JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 20, No. 2, January 2024, pp. 351-359
DOI: 10.20956/j.v20i2.32122

Comparison Predictions of the *Demam Berdarah* Dengue (DBD) using Model Exponential Smoothing: Pegel's Classification and ChatGPT

Wiwik Wiyanti[1]*, Bakti Siregar[2]*

[*]Department of Statistics, Matana University, Tangerang, Indonesia
***Email**: wiwik.chae@gmail.com[1], bakti.siregar@matanauniversity.ac.id[2]*

Correspondence author: *wiwik.chae@gmail.com*

Abstract

The evolution of AI since the Covid-19 pandemic has developed very rapidly. Until 2023, AI is claimed to be a threat to several professional jobs, especially data analysts and scientists. The purpose of this research is to check the effectiveness chat-GPT to predict about *demam berdarah* dengue (DBD) case. Method of the analyzing the data in this research is Mixed method. Quantitative method using exponential smoothing: pegel's classification and qualitative method using GPT-3. The aim of this research is to check whether ChatGPT can predict the *demam berdarah* dengue (DBD) data time series. The prediction result are check it by exponential smoothing: pegel's classification method. The benefit of this research is it can be used to reference how far the evolution of AI can be threaten the profession of data analyst or data scientist. The result of this study conclude that the ChatGPT (GPT-3) can't predict DBD'd data correctly.

Keywords: prediction, ChatGPT, pegel's, dbd, forecast

1. INTRODUCTION AND PRELIMINARIES

Pandemic Covid-19 in 2020-2021 has impact in real life, such as force the society to adapt with the technology. Quoted from Yusuf's report [19] that the Covid-19 pandemic has changed the interaction relations of the world society to be adaptive to using the digital technology. This adaptation makes the digital technology develop very rapidly, like an existence of an open AI (Artificial Inteligence). One of the AI product is ChatGPT. Based on observation from December 2022 to June 2023, media in Indonesia widely report about ChatGPT. The report was such as benefit and threat from this AI machine. One claim of the threats is impact the ChatGPT for data scientist. ChatGPT is a chatbot machine that claimed to be a threat to various professions, one of which is proffessions in the technology sectors (coding, computer programmer, software engineers, data

analysts) [1,4]. Beside that, chatGPT wrote two truth and lie about it's status in Figure 1.1. In Figure one shows that chatGPT can predict the future with 100% accuracy.

Figure 1.1. CharGPT's claim (source [13])

When this article was written, ChatGPT has a new product that is GPT-4. Quoted from the web Cornell University about GPT-4 technical report [2] is ChatGPT was develop to can accept image and text input and produce the text output. Beside that while less capable than humans in many real-world scenario, GPT-4 exhibits human level performance on various professional and benchmark, including passing a simulated bar exam with score around the top 10% of test takers. Based on arXiv report [2] GPT-4 used to testing the simulation test for human. The result is the value perfomance for answering the real exam, score of uniform bax exam (MBE+MEE+MPT) is 298/400(~90^{th}), AP Statistics is 5(85^{th}-100^{th}). One of research report of Lira & Tang [14], ChatGPT can predict stock price movement. Paper of Lira & Tang, has been referred 8 times when this research do. Beside that, Hansen & Kazinnik [5] did evaluate by empirical the ability ChatGPT how to elaborate feedspeak, that is the language to discuss about monetary policy. The result of the study is ChatGPT succesfull improvement in classification compared others methods even frequently used, it's look like BERT. The reliability of ChatGPT to contribution to establshied simple and effective prompts that improve GPT-3's reliability has been researched [17] with conclussion that is with appropriate prompts, GPT-3 is more reliable than smaller-scale supervised models on all these facets.

Some of these research are become motivation why this research do. This research do prediction the medical data by ChatGPT. To evaluate the result of the prediction, this research compared the result by Exponential smoothing:pegel's classification. Why choose the exponential smoothing method for predicting?. The exponential smoothing can be said to be effective for predict the time series data, especially time series data with trend type [11]. In Exponential smoothing method have a good model to choose the data with caracteristic trend, linearity, season, non season, and exponential. That is reinforced by previous research scientific that is using the exponential smoothing method to predict data covid vaccinate with trend characteristic got forecast error ranging from 0% to 7%[18]. Moreover, research on predictions using exponential smoothing:pegel's classification has been carried out in various areas of life, such as agricultural sector [15] that is pegel's model used is A1, with maximum error prediction is 12,216% and the minimum is 8,266%. Next, from report [16] by pegel's exponential smoothing model get the error prediction that is MAPE 7,04% has best accuration compared the other methods. Moreover [3], got best model in triple exponential smoothing winter to predict DBD's patients in RSUD Dr.Soesilo Slawi with MAPE 43,81 at the $\alpha = 0{,}3;\ \gamma = 0{,}3, \Delta = 0{,}3$.

Based on that background, this research has formulation problem's, that is, Can ChatGPT do predict DBD case patient by exponential smoothing:pegel's classification?.

The design experiment of this research is mixed method [9], that is, quantitative and qualitative methods can see in Figure 2.1. These mixed methods strategies can be described using notation "+" and "→" that has developed in mixed method field. The following notation is adapted from Morse (1991), Tashakkori and Teddlie (1998) and Creswell and Plano Clark (2007) whose suggest the following [9]:

1. A"→" indicates a sequential form of data collection, with one form (e.g.,qualitative data) building on the other (e.g., quantitative data).
2. Capitalization indicates a weight or priority on the quantitative or qualitative data, analysis, and interpretation in the study.
3. "Quant" and "Qual" stand for quantitative and qualitative, respectively, and the use the same number of letters to indicate equality between the forms of data.
4. Boxes highlight the quantitative and qualitative data collection and analysis.

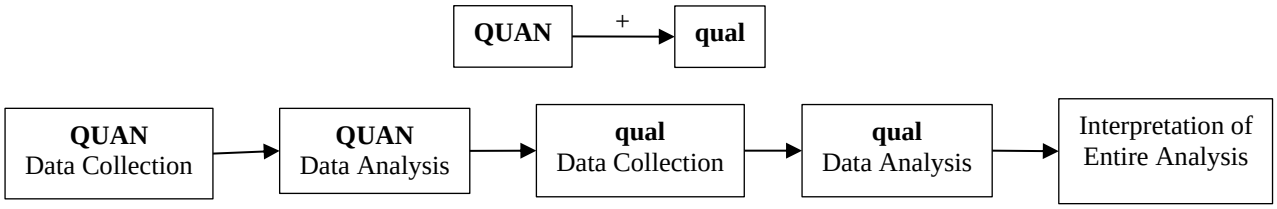

Figure 2.1. Design Mixed method sequential explanatory

The quantitative method this research is using exponential smoothing smoothing: pegel's classification and qualitative method is using ChatGPT. Reference the smoothing exponential: pegel's classification from Makridakis, Whellwright and McGee [11] and ChatGPT using GPT-3 version from [12]. The sequences of the quantitative process are:

1. Step 1, Trend and seasonal detection in the data time series. For check any season can be used plot or formula autocorrelation (r_k) at equation (1.1).

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})\,(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \tag{1.1}$$

From the pattern of autocorrelation values, it can be seen whether the data contains seasonal or not. As an illustration, for example, if the data is monthly, then the seasonal elements will be known form 12th, 24th, 36th lag the autocorrelation values and soon.

2. Step 2, Choose a mathematical model for predict using exponential smoothing: pegel's classification. In Makridakis, Whellwright and McGee [11] it's written that pegels (1969) has provided a simple but useful framework for classifiying models in two directions, as in Table 2.1.

Table 2.1. Pegel's Classification model

Trend component	Seasonal Component		
	1 (None)	2 (Linear)	3 (Non-Linear)
A (None)	A-1	A-2	A-3
B (Linear)	B-1	B-2	B-3
C (Non-linear)	C-1	C-2	C-3

Pegel's notation conversion of the nine model in Table 2.1 can summarized in the Table 2.2. Where P, Q, R and S in Table 2.2 vary according to the choice of which cell the model belongs. Nine of the exponential formula can be summarized by the formula:

$$\begin{aligned} L_t &= \alpha P_t + (1-\alpha)Q_t \\ b_t &= \beta R_t + (1-\beta)b_{t-1} \\ S_t &= \gamma T_t + (1-\gamma)S_{t-s} \end{aligned} \tag{1.2}$$

Table 2.2. Formula calculation of Pegel's Classification

Trend Component	Seasonal Component 1 (None)	2 (Linear)	3 (Non- Linear)
A (None)	$P_t = Y_t$ $Q_t = L_{t-1}$ $F_{t+m} = L_t$	$P_t = Y_t - S_{t-s}$ $Q_t = L_{t-1}$ $T_t = Y_t - L_t$ $F_{t+m} = L_t + S_{t+m-s}$	$P_t = Y_t / S_{t-s}$ $Q_t = L_{t-1}$ $T_t = Y_t - L_t$ $F_{t+m} = L_t S_{t+m-s}$
B (Linear)	$P_t = Y_t$ $Q_t = L_{t-1} + b_{t-1}$ $R_t = L_t - L_{t-1}$ $F_{t+m} = L_t + mb_t$	$P_t = Y_t - S_{t-s}$ $Q_t = L_{t-1} + b_{t-1}$ $R_t = L_t - L_{t-1}$ $T_t = Y_t - L_t$ $F_{t+m} = L_t + mb_t + S_{t+m-s}$	$P_t = Y_t / S_{t-s}$ $Q_t = L_{t-1} + b_{t-1}$ $R_t = L_t - L_{t-1}$ $T_t = Y_t / L_t$ $F_{t+m} = (L_t + mb_t) S_{t+m-s}$
C (Non-Linear)	$P_t = Y_t$ $Q_t = L_{t-1} b_{t-1}$ $R_t = L_t / L_{t-1}$ $F_{t+m} = L_t b_t^m$	$P_t = Y_t - S_{t-s}$ $Q_t = L_{t-1} b_{t-1}$ $R_t = L_t / L_{t-1}$ $T_t = Y_t - L_t$ $F_{t+m} = L_t b_t^m + S_{t+m-s}$	$P_t = Y_t / S_{t-s}$ $Q_t = L_{t-1} b_{t-1}$ $R_t = L_t / L_{t-1}$ $T_t = Y_t / L_t$ $F_{t+m} = L b_t^m S_{t+m-s}$

3. Step 3, calculate the mean absolute error (MAPE) by Makridakis, Whellwright and McGee [11].

$$\begin{aligned} PE_t &= \left(\frac{Y_t - F_t}{Y_t}\right) \times 100 \\ MAPE &= \frac{1}{n}\sum_{t=1}^{n} |PE_t| \end{aligned} \tag{1.3}$$

Note: PE_t = *Relative percentage error;* $MAPE$ = *Mean absolute percentage error*; n = count of the data.

4. Step 4 is calculate the confidence interval, because the value of prediction's data have no exact truth value. The formula of calculation by Makridakis, Whellwright and McGee [11].

$$\begin{aligned} & F_{n+1} \pm z\sqrt{MSE} \\ & MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \\ & e_t = Y_t - F_t \end{aligned} \tag{1.4}$$

Note: F_{n+1} = Value of the prediction data at $n+1$ time; *MSE= Mean Square Error;* e_t = error at t time. The value of z determines the width and probability of the prediction

interval. In this research using $z = 1.96$ gives a 95% confidence interval. The value of z depend on the alpha or confidence interval taken, summarize in table 2.3.

Tabel 2.3. Table most common values of z

z	Probability (1-α)
0.674	0.50
1.000	0.68
1.150	0.75
1.282	0.80
1.645	0.90
1.960	0.95
2.576	0.99

The qualitative method in this research using ChatGPT. The reason using ChatGPT is it truth and lie about status of ChatGPT[13]:

1. ChatGPT is experiencing high traffic at the moment.
2. The developers are working hard to accommodate all users.
3. ChatGPT can predict the future with 100% accuracy.

In this research is use GPT-3 to analyze the data.

2. MAIN RESULTS

Data in this research was using local and international data of *demam beradarah* dengue (DBD). DBD's local data gets from Jakarta's web government [6], Kendal web governement [7] and international data from philipina's data gets from kaggle [8]. The various DBD data are used to examine various type of characters from DBD data patterns found, then applied to pegels classification and chatGPT to get the predict value and calculate the effectiveness of pegel's model.

Form the quantitative data analysis of DBD local data (Jakarta), the data shows that the data pattern doesn't any trend and seasonality, so the best model pegel's classification to use is the Pegels A1 model. The formula to predict this data in equation (1.5)

$$\begin{aligned} P_t &= Y_t \\ Q_t &= L_{t-1} \\ F_{t+m} &= L_t \end{aligned} \tag{1.5}$$

After implementation those formula 1.5, an alpha value (α) 0.99 was obtained with forecasting error calculation (MAPE) is 54.98%. Next, the DBD data from Kendal government which is used for data analysis, has a data pattern that contains an upward trend and cycle. Because the pegel's classification does not involve cycle in the formula, so this cycle pattern is ignored. The best model for this condition is pegels B1. The formula to predict this data in equation (1.6)

$$\begin{aligned} P_t &= Y_t \\ Q_t &= L_{t-1} + b_{t-1} \\ R_t &= L_t - L_{t-1} \\ F_{t+m} &= L_t + mb_t \end{aligned} \tag{1.6}$$

After implementation equation (1.6), gets best α and β are 0.9 and 0.002, respectively. The forecasting error calculation (MAPE) is 45.88%. The last, the DBD data from Philipina, Region 4. The pattern of data, there is no trend but seasonality is exist. So, best model of this data is pegels A2. The formula to predict this data in equation (1.7).

$$\begin{aligned} P_t &= Y_t - S_{t-s} \\ Q_t &= L_{t-1} \end{aligned} \tag{1.7}$$

$$T_t = Y_t - L_t$$
$$F_{t+m} = L_t + S_{t+m-s}$$

After implementation equation (1.7), gets best α and γ are 0.9 and 0.1, respectively. The forecasting error calculation is 25.04%.

Analyze the data qualitatively was using GPT-3. The data DBD was using for this analyze from Philipina, Region 4. The result obtained were that GPT-3 was unable to predict this data. This conslusion got from some fact when analysing data do. First fact is GPT-3 doesn't recognize pegel's classification in exponential smoothing. Second, GPT-3 produce the autocorellation value is doesn't true in other word the value of autocorrelation is fabrication. Comparisson the value of autocorrelation by formula equation(1.1), SPSS, R and GPT-3 in Table 3.1.

Table 3.1. Autocorrelation's value of 31 lags from DBD Philipina

Lag	r_k Formula,SPSS,R	r_k GPT-3	lag	r_k Formula,SPSS,R	r_k GPT-3
0	1	1	16	-0,143	-0,045
1	0,777	0,516	17	-0,114	0,024
2	0,429	0,233	18	-0,09	0,093
3	0,175	0,143	19	-0,069	0,153
4	0,009	0,05	20	-0,05	0,251
5	-0,066	0,053	21	-0,029	0,344
6	-0,099	0,191	22	-0,003	0,444
7	-0,14	0,208	23	0,025	0,442
8	-0,165	0,202	24	0,038	0,377
9	-0,13	0,121	25	0,036	0,322
10	-0,077	0,008	26	-0,01	0,274
11	-0,026	-0,061	27	-0,078	0,239
12	-0,006	-0,083	28	-0,106	0,222
13	-0,043	-0,083	29	-0,109	0,16
14	-0,105	-0,102	30	-0,101	0,091
15	-0,159	-0,067	31	-0,088	0,023

Calculation of the autocorrelation by equation (1.1), SPSS and R at lag 12, 24 are -0,006 and 0,038, respectively. This point is inflection points. Calculation the autocorrelation by GPT-3 at lag 12 and 24 are -0,083 and 0,377, respectively. These values are false because the corect actual values should be -0,006 and 0,038. Otherwise, the calculation of autocorrelation by manual, SPSS and R gets seasonal characteristic but GPT-3 is not. Graphically, the difference of autocorrelation values by equation (1.1), SPSS, R and GPT-3 in Figure 3.1.

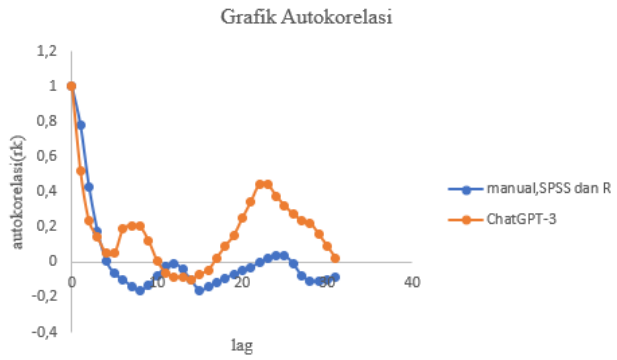

Figure 3.1. Plot autocorrelations data from equation(1.1), SPSS, R and GPT-3

The next fact from GPT-3 is ignorance the pegel's classification model. The researcher ask GPT-3 about pegel's classification model or exponential smoothing:pegel's classification. GPT-3 answer is like Figure 3.2.

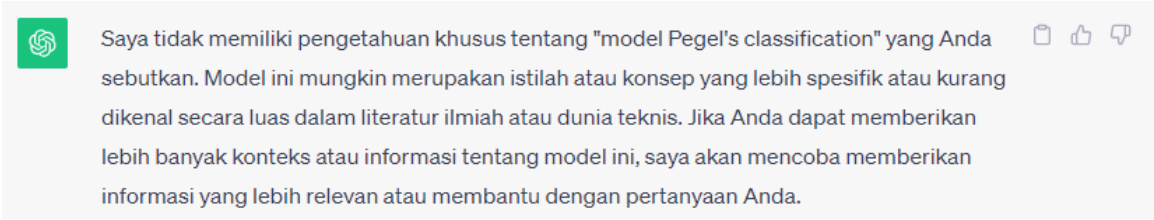

Figure 3.2. GPT-3's answer

Figure 3.2 explain that GPT-3 doesn't have knowledge about "pegel's classification model" especially. This machine said that the model may be new term or specific concept or less widely known from scientific literature. Next, this machine traying to give the more relevant information to help the question's researcher. Furthermore, even though the researcher was given the addition information about pegel's classification, GPT-3 was consistently answer don't know about pegel's classification. From this analytic, the researcher was stopped asking knowledge about "pegel's classification model" this machine.

3. CONCLUSION

The result of DBD analyze is conclude that GPT-3 cannot relied on predict the DBD's data, this case, applying the exponential smoothing: pegel's classification method for predict the DBD time series data.

For the next study may use the other method with involve the cycle character on the time series data. Next, compare two or more forecasting method to get the best method for forecast the DBD time series data. Beside that, try the GPT-4 to analyze the time series data.

ACKNOWLEDGEMENT

Thank you to our students, Alicia and Diyas whom collect the DBD data. Otherwise, thank you to Matana University to support the lecturer to do research periodically.

CONFLICT OF INTEREST

Full of results of this research is truly based on scientifically and there is no competing interests. The author is fully responsible for this matter. Thank you.

REFERENCES

[1] Dewi, I.R, 2023. Sederet Pekerjaan Terancam Punah Siap Siap Ganti Profesi. Avalaible at : https://www.cnbcindonesia.com/tech/20230316071902-37-422064/sederet-pekerjaan-terancam-punah-siap-siap-ganti-profesi (cited 19 Juni 2023)

[2] Ecoffet, A., 2023. Gpt-4 Technical Report. Cornell University. Available from: https://arxiv.org/abs/2303.08774v3 (cited 07 Juli 2023).

[3] Fitria, J.D., 2017. Peramalan Jumlah Pasien Dbd Di RSUD Dr Soesilo Slawi Dengan Menggunakan Metode Dekomposisi Dan Triple Exponential Smoothing Winter. *Tugas Akhir*. Universitas Negeri Semarang.

[4] Gunawan, A., 2023. Ancaman Chat Gpt Bagi Banyak Profesi Kreatif, Tetapi Bisa Bermanfaat Positif. Available from: https://www.kompasiana.com/ddigunn/63f235f104dff04b02235492/ancaman-chat-gpt-bagi-banyak-profesi-kreatif-tetapi-bisa-bermanfaat-positif (cited 19 Juni 2023)

[5] Hansen, A.L, & Kazinnik, S., 2023. Can Chatgpt Decipher Fedspeak. *Elsevier*. Available at: SSRN: https://ssrn.com/abstract=4399406 or http://dx.doi.org/10.2139/ssrn.4399406

[6] https://data.jakarta.go.id/dataset/data-jumlah-penderita-dbd-hasil-penyelidikan-epidemologi-di-provinsi-dki-jakarta (cited 6 Agustus 2023)

[7] https://kendalkab.bps.go.id/indicator/30/46/1/banyaknya-penderita-demam-berdarah.html. (cited 30 Agustus 2023).

[8] https://www.kaggle.com/datasets/vincentgupo/dengue-cases-in-the-philippines. (Cited 8 Agustus 2023)

[9] John, W., & Creswell, 2009. *Research Desain: Qualitative, Quantitative And Mixed Methods Approach*. Third Edition. SAGE Publications,Inc., New York.

[10] Lira, AL, & Tang, Y., 2023. Can Chatgpt Forecast Stock Price Movement?. University of Florida. Available at: https://arxiv.org/pdf/2304.07619.pdf (cited 07 Juli 2023).

[11] Makridakis, S.G., Wheelwright, S.C., & Hyndman, R.J., 1998. *Forecasting Methods And Applications*, Third Edition. Wiley: ISBN: 0471532339, 9780471532330

[12] Open, AI., 2023. Introducing Chatgpt. [internet] [cited 07 Juli 2023]; Avalibale from: https://openai.com/blog/chatgpt

[13] Open, AI., 2023. available from: https://chat.openai.com/auth/login (cited 22 November 2023).

[14] Permana, A., 2023. Mengenal Chat Gpt, Algoritma Dibalik Kemajuan Kecerdasan Buatan. Avalaible from: https://www.itb.ac.id/berita/mengenal-chat-gpt-algoritma-di-balik-kemajuan-kecerdasan-buatan/59433 (cited 9 Juni 2023).

[15] Prameswari & Ayu, R., 2015. Peramalan Tingkat Produksi Padi Di Jawa Timur Dengan Metode Exponential Smoothing: Pegels Classification Model A-2. *Sarjana Thesis*. Universitas Brawijaya.

[16] Setiawan, D.A, Wahyuningsih, S. & Goejantoro, R., 2020. Peramalan Produksi Kelapa Sawit Menggunakan Winter's Dan Pegel's Exponential Smoothing Dengan Pemantauan Tracking Signal. *Jambura Journal of Mathematics*;Vol2No1:pp1-14.

[17] Si C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Gabrier, J, & Wang, L., 2023. Prompting GPT-3 to be reliable. *Proceeding of a conference paper at ICLR*. Available from: https://arxiv.org/pdf/2210.09150.pdf

[18] Wiyanti, W., 2023. Effectiveness of single and double exponential smoothing: ses, arrses and holt's linear for time series data prediction with trend and non-seasonal characteristic (covid-19 vaccinate case): efektivitas metode exponential smoothing untuk prediksi data runtun waktu pola tren dan non-musiman (studi kasus cakupan vaksinasi covid-19. *Journal Matematika, Statistika Dan Komputasi*, Vol.20 No.1, 52-54. DOI https://doi.org/10.20956/j.v20i1.27193

[19] Yusuf., 2021. Pandemi Covid-19 Pacu Adaptasi Gunakan Teknologi Digital. Available from: https://www.kominfo.go.id/content/detail/32602/pandemi-covid-19-pacu-adaptasi-gunakan-teknologi%20digital/0/berita_satker (cited 29 Juni 2023)