

# Prediction of the COVID-19 Vaccination Target Achievement with Exponential Regression

Teja Endra Eng Tju<sup>1\*)</sup>, Dian Sa'adillah Maylawati<sup>2</sup>, Ghifari Munawar<sup>3</sup>, Suharjanto Utomo<sup>4</sup>

<sup>1</sup>Universitas Budi Luhur

<sup>2</sup>UIN Sunan Gunung Djati Bandung

<sup>3</sup>Politeknik Negeri Bandung

<sup>4</sup>Universitas Nurtanio

\*)Email: teja.endraengtju@budiluhur.ac.id

**Abstract** – The achievement of the national COVID-19 vaccination target in Indonesia is often reported to be uncertain with various existing obstacles. Prediction with exponential regression modeling is done by adopting part of the SKKNI Data Science with the stages of Data Understanding, Data Preparation, Modeling, Model Evaluation. The vaccination dataset from the Ministry of Health of the Republic of Indonesia for the period from January 13, 2021 to October 10, 2021, was randomly separated into training data of 0.8 parts and testing data of 0.2 parts. The optimal parameters of the exponential function are found using the `scipy.optimize` library in IPython. The model obtained was evaluated using MAE, RMSE, and R-Squared metrics on normalized training data, training data, test data, and recent data for seven days from 11 to 17 October 2021. The prediction results show that the vaccination target will be achieved 100 percent on January 18, 2022, while on December 31, 2021, only 80 percent will be achieved. From the recent data, it appears that more acceleration is needed, especially if it is desired to be achieved in December 2021 as determined by President Joko Widodo, there will be a shortfall of 20 percent based on the prediction results.

**Keywords** – Prediction, Exponential, Regression, Vaccination, COVID-19, SKKNI Data Science.

## I. INTRODUCTION

The achievement of the target of the Indonesian population having been fully vaccinated of COVID-19 is not certain when it will occur. Some of these obstacles include resistance from the population, areas that are difficult to reach, inefficient vaccination implementation, distribution and availability of vaccines that are not smooth. The emergence of new virus variants makes it difficult to achieve the target of herd immunity due to vaccine efficacy constraints, so now the government has changed the vaccination target to a minimum of 208,265,720 Indonesians or about 80 percent of the total population so that the outbreak can be controlled [1]. Initially the target time was set for March 2022, then President Joko Widodo wanted an acceleration to December 2021 on the grounds that the economy could run. However, some officials and figures often mention different things according to the conditions at that time [2].

Previous research has mostly been done on the spread or increase of COVID-19 sufferers, not on vaccination. Prediction of the epidemic in Egypt is done by various regression analysis [3]. The spreading trend in China is used as an exponential attractor [4]. Modeling with reverse exponential regression for daily cases in Saudi Arabia [5]. In India, modeling and forecasting growth curves using various analytical techniques [6] as well as piecewise regression techniques [7]. For the prediction of COVID-19 cases in Indonesia using the hybrid method nonlinear regression logistic – double exponential smoothing [8], single exponential smoothing and the Holt's method [9], exponential smoothing method [10].

The Data Science approach that utilizes AI (Artificial Intelligence) technology can be a solution to present insight knowledge from data or facts related to vaccination rates in

Indonesia so that it can be used as a basis or recommendation in decision making for policy makers. AI technology can help predict when Indonesia will be able to achieve herd immunity. This prediction can also lead to recommendations for further activities of the Indonesian people, such as policies related to health protocols, community activities in public places, activities in the work and school environment, and other policies related to recovering post-pandemic conditions or living side by side with Covid-19. These predictions can be processed with AI techniques in Data Science based on data on vaccination rates that have been carried out in various situations and conditions, so that from time to time it can be scientifically determined when the target number of vaccinations can be achieved. Likewise, if the target amount changes, the target time will be more easily analyzed. Therefore, this study aims to predict the achievement of the Covid-19 vaccination target in Indonesia.

## II. RESEARCH METHODOLOGY

The dataset was obtained from the website of the Ministry of Health of the Republic of Indonesia which specifically provides vaccination report [12]. The data used for modeling is the second vaccination dose from January 13, 2021 to October 10, 2021 or as many as 271 data which is the result of grouping from various regions or population demographics in Indonesia.

The research methodology refers to the Indonesian National Work Competency Standard (SKKNI) No. 299 of 2020 in the field of Artificial Intelligence, sub-field of Data science [11]. The SKKNI Data Science consists of seven main activities, namely business understanding, data understanding, data preparation, modeling, model evaluation, deployment, and evaluation. This paper adopts



four activities relevant to the research conducted, namely data understanding, data preparation, modeling, and model evaluation.

**A. Data Understanding**

The data is a time series with date as the independent variable and the second vaccination achievement as the dependent variable.

**B. Data Preparation**

For regression purposes, the date needs to be changed to an index starting from day 0 (zero) to day 270, sequentially according to the daily data date.

The dependent variable is the percentage of the daily cumulative amount, such that when it reaches 100 percent, it means that 208,265,720 or 80 percent of the total population of Indonesia have been vaccinated.

The dataset with 271 data was divided into two parts randomly, 0.8 part as the training dataset with 216 data and 0.2 part as the testing dataset with 55 data. The distribution of data and the results of data compilation are visualized [13] in Figure 1.

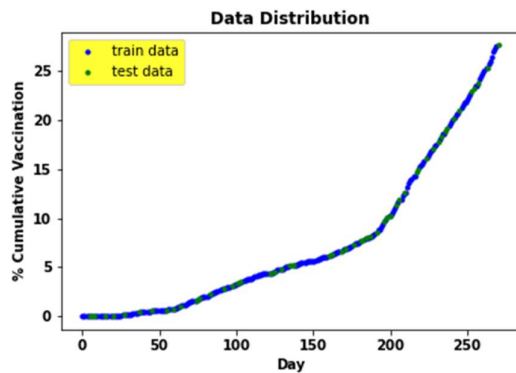


Figure 1. Data Distribution

**C. Modeling**

From the characteristics of the data with increasing growth from time to time, also driven by the government's desire to accelerate vaccination, the suitable modeling approach is Exponential Regression.

$$f(x) = ab^x + c \tag{1}$$

The Exponential Function [14] used is shown in Equation (1) where  $x$  is the independent variable,  $f(x)$  is the dependent variable, and there are three parameters  $a$ ,  $b$ ,  $c$  which the optimal values are sought with curve\_fit from the scipy.optimize library [15] in IPython [16].

Data normalization was carried out [17] to obtain the appropriate model, after which the results were returned to the original values. The model will be built from normalized training dataset.

**D. Model Evaluation**

There are three metrics used for model evaluation, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-Squared [18]. Evaluation is carried out on training dataset, testing dataset, and the recent dataset.

Recent dataset is new data from October 11, 2021 to October 17, 2021.

Furthermore, it is necessary to find out when the 100 percent vaccination target is achieved by iterating to get the value of the independent variable.

**III. RESULTS AND DISCUSSION**

The results of the model development with training data normalized to the exponential function Equation (1) resulted in the optimal fit parameter values  $a = 0.037263815344230705$ ,  $b = 28.481521431056155$ , and  $c = -0.03569552348715502$ , the visualization results are in Figure 2. The result after returning to the original value is shown in Figure 3.

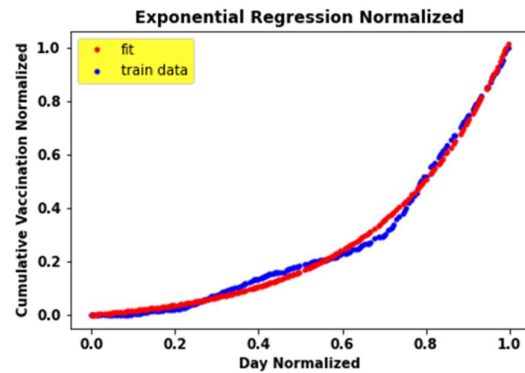


Figure 2. Exponential Regression Normalized

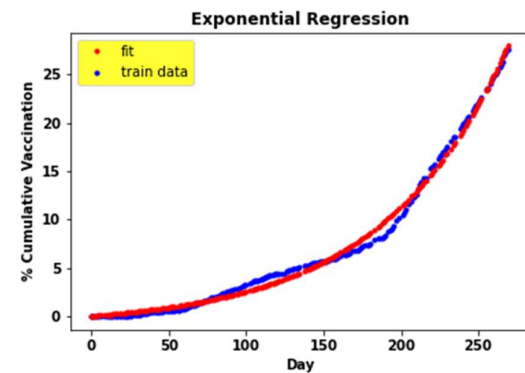


Figure 3. Exponential Regression

The results of the model evaluation on the testing dataset are visualized in Figure 4. Furthermore, the results on the new dataset for seven days from October 11, 2021 to October 17, 2021 can be seen in Figure 5.

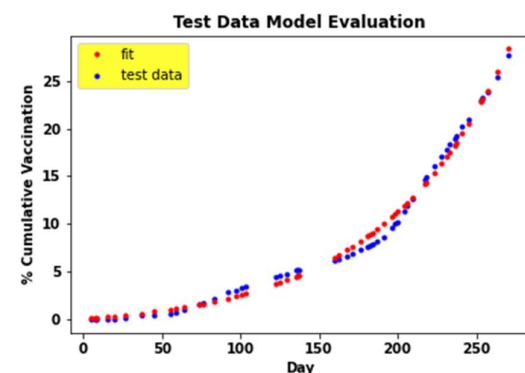


Figure 4. Test Data Model Evaluation

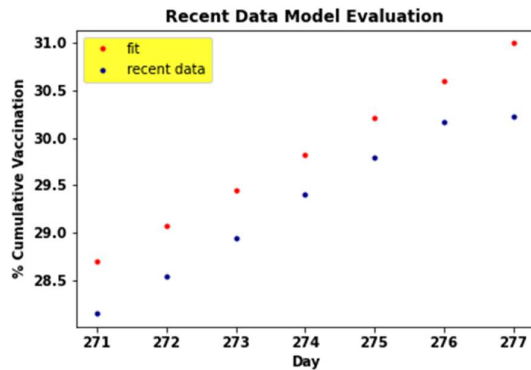


Figure 5. Recent Data Model Evaluation

Model Evaluation as measured by MAE, RMSE, R-Squared on training dataset, testing dataset, and the recent dataset is presented in Table 1. These results have been visualized in Figure 2, Figure 3, Figure 4, and Figure 5, respectively.

Table 1. Model Evaluations Result

Evaluation Metrics	Dataset			
	Normalized	Training	Testing	Recent
MAE	0.02	7.88	8.67	0.52
RMSE	0.02	10.62	10.97	0.53
R-Squared	0.99	0.99	0.99	0.51

To get the date when the vaccination target is reached 100 percent, an iteration has been carried out with the results on day 370 or on January 18, 2022. On the achievement of the recent data in Figure 5, it seems that it lags behind the fit data model (predicted results), so the implementation of daily vaccinations must be accelerated. Based on historical data, the acceleration of vaccination implementation should be maintained with reference to the predicted data, so it can be said as a prediction roadmap as shown in Figure 6. This prediction result is the optimal or moderate value of the government's target, which was originally March 2022 and then advanced to December 2021. On December 31, 2021 or day 352, 80 percent vaccination progress will be made, the Indonesian government needs to make efforts to accelerate in such a way that recent data can more often exceed the predicted results so that the December 2021 target can be achieved.

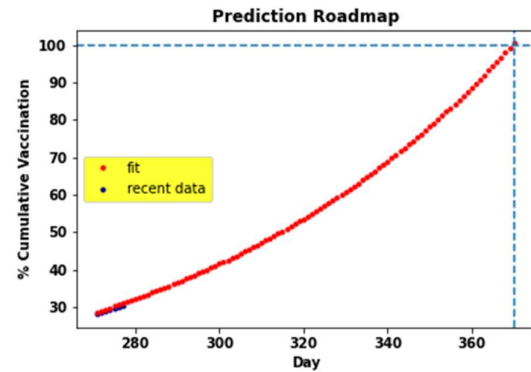


Figure 6. Prediction Roadmap

#### IV. CONCLUSION

Regression analytic approach based on historical data can be used to build a model for prediction. The characteristics of the data become an important asset, especially in terms of data growth to determine a suitable function as a model.

With exponential regression modeling able to predict the target of achieving COVID-19 vaccination in Indonesia. Functional regression can be used on continuous time series data in a variety of other problems to predict future results.

Based on predictions using the exponential regression model, the vaccination target will be reached 100 percent on January 18, 2022, while the government target for December 2021 has only reached 80 percent. This has an impact that the government needs to make various acceleration efforts so that herd immunity can occur in December 2021.

#### REFERENCES

- [1] Satuan Tugas Penanganan COVID-19, "208.265.720 Orang Target Sasaran Vaksinasi COVID-19 di Indonesia," 2021. <https://covid19.go.id/p/vaksin/208265720-orang-target-sasaran-vaksinasi-covid-19-di-indonesia>.
- [2] Kompas, "Sulit Kejar Target Vaksinasi Covid-19, Menkes: Kami Butuh Bantuan Pemda dan Swasta," 2021. <https://megapolitan.kompas.com/read/2021/06/02/13554891/sulit-kejar-target-vaksinasi-covid-19-menkes-kami-butuh-bantuan-pemda-dan-swasta>.
- [3] L. A. Amar, A. A. Taha, and M. Y. Mohamed, "Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt," *Infect. Dis. Model.*, vol. 5, 2020, doi: 10.1016/j.idm.2020.08.008.
- [4] C. C. Zhu and J. Zhu, "Spread trend of COVID-19



- epidemic outbreak in China: Using exponential attractor method in a spatial heterogeneous SEIQR model,” *Math. Biosci. Eng.*, vol. 17, no. 4, 2020, doi: 10.3934/MBE.2020174.
- [5] S. R. Al-Dawsari and K. S. Sultan, “Modeling of daily confirmed Saudi COVID-19 cases using inverted exponential regression,” *Math. Biosci. Eng.*, vol. 18, no. 3, 2021, doi: 10.3934/MBE.2021117.
- [6] V. K. Sharma and U. Nigam, “Modeling and Forecasting of COVID-19 Growth Curve in India,” *Trans. Indian Natl. Acad. Eng.*, vol. 5, no. 4, 2020, doi: 10.1007/s41403-020-00165-z.
- [7] A. Senapati, A. Nag, A. Mondal, and S. Maji, “A novel framework for COVID-19 case prediction through piecewise regression in India,” *Int. J. Inf. Technol.*, vol. 13, no. 1, 2021, doi: 10.1007/s41870-020-00552-3.
- [8] I. G. B. Ngurah Diksa, “Peramalan Gelombang Covid 19 Menggunakan Hybrid Nonlinear Regression Logistic – Double Exponential Smoothing di Indonesia dan Prancis,” *Jambura J. Math.*, vol. 3, no. 1, 2021, doi: 10.34312/jjom.v3i1.7771.
- [9] N. H. A. S. Al Ihsan, H. H. Dzakiyah, and F. Liantoni, “Perbandingan Metode Single Exponential Smoothing dan Metode Holt untuk Prediksi Kasus COVID-19 di Indonesia,” *Ultim. J. Tek. Inform.*, vol. 12, no. 2, 2020, doi: 10.31937/ti.v12i2.1689.
- [10] C. M. Gibran, S. Setiyawati, and F. Liantoni, “Prediksi Penambahan Kasus Covid-19 di Indonesia Melalui Pendekatan Time Series Menggunakan Metode Exponential Smoothing,” *J. Inform. Univ. Pamulang*, vol. 6, no. 1, 2021, doi: 10.32493/informatika.v6i1.9442.
- [11] Kementerian Ketenagakerjaan Republik Indonesia, “SKKNI Keahlian Artificial Intelligence (Data Science),” 2020. <https://skkni.kemnaker.go.id/tentang-skkni/dokumen?area=data>
- [12] Kementerian Kesehatan Republik Indonesia, “Vaksinasi COVID-19 Nasional,” 2021. <https://vaksin.kemkes.go.id/#/vaccines>.
- [13] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.
- [14] R. Torres, “Mathematical Investigation of Functions,” *J. Educ. Manag. Dev. Stud.*, vol. 1, no. 1, 2021, doi: 10.52631/jemds.v1i1.6.
- [15] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat. Methods*, vol. 17, no. 3, pp. 261–272, 2020, doi: 10.1038/s41592-019-0686-2.
- [16] P. Fernando and E. G. Brian, “IPython: A System for Interactive Scientific Computing,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 21–29, 2007.
- [17] S. Lakshmanan, “How, When, and Why Should You Normalize / Standardize / Rescale Your Data?,” *Towards AI — The Best of Tech, Science, and Engineering*, 2019. <https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff>.
- [18] J. Hale, “Which Evaluation Metric Should You Use in Machine Learning Regression Problems?,” *Towards Datascience*, 2020. <https://towardsdatascience.com/which-evaluation-metric-should-you-use-in-machine-learning-regression-problems-20cdaef258e>.

