Counseling & Humanities Review
http://bk.ppj.unp.ac.id/index.php/chr/index

Counseling & Humanities Review
Vol. 6, No. 1 Month 2026, pp. 101-117
p-ISSN: 2798-3188, e-ISSN: 2798-0316 | | http://bk.ppj.unp.ac.id/index.php/chr
DOI: https://10.24036/0001408chr%Y

Received (June 12th 2026); Accepted (July 10th 2026); Published (July 16th 2026)

Examining Counseling Search Interest Through Mental Health-Related Search Trends and Machine Learning Approaches

Mutia Yollanda[1*)], Ghea Weisha[1], Ade Herdian Putra[2]
[1] Department of Mathematics and Data Science, Universitas Andalas, Indonesia,
[2] Department of Guidance and Counseling, Universitas Negeri Padang, Indonesia
*Corresponding author, e-mail: mutia.yollanda@sci.unand.ac.id

Abstract
Mental health-related concerns have become increasingly visible in digital environments, where online search behavior reflects public awareness and information-seeking patterns. Examining temporal changes in mental health-related searches may provide valuable insights into emerging public concerns and support evidence-based planning. This study analyzes monthly Google Trends data on counseling and related mental health topics from January 2011 to June 2026. An XGBoost regression model was developed using historical counseling search interest, lagged variables, and related mental health indicators, including depression, anxiety, stress, suicide, and mental health search trends. Temporal analysis, feature importance evaluation, and SHAP interpretation were applied to investigate trend patterns and explain model predictions. The findings revealed a substantial increase in counseling-related search interest, particularly after 2018, with several periods showing notable fluctuations. The XGBoost model demonstrated satisfactory predictive performance, achieving a mean absolute percentage error of 18.78 percent and a coefficient of determination of 0.658. Historical counseling search variables were identified as the most influential predictors, indicating that previous search behavior plays a dominant role in forecasting future trends. The study highlights the potential of interpretable machine learning approaches for monitoring digital mental health trends and supporting data-driven public health strategies.

Keywords: Google Trends, counseling, mental health, XGBoost, time series forecasting

Introduction

The rapid development of digital platforms has transformed how individuals access and seek information about health, psychological conditions, and social issues. The increasing availability of online behavioral data has created new opportunities for population-level monitoring through digital traces generated from internet activities. Compared with conventional survey approaches, which are often constrained by limited observation periods, reporting bias, and delayed data availability, digital search behavior provides continuous information that reflects public attention and information-seeking patterns over time. In mental health research, these digital footprints have increasingly been utilized to complement traditional epidemiological approaches by identifying changes in public concerns, awareness, and psychological information needs. Recent evidence further suggests that mental health is shaped by broader psychosocial and environmental stressors, reinforcing the importance of monitoring population responses to changing social conditions through timely and scalable data sources (Putra & Afiati, 2026).

Google Trends has become one of the most widely used platforms for analyzing population search behavior and has contributed to the development of mental health infodemiology. This approach examines the distribution and determinants of health-related information in electronic environments to understand population behavior. Google Trends provides anonymized relative search volume data, enabling researchers to investigate sensitive mental health topics that may be difficult to capture through conventional methods. However, Google Trends represents normalized relative search interest rather than absolute search volume; therefore, interpretation requires careful consideration of keywords, categories, geographic scope, and observation periods (Alibudbud, 2023).

Previous studies have demonstrated the potential of Google Trends for monitoring mental health-related phenomena. Search patterns associated with depression have shown temporal and geographical variations consistent with established epidemiological characteristics, indicating that online search behavior may provide meaningful signals of population mental health concerns (Wang et al., 2022). Similarly, online symptom search data have been utilized to estimate suicide-related trends and provide more timely information compared with conventional surveillance systems that often experience reporting delays (Sumner et al., 2023). Recent studies have further extended Google Trends applications through predictive approaches. Machine learning models incorporating Google search patterns have demonstrated the potential to predict suicide attempt trends (Kachlik et al., 2025), while selected internet search queries, combined with advanced statistical approaches, have shown promising performance in forecasting suicide trends across different contexts (Takane, 2025).

Although Google Trends provides valuable behavioral indicators, search data are inherently sequential and dynamic, containing temporal dependencies, seasonal variations, and nonlinear patterns. Therefore, time series forecasting methods are required to capture these complex structures. Previous research has demonstrated the effectiveness of time series approaches in various forecasting applications. Seasonal autoregressive models have been applied for tourism demand forecasting (Yollanda & Devianto, 2020), while artificial neural networks have been utilized for financial time series modeling involving nonlinear relationships (Yollanda et al., 2018). Furthermore, nonlinear and hybrid time series models have shown improved forecasting performance by combining different modeling structures (Arif et al., 2022; Devianto et al., 2022). Advanced forecasting approaches incorporating fractional models, fuzzy time series, and neural networks have also demonstrated the ability to capture complex temporal patterns (Arif et al., 2024; Devianto et al., 2023; Devianto, Permana, et al., 2024; Devianto, Wahyuni, et al., 2024; Yollanda et al., 2023).

The development of machine learning algorithms has further expanded forecasting capabilities for datasets involving nonlinear relationships and multiple predictors. Unlike conventional statistical models that depend on predefined assumptions, machine learning methods can identify complex patterns through data-driven learning. Extreme Gradient Boosting (XGBoost) has gained substantial attention for its high predictive performance, ensemble tree structure, and regularization mechanisms that reduce model complexity (Chen & Guestrin, 2016). These characteristics make XGBoost suitable for analyzing behavioral datasets, including digital mental health indicators derived from online search activity.

Recent mental health studies have demonstrated the effectiveness of machine learning approaches for identifying complex relationships between predictors and psychological outcomes. The increasing recognition of work-related psychological distress further highlights the need for scalable monitoring systems capable of detecting changes in mental health information-seeking behavior beyond traditional clinical settings (Putra, Ardi, & Afiati, 2026). XGBoost-based models have been applied to predict mental health severity and identify influential factors among university students (Kong et al., 2025). Similarly, machine learning-based depression prediction studies have shown strong predictive performance, with XGBoost demonstrating the capability in identifying important depression-related predictors (Vu et al., 2025; Xia et al., 2025). However, predictive accuracy alone is insufficient for sensitive applications such as mental health, where understanding model decisions is essential.

Explainable Artificial Intelligence (XAI) has therefore become increasingly important to improve the transparency and interpretability of complex machine learning models. SHapley Additive exPlanations (SHAP) provides a framework for quantifying the contribution of individual predictors to model predictions using game-theoretic principles (Lundberg & Lee, 2017). Machine learning and SHAP have been applied in various mental health settings, including depression prediction, anxiety assessment, suicide risk identification, and mental health profiling (Benito et al., 2024; Manyam et al., 2025; Mimikou et al., 2025; Niu et al., 2025). Moreover, explainable AI approaches for healthcare time series and remote health applications emphasize the importance of interpretable predictions when models are applied to sensitive human conditions (Di Martino & Delmastro, 2023; Joyce et al., 2023). Despite the growing application of Google Trends and machine learning in mental health research, previous studies have mainly focused on identifying psychiatric disorders, estimating depression prevalence, or predicting suicide-related outcomes. At the same time, recent evidence from low- and middle-income countries has highlighted the importance of developing context-specific approaches for mental health surveillance and prevention, reflecting the need for evidence that captures changing population mental health needs in diverse socioeconomic settings (Putra, Ardi, Ifdil, et al., 2026). Nevertheless, limited attention has been given to forecasting counseling-related search interest as an indicator of population-level demand for psychological support. Furthermore, the contribution of different mental health-related search patterns, including depression, anxiety, stress, suicide, and general mental health, toward counseling search trends remains insufficiently explored using longitudinal forecasting frameworks.

Therefore, this study aims to develop an explainable machine-learning forecasting framework for counseling-related search interest, using monthly Google Trends data. Counseling search interest is modeled as the target variable, while related mental health search trends, lag variables, temporal components, and seasonal characteristics are incorporated as predictors. XGBoost is employed to capture nonlinear relationships and temporal dependencies, while SHAP analysis is applied to identify the contribution of each predictor. This study contributes to digital mental health surveillance by demonstrating how search behavior data can be integrated with machine learning and explainable artificial intelligence to understand and predict population-level interest in information about psychological counseling.

Method

Data Collection and Preparation.

This study employed monthly time-series data from Google Trends covering January 2011 to June 2026, yielding 186 observations. The objective of this study was to examine and predict counseling search interest based on mental health-related search trends using a machine learning approach. The dependent variable was counseling search interest, while the explanatory variables consisted of search interest related to depression, anxiety, stress, suicide, and mental health. Google Trends provides normalized search interest values ranging from 0 to 100, where 100 represents the peak search popularity of a term during the selected period, while lower values indicate proportionally lower search interest. Therefore, the counseling variable does not represent actual counseling utilization or the number of individuals receiving counseling services; rather, it reflects population-level attention and information-seeking behavior related to counseling topics.

The selection of mental health-related search variables was based on the assumption that individuals experiencing psychological concerns may seek information regarding emotional conditions and available support resources before accessing professional assistance. Search terms related to depression, anxiety, stress, suicide, and mental health were therefore incorporated as potential predictors of counseling search interest. Table 1 presents the variables and data sources used in this study.

Table 1. Variables and Data Sources

Variable	Description	Unit
Counselling	Search interest related to counseling	Google Trends index (0–100)
Depression	Search interest related to depression	Google Trends index (0–100)
Anxiety	Search interest related to anxiety	Google Trends index (0–100)
Stress	Search interest related to stress	Google Trends index (0–100)
Suicide	Search interest related to suicide	Google Trends index (0–100)
Mental Health	Search interest related to mental health	Google Trends index (0–100)

Before model development, the dataset underwent several preparation procedures. First, all observations were arranged chronologically to preserve the data's temporal structure. Second, data completeness was examined to ensure that no observations were missing. Third, exploratory visualization was conducted to identify general trends, fluctuations, and possible seasonal patterns in counseling search interest. Because time-series observations are inherently dependent, random data splitting was avoided. Instead, chronological ordering was maintained throughout the modeling process to prevent information leakage from future observations into the training dataset.

Feature Engineering

Feature engineering was performed to convert the original sequential Google Trends data into informative predictive variables capturing temporal dependencies, long-term trends, and recurring seasonal patterns. Since search interest data are generated over time, the model requires additional temporal information to capture behavioral changes that cannot be represented by the original variables alone. Therefore, three groups of features were constructed: deterministic trend features, seasonal features, and lagged historical features.

a. Trend Feature Construction.
   A deterministic time-trend variable was introduced to capture gradual structural changes in counseling-related search interest over the observation period. The inclusion of a trend component enables the model to detect whether public interest in counseling exhibits persistent upward or downward trends over time, rather than relying solely on short-term fluctuations. The trend variable was formulated as:
   $$Trend_t = t$$
   where $Trend_t$ represents the sequential time index at period t, t represents the monthly observation sequence, and T represents the total number of observations: t=1,2,3,...,T. For this study, the trend variable was assigned sequential monthly values from January 2011 to June 2026. The feature provides additional temporal information that helps the model distinguish between temporary variations and long-term changes in counseling search behavior.

b. Seasonal Feature Construction.
   Monthly search behavior may contain periodic patterns caused by recurring annual events, changes in social conditions, academic cycles, holidays, and other time-dependent factors. However, directly including month categories may increase model complexity and introduce discontinuities in relationships between adjacent months. Therefore, the Fourier transformation was applied to represent seasonal variation using continuous cyclical variables. The seasonal components were constructed using sine and cosine transformations:
   $$Seasonal_{sin,t} = sin\left(\frac{2\pi M_t}{12}\right)$$
   and
   $$Seasonal_{cos,t} = cos\left(\frac{2\pi M_t}{12}\right)$$
   where $M_t$represents the month number ranging from 1 to 12, 2πrepresents a complete circular cycle, and 12 represents the annual seasonal period.

The sine and cosine transformations allow the model to identify cyclical relationships between months. For example, December and January are close in the annual cycle despite having different numerical month values. The transformation preserves this continuity and enables XGBoost to capture nonlinear seasonal fluctuations in counseling-related search interest.

c. Lag Feature Construction
Because search behavior is influenced by previous public attention and information-seeking patterns, lagged variables were created to represent temporal dependence between past and current counseling search interest. Lag features allow the model to learn whether previous search activity contributes to future changes in search behavior. The lagged counseling variable was defined as:

$$C_{t-k} = C(t-k)$$

where C_t represents the counseling search interest at the time t and k represents the lag period. This study incorporated five lag structures:

$$k = \{1,2,3,6,12\}$$

The selected lag periods represent different memory effects:

- Lag 1 month represents the immediate short-term influence from the previous month.
- Lag 2 and 3 months represent delayed behavioral responses.
- Lag 6 months represents medium-term fluctuations.
- Lag 12 months represents annual recurring patterns.

The inclusion of lagged variables enables the model to incorporate historical search dynamics and improve forecasting capability by learning temporal relationships within the counseling search series.

The predictor variables were selected based on both theoretical and empirical considerations. Mental health-related search terms, including depression, anxiety, stress, suicide, and mental health, were chosen because previous studies have shown that these topics frequently precede or accompany information-seeking behavior related to psychological support. To capture temporal dynamics, deterministic trend variables, seasonal Fourier components, and lagged counseling search values were additionally incorporated. These engineered features enable the model to learn long-term trends, recurring seasonal fluctuations, and short-, medium-, and long-term temporal dependencies while reducing information loss associated with using the original Google Trends series alone.

Predictive Modeling Using XGBoost

This study employed the Extreme Gradient Boosting (XGBoost) algorithm to forecast future counseling-related search interest. XGBoost was selected because it can model nonlinear relationships, handle interactions among multiple predictors, and capture complex patterns commonly found in time-series data. Unlike traditional statistical approaches that assume linear relationships, XGBoost constructs an ensemble of decision trees sequentially. Each additional tree is developed to minimize the prediction errors generated by previous trees, allowing the model to gradually improve predictive accuracy. The prediction function of XGBoost can be expressed as:

$$\hat{Y}_t = \sum_{m=1}^{M} f_m\,(X_t)$$

where $\hat{Y}_t$represents the predicted counseling search interest, Mrepresents the total number of decision trees, $f_m$represents the prediction contribution of the m-th tree, and $X_t$represents the input feature vector at time t. The optimization process aims to minimize the objective function:

$$Obj = \sum_{i=1}^{n} L(\,y_i, \hat{y}_i) + \sum_{m=1}^{M} \Omega(\,f_m)$$

where $L(y_i, \hat{y}_i)$represents the prediction loss between observed and predicted values, $y_i$represents the actual counseling search interest, $\hat{y}_i$represents the predicted value, and $\Omega(f_m)$represents the regularization term used to control model complexity.The regularization component is formulated as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

where Trepresents the number of leaves in a decision tree, $w_j$represents leaf weights, γcontrols tree complexity, and λcontrols weight regularization. The regularization mechanism reduces overfitting by penalizing overly complex trees and improving the model's ability to generalize to unseen observations. The XGBoost model was configured with 500 estimators, a learning rate of 0.03, a maximum tree depth of 3, a subsample ratio of 0.8, and a column sampling ratio of 0.8. The squared error regression objective function was applied because the target variable represents continuous counselling search interest values. These parameter settings were selected to balance model complexity and predictive performance while maintaining the ability to capture nonlinear relationships within the time-series data.

Because the dataset consists of chronological time-series observations, conventional random cross-validation was not applied to avoid information leakage. Instead, model validation followed a chronological train-test split, where earlier observations were used for model training and the most recent observations were reserved for testing. Model performance on the testing data was subsequently evaluated using RMSE, MAE, and R² to assess predictive accuracy and generalizability.

Forecast Accuracy Evaluation

The predictive performance of the XGBoost model was evaluated using three statistical indicators: RMSE, MAE, and R^2.

a. Root Mean Squared Error (RMSE). RMSE measures the magnitude of prediction errors by assigning greater weight to larger deviations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

A smaller RMSE indicates that predictions are closer to actual observations.

b. Mean Absolute Error (MAE). MAE measures the average absolute difference between observed and predicted values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\, y_i - \hat{y}_i \,|$$

Lower MAE values indicate better predictive accuracy.

c. Coefficient of Determination (R^2). The explanatory capability of the model was evaluated using:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where $\bar{y}$ represents the mean observed counseling search interest. Higher R^2values indicate that the model explains a larger proportion of variation in counseling search interest.

Model Interpretation

Although machine learning models generally provide high predictive accuracy, interpretation is required to understand the contribution of each predictor. Therefore, this study applied feature importance analysis and SHAP interpretation.

a. XGBoost Feature Importance.

The relative contribution of each feature was measured using the gain-based importance score:

$$Importance_j = \frac{Gain_j}{\sum_{j=1}^{p} Gai\,n_j}$$

where $Gain_j$represents the improvement in prediction accuracy produced by the feature j, $\sum_{j=1}^{p} Gai\,n_j$ represents the total number of features. Higher importance values indicate stronger contributions to predicting counseling search interest.

b. SHAP Interpretation

SHAP analysis was used to explain individual predictions from the XGBoost model.

The SHAP decomposition is defined as:

$$f(x) = \phi_0 + \sum_{i=1}^{p} \phi_i$$

where $f(x)$represents the model prediction, ϕ_0 represents the baseline prediction, and ϕ_i represents the contribution of the feature i. Positive SHAP values indicate that the feature increases predicted counseling search interest, whereas negative SHAP values indicate a decreasing contribution. Through SHAP analysis, the study identifies not only which variables are important but also how each mental health-related search trend influences counseling search behavior.

Results and Discussion

1. Temporal Trends of Mental Health-Related Google Search Interest

The descriptive analysis examined the characteristics and temporal evolution of Google search interest related to mental health from January 2011 to June 2026. The dataset consisted of 186 monthly observations covering six mental health-related search indicators: counseling, depression, anxiety, stress, suicide, and mental health. All variables contained complete observations without missing values. The descriptive statistics of each variable are presented on Table 1 below.

Table 1. Descriptive Statistics of Mental Health-Related Google Search Interest

Variable	Mean	Standard Deviation	Minimum	Median	Maximum
Counselling	30.66	22.14	0	24	100
Depression	8.3	4.85	3	7	55
Anxiety	22.84	18.7	4	17	100
Stress	18.17	3.42	11	18	38
Suicide	6.49	8	3	5	100
Mental Health	11.73	14.83	0	5.5	100

Table 1 shows considerable variation among mental health-related search indicators. Counseling search interest showed the greatest variability among the measured indicators, with a standard deviation of 22.14 and a maximum of 100. Anxiety and mental health searches also showed large fluctuations, indicating periods of increased public attention toward psychological concerns. Stress-related searches presented a more stable pattern with a lower standard deviation.

The descriptive statistics also indicate substantial differences in the distribution of search interest across mental health topics. Counseling, anxiety, suicide, and mental health exhibited maximum values of 100, suggesting occasional periods of exceptionally close public attention, whereas stress-related searches remained relatively stable throughout the observation period. The noticeable disparities between the median and maximum values for several variables further indicate irregular spikes in search activity rather than consistently elevated interest. Such characteristics suggest that Google Trends data exhibit nonlinear fluctuations and varying temporal dynamics across mental health topics, supporting the use of flexible machine learning approaches that can capture complex relationships and sudden changes in search behavior. The monthly trends of mental health-related search interest are illustrated in Figure 1.

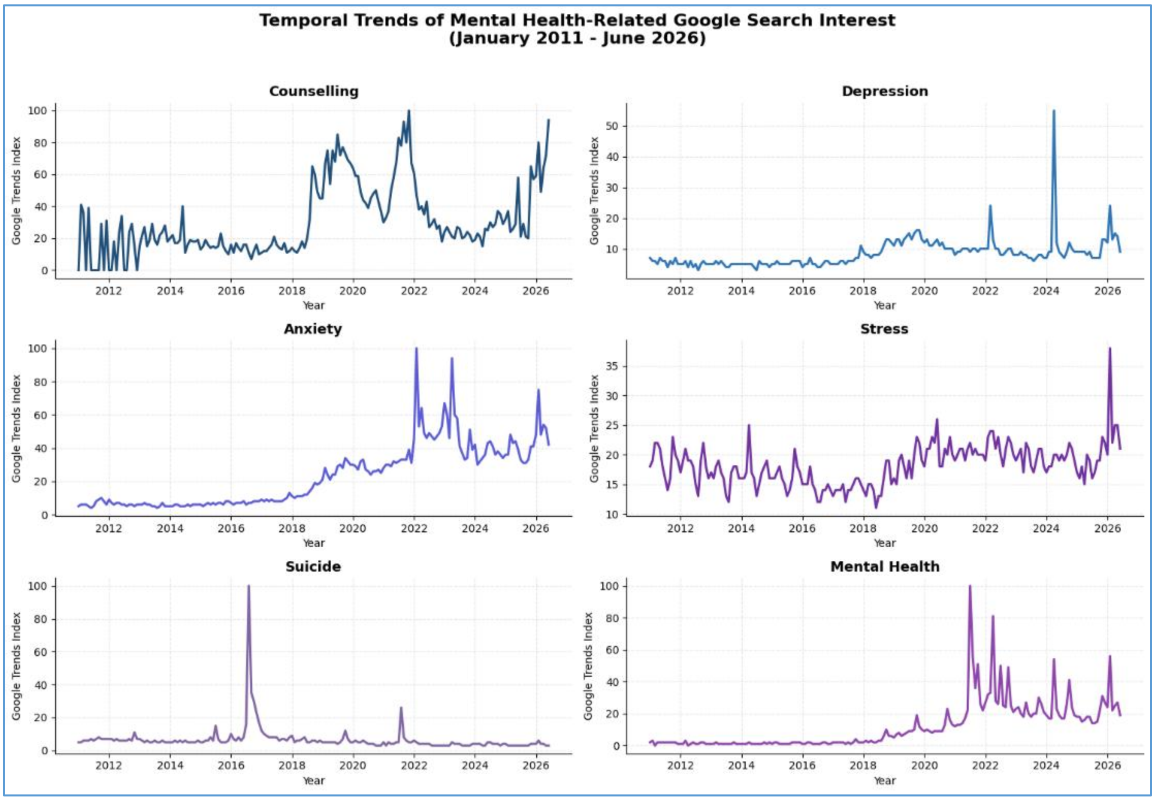

Figure 1. Temporal Trends of Mental Health-Related Google Search Interest

Figure 1 presents the temporal pattern of counseling-related search interest throughout the observation period. During 2011–2017, counseling searches remained relatively low with several short-term fluctuations, indicating a relatively stable but limited level of public engagement with counseling-related information. An increasing trend became apparent after 2018, followed by a substantial rise during 2019–2021. Several peaks were observed during this period, with search interest reaching the maximum Google Trends index value. The increase suggests a growing public interest in mental health support and counseling information through online platforms.

After the peak period, counseling-related searches remained at a higher level compared with the early years of observation. Although fluctuations occurred between 2022 and 2026, the overall pattern indicates sustained public engagement with counseling-related information. The long-term increase highlights the potential role of digital search behavior as an indicator of changes in mental health awareness and information-seeking patterns.

2. XGBoost Model Development and Feature Construction

The forecasting model was developed using an XGBoost regression framework to predict future counseling search interest based on historical search behavior and related mental health indicators. The model incorporated temporal dependencies via lag variables and contextual psychological indicators. The feature construction process involved generating lagged counseling variables representing previous search behavior. Five lag structures were initially explored, including lag-1, lag-2, lag-3, lag-6, and lag-12 months. These variables were designed to capture short-term persistence and longer-term seasonal memory within counseling search patterns. The final input variables used in the XGBoost model are summarized in Table 2.

Table 2. Input features used in the XGBoost forecasting model

Feature Category	Variables
Historical counseling pattern	Counselling lag-1, lag-2, lag-3, lag-12
Psychological indicators	Depression, Anxiety, Stress, Suicide, Mental Health
Target variable	Counseling search interest

The dataset was arranged chronologically to preserve the data's temporal structure. The first 80% of observations were assigned as training data, while the remaining 20% were used as testing data. This approach avoided random sampling procedures that could introduce future information into the training process. The XGBoost model was configured with 500 boosting iterations, a learning rate of 0.03, a maximum tree depth of 3, and a subsampling rate of 0.8. These settings were selected to balance predictive performance and model complexity while reducing the possibility of overfitting.

After completing the iteration, the contribution of each predictor variable was evaluated using the feature importance scores generated by the trained XGBoost model. The ranking of predictor importance is presented in Figure 2.

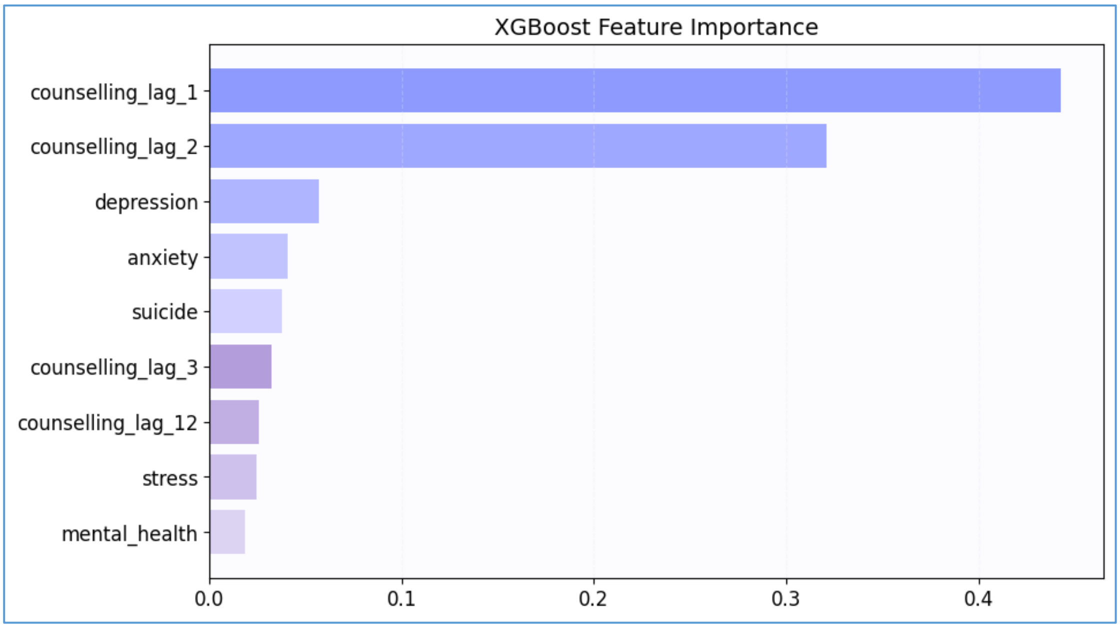

Figure 2. XGBoost feature importance for predicting counseling search interest

Figure 2 demonstrates that historical counseling patterns were the dominant predictors of future interest in counseling searches. Counseling lag-1 contributed the largest share of model importance, followed by counseling lag-2, indicating that recent search behavior provided substantial information for predicting subsequent counseling-related searches. This finding suggests the presence of temporal dependency in public mental health information-seeking behavior, where previous levels of attention toward counseling topics influence future search intensity. Similar patterns have been observed in digital mental health surveillance studies, where historical search trends and prior behavioral signals have improved the prediction of mental health-related outcomes (Kachlik et al., 2025; Sumner et al., 2023). The predictive value of temporal search patterns also supports previous findings that Google Trends data can capture dynamic changes in population-level mental health concerns through continuous, near-real-time behavioral signals.

The importance of lag variables further indicates that counseling-related searches are not isolated events but represent persistent information-seeking patterns over time. Previous studies utilizing internet search data have demonstrated that mental health-related queries exhibit temporal structures,

Vol. 6, No. 1, 2026, pp. 101-117

including seasonal fluctuations and longitudinal changes, which can be effectively modeled using forecasting approaches (Takane, 2025; Wang et al., 2022). In particular, Wang et al. (2022) showed that depression-related Google Trends patterns exhibited meaningful temporal variations consistent with known epidemiological characteristics, suggesting that search behavior contains predictive information about psychological conditions. Similarly, Takane (2025) demonstrated that integrating multiple search patterns with historical suicide data improved prediction accuracy, highlighting the importance of incorporating previous search information when modeling mental health-related outcomes.

Among psychological indicators, depression and anxiety search trends showed relatively greater contributions compared with stress, suicide, and general mental health indicators. This finding suggests that individuals seeking counseling information may be more closely associated with depressive and anxiety-related concerns, which are among the most common psychological conditions linked with help-seeking behavior. Previous machine learning studies have also identified depression and anxiety-related variables as important predictors of mental health outcomes. Kong et al. (2025) demonstrated that XGBoost combined with SHAP successfully identified influential predictors of mental health severity, while Xia et al. (2025) and Vu et al. (2025) reported that machine learning models, particularly XGBoost, could effectively identify important factors associated with depression risk. Furthermore, studies applying explainable machine learning approaches have consistently shown that psychological distress, anxiety symptoms, and depression-related characteristics contribute substantially to predictive performance in mental health models (Benito et al., 2024; Manyam et al., 2025; Mimikou et al., 2025).

The relatively lower contribution of stress, suicide, and general mental health search indicators does not necessarily indicate limited relevance but may reflect differences in information-seeking pathways. Individuals searching for counseling services may use more specific symptom-related terms, such as depression or anxiety, rather than broader psychological concepts. Previous Google Trends-based studies have shown that specific mental health-related queries can provide stronger predictive signals than general health terms because they more directly represent users' concerns or symptoms (Kachlik et al., 2025; Sumner et al., 2023). Therefore, the higher contribution of depression and anxiety variables in this study is consistent with previous evidence that symptom-specific search patterns provide valuable indicators for monitoring population mental health behavior.

The dominance of lag variables and psychological search indicators also supports the suitability of machine learning approaches for modeling digital mental health data. Unlike traditional statistical models that may assume linear relationships, XGBoost can capture nonlinear interactions between historical patterns and psychological indicators, enabling a more flexible representation of complex behavioral processes (Chen & Guestrin, 2016). The application of SHAP analysis further strengthens interpretation by quantifying how each feature contributes to model predictions, enabling identification of the most influential factors driving counseling search interest (Lundberg & Lee, 2017). Previous research has emphasized that explainability is particularly important in mental health applications because predictive models should provide understandable information regarding factors associated with psychological outcomes (Di Martino & Delmastro, 2023; Joyce et al., 2023; Noor et al., 2025).

To further investigate the relationship between predictor variables and model predictions, SHAP (SHapley Additive exPlanations) analysis was conducted. SHAP analysis provides information on the magnitude and direction of each feature's contribution to the predicted counseling search interest. The SHAP summary plot is presented in Figure 3.

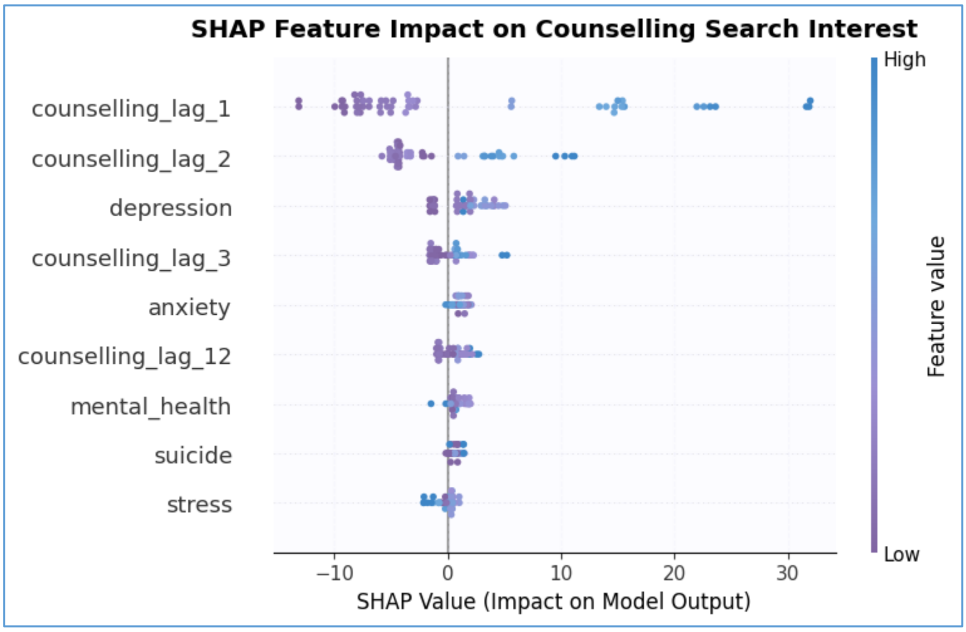

Figure 3. SHAP summary plot of predictor contributions in the XGBoost model

Figure 3 indicates that the counseling lag variables had the strongest influence on the model output, consistent with the feature importance results. Higher recent counseling search values generally contributed positively to predicted counseling interest, indicating that search behavior persists over time. Psychological variables, particularly anxiety and depression, also contributed to model predictions. Higher anxiety-related search interest tended to increase the predicted counseling demand, indicating a potential relationship between symptom-related information seeking and the intention to seek counseling resources. The SHAP interpretation confirms that the XGBoost model does not rely solely on historical trends but also captures variations associated with broader mental health-related search behavior.

3. Model Performance Evaluation

After examining the contribution and interpretation of each predictor variable, the predictive capability of the XGBoost model was evaluated on the test set. The evaluation was conducted using four performance metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). These metrics were selected to assess both prediction accuracy and the model's ability to explain variation in counseling search interest. The model performance results are presented in Table 3.

Table 3. Performance evaluation of the XGBoost forecasting model

Evaluation Metric	RMSE	MAE	MAPE (%)	R^2
Value	11.88	7.58	18.78	0.658

Table 3 shows that the XGBoost model achieved an R^2 of 0.658, explaining approximately 65.8% of the variation in counseling search interest. The RMSE of 11.88 indicates that the average prediction deviation remained within an acceptable range, given the variability of Google Trends indices, which are normalized to 0-100. The MAE value of 7.58 indicates that the model predictions differed from the observed counseling search interest by approximately 7.58 points on average. Furthermore, the MAPE of 18.78% indicates that the forecasting error remained below 20%, indicating reasonable predictive performance for monthly search behavior data characterized by sudden fluctuations. The model's predictive capability was further examined by comparing observed and predicted counseling search interest values during the testing period. The comparison is illustrated in Figure 4.

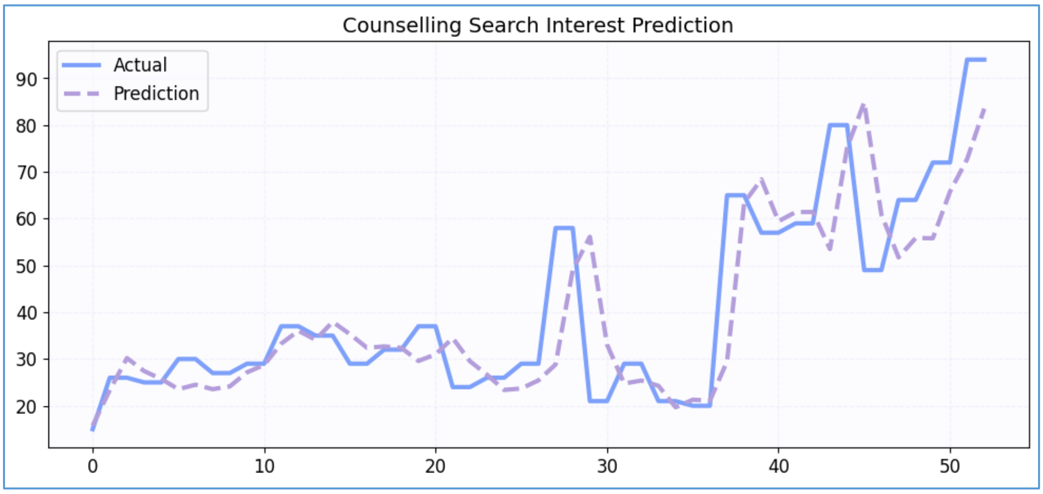

Figure 4. Comparison between observed and XGBoost-predicted counseling search interest during the testing period

Figure 4 shows that the XGBoost model captured the general trend in counseling search interest, particularly during periods of increased search activity. The predicted values followed the observed trend during relatively stable periods, while larger deviations appeared during abrupt changes in search intensity. The largest differences between actual and predicted values occurred during several sudden increases in counseling searches, particularly when the observed index increased sharply within a short period. These deviations are expected because machine learning models based on historical patterns may require additional contextual variables to fully capture unexpected behavioral changes.

Despite these limitations, the model successfully identified the overall direction and magnitude of counseling search dynamics, demonstrating its potential for short-term forecasting of public mental health information-seeking behavior.

4. Future Forecasting of Counseling Search Interest

Following the evaluation of model performance, the trained XGBoost model was applied to generate short-term forecasts of counseling search interest. The forecasting period covered three months after the final observation, corresponding to July 2026 to December 2026. The forecast values generated by the model are summarized in Table 4.

Table 4. Forecasted Mental Health-Related Google Search Interest with 95% Confidence Interval

Time	Forecast	Lower 95%	Upper 95%
Jul-26	69.0498	65.8077	74.0837
Aug-26	72.6235	67.3943	78.6458
Sep-26	59.0887	52.5880	66.2759
Oct-26	56.9902	49.4004	65.3031
Nov-26	48.5371	39.8053	57.5957
Dec-26	43.8740	34.4810	53.8115

Table 4 presents the six-month-ahead forecast of counseling-related Google search interest using the XGBoost model, with a 95% prediction interval. The forecasting results indicate that counseling search interest is projected to rise from 69.0498 in July 2026 to 72.6235 in August 2026, reflecting a short-term increase in public attention to counseling-related information. After reaching the highest predicted value in August 2026, the forecast shows a declining pattern, with search interest decreasing to 59.0887 in September 2026. The declining trend continues throughout the remainder of the forecast period, with predicted values of 56.9902 in October 2026, 48.5371 in November 2026, and 43.8740 in December 2026.

This indicates that the model anticipates a gradual reduction in counseling-related search interest following the initial increase at the start of the forecast horizon.

The 95% prediction interval reflects the uncertainty associated with future forecasts. The interval ranges from 65.8077–74.0837 in July 2026 and expands to 34.4810–53.8115 in December 2026, indicating increasing uncertainty as the forecast horizon extends. The widening prediction interval is consistent with multi-step time-series forecasting, where uncertainty accumulates as predictions are projected further into the future.

5. Implications of Machine Learning-Based Mental Health Trend Prediction

The findings demonstrate that Google Trends data combined with machine learning approaches can provide valuable insights into population-level mental health information-seeking behavior. The upward trend in counseling-related searches suggests that online search activity may reflect shifts in public interest, awareness, and demand for psychological support services. This interpretation is consistent with the concept of psychiatric and mental health infodemiology, in which digital search behavior is considered a source of population-level information on mental health concerns and service-seeking behavior. Previous studies have also demonstrated that Google Trends data can capture meaningful variations in mental health-related interests, including depression-related search patterns that correspond with known temporal and geographic characteristics of psychological conditions (Wang et al., 2022).

The strong contribution of historical counseling search variables indicates that counseling-related information-seeking behavior exhibits temporal persistence. The dominance of counseling lag-1 and lag-2 variables suggests that recent search intensity provides important information for predicting subsequent counseling interest, indicating that online mental health behavior is not randomly distributed but follows sequential patterns over time. Similar findings have been reported in previous studies using digital search data, in which historical and selected search patterns contributed substantially to improved prediction accuracy for mental health-related outcomes. For example, Sumner et al. (2023) demonstrated that online symptom search data could accurately estimate monthly suicide-related trends, highlighting the predictive value of temporally structured digital signals. Furthermore, Takane (2025) showed that integrating multiple Google Trends search patterns with appropriate variable selection methods improved suicide rate prediction, although the effectiveness depended on contextual factors and population characteristics.

The contribution of anxiety and depression-related searches further highlights the relationship between specific psychological symptoms and counseling interest. The higher importance of these predictors suggests that individuals seeking counseling information may simultaneously express concerns related to depressive and anxiety symptoms through online searches. This finding is supported by previous machine learning studies showing that mental health-related indicators, particularly depression and anxiety features, contribute significantly to predictive models of psychological outcomes. Machine learning approaches incorporating Google Trends data have demonstrated that anxiety-related and psychiatric search terms can serve as relevant predictors for suicide attempt trends (Kachlik et al., 2025). Similarly, XGBoost-based mental health prediction models have identified depression- and anxiety-related factors as important contributors when combined with interpretable machine learning techniques (Kong et al., 2025; Vu et al., 2025; Xia et al., 2025).

The ability of XGBoost to identify the relative importance of counseling history, depression, and anxiety indicators supports the suitability of nonlinear machine learning approaches for modeling complex mental health search behavior. Unlike traditional linear models, XGBoost can capture interactions among multiple predictors and identify hidden nonlinear relationships within high-dimensional behavioral data(Chen & Guestrin, 2016). Previous research has demonstrated the effectiveness of XGBoost combined with SHAP interpretation for understanding influential predictors in mental health applications, including depression prediction, anxiety assessment, and suicidal ideation modeling (Benito et al., 2024; Manyam et al., 2025; Mimikou et al., 2025). The application of SHAP in the current study provides additional interpretability by explaining how each search indicator contributes to counseling interest, which is particularly important when predictive models are applied to sensitive domains such as mental health (Di Martino & Delmastro, 2023; Joyce et al., 2023).

From a practical perspective, the forecasting framework provides an early-warning approach for identifying potential increases in public demand for psychological information and counseling services. By using historical search patterns and related psychological indicators, health organizations, policymakers, and service providers can gain additional information to inform resource planning, digital mental health campaigns, and timely intervention strategies. However, Google Trends represents relative search intensity rather than the actual prevalence of mental health conditions. Therefore, increases in search activity should be interpreted as changes in information-seeking behavior rather than direct measurements of psychological disorders (Alibudbud, 2023). Additionally, the current model does not incorporate socioeconomic factors, healthcare accessibility indicators, or major social events that may influence mental health-related searches. Future studies may improve predictive performance by integrating external variables, developing hybrid statistical-machine learning models, or applying advanced deep learning architectures for digital mental health forecasting.

Conclusion

This study developed an XGBoost-based forecasting model to predict counseling-related Google Search interest using monthly mental health-related search indicators from January 2011 to June 2026. The temporal analysis identified an upward trend in counseling searches, particularly after 2018, indicating growing public interest in psychological support and mental health information. The findings demonstrate that Google Trends data can provide useful information for examining changes in population-level mental health-related search behavior over time.

The XGBoost model successfully captured the temporal dynamics of counseling search interest by incorporating historical search patterns and related mental health indicators. Feature importance and SHAP analysis showed that previous counseling search behavior was the main factor influencing model predictions, while anxiety and depression-related searches also contributed to the forecasting process. The model demonstrated satisfactory predictive performance and captured nonlinear patterns in mental health-related search behavior.

The results highlight the potential of machine learning-based forecasting to monitor changes in public demand for mental health information and support services. Although Google Trends reflects online search behavior rather than direct measures of mental health conditions, the proposed framework may support early identification of shifts in public attention to psychological issues. Future research may improve the model by incorporating additional socioeconomic, healthcare, and contextual factors to provide a more comprehensive understanding of mental health-related trends.

Acknowledgment

The authors would like to express their sincere appreciation to all individuals and institutions that contributed to the completion of this study. The authors acknowledge the availability of publicly accessible Google Trends data and other supporting information sources that enabled the analysis of mental health-related search behavior and temporal patterns. Special gratitude is extended to colleagues and reviewers who provided constructive feedback, methodological suggestions, and valuable insights throughout the development of this research. Appreciation is also given to individuals who contributed to improving the manuscript through language refinement and critical evaluation. All acknowledged individuals and parties have granted their permission to be mentioned. The authors are grateful for the support, collaboration, and professional contributions that enhanced the quality and clarity of this study.

References

Alibudbud, R. (2023). Google Trends for health research: Its advantages, application, methodological considerations, and limitations in psychiatric and mental health infodemiology. *Frontiers in Big Data*, *6*, 1132764. https://doi.org/10.3389/fdata.2023.1132764

Arif, E., Devianto, D., Yollanda, M., & Afrimayani, A. (2022). Analysis of Precious Metal Price Movements Using Long Memory Model and Fuzzy Time Series Markov Chain. *International Journal of Energy Economics and Policy*, *12*(6), 202–214. https://doi.org/10.32479/ijeep.13531

Arif, E., Herlinawati, E., Devianto, D., Yollanda, M., & Permana, D. (2024). Hybridization of long short-term memory neural network in fractional time series modeling of inflation. *Frontiers in Big Data*, *6*, 1282541. https://doi.org/10.3389/fdata.2023.1282541

Benito, G. V., Goldberg, X., Brachowicz, N., de Cid, R., & Petrone, P. M. (2024). Machine learning for anxiety and depression profiling and risk assessment in the aftermath of an emergency. *Artificial Intelligence in Medicine*. https://doi.org/10.1016/j.artmed.2024.102991

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785

Devianto, D., Permana, D., Arif, E., Afrimayani, A., Yanuar, F., Maiyastri, M., & Yollanda, M. (2024). An innovative model for capturing seasonal patterns of train passenger movement using exogenous variables and fuzzy time series hybridization. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1), 100232. https://doi.org/10.1016/j.joitmc.2024.100232

Devianto, D., Ramadani, K., Maiyastri, Asdi, Y., & Yollanda, M. (2022). The hybrid model of autoregressive integrated moving average and fuzzy time series Markov chain on long-memory data. *Frontiers in Applied Mathematics and Statistics*, *8*, 1045241. https://doi.org/10.3389/fams.2022.1045241

Devianto, D., Wahyuni, E., Maiyastri, M., & Yollanda, M. (2024). The seasonal model of chili price movement with the effect of long memory and exogenous variables for improving time series model accuracy. *Frontiers in Applied Mathematics and Statistics*, *10*, 1408381. https://doi.org/10.3389/fams.2024.1408381

Devianto, D., Yollanda, M., Maiyastri, M., & Yanuar, F. (2023). The soft computing FFNN method for adjusting heteroscedasticity on the time series model of currency exchange rate. *Frontiers in Applied Mathematics and Statistics*, *9*, 1045218. https://doi.org/10.3389/fams.2023.1045218

Di Martino, F., & Delmastro, F. (2023). Explainable AI for clinical and remote health applications: A survey on tabular and time series data. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-022-10304-3

Joyce, D. W., Kormilitzin, A., Smith, K. A., & Cipriani, A. S. (2023). Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *Npj Digital Medicine*. https://doi.org/10.1038/s41746-023-00751-9

Kachlik, Z., Walaszek, M., Nazar, W., Sokołowska, M., Karbiak, A., Pilarska, E., & Cubała, W. J. (2025). Predicting Suicide Attempt Trends in Youth: A Machine Learning Analysis Using Google Trends and Historical Data. *Journal of Clinical Medicine*, *14*(18), 6373. https://doi.org/10.3390/jcm14186373

Kong, W., Pei, Z., Guo, Z., Xu, R., & Zhao, J. (2025). Relationship matters: Using machine learning methods to predict the mental health severity of Chinese college freshmen during the pandemic period. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2024.09.168

Lundberg, S. M., & Lee, S. (2017, January 1). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*.

Manyam, R., Yun, J., Lou, P., Scott, L. N., & Krafty, R. T. (2025). Predicting Risk Factors for Suicidal Ideation Using Interpretable Artificial Intelligence Methods. *Proceedings - 2025 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2025*. https://doi.org/10.1109/BIBM66473.2025.11357209

Mimikou, C., Kokkotis, C., Tsiptsios, D. I., Anagnostopoulos, K. P., & Trıotapsianis, G. A. (2025). Explainable Machine Learning in the Prediction of Depression. *Diagnostics*. https://doi.org/10.3390/diagnostics15111412

Niu, B., Wan, M., & Zhou, Y. (2025). Development of an explainable machine learning model for predicting depression in adolescent girls with non-suicidal self-injury: A cross-sectional multicenter study. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2025.03.080

Noor, A. A., Manzoor, A., Mazhar Qureshi, M. D., Qureshi, M. A., & Rashwan, W. (2025). Unveiling Explainable AI in Healthcare: Current Trends, Challenges, and Future Directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. https://doi.org/10.1002/widm.70018

Putra, A. H., & Afiati, E. (2026). Beyond the Australian Context: Reconsidering the Relationship Between Climate Exposure and Psychosocial Stressors in Youth Mental Health. *Early Intervention in Psychiatry*, *20*(7), e70196. https://doi.org/10.1111/eip.70196

Putra, A. H., Ardi, Z., & Afiati, E. (2026). Perceived organisational support is a condition, not a cure: The case for acceptance and commitment therapy in child protection worker mental health frameworks. *European Journal of Trauma & Dissociation*, *10*(3), 100714. https://doi.org/10.1016/j.ejtd.2026.100714

Putra, A. H., Ardi, Z., Ifdil, Afiati, E., & Herawati, A. A. (2026). Control-group deterioration and inflated effect sizes in LMIC school-based well-being trials: A methodological warning for adolescent mental health prevention research. *Mental Health & Prevention*, *43*, 200534. https://doi.org/10.1016/j.mhp.2026.200534

Sumner, S. A., Alic, A., Law, R. K., Idaikkadar, N., & Patel, N. (2023). Estimating national and state-level suicide deaths using a novel online symptom search data source. *Journal of Affective Disorders*, *342*, 63–68. https://doi.org/10.1016/j.jad.2023.08.141

Takane, S. (2025). Can the integrated and selected pattern approach effectively predict suicide rates? A study using internet search queries. *Journal of Affective Disorders Reports*, *21*, 100934. https://doi.org/10.1016/j.jadr.2025.100934

Vu, T., Dawadi, R., Yamamoto, M., Tran Ngoc Hoang, P., & Araki, M. (2025). Prediction of depressive disorder using machine learning approaches: Findings from the NHANES. *BMC Medical Informatics and Decision Making*. https://doi.org/10.1186/s12911-025-02903-1

Wang, A., McCarron, R., Azzam, D., Stehli, A., Xiong, G., & DeMartini, J. (2022). Utilizing Big Data From Google Trends to Map Population Depression in the United States: Exploratory Infodemiology Study. *JMIR Mental Health*, *9*(3), e35253. https://doi.org/10.2196/35253

Xia, F., Ren, J., Liu, L., Cui, Y., & He, Y. (2025). A machine learning-based depression risk prediction model for healthy middle-aged and older adult people based on data from the China health and aging tracking study. *Frontiers in Public Health*. https://doi.org/10.3389/fpubh.2025.1515094

Yollanda, M., & Devianto, D. (2020). Hybrid Model of Seasonal ARIMA-ANN to Forecast Tourist Arrivals through Minangkabau International Airport. *Proceedings of the Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia*. Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia. https://doi.org/10.4108/eai.2-8-2019.2290473

Yollanda, M., Devianto, D., & Yozza, H. (2018). Nonlinear Modeling of IHSG with Artificial Intelligence. *2018 International Conference on Applied Information Technology and Innovation (ICAITI)*, 85–90. https://doi.org/10.1109/ICAITI.2018.8686702

Yollanda, M., Harjupa, W., Devianto, D., Andarini, D. F., Nauval, F., Saufina, E., Purwaningsih, A., Harjupa, W., Trismidianto, Harjana, T., Risyanto, Rahmatia, F., Pratama, R., & Satiadi, D. (2023).

Prediction of CENS, MJO, and Extreme Rainfall Events in Indonesia Using the VECM Model. In A. Basit, E. Yulihastin, S. Y. Cahyarini, H. Santoso, W. S. Pranowo, L. Slamet S., & H. A. Belgaman (Eds.), *Proceedings of the International Conference on Radioscience, Equatorial Atmospheric Science and Environment and Humanosphere Science* (Vol. 290, pp. 367–383). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-9768-6_35