Implementation of the prophet model in COVID-19 cases forecast

Rodiah [a,1]; Eka Patriya [a,2,*]; Diana Tri Susetianingtias [a,3]; Ety Sutanty [a,4]

[a] *Gunadarma University, Margonda Raya Street 100, Depok and 16424, Indonesia*

[1] *rodiah@staff.gunadarma.ac.id;* [2] *ekapatriya@staff.gunadarma.ac.id;* [3] *diants@staff.gunadarma.ac.id;* [4]*ety_s@staff.gunadarma.ac.id*

* *Corresponding author*

Abstract

One of the steps to understanding this pandemic is to look at the spread of the data by predicting an increase in cases in various countries so that prevention can be carried out as early as possible. One way to see fluctuations in COVID-19 pandemic data is to predict the rate of cases using forecasting methods so that conclusions can be drawn on the spread of COVID-19 pandemic data around the world to be processed using statistical models. This study will implement the use of the Prophet Model in seeing the rate of development of COVID-19 in the world using four features in the forecasting process such as the number of confirmed cases, the number of cases of recovered patients, the number of cases of death, and the number of active cases. The results of this study produce forecasting data on the number of cases of the COVID-19 pandemic that can be viewed daily, weekly, and even monthly. Forecasting results show the first spike at the end of March until the number of cases reached around 10,275,800 million as of June 29, 2020, where the number of cases grew exponentially until June 29, 2020. The case rate of growth in many instances experienced significant growth until the end of October, touching the number in the range of 34,507,150 million as of October 25, 2020. After June 29, 2020, a very high spike was different from the increase in cases in the previous months. Forecasting results show no point decline because historical data on the number of daily confirmed cases of the COVID-19 pandemic has not decreased. The forecasting results in this study are expected to be able to systematically predict events or events that will occur in the COVID-19 pandemic around the world with the help of valid periodic data so that some information can be obtained for preventive measures related to the COVID-19 pandemic.

Keywords: COVID-19; Features; Cases; Forecasting; Prophet Model.

Introduction

The 2019 Coronavirus Disease (COVID-19) pandemic is an international health outbreak caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1]. This pandemic was started by cases of the outbreak of residents of Wuhan, China, and spread throughout the world [2]. Through human activities, this virus so quickly infects victims through physical contact. Therefore, it spreads quickly in a short time [3]. The virus threatens part of the world's population, affecting several countries' health, socioeconomic, and finances [4], including Indonesia. Some of them have experienced an increase in cases every day since this pandemic emerged and are ranked the highest in the world, with the most cases of the COVID-19 pandemic [5] during a certain period. Various efforts and government policies in each country have been carried out to reduce the number of pandemic cases [6] and look for healing methods simultaneously so that the infectious virus does not continue to spread [7] and disrupt the economy in each country. The results obtained are used to see the movement of the data [8] of the infected population and can be used for research purposes in general.

Research related to the implementation of forecasting models in seeing the rate of cases of the spread of COVID-19 was carried out by previous researchers. Research [9] implemented a combined autoregressive integrated moving average model (ARIMA) and Wavelet-based forecasting model (WBF Model) to predict the latest number of COVID-19 pandemic cases from five countries consisting of Canada, France, India, South Korea, and Britain. Research [9] obtained the best accuracy value for the Hybrid ARIMA-WBF Model using the root mean square error (RMSE) of 621.91 and (MAE) of 306.78. Research [10] implemented the autoregressive time series model (TP-SMN-AR) in predicting the number of confirmed cases and the number of recovered patients. Research [10] obtained the best accuracy for the TP-SMN-AR model using the mean absolute percentage error (MAPE) for the confirmed case rate of 0.22% and the cured patient case rate of 1.6%. Research [11] collects world data from cases

of the COVID-19 pandemic to see the relationship between data collected daily. The correlation between the total number of confirmed cases, active cases, deaths, and the number of positive cases is stated in this study. It shows the results of the comparison of the two regression models validated by R2. There was a 68% increase in active cases in Odisha for every one-unit increase in positive cases in India, representing a 57% increase in active cases of the COVID-19 pandemic [11]. The score represents the value of R2, namely 0.995588 and 0.974855, which states that this is a strong forecasting model [11]. Research [12] Long Short-Term Memory model predicts the total number of deaths and real-world cases selected based on particular countries and their performance comparisons. Research [12] uses two types of datasets, including the actual dataset sourced from ourworlddata.org for univariate time series forecasting and data visualization, starting from January 1, 2020, to April 22, 2020. The simulated data is connected to actual data from countries that are ranked as countries with COVID-19 dangerous zones from January to March 2020. The study [13] used a logistic growth model to find the stability of the pandemic and the prophet model to forecast the total number of confirmed cases caused by the COVID-19 pandemic in India. The study [13] used parameters: hyperparameter (hp), multiplicative factor (mf), and the maximum value of infected patients, which is the carrying capacity in this case (cc). Research [13] found that using logistic curve trends provides a much better estimate for the model used to analyze the spread of the COVID-19 pandemic.

In this study, the prophet model was implemented to spot the number of COVID-19 developments based on the number of cases that occurred worldwide, including in Indonesia. This study uses four features in the forecasting process such as the number of confirmed cases, the number of cases of recovered patients, the number of cases of death, and the number of active cases, which will then be presented with the results for each feature starting from a certain period. The data is then entered into the prophet model to produce forecasting data on the number of COVID-19 pandemic cases that can be viewed daily, weekly, and even monthly. The model's accuracy is determined based on the forecasting results obtained by taking the highest accuracy value using the mean squared error (MSE). The forecasting results in this study are also expected to be able to systematically predict events or events that will occur in the COVID-19 pandemic around the world with the help of valid periodic data so that some information can be obtained for preventive measures related to this pandemic.

Method

This research was carried out with 7 stages in this study, namely determining the source data, initializing predictive variables, calculating additive regression models, calculating non-linear trend models, determining the location of change points, calculating error values using MSE, and visualizing the results of forecasting and scientometrics data from the COVID-19 pandemic. 19. The stage of determining the source data is the stage to determine the input data which consists of several important features to be used as processing data in presenting the visualization of forecasting results and processing forecasting with the prophet model. The data is drawn from the source, including Kaggle for the purposes of forecasting visualization and then it will be passed to the next stage, namely the stage of initializing the forecasting variable. This study uses the prophet model [14] to observe the pattern of functional relationships between response variables and predictor variables [15]. In general, the stages of this research can be seen in **Figure 1**.
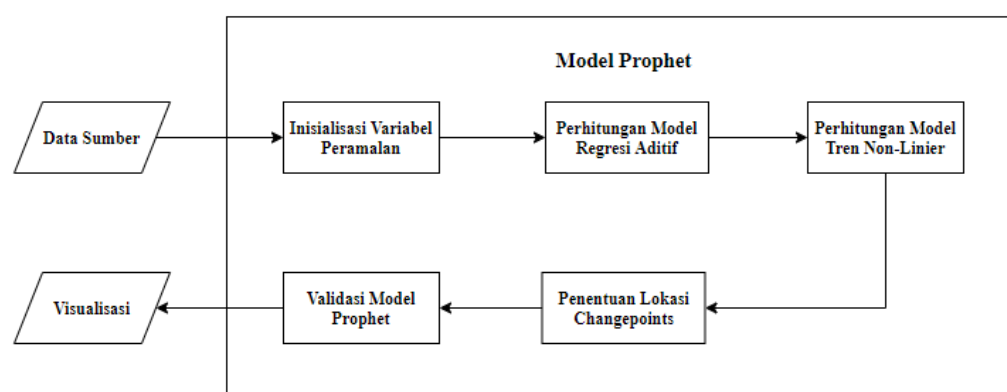

Figure 1 . Stage of COVID-19 Forecasting Research

A. Source Data

The data obtained is a public dataset [13] using three datasets, including the "covid_19_clean_complete.csv" dataset to be used in the overall visualization and forecasting. The "day_wise.csv" dataset is only used to present daily data on the number of cases of the COVID-19 pandemic in the world as of a date which consists of the number of confirmed cases, the number of cases of recovered patients, the number of cases of death, and the number of

active cases of the COVID-19 pandemic in the world through scientometrics visualization to see the movement of the pandemic data. This study also uses the "country_wise_latest.csv" asset which is also used only to see daily data per country name that has the world's daily number of COVID-19 pandemic cases consisting of confirmed case numbers, recovered patient cases, death cases, and active cases of the COVID-19 pandemic in the world. **Figure 2** is an example of a sample dataset from covid_19_clean_complete.csv [13].

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Active	WHO Region
49063	NaN	Sao Tome and Principe	0.186400	6.613100	2020-07-27	865	14	734	117	Africa
49064	NaN	Yemen	15.552727	48.516388	2020-07-27	1691	483	833	375	Eastern Mediterranean
49065	NaN	Comoros	-11.645500	43.333300	2020-07-27	354	7	328	19	Africa
49066	NaN	Tajikistan	38.861000	71.276100	2020-07-27	7235	60	6028	1147	Europe
49067	NaN	Lesotho	-29.610000	28.233600	2020-07-27	505	12	128	365	Africa

Figure 2. Sample dataset "covid_19_clean_complete.csv" [13]

B. Formation of Prophet Model

This study uses the prophet model in predicting periodic data from the number of cases of the COVID-19 pandemic to obtain an estimated number of cases that occur within a certain time. Using the Prophet Model in historical forecasting data as an ingredient in determining forecasting the spread of COVID-19 numbers is the proper method where this method has a change point to see any trend changes that suddenly fluctuate in certain months of spike cases of COVID-19 figures. Forecasting in this study will focus on two important cases, including the number of cases of the global COVID-19 pandemic and the number of cases of the COVID-19 pandemic in Indonesia. This study focuses on issues surrounding the number of cases of the COVID-19 pandemic using the features of "Confirmed" to represent the number of confirmed cases of infected patients, "Deaths" to describe the number of cases of death of patients infected with the COVID-19 pandemic, "Recovered" to represent the number of cases of recovered patients who contracted the COVID-19 pandemic, and "Active" to describe the number of active cases of patients infected with the COVID-19 pandemic sorted by date of occurrence. The results obtained can be used as an early warning for cases of the COVID-19 pandemic around the world.

The following are the stages of forming the prophet model that is used to see the spread of COVID-19 figures based on the number of cases that occur in the world, including Indonesia:

1. Initialization of Forecasting Variables

This stage is the initial stage of forecasting the number of cases of the COVID-19 pandemic by initializing predictive variables from the related features used in this process. To obtain the data in the variables, it is necessary to read the dataset after collection. Because the data used is data in .csv format, a command is needed to read the data. Here is a command snippet to read a .csv formatted dataset using the Pandas library.

Covid is a variable used to accommodate the features contained in the "covid_19_clean_complete.csv" dataset by reading data in the read_csv() function. A pd variable is an object for accessing Pandas' built-in attributes and functions using *pseudocode*:

```
covid = pd.read_csv('/content/files/covid_19_clean_complete.csv')
```

The next step is to adjust the feature name to be included in the coding because two features are obtained that do not follow the variable naming rules. This is done by renaming the column or feature in the dataset. These features are "Province/State," which is changed to "Province_State" and "Country/Region," which is changed to "Country_region." Forecasting in this study uses four features that represent existing problems consisting "Confirmed" for the number of confirmed cases of infected patients, "Deaths" for the number of cases of death of patients infected with COVID-19, "Recovered" for the number of cases of recovered patients infected with COVID-19. -19, and "Active" for the number of active cases of patients infected with COVID-19.

2. Calculation of Additive Regression Model

The Additive Regression Model is a statistical regression model in which the systematic component is the arithmetic sum of the individual effects of the predictors [16]. At the initial stage, to predict a feature, if you want to predict the number of COVID-19 cases in the world, you can enter the world_pred variable into a variable called data. On the other hand, if the data you want to predict is the number of COVID-19 cases in Indonesia, the indonesia_pred variable is entered into the same variable, namely the data variable. Furthermore, every one of the 4 features that have been previously selected is entered into a variable representing one problem, namely a variable

named feature. From the results of sorting the data on the command data.loc[:, ['Date', feature]] then stored into the pred variable. It should be noted that the input in the prophet model only has two features [17], namely "ds" and "y". The "ds" feature is a feature that will store periodic time data. This feature is always paired with the "y" feature which is a feature that holds the data you want to predict or the target data. The illustration of the ds and y columns used in forecasting by the prophet model can be seen in **Table 1**.

Table 1. Illustration of ds and y columns for forecasting data

	ds	**Y**
0	2021-01-01	146376
1	2021-02-01	147079
2	2021-03-01	159336
3	2021-04-01	163669
4	2021-05-01	170068

When instantiating an object named model in the command model = Prophet (changepoint_prior_scale=0.5) then the following additive regression model equation is executed in that command given that the prophet model uses equation [18]:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \qquad (1)$$

```
pred = data.loc[:, ['Date', feature]]
pred.columns = ['ds', 'y']
model = Prophet(changepoint_prior_scale=0.5)
model.fit(pred)
```

So in *psudocode* the equation is written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (2)$$

So that the final result is obtained in the form of a sum of g(t) as a trend that has non-periodic changes [19] where the trend describes the movement of time series data over a long period or long enough and tends to go up or down in one direction, s(t) as seasons that change periodically from daily to yearly with the help of Fourier series, h(t) series effect, and (t) which is an error, the above command takes arguments to configure the desired model type, such as growth type, seasonal type, and more. Next, the model of fit (pred) command is executed to train the previously entered data to be prepared and sent to the following process. The function retrieves a data frame from time series data.

3. Calculation of Non-Linear Trend Model

At this stage, a non-linear trend model is calculated where the trend line should not and is not always a linear line [20]. There are also non-linear trend lines. Each trend illustrates the movement on average [21] or as a whole. The non-linear trend is a measure of the tendency with a model with an equation of the power of two, the power of three, and so on [22]. The advantage of this method is that it is perfect for long-term data, and the forecast results are close to the actual value, but this method is not suitable for short-term data [23]. This study uses a non-linear trend method because the data used are long-term data, and based on the historical data plot, it is known that the data is not continuous and is non-linear. The prophet model will provide a forecast interval for the trend component by simulating future trend changes on the time series [24] using the following pseudocode:

The make_future_dataframe() function generates forecast data frames from model variables with a period of 90 days ahead. The data frame obtained is an extension of the "ds" feature, which consists of a specific number of days. This function describes the frequency and number of periods that people want to forecast in the future. For example, **Table 2** shows the forecasting results obtained after entering the future variable into the prediction (future) function.

Table 2. Illustration of Future Variable Forecasting Results

	ds	**yhat**	**yhat_lower**	**yhat_upper**
294	2021-06-30	13.057100	13.037462	13.075858
295	2021-07-31	13.081213	13.061035	13.100568

	ds	yhat	yhat_lower	yhat_upper
296	2021-08-31	13.014274	12.993798	13.034220
297	2021-09-30	13.036700	13.016954	13.056166
298	2021-10-31	13.054154	13.031567	13.074649

4. Determining the Location of Changepoints

In forecasting, real-time often has sudden changes in its trajectory. By default, the prophet model will automatically detect this changepoint and allow the trend to adapt accordingly. The prophet model detects change points by first determining a large number of potential change points at which the rate of change is allowed to change [25]. By default, the prophet model defines 25 possible changepoints that are uniformly placed in the first 80% of the time series. The vertical lines in **Figure 3** indicate where changepoints are defined.
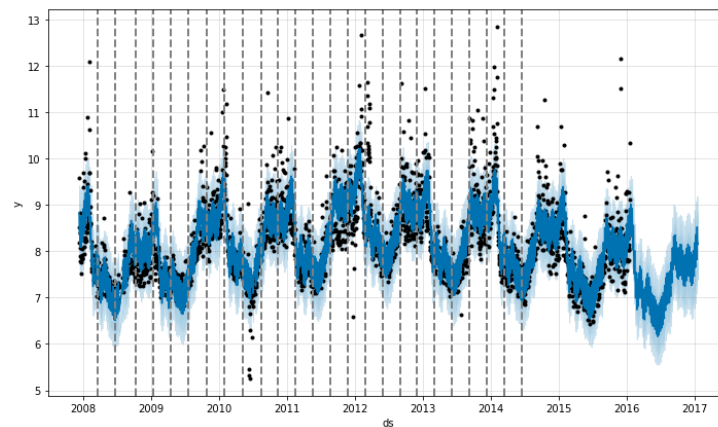

Figure 3. Illustration of locating changepoints

```
future = model.make_future_dataframe(periods=90)
forecast = model.predict(future)
metric_df = forecast.set_index('ds')[['yhat']].join(pred.set_index
('ds').y).reset_index()
metric_df.dropna(inplace=True)
print ('mse score :', mean_squared_error(metric_df.y, metric_df.yhat))
```

By default, inflection points are only inferred for the first 80% of the time series to project the trend forward and to avoid overfitting fluctuations at the end of the time series. The following is the determination of changepoints in the prophet model.

Suppose the trend change becomes overfit, has too much flexibility, underfit, or is not flexible enough. In that case, the strength of the sparse prior can be adjusted using the changepoint prior_scale input argument. By default, this parameter is set to 0.05, where increasing it will make the chart trend more flexible.

5. Prophet Model Validation

In evaluating the forecasting results, this study uses the mean squared error (MSE) [26] to measure the mean squared error, including the difference in the average square between the estimated value and the actual value where each error or remainder is squared using the **Equation 3**:

$$MSE = \frac{\sum_{t=1}^{n} |Y_t - \hat{Y}_t|^2}{n} = \frac{\sum \text{forecast errors}^2}{n} \qquad (3)$$

This study will measure the error value of the two data using periodic data from August 24, 2020, to August 26, 2020, for several cases, such as the value of the error in the world number of confirmed cases, the number of world deaths, the number of cases of recovered patients in the world, the number of active cases in the world, the number of cases Indonesia's confirmation, Indonesia's death rate, Indonesia's number of recovered cases, Indonesia's active case number. The following is pseudocode for calculating errors using MSE using the sklearn library. The MSE value is calculated by calling the default function from sklearn [27] using mean_squared_error(metric_df.y, metric_df.yhat) followed by inputting parameters in the form of the actual value for the y variable and the forecasted value for the yhat variable.

6.   Visualization of Scientometrics COVID-19 Pandemic Data using Map

The COVID-19 pandemic has hit every part of the world by spreading fast in a certain amount of time. In this study, the distribution of the pandemic data through visualization using geospatial plots [28] can be done interestingly through scientometrics visualization. The visualization of this study uses data on the spread of COVID-19 case numbers around the world through geographic displays to find out which countries have the potential to spread this virus and infect the human population in that country exponentially. The distribution of data to be seen includes the number of confirmed cases, the number of cases of recovered patients, the number of cases of death, the number of active cases, the number of the latest confirmed cases, the number of the latest death cases, the number of fatalities per 100 confirmed cases, the number of cases of recovered patients per 100 number of confirmed cases, the number of cases of death per 100 number of cases of recovered patients, the number of confirmed cases last week since the end of this data was updated on April 20, 2022, changes in the number of confirmed cases of COVID-19 in the world for one week, an increase the number of confirmed cases of COVID-19 for one week with a percentage count as in the following pseudocode.

```
future = model.make_future_dataframe(periods=90)
forecast = model.predict(future)
def plot_map(df, col, pal):
        df = df[df[col]>0]
        fig = px.choropleth(df, locations="Country/Region", locationmode='country names',
        color=col, hover_name="Country/Region", title=col, hover_data=[col], color_continuous_scale=pal)
        fig.show()
plot_map(country_wise, 'Deaths', 'matter')
plot_map(country_wise, 'Recovered', 'matter')
plot_map(country_wise, 'Active', 'matter')
plot_map(country_wise, 'New cases', 'matter')
plot_map(country_wise, 'New deaths', 'matter')
plot_map(country_wise, 'Deaths / 100 Cases', 'matter')
plot_map(country_wise, 'Recovered / 100 Cases', 'matter')
plot_map(country_wise, 'Deaths / 100 Recovered', 'matter')
plot_map(country_wise, 'Confirmed last week', 'matter')
plot_map(country_wise, '1 week change', 'matter')
plot_map(country_wise, '1 week % increase', 'matter')
```

In the code snippet above, the scientometrics visualization of the COVID-19 pandemic data in the world is carried out using the country_wise variable from the "country_wise_latest.csv" dataset for geospatial visualization. The first step is to pass the country_wise variable into the parameter of the plot_map() function, followed by the name of the feature to be presented in a geospatial plot, and then pass the value named 'matter' into the function parameter to apply the color scale. After that, the country_wise variable will be passed into the choropleth() function to form a geospatial plot model, as seen in **Figure 4**.
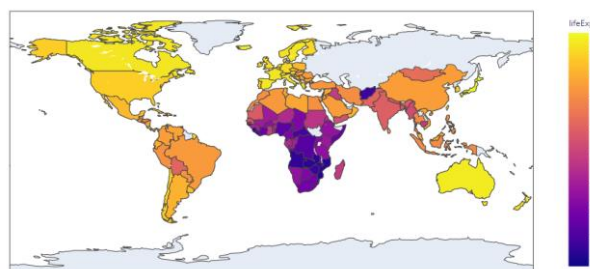

Figure 4. Illustrations of geospatial plots of COVID-19 pandemic data

Then it is also displayed a geospatial plot in the form of geographic points from countries affected by the COVID-19 pandemic in the world to describe which countries were affected by the pandemic case through the goepandas library [29] [30] using the following pseudocode.

```
geo_covid = gpd.GeoDataFrame(covid,geometry=gpd.points_from_xy(covid['Long'], covid['Lat']))
points = [Point(x,y) for x,y in zip(covid.Long, covid.Lat)]
fig,ax = plt.subplots(figsize=(20,10))
geo_covid.plot(cmap='hsv',ax=ax)
world.geometry.boundary.plot(color=None,edgecolor='k',
linewidth=2,ax=ax)
```

Geographic points are saved to a variable named points after iterating over geographic points on the x and y axes on the command line points = [Point(x,y) for x,y in zip(covid.Long, covid.Lat)]. The next step is to visualize geospatial scientometrics by assigning color to geographic points in the geo_covid.plot (cmap='hsv',ax=ax) command and display a graphical plot of the COVID-19 pandemic data in the world by running the command: *world.geometry.boundary.plot(color=None,edgecolor='k',linewidth=2,ax=ax)*. **Figure 5** is a geospatial plot in presenting the distribution of data on the COVID-19 pandemic in the world.
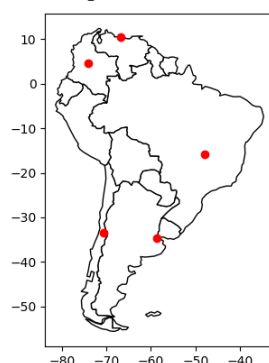

Figure. 5 Illustrations of geospatial plots of COVID-19 pandemic data

Results and Discussion

The prophet model stage is the initial stage conducted before carrying out the visualization process. This stage consists of three processes: initializing prediction variables, calculating additive regression models, calculating non-linear trend models, determining the location of changepoints, validating prophet models, and visualizing forecasting and scientometrics results of COVID-19. The results at this stage are in the form of forecasting the number of COVID-19 cases worldwide and in Indonesia. **Figure 6** and **Figure 7** are the results of the number of confirmed cases of the COVID-19 pandemic in the forecasting world on a variable called forecast the number of confirmed cases of the COVID-19 pandemic in the world. It can be clearly seen below, on October 25th, 2020, that the trend column has a value of 3.449991e+07 while the yhat column, which is the result of forecasting, gets a value of 3.450715e+07.

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upper	weekly
273	2020-10-21	3.388185e+07	3.055497e+07	3.745101e+07	3.058703e+07	3.737808e+07	-10667.120330	-10667.120330	-10667.120330	-10667.120330
274	2020-10-22	3.408791e+07	3.066358e+07	3.770691e+07	3.070828e+07	3.764865e+07	-770.888053	-770.888053	-770.888053	-770.888053
275	2020-10-23	3.429398e+07	3.086315e+07	3.797288e+07	3.082632e+07	3.792738e+07	10375.003141	10375.003141	10375.003141	10375.003141
276	2020-10-24	3.450004e+07	3.094967e+07	3.821762e+07	3.096369e+07	3.820144e+07	12797.019219	12797.019219	12797.019219	12797.019219
277	2020-10-25	3.470610e+07	3.112599e+07	3.843816e+07	3.109884e+07	3.846992e+07	6791.277065	6791.277065	6791.277065	6791.277065

Figure 6. Forecasting result on forecast variables from the number of confirmed cases of the COVID-19 pandemic worldwide

weekly_lower	weekly_upper	multiplicative_terms	multiplicative_terms_lower	multiplicative_terms_upper	yhat
-10667.120330	-10667.120330	0.0	0.0	0.0	3.387119e+07
-770.888053	-770.888053	0.0	0.0	0.0	3.408714e+07
10375.003141	10375.003141	0.0	0.0	0.0	3.430435e+07
12797.019219	12797.019219	0.0	0.0	0.0	3.451283e+07
6791.277065	6791.277065	0.0	0.0	0.0	3.471289e+07

Figure 7. Forecasting results on forecast variables from the number of confirmed cases of the COVID-19 pandemic worldwide

Figure 8 and **Figure 9** are the results of the number of confirmed cases of the COVID-19 pandemic in the forecasting world on a variable called forecast the number of confirmed cases of the COVID-19 pandemic in the world. It can be seen below on October 25, 2020, that the trend column has a value of 2.166477e+07 while the yhat column, which is the result of forecasting, gets a value of 2.166368e+07.

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upper	weekly
273	2020-10-21	2.111122e+07	1.865535e+07	2.378946e+07	1.866757e+07	2.377059e+07	-4700.966196	-4700.966196	-4700.966196	-4700.966196
274	2020-10-22	2.124960e+07	1.872443e+07	2.395417e+07	1.878341e+07	2.394361e+07	1428.074741	1428.074741	1428.074741	1428.074741
275	2020-10-23	2.138799e+07	1.888904e+07	2.414327e+07	1.889224e+07	2.411338e+07	6270.307854	6270.307854	6270.307854	6270.307854
276	2020-10-24	2.152638e+07	1.896047e+07	2.426498e+07	1.898134e+07	2.429102e+07	8934.756091	8934.756091	8934.756091	8934.756091
277	2020-10-25	2.166477e+07	1.906280e+07	2.447117e+07	1.908381e+07	2.447898e+07	-1090.464121	-1090.464121	-1090.464121	-1090.464121

Figures 8. Forecasting results on forecast variables from the number of active patient cases in the COVID-19 pandemic in the world

weekly_lower	weekly_upper	multiplicative_terms	multiplicative_terms_lower	multiplicative_terms_upper	yhat
-4700.966196	-4700.966196	0.0	0.0	0.0	2.110651e+07
1428.074741	1428.074741	0.0	0.0	0.0	2.125103e+07
6270.307854	6270.307854	0.0	0.0	0.0	2.139426e+07
8934.756091	8934.756091	0.0	0.0	0.0	2.153532e+07
-1090.464121	-1090.464121	0.0	0.0	0.0	2.166368e+07

Figure 9. Forecasting Results on forecast variables from the number of active patient cases for the COVID-19 pandemic worldwide

A.   Results of Visualization of Data Forecasting COVID-19 Pandemic Data

This stage is the final stage that produces scientometrics visualization of the number of COVID-19 cases in the world and in Indonesia. It is also generates scientometrics visualization of COVID-19 data in the world through interactive graphic plots. As clearly seen in **Figure 10** is the result of forecasting the number of confirmed cases of COVID-19 in the world as well as the results of trends and seasonality.

Figure 10. Results Forecasting Number of Confirmed Cases of the COVID-19 Pandemic in the World

It can be observed in **Figure 10** that the number of confirmed cases of the COVID-19 pandemic in the world began in February, and there have been no signs of growing pandemic cases since the international community declared the danger of this pandemic. Finally, the data showed the first spike at the end of March until the number of cases touched around 10,275,800 million as of June 29, 2020, as seen in **Figure 11**. The number of cases grew exponentially until June 29, 2020. Finally, the number of cases increased significantly until the end of October, touching 34,507,150 million as of October 25, 2020, as shown in **Figure 12**. After June 29, 2020, it can be seen in the graph above that there is a very high spike that is different from an increase in the number of cases in the previous months. The prediction results do not seem to have decreased by a single point because historical data shows that the daily number of confirmed cases of the COVID-19 pandemic does not offer any decline concerning this data. **Figure 13** shows the trend pattern in the number of confirmed cases of COVID-19 in the world and the movement of weekly data related to this pandemic.
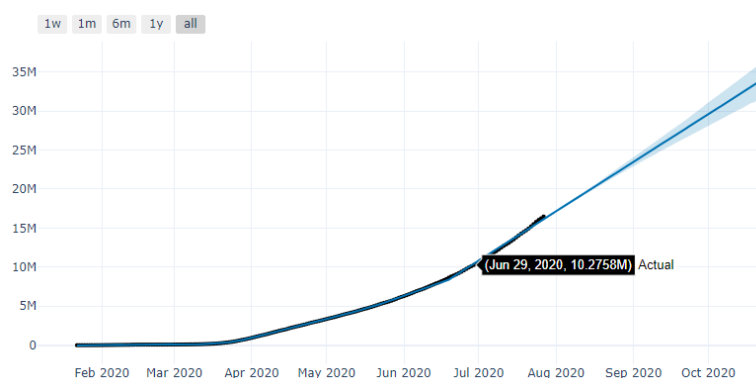

Figure 11. Actual Data on the Growth in the Number of Confirmed Cases of the COVID-19 Pandemic in the World
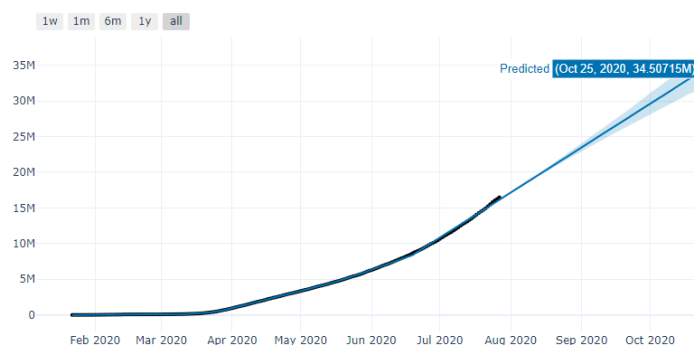

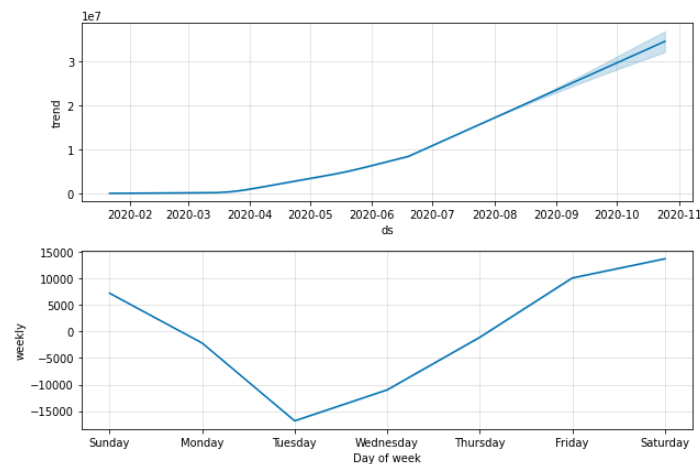**Figure 12.** Forecasting Data for Confirming COVID-19 Pandemic Cases in the World

Figure 13. Trending Patterns in the Number of Confirmed COVID-19 Cases in the World and Weekly Data Movements

The results of the weekly forecast explain that from Sunday to Tuesday, it is envisioned that there will be a decline in the range of 15,000 cases, then the figure will bounce high until Friday. Then it slowly incline approaching Saturday in the range of 15,000 cases. **Figure 14** is the result of forecasting the number of cases of recovered patients from the COVID-19 pandemic in the world, as well as seasonal trends.

Figure 14. Results of Forecasting Case Numbers of Patients Recovering from the COVID-19 Pandemic in the World

The graphic above shows that the results of forecasting the number of recovered cases of the COVID-19 pandemic in the world experienced a small number of recovered cases from February to April. At the beginning of May, there was good news, namely a spike in the number of confirmed cases in line with the increasing number of cases of recovered patients, which indicated that the patients had been handled well. Until July 27, 2020, the number of recovered patients continued to grow by touching the figure of 9,468,087 million, and it is predicted that it will continue to increase until the end of October with a range of 21,663,680 people. Actual data on the number of cases of patients recovering from the COVID-19 pandemic in the world can be seen in **Figure 15**.
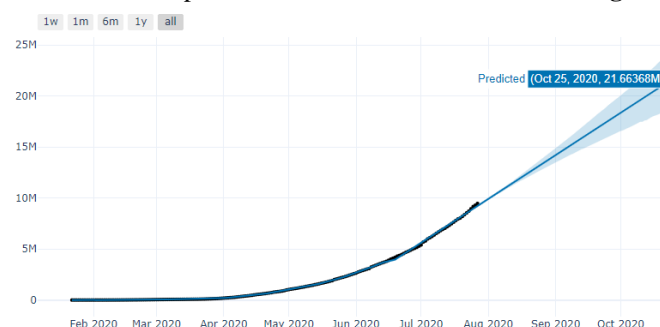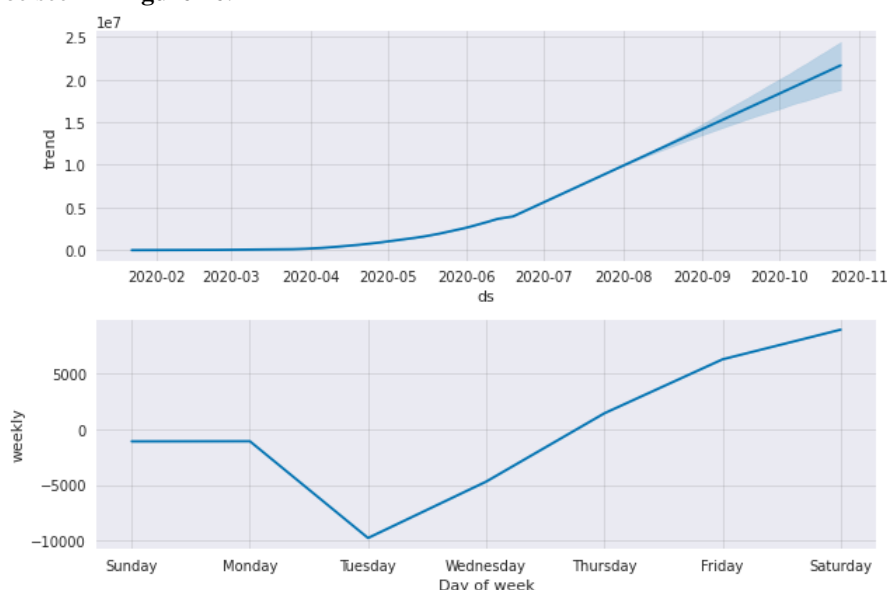

Figure 15. Actual Data on the Growth in Cases of Patients Recovering from the COVID-19 Pandemic in the World

The trend pattern in the number of cases of recovered COVID-19 patients in the world and the movement of weekly data can be seen in **Figure 16**.

Picture 16. Trending Patterns in the Number of Cases of Recovered COVID-19 Patients in the World and Weekly Data Movements

B. *Prophet Model Validation Results*

The prophet model was successfully validated by generating prophet accuracy values using manual calculations or sklearn's function, hence, manual accuracy results were obtained of 2,068,881,057,325,967 for the number of confirmed cases of COVID-19 in the world, 1,398,832,085.6 for the number of deaths caused by COVID-19 in the world, 5,906,937,881,429.5 for the number of recovered COVID-19 patients around the globe, 1,015,331,005,936.9 for the number of active COVID-19 cases in the world. Furthermore, there is a calculation of the error value for the number of confirmed cases of COVID-19 in Indonesia of 280,281,948.2, the number of cases of death caused by COVID-19 in Indonesia of 155,420.7, the number of cases of recovered COVID-19 patients of 560,918.811.8, and the number of recovered cases of COVID-19 active cases of COVID-19 in Indonesia amounted to 45,656,877.6.

Conclusion

The formation of the prophet model for forecasting the number of COVID-19 cases in the world and Indonesia was successfully carried out by performing calculations using an additive regression model using two columns, ds, and y, as dataframes for forecasting variables. The calculation of the non-linear trend has succeeded in producing a new dataframe in the form of the latest ds column that already has a periodic date that will be predicted in 90 days. The calculation of the error value resulted in the number of confirmed cases of COVID-19 in the world at 6151932550.422536, the number of deaths from COVID-19 in the world was 2341932.9982180223, the number of cases of recovered COVID-19 patients around the world was 3745400500.1680226, and the number of active cases of COVID-19 in the world was 1273740778.980077. In addition, there is a calculation of the number of confirmed COVID-19 cases in Indonesia of 407601.9973343974, the number of COVID-19 deaths in Indonesia of 2684.5462191016004, the number of cases of recovered COVID-19 patients in Indonesia is 1143303.7984896912, and the number of active cases in Indonesia is 420645.4986416476. The chart trend shows a growth in the number of confirmed cases which began at the end of March and continued to grow until June 29, 2020. However, the rate of cases did not seem to have decreased until after June 29, 2020. The case rate increased significantly until the end of October 2020.

Further development can be done to improve the visualization of COVID-19 data forecasting using the prophet model assisted by applying an additive regression model.

References

[1] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan," China.Lancet. 2020;395(10223):497-506.

[2] World Health Organization. "Naming the coronavirus disease (COVID-19) and the virus that causes it [Internet]," Geneva: World Health Organization; 2020 [cited 2020 March 29]. Available from:

https://www.who.int/emergencies/diseases/novelcoronavirus-2019/technical-guidance/naming-the-coronavirusdisease-(covid-2019)-and-the-virus-that-causes-it.

[3] Wu Z, McGoogan JM. "Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19)," Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. JAMA. 2020; published online February 24. DOI:10.1001/jama.2020.2648.

[4] Kementerian Kesehatan Republik Indonesia." Info infeksi emerging Kementerian Kesehatan RI [Internet]," 2020 [updated 2020 March 30; cited 2020 March 31]. Available from: https:// infeksiemerging.kemkes.go.id/

[5] Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., & Chowell, G. "Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang," China: February 13–23, 2020. Journal of Clinical Medicine, 9(2), 596.

[6] Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, Nanda C, Sharma S, Sharma YD, Rabaan AA, Rahmani J, Kumar P. "Prediction of the COVID-19 Pandemic for the top 15 affected countries: Advanced Autoregressive Integrated Moving Average (ARIMA) model," JMIR Public Health Surveill. 2020 May 13;6(2):e19115. doi: 10.2196/19115. PMID: 32391801; PMCID: PMC7223426.

[7] Baloch S, Baloch MA, Zheng T, Pei X. "The Coronavirus Disease 2019 (COVID-19) Pandemic," Tohoku J Exp Med. 2020 Apr;250(4):271-278. doi: 10.1620/tjem.250.271. PMID: 32321874.

[8] Wang, X., Liu, S., & Huang, Y. A." Study on the rapid parameter estimation and the grey prediction in richards model," Journal of Systems Science and Information, 4(3), 223-234.2016

[9] Chakraborty T, Ghosh I. "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis," Chaos Solitons Fractals. 2020 Jun;135:109850. doi: 10.1016/j.chaos.2020.109850. Epub 2020 Apr 30. PMID: 32355424; PMCID: PMC7190506.

[10] Maleki M, Mahmoudi MR, Wraith D, Pho KH. "Time series modelling to forecast the confirmed and recovered cases of COVID-19," Travel Med Infect Dis. 2020 Sep-Oct;37:101742. doi: 10.1016/j.tmaid.2020.101742. Epub 2020 May 13. PMID: 33081974.

[11] Rath S, Tripathy A, Tripathy AR. "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model," Diabetes Metab Syndr. 2020 Sep-Oct;14(5):1467-1474. doi: 10.1016/j.dsx.2020.07.045. Epub 2020 Aug 1. PMID: 32771920; PMCID: PMC7395225.

[12] Chatterjee, A.; Gerdes, M.W.; Martinez, S.G. "Statistical explorations and Univariate Timeseries Analysis on COVID-19 datasets to understand the trend of disease spreading and death," Sensors 2020, 20, 3089. https://doi.org/10.3390/s20113089

[13] Devakumar K. P. "COVID-19 - analysis, visualization & comparisons, " Kaggle COVID-19 dataset, available from: https://www.kaggle.com/imdevskp/covid-19-analysis-visualization-comparisons/data, 2020

[14] V. Tulshyan, D. Sharma, and M. Mittal. "An eye on the future of COVID-19: prediction of likely positive cases and fatality in India over a 30-day horizon using the Prophet model," Disaster Medicine and Public Health Preparedness, pp. 1–7, 2020.

[15] P. Mishra, A. Mohammad, G. Al Khatib et al., "Modelling and forecasting of COVID-19 in India," Journal of Infectious Diseases and Epidemiology, vol. 6, no. 5, 2020.

[16] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the ARIMA model on the COVID-2019 epidemic dataset," Data in Brief, vol. 29, p. 105340, 2020.

[17] M. Indhuja and P. P. Sindhuja, "Prediction of covid-19 cases in India using prophet," International Journal of Statistics and Applied Mathematics, vol. 5, no. 4, pp. 103–106, 2020.

[18] Duccio Fanelli and Francesco Piazza, "Analysis and forecast of Covid-19 spreading in China, Italy and France," Chaos, Solitons & Fractals, 134:109761, 2020.

[19] Milad Haghani, Michiel CJ Bliemer, Floris Goerlandt, and Jie Li, "The scientific literature on coronaviruses, covid-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review," Safety Science, page 104806, 2020

[20] Github Inc. Covid-19 cases. https://github.com/cssegisanddata/covid-19 (accessed in 21 may, 2021).

[21] Lin Jia, Kewen Li, Yu Jiang, Xin Guo, et al, "Prediction and analysis of coronavirus disease 2019." arXiv preprint arXiv:2003.05447, 2020.

[22] Parikshit N Mahalle, Nilesh P Sable, Namita P Mahalle, and Gitanjali R Shinde, "Data analytics: Covid-19 prediction using multimodal data," Preprints, 2020

[23] Manotosh Mandal, Soovoojeet Jana, Swapan Kumar Nandi, Anupam Khatua, Sayani Adak, and TK Kar, "A model based study on the dynamics of covid-19: Prediction and control," Chaos, Solitons & Fractals, page 109889, (accessed in 21 may, 2021).

[24] [Prophet: forecasting at scale - Facebook Research. https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale, (accessed in 21 may, 2021).

[25] 9] Al-Rousan, N., Al-Najjar, H, "Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases," Journal of Medical Virology, 92(9): 1603-1608. https://doi.org/10.1002/jmv.25850, 2020

[26] Gambhir, E., Jain, R., Gupta, A. & Tomer, U, "Regression analysis of COVID-19 using machine learning algorithms, " Proceedings of the International Conference on Smart Electronics and Communication, pp. 65–71. https://doi.org/10.1201/9781351073974, 2020

[27] Shinde G.R., Kalamkar A.B., Mahalle P.N., Dey N., Chaki J., Hassanien A.E, "Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art", SN Comp Sci. 2020;1:197. doi: 10.1007/s42979-020-00209-9.

[28] 8. Wang C., Horby P.W., Hayden F.G., Gao G.F, "A novel coronavirus outbreak of global health concern," Lancet North Am Ed. 2020;395(10223):470–473. 10.1016%2FS0140-6736(20)30185-9.

[29] McKinney W, "Data structures for statistical computing in python,", Pp. 51–56, Proceedings of the 9th Python in Science Conference. Austin, TX. Vol. 445. 2010.

[30] Oliphant T.E, "A guide to NumPy, " Vol. 1. Trelgol Publishing USA, 2006.