Novel Coronavirus Pandemic in Indonesia: Cases Overview and Daily Data Time Series using Naïve Forecast Method

Annisa Puspa Kirana¹, Adhitya Bhawiyuga²
¹ Department of Information Technology, State Polytechnic of Malang, Indonesia
  Jl. Soekarno Hatta No.9, Malang, Jawa Timur 65141, Indonesia
² Faculty of Computer Science, University of Brawijaya Malang, Indonesia
  Jl. Veteran, Malang, Jawa Timur 65145, Indonesia

I. INTRODUCTION

A pandemic like COVID-19 is not a new health disaster. The World Health Organization (WHO) office of China received information on the first coronavirus case on December 31, 2019, in Wuhan, China. Coronavirus spread quickly to various countries in the world, including in Indonesia. Pandemic previously, e.g., Black Death "(1347-1351) or the" Spanish flu "(1918-1919), caused a lot of death and destruction in many cities and countries [1]. Indonesia is the fourth most populous country globally, so some experts predict that the COVID-19 pandemic has a significant negative impact and a longer pandemic period than other less populous countries. When the new Coronavirus, SARS-CoV-2, hit China during December 2019 - February 2020, the Indonesian government did not announce any cases of COVID-19 at all. On March 2, 2020, President Joko Widodo announced 2 cases of COVID-19 in Indonesia for the first time. The spread of the Coronavirus in Indonesia has increased significantly since the government announced the first cases. On April 2, 2020, the United States announced that there had been 1,790 positive cases of COVID-19, 113 people died, and 112 cases recovered [2]. After 9 months since the announcement of the first cases in Indonesia on December 9, 2020, the total number of positive cases of COVID-19 in Indonesia reached 592,900, making Indonesia the country with the highest COVID-19 cases in Southeast Asia [3].

The highest daily cases in Indonesia in 2020 occurred on December 4 2020, reaching 8,363 new cases in just one day [4]. Indonesia's death rate is a very extreme case, increasing to 17,199 deaths with a Case Fatality Rate (CFR) of 3.06% in December 2020 [5]. On December 31 2020, the total cases positively are 743,198, recovered 611,097, and died 22,138 [6].

This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).
The problems most faced by the Indonesian government are the failure to monitor information and the need for freely available data related to the spread of COVID-19 [7]. The government must prepare massive preparations to deal with the high spread of COVID-19 and reduce this pandemic's negative impact [8]. The government needs to take more careful action and react more quickly to solve the problem of the COVID-19 virus so that it does not worsen. Since the emergence of COVID-19 for the first time, several websites have released daily data updates about COVID-19. One of which is from the World Health Organization (WHO). A large number of daily has helped the emergence of mathematical analysis right about the development of COVID-19 in Indonesia to assist the government and authorities in decision-making for solving this pandemic. For that reason, the researchers took the title "Coronavirus (COVID-19) Pandemic in Indonesia: Cases Overview and Daily Time Series Data using Naïve Forecast Method".

Some researchers have analyzed the spread pattern of various diseases such as SARS, Ebola, Influenza pandemic, and dengue fever. Many previous studies use the COVID-19 daily dataset to analyze active cases, death rates, and recovered cases. S. Mahmud et al. in [9] have predicted and analyze COVID-19 data based on time series data. In that study, they used active daily data of COVID-19 to get data trends and predict the number of active cases. Several studies on the spread of COVID-19 have developed, especially in China, in [9]–[12]. Another study about COVID-19 in Indonesia is observed by [11]. This study predicts that the number of cases of the disease COVID-19 in Indonesia reaches more than 8,000 cases and end in April 2020. X.Song et al. in [13] analyzed influenza cases in one province in China starting from 2004 up to 2011. They analyzed influenza cases to get a trend of active cases. Besides of that, that study tried to analyze the distribution of region that has the highest cases. X.Song et al. in [13] authors used ARIMA to transform and perform analysis of time series. V.Tandon et al. predict the number of COVID-19 cases in India in [14]. That study result has a conclusion that the case in India increases exponentially. N.Shinde et al. in [15] used machine learning and statistical approaches to get a pattern and trend of several diseases.

Based on the background above, this study tries to summarize and predict the daily cases of COVID-19 in Indonesia graphically. The time series of data taken from March 2020 until January 2021. We try to analyze the COVID-19 data trend daily using daily data from the WHO and the algorithm ARIMA. We obtain data from www.covid19.who.int. We use the number of accumulated COVID-19 daily cases for our analysis. This study consists of five parts. (1) The first section briefly presents information about this research background and a general overview of Indonesia's COVID-19 cases. (2) The second section describes material and methods, including our dataset, area study, and proposed method. (3) The third sections explain this study results, including the cumulative number of cases, active cases, and the number of death cases due to the impact of COVID-19 for March 2 2020, up to January 31 2021. (4) The fourth section about discussion related to this research topic. (5) The fifth section consists of conclusions and suggestions. The analysis result gives us an overview of COVID-19 cases in Indonesia. This study also predicts the occurrence of COVID-19 cases. We hope this study can help the government as one of the analysis references in solving the coronavirus pandemic.

II. MATERIALS AND METHODS

This research examines the trend of COVID-19 cases in Indonesia. Researchers used daily time series data for COVID-19 cases in Indonesia taken from the first day to the last day in 2020.

1) Data

We obtained our training dataset and test dataset from the number of confirmed positive cases of COVID-19 in Indonesia. We use the dataset from March 2, 2020, to January 31, 2021. We obtained it from the World Health Organization (covid19.who.int), which is updated daily. Our analysis variables are daily active cases and daily death cases.

2) Studi Area

Indonesia is the fourth most populous country in the world, with 238,452,952 people [16]. As of January 27, Indonesia has reported 1,024,298 cases of COVID-19, the second-highest in Southeast Asia. In terms of passing rates, Indonesia ranks third in Asia and 17th globally [17]. The total number of cases died as of January 27 2021, as many as 28,855 people, and recovered as many as 831,330 people [18]. Indonesia has tested 3,907,273 people out of its 269 million population so far, or around 6,885 people per million [19]. The largest increase in new cases in one day occurred on December 3.

Fig 1. Geographical distribution of the cumulative number of confirmed COVID-19 cases in Indonesia [18]
The government announced 8,363 active cases and 171 died cases within 24 hours [20]. In this study, researchers analyzed positive cases of COVID-19 in Indonesia from March 2, 2020, to December 31, 2020. Figure 1 illustrates the geographic distribution of the cumulative number of confirmed COVID-19 cases in Indonesia across provinces as of January 27, 2021.

B. Methods

Time series analysis is a prediction method that focuses on analyzing random variables in the past. Time series forecasting creates a statistic model to predict the future value of random variables [21]. In this study, we used the Naïve Forecast method. Naïve forecast is one of the simplest forecasting methods, and it is very useful to be considered as a benchmark method for comparing models [22]. One of the time series forecasting methods is the Naïve Forecast method. The Naïve Forecast method observes a particular value of time before analyzing it as a current time model. For example, the forecast for 2020 is the observed value for 2019. The forecast for 2021 is the observed value for 2020 and so on. The authors argue that the Naïve Forecast method is one of the easy and very effective forecasting methods.

Naïve 1. The Naïve Forecast 1 method states that the forecast value for this period (t) is the same as the observed value for the last period (t-1) [23]

Naïve 2. The Naïve Forecast 2 method requires that period t is obtained by multiplying the number of cases occurring at this time then compared with the growth rate between previous cases in the period (t-1) [24]

The Naïve Forecast method is the most efficient prediction model and provides benchmarks against many improved models that are frequently compared. This forecasting technique is only suitable for statistical data. Naïve approaches are capable of producing the last observed value. This method works quite well for time series where patterns are difficult to predict. If the predicted time series have a seasonality pattern, the Seasonal Naïve Forecast approach is suitable. A variety of the Naïve Forecast method allows forecasting to increase or decrease over time. The amount of change over time is set to be the average change seen in historical data. Formulas for the Naïve method as follows (Formula 1) [24]:

\[ Naive\ Method = X_{t-1} \]

Where \(X_t\) = actual data in period t. So, the forecast value for the following data is the same as the last actual data. The number of forecasting test data for a single day was used to develop a chart for the number of positive COVID19 cases with respect to the number of tests conducted. We analyze the predicted value of positive cases based on the reported number of tests carried out per day and the average positivity rate. This research uses the daily active cases COVID-19 dataset from March 2020 to January 2021 from confirmed positive COVID-19 cases in Indonesia. We predict COVID-19 daily active cases in Indonesia from February 2021 to June 2021 using the Naïve Forecast method.

III. RESULTS

This research focused on cumulative daily figures collected globally from two main variables of interest: daily active cases and daily mortality cases. We obtained the COVID-19 dataset from open-access sites at ‘covid19.who.int’. Figure 2 illustrates the number of COVID-19 cases per one million population by provinces in Indonesia on January 27 2021.

1) Confirmed Cases from 2 Maret 2020 to January 31 2021

Figure 2 shows that DKI Jakarta, East Kalimantan, and North Borneo has the highest cumulative confirmed cases of COVID-19. As of February 2, 2021, Indonesia has reported 1,099,687 cases (Figure 3), the highest in Southeast Asia, above the Philippines. The rate of increase in the case of daily getting high within time. Additional cases per day experienced a trend in over 10,000 cases from the second week on January 2021.

![Cumulative confirmed cases of COVID-19 per one million population by provinces in Indonesia, as of January 27, 2021][18]

---

Accredited by Ministry of Research and Technology /National Research and Innovation Agency
Decree No: 200/M/KPT/2020
Figure 4 shows the distribution of the total of confirmed cases by province across the country. COVID-19 data released by the Ministry of Health of the Republic of Indonesia in February 2021 recorded 1,099,687 cases all over Indonesia. Figure 4 seen that regions that have darker colors have higher densities case. The region that has the highest case is still in Java island. Jakarta, West Java, Central Java, Yogyakarta, and East Java have the highest cumulative confirmed cases of COVID-19.

2) Death Cases from 2 Maret 2020 to January 31 2021

As of February 2 2021, total deaths reached 30,581 people, thus putting Indonesia in third place in Asia and 17th in the world [17]. The Case Fatality Rate (CFR) in Indonesia is much higher than in China (8.9% versus 4%) [26]. Figure 5 shows an illustration of death cases due to the coronavirus by the province in Indonesia. COVID-19 dataset obtained on February 2, 2021. It suggests that some provinces in Java and Kalimantan suffered the highest death cases[25].
3) Monthly Active Cases from 2 March 2020 to January 31 2021

We summarized the daily active cases into monthly active cases in Figure 7 based on new daily active cases in Indonesia. In October, 59.5% of confirmed cases were in Java: DKI Jakarta, East Java, West Java, and Central Java were the top four provinces in terms of the number of confirmed cases. South Sulawesi is the only province outside Java that in the top five provinces in terms of the number of confirmed cases. Cumulative number of confirmed COVID-19 cases by province[28]

From March to August 2020, the monthly death cases increased dramatically from 136 to 2,286. From August to January 2021, death cases increased significantly from 2,286 to 6,432 in January 2021 (Figure 9). The COVID-19 death rate in Indonesia illustrates the ability of the health service system to handle COVID-19 patients. Various factor contributes it for example availability of beds, medical personnel, medical equipment, and medicines. In December, the mortality rate, at 3.1%, was also higher than the global figure of 2.4% [29].

Fig 7. Monthly active cases of COVID-19 in Indonesia from March 2020 to January 2021

New active cases increases significantly from August to September and November to December. After 6 months of the first Coronavirus pandemic hitting Indonesia, new positive cases increased significantly. In the first month, March 2020, new cases in all regions in Indonesia reached around 1,528 cases per month. In just 6 months, new cases reached about 66,420 or 60 times higher than the first month. In January 2021, the new cases in one month reached 277,911 cases of new activities is nearly 5 times higher than in August 2020.

4) Monthly Death Cases from 2 March 2020 to January 31 2021

Indonesia continues to record a high death toll from COVID-19 higher than the global average. As of October 7, 2020, Jakarta's death toll has 172 confirmed death cases per one million population was the highest in Indonesia, followed by East Kalimantan, South Kalimantan, East Java, North Sulawesi, and Bali. Not a single province in Java has shown a decrease in the number of deaths among confirmed and likely cases since October 2020. DKI Jakarta has a higher number of deaths compared to deaths among confirmed cases [28]. Figure 8 illustrates the new monthly death cases based on new cases active daily in Indonesia from March 2020 to January 2021.

Various factors contribute to the high mortality rate from COVID-19 in Indonesia, including public health facilities and gaps in the health care system. Indonesia is still lagging behind other countries in testing COVID-19 [29]. Indonesia has only conducted 67,828 total tests until April 25, 2020. This figure is below other countries in the Southeast Asia region. Vietnam (206,253), Thailand (142,589), Malaysia 126,970, Singapore (121,775), the Philippines (81,292). Other factors in the equation include the slow response from authorities and local authorities in the early stages of an outbreak. In early March, there was only one laboratory in the entire country capable of carrying out a COVID-19 test. Few laboratory facilities in the first month of COVID-19 pandemic make a bad impact for Indonesia because of late patient treatment [29]. Thus, a fast and effective response from the government is needed most to decrease this pandemic effect in Indonesia.


We used the Naïve forecast method for the prediction of positive cases of COVID -19. We used the daily cumulative active cases dataset from all provinces in Indonesia (Figure 3) from March 2, 2020, to January 31, 2020. Different provinces may have different models according to the monthly incidence data. However, in this study, we selected all provinces to predict
daily active cases of COVID-19. The prediction is a daily case prediction for 5 months, starting from February 1, 2021, to June 31, 2021.

This research consists of 3 steps to build a prediction model using the Naïve Forecast algorithm. (1) The first step is to calculate the cumulative number of cases for Indonesia. Researchers first started with reported data (i.e., the cumulative daily number of new cases from March 2, 2020, to January 31, 2021, in Indonesia, as reported by the Indonesian Ministry of Health). (2) The second step is to apply the Naïve Forecast forecasting method. (3) The third step is to use predictive data based on the second step; We compare the predicted results with the results of real active cases then calculate the accuracy of the forecast results.

Figure 10 illustrates the daily real cases from March 2, 2020, to January 31, 2021, marked with a blue line. The red line illustrated forecasting daily cases starting from February 1, 2021, to June 30, 2021.

Figure 11 is a prediction graph for the monthly average cases of COVID-19. It can be seen that based on the results of the researched predictions, there will continue to be an increase in cases during the first 6 months of 2021 cases. The highest average case prediction is in June, which is 35,662 daily active cases. The highest increase in the average monthly case was from May 2021 to June 2021, where the difference in the average monthly cases reached 5,665 people (Figure 11).

IV. DISCUSSION

Naïve Forecast method gives good forecasts for medium to long term prediction. Regarding the total number of infected cases, the method gives good prediction up to 5 months for COVID-19 cases in Indonesia. As we can see on Figure 10 and Figure 11 daily predicted cases increased significantly with the difference in the average monthly cases reached 5,665 people. The result demonstrates that given data processed using time series Naïve Forecast method can learn and create model to precisely estimate the number of total populations that will become affected in the future. To find out the model performance, we compare the daily positive case output results with the generated models. Analysis of the error rate of our model using the Mean Absolute Percent Error (MAPE) index. MAPE is a calculation used to calculate the average absolute error percentage, with the Formula 2 [21]:

$$MAPE = \frac{\sum |Actual - Forecast| \times 100}{Actual}$$

The formula above means that $\sum |Actual - Forecast|\over Actual$ is the result of the reduction between the actual and forecast values.
that have been absolute. It is divided by the actual value for each period, and n is the number of periods used for calculation. The lower the MAPE value means that the forecasting model is good. Based on the prediction of the number of daily cases of COVID-19, the MAPE value is 15.85%. Based on the prediction, it can be concluded that in the first 6 months of 2021, the increase in corona cases is still high.

Naïve Forecast method is a novel method and can be applied to forecast other diseases or virus pandemic in future. We hope that our forecasts will be a useful tool for governments or related institution who are attempting to do prediction right now for the next few months. Furthermore, to reduce the number of positive cases of corona in Indonesia, the government needs to act decisively against people who violate the COVID-19 health protocol. The government is expected to continue to enforce strict health protocols indiscriminately. Also, maintain two-way communication between the government and the community. The public is also encouraged to have a high and collective responsibility to comply with health protocols. Because to suppress the corona outbreak, it starts from reducing the transmission rate. For this reason, the government emphasizes the importance of COVID-19 preventive behavior, such as wearing masks, washing hands, maintaining distance, avoiding crowds, and limiting mobility.

V. CONCLUSION

In this study, we analyzed daily data on COVID-19, especially in Indonesia. Our dataset is active daily case data from March 2, 2020, to January 31, 2020. We obtain our dataset from the World Health Organization website (covid19.who.int). We focused on cumulative daily figures collected globally from two main variables of interest: daily active cases and daily mortality—we predict positive cases of COVID-19 using the Naïve Forecast Method. The prediction is a daily case prediction for 5 months, starting from January 1, 2021, to June 31, 2021. The data used is daily data on cumulative active cases from all provinces in Indonesia. In this research, there are 3 steps taken to build a prediction model using the Naïve Forecast algorithm. (1) calculating the cumulative number of cases for Indonesia. (2) applying the Naïve Forecast forecasting method (3) using predictive data based on step number 2, we compare the prediction results with the results of real active cases then calculate the accuracy of the forecast results.

The highest average case prediction is in June, which is 23,479 daily active cases. The highest increase in the average monthly case was from May 2021 to June 2021, where the difference in the average monthly cases reached 5,665 people. Based on the prediction of the number of daily cases of COVID-19, the MAPE value is 15.85%. Based on predictive simulation calculations, we estimate that in 2021 the number of positive cases of COVID-19 increases significantly day by day. In the first 6 months, based on the prediction model, the highest cases occur in June and keep increasing in the following months. Therefore, the government is expected to continue to enforce strict health protocols indiscriminately. Useful and accurate policies to respond to the COVID-19 outbreak are crucial[30]. Some countries have succeeded in implementing strategic procedures to respond to it. At least three points should be considered deliberately by the governments to survive the disruption caused by the COVID-19 outbreak in the short run: (1) Increasing COVID-19 tests is a must. (2) Flattening the curve of COVID-19 cases through social or physical distancing is a crucial measure. (3) The government must protect the most vulnerable groups such as poor people or unemployed people and vulnerable sectors such as tourism, small and medium enterprise, or manufacturing sectors during the pandemic [30]. At last, future modifications from this work is needed to improve the predictive accuracy of the forecasting models and forecasting performance. For future work, combining various variable such as location, duration of social distance, and smaller region such as cases prediction in each province can be useful to generate more accurate model for each region.

REFERENCES


