COMPARISON OF RANDOM FOREST AND DECOMPOSITION METHODS IN THE PREDICTION OF CENTRAL OXYGEN SUPPLY AT dr. RADEN SOEDJONO SELONG HOSPITAL

Rahmat Rifki Aolia Akhzami[1], Wiwit Pura Nurmayanti[1*], Umam Hidayaturrohman[1], M. Hadiyan Amaly[1], Siti Hadijah Hasanah[2]

[1]Department of Statistics, Hamzanwadi of University (INDONESIA)
[2]Study Program of Statistics, Faculty of Sciences and Technology, Universitas Terbuka (INDONESIA)

**Corresponding author: wiwit.adiwinata3@gmail.com*

Abstract

Forecasting is a method for estimating a future value using past data. In forecasting, there are several methods, including Random Forest and Decomposition. Both methods have the advantage that they do not require assumptions compared to other forecasting methods. Issues related to Covid-19 have not yet been resolved and have caused many negative effects, one of which is the scarcity of medical gases such as oxygen. In mid-2021, oxygen to scarcity occurred in Java and Bali because of the Covid-19. This scarcity affects the supply of oxygen to other central hospitals in Indonesia, one of which is RSUD dr.R.Soedjono Selong NTB. For this reason, a strategy is needed in dealing with cases of central oxygen scarcity. One strategy that can be done is to make predictions to estimate the amount of oxygen supply each week, and the methods that can be used are Random Forest and Decomposition. The purpose of this study was to compare the Random Forest and Decomposition on the prediction of oxygen supply at RSUD dr. Raden Soedjono. The results of the analysis show that Decomposition is better than Random Forest. This

is because the MAPE value produced by Decomposition (25.68%) is smaller than Random Forest (52.31%).
Keywords: Forecasting, Random Forest, Decomposition, Covid-19, oxygen

1 INTRODUCTION

Prediction is a process to estimate the possibility that will occur in the future based on information or data in the past and currently owned, so that errors formed can be minimized [1]. There are several methods of forecasting between Random Forest and Decomposition. The Random Forest method is a set of decision trees that are used for classification and prediction of data by entering inputs into the root part that is above and then down to the part of the leaf that is below [2]. Decomposition is a method that uses four main components in forecasting future values. The four components include trend, seasonality, cycles, and errors [3].

This forecasting method can be applied in various fields, one of which is the health sector to determine the amount of oxygen supply in the central hospital in the future. Hospitals are health service institutions that provide complete individual health services that provide inpatient, outpatient, and emergency services. In April 2021, Indonesia was recorded to have 3,039 hospitals. This addition occurred not only due to the need for health services in general, but also because Indonesia, which is currently still hit by the Covid-19 pandemic, which requires an increase in the number of hospitals or health service institutions [4].

The spike in Covid-19 cases has resulted in many impacts, one of which is the scarcity of medical gas in the form of oxygen [5]. In the middle of 2021, oxygen scarcity occurred in Java and Bali due to the Covid-19 pandemic [4]. This scarcity has an impact on the supply of oxygen to other hospitals in Indonesia. One of them is RSUD dr. R. Soedjono Selong in NTB Province. RSUD dr. R. Soedjono Selong usually gets a supply of 100 tubes per day, but due to the impact of oxygen scarcity on the islands of Java and Bali, dr. R. Soedjono Selong gets an oxygen supply of 60 to 80 tubes per day. This is also

caused by the diversion of oxygen which was originally to be supplied to RSUD dr. R. Soedjono Selong, however, the fulfillment of oxygen needs for other hospitals in the province of NTB made the oxygen supply allocation for dr. R. Soedjono Selong reduced [6].

There are several previous studies that used the Random Forest and Decomposition method for forecasting including those conducted by Primajaya & Sari [7] with the title Random Forest Algorithm for Prediction of Precipitation, the results in this study are the MAE value of 0.35, RMSE of 0.46, and accuracy of 99.45%. A similar study was conducted by Siburian & Mulyana [8], namely the prediction of cell phone prices using the Random Forest method, the result of this research is the prediction accuracy rate using the Random Forest method is 81%. Another study using the Decomposition method was carried out by Satyawati et al [9] with the title prediction of the poor in Indonesia using decomposition analysis, the results of this study indicate that the additive decomposition model is better than the multiplicative decomposition model, this is due to the accuracy of the additive decomposition model (5 ,96%) is 10% smaller than the multiplicative decomposition model. Based on these previous studies, the purpose of this research is to see the results of predictions and comparisons of the Random Forest and Decomposition methods on the prediction of central oxygen supply at RSUD dr. Raden Soedjono Selong.

2 METHODOLOGY

The data used in this study is secondary data, namely central oxygen supply data at RSUD dr. Raden Soedjono Selong for the period from January to November 2021. The data used is sourced from RSUD dr. Raden Soedjono Selong. The method used in this research is the Random Forest and Decomposition method with the help of R Studio and Minitab software.

1.1 Random Forest

The Random Forest method is a development of the Classification and Regression Tree (CART) method by applying the Boostrap

Aggregating (Bagging) and Random Feature Selection methods. In conducting the analysis using Random Forest, there are no certain assumptions that must be met [10]. Random Forest is a method that consists of a structured set of trees that each casts a vote unit for the class and the results obtained are based on the most decisions. The basic technique used by Random Forest is Decision Tree. In other words, a random forest is a set of decision trees that are used for classification and prediction of data by entering input into the roots above and then down to the leaves below [2].

Random Forest uses an ensemble bagging strategy that can overcome the overfitting problem that occurs if the train data is small [11]. The results of the Random Forest analysis for classification are the mode of each tree of the forest built, while the prediction results are obtained from the average value of each tree [12]. The algorithm to follow when constructing a tree using a Random Forest is divided into two parts. The first is the creation of "n" trees to form a random forest. The second is to make predictions from Random Forests that have been made [2].

Input:

- D, a dataset consisting of d rows
- k, the number of trees

The Random Forest method process in constructing a tree:

a. Generate sample data Di data by taking random data from dataset D with replacement.
b. Use sample data Di to build a tree to i (i=1,2,…k)
c. Steps 1 and 2 are repeated k times

In the classification process, the individuals are based on the vote of the most votes in the tree population collection, while for the regression using the average results of the tree population. Stages of analysis of the Random Forest method

1) In Random Forest analysis, the first step is to input data into the R Studio software.
2) Divide the data into training and testing data. Then identify the Random Forest model with the ntree value (number of trees) that

has been determined using training data, testing data is used to see the error rate of the model made.

3) Identify the model to get the best model with the smallest error value.
4) Evaluating the best model and forecasting the central oxygen supply using the best model with the smallest error value.

1.2 Decomposition

According to Makkulau [3] the decomposition method is a method that uses four main components in predicting future values. The four components include trend, seasonality, cycle, and error. The decomposition method is based on the assumption that the existing data is a combination of several components, which is simply described as Equation (1) [13].

$$\begin{aligned} \text{Data} &= \text{Pattern} + \text{error} \\ &= f(\text{trend, cycle, seasonal}) + \text{error} \end{aligned} \quad (1)$$

The assumption above means that there are four components that affect a time series, namely three components that can be identified due to certain patterns, including trends, cycles and seasonality. The general mathematical equation of the decomposition approach is in Equation (2), where Xt = time series value (actual data) in period, Tt = trend component in period t, St = seasonal component in period t, Ct = cyclical component in period t, It = component of irregular error in period t, t = period (time).

$$X_t = f(T_t, S_t, C_t, I_t) \quad (2)$$

The research stages are generally described in the flow chart as follows:

1) The first step in the Decomposition analysis is to input data into the Minitab software.
2) Start the analysis by selecting the Stat tools, then select Time Series, then select decomposition.
3) Then input the name of the data variable to be analyzed and fill in the Generate Forecasts with how long the forecasting will take, then click Ok.

1.3 Measurement of Forecasting Error

Mean Absolute Percentage Error (MAPE) is a standard measure that is often used in measuring the suitability of a forecasting or prediction method. MAPE is used to see how far (in %) the prediction results deviate from the actual data. The smaller the MAPE generated from a forecasting method, the better the forecasting method will be calculated by Equation (3) [14], where Xt = Actual data, Ft = Forecasting data, n = Number of data.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|X(t)-F(t)|}{X(t)}\ x\ 100\% \quad (3)$$

3 RESULTS

1.4 Descriptive Analysis

Descriptive analysis in this study was used to see an overview of the central oxygen supply data at RSUD dr. Raden Soedjono Selong January to November 2021.

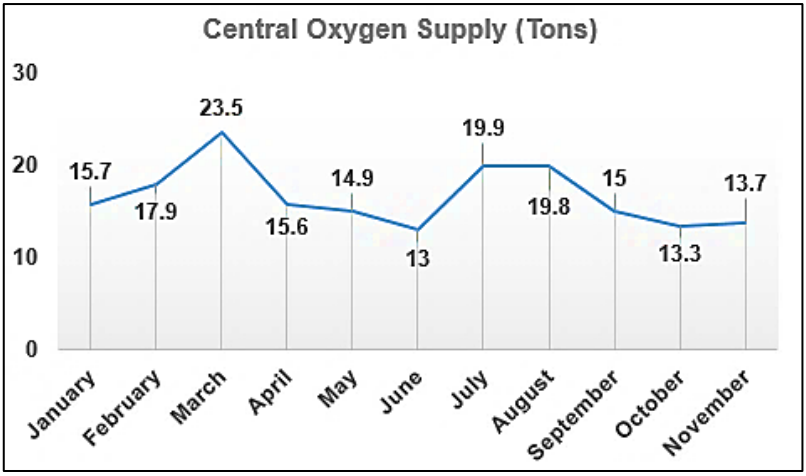

Figure 1. *Central Oxygen Supply.*

Based on ***Figure. 1***, it can be seen that the central oxygen supply of RSUD dr. Raden Soedjono Selong in 2021, where the highest supply of central oxygen is in March with a total of 23.5 tons and the lowest

supply in June of 13.0 tons. This decrease occurred due to the surge in Covid-19 cases which resulted in a lack of central oxygen supply at the dr. Raden Soedjono Selong.

1.5 Random Forest

Forecasting analysis using the Random Forest method requires training data and testing data. Training data is used to create a model and testing data is used to see the accuracy of the model. In forecasting analysis using Random Forest, it is used as well as comparing several Ntree values in modeling using training data. This is done to see the best model based on the smallest MAPE value. The following is the MAPE value based on the selection of several Ntree values.

Table 1. *Value of Ntree and MAPE Random Forest*

Ntree	25	50	100	150	200	250	300	500
MAPE	52.86	53.07	52.31	52.34	52.73	52.59	52.63	52.57

Based on ***Table. 1***, the smallest MAPE value is obtained with an Ntree 100 value of 52.31. The following is a graph of the prediction results using the Random Forest method using the Ntree value of 100.

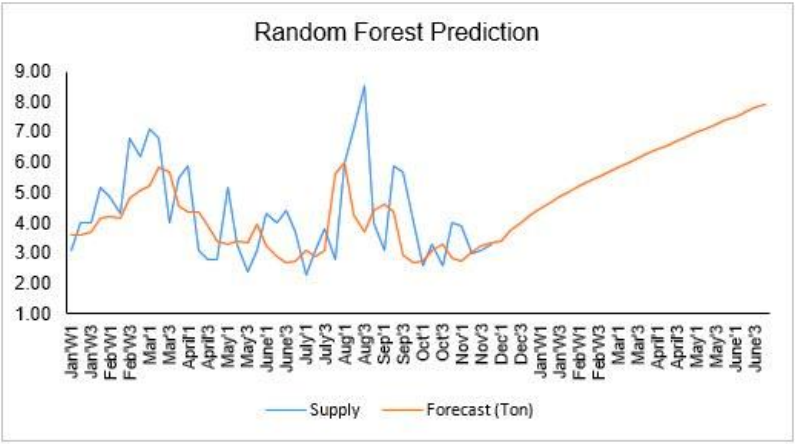

Figure 2. *Plot of Random Forest Prediction Results.*

The graph in ***Figure. 2***. shows that the prediction results for each week from the first week of December to the 4th week of June 2022 continue to increase until it reaches 7.9 tons of central oxygen. The following are the results of forecasting using the Random Forest method in ***Table. 2.***

Table 2. *Random Forest Prediction.*

Year	Month	Week	Supply (Ton)
2021	December	1	4.1
		2	4.4
		3	4.7
		4	4.9
2022	January	1	5.1
		2	5.2
		3	5.4
		4	5.6
	February	1	5.7
		2	5.8
		3	6.0
		4	6.1
	March	1	6.2
		2	6.4
		3	6.5
		4	6.6
	April	1	6.7
		2	6.8
		3	6.9
		4	7.0
	May	1	7.1
		2	7.3
		3	7.4
		4	7.5
	June	1	7.6
		2	7.7
		3	7.8
		4	7.9

2.3 Decomposition

Following are the prediction results using the Decomposition method as shown in ***Figure. 3.***

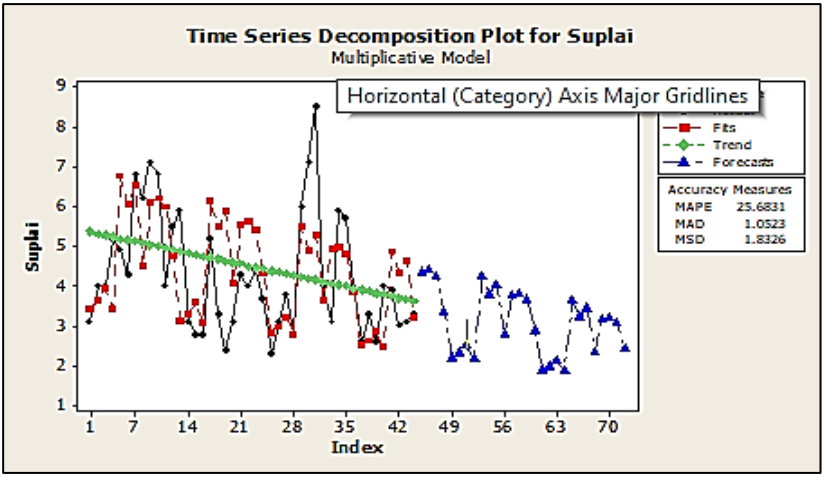

Figure 3. *Decomposition Output.*

Figure. 3. shows the output of the predicted central oxygen supply using the Decomposition method. Descriptively, the central oxygen supply is seasonal, seen from the black line that shows the actual data, the red line shows the estimated value, while the blue line shows the predicted results. With an error value of MAD = 1.0523, MSD = 1.8326, and MAPE = 25.68.

Table 3. *Decomposition Forecasting Results.*

Year	Month	Week	Supply (Ton)
2021	December	1	4.36
		2	4.41
		3	4.24
		4	3.37
2022	January	1	2.20

Year	Month	Week	Supply (Ton)
		2	2.32
		3	2.50
		4	2.16
	February	1	4.25
		2	3.78
		3	4.05
		4	2.79
	March	1	3.77
		2	3.81
		3	3.66
		4	2.90
	April	1	1.89
		2	1.99
		3	2.14
		4	1.85
	May	1	3.62
		2	3.22
		3	3.43
		4	2.36
	June	1	3.18
		2	3.20
		3	3.07
		4	2.43

Table. 3 showed the prediction results using the Decomposition method in December 2021 to June 2022. The highest prediction results occur in the second week of December 2021 with a figure of

4.41 tons, while the lowest prediction occurs in the fourth week of April 2022 with a number 1, 85 tons.

2.4 Comparison Of Error Values

Table 4. *Comparison of MAPE Values.*

	Decomposition	***Random Forest***
MAPE	25.68	52.31

Table. 4 showed that the smallest MAPE value is generated by the Decomposition method with a value of 25.68, while the Random Forest method produces a MAPE value of 52.31. This shows that the Decomposition method is the best method because it is able to produce the smallest error value in the prediction of central oxygen demand at RSUD dr. Raden Soedjono Selong was compared with the Random Forest method.

4 CONCLUSION

General description of central oxygen supply data at RSUD dr. Raden Soedjono Selong in 2021, where the highest supply of central oxygen is in March with a total of 23.5 tons and the lowest supply in June of 13.0 tons. The smallest MAPE value is generated by the Decomposition method with a value of 25.68, while the Random Forest method produces a MAPE value of 52.31. This shows that the Decomposition method is the best method because it is able to produce the smallest error value in the prediction of central oxygen demand at RSUD dr. Raden Soedjono Selong was compared with the Random Forest method.

REFERENCES

[1] M. L. Suliztia, "Penerapan Analisis Random Forest pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask," Universitas Islam Indonesia, 2020.

[2] R. A. Haristu dan P. H. P. Rossa, "Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unknown Battleground," *MEANS (Media Inf. Anal. dan Sist.*, vol. 4, no.

2, hal. 120–128, 2019, doi: 10.54367/means.v4i2.545.

[3] Makkulau, R. Raya, dan S. Marlinda, "Aplikasi Metode Dekomposisi Pada Peramalan Jumlah Kelahiran," *Semin. Nas.*, hal. 535–545, 2017.

[4] D. M. Purnamasari, "PERSI: Rumah Sakit di Indonesia Kewalahan Terima Pasien Covid-19," *Kompas.com*. 2021, [Daring]. Tersedia pada: https://nasional.kompas.com/read/2021/06/20/12163011/persi-rumah-sakit-di-indonesia-kewalahan-terima-pasien-Covid-19?page=all.

[5] I. Ayu, "Ternyata Ini Penyebab Tabung Oksigen Langka, Ada Penimbunan?" Bisnis.com, 2021.

[6] Poroslombok.com, "Duh! Stok Oksigen di RS dr. Soedjono Selong Mulai Menipis." Poroslombok.com, 2021.

[7] A. Primajaya dan B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, hal. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.

[8] V. W. Siburian dan I. E. Mulyana, "Prediksi Harga Ponsel Menggunakan Metode Random Forest," *Pros. Annu. Res. Semin. 2018*, vol. 4, no. 1, 2018.

[9] N. M. W. Satyawati, I. M. Candiasa, dan N. M. S. Mertasari, "Prediksi Penduduk Miskin Di Indonesia Menggunakan Analisis Dekomposisi," *J. Ilm. Pendidik. Mat.*, vol. 9, no. 1, hal. 77–88, 2021.

[10] M. J. Wulansari, "Analisis Faktor - faktor Yang Mempengaruhi Seseorang Terkena Penyakit Diabetes Melitus Menggunakan Regresi Random Forest," Universitas Islam Indonesia, 2018.

[11] A. yudha Samudra, "Pendekatan Random Forest Untuk Model Peramalan Harga Tembakau Rajang di Kabupaten Temanggung," Universitas Sanata Dharma, 2019.

[12] R. Dwi Lingga P., C. Fatichah, dan D. Purwitasari, "Deteksi

Gempa Berdasarkan Data Twitter Menggunakan Decision Tree, Random Forest, dan SVM," *J. Tek. ITS*, vol. 6, no. 1, hal. 159–162, 2017, doi: 10.12962/j23373539.v6i1.22037.

[13] O. J. Kendek, J. D. Prang, dan M. Paendong, "Prediksi Jumlah Pengunjung Perpustakaan Universitas Sam Ratulangi Manado Menggunakan Metode Dekomposisi," *d'CARTESIAN*, vol. 3, no. 1, hal. 73, 2014, doi: 10.35799/dc.3.1.2014.4000.

[14] A. yunita Resnaningtyas, "Penerapan Metode Fuzzy Untuk Prediksi Tingkat Akurasi Pada Data Mackey-Glass," Universitas Brawijaya, 2017.