TECHNIUM
SOCIAL SCIENCES JOURNAL

Vol. 16, 2021

A new decade for social changes

www.techniumscience.com

ISSN 2668-7798
9 772668 779000

Technium Social Sciences Journal
Vol. 16, 618-630, February, 2021
ISSN: 2668-7798
www.techniumscience.com

Forecasting Method Using the Minitab Program

Sunarsan Sitohang[1], Very Karnadi[2]

[1] [2]Putera Batam University of Informatic Enginering, Batam, Indonesia

Sunarsan@puterabatam.ac.id[1]

Abstract. The Corona Virus 2019 (COVID-19) pandemic has yet to subside. This epidemic has spread to almost all countries in the world. This pandemic has resulted in the decrease of the activities of people and the economy. The COVID-19 pandemic itself spread in Indonesia on March 2, 2020, to be precise. Two people tested positive for COVID-19, and they were referred to as case 1 and 2. After the detection of the COVID-19 pandemic in Indonesia, Indonesia experienced additional positive cases of COVID-19 every day. This study aims - to build a model to predict the development of COVID-19 cases based on time series data and the number of COVID-19 sufferers, using seven forecasting methods, namely, Exponential Smoothing, Exponential Smoothing with Trend, Linear Regression/Least Squares, Moving Average, Trend Analysis (Regression Over Time), Additive Decomposition (Seasonal), and Multiplicative Decomposition (Seasonal), with the Minitab program, including in Indonesia. The results showed that the Exponential Smoothing with Trend method, has mean absolute deviation (MAD) of 220.6and mean squared error (MSE) of 69,994.2; the forecast resulted in 177,621 number of COVID-19 cases, which is predicted to occur on September 1, 2020.

Keywords. forecasting, Minitab, seven method, COVID-19

1 Preliminary

1.1 Background

The world is currently hit by a very serious outbreak of Corona Virus 2019 or COVID-19. The COVID-19 pandemic has spread to almost all countries in the world [1]. As it spreads from person to person, this pandemic spreads very quickly [2]. So far, no drugs or vaccines that can be used to overcome the spread of the COVID-19 pandemic have been found [3]. The community is advised to stay at home and to not leave the house except in emergency situations. The COVID-19 pandemic spread in Indonesia at the beginning of March precisely on March 2, 2020, in which two people tested positive for COVID-19; they were called case 1 and 2. After the detection of these two cases, there were additional positive cases of COVID-19 every day. The number of additional cases each day cannot be estimated so subscribers such as service providers, facilities, and medical personnel cannot be predicted. From these cases, we conducted research using forecasting methods with the Minitab program to obtain a model in predicting the number of new cases of COVID-19 in Indonesia.

The use of forecasting methods in business is increasingly growing [4]. In actuality, we indirectly use forecasting every day, for example, in predicting our goals, predicting business sales, and projecting our business income[5]. There are many forecasting methods that can be used, including the QM for Windows program and manual forecast,, one of which is with the help of the Minitab program [6]. The process of using the forecasting method can be completed if it is supported by accurate historical data and the more the historical data over time, the better and closer the correctness of the forecasting results [7]. In this study, data from the website www.covid19.go.id were used, with historical data for 31 days, starting from August 1, 2020 to August 31, 2020.

1.2 Identification of Problems

From the description above, the problems that will be resolved in this study are as follows: 1) How big is the forecast for the number of sufferers infected with COVID19 using the Minitab program? 2) Of the seven Minitab forecasting methods used, which method has the smallest mean absolute deviation (MAD) and mean squared error (MSE)?

1.3 Research Purposes

This study aim to answer the problems identified above: how much is the forecast for the number of sufferers infected with COVID-19 with the Minitab program and which of the seven forecasting methods with the Minitab program has the smallest MAD and MSE.

2 Theoretical Basis

2.1 The COVID-19 Pandemic

At the end of December 2019, Corona Virus 2019 or COVID-19 first appeared in Wuhan City, Hubei Province, China [8]. On March 11, 2020, the World Health Organization (WHO) declared the 2019 novel coronavirus a global pandemic [8]. A person infected with this virus will show mild symptoms such as flu, making it difficult for this virus to be detected. The COVID-19 virus attacks the respiratory tract, causing the sufferer to have difficulty breathing. Until now, vaccines and drugs that prevent the spread of COVID-19 have not been found [9]. The only way to reduce the spread of this pandemic is to stay at home, not be in a crowd, diligently wash your hands before and after doing activities, and always use a mask [10]. In Indonesia, the number of COVID-19 cases continues to increase and it is uncertain when the COVID-19 pandemic will end.

2.2 Time Series Data

Time series data are data in sequential order at different times [11]. One example of time series data that can be used to make predictions is data that is sequentially presented per day [12]. This research is based on several studies conducted using time series data. Research conducted by [13] uses a statistical model to test the time series database. Research by [14] uses a hybrid algorithm in which the Boltzmann engine and the backpropagation method were combined to produce a model; the results obtained in this study are promising. Research conducted by [15] aimed to predict confirmed positive and recovered COVID-19 cases using the TP-SMN-AR time series model. The study results indicate that the proposed method can make predictions in confirming positive and recovered COVID-19 cases well in a number of countries in the world. The research was carried out with the aim of predicting the current trends and timing of stopping the COVID-19 outbreak in Canada and around the world. In this study [16] short-term memory network (LSTM) was used. Based on the results of the LSTM, it is estimated that this outbreak will stop around the end of June 2020 based on data from the start of the COVID-19 cases. Research conducted in Italy, Spain, and France, the most influential countries in Europe, aimed at predicting the epidemiological trend of the prevalence of COVID-19 using time series data in the form of COVID-19 prevalence data from February 21, 2020, to April 15, 2020 collected from the WHO website using the Autoregressive Integrated Moving Average model. The study results can explain the trend of the outbreak and provide an overview of the epidemiological stage of the area. In addition, predictions of COVID-19 prevalence trends in Italy, Spain and France can help take preventive measures and formulate policies for this epidemic in other countries.

2.3 Minitab
Minitab is a software package for statistics and is similar to SPSS, SAS, and several other statistical software packages [4].

2.4 Forecasting
According to [17], forecasting is the art and science of future event prediction in which collected past data are used and placed into the future with a systematic model [18]. According to [19], forecasting is also the initial part of a decision-making process.

2.4.1 Quantitative Forecasting Divisions
According to [20], quantitative forecasting is divided into the following:
1. Time series
2. Causal method

2.4.2 Calculating Forecasting Errors
According to [21], there are two formulas for calculating forecast errors: MAD is the first measure of the overall forecast error for a model. This value is calculated by taking the absolute value of the forecast error in the number of data periods (n).
MSE is a second way of measuring the overall forecast error. MSE is the average of the difference in squares between the predicted and observed values. The disadvantage of using MSE is that it tends to accentuate large deviations due to squares.

3 Research Methodology
In this study, the Minitab program was used. [22] and [23] state that the general research flow and those carried out in this study are as follows:

1. Literature Review
In the early stages of working on this research, we gathered information about COVID-19, as well as appropriate methods for modeling. Information was gathered from journals, books, and related websites.

2. Data Search
In the next stage, we looked for COVID-19 data, namely, data on the increasing number of patients in Indonesia per day.

3. Observation Method
After data collection, we studied the methods that can be used to create models and make predictions. At this stage, we also observed the development of a method to determine the learning rate and momentum in the backpropagation method using the Tsukamoto fuzzy method.

4. Implementing the System
In the last stage, implementation was carried out by applying the backpropagation method and the Tsukamoto fuzzy method to the COVID-19 data to obtain a prediction model for COVID-19 cases in Indonesia. Figure 1 shows the general research flow.

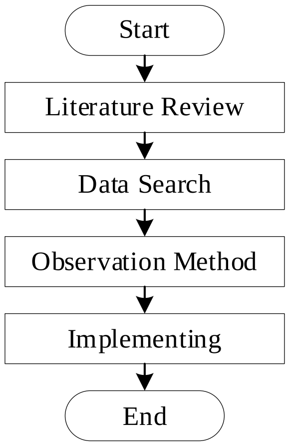

Fig. 1. General research flow [24]

Table 1 shows a summary of the research methodology used in conducting this research.

Table 1. Research methodology

1	Forecasting	Predictions of patients with COVID-19 in September 2020	Data on patients infected with COVID19 in Indonesia from August 1, 2020 to August 31, 2020
2	Forecasting model	The prediction method that will be applied in this research is Indonesian.	Exponential Smoothing, Exponential Smoothing with Trend, Linear Regression/Least Squares, Moving Average, Trend Analysis (Regression Over Time), Additive Decomposition (Seasonal), and Multiplicative Decomposition (Seasonal) [25]
3	Forecasting results	The smallest measure of accuracy for forecasting	MAD and MSE results

4 Results and Discussion

Table 2 shows the number of patients infected with COVID-19 in Indonesia from August 1, 2020, to August 31, 2020.

Table 2. Number of COVID-19 sufferers in Indonesia in August 2020

Date	Number of	Date	Number of patients
01/08/20	109.936	17/08/20	141.370
02/08/20	111.455	18/08/20	143.043
03/08/20	113.134	19/08/20	144.945
04/08/20	115.056	20/08/20	147.211
05/08/20	116.871	21/08/20	149.408
06/08/20	118.753	22/08/20	151.498
07/08/20	121.226	23/08/20	153.535
08/08/20	123.503	24/08/20	155.412
09/08/20	125.396	25/08/20	157.859
10/08/20	127.083	26/08/20	160.165

11/08/20	128.776	27/08/20	162.884
12/08/20	130.718	28/08/20	165.887
13/08/20	132.816	29/08/20	169.195
14/08/20	135.123	30/08/20	172.053
15/08/20	137.468	31/08/20	174.796
16/08/20	139.549		

Source: [26]

To calculate the forecast, if we use the manual approach or the QM for Windows program, we can solve with 11 methods. However, if we use the Minitab program, according to the author it can only be solved by seven methods, namely, Exponential Smoothing, Exponential Smoothing with Trend, Linear Regression/Least Squares, Moving Average, Trend Analysis (Regression Over Time), Additive Decomposition (Seasonal), and Multiplicative Decomposition (Seasonal).

1. Exponential Smoothing

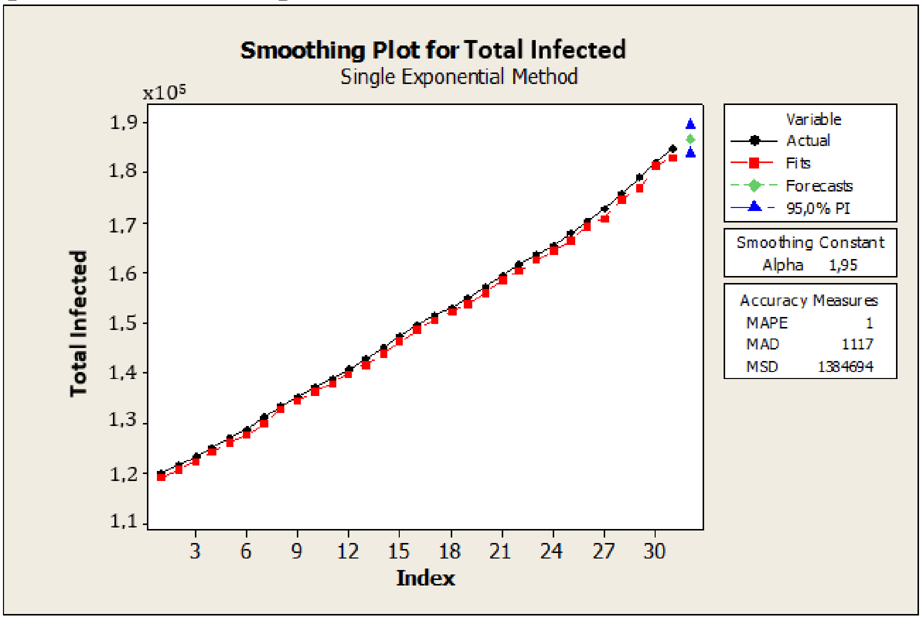

Fig 2. Single Exponential Smoothing

Single Exponential Smoothing for the Total Number of People Infected with COVID-19

Data total infected
Length 31
Smoothing Constant
Alpha 1,95
Accuracy Measures
MAPE 1
MAD 1117
MSD 1384694

Time	Total infected	Smooth	Predict	Error
1	109936	110653	109182	754,33
2	111455	112217	110653	802,39
3	113134	114005	112217	916,73

Technium Social Sciences Journal
Vol. 16, 618-630, February, 2021
ISSN: 2668-7798
www.techniumscience.com

⋮

29	169195	171312	166967	2228,10
30	172053	172757	171312	741,31
31	174796	176733	172757	2038,76

Forecasts

Period	Forecast	Lower	Upper
32	176733	173995	179471

2. Exponential Smoothing with Trend

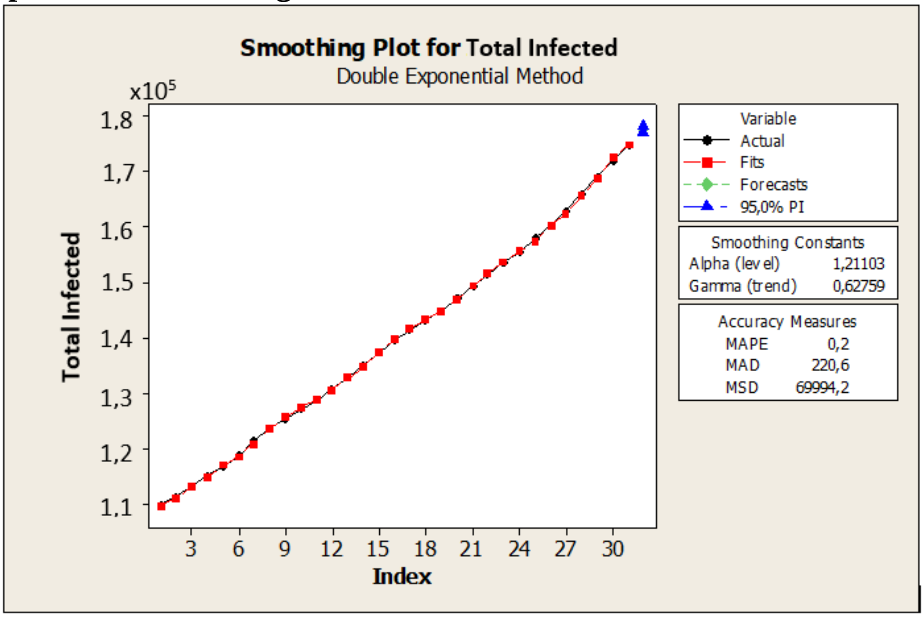

Fig 3. Exponential Smoothing with Trend

Double Exponential Smoothing for Total Infected

Data Total infected
Length 31
Smoothing Constants
Alpha (level) 1,21103
Gamma (trend) 0,62759

Accuracy Measures
MAPE 0,2
MAD 220,6
MSD 69994,2

Time	Total infected	Smooth	Predict	Error
1	109936	109978	109736	200,430
2	111455	111518	111158	296,960
3	113134	113166	112983	150,855
⋮				
29	169195	169284	168772	423,460
30	172053	171972	172435	−381,769
31	174796	174788	174833	−36,687

Forecasts

Period	Forecast	Lower	Upper
32	177621	177080	178161

3. Linear Regression / Least Squares

Regression Analysis: Total Number of Infected Versus Period

The regression equation is

Total Infected = 106107 + 2110 Periode

Predictor	Coef	SE Coef	T	P
Constant	106107	433	245,13	0,000
Periode	2110,48	23,61	89,37	0,000

S = 1175,99   R-Sq = 99,6%   R-Sq(adj) = 99,6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	11046243034	11046243034	7987,44	0,000
Residual Error	29	40105617	1382952		
Total	30	11086348651			

Unusual Observations

Obs	Periode	Total infected	Fit	SE Fit	Residual	St Resid
30	30,0	172053	169422	392	2631	2,37R
31	31,0	174796	171532	412	3264	2,96R

R denotes an observation with a large standardized residual.

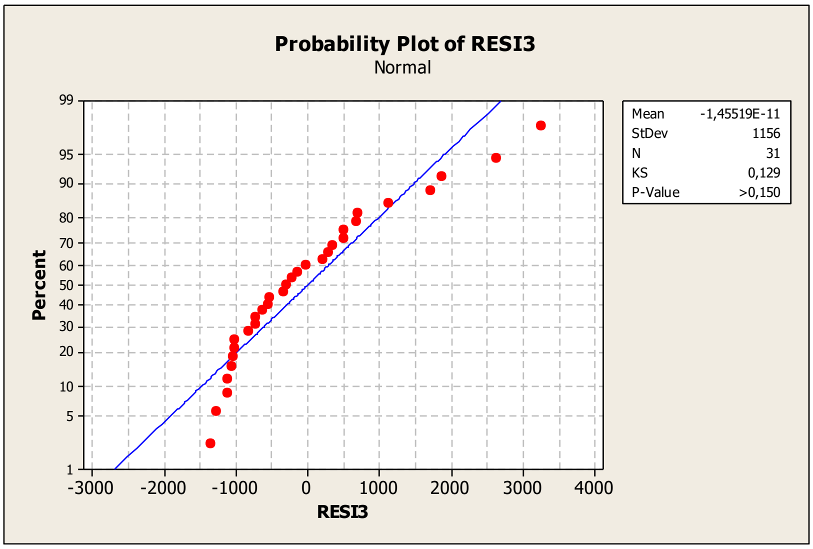

Fig. 4. Linear regression

4. Moving Average

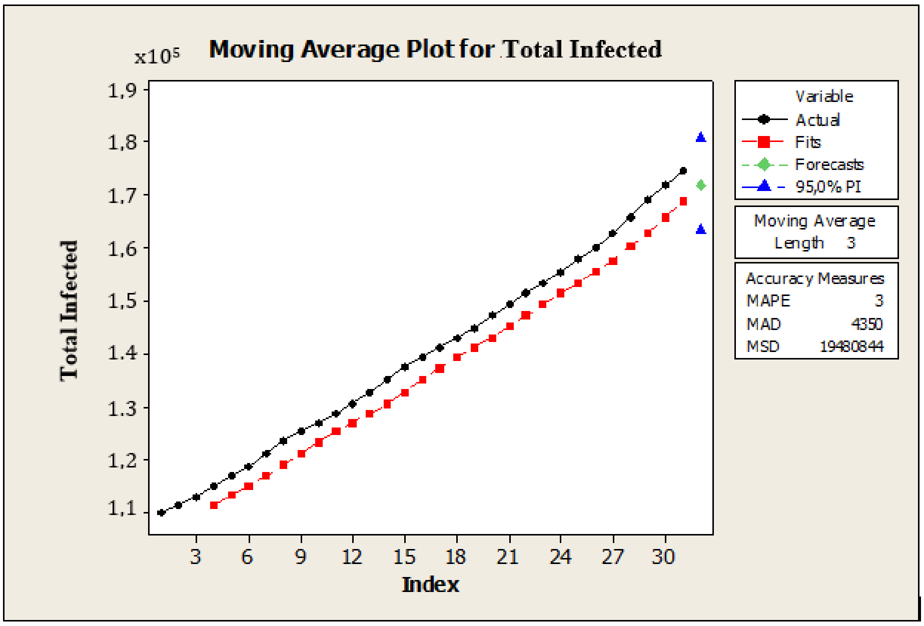

Fig 5. Moving Average

Moving Average for Total Infected

Data Total Infected
Length 31
NMissing 0
Moving Average
Length 3
Accuracy Measures
MAPE 3
MAD 4350
MSD 19480844

Time	Total infected	MA	Predict	Error
1	109936	*	*	*
2	111455	*	*	*
3	113134	111508	*	*
⋮				
29	169195	165989	162979	6216,33
30	172053	169045	165989	6064,33
31	174796	172015	169045	5751,00

Forecasts

Period	Forecast	Lower	Upper
32	172015	163364	180665

5. Trend Analysis (Regression Over Time)

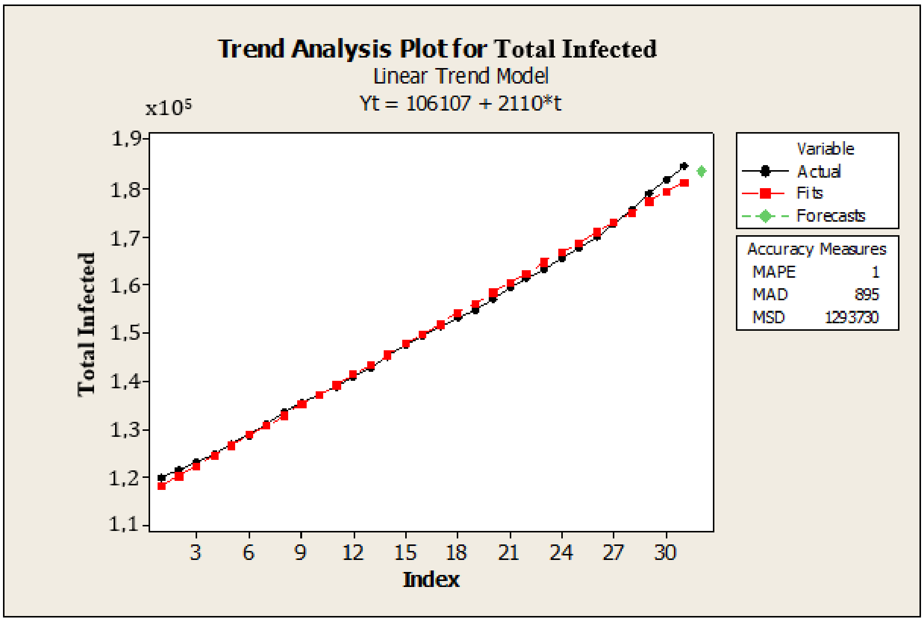

Fig 6. Trend Analysis (Regression Over Time)

Trend Analysis (Regression Over Time)

Data Jumlah Terinfeksi
Length 31
NMissing 0
Fitted Trend Equation
Yt = 106107 + 2110*t
Accuracy Measures
MAPE 1
MAD 895
MSD 1293730

Time	Total infected	Trend	Detrend
1	109936	108218	1718,25
2	111455	110328	1126,77
3	113134	112439	695,29
⋮			
29	169195	167311	1883,78
30	172053	169422	2631,30
31	174796	171532	3263,82

Forecasts

Period	Forecast
32	173643

TECHNIUM SOCIAL SCIENCES JOURNAL

Technium Social Sciences Journal
Vol. 16, 618-630, February, 2021
ISSN: 2668-7798
www.techniumscience.com

6. Additive Decomposition (Seasonal)

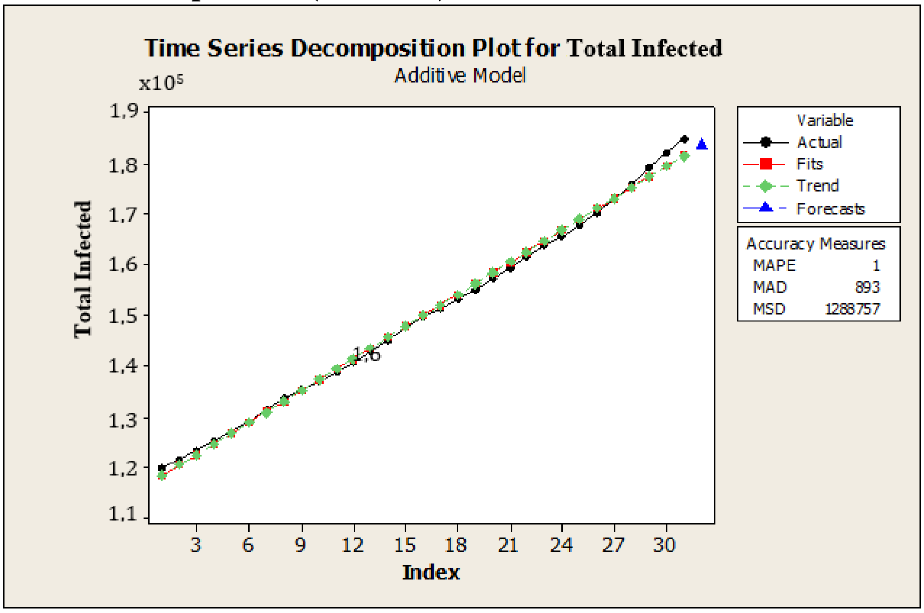

Fig 7. Time series decomposition

Time Series Decomposition for Total Number of Infected

Additive Model
Data Total Infected
Length 31
NMissing 0
Fitted Trend Equation
Yt = 106104 + 2111*t
Seasonal Indices

Period	Index
1	31,4444
2	18,9444
3	−50,3889

Accuracy Measures

MAPE	1
MAD	893
MSD	1288757

Time	Total infected	Trend	Seasonal	Detrend	Deseason	Predict	Error
1	109936	108215	31,4444	1721,36	109905	108246	1689,91
2	111455	110325	18,9444	1129,74	111436	110344	1110,79
3	113134	112436	−50,3889	698,12	113184	112385	748,51
⋮							
29	169195	167312	18,9444	1882,98	169176	167331	1864,03
30	172053	169423	−50,3889	2630,35	172103	169372	2680,74
31	174796	171533	31,4444	3262,73	174765	171565	3231,29

Forecasts

Period	Forecast
32	173663

7. Multiplicative Decomposition (Seasonal)

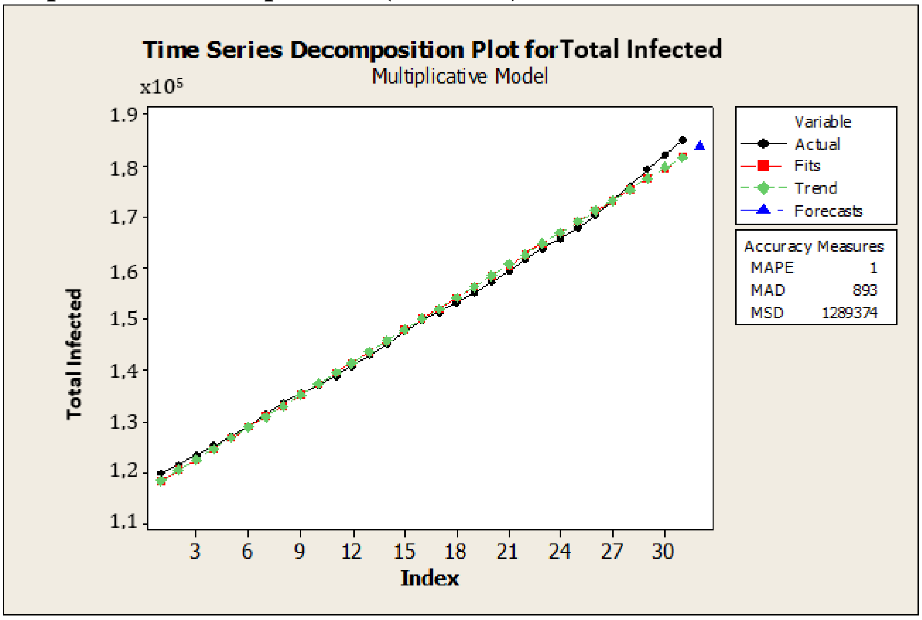

Fig 8. Multiplicative decomposition

Time Series Decomposition for Total Infected

Multiplicative Model

Data Total Infected

Length 31

NMissing 0

Fitted Trend Equation

Yt = 106105 + 2111*t

Seasonal Indices

Period	Index
1	1,00022
2	1,00014
3	0,99964

Accuracy Measures

MAPE	1
MAD	893
MSD	1289374

Time	Jumlah Terinfeksi	Trend	Seasonal	Detrend	Deseason	Predict	Error
1	109936	108215	1,00022	1,01590	109912	108240	1696,45
2	111455	110326	1,00014	1,01023	111440	110341	1113,82
3	113134	112437	0,99964	1,00620	113175	112396	737,81
⋮							
29	169195	167312	1,00014	1,01126	169172	167335	1860,44
30	172053	169422	0,99964	1,01553	172115	169361	2691,75
31	174796	171533	1,00022	1,01902	174757	171571	3225,06

Forecasts

Period	Forecast
32	173667

Table 3. Summary of methods

No	Forecasting method	MAD	MSE	Result
1	Exponential Smoothing	1117	1384694	176733
2	Exponential Smoothing with Trend	220,6	69994,2	177621
3	Linear Regression / Least Squares	895	1293730	173627
4	Moving Average	4350	19480844	172015
5	Trend Analysis (Regression Over Time)	895	1293730	173643
6	Additive Decomposition (Seasonal)	893	1288757	173663
7	Multiplicative Decomposition (Seasonal)	893	1289374	173667

Note: For the Linear Regression/Least Squares method and Trend Analysis (Regression Over Time) method, the results of MAD and MSE are usually always the same.
From the summary results in Table 2, it can be seen that the Exponential Smoothing with Trend method has MAD of 220.6, which is the smallest among the six existing methods, and MSE of 69994.2; the forecasting results are 177,621 cases of COVID-19 infections, which is predicted to occur on September 1, 2020.

Conclusion

1. From the results of exponential calculations using the Minitab program, according to the author, only seven methods can be calculated, namely, Exponential Smoothing with forecast results of 176733, Exponential Smoothing with Trend with forecast results of 177621, Linear Regression/Least Squares with forecast results of 173623, Moving Average with forecast results of 172015, Trend Analysis (Regression Over Time) with forecast results of 173643, Additive Decomposition (Seasonal) with forecast results of 173663, and Multiplicative Decomposition (Seasonal) with forecast results of 173667.
2. Of the seven forecasting methods, it turns out that the Exponential Smoothing with Trend method produces the smallest MAD, which is 220.6, and MSE, which is 69994.2, and the forecasting results are 177,621 patients, this is what will be used for further calculations.

References

[1] A. E. E. Eltoukhy, I. A. Shaban, F. T. S. Chan, and M. A. M. Abdel-Aal, "Predicting the Epidemiological Outbreak of the Coronavirus Disease 2019 (COVID-19) in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 17, no. 19, p. 7080, 2020, doi: 10.3390/ijerph17197080.
[2] G. Perone, "An ARIMA Model to Forecast the Spread of COVID-2019 Epidemic in Italy," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3564865.
[3] D. Alboaneen, B. Pranggono, D. Alshammari, N. Alqahtani, and R. Alyaffer, "Predicting the epidemiological outbreak of the coronavirus disease 2019 (COVID-19) in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, pp. 1–10, 2020, doi: 10.3390/ijerph17124568.
[4] A. A. Mattjik, *Perancangan Percobaan dengan Aplikasi SAS dan Minitab*. IPB Press, 2013.
[5] A. Widodo, *Pengantar Statistika*. Universitas Brawijaya Press, 2017.

[6] S. Santoso, *Kupas Tuntas Riset Eksperimen dengan Excel 2007 dan Minitab 15*. ElexMedia Komputindo, 2013.
[7] S. Siregar, *Statistika Terapan Untuk Perguruan Tinggi*. Kencana, 2017.
[8] Q. Li, W. Feng, and Y. H. Quan, "Trend and forecasting of the COVID-19 outbreak in China," *J. Infect.*, vol. 80, no. 4, pp. 469–496, 2020, doi: 10.1016/j.jinf.2020.02.014.
[9] B. Paixão *et al.*, "Estimation of COVID-19 under-reporting in Brazilian States through SARI," 2020.
[10] A. R. Tuite, D. N. Fisman, and A. L. Greer, "Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada," *Cmaj*, vol. 192, no. 19, pp. E497–E505, 2020, doi: 10.1503/cmaj.200476.
[11] S. Sitohang, A. S. Girsang, and Suharjito, "Prediction of the number of airport passengers using fuzzy C-means and adaptive neuro fuzzy inference system," *Int. Rev. Autom. Control*, vol. 10, no. 3, pp. 280–287, 2017, doi: 10.15866/ireaco.v10i3.12003.
[12] M. L. Ashari and M. Sadikin, "Prediksi Data Transaksi Penjualan Time Series Menggunakan Regresi Lstm," *J. Nas. Pendidik. Tek. Inform.*, vol. 9, no. 1, p. 1, 2020, doi: 10.23887/janapati.v9i1.19140.
[13] F. Pahala, "Prediksi Lalulintas Penumpang Bandar Udara Soekarno Hatta dengan Teknik Time Series Trend Forcasting," pp. 1–10, 2019, doi: 10.1017/CBO9781107415324.004.
[14] Sutarti, "Penggunaan Metode Runtun Waktu dengan Bantuan Minitab 11 for window untuk Forcasting Produksi Textil Pada PT. Primatexco Indonesia Kabupaten Batang Tahun 2009," 2019.
[15] F. S. D. Arianto and N. P, "Prediksi Kasus COVID-19 di Indonesia Menggunakan Metode Backpropagation dan Fuzzy Tsukamoto," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 120–127, 2020.
[16] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons and Fractals*, no. January, 2020.
[17] M. Genta, A. Shandi, R. Adhitama, and A. B. Arifa, "Penerapan Long Short Term Memory untuk Memprediksi Flight Delay," vol. 1, no. 10, pp. 1–3, 2020.
[18] C. P. Da Veiga, C. R. P. Da Veiga, A. Catapan, U. Tortato, and W. V. Da Silva, "Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models," *WSEAS Trans. Bus. Econ.*, vol. 11, no. 1, pp. 608–614, 2014.
[19] I. N. K. Wardana, N. Jawas, I. K. Agus, and A. Aryanto, "Prediksi Penggunaan Energi Listrik pada Rumah Hunian Menggunakan Long Short-Term Memory," vol. 1, no. 1, 2020.
[20] Adhitio Satyo Bayangkari Karno, "Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory)," vol. 1, no. 1, pp. 1–8, 2020.
[21] A. N. Irfansyah, "Prediksi polusi udara perkotaan di surabaya menggunakan," vol. 18, pp. 102–114, 2020.
[22] T. I. M. Penyusun, *Fakultas Teknik Universitas Negeri Semarang 2017*. 2017.
[23] S. Gowrishankar and P. S. Satyanarayana, "A Time Series Modeling and Prediction of Wireless Network Traffic," *Int. J. Interact. Mob. Technol.*, vol. 3, no. 1, pp. 53–62, 2009, doi: 10.3991/ijim.v3i1.284.
[24] D. Didiharyono and M. Syukri, "Forecasting with arima model in anticipating open unemployment rates in south sulawesi," *Int. J. Sci. Technol. Res.*, vol. 9, no. 3, pp. 3838–3841, 2020.
[25] B. Rai, A. Shukla, and L. K. Dwivedi, "Estimates of serial interval for COVID-19: A systematic review and meta-analysis," *Clin. Epidemiol. Glob. Heal.*, no. August, pp. 1–5, 2020, doi: 10.1016/j.cegh.2020.08.007.
[26] BNPB, "www.covid19.go.id," 2020. .