

Forecasting Number of New Cases Daily COVID-19 in Central Java Province Using Exponential Smoothing Holt-Winters

Dinda Fitri Irandi^{#1}, Aniq A. Rohmawati^{#2}, P. H. Gunawan^{#3}

School of Computing, Telkom University Jl. Telekomunikasi, Bandung, Jawa Barat, Indonesia

> ¹ dindafitri@student.telkomuniversity.ac.id, ²aniqatiqi@telkomuniversity.ac.id, 3 phgunawan@telkomuniversity.ac.id

Abstract

There is hard to mention how long the COVID-19 pandemic will discontinue. There are some factors, including the public's efforts to slow spread and researchers' work to observe more about this outbreak. From the beginning of the health crisis, particularly following the announcement of the first positive case In Indonesia due to the COVID-19 on March 2, 2020. Afterwards, the number of daily cases increase simultaneously in other regions in Indonesia until today. Due to the fact that the significant mobility of the people, Central Java has contributed the 3rd rank of potential number of COVID-19 positive cases in Indonesia. This study aims to forecast the number of COVID-19 daily new cases in Central Java to assist the government in preparing the necessary resources and controlling the spread of the COVID-19 virus in Central Java Province. We proposed Exponential Smoothing Holt-Winters with the Additive model with seasonal addition considering trend and seasonal factors. The dataset during March 14 to April 17, 2021, revealed fluctuation of trend and seasonal patterns. Our simulation studies indicate that Exponential Smoothing Holt-Winters provides sharp and well performance for forecasting daily new cases of COVID-19 in Central Java province with MAPE less than 10%.

Keywords: Additive, COVID-19, Forecasting, Holt-Winters, Seasonal

Abstrak

Pandemi COVID-19 masih belum dapat dipastikan kapan akan berhenti. Ada beberapa faktor, antara lain upaya masyarakat untuk memperlambat penyebaran dan upaya para peneliti untuk mengamati lebih jauh tentang wabah ini. Sejak awal krisis kesehatan, terutama setelah diumumkannya kasus positif pertama di Indonesia akibat COVID-19 pada 2 Maret 2020, jumlah kasus harian meningkat ecara simultan di wilayah lain di Indonesia hingga saat ini. Karena mobilitas penduduk yang signifikan, Jawa Tengah menempati urutan ke-3 terbanyak jumlah kasus positif COVID-19 di Indonesia. Penelitian ini bertujuan untuk meramalkan jumlah kasus baru harian COVID-19 di Jawa Tengah untuk membantu pemerintah dalam mempersiapkan sumber daya yang diperlukan dan pengendalian penyebaran virus COVID-19 di Provinsi Jawa Tengah. Kami menggunakan Exponential Smoothing Holt-Winters dengan pendekatan aditif dengan penambahan musiman dengan mempertimbangkan faktor trend dan musiman. Dataset selama 14 Maret 2021 hingga 17 April 2021 memiliki fluktuasi pola trend dan musiman. Studi simulasi pada paper ini menunjukkan bahwa model Exponential Smoothing Holt-Winters memiliki performansi yang baik untuk peramalan kasus baru COVID-19 harian di Provinsi Jawa Tengah dengan nilai MAPE di bawah 10%.

Kata Kunci: Additive, COVID-19, Holt-Winters, Musiman, Peramalan

I. INTRODUCTION

OVID-19 is a disease caused by a new coronavirus that was not previously identified in humans. There is hard to mention how long the COVID-19 pandemic will discontinue. There are some factors, including the public's efforts to slow spread and researcher's work to observe more about this outbreak. From the beginning of the health crisis, particularly following the announcement of the first positive case In Indonesia due to the COVID-19 on March 2, 2020. Afterwards, the number of daily cases increase simultaneously in other regions in Indonesia until today. For some people, the symptoms of this virus can cause pneumonia or difficulty breathing and can lead to someone's death [1]. In Indonesia, the first and second confirmed cases of COVID-19 were announced by President Joko Widodo on March 2, 2020, [2]. Due to the fact that the significant mobility of the people, Central Java has contributed the $3rd$ rank of potential number of COVID-19 positive cases in Indonesia according to official website kawalcovid19.id. **C**

Meanwhile, time series analysis is an analysis of a series of data that is sequential in time in the past and is useful for forecasting in the future [3] [3]. COVID-19 data is quantitative data that has a pattern of relationships between observation variables and time variables and the data is sorted in time. Therefore, that COVID-19 data can be carried out in time series research such as forecasting. The forecasting of COVID-19 in the Gorontalo region from April 10, 2020 – October 13, 2020 (especially total cases) using the Holt-Winters Exponential Smoothing Method which was studied by I Djakaria and S E Saleh [4]. The results of the forecasting have a MAPE value of 6.14% and the value is below 10%, which means that the forecasting competence using the Holt-Winters Exponential Smoothing method is very good [5]. Hansen Wiguna, et al [6] proposed time series model in analyzing and predicting the spread of COVID-19 in Jakarta using the ARIMA method. The MAPE value obtained is 20.97% so that the predictive competence is said to be feasible, since the MAPE value ranges from 20-50% [5]. The Exponential Smoothing Holt-Winters method is used to overcome time series data that have trend patterns and seasonal patterns [7]. Evelina Padang, et al [8] used the Exponential Smoothing Holt-Winters method in forecasting the number of passengers on the Medan-Rantau train. The data used is data on the number of train passengers for the period January 2007-December 2011 using the Holt-Winters model and the MAPE values obtained are 0.81 and 0.6 so that the forecasting results have very good competence [5]. Also, Tias Safitri, et al [9] compared the Exponential Smoothing Holt-Winters and ARIMA methods in forecasting the number of foreign tourist visits to Bali Ngurah Rai. Based on the results of the comparison of MAPE values, the Exponential Smoothing Holt-Winters method is the better method because it has a MAPE value of 8.86198% while ARIMA has a MAPE value of 9.40981%. A comparison of forecasting Moving Average and Exponential Smoothing Holt-Winters methods by Yuseva Rismawanti and Moh Yamin Darsyah [10], is also carried out to determine inflation forecasting in Indonesia. Based on the comparison result, the Holt-Winters Exponential Smoothing method is also better than the Moving Average method.

The number of daily COVID-19 cases reveal fluctuations, several seasonal and noise patterns. As the significant severity of this pandemic, estimating the future number of daily COVID-19 become a major concern to support information and maintain essential health services. Central Java is the province with the third highest population density in Indonesia and predicted to be affected significantly over a particular period of this outbreak. It generally accepted that forecasting of time series observation is involving temporal dimension. Thus, we consider to apply Exponential Smoothing Holt-Winters to reveal the possibly trend pattern and seasonal pattern of the dataset of COVID-19 daily new cases. This paper is organized in the following sections: the first about the framework of Exponential Smoothing Holt-Winters and and the accuracy with MAPE. Next section is detailed procedure of forecasting based Exponential Smoothing Holt-Winters, we estimate the parameters of Holt-Winters with additive model. In the third section, we compare the MAPE corresponding to scenarios of splitting data. And, the last section is conclusion and proposing possibly future work.

II. LITERATURE REVIEW

A. Exponential Smoothing Holt-Winters

Periodic data (time series) is data based on time sequences or data collected from time to time and the data used can be in the form of daily, weeks, months, and years [11]. Time series data is data that is commonly used to determine the development of a situation and can predict the value of time series data at a certain period in the future. In research [12] explained that the most important component in choosing an appropriate forecasting method for a time series data is a pattern. There are four types of data patterns, namely trend patterns, seasonal patterns, cyclical patterns, and horizontal patterns. Trend patterns are time series data that experience growth or decline over several periods. Seasonal patterns are patterns that are influenced by seasonal factors such as weeks, years, months. A cyclical pattern is data that increases and decreases from a fixed period. Horizontal pattern is data whose average value is constant. The Holt-Winters method is used by data that has trend patterns and seasonal patterns. This method is often called the exponential smoothing method which approaches [13]. This method uses three important parameters, namely the overall smoothing of the data (α) , trend smoothing (β) , and seasonal smoothing (γ) , where the parameter value is 0 to 1 [14]. The Exponential Smoothing Holt-Winters method is divided into two models, namely Exponential Smoothing Holt-Winters with Additive model and Exponential Smoothing Holt-Winters with Multiplicative model.

1) Exponential Smoothing Holt-Winters Additive

Exponential Smoothing Holt-Winters with Additive model is a seasonal addition method. This model is used on constant seasonal data variation [13]. These following are the equation of Exponential Smoothing Holt-Winters Additive [12].

(1) Smoothing level

$$
L_t = \alpha(y_t - S_{t-s}) + (1 - a)(L_{t-1} + b_{t-1})
$$
\n(1)

(2) Smoothing trend pattern

$$
b_t = \beta (L_t - L_{t-1}) + (1 - \beta) b_{t-1}
$$
 (2)

(3) Smoothing seasonal pattern

$$
S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-s}
$$
\n(3)

(4) Forecast for the next period

$$
F_{t+m} = L_t + b_t m + S_{t-s+m} \tag{4}
$$

where

2) Exponential Smoothing Holt-Winters Multiplicative

Exponential Smoothing Holt-Winters with the Multiplicative model is a seasonal multiplication model. This model is used on seasonal data that experiences an increase or decrease volatility. These following are the equation of Exponential Smoothing Holt-Winters Multiplicative [12].

Irandi et al. FORECASTING NUMBER OF NEW CASES DAILY... 26

(1) Smoothing level

$$
L_t = \alpha \frac{y_t}{s_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1})
$$
\n(5)

(2) Smoothing trend

$$
b_t = \beta (L_t - L_{t-1}) + (1 - \beta) b_{t-1}
$$
 (6)

(3) Smoothing seasonal

$$
S_t = \gamma \frac{y_t}{l_t} + (1 - \gamma) S_{t-s} \tag{7}
$$

(4) Forecast for the next period

$$
F_{t+m} = (L_t + b_t m)S_{t-s+m} \tag{8}
$$

The early process is initializing the initial value of the parameter. It is essential to observe the period and capturing seasonal period. In initializing the parameter values of the Exponential Smoothing Holt-Winters, there are the following formulas:

(1) Estimation of smoothing level

$$
L_s = \frac{1}{s}(y_1 + y_2 + \dots + y_s)
$$
\n(9)

(2) Estimation of smoothing trend

$$
b_{s} = \frac{1}{s} \left(\frac{y_{s+1} - y_1}{s} + \frac{y_{s+2} - y_2}{s} + \dots + \frac{y_{s+s} - y_s}{s} \right) \tag{10}
$$

(3) Estimation of smoothing the seasonal pattern Additive model

$$
S_k = (y_k - L_s) \tag{11}
$$

(4) Estimation of smoothing the seasonal pattern Multiplicative model

$$
S_k = \frac{y_k}{L_s} \tag{12}
$$

B. Mean Absolute Percent Error (MAPE)

In this research, the calculation of the accuracy value uses the MAPE value, where the forecast is in the form of a percentage. MAPE is a calculation used to calculate the average absolute error percentage [15]. In measuring the MAPE value, the percentage of MAPE corresponding to system that well-performance if it is below 10%, MAPE values with a good model are between 10-20%, MAPE with a range of 20%-50% are included in the appropriate or adequate category, and MAPE which is more than 50% included in the bad category [8]. The following is the equation of MAPE [16].

$$
MAPE = \frac{\sum_{actual-foreast1}^{|actual}\times100}{n}
$$
 (13)

III. RESEARCH METHOD

We present an overview of design system to forecast the new cases of daily COVID-19 cases in Central Java between March and April 2021 in Figure 1 considering seasonal period that captured during those period. We consider to apply Exponential Smoothing Holt-Winters to reveal the possibly trend and seasonal pattern of the dataset of COVID-19 daily new cases. Moreover, to capture unprecedented volatility of daily COVID-19 cases to support maintain essential health services. We propose three scenarios of splitting data in our simulation and compare the MAPE for each.

Fig. 1. Flowchart System of Forecasting Using Exponential Smoothing Holt-Winters Model

A. Dataset

The dataset used for forecasting new cases of COVID-19 in Central Java Province were obtained from the official website kawalcovid19.id. In this study, corresponding to seasonal pattern in Holt-Winters approach, we observe that the new cases during March 14 to April 17, 2021 captured this pattern. We split data into three scenarios of testing and training data, Figure 2. As initial process, we extract the dataset according to three main components: trend, seasonal, and noise. Observing those patterns help to choose among additive and multiplicative approach. Based on Figure 3, there is an upward trend pattern and a constant seasonal pattern so that the model used in this study is the Holt-Winters Smoothing Exponential Additive Model. The data has 7 seasonal length, the seasonal periodic repeats every 7 days. Therefore, the ratio in each selected scenario has adjusted to the length of the season where the maximum testing data is 7 data.

Fig. 2. Plot Data of Daily New Cases of COVID-19 in Central Java Province

B. Initialize the Parameter Value

Based on Figure 3 on the time series decomposition, the data has an upward trend pattern and a constant seasonal pattern so that the forecasting model used is the Holt-Winters Additive model. Time series decomposition can identify variables by separating time series data so that levels, trends, and seasonality can be found in the data. The stage in forecasting the Holt-Winters Additive model is to get the initial value of smoothing level (L_s) , initial value of trend pattern smoothing (b_s) , and the initial value of seasonal pattern smoothing (S_k) . The following is the result of the calculation of the initial smoothing level value using Equation 9, 10 and 11.

$$
L_s = \frac{1}{7}(144 + 700 + 560 + 647 + 658 + 611 + 219) = 505.57
$$

\n
$$
b_s = \frac{1}{7}\left(\frac{102 - 144}{7} + \frac{741 - 700}{7} + \frac{732 - 560}{7} + \frac{741 - 700}{7} \dots + \frac{y_{s+7} - y_7}{7}\right) = -13.3
$$

\n
$$
S_1 = -361,57, S_2 = 194,43, S_3 = 54,43, S_4 = 141,43, S_5 = 152,43, S_6 = 105,43, S_7 = -286,57.
$$

The result of the initial value of the smoothing level(L_s), trend pattern smoothing (b_s), and seasonal pattern smoothing (S_k) used to obtain the overall value of the smoothing level (L_t) , smoothing trend pattern (b_t) , and smoothing seasonal pattern (S_t) .

Fig. 3. Decomposing Pattern of New Cases COVID-19 Dataset

C. Parameter Estimation α , β , γ

In this process, parameter values are obtained from smoothing level (α) , smoothing trend pattern (β) , and smoothing seasonal pattern (y) by inputting random values with a range of 0 to 1 for each parameter and these values are used to obtain optimal values for each parameter. The procedure for obtaining the optimal parameters of smoothing level (α) , smoothing trend pattern (β) , and seasonal smoothing pattern is using Nonlinear GRG technique. The optimal parameters in Table I are considering to calculate the overall value of smoothing levels (L_t) , smoothing trend patterns (b_t) , and smoothing seasonal patterns (S_t) . The formula to determine the overall smoothing of level value (L_t) , trend pattern (b_t) , and seasonal pattern (S_t) are based on Equation 1, 2 and 3, respectively.

TABLE I HOLT-WINTERS ADDITIVE PARAMETER DETERMINATION

	Data	Data	Parameter Estimation		
Scenarios	Training	Testing	α		
80:20			0.09	0.44	0.22
83:17			0.19	0.35	0,36
86:14).09	0.53	0.69

IV. RESULTS AND DISCUSSION

Forecasting daily new cases for COVID-19 uses the Holt-Winters Additive model using the calculations described in Equation 4. We employ training data to construct the Holt-Winters model for each scenario, and analyze the model performance using testing dataset based on error percentage, MAPE.

TABLE III HOLT-WINTERS ADDITIVE MODEL PREDICTION RESULT

Scenarios	Time Period (daily)	Actual Data	Forecast (F_{t+m})
80:20	$2021 - 04 - 11$	205	205

Based on the forecasting results in Table III, the scenario (80:20) has 7 of testing data with the highest new case forecasting results on April 15, 2021, which is 843 cases and the lowest on April 11, 2021, which is 205 cases. The scenario (83:17) has 6 testing data with the highest forecasting results on April 15, 2021, which is 879 cases and the lowest on April 17, 2021, which is 461 cases. The scenario (86:14) has 5 testing data with the highest new case forecasting results on April 15, 2021, which is 910 cases and the lowest on April 17, 2021, which is 477 cases. Figure 5 is a plot of the testing data and the forecasting results of each test scenario. Corresponding to Figure 5, from the three test scenarios, on April 15, 2021, there was an increase in daily new cases and the forecasting results for that period were the highest cases. From April 16, 2021 to April 17, 2021, from the forecasting results, the number of new cases decreased. The next process is to calculate the MAPE accuracy value to see the performance of the Holt-Winters Additive model and compare which scenario has the lowest MAPE value. The calculation of MAPE accuracy is using Equation 13 and the results are shown in Table IV.

Fig. 4. Plot of Daily New Case Forecasting Results from the Three Scenarios

Scenarios	MAPE
80:20	4,90%
83:17	3.94%
86:14	4.13%

TABLE IV MAPE ACCURACY VALUE

In Table IV, the MAPE accuracy values from the three scenarios all show results below 10%. According to the results of the MAPE values from the three scenarios, the smallest MAPE value is in the data scenario (83:17) with a total of 29 training data in the range of March 14, 2021 to April 11, 2021 and 6 testing data in the range of April 12, 2021 to April 17, 2021. From the data scenario (83:17), the MAPE value obtained is 3.94% and after being compared with other scenarios, this scenario revealed the well perfomrance in forecasting daily new cases of COVID-19 using the Exponential Smoothing Holt-Winters Additive model. Forecasting results from the data scenario (83:17) show that the highest cases is on April 15, reported 879 cases and the lowest case point to April 17 with 461 cases. Our simulation studies indicate that the Exponential Smoothing Holt-Winters provides sharp and well estimation for forecasting daily new cases of COVID-19 in Central Java province. MAPE value that reported in this study is smaller, 3.94%, while the MAPE value in research I Djakaria and SE Saleh [4] is 6.14%. The method in this study also has a better performance than the research by Hansen Wiguna, et al [6] which uses the ARIMA method in analyzing and predicting the spread of COVID-19 in Jakarta.

V. CONCLUSION

Due to the fact that the significant mobility of the people, Central Java has contributed the 3rd rank of potential number of COVID-19 positive cases in Indonesia. From the beginning of the health crisis, particularly following the announcement of the first positive case In Indonesia due to the COVID-19 on March 2, 2020. Afterwards, the number of daily cases increase simultaneously in other regions in Indonesia until today. It generally accepted that forecasting of time series observation is involving temporal dimension. Thus, we consider to apply Exponential Smoothing Holt-Winters to reveal the trend and seasonal pattern the dataset of COVID-19 daily new cases. We observe that the new cases during March 14 to April 17, 2021 captured this pattern. Our simulation studies indicate that Exponential Smoothing Holt-Winters provides sharp and well estimation for forecasting daily new cases of COVID-19 in Central Java province, the MAPE based on three scenarios of splitting data reported small, 3.94%. This result may help the authorities and public health services in preparing the necessary resources and controlling the spread of the COVID-19 virus in Central Java Province. Therefore, under the high volatility of this outbreak, finding an appropriate model to estimate upper prediction, such assessing upper limit prediction is desirable and should be considered.

REFERENCES

- [1] W. H. Organization, "Materi Komunikasi Risiko COVID-19 untuk Fasilitas Pelayanan Kesehatan," 2 March 2020. [Online]. Available: https://www.who.int/docs/defaultsource/searo/indonesia/covid19/risk-communication-for-healthcare-facility.pdf?sfvrsn=9207787a_2. [Accessed 9 June 2021].
- [2] A. Almuttaqi, "Kekacauan Respons terhadap COVID-19 di Indonesia," *THC INSIGHTS,* pp. 1-7, 23 March 2021.
- [3] A. Triangga, "Analisis Curah Hujan dengan Pemodelan Deret Waktu pada Das Walanae," p. 23, 2020.
- [4] I. Djakaria and S. Saleh, "Covid-19 forecast using Holt-Winters exponential smoothing," *Journal of Physics,* vol. 1882, 2021.
- [5] I. Nabillah and I. Ranggadara, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," *Journal of Information System,* vol. 5, no. 2, p. 254, 2020.
- [6] H. Wiguna, Y. Nugraha, F. R. R, A. Andika, J. I. Kanggrawan and A. L. Suherman, "Kebijakan Berbasis Data: Analisis dan Prediksi Penyebaran COVID-19 di Jakarta dengan Metode Autoregressive Integrated Moving Average (ARIMA)," *Jurnal Sistem Cerdas,* vol. 03, no. 02 02, pp. 74-83, 2020.
- [7] D. A. W. S. Putra, K. D. Hartomo and R. Tanone, "Model Prediksi Kekeringan Menggunakan Metode Holt-Winters (Studi Kasus : Wilayah Kabupaten Boyolali)," *Journal of Modeling and Computing 1,* p. 38, 2018.
- [8] E. Padang, G. Tarigan and U. Sinulingga, "PERAMALAN JUMLAH PENUMPANG KERETA API MEDAN-RANTAU PRAPAT DENGAN METODE PEMULUSAN EKSPONENSIAL HOLT-WINTERS," *Journal of Saintia Matematika,* vol. 1, no. 2, pp. 161-174, 2013.
- [9] T. Safitri, N. Dwidayati and S. , "PERBANDINGAN PERAMALAN MENGGUNAKAN METODE EXPONENTIAL SMOOTHING HOLT-WINTERS DAN ARIMA," *Journal of Mathematics,* vol. 6, no. 1, pp. 48-58, 2017.
- [10] Y. Rismawanti and M. Y. Darsyah, "Perbandingan Peramalan Metode Moving Average dan Exponential Smoothing Holt Winter Untuk Menentukan Peramalan Inflasi di Indonesia," in *Prosiding Seminar Nasional Mahasiswa Unimus*, 2018.
- [11] A. A. Anwary, "Prediksi Kurs Rupiah Terhadap Dollar Amerika Menggunakan Metode Fuzzy Time Series," p. 5, 2011.
- [12] P. J. Kristianti, PENERAPAN METODE HOLT-WINTERS UNTUK PERAMALAN TINGKAT INFLASI DI INDONESIA, 2020.
- [13] G. Pongdatu , E. Abinowi and W. , "PERAMALAN TRANSAKSI PENJUALAN DENGAN METODE HOLT-WINTER'S EXPONENTIAL SMOOTHING," *Jurnal Ilmiah Teknologi Informasi Terapan,* vol. 6, no. 3, pp. 228-233, 2020.
- [14] N. Parwati, PRAKIRAN JUMLAH PENUMPANG MENGGUNAKAN HOLT-WINTER'S EXPONENTIAL SMOOTHING, 2020.
- [15] M. A. Maricar, "Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ," *Jurnal Sistem dan Informatika,* vol. 13, no. 2, pp. 36-45, 2019.
- [16] H. Tannady, R. Mulyadi and R. Cahyadi, "PENENTUAN JENIS INVESTASI DENGAN ANALISA EKONOMI TEKNIK DAN FORECASTING," *Jurnal Ilmiah Teknik Industri,* vol. 13, no. 2, pp. 134-140, 2014.