graph that the line touches $y = 0.0$ at $x = 1$. Therefore $p = 1$.

3.2. Forecasting with PROPHET

PROPHET is a procedure for forecasting time series data that was created by Facebook's Core Data Science team. It aims to be able to forecast 'at scale', meaning PROPHET wants to be the forecasting tool that is automated in nature, giving more ease of use in tuning time series methods and enabling analysts from any background or people with little to (possibly) no prior knowledge in forecasting to be able to forecast successfully.

According to Facebook, PROPHET "works best with time series that have strong seasonal effects and several seasons of historical data and is robust to outliers and shifts in the trend." In this case, our data has no seasonality but performs well. And this is what PROPHET is good for. Its automatic nature gives flexibility into time series data that have dramatic changes and hence analysts do not have to worry about their data being not suitable for forecasting with PROPHET[30]. PROPHET is straightforward to use. Analysts need to prepare the dataset and create a data frame with two columns: '*ds*' or date stamp (in datetime format), and '*y*' or the forecasting measurement that must be in numerical values. Then, analysts need to create an object from the `Prophet()` class, in which the data frame is fitted into the object. After that analysts can choose the desired period to be forecasted and then proceed into forecasting. The forecasting result will have several columns, in which what analysts want to look at are the '*ds*' and '*yhat*' columns. '*yhat*' is a column containing the forecasted results of '*y*' in the historical data frame. '*ds*' and '*yhat*' can be plotted to show features such as future trend or seasonality.

4. Results and Discussion

Using ARIMA and PROPHET, we forecasted the confirmed, deaths, and recovered cases of COVID-19 in Indonesia within a 30-day period. We began forecasting in April 22, 2020 and ended in May 21, 2020. For PROPHET, we used the default parameters that are automatically tuned by itself. As for ARIMA, the only hyperparameters that were tuned are the order of p, d, and q.

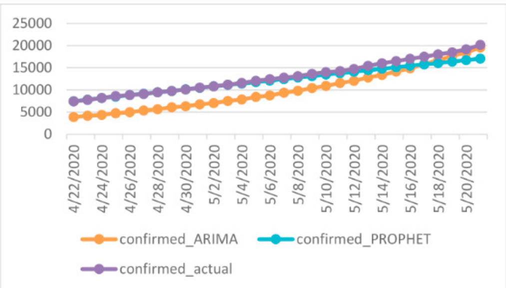


Fig. 5. Visualization of Indonesia COVID-19 confirmed cases

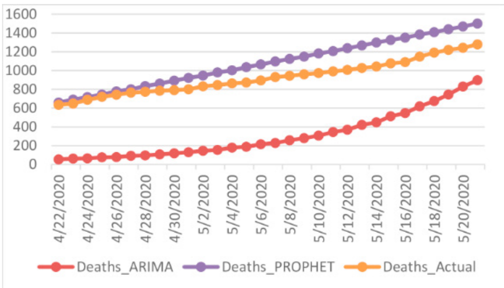


Fig. 6. Visualization of Indonesia COVID-19 Death Cases

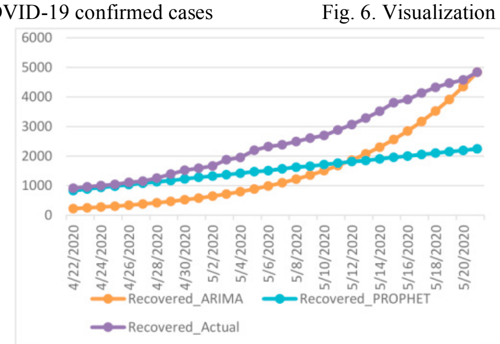


Fig. 7. Visualization of Indonesia COVID-19 Recovered Cases

Table 3. Evaluation Metrics for the ARIMA Model

ARIMA Set	R^2	MSE	MAE	MFE
Confirmed Cases	0.255996	9814486.067	2967	2967
Deaths	-11.406655	388366.6667	617.6	617
Recovered	0.310856	1024253.9	961.1666667	961.166667

Table 4. Evaluation Metrics for the Prophet Model

Prophet Set	R^2	MSE	MAE	MFE
Confirmed Cases	0.911158	1171958.567	686.3	671.766667
Deaths	0.050148	29733.3	153.1	-153.1
Recovered	-0.052854	1564826.033	955.833333	955.833333

The forecast data with PROPHET shows high accuracy at the beginning with only minimal differences between the actual data. However, as the time progresses, the differences tend to increase, creating a visible gap. Nevertheless, the trend in both actual and PROPHET's result are upwards, though in the recovered cases PROPHET shows trend in a more linear nature. The most visible gap between the actual data and PROPHET's prediction is at the recovered cases part. ARIMA was rather complicated to use. The key lies in choosing the best fit with tuning the p, d, and q order. However, we managed to make the dataset stationary so that order d can be considered 1 but still, ARIMA has less accuracy than PROPHET even since the start of forecast. Table 3 and 4 shows that both model do not fit very well. Though in comparison PROPHET still performs better than ARIMA even though no tunings were made as seen in Fig. 5, 6, and 7.

5. Conclusion

Our objective for this study is to compare how well can ARIMA and PROPHET handle time-series data with no seasonality, has random patterns, and with minimum observations by using COVID-19 cases data. The forecast was done in a 30-day time period for both model from April 22, 2020 to May 21, 2020. Both ARIMA and PROPHET were found to be fairly inaccurate in forecasting as time progresses.

Both models have Mean Forecast Error (MFE) which indicate that both models have positive biases except in the PROPHET's deaths part where it shows negative bias. Positive bias shows that the model is under-forecast, which more often than not, the forecast is less than the actual data. Negative bias shows vice versa, which more often than not, the forecast is more than the actual data.

PROPHET has good accuracy in predicting the confirmed cases with 91% precision, while ARIMA did not even pass through half precision. Both model also have negative R2 values. This means that the Sum of Squared Errors of the regression (the distances between actual data points and the regression line) are very far and even greater than the total of Sum of Squared Error. In other words, the regression line going to the way that does not match the harmony of the actual data (going further than the actual data). The greater the negative number of the R2, the greater the distance between the actual data and the predicted data, and the less accurate the predicted data will be.

References

1. Almasarweh M, Wadi SAL. ARIMA Model in Predicting Banking Stock Market Data. Modern Applied Science. 2018 Oct; 12: p. 309.
2. Petrevska B. Predicting tourism demand by A.R.I.M.A. models. Economic Research-Ekonomska Istraživanja. 2017 Jan; 30: p. 939–950.
3. Petrevska B. Forecasting international tourism demand: The evidence of Macedonia. UTMS Journal of Economics. 2012; 3: p. 45–55.

4. Miranda GHB, Baetens JM, Bossuyt N, Bruno OM, Baets BD. Real-time prediction of influenza outbreaks in Belgium. Epidemics. 2019 Sep; 28: p. 100341.
5. Yang W, Cowling BJ, Lau EHY, Shaman J. Forecasting Influenza Epidemics in Hong Kong. PLOS Computational Biology. 2015 Jul; 11: p. e1004383.
6. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease. 2020 Mar; 12: p. 165–174.
7. Yang Q, Yi C, Vajdi A, Cohnstaedt LW, Wu H, Guo X, et al. Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei Province, China. 2020 Mar..
8. Kim L, Fast SM, Markuzon N. Incorporating media data into a model of infectious disease transmission. PLOS ONE. 2019 Feb; 14: p. e0197646.
9. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLOS ONE. 2020 Mar; 15: p. e0230405.
10. Leo J, Luhanga E, Michael K. Machine Learning Model for Imbalanced Cholera Dataset in Tanzania. The Scientific World Journal. 2019 Jul; 2019: p. 1–12.
11. Che C, Xiao C, Liang J, Jin B, Zho J, Wang F. An RNN Architecture with Dynamic Temporal Matching for Personalized Predictions of Parkinson\textquotesingles Disease. In Proceedings of the 2017 SIAM International Conference on Data Mining.: Society for Industrial and Applied Mathematics; 2017. p. 198–206.
12. Vijayarani S, Dhayanand S. Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR). 2015; 4: p. 816–820.
13. Detecting Kidney Disease using Naïve Bayes and Decision Tree in Machine Learning. International Journal of Innovative Technology and Exploring Engineering. 2019 Nov; 9: p. 498–501.
14. Al-Turaiki I, Alshahrani M, Almutairi T. Building predictive models for MERS-CoV infections using data mining techniques. Journal of Infection and Public Health. 2016 Nov; 9: p. 744–748.
15. Sathler C, Luciano J. Predictive modeling of dengue fever epidemics: A Neural Network Approach. 2017.
16. Nguyen M, Sun N, Alexander DC, Feng J, Yeo BTT. Modeling Alzheimer's disease progression using deep recurrent neural networks. In 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI); 2018 Jun: IEEE.
17. Nan Y, Gao Y. A machine learning method to monitor China's AIDS epidemics with data from Baidu trends. PLOS ONE. 2018 Jul; 13: p. e0199697.
18. Akhtar M, Kraemer MUG, Gardner LM. A dynamic neural network model for predicting risk of Zika in real time. BMC Medicine. 2019 Sep; 17.
19. Lega J, Brown HE. Data-driven outbreak forecasting with a simple nonlinear growth model. Epidemics. 2016 Dec; 17: p. 19–26.
20. Ndanguza D, Mbalawata IS, Haario H, Tchuenche JM. Analysis of bias in an Ebola epidemic model by extended Kalman filter approach. Mathematics and Computers in Simulation. 2017 Dec; 142: p. 113–129.
21. Zhu X, Fu B, Yang Y, Ma Y, Hao J, Chen S, et al. Attention-based recurrent neural network for influenza epidemic prediction. BMC Bioinformatics. 2019 Nov; 20.
22. Chae S, Kwon S, Lee D. Predicting Infectious Disease Using Deep Learning and Big Data. International Journal of Environmental Research and Public Health. 2018 Jul; 15: p. 1596.
23. Kirkeby C, Halasa T, Gussmann M, Toft N, Græsbøll K. Methods for estimating disease transmission rates: Evaluating the precision of Poisson regression and two novel methods. Scientific Reports. 2017 Aug; 7.
24. Zhang J, Nawata K. Multi-step prediction for influenza outbreak by an adjusted long short-term memory. Epidemiology and Infection. 2018 Apr; 146: p. 809–816.
25. Ye Z. Air Pollutants Prediction in Shenzhen Based on ARIMA and Prophet Method. E3S Web of Conferences. 2019; 136: p. 05001.
26. Yenidogan I, Cayir A, Kozan O, Dag T, Arslan C. Bitcoin Forecasting Using ARIMA and PROPHET. In 2018 3rd International Conference on Computer Science and Engineering (UBMK); 2018 Sep: IEEE.
27. Tyralis H, Papacharalampous GA. Large-scale assessment of Prophet for multi-step ahead forecasting of monthly streamflow. Advances in Geosciences. 2018 Aug; 45: p. 147–153.
28. Alabi RO, Siemuri A, Elmusrati M. COVID-19: Easing the coronavirus lockdowns with caution. 2020 May.
29. Bögl M, Aigner W, Filzmoser P, Lammarsch T, Miksch S, Rind A. Visual analytics for model selection in time series analysis. IEEE transactions on visualization and computer graphics. 2013 Oct; 19(12):2237-2246.
30. Taylor SJ, Letham B. Forecasting at scale. 2017 Sep.