Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:63
https://doi.org/10.28919/cmbn/8579
ISSN: 2052-2541

# PREDICTIVE ANALYSIS USING GAUSSIAN PROCESSES REGRESSION AND MAHALANOBIS DISTANCE APPROACH: ANTICIPATION OF COVID-19 SPIKE IN BANDUNG CITY

SRI WINARNI[1,*], SAPTO WAHYU INDRATNO[2], HERLINA ROSELINE[3], RESTU ARISANTI[1],

RESA SEPTIANI PONTOH[1]

[1]Department of Statistics, Faculty of Mathematics and Natural Science, Padjadjaran University, Indonesia

[2]Statistics Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia

[3]The agency for development, planning, research and development of Bandung City Government, Indonesia



**Abstract.** This paper presents research on the use of Gaussian process regression integrated with Mahalanobis distance to model the number of active COVID-19 cases, healed rates, and deaths from COVID-19 in Bandung City, Indonesia. Gaussian process regression as a machine learning method is used to predict the number of COVID-19 cases, while Mahalanobis distance is used to determine outliers, which are values that exceed a safe threshold as an early warning system. This analysis is important to anticipate possible spikes in cases that could occur in the future. Analysis was also carried out using a Pareto diagram to find out which region experienced the biggest spike. In addition, this study also considers the availability of beds in hospitals as health facilities that support the handling of COVID-19. The analysis results show that both analysis methods can model the COVID-19 case well. These findings can serve as a basis for policy-making related to COVID-19 handling.

**Keywords**: COVID-19; Gaussian process regression; Mahalanobis distance; beds monitoring; early warning system.

**2020 AMS Subject Classification:** 92C60.

---

[*]Corresponding author
E-mail address: sri.winarni@unpad.ac.id
Received April 05, 2024

## 1. INTRODUCTION

COVID-19 first entered Indonesia in early March 2020, when two Indonesian citizens tested positive after traveling from abroad. This started the widespread spread of the virus across the country, prompting the government to take strict prevention and containment measures. Nonetheless, great challenges remain in controlling the spread of this virus amidst social and economic dynamics [1]–[3]. The COVID-19 pandemic has had a huge impact in Indonesia, including in the city of Bandung. The wave of infections, which claimed thousands of lives and disrupted economic and social life, makes it a serious challenge. The unique characteristics of this virus, such as high transmission rates and variable symptoms, make modeling and prediction more complex [4]–[7]. Therefore, efforts to develop better methods to understand and forecast the spread of COVID-19 in Bandung are crucial to respond more effectively to the pandemic.

During the COVID-19 pandemic, Bandung experienced fluctuations in the number of active cases of the virus. While there was an increase in the number of recoveries, there was also an increase in the number of deaths from the virus. This emphasizes the importance of constant attention to public health and the need for ongoing prevention efforts to overcome the pandemic in Bandung. With 30 sub-districts in Bandung City, it is important to detect significant spikes in COVID-19 cases in each area. Analysis of active case data in each sub-district can help identify areas that need more serious attention. With this information, prioritization in resource allocation and response strategies can be done more effectively. With the surge of COVID-19 cases in various sub-districts of Bandung City, hospitals need additional facilities and resources. Bed monitoring in the treatment room is important to monitor the patient's condition continuously and provide timely treatment [8], [9].

Various statistical and computational mathematical models [10]–[16] have been developed to predict the spread of the virus [17], [18], understand epidemiological dynamics [19], and evaluate the effectiveness of various control strategies. These models use data on epidemiology [20], demographics [21][22], human mobility [23][24] and other factors [25]–[28] to evaluate the impact of public policies [29], such as movement restrictions [30], social distancing [31], early detection, and mass vaccination. While the COVID-19 pandemic appears to be winding down, it is important to continue modeling and studies to anticipate the possibility of similar events in the future. By understanding the patterns of virus spread and the factors that influence them, we can be better prepared for potential new waves or emerging virus variants. This is important for designing more effective policies and response strategies as well as preparing the health system for the challenges that may occur in the future.

In previous studies, prediction analysis was often done separately from surge detection [32][33] and surge region detection. This study does the work sequentially. This study uses Gaussian Process Regression (GPR) to predict the number of active cases, healeds, and deaths of COVID-19. The data on the number of active cases, healeds, and deaths were first transformed into return values to obtain continuous data. GPR is a statistical method that models the relationship between input and output variables by assuming a Gaussian distribution. It produces accurate predictions and accounts for uncertainty. GPR is often used for non-linear and complex data, including in COVID-19 prediction. With this approach, we hope to provide accurate projections based on historical data.

In analyzing predicted data, outlier detection or spikes in COVID-19 numbers can be done using the Mahalanobis distance [34]. The Mahalanobis distance measures how far a data point is from the center of the multivariate data distribution, taking into account the covariance between variables [35]. By looking at the Mahalanobis distance value of each data point, we can identify whether any data is significantly different from the general pattern of the data. Large spikes or outliers in COVID-19 numbers can be identified based on significant Mahalanobis distance values. Furthermore, the use of Pareto diagrams can help identify regions that have experienced significant spikes in COVID-19 cases. By analyzing the relative contribution of each region to the total cases, Pareto diagrams can provide a clear view of which regions need special attention in handling the surge of cases. Before entering the analysis, the data needs to be smoothed using a moving average (MA) to overcome data fluctuations.

## 2. METHOD

### 2.1. Moving average (MA) smoothing and return value transformation

MA smoothing helps create a smoother representation of trends or patterns in the data, making it easier to identify long-term trends while filtering out short-term fluctuations. The formulation of MA smoothing is given as follows:

$$X_t = \frac{1}{m}\sum_{i=0}^{m-1} X_{(o)t-i} \quad (1)$$

With $X_t$ as the Moving Average value at time $t$, then $X_{(o)t-i}$ is the original data value at time $t-i$, and $m$ is the Moving Average calculation period [36]. When the data has been smoothed, the next transformation is done in the form of return values.

The return value of COVID-19 is the value of growth or decline in COVID-19 cases from one period to the next. The use of this return value helps in measuring how fast COVID-19 cases increase or decrease over time. This transformation is done with the following equation:

$$R_t = \frac{X_t - X_{t-1}}{X_{t-1}+1} \quad (2)$$

With $R_t$ as the return value at time $t$, and $X_t$ is the smoothing value at time $t$ and $X_{t-1}$ is the smoothing value at time $(t-1)$. Adding the number one to the denominator is used to avoid division by zero.

This return value can be positive or negative. Positive values indicate growth, and negative values indicate decline. By monitoring and analyzing the return value of COVID-19 over time, health experts and researchers can gain insights into trends in case growth or decline, estimate the speed of virus spread, and plan response actions accordingly. In addition, these return values can also help in modeling and predicting the future spread of COVID-19.

## 2.2. Gaussian Process Regression (GPR)

GPR is an interesting regression method because it not only predicts the output but also the distribution of the output. The number of parameters in GPR depends on the amount of data used, making it more flexible in describing data patterns. GPR produces point estimates along with their distributions and is often used in the analysis of large complex data. GPR model building depends on the selection of the covariance function, with the squared exponential covariance function being a common choice. Hyperparameter optimization is performed using the maximum likelihood method.

The training data $\mathcal{D} = (\mathrm{X}, \mathbf{y})$ consisting of input matrix X and target vector $\mathbf{y}$ has an input vector $\mathrm{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n]^T$ where each vector $\mathbf{x}_i \in \mathbb{R}^D$ with $i = 1,2,\ldots,n$ corresponds to the corresponding target $y_i$. The target vector $\mathbf{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]^T$ with $n$ is the number of training data.

A Gaussian process is a collection of random variables, where a finite number of them have a common Gaussian distribution [37]. The Gaussian process is completely defined by its mean function $m(x)$ and its covariance function $k(x_i, x_j)$. The Gaussian process function is written as follows:

$$f(x) \sim GP\left(m(x), k(x_i, x_j)\right)$$

This means that $f(x)$ follows a Gaussian process with mean function $m(x)$ and covariance function $k(x_i, x_j)$. This means that for each input value x, the corresponding output value $f(x)$ forms a Gaussian process with the specified mean and covariance properties. The covariance function $k(x_i, x_j)$ in a Gaussian process is a metric that measures how close the relationship is between two data points $x_i$ and $x_j$ in the input space. Proper selection of the covariance function is important to accurately model the data in the Gaussian process. A commonly used covariance function is the squared exponential function, or Radial Basis Function (RBF) [38].

RBF measures the correlation between two points in the input space based on the euclidean distance between the two points. In a quadratic exponential covariance function, the correlation between two points will decrease exponentially as the distance between them increases. The mathematical equation of the squared exponential covariance function is as follows:

$$k(x_i, x_j) = \sigma_f^2 exp\left(-\frac{1}{2\ell^2}\left|x_i - x_j\right|^2\right) \quad (3)$$

with $\sigma_f^2$ and $\ell$ being the hyperparameters of signal variance and scale length. The signal variance controls how fluctuating the resulting stochastic process is, while the scale length controls how fast the correlation between the points decreases as the distance between them increases. Furthermore, $\left|x_i - x_j\right|$ is the Euclidian distance between two points in the input space. Estimation of the hyperparametric value is done by optimizing the following likelihood function:

$$\mathrm{Log}\left(p(\mathbf{y}|X)\right) = -\frac{1}{2}\mathbf{y}^T[K(X,X) + \sigma_n^2 I]^{-1}\mathbf{y} - \frac{1}{2}log|K(X,X) + \sigma_n^2 I| - \frac{n}{2}\log(2\pi) \quad (4)$$

where $K(X,X)$ is the covariance matrix of the training input data, and $\sigma_n^2$ is the noise variance that describes how much uncertainty is associated with the target.

Suppose the test data $\mathcal{D}_* = (X_*, \mathrm{y}_*)$ consists of the input value $X_*$ and the output value to be predicted $y_*$. The prediction of the output value $y_*$ on the test input value $X_*$ follows a Gaussian distribution with mean $\mu_*$ and variance $\sigma_*^2$ which is written as follows:

$$y_*|X, y, X_* \sim \mathcal{N}(\mu_*, \sigma_*^2) \tag{5}$$

Mean $\mu_*$ and variance $\sigma_*^2$ obtained as follows:

$$\mu_* = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \tag{6}$$

$$\sigma_*^2 = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \tag{7}$$

with $K(X_*, X)$ is the covariance matrix between test data and training data, while $K(X, X_*)$ is the transpose matrix of $K(X_*, X)$. Furthermore $K(X_*, X_*)$ is the covariance matrix of the test data [39].

### 2.3. Mahalanobis distance and Pareto diagram

Mahalanobis distance is a distance metric between data points in a multidimensional space that considers the covariance between variables. It is useful in pattern recognition, data clustering, and outlier detection in data analysis [40][41]. Formulation of Mahalanobis distance for a data point $x$ from the center of distribution $(\mu)$ in multidimensional space with covariance matrix $(\Sigma)$ [42]:

$$D(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \tag{8}$$

Since the covariance $\Sigma$ is a symmetric and positive semi-definite matrix, so is its inverse. Therefore, we can use the square root since all values are positive. However, before that, we have to assume that the data samples come from a multivariate normal distribution [43]. The squared value of the Mahalanobis distance follows the chi-square distribution with degrees of freedom $p$ (number of variables) as follows:

$$(x - \mu)^T \Sigma^{-1} (x - \mu) \sim \chi^2 \tag{9}$$

to identify outliers, we chose the threshold value for the Mahalanobis distance as the 0.975th quantile of $\chi_p^2$.

A Pareto diagram is a visual tool used to identify and prioritize the main causes of problems or imperfections in a system [44]. Data is presented in the form of a bar graph that presents the relative contribution of each factor to the overall total, thus allowing stakeholders to quickly see the most influential factors. This helps in making more informed and efficient decisions in quality improvement or control efforts

## 3. RESULTS AND DISCUSSION

In this study, the data used consists of COVID-19 data which includes the number of active cases, recoveries, and deaths. In addition, bad monitoring data is also used to enrich the analysis. The training data was obtained from the period of February 1, 2021, to July 21, 2022, while the testing data came from the period July 22-July 31, 2022. Predictions were made for the period August 1-10, 2022.

### 2.1. Results of MA smoothing and prediction using GPR

Before conducting a prediction analysis with Gaussian Process Regression (GPR), the data first undergoes a Moving Average (MA) smoothing process using Equation 1. The results of this smoothing process are as follows:

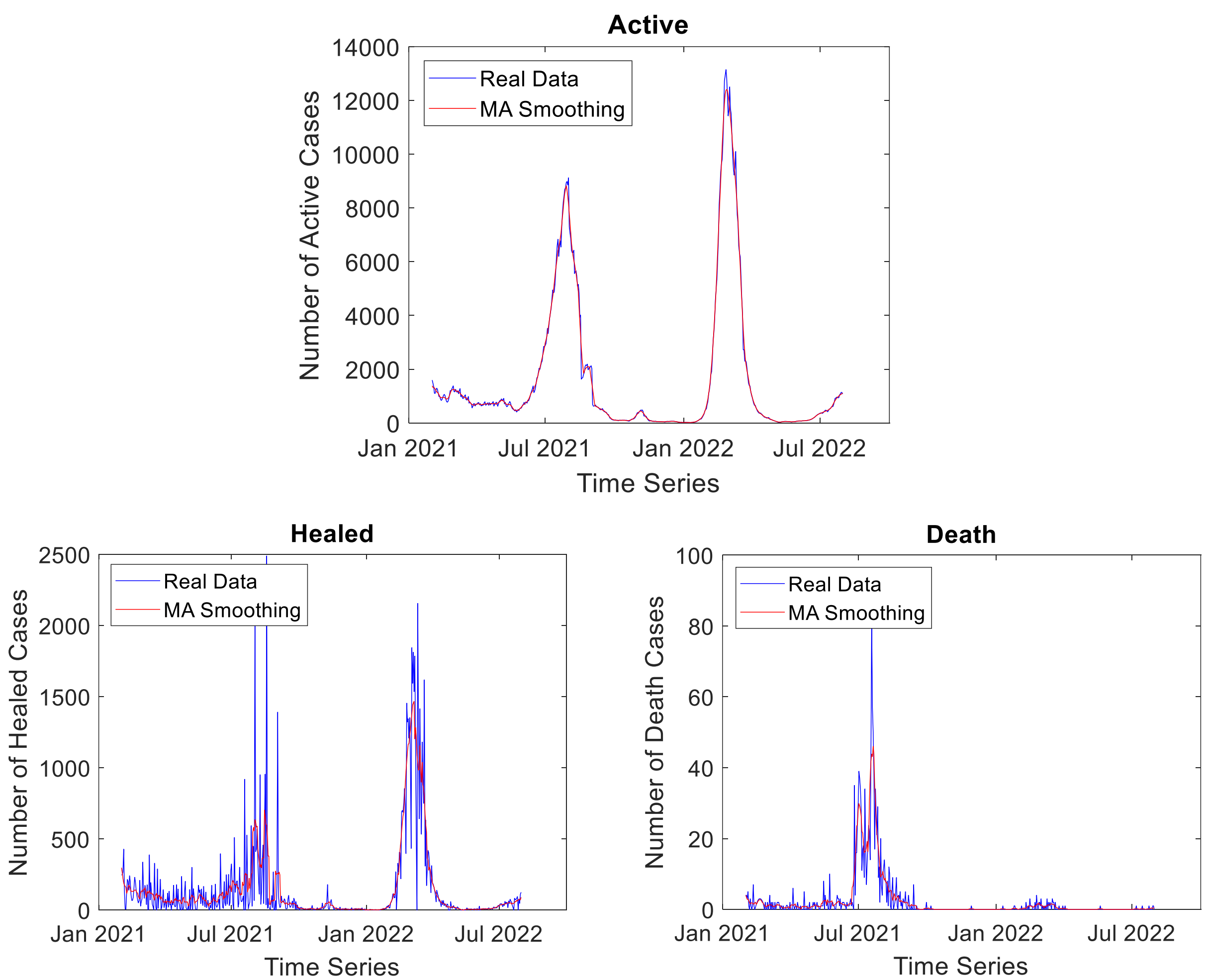


**FIGURE 1.** The plot of real data and data smoothing of active (a), recovered (b), and death cases (c) in Bandung

This process aims to reduce noise or fluctuations in the data before prediction analysis with GPR. There is a large spike in the number of COVID-19 cases around July 2021 and April 2022. This spike was caused by various factors, including the spread of new variants of the virus, changes in government policies related to social restrictions or community movement, as well as the level of compliance with health protocols such as mask use and social distancing. Increased mobility and social interaction may also contribute to the increase in cases. After MA smoothing, data fluctuations are reduced, especially in cases of recovery and death.

The covariance function used is a quadratic exponential covariance function, as described in Equation 3. The model hyperparameters are estimated by maximizing the likelihood function, as described in Equation 4. Prediction of the new data is done using Equations 5-7. The prediction results can be given as follows:

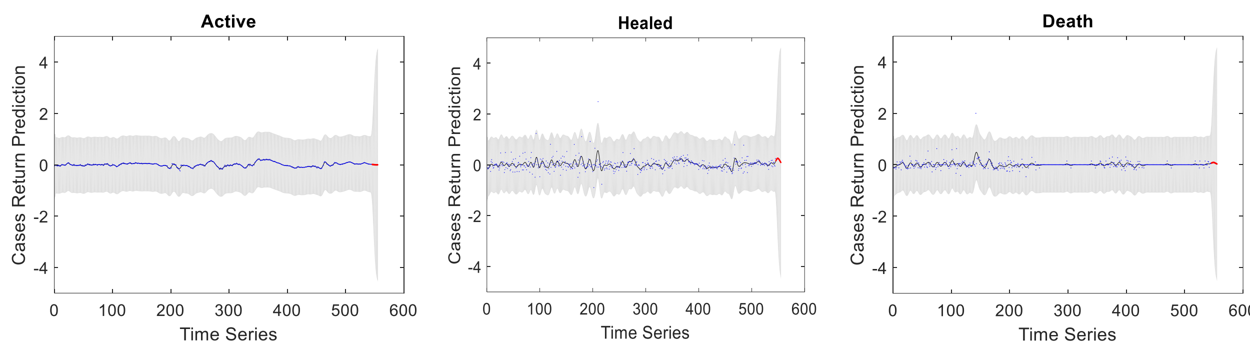


**FIGURE 2.** Prediction results of active (a), recovered (b), and death cases (c) in Bandung using GPR

The prediction results obtained provide an overview of the estimated number of active cases, recoveries, and COVID-19 deaths. The prediction results using GPR not only provide a single-point estimate but also include a confidence interval. The red dots in the graph represent the predicted values generated by the model, while the surrounding gray area shows the confidence interval of the prediction. The confidence interval provides information on how confident we are in the predictions given by the model. The wider the confidence interval, the greater the uncertainty associated with the prediction. This additional explanation helps to understand that the prediction result is not just a single number but also includes information about how much we can trust the prediction.

2.2. Spike detection results using Mahalanobis distance

The prediction results obtained from GPR are not only used to determine the estimated number of cases but also to detect whether a spike has occurred or not. Spike detection analysis is performed using the Mahalanobis distance formulated in Equation 8 and threshold determination using Equation 9. Spike detection aims to identify significant changes in case trends that may indicate a spike in infection. The results of spike detection provide additional insight into sudden changes in the number of cases that may require a rapid response or action in handling the COVID-19 pandemic. Below are the surge detection results on active cases:

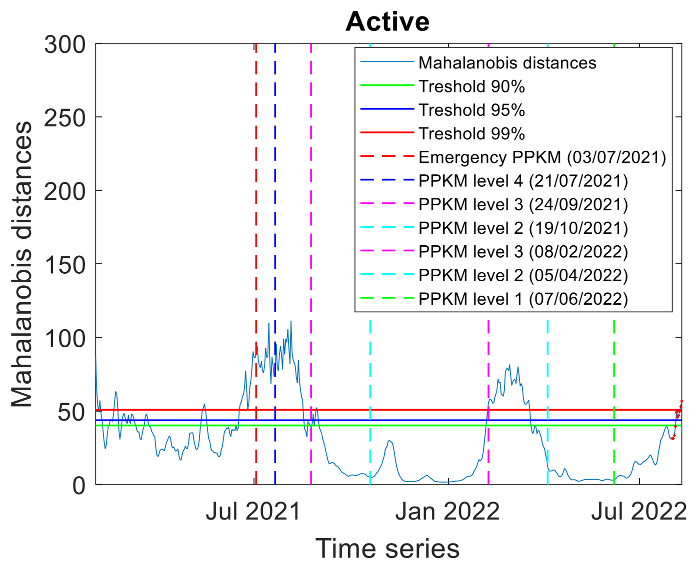


**FIGURE 3.** Spike detection results of active cases using Mahalanobis distance

In Figure 3, the Mahalanobis distance plot is used to identify spikes in active cases. The horizontal lines in the graph indicate thresholds at several levels of significance. Points that cross these thresholds are considered outliers or spikes in cases. Along the graph, there is a vertical line that marks the time of implementation of policies related to handling COVID-19 cases in Bandung City, such as the Enforcement of Restrictions on Community Activities, or " Pemberlakuan Pembatasan Kegiatan Masyarakat" (PPKM). The higher the PPKM level, the stricter the policy applied.

Spike detection results on active case prediction show an increase in red dots in the graph, indicating an increase in cases. These dots begin to exceed the Mahalanobis distance threshold, which provides an early warning of a possible spike in cases. In other words, a sudden change in the trend of cases past the Mahalanobis distance threshold can be considered an early indicator of a potential spike in cases to watch out for. There was also an increase in cured cases, which is in line with the spike in active cases. Below are the results of the Mahalanobis distance plots for recovered and deceased cases:

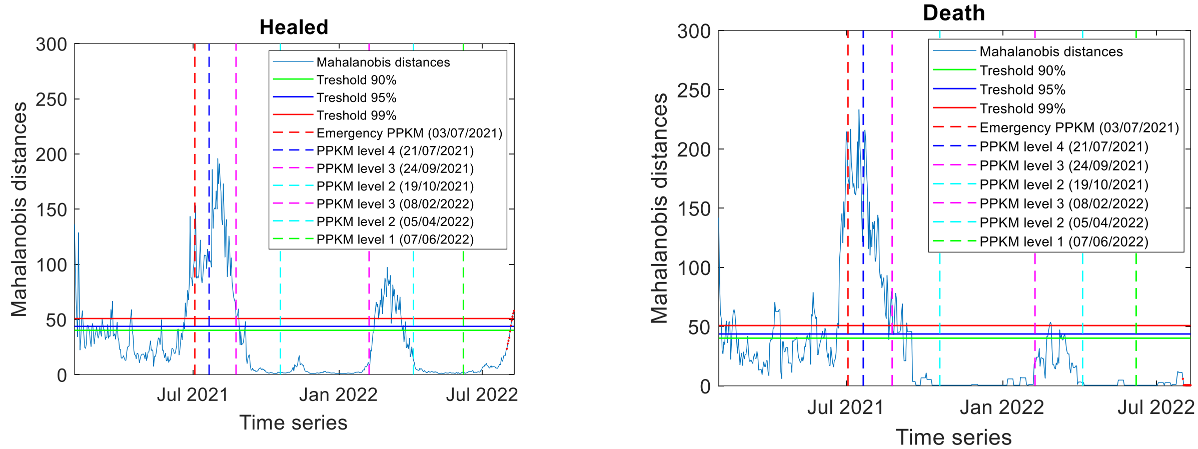


**FIGURE 4.** Spike detection results of healed and death cases using Mahalanobis distance

In addition to finding an overall spike in cases, this study also identified sub-districts in Bandung City that experienced a significant increase. Based on the Pareto diagram in Figure 5, it was found that Antapani, Coblong, Rancasari, and Arcamanik sub-districts had the largest spike in cases. This analysis provides important additional insights to focus treatments and policies in areas that are significantly affected.

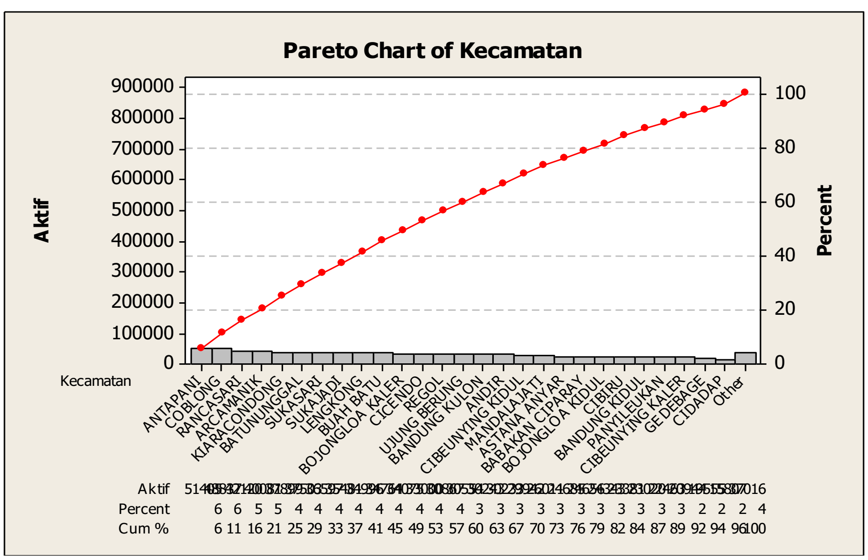


**FIGURE 5.** Pareto diagram results for the detection of Sub-districts that experience large spikes in active cases

An analysis was also conducted on the availability of beds in hospitals for isolation rooms, as a supporting facility in dealing with the surge in COVID-19. The results of the Mahalanobis distance calculation for isolation rooms in all hospitals in Bandung City show a surge in the use of beds for isolation rooms. This finding highlights the importance of measures to increase the availability of beds in hospitals to cope with the surge in cases. The Mahalanobis distance plot results are as follows:

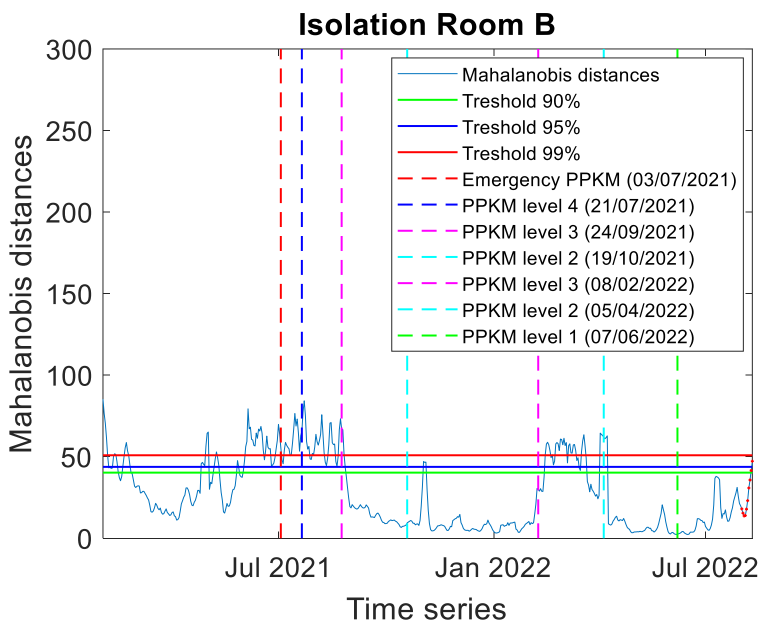


**FIGURE 6.** Spike detection results of hospital bed utilization cases using Mahalanobis distance

Analyzing the use of hospital beds for isolation rooms is an important part of anticipating COVID-19 spikes. Through the calculation of Mahalanobis distance, we were able to identify spikes in bed utilization for isolation rooms across hospitals in Bandung City. This finding underscores the urgency to increase hospital capacity to cope with the surge in cases. By ensuring adequate availability of health facilities, including beds for isolation rooms, it is expected that the health system can respond to the surge in cases more effectively and efficiently. Measures such as increasing bed capacity, procuring medical equipment, and improving human resources in hospitals can be prioritized in responding to the changing pandemic situation.

The results of this study are in line with previous studies that have used Gaussian processes to predict Covid-19 cases [45]–[51], as well as with other studies that utilize Mahalanobis distance and Pareto diagrams [52] to detect spikes in Covid-19 cases. However, what sets this study apart is its continuous approach. Here, the results of case prediction using Gaussian process regression are used as the basis to detect a continuous spike in COVID-19 cases.

## CONCLUSION

This research makes an important contribution to understanding and anticipating the dynamics of COVID-19 spread in Bandung City, Indonesia. Using the GPR method and spike detection using Mahalanobis distance, we successfully identified significant spikes in cases and sub-districts. These findings have direct implications for the development of COVID-19 response policies, including resource allocation and implementation of more effective prevention measures. With a better understanding of the spread patterns and early detection of surges, it is expected that future pandemic response efforts can be more targeted and responsive to sudden changes in the epidemiological situation. The approach used in this study can be used as a form of anticipation of the future.

## ACKNOWLEDGMENT

The authors thank to Universitas Padjadjaran for their support through the Research Data Library and Online or Riset Data Pustaka dan Daring (RDPD) program, titled "Exploring MNIST: Comparing Image Characteristics through Distribution Function Features". Thank you to the research team, Susi Setiyowati and Abdullah Sonhaji for their good cooperation in completing this research.

## CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

## REFERENCES

[1] J. Gonzalvez, E. Lezmi, T. Roncalli, et al. Financial applications of Gaussian processes and Bayesian optimization, SSRN, (2019). https://doi.org/10.2139/ssrn.3344332.

[2] E.A. Severo, J.C.F. De Guimarães, M.L. Dellarmelin, Impact of the COVID-19 pandemic on environmental

awareness, sustainable consumption and social responsibility: Evidence from generations in Brazil and Portugal, J. Cleaner Product. 286 (2021), 124947. https://doi.org/10.1016/j.jclepro.2020.124947.

[3] P. Gulati, A. Kumar, R. Bhardwaj, Impact of Covid19 on electricity load in Haryana (India), Int. J. Energy Res. 45 (2020), 3397–3409. https://doi.org/10.1002/er.6008.

[4] A. Hernandez-Matamoros, H. Fujita, T. Hayashi, et al. Forecasting of COVID19 per regions using ARIMA models and polynomial functions, Appl. Soft Comput. 96 (2020), 106610. https://doi.org/10.1016/j.asoc.2020.106610.

[5] T. Alamo, D. Reina, M. Mammarella, et al. Covid-19: Open-data resources for monitoring, modeling, and forecasting the epidemic, Electronics 9 (2020), 827. https://doi.org/10.3390/electronics9050827.

[6] T.T. da Silva, R. Francisquini, M.C.V. Nascimento, Meteorological and human mobility data on predicting COVID-19 cases by a novel hybrid decomposition method with anomaly detection analysis: A case study in the capitals of Brazil, Expert Syst. Appl. 182 (2021), 115190. https://doi.org/10.1016/j.eswa.2021.115190.

[7] J.O. Odhiambo, P. Ngare, P. Weke, et al. Modelling of COVID-19 transmission in Kenya using compound Poisson regression model, J. Adv. Math. Comput. Sci. (2020), 101–111. https://doi.org/10.9734/jamcs/2020/v35i230252.

[8] J. De-Graft Ankamah, E. Okyere, S.T. Appiah, et al. Nonlinear dynamics of COVID-19 SEIR infection model with optimal control analysis, Commun. Math. Biol. Neurosci. 2021 (2021), 13. https://doi.org/10.28919/cmbn/5095.

[9] M.S. Bryan, J. Sun, J. Jagai, et al. Coronavirus disease 2019 (COVID-19) mortality and neighborhood characteristics in Chicago, Ann. Epidemiol. 56 (2021), 47-54.e5. https://doi.org/10.1016/j.annepidem.2020.10.011.

[10] A.B. Gumel, E.A. Iboi, C.N. Ngonghala, et al. A primer on using mathematics to understand COVID-19 dynamics: Modeling, analysis and simulations, Infect. Dis. Model. 6 (2021), 148–168. https://doi.org/10.1016/j.idm.2020.11.005.

[11] R.C. Reiner Jr, R.M. Barber, J.K. Collins, et al. Modeling COVID-19 scenarios for the United States, Nat. Med. 27 (2020), 94–105. https://doi.org/10.1038/s41591-020-1132-9.

[12] P.C. Jentsch, M. Anand, C.T. Bauch, Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study, The Lancet Infectious Diseases 21 (2021) 1097–1106. https://doi.org/10.1016/s1473-3099(21)00057-8.

[13] L. Benahmadi, M. Lhous, A. Tridane, Mathematical modeling of COVID-19 in Morocco and the impact of controlling measures, Commun. Math. Biol. Neurosci. 2021 (2021), 53. https://doi.org/10.28919/cmbn/5697.

[14] A.I. Abioye, M.D. Umoh, O.J. Peter, et al. Forecasting of COVID-19 pandemic in Nigeria using real statistical data, Commun. Math. Biol. Neurosci. 2021 (2021), 2. https://doi.org/10.28919/cmbn/5144.

[15] G. Guan, Y. Dery, M. Yechezkel, et al. Early detection of COVID-19 outbreaks using human mobility data, PLoS ONE 16 (2021), e0253865. https://doi.org/10.1371/journal.pone.0253865.

[16] K. Leung, J.T. Wu, G.M. Leung, Real-time tracking and prediction of COVID-19 infection using digital proxies of population mobility and mixing, Nat. Commun. 12 (2021), 1501. https://doi.org/10.1038/s41467-021-21776-2.

[17] D. Murakami, T. Matsui, Improved log-Gaussian approximation for over-dispersed Poisson regression: Application to spatial analysis of COVID-19, PLoS ONE 17 (2022), e0260836. https://doi.org/10.1371/journal.pone.0260836.

[18] M.D. Hssayeni, A. Chala, R. Dev, et al. The forecast of COVID-19 spread risk at the county level, J. Big Data 8 (2021), 99. https://doi.org/10.1186/s40537-021-00491-1.

[19] M. Shahzad, A.H. Abdel-Aty, R.A.M. Attia, et al. Dynamics models for identifying the key transmission parameters of the COVID-19 disease, Alexandria Eng. J. 60 (2021), 757–765. https://doi.org/10.1016/j.aej.2020.10.006.

[20] H. Zhao, N.N. Merchant, A. McNulty, et al. COVID-19: Short term prediction model using daily incidence data, PLoS ONE 16 (2021), e0250110. https://doi.org/10.1371/journal.pone.0250110.

[21] G.J.B. Sousa, T.S. Garces, V.R.F. Cestari, et al. Mortality and survival of COVID-19, Epidemiol. Infect. 148 (2020), e123. https://doi.org/10.1017/s0950268820001405.

[22] M. Ujiie, S. Tsuzuki, N. Ohmagari, Effect of temperature on the infectivity of COVID-19, Int. J. Infect. Dis. 95 (2020), 301–303. https://doi.org/10.1016/j.ijid.2020.04.068.

[23] L. Wang, X. Ben, A. Adiga, Using mobility data to understand and forecast COVID19 dynamics, medRxiv 2020.12.13.20248129. https://doi.org/10.1101/2020.12.13.20248129.

[24] A. Schwabe, J. Persson, S. Feuerriegel, Predicting COVID-19 spread from large-scale mobility data, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, ACM, Virtual Event Singapore, 2021: pp. 3531–3539. https://doi.org/10.1145/3447548.3467157.

[25] T. Varsavsky, M.S. Graham, L.S. Canas, et al. Detecting COVID-19 infection hotspots in England using

large-scale self-reported data from a mobile application: a prospective, observational study, Lancet Public Health 6 (2021), e21–e29. https://doi.org/10.1016/S2468-2667(20)30269-3.

[26] E.A. Rashed, A. Hirata, One-year lesson: Machine learning prediction of COVID-19 positive cases with meteorological data and mobility estimate in Japan, Int. J. Environ. Res. Public Health 18 (2021), 5736. https://doi.org/10.3390/ijerph18115736.

[27] J. Liu, Z. Zhang, X. Fan, et al. Power system load forecasting using mobility optimization and multi-task learning in COVID-19, Appl. Energy 310 (2022), 118303. https://doi.org/10.1016/j.apenergy.2021.118303.

[28] B. Rostami-Tabar, J.F. Rendon-Sanchez, Forecasting COVID-19 daily cases using phone call data, Appl. Soft Comput. 100 (2021), 106932. https://doi.org/10.1016/j.asoc.2020.106932.

[29] Y. Karadayi, M.N. Aydin, A.S. Ogrenci, Unsupervised anomaly detection in multivariate spatio-temporal data using deep learning: Early detection of COVID-19 outbreak in Italy, IEEE Access 8 (2020), 164155–164177. https://doi.org/10.1109/access.2020.3022366.

[30] I.A. Baba, A. Yusuf, K.S. Nisar, et al. Mathematical model to assess the imposition of lockdown during COVID-19 pandemic, Results Phys. 20 (2021), 103716. https://doi.org/10.1016/j.rinp.2020.103716.

[31] D. J. McGrail, J. Dai, K. M. McAndrews, and R. Kalluri, "Enacting national social distancing policies corresponds with dramatic reduction in COVID19 infection rates," PLoS One, vol. 15, no. 7 July, pp. 1–9, 2020, doi: 10.1371/journal.pone.0236619.

[32] E.A. Rashed, A. Hirata, Infectivity upsurge by COVID-19 viral variants in Japan: Evidence from deep learning modeling, Int. J. Environ. Res. Public Health 18 (2021), 7799. https://doi.org/10.3390/ijerph18157799.

[33] M. Liu, R. Thomadsen, S. Yao, Forecasting the spread of COVID-19 under different reopening strategies, Sci Rep 10 (2020), 20367. https://doi.org/10.1038/s41598-020-77292-8.

[34] E. Giménez, M. Crespi, M.S. Garrido, et al. Multivariate outlier detection based on robust computation of Mahalanobis distances. Application to positioning assisted by RTK GNSS Networks, Int. J. Appl. Earth Observ. Geoinf. 16 (2012), 94–100. https://doi.org/10.1016/j.jag.2011.11.011.

[35] E. Cabana, R.E. Lillo, H. Laniado, Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators, Stat. Papers 62 (2019), 1583–1609. https://doi.org/10.1007/s00362-019-01148-1.

[36] V. Chaurasia, S. Pal, Application of machine learning time series analysis for prediction COVID-19 pandemic, Res. Biomed. Eng. 38 (2020), 35–47. https://doi.org/10.1007/s42600-020-00105-4.

[37] C.E. Rasmussen, C.K.I. Williams, Gaussian processes for machine learning, The MIT Press, Cambridge, 2006.

[38] S. Winarni, S.W. Indratno, Gaussian process regression with average hyperparameter, J. Phys.: Conf. Ser. 1496 (2020), 012002. https://doi.org/10.1088/1742-6596/1496/1/012002.

[39] S. Winarni, S.W. Indratno, K.N. Sari, Gaussian process regression using partition resampling model, Int. J. Math. Computer Sci. 17 (2022), 853–864.

[40] Q. Yan, J. Chen, L. De Strycker, An outlier detection method based on mahalanobis distance for source localization, Sensors 18 (2018), 2186. https://doi.org/10.3390/s18072186.

[41] P. Filzmoser, Identification of multivariate outliers: A performance study, Austrian J. Stat. 34 (2016), 127–138. https://doi.org/10.17713/ajs.v34i2.406.

[42] C. Leys, O. Klein, Y. Dominicy, et al. Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance, J. Exp. Soc. Psychol. 74 (2018), 150–156. https://doi.org/10.1016/j.jesp.2017.09.011.

[43] K.I. Penny, Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance, Appl. Stat. 45 (1996), 73. https://doi.org/10.2307/2986224.

[44] L. Wilkinson, Revising the pareto chart, Amer. Statistician 60 (2006), 332–334. https://doi.org/10.1198/000313006x152243.

[45] Y. Lei, K. Ozbay, K. Xie, Safety analytics at a granular level using a Gaussian process modulated renewal model: A case study of the COVID-19 pandemic, Accid. Anal. Prev. 173 (2022), 106715. https://doi.org/10.1016/j.aap.2022.106715.

[46] R.M.A. Velásquez, J.V.M. Lara, Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression, Chaos Solitons Fractals 136 (2020), 109924. https://doi.org/10.1016/j.chaos.2020.109924.

[47] S. Cui, E.C. Yoo, D. Li, et al. Hierarchical Gaussian processes and mixtures of experts to model COVID-19 patient trajectories, Biocomputing 2022 (2021), 266-277. https://doi.org/10.1142/9789811250477_0025.

[48] I. Hawryluk, H. Hoeltgebaum, S. Mishra, et al. Gaussian process nowcasting: Application to COVID-19 mortality reporting, in: Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, PMLR 161:1258-1268, 2021.

[49] S. Ketu, P.K. Mishra, Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection, Appl. Intell. 51 (2020), 1492–1512. https://doi.org/10.1007/s10489-020-01889-9.

[50] S. Dhamodharavadhani, R. Rathipriya, COVID-19 mortality rate prediction for India using statistical neural

networks and gaussian process regression model, Afr. Health Sci. 21 (2021), 194–206. https://doi.org/10.4314/ahs.v21i1.26.

[51] E. Gurjao , E. Guedes , P.L. Junior, et al. Autoregressive short-term COVID-19 cases forecasting with Gaussian processes regression, TechRxiv, 2022. https://doi.org/10.36227/techrxiv.18318638.v1.

[52] A. Rikasoraya, H. Purwoko, M.I. Firdaus, et al. Pareto analysis implementation for Covid-19 outbreak humanitarian supply chain by Indonesian national disaster agency (Bnpb) March-June 2020, Global Research on Sustainable Transport and Logistics, 2020.