WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Modeling Google Search Trends for Leading Causes of Death in Indonesia Using the Loop Prophet Method

*Anggun Yuliarum Qur'ani[1*,2], Chandra Sari Widyaningrum[2]*

[1*]Universitas Udayana, Indonesia, anggunyuliarum@unud.ac.id

[2]PT. Barokah Ilmu Indonesia, Indonesia, chandrasari@biiscience.com

**A B S T R A C T :**

The *Loop Prophet* method is a time series forecasting approach that utilizes the Facebook Prophet model iteratively (looping) to automatically process multiple variables or keywords. This study aims to predict the weekly search trend on Google for the leading causes of death in Indonesia, taking into account trend, yearly seasonality, and weekly seasonality components, as well as to evaluate model performance using error metrics such as NMAE, NRMSE, and MAPE. The dataset consists of weekly Google Trends search data for selected diseases during the period from September 2020 to July 2025. The results indicate that the Loop Prophet model successfully captures both trend and seasonal patterns. Based on the evaluation criteria, models categorized as "very good" were obtained for the keywords "*Stroke*", "*Diabetes*", and "*Diare*". The "good" category was obtained for "*Serangan Jantung*", "*Sirosis Hati*", "*Penyakit Paru*", "COPD", dan "*Neonatal*". The keyword of "*Tuberkulosis*" was categorized as "good enough". Meanwhile, the "poor" category was found for "*Kanker Paru-Paru*" dan "*Pneumonia*", which tend to have fluctuating patterns influenced by incidental events. These findings demonstrate that the Loop Prophet approach proves effective in analyzing time series with complex seasonal structures; however, its accuracy declines when applied to diseases exhibiting highly irregular search behaviors.

**Keywords**: Loop Prophet Model, Forecasting, Google Trends, Death, Diseases

## 1. Introduction

Monitoring annual mortality rates is crucial to understanding their causes and adjusting health systems to respond effectively. Knowing the causes of death reveals how people live and is vital for strengthening healthcare and lowering preventable mortality, particularly in Indonesia during epidemiological shifts.

The (World Health Organization, 2024) classifies causes of death into three major categories: (1) Infectious and related conditions—such as maternal, perinatal, and nutritional disorders; (2) chronic forms of non-communicable diseases, encompassing heart disease, stroke, diabetes, and cancer; and (3) injuries, notably those caused by road accidents and violence. In 2021, the list of the ten major causes of death globally was updated by WHO through the Global Health Estimates (GHE) framework. Worldwide, ischemic heart disease is the leading cause of death, with COVID-19 and stroke next in rank. In Indonesia, the main causes are non-communicable diseases, especially heart disease, stroke, and diabetes. For children under five, (UNICEF, 2024) reported that pneumonia and diarrhea are the two top causes of child deaths worldwide, each taking more than 2,000 young lives every day.

As access to the internet and health-related information expands, people's search behavior on digital platforms such as Google has begun to reflect their attention to certain health issues. Data from Google Trends provides real-time and historical information about users' search interests for specific keywords. By analyzing search trends for major diseases identified by WHO and UNICEF, such as heart disease, stroke, diabetes, pneumonia, and diarrhea, we can better understand the dynamics of public health awareness over time. The integration of official epidemiological data (such as WHO's GHE and UNICEF reports) with Google Trends analysis has the potential to serve as an early indicator of public attention, shifts in awareness, and potential surges in cases. This is particularly important as a basis for evidence-based health policy planning, health promotion, and strengthening early warning systems.

To support evidence-based policy, analytical methods are needed that not only interpret historical trends but also generate accurate short- to medium-term forecasts. A commonly applied approach for time series forecasting is Facebook Prophet (Prophet model), an additive model based on trend, seasonality, and holiday components, which is known for its flexibility and accuracy in handling daily, monthly, and yearly data. The Prophet forecasting model has also been applied by (Brykin, 2024; Oktavia & Witanti, 2024) to predict sales from Tinkoff data and to forecast air quality in Yogyakarta. Both studies produced highly accurate predictions, with one yielding a MAPE value of 0.81%, demonstrating the model's outstanding ability to deliver precise forecasts across different sales categories. Several years earlier, applied Prophet and Holt-Winters models for long-term electricity load forecasting in Kuwait, reporting that Prophet proved to be more robust against noise than the Holt-Winters model. In the same year, (Oo

& Phyu, 2020) forecasted temperature in Myitkyina using Facebook Prophet, and the findings indicated that the proposed model delivered accurate and efficient temperature forecasts.

In this study, the **Loop Prophet approach** is applied, an iterative process of automatically implementing the Prophet model across multiple disease-related keywords. This approach allows systematic and efficient analysis of dozens of disease keyword variables, producing comprehensive trend visualizations and forecasts. The method enables sharper decision-making regarding which diseases require priority attention in public health interventions.

Forecasting disease search interest trends for several weeks or months ahead using the Loop Prophet model provides an early picture of potential increases in specific health risks in the community. Such information can be used to design health education strategies, disease prevention campaigns, and more targeted public awareness initiatives. These findings also contribute to supporting Indonesia's goal of strengthening its national health system and meeting the Sustainable Development Goals (SDGs), especially Goal 3 on health and well-being for all age groups. (United Nations, 2015).

## 2. Literature Review

**The Prophet model** is a forecasting approach that typically breaks down a time series into three core components: trend, seasonality, and holidays. (Taylor & Letham, 2017). This model belongs to the class of continuous-function–based forecasting models, which treats time as a continuous numerical variable. In general, the model may be represented in mathematical form as:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{2.1}$$

where $g(t)$ is defined as the trend function determined by time $t$ which models non-periodic changes, $s(t)$ is the seasonal function (periodic changes) at time $t$, $h(t)$ is a function showing the effect of holidays at time $t$ (irregular, can span one or more days), and $\varepsilon_t$ is the error term not captured by the model, assumed to follow a Normal distribution.

The types of trends commonly used in the Prophet model include **saturating growth**, usually modeled with logistic growth, and **piecewise linear growth**, defined by a constant growth rate within segments. In this study, the piecewise linear trend type is implemented. This trend is used when there is no indication of logistic growth in the data. The growth rate in the piecewise linear trend is segmented into specific time intervals (piecewise), making the number of parameters efficient for modeling.

Consider the case in which there are $S$ changepoints at time $s_j$, with $j = 1,2,\dots,S$. A vector denoting the adjustments applied to the growth rate $\boldsymbol{\delta} \in \mathbb{R}^S$, where $\delta_j$ is the change in rate at time $s_j$. The growth rate at time $t$ is the base rate $k$ plus all adjustments up to that point, expressed as:

$$g_0(t) = k + \sum_{j:t\geq s_j} \delta_j \tag{2.2}$$

This is more easily represented by defining a vector $\boldsymbol{a}(t) \in \{0,1\}^S$, such that

$$a_j(t) = \begin{cases} 1, & \text{jika } t \geq s_j \\ 0, & \text{selainnya} \end{cases} \tag{2.3}$$

Thus, the growth rate at time $t$ becomes:

$$g_0(t) = k + \boldsymbol{a}(t)^T\boldsymbol{\delta} \tag{2.4}$$

The piecewise linear trend can then be written as the $t$ growth rate $g_0(t)$ multiplied by $t$, plus an offset parameter, yielding:

$$g(t) = (k + \boldsymbol{a}(t)^T)\boldsymbol{\delta}t + (m + \boldsymbol{a}(t)^T)\boldsymbol{\gamma} \tag{2.5}$$

where $m$ is the offset parameter, and $\gamma_j$ is set to $-s_j\delta_j$ to ensure continuity.

The seasonal effect in Prophet is represented using a Fourier series expansion:

$$s(t) = \sum_{n=1}^{N} \left( a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \tag{2.6}$$

where $P$ is the expected period of the time series (e.g., 365.25 for yearly data, 7 for weekly data, expressed in days). To estimate seasonality, $2N$ parameters must be determined: $\boldsymbol{\beta} = [a_1, b_1, a_2, b_2, \dots, a_N, b_N]^T$. The seasonal vector matrix for each $t$ can be constructed in the historical data (original data) and future data (forecast results), for example if the data is known for weekly seasonality and $N = 10$ (10 is an example of the number of N proposed for seasonal smoothing priors), then

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{7}\right), \sin\left(\frac{2\pi(1)t}{7}\right), \dots, \cos\left(\frac{2\pi(10)t}{7}\right), \sin\left(\frac{2\pi(10)t}{7}\right)\right] \tag{2.7}$$

Thus, seasonality can be written as:

$$s(t) = X(t)\boldsymbol{\beta} \tag{2.8}$$

with the assumption $\boldsymbol{\beta} \sim N(0, \sigma^2)$ in the generative model, enforcing a smoothing prior. This additive seasonality technique resembles the method applied in exponential smoothing (Gardner, 1985). In contrast, multiplicative seasonality where seasonality scales the trend, can be applied using a log transform. In this study, holiday effects are excluded because disease-related search trends are not strongly affected by national holidays. Multiplicative seasonality is used instead. From (2.1), the resulting model is:

$$y(t) = g(t)s(t) + \varepsilon_t \tag{2.9}$$

Thus, if the logarithmic value of equation (2.9) is taken, we get

$$\log\big(y(t)\big) = \log\big(g(t)\big) + \log\big(s(t)\big) + \log(\varepsilon_t) \tag{2.10}$$

A set of time series $y_z(t)$ with $z = 1,2,\dots,Z$, $t = 1,2,\dots,T$, and $z$ is the number of time series variables, $t$ is a continuous time variable, is called **a multiple time series**, and $\hat{y}_z(T+h)$ is the forecasting function of the multiple time series data. As in the univariate case, one of the main objectives of multiple time series analysis is to determine the set of cumulative functions $f_1, f_2, \dots, f_z$ that can be used to obtain forecasting results from a model (Lütkepohl, 2007). The forecasting function of $\hat{y}_z(T+h)$ at the last time $T$ from $z = 1,2,\dots,Z$ can be expressed as follows.

$$\hat{y}_1(T+h) = f_1(y_1(T), y_2(T), \dots, y_Z(T), y_1(T-1), y_2(T-1), \dots, y_Z(T-1), \dots, y_1(T-2), y_2(T-2), \dots) \tag{2.11}$$

$$\hat{y}_2(T+h) = f_2(y_1(T), y_2(T), \dots, y_Z(T), y_1(T-1), y_2(T-1), \dots, y_Z(T-1), \dots, y_1(T-2), y_2(T-2), \dots)$$

$$\dots$$

$$\hat{y}_Z(T+h) = f_Z(y_1(T), y_2(T), \dots, y_Z(T), y_1(T-1), y_2(T-1), \dots, y_Z(T-1), \dots, y_1(T-2), y_2(T-2), \dots)$$

Thus, from (2.10) and (2.11) for multiple time series with $z = 1,2,\dots,Z$ can be written with the equation

$$\begin{aligned} \log(\hat{y}_z(T+h)) &= \log(\hat{g}_z(T+h)) + \log(\hat{s}_z(T+h)) \\ &= \log\big(f_z(g_1(T), g_2(T), \dots, g_Z(T), g_1(T-1), g_2(T-1), \dots, g_Z(T-1), \dots, \quad g_1(T-2), g_2(T-2), \dots)\big) \\ &\quad + \log\big(f_z(s_1(T), s_2(T), \dots, s_Z(T), s_1(T-1), s_2(T-1), \dots, s_Z(T-1), \dots, \quad s_1(T-2), s_2(T-2), \dots)\big) \end{aligned} \tag{2.12}$$

The function in (2.12) is a forecasting function obtained from an iterative or looping process using the Prophet model. In other words, $\log(\hat{y}_z(T+h))$ is called the **Loop Prophet model forecasting function with a multiplicative seasonal form and without holiday effects.**

Parameter estimation in the Loop Prophet model uses Maximum A Posteriori (MAP), which is a method to find parameter values that maximize posterior probability. In the Loop Prophet model, each model parameter has a prior distribution. This MAP estimation can be written mathematically as follows.

$$\begin{aligned} \hat{\theta}_{MAP} &= \arg\max_{\theta} f\big(\theta | y_z(t)\big) \\ &= \arg\max_{\theta} \frac{f(y_z(t)|\theta)h(\theta)}{\int_{\Theta} f(y_z(t)|\vartheta)\, h(\vartheta) d\vartheta} \\ &= \arg\max_{\theta} f(y_z(t)|\theta)h(\theta) \end{aligned} \tag{2.13}$$

with $\hat{\theta}_{MAP}$ being the parameters estimated using MAP (Indian Institute of Technology Madras, 2022), $f(y_z(t)|\theta)$ being the likelihood function, $h(\theta)$ being the prior function of each parameter, and $\theta = \{k, m, \delta_j, \beta, \sigma\}$ being the set of linear growth trend parameters, seasonal parameters, and error parameters.

Model evaluation is needed to determine the accuracy of the model that has been formed in forecasting. In research using several evaluation measures, namely Normalized Mean Absolute Error (NMAE), Normalized Root Mean Squared Error (NRMSE), and Mean Absolute Percentage Error (MAPE). Normalized Mean Absolute Error (NMAE) is the standardization value of MAE against the mean actual value in percent units which can be expressed in the following equation.

$$\text{NMAE}_z(\%) = \frac{\frac{\sum_{t=1}^{T} |y_z(t) - \hat{y}_z(t)|}{T}}{\bar{y}_z} \times 100 \tag{2.14}$$

Normalized Root Mean Squared Error (NMRSE) is a standardization of the RMSE value with units in percentage. NRMSE or RMSE in percent units can be written in the following equation (Kambezidis, 2012).

$$\text{NRMSE}_z(\%) = \frac{\sqrt{\frac{\sum_{t=1}^{T} (y_z(t) - \hat{y}_z(t))^2}{T}}}{\bar{y}_z} \times 100 \tag{2.15}$$

MAPE measures the average prediction error expressed in percentage terms. (Žunić et al., 2020) wrote the MAPE value in the form of the following equation.

$$\text{MAPE}_z(\%) = \frac{\sum_{t=1}^{T}\left|\frac{y_z(t)-\hat{y}_z(t)}{y_z(t)}\right|}{T} \times 100; \qquad z = 1,2, \dots Z; t = 1,2, \dots, T \tag{2.16}$$

where $z$ is the number of time series variables $y$, and $y_z(t)$ is the actual value of the $-z$th variable at time $t$, also $\hat{y}_z(t)$ mis the predicted value of the$-z$ th at the time $t$ from the Prophet model. From the results obtained from the three evaluation measures, the criteria for the percentage values of NMAE, NRMSE, and MAPE are formed (Despotovic et al., 2016; Lewis, 1982) namely values below 10% are categorized as "**Excellent**" models, between 11%-20% are categorized as '**good**' models, between 21%-30% are categorized as "**fairly good**" models, and above 30% are categorized as "**bad**" models when used for forecasting.

To understand the results of the analysis, it is necessary to first explain the definition of each disease keyword used in the analysis. Each keyword represents a specific search term in Google Trends that is assumed to reflect the level of public concern about a disease. The first disease keyword is "*Serangan Jantung*". Based on the explanation (Jannah, 2018) "***serangan jantung***" (heart attack) is a condition when the heart muscle suddenly does not get blood supply due to sudden blockage of the coronary arteries by blood clots due to plaque rupture. Furthermore, the disease keyword **"*Stroke*"** according to (Kementerian Kesehatan Republik Indonesia, 2019b) is an acute clinical event caused by neurological dysfunction in the brain, spinal cord, or retina, persisting for at least 24 hours or resulting in death due to vascular disturbance. Meanwhile (Kementerian Kesehatan Republik Indonesia, 2024) defines "**diabetes**" as a multifactorial metabolic disease defined by ongoing high blood sugar due to problems with insulin secretion, insulin function, or their combination.

Furthermore, (Kementerian Kesehatan Republik Indonesia, 2019a) mentions that "***Tuberkulosis***" (Tuberculosis) is a long-term infectious disease caused by the bacterium *Mycobacterium tuberculosis*. Meanwhile, the definition of "***Sirosis Hati***" (Liver Cirrhosis) is a condition characterized by the formation of scar tissue in the liver organ as a result of prolonged damage that disrupts the normal functioning of the liver (Kementerian Kesehatan Republik Indonesia, 2022). As for "***Penyakit Paru***" (Lung Disease) it is defined as any type of disease that attacks the lung organ. In lung disease, there are 3 complaints that are often found, namely coughing, shortness of breath and chest pain (FK Unsoed, 2013; Rahmawati, 2016).

Furthermore, the keyword "***Chronic Obstructive Pulmonary Disease* (COPD)**" or *Penyakit Paru Obstruktif Kronik (PPOK)* is a long-term airway blockage caused by emphysema and chronic bronchitis (Paramitha, 2020). Then the definition for "***Kanker Paru-Paru***" (Lung Cancer) based on (RSU Tjokronegoro, 2021) is a type of cancer that starts in the lungs. This condition is even one of the leading causes of cancer deaths worldwide. People are more at risk of developing lung cancer if they have a history of COPD, Tuberculosis, and chronic brochitis.

Next, "**Neonatal**" ys a baby within the first 28 days of life. Neonatal mortality is extremely vulnerable, the disease kills 2 million babies and children worldwide (Tidy, 2023; Tim detikHealth, 2022). The next disease keyword definition is "***Pneumonia***" hich is inflammation of the lungs due to acute infection of the respiratory tract, caused by viruses, bacteria, or fungi (Aji Muhawarman, 2024). And finally, "***Diare***" (Diarrhea) is characterized as a condition where a person experiences an increased frequency of bowel movements with liquid or watery stools. It can be accompanied by other symptoms such as nausea, vomiting, abdominal cramps, and sometimes weight loss (Kementerian Kesehatan Republik Indonesia, 2023).

## 3. Research Methods

The following describes in detail the data used in the study as well as the data analysis procedures carried out.

### *3.1. Data*

This study uses secondary data in the form of weekly search volume of Indonesian people on various diseases that cause the highest mortality, obtained from Google Trends. The data was collected for the last five years from the first week of September 2020 to the last week of July 2025 and has been normalized by Google based on the total search volume each week. Each keyword used represents each disease studied. The selection of keywords refers to the list of diseases released by (UNICEF, 2024) and (World Health Organization, 2024). From this data, the analysis was carried out using the Prophet Loop approach, which is the application of the Prophet time series model iteratively for each disease keyword. The following is a list of disease keywords used in this study as a representation of search variables in Google Trends.

**Table 1 Research Variables**

| No. | Disease Keywords | No. | Disease Keywords |
|---|---|---|---|
| 1. | *Serangan Jantung* | 7. | COPD |
| 2. | *Stroke* | 8. | *Kanker Paru-Paru* |
| 3. | *Diabetes* | 9. | *Neonatal* |
| 4. | *Tuberkulosis* | 10. | *Pneumonia* |
| 5. | *Sirosis Hati* | 11. | *Diare* |
| 6. | *Penyakit Paru* | | |

Source: (UNICEF, 2024; World Health Organization, 2024)

### *3.2. Data Analysis Procedure*

The data analysis procedure in this study was carried out through several systematic stages. The steps of data analysis using Loop Prophet analysis in detail are presented as follows.

1. Preparing search trend data obtained by downloading it through the Google Trends platform (https://trends.google.com/trends/explore?date=all&geo=ID&hl=en-GB), with the setting of the Indonesian region and with a time span of the past five years. This data was then filtered based on keywords that represent each disease keyword that is the focus of the research.
2. Descriptive analysis based on the plotting results of the research data.
3. Estimating the Loop Prophet model parameters using the "prophet" package by (Taylan, 2023).
4. Evaluate the Loop Prophet model using the NMAE, NRMSE, and MAPE evaluation metrics in Equations (2.14) - (2.16) respectively.
5. Categorize the Loop Prophet model evaluation results.
6. Perform forecasting using the Loop Prophet model with the "Good" and "Excellent" model evaluation result categories.
7. Interpretation of results and discussion.

## 4. Results And Discussion

To understand the dynamics of public attention to various diseases that are the main causes of death in Indonesia, data plotting of the number of Google searches from the beginning of September 2020 to mid-2025 was carried out. Graphs 1a and 1b below show the weekly search patterns for several disease keywords representing the non-communicable disease and infectious disease groups. Each graph represents the level of public search interest in each disease keyword over time, which may reflect awareness, concern, or the influence of certain events on health information search behavior.

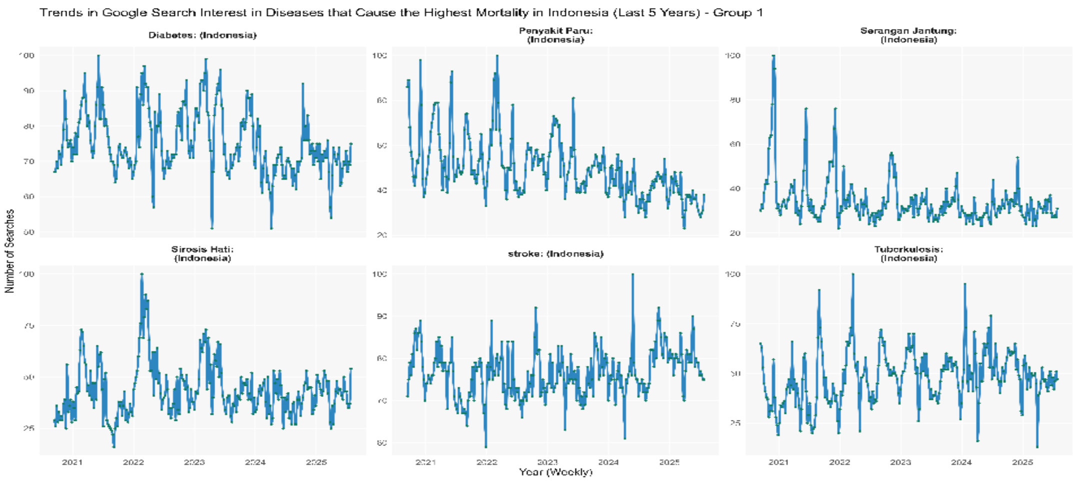


**Figure 1a. Trends in Google Search Interest in Keyword Diseases that Cause the Most Deaths in Indonesia (Weekly)**

Based on the visualization results of Figure 1a, a plot of Google searches in Indonesia for the six diseases that cause the highest mortality, there are variations in search patterns that reflect the different levels of public attention in a span of 5 years per week, from the beginning of September 2020 to mid-2025. The keyword "*Diabetes*" shows a relatively stable search pattern throughout the observation period with fairly consistent seasonal fluctuations. This indicates that diabetes is a health issue that receives continuous attention from the public, most likely due to its high prevalence and annual campaigns such as World Diabetes Day. In contrast, the keyword "*Penyakit Paru*" experienced a high spike in searches especially in 2020 to 2022, which temporally correlates with the peak of the COVID-19 pandemic. This suggests that public interest in lung disease tends to increase when there is a high risk of respiratory distress due to the virus.

A different pattern is seen for "*Serangan Jantung*", with sporadic and temporary spikes in searches. These peaks are most likely caused by public events or media coverage, such as heart attack cases in public figures, which trigger reactive searches by the public. "*Stroke*" shows a significant increase during 2022, then declines and tends to stabilize. This could indicate increased public awareness of mild neurological symptoms post-pandemic, or increased exposure to medical information about non-fatal strokes. For the keyword "*Sirosis Hati*", search trends appear flatter and do not show strong seasonal peaks, with relatively low and consistent interest. This suggests that although cirrhosis is a serious disease, public concern about it is not as strong as it is for other non-communicable diseases.

The trend of public searches related to "*Tuberkulosis*" shows a pattern that fluctuates sharply from year to year, with some spikes at certain moments that may be influenced by health campaigns or media coverage. However, overall, there is no consistent upward or downward trend.

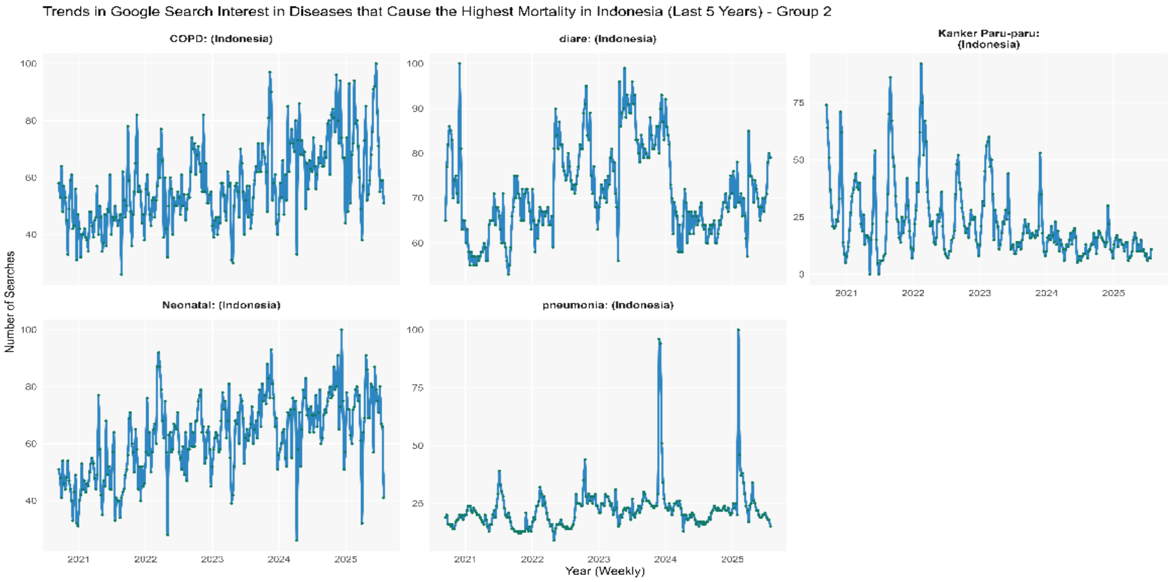


**Figure 1b. Trends in Google Search Interest in Keyword Diseases that Cause the Most Deaths in Indonesia (Weekly)**

From Figure 1b, public searches on the keyword "COPD (Chronic Obstructive Pulmonary Disease)", there is a gradual increase in searches from year to year, although still accompanied by considerable weekly fluctuations. This pattern indicates an increased public awareness of lung diseases, possibly because of the COVID-19 pandemic expanding public understanding of long-term respiratory diseases.

The keyword "*Diare*" shows a sharply fluctuating search pattern over time, with some significant peaks and rapid changes. This reflects that people tend to search for information on diarrhea seasonally, possibly triggered by a spike in cases during the rainy season or an outbreak of a sanitation-based disease. As an infectious disease that often affects children and vulnerable groups, diarrhea seems to trigger strong search reactions at certain times but does not show sustained awareness outside of those periods.

Meanwhile, the search trend for "*Kanker Paru-Paru*" shows a gradual decline since 2020 with some incidental peaks that tend not to recur. This phenomenon indicates that public awareness of lung cancer is still uneven and tends to be triggered by external momentum such as celebrity news or annual health campaigns. In fact, epidemiologically, lung cancer is one of the highest causes of death that should receive serious attention from the promotive and preventive side.

An interesting pattern emerges in the keyword "*Neonatal*", which shows an increasing search trend from year to year. This increase shows a growing public interest and concern for newborn health issues, which could be due to increased access to digital information among young mothers, as well as more active promotion of maternal and child health services through digital channels.

Finally, "*Pneumonia*" shows a flat and low search pattern, except for two sharp spikes that are likely related to the peak of the COVID-19 pandemic and the emergence of new variants targeting the lower respiratory tract. Outside of these two periods, searches for pneumonia are relatively low, suggesting that the disease is only of concern to the public when in the context of a pandemic or major outbreak.

In general, the results from both plots show that people's search patterns on health issues are strongly influenced by momentum, seasonality, media and situational context. Diseases that are pandemic-related or receive high exposure in the media tend to show a sharp spike in searches, while non-communicable diseases such as COPD show a slow but consistent increase in interest. In contrast, infectious diseases such as diarrhea and pneumonia tend to be searched for reactively, when there is an increase in cases or an outbreak. This finding confirms the importance of health communication strategies that are based on digital data and adapted to the dynamics of public search behavior, so that health information can reach the public in a timely manner, not only in times of crisis but also preventively and sustainably. The implications of these results are very relevant for planning public health campaigns that are more strategic, long-term, and adaptive to changes in public attention patterns.

After conducting descriptive analysis to understand the characteristics of the data, the next step is to estimate the model parameters using Loop Prophet. In this study, the changepoint $S = 8$, annual seasonality $N_{\text{tahunan}} = 10$ and weekly seasonality $N_{\text{mingguan}} = 3$ were selected, so the estimated $\boldsymbol{\beta}$ parameters were $2N_{\text{tahunan}} + 2N_{\text{mingguan}} = 2(10) + 2(3) = 26$. The model parameter estimation results for each variable $z$ are shown in Table 2, which includes the value of $k$ as the initial trend slope, $m$ as the trend offset, $\boldsymbol{\delta}$ which shows the magnitude of the slope change at each changepoint, $\boldsymbol{\beta}$ s the Fourier coefficients that form the annual and weekly seasonal patterns, and $\sigma$ as the standard deviation of the model residuals. These values become numerical representations of the trend, seasonal, and error components that Prophet uses to generate forecasts.

The parameter estimation results obtained are detailed in Table 2, Figure 2, and Figure 3, which present the results of the Loop Prophet model parameter estimation for various disease keywords that are the main causes of death in Indonesia. The parameter estimation values of $k_z$, $m_z$, $\sigma_z$, as well as the estimates of $\boldsymbol{\delta}$, and $\boldsymbol{\beta}$ vary between types of disease keywords, indicating differences in the pattern of public search trends. For example, the keyword Diabetes has a positive $k$ estimate (0.0390) with a relatively low σ estimate (0,0695), indicating a search trend that tends to increase steadily. This is also the case for the keywords "*Tuberkulosis", "COPD"," Sirosis Hati", "Neonatal*", dan "*Pneumonia*". In contrast, *Diare* shows a negative $k$ estimate (-0.2747) with a small σ estimate (0.0590), indicating a downward trend in the search trend at a rate of 0.2747. Non-communicable diseases such as *Stroke, Diare,* and *Serangan Jantung* show negative $k$ estimates, but with high variation in search. On the other hand, infectious diseases such as *Penyakit Paru and Kanker Paru-Paru* also have negative $k$ negatif and relatively larger σ relatif estimates, indicating a decreasing search pattern with higher fluctuations.

**Table 2 Loop Prophet Parameter Estimation Results**

| Disease Keywords | Parameter Estimation Values | | |
|---|---|---|---|
| | $k_z$ | $m_z$ | $\sigma_z$ |
| *Serangan Jantung* | -0.196705424 | 0.378927 | 0.06352892 |
| *Stroke* | -0.148460585 | 0.67799 | 0.041862614 |
| *Diabetes* | 0.039011636 | 0.625003 | 0.069560522 |
| *Tuberkulosis* | 0.13329390 | 0.14517674 | 0.10151681 |
| *Sirosis Hati* | 0.116974519 | 0.086248 | 0.086821521 |
| *Penyakit Paru* | -0.288927561 | 0.832943 | 0.072382002 |
| *COPD* | 0.292806931 | 0.443978 | 0.099622605 |
| *Kanker Paru-Paru* | -0.632913064 | 0.825442 | 0.095710417 |
| *Neonatal* | 0.41723421 | 0.331738 | 0.0936311 |
| *Pneumonia* | 0.120504644 | 0.171076 | 0.085662636 |
| *Diare* | -0.274652221 | 0.124587 | 0.05903535 |

Source: Loop Prophet Parameter Estimation Results Using R

The estimated **δ** parameter in Figure 2 shows that some diseases experience different search trend changepoints. The eight $\boldsymbol{\delta}$ parameters corresponding to the number of changepoints ($S = 8$) indicate the magnitude of the change in trend slope at certain changepoints; positive $\boldsymbol{\delta}$ estimation values indicate trend acceleration, while negative values indicate trend deceleration.

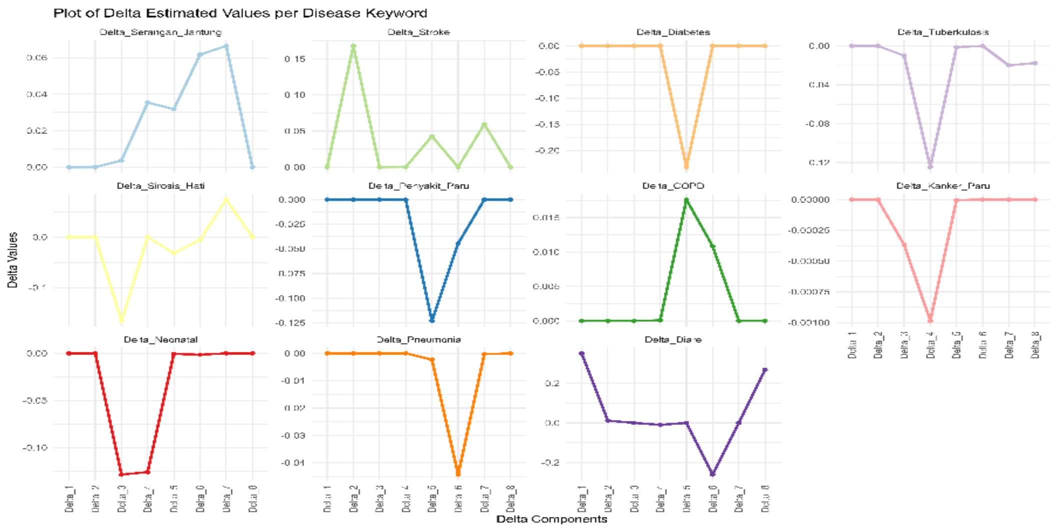


**Figure 2. Plot of Estimated Value of $\delta$ of Prophet Loop Model Per Keyword of Diseases Causing Highest Mortality in Indonesia**

The estimated **δ** values for the keywords "*Serangan Jantung", "Stroke*", and "*Sirosis Hati*" are generally positive or fluctuate up and down with considerable amplitude. This means that the trend of public search tends to increase, although there are slight fluctuations. This consistency may reflect the gradual increase in public awareness. The keywords "*Diabetes", "Kanker Paru", "Penyakit Paru*" show sharp negative **δ** estimation values in certain periods (e.g. *Diabetes* -0.21, *Penyakit Paru* -0.12, *Kanker Paru* around -0.0001). This indicates a period when public attention to the disease decreases dramatically. After that, the value returns to 0, meaning that the trend stabilizes again. And for "*Diare*", "*Pneumonia*", and "*Tuberkulosis*", the estimated **δ** values are deeply negative and then experience a large positive rebound. This means that the search trend is reactive to issues or outbreaks, not a long-term trend.

For "Neonatal", the estimated **δ** value drops sharply to -0.129 and then stabilizes again. This indicates a momentary public attention (e.g. triggered by case news). For "COPD", it increases and decreases sharply until the estimated **δ** value is close to zero. This means that the trend is stable and not much influenced by momentum.

The estimation of the 26 parameters $\boldsymbol{\beta}$ represents the Fourier coefficients that construct the annual and weekly seasonal patterns, with the variation in the estimated values of $\boldsymbol{\beta}$ reflecting the amplitude and phase (wave shift) of these seasonal patterns. From the seasonal perspective, the estimated $\boldsymbol{\beta}$ parameters also reveal interesting variations, as shown in Figure 3. Disease-related keywords such as *Stroke, Diabetes, Penyakit Paru, Neonatal,* and *COPD* m exhibit weak seasonal patterns, indicated by stable $\boldsymbol{\beta}$ values around zero to $\pm$0,05. This suggests that public search trends for these diseases are not strongly influenced by specific seasons or periods but instead remain relatively consistent throughout the year.

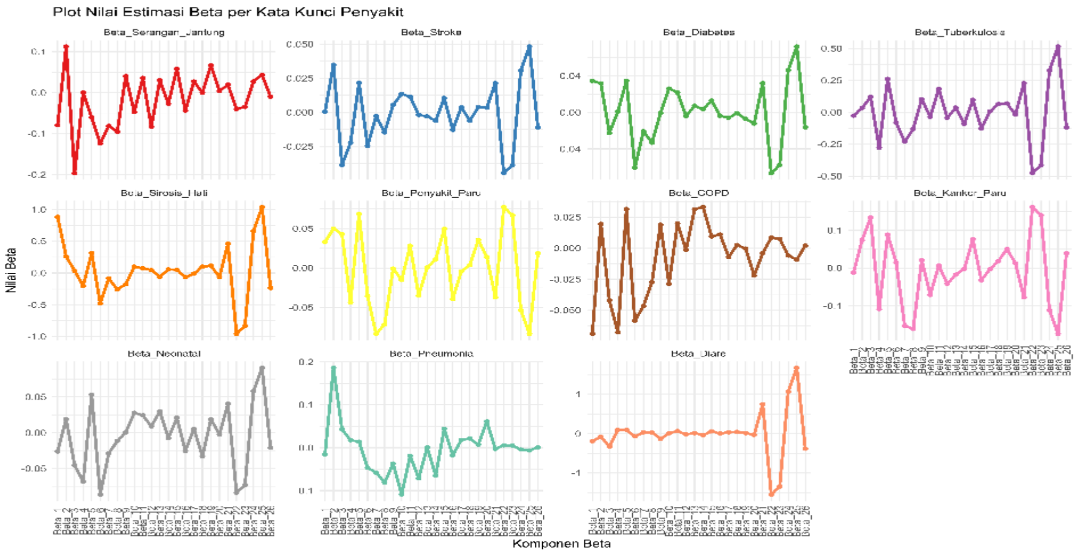


**Figure 3. Plot of Estimated Value of $\beta$ Prophet Loop Model Per Keyword of Diseases Causing Highest Mortality in Indonesia**

Thus, the combination of positive and negative estimated values for the β parameter illustrates that the trend of disease search is not only determined by long-term patterns, but is also affected by seasonal fluctuations, i.e. at certain times the search increases sharply while at other times it decreases.

The performance of the public search prediction model for diseases in Indonesia was evaluated using three forecasting accuracy measures: MAPE, NMAE, and NRMSE. Table 3 presents the model evaluation results for each disease keyword, based on each accuracy measure to facilitate interpretation of model quality.

Based on Table 3, the results of evaluating the Loop Prophet model against various disease keywords in Indonesia show that the accuracy of the model varies depending on the type of disease being analyzed.

**Table 3 Prophet Loop Model Accuracy Evaluation Per Disease Keyword (in percent)**

| Disease Keywords | NMAE | NRMSE | MAPE | Model Categories |
|---|---|---|---|---|
| *Serangan Jantung* | 13.701528 | 18.532183 | 13.407745 | Good |
| *Stroke* | 4.070858 | 5.456134 | 4.078192 | Excellent |
| *Diabetes* | 7.371650 | 9.194460 | 7.506121 | Excellent |
| *Tuberkulosis* | 15.576272 | 21.113151 | 17.736408 | Fairly Good |
| *Sirosis Hati* | 15.853879 | 19.987965 | 17.055368 | Good |
| *Penyakit Paru* | 11.633907 | 14.454300 | 12.003469 | Good |
| *COPD* | 13.653578 | 17.295911 | 14.578987 | Good |
| *Kanker Paru-Paru* | 29.973561 | 39.954285 | Inf | Poor |
| *Neonatal* | 10.890245 | 15.083067 | 12.085282 | Good |
| *Pneumonia* | 21.319549 | 38.264519 | 20.461053 | Poor |
| *Diare* | 6.133652 | 8.227786 | 6.121636 | Excellent |

Source: Prophet Loop Model Evaluation Using R

The model performed Excellent on the keywords *Stroke, Diabetes* and *Diare*, characterized by low NMAE and NRMSE values (<10% each) and MAPE below 10%. This indicates that people's search patterns for these two diseases are quite stable and easily predicted by the model. In contrast, the poorest model performance was shown for the keywords *Kanker Paru-Paru,* and *Pneumonia*, with very high NMAE and NRMSE values and infinite MAPE values (Inf), indicating a mismatch between the model and the actual data. This is most likely due to extreme fluctuations or many zero values in the search data, making it difficult for the model to recognize consistent patterns.

The *Tuberkulosis* keyword is categorized as a fairly good model with an NRMSE value of more than 20%. This shows that although the model can capture the general pattern of search trends, there is still a relatively large deviation between the forecasting results and the actual data. Thus, the forecasting results for this keyword can be used as an overview of the general trend, but it is not optimal if used as a reference for long-term predictions.

Meanwhile, for diseases such as *Serangan Jantung, Sirosis Hati, Penyakit Paru, COPD*, and *Neonatal*, the model performs well, with prediction error rates that are still within reasonable limits and acceptable in the context of modeling trends in Google Trends data. These results generally indicate that the Loop Prophet model works optimally for diseases with consistent trends, while its performance degrades when dealing with unstable data.

Thus, these results emphasize the importance of evaluating models per keyword individually before drawing general conclusions, as well as the need for specialized approaches or additional preprocessing for diseases characterized by erratic search data.

Figure 4 presents the results of forecasting public search interest on Google for some of the diseases that cause the highest mortality in Indonesia. This visualization illustrates the fluctuation of the number of searches for 52 weeks from August 2025 to July 2026 for each disease based on the Loop Prophet predictive model with model evaluation categories of "Excellent" and "Good". Each line represents one disease with a different color, making it easier to observe seasonal patterns, trends, and differences in search rates between diseases.

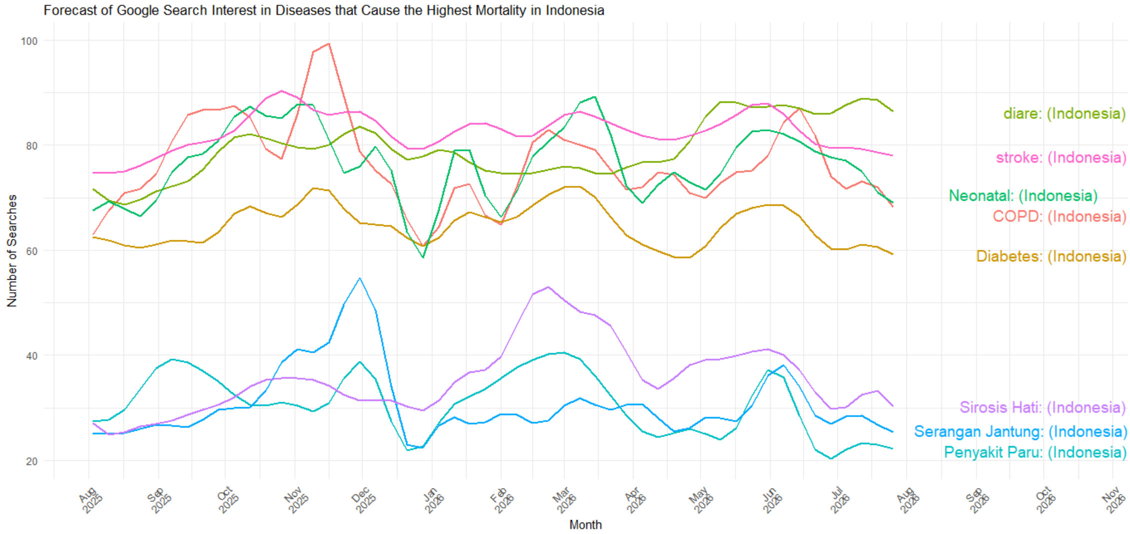


**Figure 4. Forecast Loop Prophet Model of Google Searches for Keyword Diseases that Cause Highest Mortality in Indonesia**

Based on the forecasting graph (Figure 4), Google search interest in the diseases that cause the highest mortality in Indonesia, there are different trend patterns for each disease. disease keywords such as *diare, stroke*, and *COPD* are estimated to occupy the top position in terms of the number of searches, with significant fluctuations especially in certain months such as December, March and June. This indicates that people's interest in these disease keywords is seasonal and most likely influenced by environmental factors, media campaigns, or an increase in cases.

Meanwhile, *diabetes* and *neonatal* care tended to show a more stable trend throughout the year, although there were still slight increases in some months. This could reflect the characteristics of non-communicable diseases where public awareness of them tends to be consistent. On the other hand, disease keywords such as *sirosis hati, serangan jantung,* and *penyakit paru* showed lower search volumes overall.

In general, this graph (Figure 4) shows that seasonal infectious diseases such as diarrhea are easier to predict due to their relatively consistent up-and-down patterns, while non-communicable diseases such as *diabetes* and *serangan jantung* require a different prediction approach as their search trends are flatter and more stable. This information is important for designing timely public health communication strategies, by optimizing information dissemination according to peaks in public interest.

Figure 5 shows the results of forecasting Google search interest in the eight diseases that cause the highest mortality in Indonesia, categorized as having "Good" and "Excellent" model performance. Each panel shows a comparison between the actual data (in red) and the prediction results using the Loop Prophet model (in blue) for each disease keyword over the period 2020 to mid-2026. This visualization illustrates the extent to which the model can follow historical search patterns and forecast future trends, which is useful in supporting early detection policies and digital data-based public health communication planning.

Based on Figure 5, a comparison between the actual data and the prediction results of Google search interest in nine disease keywords that cause the highest mortality in Indonesia during the period 2020 to 2026, the Loop Prophet model can follow historical trend patterns quite well for most disease keywords. The blue line (prediction) generally follows the pattern of the red line (actual data), especially for disease keywords such as *Diabetes, Diare,* and *Stroke*, which shows the stability of search patterns and relatively clear trends. This shows that the model is quite reliable in mapping and predicting search fluctuations for keywords of non-communicable diseases or those with consistent patterns.

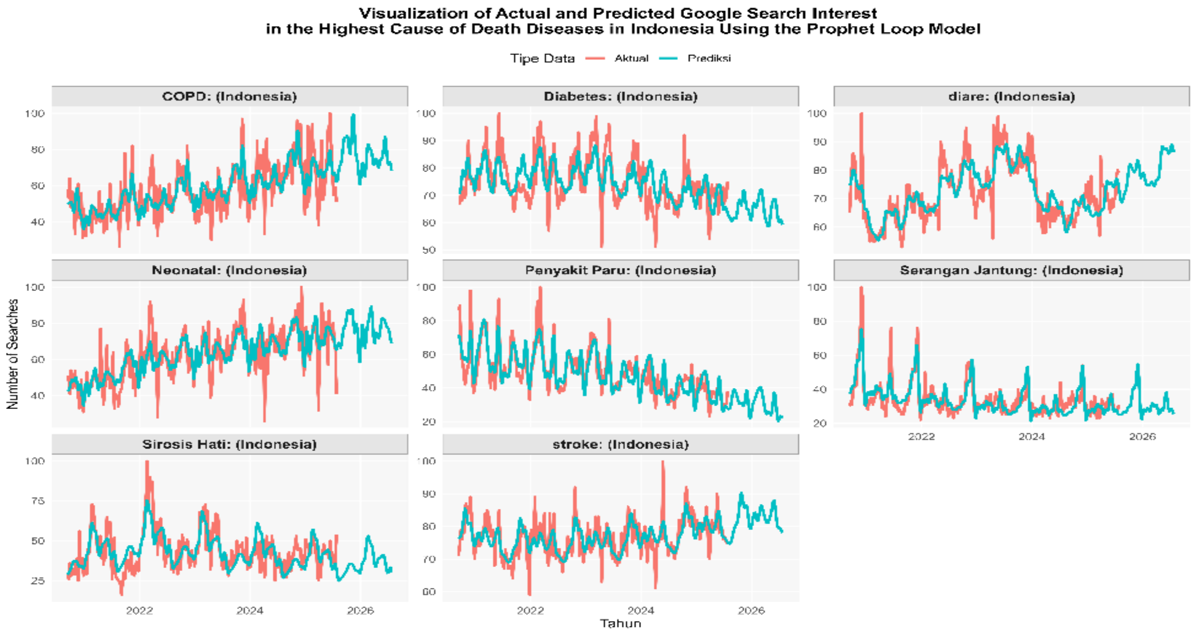


**Figure 5. Visualization of Actual and Predicted Data of Google Search Interest in Diseases that Cause the Highest Mortality in Indonesia Using the Prophet Loop Model (2020-2026)**

However, for some other disease keywords such as *Serangan Jantung, Penyakit Paru,* and *Sirosis Hati*, the model showed slight deviations between predictions and actual data, especially for sharp spikes that the model could not accurately anticipate. This is most likely due to sudden search spikes that are contextual or related to current issues (e.g. health news or local outbreaks) that are not reflected in historical patterns. Thus, the use of the Loop Prophet model used in this analysis is sufficiently helpful to see an overview of search trends for disease keywords on Google.

## 5. Conclusion

Based on the results of the analysis using the Loop Prophet model on weekly Google Trends data, it can be concluded that this model is able to provide a reliable prediction of the search interest of keywords for diseases that cause the highest mortality in Indonesia, especially for diseases with stable search patterns such as *diabetes, stroke,* and *diare*. The model evaluation category of "Excellent" is achieved for some disease keywords, indicating that the search patterns tend to be regular and can be predicted well. However, for disease keywords that have fluctuating search trends, such as *tuberkulosis, kanker paru-paru,* and *pneumonia*, the accuracy of the model still needs to be further tested as the evaluation results show less than optimal performance. The findings also show that people's search behavior is strongly influenced by situational contexts, such as seasons, media coverage, or extraordinary events such as pandemics.

### REFERENCES

1. Aji Muhawarman. (2024, November 18). *Pneumonia Terus Ancam Anak-anak* [Rilis Berita]. https://kemkes.go.id/id/pneumonia-terus-ancam-anak-anak
2. Brykin, D. (2024). Sales Forecasting Models: Comparison between ARIMA, LSTM and Prophet. *Journal of Computer Science*, *20*(10), 1222–1230. https://doi.org/10.3844/jcssp.2024.1222.1230
3. Despotovic, M., Nedic, V. M., Despotovic, S., & Cvetanovic, S. (2016). Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renewable and Sustainable Energy Reviews*, *56*, 246–260. https://doi.org/10.1016/j.rser.2015.11.058
4. FK Unsoed. (2013, February 13). *Pemeriksaan Paru*. Lab. Ketrampilan Medik PPD Unsoed. https://fk.unsoed.ac.id/wp-content/uploads/modul%20labskill/modul%20ganjil%20I/Ganjil%20I%20-%20Pemeriksaan%20Fisik%20paru.pdf
5. Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*, 1–28.
6. Indian Institute of Technology Madras. (2022). *Maximum a posteriori estimation [PDF lecture slides]*. https://www.cse.iitm.ac.in/~vplab/courses/PRML/MAP.pdf
7. Jannah, Y. (2018). *Asuhan Keperawatan Pada Pasien Dengan Penyakit Jantung Koroner Di Ruangan CVCU RSUP Dr. M. Djamil Padang* [Karya Tulis Ilmiah, Poltekkes Kemenkes Padang]. https://pustaka.poltekkes-pdg.ac.id/index.php?p=fstream-pdf&fid=1423&bid=5249
8. Kambezidis, H. D. (2012). 3.02—The Solar Resource. In *Earth Systems and Environmental Sciences—Comprehensive Renewable Energy* (Vol. 3, pp. 27–84). Elsevier. https://doi.org/10.1016/B978-0-08-087872-0.00302-4
9. Kementerian Kesehatan Republik Indonesia. (2019a). *Pedoman Nasional Pelayanan Kedokteran Tata Laksana Tuberkulosis* [Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/755/2019]. Kemenkes RI.

**10.** Kementerian Kesehatan Republik Indonesia. (2019b). *Pedoman Nasional Pelayanan Kedokteran Tata Laksana Stroke* [Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/394/2019.]. Kemenkes RI.

**11.** Kementerian Kesehatan Republik Indonesia. (2022). *Sirosis*. Topik Kesehatan. https://ayosehat.kemkes.go.id/topik-penyakit/penyakit-pencernaan/sirosis

**12.** Kementerian Kesehatan Republik Indonesia. (2023). *Diare* [Topik Kesehatan]. Ayo Sehat Kemenkes. https://ayosehat.kemkes.go.id/topik-penyakit/penyakit-pencernaan/diare

**13.** Kementerian Kesehatan Republik Indonesia. (2024). *Pedoman Nasional Pelayanan Klinis Tata Laksana Diabetes Melitus Pada Anak* [Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/2009/2024]. Kemenkes RI.

**14.** Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth Scientific.

**15.** Lütkepohl, H. (2007). *New introduction to multiple time series analysis: With ... 36 tables* (1. ed., corr. 2. print). Springer.

**16.** Oktavia, F., & Witanti, A. (2024). *Implementasi Prophet Forecasting Model Dalam Prediksi Kualitas Udara Daerah Istimewa Yogyakarta*. *11*(1). https://doi.org/https://doi.org/10.35957/jatisi.v11i1.6804

**17.** Oo, Z. Z., & Phyu, S. (2020). Time Series Prediction Based on Facebook Prophet: A Case Study, Temperature Forecasting in Myintkyina. *International Journal of Applied Mathematics Electronics and Computers*, *8*(4), 263–267. https://doi.org/10.18100/ijamec.816894

**18.** Paramitha, P. (2020). *Respon Pasien Penyakit Paru Obstruktif Kronis (PPOK) Dengan Gangguan Pemenuhan Kebutuhan Oksigenasi Terhadap Penerapan Fisioterapi Dada Di Rumah Sakit Khusus Paru "Respira"* [Diploma thesis, Poltekkes Kemenkes Yogyakarta]. https://eprints.poltekkesjogja.ac.id/2512/4/Chapter2.pdf

**19.** Rahmawati, E. (2016). *Sistem Pakar Diagnosis Penyakit Paru-Paru Menggunakan Metode Forward Chaining*. *8*(2).

**20.** RSU Tjokronegoro. (2021, August 2). *Kanker Paru*. RSU Tjokronegoro. https://rsutjokronegoro.purworejokab.go.id/download/file/KANKER_PARU.pdf

**21.** Taylan, A. S. (2023). Automatic Time Series Analysis and Forecasting using the Ata Method. *CRAN R*. https://cran.r-project.org/web/packages/ATAforecasting/ATAforecasting.pdf

**22.** Taylor, S., & Letham, B. (2017). *prophet: Automatic Forecasting Procedure* (p. 1.0) [Dataset]. https://doi.org/10.32614/CRAN.package.prophet

**23.** Tidy, C. (2023, May 9). *Stillbirth and neonatal death* [Professional articles]. Paediatrics. https://patient.info/doctor/paediatrics/stillbirth-and-neonatal-death

**24.** Tim detikHealth. (2022, December 22). *WHO Rilis 10 Penyakit Mematikan di Dunia, Jantung hingga Diare* [Berita]. detikSulsel. https://www.detik.com/sulsel/berita/d-6476113/who-rilis-10-penyakit-mematikan-di-dunia-jantung-hingga-diare

**25.** UNICEF. (2024). *A child dies of pneumonia every 43 seconds*. UNICEF: For Every Child. https://data.unicef.org/topic/child-health/pneumonia/

**26.** United Nations. (2015). *Global Sustainable Development Report, 2015 edition (Advanced Unedited Version)*. United Nations General Assembly. https://sustainabledevelopment.un.org/content/documents/1758GSDR%202015%20Advance%20Unedited%20Version.pdf

**27.** World Health Organization. (2024). *Global Health Estimates 2021: Deaths by cause, age, sex, by country and by region, 2000-2021.* Global Health Observatory. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death

**28.** Žunić, E., Korjenić, K., Hodžić, K., & Đonko, D. (2020). Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data. *International Journal of Computer Science and Information Technology*, *12*(2), 23–36. https://doi.org/10.5121/ijcsit.2020.12203