

Prediction of Covid-19 Cases in Central Java using the Autoregressive (AR) Method

Tangguh Widodo¹, Siti Maghfiroh², Surya Haganta Brema Ginting³, Alif Aryaputra⁴, Sudioanto Sudioanto^{5*}

^{1, 2, 3, 4, 5*} Department of Informatics, Faculty of Informatics, Institut Teknologi Telkom Purwokerto

¹20102186@ittelkom-pwt.ac.id, ²20102176@ittelkom-pwt.ac.id, ³20102182@ittelkom-pwt.ac.id, ⁴20102008@ittelkom-pwt.ac.id, ^{5*}sudioanto@ittelkom-pwt.ac.id

Abstract

Since the beginning of the Covid-19 case in Indonesia in March 2020, more than 6 million confirmed cases had been confirmed. The rapid development of this case can be accessed through the covid19.go.id page. In Central Java province, confirmed cases as of July 6, 2022, reached 628,393 people, with the number of recovered patients reaching 594,783 people and the number of patients dying as many as 33,215 people. With this data, a prediction is needed to help the government anticipate an increase in Covid-19 cases in Central Java Province. This study aims to create a forecasting model using the Autoregressive (AR) method by optimizing the function parameters. Then Mean Squared Error (MSE) to analyze the results of forecasting data errors. The results are the best parameter functions on AR (30) with the smallest MSE. Furthermore, predictions are made from July 1 to August 30, 2022, showing an increase in cases.

Keywords: Covid-19, Autoregressive, Forecasting, Mean Squared Error.

© 2023 Journal of DINDA

1. Pendahuluan

At the end of 2019, a new virus called the Covid-19 virus (*Corona Virus Disease-2019*) emerged. The virus was first confirmed to be in Wuhan, China and spreads daily to other countries. Until November 2020, there were more than 200 countries that reported this Covid-19 case to the *World Health Organization (WHO)*, including Indonesia. Thus, Covid-19 was designated as a pandemic by the World Health Organization (WHO) on March 11, 2020.

The first Covid-19 case that appeared in Indonesia was in Depok, West Java and was announced on March 2, 2020. In the aftermath of these cases, the addition of Covid-19 cases is increasingly unpredictable every day. As of July 6, 2022, Indonesia confirmed 6,100,671 positive cases, with a total number of recovered patients of 5,925,853 people, and the number of patients dying as many as 156,770 people. Central Java Province is one of the provinces with the highest number of positive cases of Covid-19 in Indonesia. In Central Java province, confirmed positive cases to date have reached 628,393 people, the number of recovered patients has reached 594,783 people, and the number of patients has died as many as 33,215 people [7].

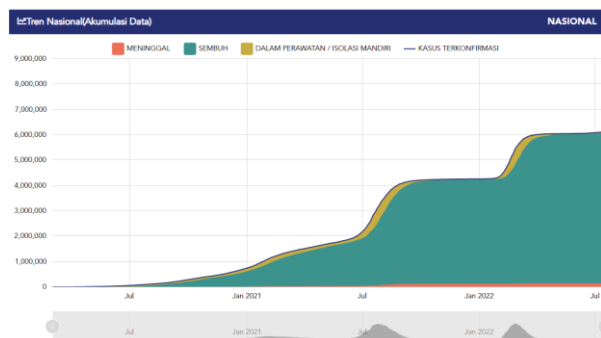


Figure 1. Graph of Covid-19 Cases in Indonesia in 2020-2022

Figure 1 is a graph of Covid-19 cases in Indonesia until July 6, 2022. Confirmed cases have stabilized in the period from March to July 2022. Then there were 156,770 deaths. Then there were 5,925,853 recovered cases.

Based on this data, a prediction is needed to find out the number of additional Covid-19 cases in the future. Where the results of these predictions can be used as an illustration for consideration in making a policy by the government in handling Covid-19 cases. According to

Herdianto (2013), prediction is a systematic process of estimating what may happen in the future and can minimize errors based on the data that the author currently has [11]. Predictions can help in analyzing data patterns in the past so that they can find solutions to solve problems in the future through prediction. Data prediction can be done using historical data (time series u) and *time series* methods [4].

The result of a prediction is not required to give a definite answer but is sought to be able to give the closest answer to future events. By minimizing prediction errors, the closest answer can be found. Prediction error can be calculated by the *Mean Squared Error* (MSE) method [4]. In this study, the authors used the *Autoregressive* method to complete their research. The author aims to predict an increase in Covid-19 cases in July to August 2022.

In a previous study entitled, "Forecasting the Increase of Covid-19 Patients Using *Support Vector Regression*," [3]. The study was conducted to help anticipate a surge in Covid-19 cases in East Java using the *Support Vector Regression* method. However, in the study, there was a lack of visualization in the form of graphs for forecasting. Based on the study, the authors used the *Autoregressive* method because this method is suitable for use in short-term forecasting.

Based on the reasons and research, there is a question that is used as a formulation of the problem in this study, described in the following: (a) How to obtain the optimum parameters for the percentage of errors when making predictions using the *Mean Squared Error*? (b) How to predict Covid-19 cases using the *Autoregressive* method for July to August 2022

2. Research Method

This research method was carried out with several stages to predict the increase in Covid-19 cases shown in Figure 2.

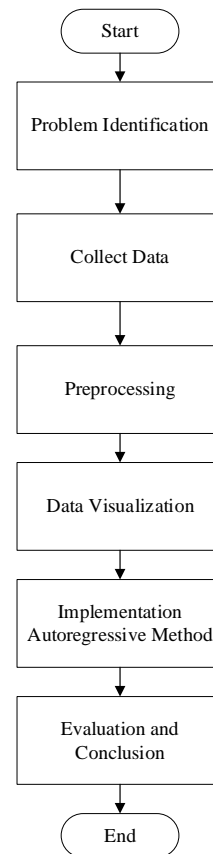


Figure 2. Flowchart Research

2.1. Problem Identification and Formulation

The first step to conduct this research is to identify and formulate problems, namely predicting the increase in Covid-19 cases in Central Java from July to August 2022 using the *Autoregressive* method and the data used is confirmed case data.

2.2. Data Collection

Researchers collected data on Covid-19 cases on the website <https://data.humdata.org/dataset/Indonesia-covid-19-cases-recoveries-and-deaths-per-province> [8] and <https://www.covid19.go.id/> [9]. Researchers choose data from the website because the data is real data and has been listed for every date and month in 2020 to 2021. The data contains data on confirmed, active, recovered, and deceased cases, as well as data for the validation table for the months of October to November 2021.

csv_for_prediction_cases_200516	7/4/2022 7:57 AM	Microsoft Excel Co...	10 KB
csv_for_validate_cases_200516	7/4/2022 8:39 AM	Microsoft Excel Co...	2 KB
csv_tabel_aktif_200516	7/4/2022 1:22 AM	Microsoft Excel Co...	89 KB
csv_tabel_konfirmasi_200516	7/4/2022 1:22 AM	Microsoft Excel Co...	104 KB
csv_tabel_meninggal_200516	7/4/2022 1:22 AM	Microsoft Excel Co...	76 KB
csv_tabel_sembuh_200516	7/4/2022 1:23 AM	Microsoft Excel Co...	100 KB

Figure 3. Data for Visualization and Prediction

2.3. Preprocessing

The *preprocessing* stage in this study was used by converting the date column into a datetime data type, this was used to facilitate the process of predicting the increase in Covid-19 cases.

```
<Day>
DatetimeIndex(['2020-03-19', '2020-03-20', '2020-03-21', '2020-03-22',
                '2020-03-23', '2020-03-24', '2020-03-25', '2020-03-26',
                '2020-03-27', '2020-03-28',
                ...
                '2021-09-21', '2021-09-22', '2021-09-23', '2021-09-24',
                '2021-09-25', '2021-09-26', '2021-09-27', '2021-09-28',
                '2021-09-29', '2021-09-30'],
              dtype='datetime64[ns]', length=561, freq='D')
```

Figure 4. The result of converting a date column to a Datetime data type

2.4. Data Visualization

Data Visualization is used to display data on confirmed, active, recovered, and deceased cases in the period from March 2020 to September 2021.

2.5. Application of Autoregressive Methods

The flow from the implementation of the *Autoregressive Method* will be shown in Figure 5. This flow starts by inputting data, then making predictions, then by dividing the data into train data and test data. Next, an AR Test is held and uses the *Autoregressive* method to get a prediction based on the smallest error value. Then compare the prediction data with the validation data. After everything is done, the author gets the results for predictions in July to August 2022.

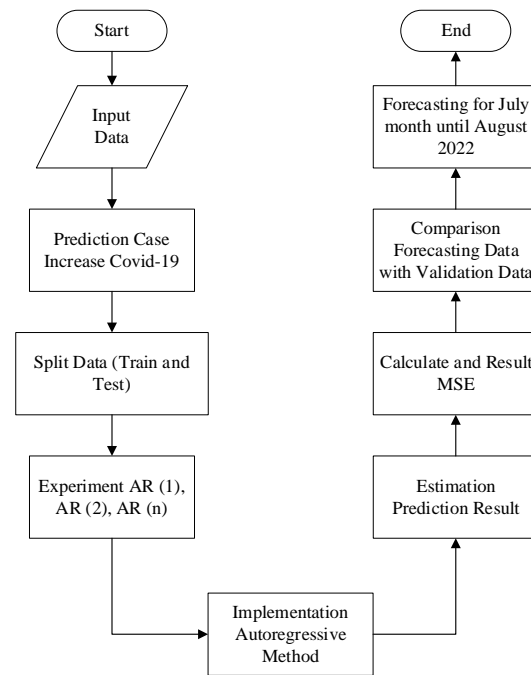


Figure 5. Application of Autoregressive Methods

2.6. Evaluation and Conclusion

The final stage is to conduct evaluations and conclusions. After an analysis for prediction, conclusions can be made from all series of research.

3. Result and Discussion

The data used by the author is secondary data which is Covid-19 case data including data on confirmed patients, active cases, recoveries, and deaths. The case data that the author uses is Covid-19 case data from March 2020 to September 2021. After that, the author visualizes the data, then predicts covid-19 cases from October to November 2021. Then, at the end the author will make predictions for the covid case which will occur in July to August 2022.

3.1. Data Collection

The author took data on Covid-19 cases on the data.humdata.org website. The author chose data from the website because the data is real data and has been listed for every date and month in 2020 to 2021.

3.2. Covid-19 Case Data Visualization

The first step to create a Covid-19 data visualization is to determine the library to be used. Next is to create a data plot to display all cases such as confirmed, active, dead, and recovered cases. For confirmed cases shown in Figure 1.

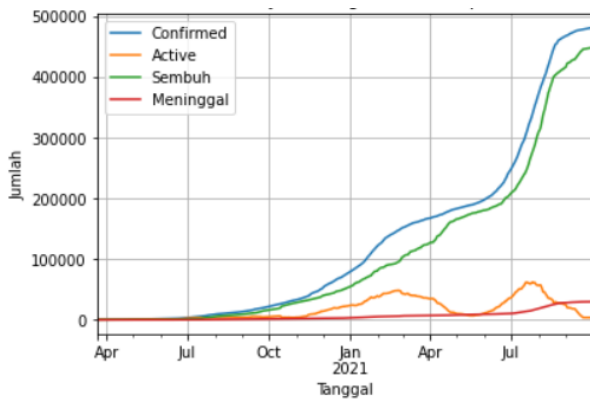


Figure 6. Visualization of Covid-19 Cases in Central Java

3.3. Forecasting Covid-19 Cases in Central Java

The first step to make a Prediction of Covid-19 Cases is to call the data to be used, namely `csv_for_prediction_cases_200516.csv`. The data used for prediction is 561 data, then divided the data into train and test data. Data train is used to train the model, the data used is 520 data. The test data was used to test the model, the data used was 41 data.

First, the author tested the first model called AR (1) Predictions with the maxlag being 1. AR (1) Predictions will be shown in Figure 2.

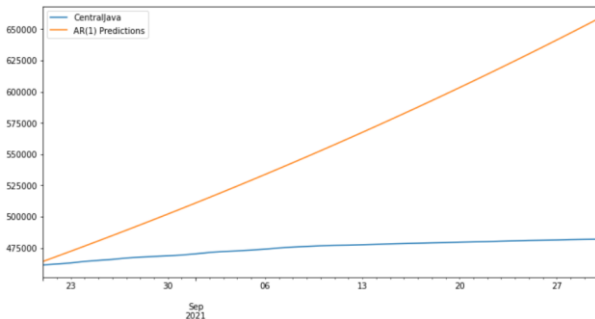


Figure 7. AR (1) Predictions

Second, the author tested a second model called AR (2) Predictions with the maxlag being 2. AR (2) Predictions will be shown in Figure 3.

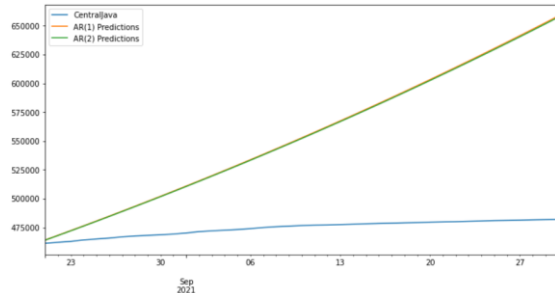


Figure 8. AR (2) Predictions

Next, the author tested the AR(n) model with a maxlag of 30. AR (30) Predictions will be shown in Figure 4.

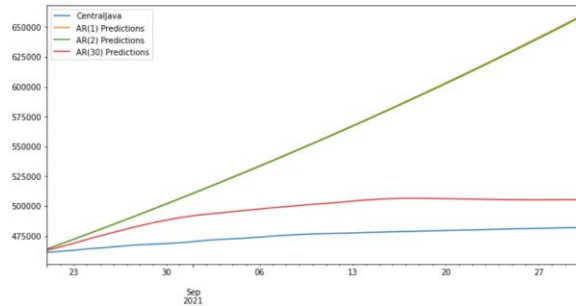


Figure 9. AR (30) Predictions

Then using the *Mean Squared Error*, you can find out the *error* of each trial which will be displayed in Table 1.

Table 1 *Error* calculation using MSE

<i>Error</i>	Count
AR (1)	9.270063015e+09
AR (2)	9.150384591e+09
AR (30)	504093762.2

Using AR (30) which has the smallest *error* value , a forecasting for October to November 2021 is obtained as shown by Figure 5.

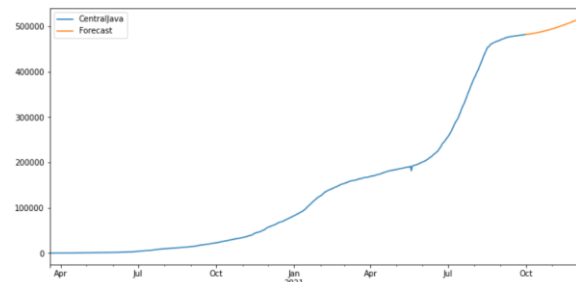


Figure 10. Forecasting Covid-19 Cases from October to November 2021

The number of *forecasting* data is 61 data, then it will be compared with validation data.

3.4. Comparison of Forecasting Data with Validation Data

The initial step for data validation is to call the csv_for_validate_cases_200516.csv data. This data is recapitulated by retrieving it from covid19.go.id website. The results of comparing forecasting data with validation data will be shown in Figure 6.

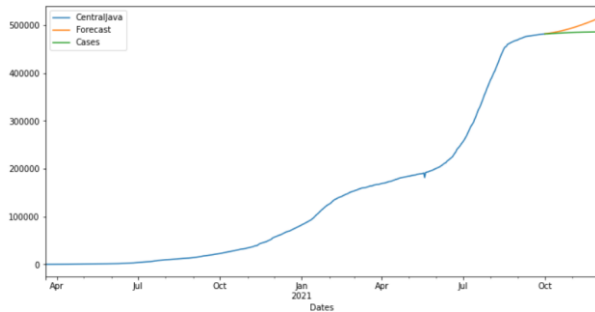


Figure 11. Comparison Chart of Forecasting data with Validation data

A comparison of the contents of the data is shown in Tabel 2.

Table 2 Comparison of Forecasting data with Validation data

Date	Predictions	Validation
1-Oct-21	482306.57	482116
2-Oct-21	482554.084	482221
3-Oct-21	482808.987	482267
4-Oct-21	483069.695	482353
5-Oct-21	483348.555	482444
6-Oct-21	483632.591	482610
7-Oct-21	483922.012	482769
...
28-Nov-21	512947.904	486330
29-Nov-21	513709.311	486385
30-Nov-21	514475.727	486385

3.5. Forecasting Covid-19 Cases from July to August 2022

The prediction of an increase in Covid-19 cases in July to August 2022 was obtained after validating the data at point number 3.4. The following is a display of the prediction results in the form of a graph shown in Figure 12.

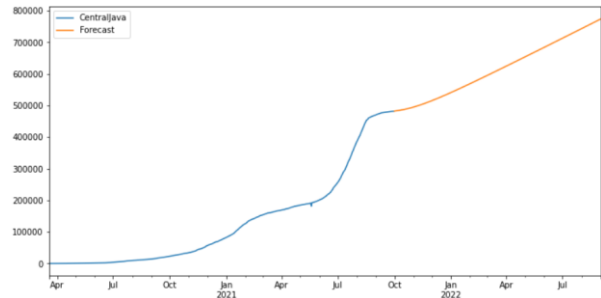


Figure 12. Graph of Forecasting Covid-19 Cases from July to August 2022

The following is the content of the forecasting data that will be displayed in Table 3.

Table 3 Forecasting Covid-19 Cases from July to August 2022

Date	Predictions
1-Jul-22	712681.3277
2-Jul-22	713656.6023
3-Jul-22	714632.0224
4-Jul-22	715607.5877
5-Jul-22	716583.2976
6-Jul-22	717559.1519
...	...
29-Aug-22	770462.3166
30-Aug-22	771445.7476
31-Aug-22	772429.3126

4. Conclusion

Based on the results of data processing and research analysis that has been carried out, it can be concluded that the results of the prediction of the number of Covid-19 patients in Central Java Province in July to August 2022 will increase from before. The Central Java Provincial Government must continue to carry out the 3M movement to prevent the transmission of Covid-19 in Central Java Province. Based on the results of the forecasting, it is estimated that on August 31, 2022, the increase in Covid-19 cases will reach 772429 people. In addition, the best function parameter in AR(30) with the least error.

Acknowledgments

The author would like to thank all parties who have indirectly assisted in the completion of this research.

References

- [1] Widiarni, A., dan Mustakim, 2021. Penerapan Algoritma Support Vector Regression untuk Prediksi

- Jumlah Pasien Covid-19 di Provinsi Riau. *Building of Informatics, Technology, and Science (BITS)*, vol. 3, no. 2, pp. 71-78.
- [2] Supriatun, dkk., 2020. Peramalan Jumlah Kasus Covid-19 di Semarang Menggunakan Metode Autoregressive Integrated Moving Average. *Prosiding Seminar Edusaintech FMIPA UNIMUS*. 2020.
- [3] Lestari, M.P., Witarsyah, D.J., dan Hamami, F., "Peramalan Pertambahan Pasien Covid-19 Menggunakan Support Vector Regression," *E-Proceeding of Engineering.*, vol. 8, no. 1, pp. 9497–9507, 2021.
- [4] Ainiyah, L., dan Bansori, M., "Prediksi Jumlah Kasus Covid-19 Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) (Studi Kasus Kabupaten Sidoarjo)," *Jurnal Sains Dasar*, vol. 10, no. 2, 2021, pp. 159–165.
- [5] Kurniasih, D., Mariani, S., dan Sugiman., "Efisiensi Relatif Estimator Fungsi Kernel Gaussian Terhadap Estimator Polinomial Dalam Peramalan USD Terhadap JPY," *UNNES Journal of Mathematics*, 2013, vol. 2, no. 2, doi: 10.1063/1.5042998.
- [6] Yuriska, R., Rohmawati, A.T., dan Aditsania, A., 2021. Forecasting Jumlah Kasus Harian Covid-19 di Provinsi Jawa Barat Menggunakan Model Vector Autoregressive (VAR). *e-Proceeding of Engineering*. 2021, vol. 8, no. 5, pp. 11376-11387.
- [7] Tanggap Covid-19 Jawa Tengah, 2020. <https://corona.jatengprov.go.id/data>. (accessed Jul. 6, 2022).
- [8] HDX – Javier Terans, 2021. <https://data.humdata.org/dataset/indonesia-covid-19-cases-recoveries-and-deaths-per-province>. (accessed Jul. 9, 2022).
- [9] Kementerian Republik Indonesia, 2020. <https://covid19.go.id/>. (accessed Jul. 7, 2022).
- [10] Tadlaru, R.M., "Analisis Infografis Dampak Pandemi Virus Corona (COVID-19) Terhadap Perekonomian Indonesia Pada Portal Berita Katadata.co.id," Universitas Pembangunan Nasional "Veteran" Yogyakarta, 2021.
- [11] Rohmawati, F., Rohman, G., dan Mujilawati, S., 2017. Sistem Prediksi Jumlah Pengunjung Wisata Wego Kec.Sugio Kab.Lamongan Menggunakan Metode Fuzzy Time Series. *JOUTICLA*, 3(2).