Signal and Image Processing Letters
Vol. 7., No. 1, 2025, pp. 72-85
ISSN 2714-6677
https://simple.ascee.org/

Protecting South Sulawesi: Combination of Random Forest Regressor and SEIRS Mathematical Model in the Analysis and Prediction of the Spread of Covid-19

Nur Qadri Bahar [a,1,*], Syafruddin Side [b,2], Muhammad Abdy [b,3]

[a] Institut Teknologi Kalimantan, Balikpapan, Indonesia
[b] Makassar State University, Makassar, Indonesia
[1] nur.bahar@lecturer.itk.ac.id; [2] syafruddin@unm.ac.id; [3] muh.abdy@unm.ac.id
* corresponding author

ARTICLE INFO

ABSTRACT

This applied research seeks to examine the dynamics of Covid-19 transmission in South Sulawesi Province. Additionally, this study will also forecast the future spread of Covid-19. This research comprises two phases: the investigation of Covid-19 transmission utilizing the SEIRS mathematical model, incorporating Vaccination and PPKM (Enforcement of Community Activity Restrictions), and the prediction of Covid-19 spread through machine learning techniques. The utilized data is secondary data sourced from the South Sulawesi Provincial Health Office. This study developed a machine learning model utilizing a random forest regressor algorithm due to its proficiency in identifying nonlinear data patterns. This model effectively accounts for the variability of the dataset (target variable) with a R^2 score of 95%. The evaluation findings of the random forest regressor model indicated satisfactory performance, with a mean absolute error (MAE) of 29.57 and a root mean square error (RMSE) of 54.42 during training, and an MAE of 58.55 and an RMSE of 98.67 during testing. Given the significant variability in Covid-19 data and the prevalence of zeros in the dataset, the MAE and RMSE values for both training and testing are deemed acceptable. This model is designed to forecast daily Covid-19 instances in the future. This work not only employs machine learning but also analyzes the dissemination of Covid-19 through the SEIRS mathematical model, incorporating vaccination and PPKM factors. The SEIRS model analysis indicates that the disease-free equilibrium point is stable when $R_0 < 1$ and unstable when $R_0 > 1$. The fundamental reproduction rate derived from the vaccine's efficacy is $v = 65.3\%$, and the compliance rate with PPKM is $\theta = 1\%$. Consequently, $R_0 = 1.2407288$, indicating that Covid-19 will propagate in South Sulawesi Province. For $\theta = 11\%$, $R_0 = 1.0167230$. This indicates that Covid-19 will stabilize, and with $\theta = 69\%$, $R_0 = 0.2338689$, suggesting that Covid-19 will vanish from South Sulawesi Province.

Keywords
Machine Learning
Random Forest Regressor
SEIRS Model
Covid-19

1. Introduction

With a high rate of spread, cases of Covid-19 infection have occurred in almost all countries in early 2020, including Indonesia [1]. One of the provinces in Indonesia with high positive cases of Covid-19 is South Sulawesi Province. The first positive case of Covid-19 infection in South Sulawesi was announced on March 20, 2020. Since then, new cases of Covid-19 have begun to increase fluctuately with an increasing trend. Various anticipatory measures have been taken to suppress the spread of Covid-19, including calls for physical distancing, the Enforcement of Community Activity Restrictions (PPKM), the closure of shopping centers and tourist attractions, the use of masks, and mass vaccination [2]. Although the status of Covid-19 is no longer said to be a pandemic, the potential for a Covid-19 pandemic still exists. In addition, the negative impacts are still present today, including in South Sulawesi [3].

To assist the government in maintaining and protecting the community, understanding the behavior of the spread of Covid-19 is important. By understanding the behavior of the spread of Covid-19, the government can make quick and appropriate decisions in the event of a Covid-19 pandemic in the future. In this study, the behavior of the spread of Covid-19, the prediction of positive cases in the future and what initial steps the government can take if positive cases of Covid-19 increase in the future. Many Covid-19 research in Indonesia has been conducted, including the analysis of the spread of Covid-19 using the SEIR mathematical model and also the SIR mathematical model with fuzzy parameters [4][5]. Similar research using the SEIRD mathematical model has also been conducted with a focus on the region in South Kalimantan [6]. In addition to using a mathematical model of disease spread, research related to Covid-19 prediction is also carried out by applying machine learning [7][8][9].

This study will combine two methods, namely the SEIRS mathematical model and the random forest regressor model to produce an in-depth analysis and provide accurate prediction results. The SEIRS model is a modification of the SEIR model with the addition of vaccination and PPKM parameters and the addition of the assumption that people who have recovered from Covid-19 are still potentially infected. The SEIRS mathematical model built is expected to be able to describe the real situation well, so that the results obtained can be used as policy recommendations in handling endemic cases such as Covid-19 and similar diseases. In addition to the mathematical model of disease spread, another widely used method for disease spread analysis, disease classification, and even disease spread prediction is machine learning models. Machine learning models are widely applied to several types of diseases, including cardiovascular disease, diabetes, and cancer [10][11][12][13][14]. Machine learning models, especially random forest regressors, are machine learning algorithms that are quite good at recognizing nonlinear data patterns, so they are very suitable for predicting daily positive Covid-19 cases [8]. These two methods provide reinforcement in terms of analysis and prediction of Covid-19.

2. Method

In general, this research integrates two approaches: predicting Covid-19 case numbers using machine learning techniques and analyzing the transmission dynamics of Covid-19 through the SEIRS mathematical model. The study utilizes secondary data, specifically daily records of Covid-19 patients from districts and municipalities in South Sulawesi Province that implemented the Community Activity Restrictions (PPKM) policy during the pandemic, spanning the period from March 20, 2020, to December 31, 2021.

This study employs the machine learning methodology known as CRISP-DM (Cross Industry Standard Process for Data Mining). CRISP-DM offers a standardized data mining methodology as a universal problem-solving approach for the business or subject under investigation. This study employs CRISP-DM by adhering to the subsequent steps:

1. Business and Data Understanding

   This initial stage is to understand the goals and needs of the researcher. Comprehensive knowledge of the research object will make it easier to establish the structure and flow of research to the drawing of *insights* and conclusions. This study utilizes daily data on Covid-19 cases in South Sulawesi from March 20, 2020, to December 31, 2021. Data was obtained from the South Sulawesi Provincial Health Office. Daily data on positive Covid-19 cases in South Sulawesi Province is data with wide intervals, this is because positive cases of Covid-19 can be very high on certain days, and vice versa on other days there is not a single positive case of Covid-19. Data characteristics like these require a model that can read data patterns with large intervals and high fluctuations. One model that is quite good at handling this type of data is the Random Forest Regressor. The Random Forest Regressor model is able to recognize nonlinear data patterns well, so this model was chosen to predict daily positive cases of Covid-19 in South Sulawesi Province.

2. Data Preparation and Basic Exploratory Data Analysis (EDA)

This stage is a series of activities to prepare the dataset, starting with data cleaning, handling missing values and outliers, data transformation to exploratory data analysis. The initial processing of the raw data results in a final dataset that is ready to be modeled. The EDA stage will help us identify interesting and potential parts of the data that can then be used for initial hypotheses.

3. Modelling and Evaluation

After conducting Exploratory Data Analysis (EDA), the modeling phase begins by dividing the data into two subsets: 80% for training and 20% for testing. The training subset is then used to build a prediction model using the Random Forest Regressor algorithm. Once the model is trained, its accuracy is evaluated using two performance indicators: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). If these values fall within an acceptable range, the model proceeds to the testing phase, where the same metrics are recalculated to assess its predictive ability. A robust model should consistently produce low MAE and RMSE values during both training and testing. This stage generally reflects model performance according to standard practices in machine learning as well as criteria tailored to the specific algorithm applied [10].

The second approach employed in this research involves examining the transmission of Covid-19 through the SEIRS mathematical model, which incorporates both vaccination efforts and the implementation of Community Activity Restrictions (PPKM). This model extends the traditional SEIR framework by factoring in the effects of vaccination and mobility restrictions. The process begins with the formulation of the SEIRS model to study the disease's spread while accounting for these interventions. After that, an analysis of the model's equilibrium points is performed, followed by simulations to illustrate the dynamics of Covid-19 transmission in South Sulawesi.

3. Results and Discussion

3.1. Machine Learning Model

3.1.1. Data Preparation and Basic Exploratory Data Analysis (EDA)

Before entering the modeling stage with machine learning, data preparation is first carried out including data cleaning, handling outliers and ensuring that there is no multicollinearity between each feature. The dataset used consists of several variables presented in Table 1. The verification findings indicate the absence of missing values and duplicate entries in the dataset. For subsequent analysis, the variables 'Date' and 'N' were eliminated, leaving 'Susceptible (S)', 'Exposed (E)', 'Infected (I)', 'Recovered (R)', 'Died', and 'Total cases' as the variables for further examination.

Table 1. Variable Dataset of Daily Covid-19 Cases in South Sulawesi

Variable	Information
Date	Date of observation
Susceptible (S)	Number of sub-populations vulnerable to covid-19
Exposed (E)	Number of sub-populations with covid-19 symptoms
Infected (I)	Number of sub-populations infected with covid-19
Recovered (R)	Number of sub-populations recovered from covid-19
Die	Number of deaths
Total cases	Total cases
N	Total population

A sample of the dataset for each variable in Table 1 can be seen in Table 2.

Table 2. Sample of Daily Covid-19 Case Data in South Sulawesi

	Date	Susceptible (S)	Exposed (E)	Infected (I)	Recovered (R)	Die	Total cases	N
516	17-Aug-21	9,047,710	11,056	476	758	23	12,313	9,060,000
603	12-Nov-21	9,059,836	155	5	4	0	164	9,060,000
507	08-Aug-21	9,045,538	13,003	735	724	23	14,485	9,060,000
374	28-Mar-21	9,058,597	1,230	35	138	0	1,403	9,060,000
315	28-Jan-21	9,055,618	3,148	678	556	3	4,385	9,060,000
258	02-Dec-20	9,058,331	1,472	123	74	1	1,670	9,060,000
463	25-Jun-21	9,058,765	1,065	151	19	3	1,238	9,060,000
625	04-Dec-21	9,059,990	0	0	10	0	10	9,060,000
384	07-Apr-21	9,058,753	778	225	244	0	1,247	9,060,000
544	14-Sep-21	9,056,937	2,684	159	220	3	3,066	9,060,000

The predictor variables in this study are 'Susceptible (S)', 'Exposed (E)', 'Recovered (R)', 'Die', and 'Total cases', whereas the goal variable is 'Infected (I)'. The remaining variables are then examined for a statistical summary, whether there are any abnormal values, whether the distribution is symmetrical or asymmetrical. A summary of the statistics of each variable is shown in Table. 3.

Table 3. Summary of Daily Covid-19 Cases in South Sulawesi

	Susceptible (S)	Exposed (E)	Infected (I)	Recovered (R)	Die	Total cases
count	6.53E+08	653.00	653.00	653.00	653.00	653.00
mean	9.06E+12	1,800.74	168.45	164.97	3.43	2,137.59
std	3.01E+09	2,614.99	235.06	232.30	5.64	3,010.52
min	9.04E+12	0.00	0.00	0.00	0.00	1.00
25%	9.06E+12	69.00	27.00	14.00	0.00	201.00
50%	9.06E+12	610.00	83.00	74.00	2.00	753.00
75%	9.06E+12	2,747.00	180.00	206.00	4.00	3,126.00
max	9.06E+12	13,022.00	1,359.00	1,763.00	42.00	15,170.00

Table 3 indicates that the minimum and maximum values for all variables are plausible. However, the mean and median of all variables are not the same even some variables have vastly different means and medians, this shows that the distribution of all variables is asymmetrical. Fig. 1 illustrates the unequal distribution of data.

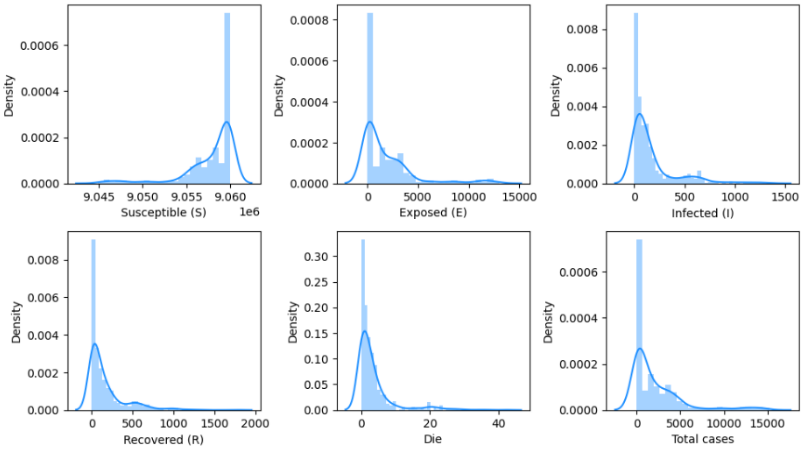

Fig. 1. Variable Distribution of Covid-19 Daily Case Data in South Sulawesi

By seeing that all variables are asymmetrical and even have a fairly extreme slope (long distribution tail), logarithmic transformations are carried out so that the distribution of each variable is close to the normal distribution. From the results of the logarithmic transformation, new variables were obtained, namely 'Susceptible (S)_log', 'Exposed (E)_log', 'Infected (I)_log', 'Recovered (R)_log', 'Die'_log, and 'Total cases_log'. The comparison of each variable with the transformation result variable is presented in the form of a boxplot which can be seen in Fig. 2.

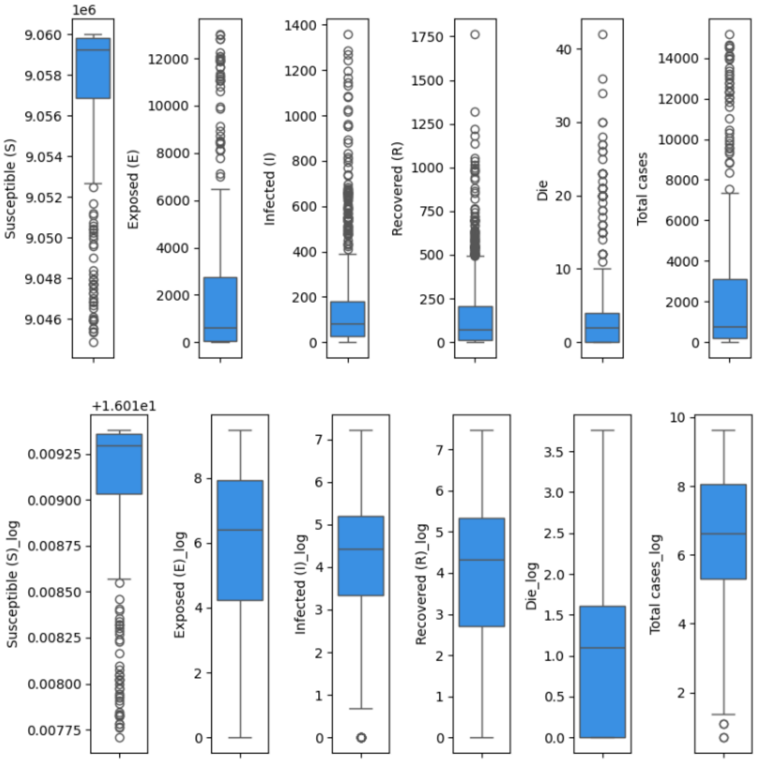

Fig. 2. Boxplot of Variable Daily Covid-19 Case Data in South Sulawesi

The last stage before getting into modeling with machine learning is the multicollinearity study. To see if any of the predictor variables (features) are highly correlated with each other, heatmap visualization is used. A heatmap visualization can be seen in Fig. 3.

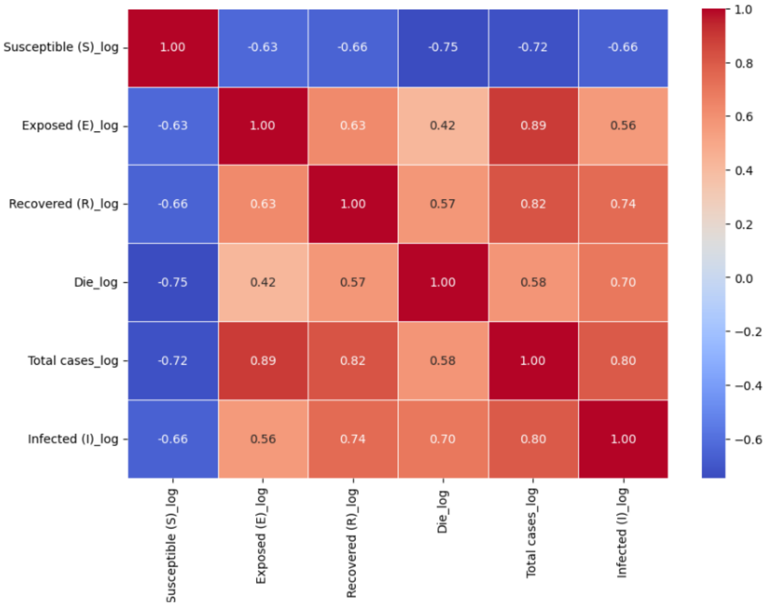

Fig. 3. Correlation Heatmap Between Each Feature

Correlation values between features that are greater than |0.8| is a sign of multicollinearity between features. From Fig. 3 it is found that 'Total kasus_log' has a correlation of > |0.8| against 'Exposed (E)_log' and 'Recovered (R)_log', so two of these three features should be discarded. The 'Total cases_log' variable was selected to be retained because it had the highest correlation value to the target variable, while 'Exposed (E)_log' and 'Recovered (R)_log' were discarded. To ensure that the remaining features are no longer highly correlated with each other, a re-check is carried out with the heatmap as shown in Fig. 4.

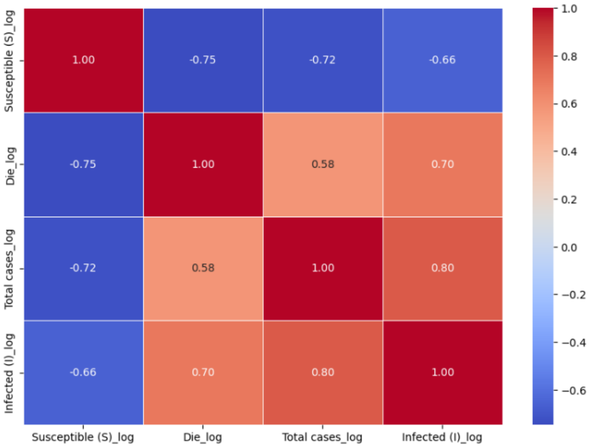

Fig. 4. Correlation Heatmap Between Each Feature That Has Been Transformed

Fig. 4 shows that there is no high correlation between each feature, so the data is ready for use for the modeling stage.

3.1.2. Modeling and Evaluation

Modeling commences by partitioning the dataset into two segments: training data and test data, allocated in an 80%:20% proportion. The Random Forest model was subsequently trained on the prepared data to produce the finalized model. The evaluation of the trained model yielded an MAE score of 29.57 and an RMSE of 54.42. The trained model was then tested using test data and evaluated. From the results of the evaluation using test data, an MAE score of 58.55 and an RMSE of 98.67 were obtained. The model is deemed adequate for predicting Covid-19 cases in South Sulawesi based on the assessment metrics obtained throughout the training and testing processes.

3.1.3. Predictions

The Random Forest Regressor model was used to simulate Covid-19 predictions in South Sulawesi. The results of the prediction simulation can be seen in Fig. 5.

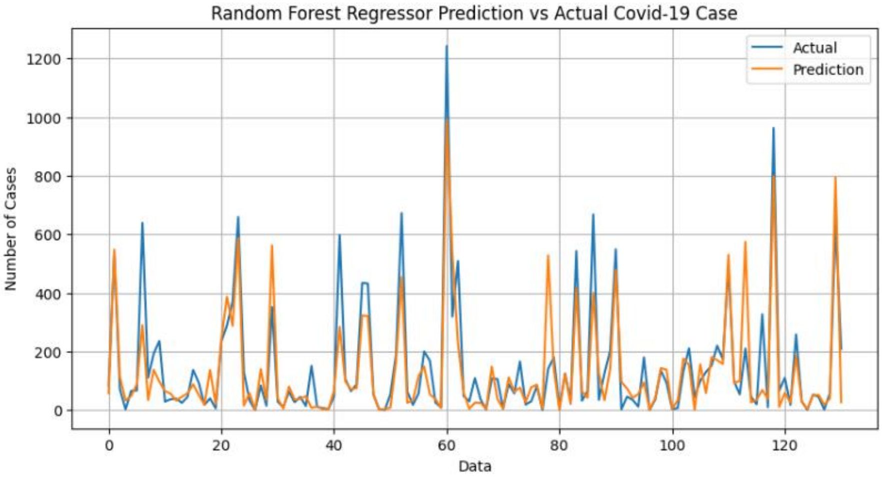

Fig. 5. Covid-19 Case Prediction Simulation

3.2. SEIRS Mathematical Model

3.2.1. Formulation of SEIRS Mathematical Model on the Spread of Covid-19 in South Sulawesi Province

The formation of the SEIRS model began by dividing the population into four sub-populations, namely Susceptible (S), Exposed (E), Infected (I), and Recovered (R). Individuals who are susceptible to contracting Covid-19 can become infected due to interaction with previously infected individuals. The changes that occur in each human subpopulation in cases of Covid-19 transmission in South Sulawesi using the SEIRS model can be interpreted into the diagram in Fig. 6.

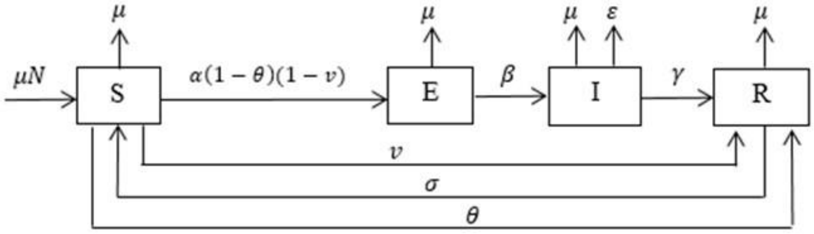

Fig. 6. Transmission Scheme of the SEIRS Model Covid-19 Spread

Table 3 displays the variables and parameters utilized in the SEIRS mathematical model for the transmission of Covid-19.

According to the SEIRS epidemic model depicted in Fig. 6, the rate of change in the number of individuals within each sub-population can be elucidated by equations (1)-(4)

$$\frac{dS}{dt} = \mu N + \sigma R - \big(\alpha I\big((1-\theta)(1-v)\big) + v + \theta + \mu\big)S \tag{1}$$

$$\frac{dE}{dt} = \alpha I\big((1-\theta)(1-v)\big)S - (\beta + \mu)E \tag{2}$$

$$\frac{dI}{dt} = \beta E - (\mu + \varepsilon + \gamma)I \tag{3}$$

$$\frac{dR}{dt} = (\theta + v)S + \gamma I - (\mu + \sigma)R \tag{4}$$

Table 3. Variables and Parameters

Parameter	Information
N	Number of populations
S	Number of susceptible sub-populations
E	Number of exposed sub-populations
I	Number of infected sub-populations
R	Number of recovered sub-populations
α	the rate of migration from a group of people who are susceptible to Covid-19 (susceptible) to a group of people who are starting to have symptoms of Covid-19 (exposed).
β	the rate of movement from a group of people who are starting to have symptoms of Covid-19 (exposed) to a group of people who have been confirmed to have Covid-19 (infected).
γ	the rate of transfer from a group of people who have been infected with Covid-19 (infected) to a group of recovered people (recovered)
μ	Natural birth and death rate
θ	The proportion of individuals who comply with PPKM
ε	Covid-19 death rate
v	Vaccine effectiveness in susceptible sub-population
σ	The rate of transfer from recovered human groups to human groups that are susceptible to being infected with Covid-19 (susceptible) both in terms of the lack of vaccine effectiveness and individuals who have been infected with Covid-19 who can be susceptible again

3.2.2. SEIRS Mathematical Model Analysis of the Spread of Covid-19 in South Sulawesi

1) Model Equilibrium Point

The disease-free equilibrium point and the endemic equilibrium point are obtained by making each equation in equation (1)-(4) equal to zero, i.e. $\frac{dS}{dt} = 0, \frac{dE}{dt} = 0, \frac{dI}{dt} = 0$, and $\frac{dR}{dt} = 0$, so that equations (5)-(8) are obtained.

$$0 = \mu N + \sigma R - \left(\alpha I\left((1-\theta)(1-v)\right) + v + \theta + \mu\right)S \tag{5}$$

$$0 = \alpha I\left((1-\theta)(1-v)\right)S - (\beta + \mu)E \tag{6}$$

$$0 = \beta E - (\mu + \varepsilon + \gamma)I \tag{7}$$

$$0 = (\theta + v)S + \gamma I - (\mu + \sigma)R \tag{8}$$

A direct substitution method will be utilized to determine the values of S, E, I, and R for both the disease-free and endemic equilibrium states of the SEIRS model. The disease-free equilibrium point is a state in which the transmission of Covid-19 halts, leading to $I = 0$. Equations (9)-(12) are determined through little algebraic manipulation of equations (5)-(8):

$$S = \frac{\mu + \sigma r}{\alpha i\left((1-\theta)(1-v)\right) + v + \theta + \mu} \tag{9}$$

$$E = \frac{\alpha i\left((1-\theta)(1-v)\right)s}{\beta + \mu} \tag{10}$$

$$I = \frac{\beta e}{(\mu + \varepsilon + \gamma)} \tag{11}$$

$$R = \frac{(\theta + v)s + \gamma i}{\mu + \sigma} \tag{12}$$

By setting $I = 0$, the disease-free equilibrium point can be derived from equations (9)-(12), specifically

$$(s, e, i, r) = \left(\frac{\mu\left((v+\theta+\mu)(\mu+\sigma) - (\theta+v)\sigma\right) + (\sigma(\theta+v)\mu)}{(v+\theta+\mu)\left((v+\theta+\mu)(\mu+\sigma) - (\theta+v)\sigma\right)}, 0, 0, \frac{(\theta+v)\mu}{(v+\theta+\mu)(\mu+\sigma) - (\theta+v)\sigma}\right) \tag{13}$$

Furthermore, in the same way, namely substituting each equation in equations (9)-(12), the endemic equilibrium point value of the SEIRS model for the spread of Covid-19 in South Sulawesi is obtained, namely

$$(S^*, E^*, I^*, R^*) = \begin{pmatrix} \dfrac{(\mu+\varepsilon+\gamma)(\beta+\mu)}{\beta\left(\alpha\left((1-\theta)(1-v)\right)\right)} \\ \dfrac{\begin{array}{c}\alpha[\mu\beta\alpha(1-\theta)(1-v)(\mu+\sigma) + \sigma(\theta+v)(\mu+\varepsilon+\gamma)(\beta+\mu) - (v+\theta+\mu)(\mu+\varepsilon+\gamma)(\beta+\mu)(\mu+\sigma)] \\ \left((1-\theta)(1-v)\right)(\mu+\varepsilon+\gamma)(\beta+\mu)\end{array}}{(\beta+\mu)[\alpha(\mu+\varepsilon+\gamma)(\beta+\mu)(\mu+\sigma)(1-\theta)(1-v) - \sigma\gamma]\left[\beta\left(\alpha(1-\theta)(1-v)\right)\right]} \\ \dfrac{\mu\beta\alpha(1-\theta)(1-v)(\mu+\sigma) + \sigma(\theta+v)(\mu+\varepsilon+\gamma)(\beta+\mu) - (v+\theta+\mu)(\mu+\varepsilon+\gamma)(\beta+\mu)(\mu+\sigma)}{\alpha(\mu+\varepsilon+\gamma)(\beta+\mu)(\mu+\sigma)(1-\theta)(1-v) - \sigma\gamma} \\ \dfrac{(\mu+\varepsilon+\gamma)(\beta+\mu)\left(\alpha i\left((1-\theta)(1-v)\right) + v + \theta + \mu\right) - \mu\beta\left(\alpha\left((1-\theta)(1-v)\right)\right)}{\sigma\beta\left(\alpha\left((1-\theta)(1-v)\right)\right)} \end{pmatrix} \tag{14}$$

2) Basic Reproduction Number

The basic reproduction number is derived using the next generation matrix approach. This method involves analysing both contributing and inhibiting factors in the transmission rate, specifically focusing on the exposed and infected groups within the population. The mathematical expression used to obtain this value is presented in Equation (15).

$$K = F' \cdot (V')^{-1} \tag{15}$$

Based on the system of equations (2) and (3), then:

$$\frac{de}{dt} = \alpha i\big((1-\theta)(1-v)\big)s - (\beta+\mu)e$$

$$\frac{di}{dt} = \beta e - (\mu+\varepsilon+\gamma)i$$

So that it is obtained

$$F = \begin{pmatrix} \alpha i\big((1-\theta)(1-v)\big)s \\ 0 \end{pmatrix}$$

$$\bar{F} = \begin{pmatrix} 0 & \alpha\big((1-\theta)(1-v)\big)s \\ 0 & 0 \end{pmatrix} \tag{16}$$

$$V = \begin{pmatrix} (\beta+\mu)e \\ (\mu+\varepsilon+\gamma)i - \beta e \end{pmatrix}$$

$$\bar{V} = \begin{pmatrix} \beta+\mu & 0 \\ -\beta & \mu+\varepsilon+\gamma \end{pmatrix} \tag{17}$$

So the inverse of the equation matrix (17) is obtained, namely

$$\bar{V}^{-1} = \begin{pmatrix} \dfrac{1}{\beta+\mu} & 0 \\ \dfrac{\beta}{(\beta+\mu)(\mu+\varepsilon+\gamma)} & \dfrac{1}{\mu+\varepsilon+\gamma} \end{pmatrix} \tag{18}$$

Next, the eigenvalue of the matrix K will be determined, based on equation (15)

$$K = \begin{pmatrix} 0 & \alpha\big((1-\theta)(1-v)\big)s \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \dfrac{1}{\beta+\mu} & 0 \\ \dfrac{\beta}{(\beta+\mu)(\mu+\varepsilon+\gamma)} & \dfrac{1}{\mu+\varepsilon+\gamma} \end{pmatrix}$$

$$K = \begin{pmatrix} \dfrac{\alpha\beta\big((1-\theta)(1-v)\big)s}{(\beta+\mu)(\mu+\varepsilon+\gamma)} & \dfrac{\alpha\big((1-\theta)(1-v)\big)s}{\mu+\varepsilon+\gamma} \\ 0 & 0 \end{pmatrix} \tag{19}$$

Upon acquiring the matrix K in equation (19), the eigenvalue will thereafter be determined using the formula $Det(\lambda I - K) = 0$, where I represents the identity matrix. The fundamental quantity of reproductions will be ascertained according to the magnitude of the maximal eigenvalue (λ).

$$\left|\lambda\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}-\begin{bmatrix}\dfrac{\alpha\beta\big((1-\theta)(1-v)\big)s}{(\beta+\mu)(\mu+\varepsilon+\gamma)} & \dfrac{\alpha\big((1-\theta)(1-v)\big)s}{\mu+\varepsilon+\gamma}\\ 0 & 0\end{bmatrix}\right| = 0 \tag{20}$$

So that the eigenvalue is obtained based on equation (20), namely

$$\lambda_1 = \frac{\alpha\beta\big((1-\theta)(1-v)\big)s}{(\beta+\mu)(\mu+\varepsilon+\gamma)}$$

$$\lambda_2 = 0$$

So the greatest eigenvalue is applied, namely $\lambda_1 = \frac{\alpha\beta\left((1-\theta)(1-v)\right)s}{(\beta+\mu)(\mu+\varepsilon+\gamma)}$

So that the basic reproduction number is obtained after the substitution of the value of the disease-free equilibrium point as in equation (21)

$$R_0 = \frac{\alpha\beta\big((1-\theta)(1-v)\big)\big((\mu+\sigma)(\theta+v)\mu\big)}{\big((\beta+\mu)(\mu+\varepsilon+\gamma)\big)(v+\theta+\mu)\big((v+\theta+\mu)(\mu+\sigma)-(\theta+v)\sigma\big)} \tag{21}$$

3.2.3. Equilibrium Point Stability Analysis

Based on equations (1)-(4) the following Jacobian matrix (J) can be formed:

$$J = \begin{bmatrix} -\big(\alpha i\big((1-\theta)(1-v)\big)+v+\theta+\mu\big) & 0 & -\big(\alpha(1-\theta)(1-v)\big)s & \sigma\\ \alpha i\big((1-\theta)(1-v)\big) & -(\beta+\mu) & \alpha(1-\theta)(1-v)s & 0\\ 0 & \beta & -(\mu+\varepsilon+\gamma) & 0\\ \theta+v & 0 & \gamma & -(\mu+\sigma)\end{bmatrix} \tag{22}$$

Theorem 1

The disease-free equilibrium point of the mathematical model for the transmission of Covid-19 is considered stable if $R_0 < 1$ and unstable if $R_0 > 1$.

Proof:

Substitution of disease-free equilibrium points to the matrix J of equation (22), so that a new matrix is obtained as in equation (23)

$$J = \begin{bmatrix} -(v+\theta+\mu) & 0 & -\big(\alpha(1-\theta)(1-v)\big)s & \sigma\\ 0 & -(\beta+\mu) & \alpha(1-\theta)(1-v)s & 0\\ 0 & \beta & -(\mu+\varepsilon+\gamma) & 0\\ \theta+v & 0 & \gamma & -(\mu+\sigma)\end{bmatrix} \tag{23}$$

Then the eigenvalue of the equation matrix (23) is searched with the following description:

$$|\lambda I - J| = 0$$

$$\left|\begin{bmatrix} \lambda+v+\theta+\mu & 0 & \big(\alpha(1-\theta)(1-v)\big)s & -\sigma\\ 0 & \lambda+\beta+\mu & -(\alpha(1-\theta)(1-v)s) & 0\\ 0 & -\beta & \lambda+\mu+\varepsilon+\gamma & 0\\ -(\theta+v) & 0 & -\gamma & \lambda+\mu+\sigma\end{bmatrix}\right| = 0 \tag{24}$$

Next, substitute S in equation (24) and make a translation so that it is obtained

$$\lambda^4 + A\lambda^3 + B\lambda^2 + C\lambda + D = 0 \tag{25}$$

According the Routh-Hurwitz criterion, equation (25) will possess a negative root if all the signs in each of its terms are positive. The disease-free equilibrium point is stable if $R_0 \leq 1$ and unstable if $R_0 > 1$.

Theorem 2

The equilibrium point of the endemic mathematical model for the propagation of Covid-19 is stable and asymptotic.

Proof:

The endemic equilibrium point is applicable when $I \neq 0$; hence, a new matrix is derived from matrix J in equation (22), as presented in equation (26)

$$J = \begin{bmatrix} -(Pi+Q) & 0 & -Ps & \sigma \\ Pi & -X & Ps & 0 \\ 0 & \beta & -Y & 0 \\ W & 0 & \gamma & -Z \end{bmatrix} \tag{26}$$

With,

$$P = \alpha(1-\theta)(1-v)$$

$$Q = v + \theta + \mu$$

$$W = \theta + v$$

$$X = \beta + \mu$$

$$Y = \mu + \varepsilon + \gamma$$

$$Z = \mu + \sigma$$

Then the eigenvalue is searched with the following description:

$$|\lambda I - J| = 0$$

$$\begin{bmatrix} \lambda + Pi + Q & 0 & Ps & -\sigma \\ -Pi & \lambda + X & -Ps & 0 \\ 0 & -\beta & \lambda + Y & 0 \\ -W & 0 & -\gamma & \lambda + Z \end{bmatrix} = 0 \tag{27}$$

Obtained

$$\lambda^4 + K\lambda^3 + L\lambda^2 + M\lambda + N = 0 \tag{28}$$

According to the Routh-Hurwitz criterion, if every root of the characteristic equation (λ) has a positive value, the equilibrium point is considered stable.

3.2.4. SEIRS Mathematical Model Simulation of the Spread of Covid-19 in South Sulawesi Province

The simulation of the SEIRS mathematical model for Covid-19 spread in South Sulawesi was carried out using software. The initial conditions for S(0), E(0), I(0), and R(0), as well as the model's parameter values employed in this simulation, are presented in Table 4 and Table 5.

Table 4. Initial Value of South Sulawesi's Covid-19 SEIRS Model

Variabel	Initial Value	Source
S	0.837313	Provincial Health Office. South Sulawesi
E	0.1389560	Provincial Health Office. South Sulawesi
I	0.01219	Provincial Health Office. South Sulawesi
R	0.01194	Provincial Health Office. South Sulawesi

Table 5. Parameter values of the South Sulawesi Covid-19 SEIR Model

Parameter	**Parameter Values**	**Source**
α	0.55	[15]
β	0.00095	[16]
γ	0.000667	[4]
σ	0.0047	[17]
v	0.653	[18]
θ	0.69	[19]
μ	0.00625	[4]
ε	0.00000208	[20]

The equilibrium point values of the SEIRS model are ascertained by replacing the uncontrolled parameter values from Table 3 and Table 4 into Equations (1)-(4), which are set to zero, resulting in the following system of equations (29)-(32)

$$\frac{ds}{dt} = 0.00625 + 0.347r - 0.0591635si - 1.34925s \tag{29}$$

$$\frac{de}{dt} = 0.0591635si - 0.00720e \tag{30}$$

$$\frac{di}{dt} = 0.00095e - 0.00691908i \tag{31}$$

$$\frac{dr}{dt} = 0.000667i - 0.35325r + 1.343s \tag{32}$$

equations (29)-(32) provide balance points of the SEIRS model endemic to the spread of Covid-19, namely:

$$(s, e, i, r) = (0.2082535, 0, 0, 0.7917465) \tag{33}$$

The results of the SEIRS model simulation of the spread of Covid-19 in South Sulawesi can be seen in Fig. 7.

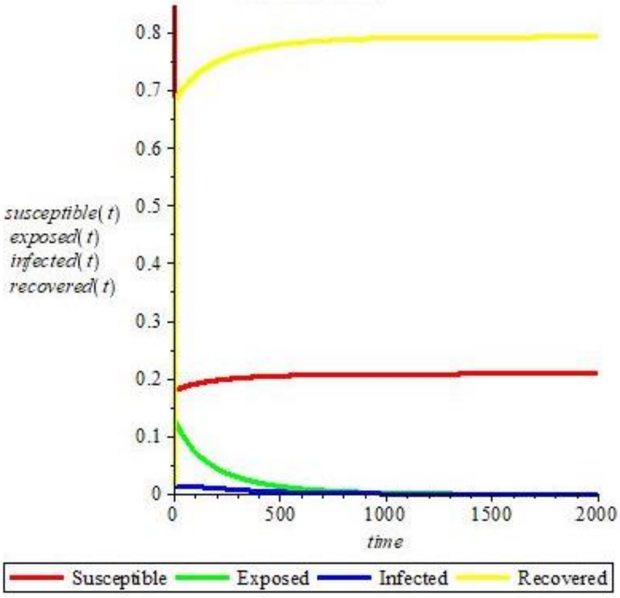

Fig. 7. Graph of Covid-19 Cases in South Sulawesi Province

Fig. 7 shows that sub-population S or vulnerable individuals, sub-population E or symptomatic individuals, and sub-population I or infected individuals decreased over time, while sub-population R or recovered individuals increased. The results of Covid-19 research in Indonesia show that with the administration of vaccines, the basic reproduction number (R_0) 1% produced are $R_0 = 3.2094$, this means that if someone is infected with Covid-19, that person can transmit it to three other people. However, if the vaccine administration is maximized to 100%, the spread of Covid-19 can be stopped [4]. In addition to vaccine administration, other things that also affect the acceleration of the decline in the number of Covid-19 infections are the implementation of health protocols and also the level of controlled vaccine effectiveness [5].

This study diverges from prior research that focused solely on a single parameter, such as vaccine utilization or the enforcement of health protocols subsequently modified to PPKM. Instead, it examines two critical parameters in the government's strategy to manage Covid-19 transmission in Indonesia, particularly in South Sulawesi Province: the efficacy of the administered vaccines and the compliance rate with PPKM among the populace in South Sulawesi Province. The Sinovac vaccine, exhibiting an efficacy rate of 65.3%, combined with a compliance rate of 69% for PPKM, yields a basic reproduction number of $R_0 = 0.2338689$, indicating the cessation of Covid-19 transmission and the eventual eradication of the virus in South Sulawesi Province.

4. Conclusion

The Random Forest Regressor model is able to recognize patterns from the daily Covid-19 case dataset in South Sulawesi accurately enough so that it can be used to predict Covid-19 cases in South Sulawesi Province. The results of the evaluation of the random forest regressor model showed the performance of the model with a MAE of 29.57 and RMSE of 54.42 in the training process and a MAE of 58.55 and RMSE of 98.67 in the testing process. Although the MAE and RMSE values showed less impressive results, considering that the Covid-19 dataset used contained a value of 0 (zero) which contributed to the increase in the value of MAE and RMSE, this Random Forest Regresor Model is acceptable.

Quite good results were also obtained from the SEIRS mathematical model. With vaccination with a vaccine effectiveness rate of 65.3% and the proportion of people who comply with PPKM of 69%, a value of $R_0 = 0.2338689$ was obtained, which means that Covid-19 transmission will decrease over time and will disappear from the population. This can be seen from the group of vulnerable individuals (S), symptomatic individuals (E), and infected individuals (I) which continues to decrease while for recovered individuals (R) continues to increase over time.

The combination of these two methods provides significant reinforcement compared to using only one of the two to analyze or predict. The SEIRS mathematical model is able to represent the real situation while the Random Forest Regressor is able to describe the upcoming conditions well according to the dataset it has. So that the results of this research can be used as a basis for taking preventive steps or quick and effective handling steps to protect South Sulawesi Province from the endemic Covid-19.

Acknowledgment

We sincerely thank the Mathematics Study Program, Institut Teknologi Kalimantan for their support during this research. We are also grateful to those who have provided the dataset and our colleagues for the in-depth feedback, which greatly contributed to the success of this study.

References

[1] D. F. N. Anisa, I. Mukhlash, and M. Iqbal, “Deteksi Berita Online Hoax Covid-19 Di Indonesia Menggunakan Metode Hybrid Long Short Term Memory dan Support Vector Machine,” *J. Sains dan Seni ITS*, vol. 11, no. 3, 2023, doi: 10.12962/j23373520.v11i3.83227.

[2] M. I. Mubarok, *Hasil Survei Sosial Demografi Dampak Covid-19 Provinsi Sulawesi Selatan 2020*. Badan Pusat Statistik Provinsi Sulawesi Selatan, 2020.

[3] N. Q. Bahar and N. M. Rauf, “Analisis Penyebaran COVID-19 di Sulawesi Selatan Menggunakan Machine Learning (Regression Models),” *EQUIVA J. Math. Inf. Technol.*, vol. 2, no. 2, 2024, [Online]. Available: https://journal.itk.ac.id/index.php/equiva/article/view/1258

[4] S. Annas, M. Isbar Pratama, M. Rifandi, W. Sanusi, and S. Side, “Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia,” *Chaos, Solitons and Fractals*, vol. 139, p. 110072, 2020, doi: 10.1016/j.chaos.2020.110072.

[5] M. Abdy, S. Side, S. Annas, W. Nur, and W. Sanusi, “An SIR epidemic model for COVID-19 spread with fuzzy parameter: the case of Indonesia,” *Adv. Differ. Equations*, vol. 2021, no. 1, 2021, doi: 10.1186/s13662-021-03263-6.

[6] M. A. Karim and Y. Yulida, “Analisis Kestabilan dan Sensitivitas pada Model Matematika SEIRD dari Penyebaran COVID-19: Studi Kasus di Kalimantan Selatan,” *J. Binawakya*, vol. 16, no. 5, p. 7003, 2021, [Online]. Available: http://ejurnal.binawakya.or.id/index.php/MBI

[7] H. Lamouadene et al., “Detection of COVID-19, lung opacity, and viral pneumonia via X-ray using machine learning and deep learning,” Comput. Biol. Med., vol. 191, p. 110131, 2025, doi: 10.1016/j.compbiomed.2025.110131.

[8] M. Rostami and M. Oussalah, “A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest,” Informatics Med. Unlocked, vol. 30, p. 100941, 2022, doi: 10.1016/j.imu.2022.100941.

[9] A. M. Fayyaz, S. Jadid Abdulkadir, S. U. Hassan, S. Al-Selwi, E. Sumiea, and L. Talib, “The Role of Advanced Machine Learning in COVID-19 Medical Imaging: A Technical Review,” Results Eng., vol. 26, p. 105154, 2025, doi: 10.1016/j.rineng.2025.105154.

[10] K. Sumwiza, C. Twizere, G. Rushingabigwi, P. Bakunzibake, and P. Bamurigire, “Enhanced cardiovascular disease prediction model using random forest algorithm,” Informatics Med. Unlocked, vol. 41, p. 101316, 2023, doi: 10.1016/j.imu.2023.101316.

[11] Z. Sun *et al.*, “Applications of resting-state fMRI and machine learning in cognitive impairment in type 2 diabetes mellitus: A scoping review,” *Meta-Radiology*, vol. 3, no. 2, p. 100136, 2025, doi: 10.1016/j.metrad.2025.100136.

[12] K. Pang, “A comparative study of explainable machine learning models with Shapley values for diabetes prediction,” *Healthc. Anal.*, vol. 7, p. 100390, 2025, doi: 10.1016/j.health.2025.100390.

[13] C. H. Feng, M. L. Disis, C. Cheng, and L. Zhang, “Multimetric feature selection for analyzing multicategory outcomes of colorectal cancer: random forest and multinomial logistic regression models,” *Lab. Investig.*, vol. 102, no. 3, pp. 236–244, 2022, doi: 10.1038/s41374-021-00662-x.

[14] K. Puttegowda *et al.*, “Enhanced machine learning models for accurate breast cancer mammogram classification,” *Glob. Transitions*, vol. 7, pp. 276–295, 2025, doi: 10.1016/j.glt.2025.04.007.

[15] A. Intissar, “A Mathematical Study of a Generalized SEIR Model of COVID-19,” *SciMedicine J.*, vol. 2, pp. 30–67, 2020, doi: 10.28991/scimedj-2020-02-si-4.

[16] S. He, Y. Peng, and K. Sun, “SEIR modeling of the COVID-19 and its dynamics,” *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1667–1680, 2020, doi: 10.1007/s11071-020-05743-y.

[17] P. Brouqui *et al.*, “COVID-19 re-infection,” *Eur. J. Clin. Invest.*, vol. 51, no. 5, pp. 4–8, 2021, doi: 10.1111/eci.13537.

[18] J. Beladiya et al., “Safety and efficacy of COVID-19 vaccines: A systematic review and meta-analysis of controlled and randomized clinical trials,” Rev. Med. Virol., vol. 34, no. 1, Jan. 2024, doi: 10.1002/RMV.2507.

[19] M. Yamani, “Tingkat Kepatuhan Protokol Kesehatan Masa PPKM - jatengdaily,” *jatengdaily*, 2021. [Online]. Available: https://jatengdaily.com/2021/tingkat-kepatuhan-protokol-kesehatan-masa-ppkm/

[20] P. Wintachai and K. Prathom, “Stability analysis of SEIR model related to efficiency of vaccines for COVID-19 situation,” Heliyon, vol. 7, no. 4, Apr. 2021, doi: 10.1016/J.HELIYON.2021.E06812.