International Journal of Scientific Research in Computer Science and Engineering
Vol.**13**, Issue.**2**, pp.**21-29**, April **2025**
E-ISSN: 2320-7639
Available online at: www.isroset.org

Research Article

Comparison of NAIVE, SARIMA, and SARIMAX Models in Short-Term and Long-Term Forecasting of Google Search Trends

Regita Putri Permata[1*], Ananda Taqhsya Dwiyana[2]

[1,2]Data Science, School of Informatic, Telkom University, Surabaya Campus, Surabaya, 60231, Indonesia

**Corresponding Author:*

Received: 13/Mar/2025; **Accepted**: 07/Apr/2025; **Published**: 30/Apr/2025. | **DOI**: https://doi.org/10.26438/ijsrcse.v13i2.665

Abstract— This study examines forecasting methods—Naïve, SARIMA, and SARIMAX—to enhance the accuracy of Google search trend predictions in the fitness industry, specifically for the keyword "Gym." The Naïve method serves as a baseline comparison, SARIMA incorporates seasonal, autoregressive, and moving average components for improved trend detection, while SARIMAX extends SARIMA by integrating exogenous variables. Historical Google search data from 2005 to 2015 is used for model evaluation, with performance assessed using MAE, RMSE, and MAPE metrics. The findings indicate that SARIMAX, which accounts for external influences, provides the highest accuracy. Additionally, short-term models exhibit greater responsiveness to seasonal variations. This study offers insights into the strengths and limitations of each method, assisting practitioners in selecting the most suitable approach for improving search trend forecasting and supporting data-driven decision-making in the fitness industry.

Keywords— Google Search Trends, Forecasting Methods, Naïve Method, SARIMA, SARIMAX, Time Series Analysis.

1. Introduction

Nowadays, a lot of individuals rely on the internet as their main information source. Google, one of the top search engines, gathers and keeps track of user-inputted terms. Online trends, inclinations, and user preferences are reflected in this search data. As a result, search data analysis is essential for many areas, including content strategy creation, business planning, and digital marketing [1].

Forecasting shifts in Google keyword demand has a big influence on company choices, particularly in sectors where market trends are crucial. The fitness and gym sector is one such area that is impacted by search patterns. Seasonal patterns and other outside variables can cause fluctuations in the demand for goods and services in this sector [2]. Therefore, in order to assist stakeholders in creating more successful business plans, an accurate forecasting tool is required to predict future trends in Google search [3].

Several previous studies have explored Google search predictions in various contexts. For example, a study on Google search prediction for hijab brands in Indonesia used a combination of classical methods and machine learning, showing that the combined approach had lower errors than single methods [4]. Additionally, Google Trends data has been used to forecast the number of dengue fever cases using the ARIMAX method, with results indicating improved prediction accuracy and minimized forecasting errors [5]. Another study utilized Google Trends to analyze inflation in Indonesia, with ARIMAX being identified as the best model for this context [6].

Many forecasting techniques have been used in a variety of industries, but there are still a number of unanswered research questions. The majority of current research concentrates on a single forecasting technique or a particular hybrid approach without methodically contrasting straightforward techniques like the Naive Method with more intricate models like SARIMA and SARIMAX. It is challenging to identify the best strategy for various forecasting scenarios due to the paucity of comparative studies that evaluate the advantages and disadvantages of each method [7].

Additionally, there is currently a dearth of research that uses Google Trends data to forecast trends in the fitness sector. There is an urgent need for specialist research on search trend forecasts in this field because of the industry's quick trend fluctuations, which are driven by things like marketing campaigns, seasonality, and changes in lifestyle [8, 9], [10].

The inclusion of external factors, which can improve prediction accuracy, is another benefit of the SARIMAX model. Studies determining the most pertinent external factors in relation to Google search trends for the fitness sector are still few, nevertheless. Although this component has not been fully examined in the literature to yet, a deeper comprehension of these variables might greatly enhance predicting ability [4, 11].

In the study conducted by Nisha Thakur and Sanjeev Karmakar were approached using Long Short-Term Memory (LSTM) models. The model's performance was assessed using evaluation metrics such as Mean Absolute Deviation (MAD), Mean Square Error (MSE), Root Mean Square Error (RMSE), Cosine Similarity (CS), and Correlation Coefficient (r), across varying learning rates and epoch configurations. While the study demonstrated the superior performance of LSTM compared to traditional ANN models, it also highlighted a broader challenge in the field: the lack of a consistent evaluation framework. Many studies, including this one, adopt different error metrics—such as MAE, MAPE, or MAD—making it difficult to conduct objective and standardized comparisons across forecasting models in real-world applications [12]. .

This research compares three forecasting methods—the Naive Method, SARIMA, and SARIMAX—for predicting Google search patterns in the fitness and club business. The comparison is made to determine the accuracy of each approach in anticipating Google search trends, as well as to assess the benefits and drawbacks of basic methods (Naive) vs more complicated statistical model-based methods (SARIMA and SARIMAX). Furthermore, this study explores if include external variables in SARIMAX enhances accuracy when compared to other approaches [13]. By understanding the effectiveness of each method, this study also aims to contribute to the broader literature on Google search trend forecasting.

2. Related Work

The application of forecasting models in a variety of sectors has been the subject of several research, particularly in relation to the analysis of trends using time-series data. Forecasting methods such as Naive, SARIMA, and SARIMAX have been widely employed to predict financial, healthcare, and marketing trends. However, there is still a dearth of research on Google search trends in particular, especially in the fitness industry. The review is divided into three primary sections: (1) Utilizing Google Trends for Forecasting; (2) Comparative Analysis of Forecasting Techniques; and (3) Applying SARIMA and SARIMAX Models.

2.1 Application of SARIMA and SARIMAX Models

In a research published in 2023, Karim et al. investigated how well the SARIMA and SARIMAX models predicted macroeconomic indices in Australia. The purpose of the study was to ascertain if using data from Google Trends might increase predicting accuracy. According to the results, SARIMAX models—which incorporate external variables—produced forecasts that were more accurate than those of conventional SARIMA models [14]. SARIMA models were used in a different research by Jenny Holm (2021) to predict monthly weather trends. Evaluating SARIMA's ability to capture seasonal fluctuations in precipitation and temperature was the goal. The findings showed that seasonal trends were accurately predicted using SARIMA models, improving prediction accuracy.

2.2 Use of Google Trends in Forecasting

The real-time nature of Google Trends data and its relevance to consumer behavior have led to its increasing use in predicting applications. The use of Google Trends data to predict influenza-like symptoms in South Africa was investigated in a research by Olukanmi et al. (2021) [15]. The purpose of the study was to determine whether adding search data may improve conventional models' capacity for prediction. According to the findings, models that used data from Google Trends performed better than those that only used historical sickness data. Boone et.al (2018) looked into how Google Trends data may be used to enhance sales projections in the business sector. The goal was to ascertain whether search trends might be used as predictive markers of customer buying patterns [16].

This study aims to fill these gaps by evaluating Naive, SARIMA, and SARIMAX methods in predicting Google search trends related to the fitness industry. By incorporating comparative analysis and assessing the impact of external variables, this research contributes to the literature on search trend forecasting and provides insights for business and marketing applications.

3. Methodology

Trend analysis, including a case study centered on the term "Gym." Monthly search volume statistics for the term "Gym" (represented by Y) from January 2005 to December 2015 (represented by X) make up the dataset. Jupyter Notebook is used for data pretreatment and analysis, which makes it easier to create models, evaluate them, and visualize the predicting results. Each model's performance is evaluated using common error metrics like MAE, RMSE, and MAPE in order to determine how well it predicts both short-term and long-term trends. This study also employs Exploratory Data Analysis (EDA) to identify patterns, trends, and potential anomalies within Google Trends time series data. EDA plays a crucial role in ensuring data quality prior to modeling, including handling missing values, outliers, and variable transformations. It supports better prediction accuracy and enables more objective interpretations in time series forecasting [17].

3.1 Seasonal Naïve Method

The seasonal naive method is a simple approach to handling seasonal elements in time series data. In this context, "naive" refers to a direct and minimalistic approach that does not consider further patterns or complexities in the data, following the model below [4]:

$$\hat{Y}t+1=Y(t+1)-s \quad (1)$$

3.2 SARIMA

The ARIMA method is developed to analyze repetitive or seasonal data patterns that occur at fixed intervals such as quarterly, semi-annually, and annually. In general, the SARIMA model $(p,d,q)(P,D,Q)^S$ is formulated as follows [5][6]:

$$\varphi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Z_t=\theta_q(B)\Theta_Q(B^S)e_t \quad (2)$$

where:

p,d,q	: Orders of non-seasonal AR, differencing, and MA
P,D,Q	: Orders of seasonal AR, differencing, and MA
$(1-B)^d$	: Non-seasonal differencing
$(1-B^S)^D$	: Seasonal differencing
$\varphi_P(B)$	: Non-seasonal autoregressive order p
$\theta_q(B)$	: Non-seasonal moving average order q
Φ_P	: Seasonal autoregressive order P
$\Theta_Q(B^S)$	: Seasonal moving average order Q
B	: Backshift operator

3.3 SARIMAX

The seasonal time series model can be enhanced by incorporating several exogenous variables that are considered to have a significant influence on the data, thereby increasing forecasting accuracy. The SARIMAX model is expressed as follows:

$$\varphi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Z_t=\theta_q(B)\Theta_Q(B^S)e_t+a_1x_{1t}+a_2x_{2t}+\ldots+a_kx_{kt} \quad (3)$$

where x_{kt} represents the exogenous variable k at time t, while the other symbols remain consistent with the SARIMA model.

3.4 Mean Absolute Percentage Error (MAPE)

MAPE is a relative accuracy measure in percentage form that evaluates the forecasting performance. A lower MAPE value indicates a better level of accuracy. The formula for calculating MAPE is as follows:

$$MAPE=\frac{1}{n}\sum_{t=1}^{n}\frac{|z_t-\hat{z}_t|}{z_t}\times 100\% \quad (4)$$

where:

z_t	: Actual value at period t
$\hat{z}_t$	: Predicted value at observation t
n	: Number of observations

3.5 Mean Squared Error (MSE)

MSE measures the average squared difference between predicted and actual values. The formula is:

$$MSE=\frac{1}{n}\sum_{t=1}^{n}\left(z_t-\hat{z}_t\right)^2 \quad (5)$$

3.6 Root Mean Squared Error (RMSE)

RMSE is the square root of MSE. It provides an error measure comparable to the scale of the target variable. This is useful because it presents errors in the same unit as the target variable. The formula is:

$$RMSE=\sqrt{MSE} \quad (6)$$

3.6 Mean Percentage Error (MPE)

MPE measures the average percentage difference between predicted and actual values. The formula is:

$$MPE=\frac{1}{n}\sum_{t=1}^{n}\frac{z_t-\hat{z}_t}{z_t}\times 100\% \quad (7)$$

The forecasting process in this study consists of the following steps:

1. Data Preparation: Historical rainfall data is collected, cleaned, and split into training and testing sets for short-term and long-term forecasting analysis.
2. Exploratory Data Analysis (EDA): EDA is performed to understand trends, seasonality, and patterns in the data through visualizations such as time series plots and seasonal decomposition.
3. Naïve Model Implementation: Seasonal naïve models (short-term and long-term) are applied by repeating past seasonal values as forecasts for future periods.
4. Naïve Model Evaluation: Performance metrics such as MSE, RMSE, and MAPE are used to compare short- and long-term naïve forecasts. The short-term model shows better accuracy.
5. Stationarity Testing: Augmented Dickey-Fuller (ADF) test is conducted to assess stationarity. Differencing is applied to achieve a stationary time series for model fitting.
6. ACF and PACF Analysis: Autocorrelation and partial autocorrelation plots are analyzed to determine the AR (autoregressive), MA (moving average), and seasonal parameters for SARIMA/SARIMAX models.
7. SARIMA Model Construction: Short-term and long-term SARIMA models are developed using the identified parameters, followed by training and forecasting.
8. SARIMAX Model with Exogenous Variables: SARIMAX models are constructed by integrating

external variables. Forecasting is performed for both short- and long-term horizons.

9. Model Performance Evaluation: All models are evaluated using MSE, RMSE, MAE, and MAPE. The short-term SARIMAX model yields the best performance due to its ability to capture recent patterns and external influences.
10. Conclusion.

4. Results and Discussion

Figure 1 illustrates the trend of search interest for the keyword "Gym" on Google over a specific time period. The plot reveals clear seasonal patterns and an overall upward trend, indicating a growing public interest in fitness-related topics. The recurring peaks may correspond to common periods of heightened fitness interest, such as the beginning of the year or pre-summer months. This pattern supports the relevance of using time series forecasting methods to model and predict future interest.

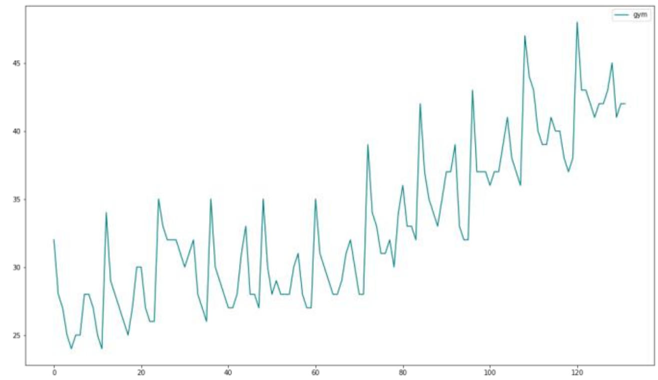

Figure 1. Plot of the word "Gym" on Google

Based on the plot, the blue line called “gym” shows the variation in values that occur in the time range of 25 to 45. This graph presents the search data for the word “gym” on Google between 2005 and 2015. There is significant variation in the data over time, as shown by the fluctuations in the Figure 1.

4.1 Naive Model

4.1.1 Short-Term and Long-Term Naive Model Forecasting

The periodic fluctuations are displayed in the short-term naive seasonal predictions in Figure 2; the comparison between the original data (blue line and dots) and the predictions (orange line and dots). Although there are some similarities, there are significant differences. These patterns help determine upcoming seasonal patterns.

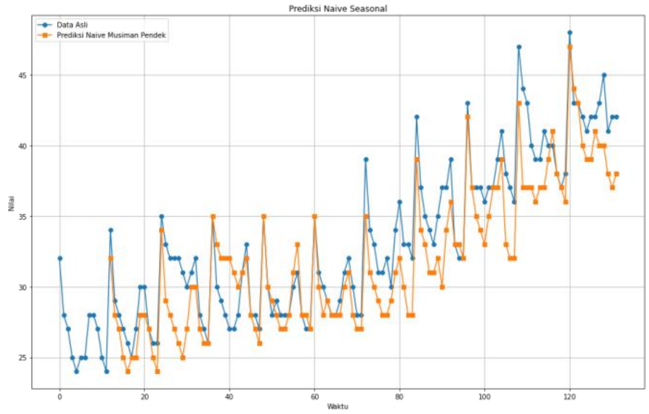

Figure 2. Short-Term Naive Seasonal Prediction

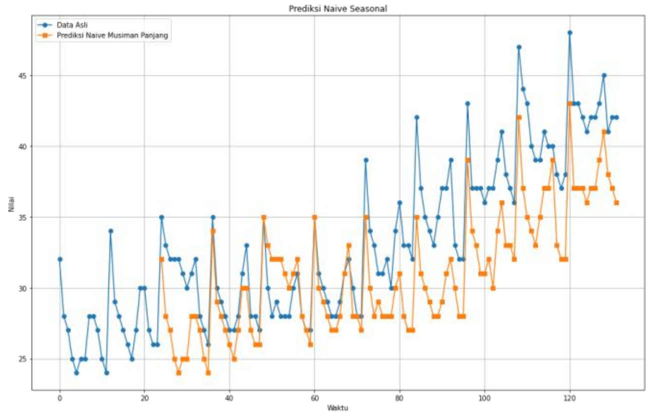

Figure 3. Long-Term Naive Seasonal Prediction

Figure 3 Long-term naive seasonal predictions show regular fluctuation patterns. The original data (blue line and dots) shows significant variation, while the predictions (orange line and dots) tend to be stable and follow seasonal patterns. This model can identify seasonal patterns for long-term predictions.

The plot shows the original data (blue) and the short-term (green) and long-term (orange) naive seasonal predictions. Short-term naive seasonal predictions are smoother and more accurate as they follow the original data patterns more closely.

4.1.2 Evaluation of Naïve Model

Table 1. Naïve Evaluation

	Model	MSE	RMSE	MAPE
0	Naïve Short-Term Seasonal	7.37	2.71	6.17
1	Naïve Short-Term Seasonal	17.66	4.20	10.18

The evaluation results show that the Naive Short-Term Seasonal model is better than the Naive Long-Term Seasonal model, with lower error values on all metrics (MSE, RMSE, and MAPE%). This indicates that the Naive Short-Term Seasonal model has better predictive fit than the Naive Long-Term Seasonal model.

4.2 SARIMA Model

4.2.1 Stationarity

The Augmented Dickey-Fuller (ADF) test statistic value is 0.83, well above the critical value for all levels of significance. The p-value is also very large, at 0.99, far above the 0.05 or 0.01 threshold. This indicates that it fails to reject the null hypothesis that the time series has a unit root, leading to the conclusion that the time series is not stationary.

4.2.2 ACF and PACF Before Differencing

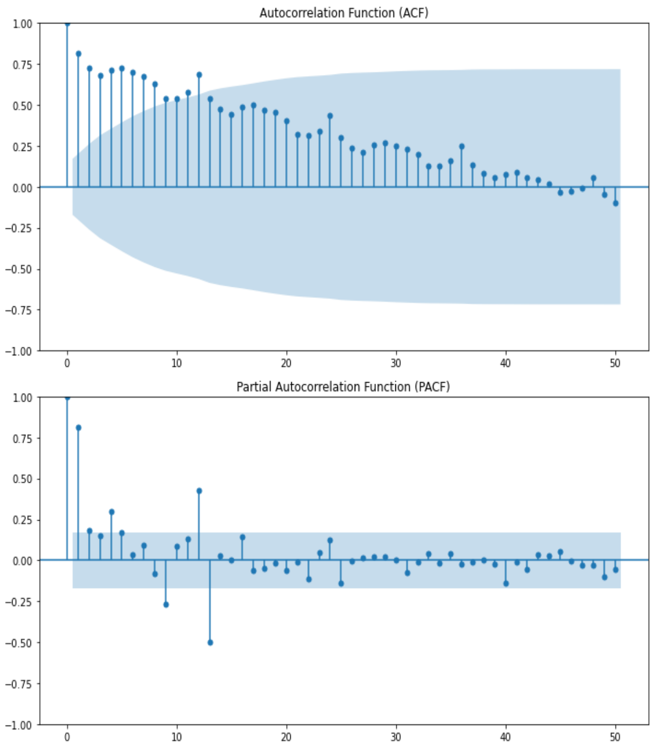

Figure 4. ACF and PACF plot before differencing

While the ACF plot in Figure 4 shows strong autocorrelation at the initial lags that gradually decreases, the PACF plot shows a significant spike at lag 1, which fits the AR(1) model. As a result, the ARIMA(1,0,0) model is appropriate. According to the first differencing results, an ADF value of -3.51 and a p-value of 0.01 indicates stationarity. According to the 12th differencing results, an ADF value of -5.29 and a p-value of 5.77×10^{-6} also indicates stationarity. Therefore, 11 lags will be used with 118 research observations.

4.2.3 ACF and PACF After Differencing

After applying differencing to achieve stationarity, further analysis was conducted using the following ACF and PACF plots shown in Figure 5.

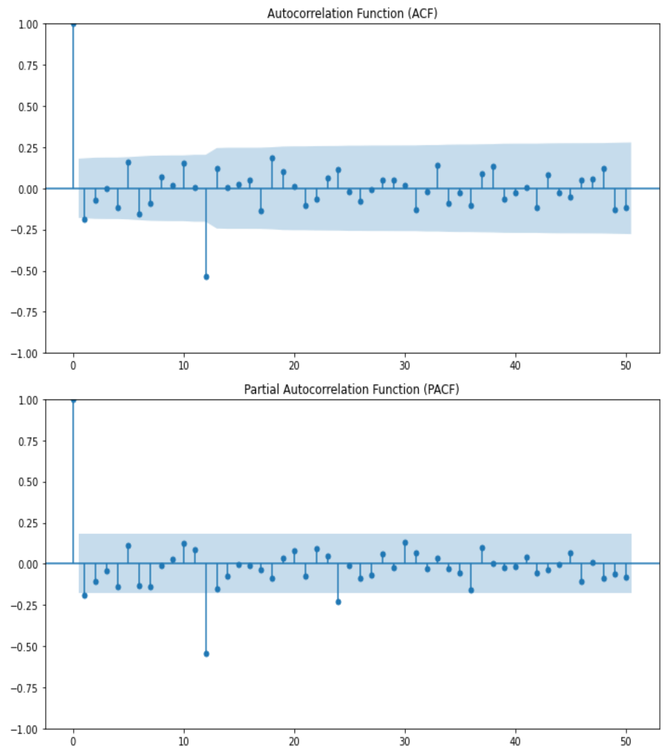

Figure 5. ACF and PACF plot after differencing

Positive autocorrelation is seen at the initial lags in the PACF is shown in Figure 5, while ACF and PACF show a sharp decline at the initial lags, indicating that the data may have a trend or be non-stationary. $P = 2$ (PACF), $D = 2$ (differencing), and $q = 2$ (ACF) determine the seasonal order.

5.2.2 Long-Term SARIMA Forecast

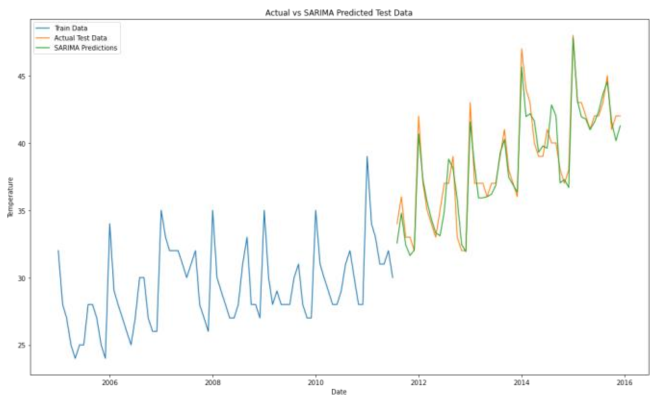

Figure 6. Comparison graph between actual test data and long-term SARIMA model prediction results

The graph in Figure 6 shows the comparison between the actual test data and the SARIMA model predictions from 2005 to 2016, with the training data used as a reference. However, there are significant differences between the green line (test data) and the orange line (predictions). Significant prediction errors are shown at several points, indicating that the SARIMA model does not capture all patterns or changes in the data.

5.2.3 Short-Term SARIMA Forecast

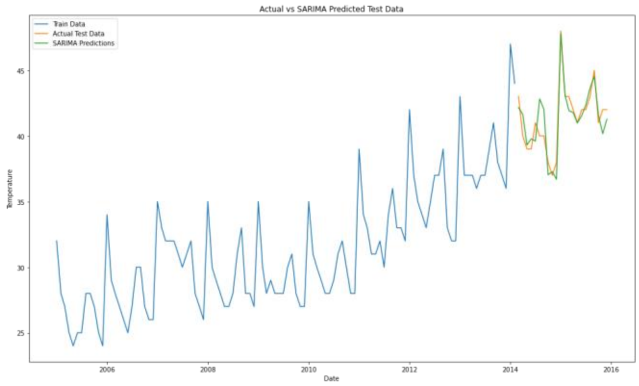

Figure 7. Comparison graph between actual test data and short-term SARIMA model prediction results

According to the actual test data versus SARIMA predictions plot, it appears that the training data was used to create the SARIMA prediction model. This plot shows the comparison between actual test data and predictions made by the SARIMA model from 2012 to 2016. We can see from this plot that, despite some deviations, the SARIMA predictions are quite accurate in following the actual temperature fluctuation patterns.

5.2.4 Evaluation of Short-Term and Long-Term SARIMA Forecast

Table 2. SARIMA Model Evaluation

	Model	MSE	RMSE	MAE	MAPE
0	SARIMA Short-Term	1.23	1.11	0.86	2.12
1	SARIMA Long-Term	1.31	1.14	0.90	2.34

By comparing the MSE, RMSE, MAE, and MAPE values, it is found that the short-term SARIMA method has lower values for all model evaluations. Therefore, the short-term SARIMA method is more accurate in predicting the data. This indicates that the short-term method is more responsive to seasonal changes as it focuses on the last values during the seasonal period. Conversely, the long-term SARIMA method is smoother as it ignores changes and assumes future values are averages of the values over the seasonal period.

4.3 SARIMAX Model

4.3.1 Long-Term SARIMAX Forecast

For long-term analysis, the data was split into training and testing sets with two proportions. 60% for training and 40% for testing, and 85% for training and 15% for testing. Then, predictions were made using OLS. ADF test results showed a p-value of 0.03, which is lower than the 0.05 significance level, indicating that the data is stationary. The ADF test statistic of -3.02, which is lower than the critical values at the 1%, 5%, and 10% significance levels, supports the conclusion that the time series is stationary.

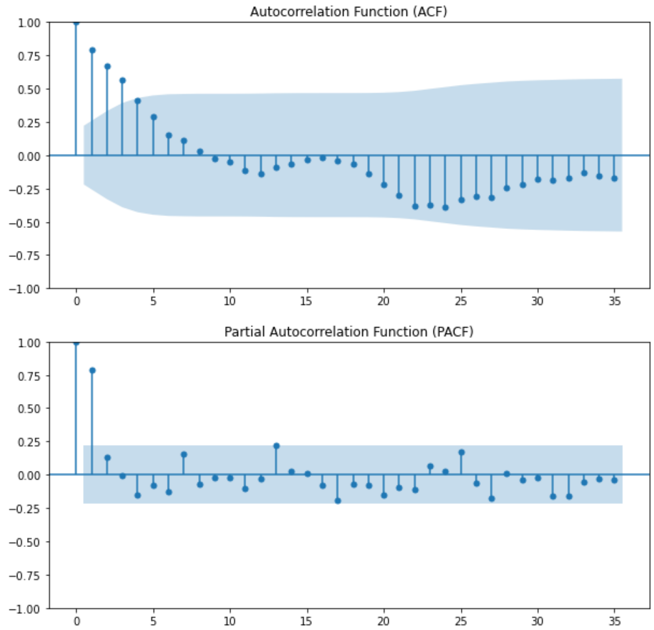

Figure 8. ACF and PACF plot before differencing

In time series models, both plots help determine the order of the MA (Moving Average) and AR (Autoregressive) components in the time series model. The ACF plot shows significant correlation at the initial lags, indicating a seasonal or trend component. The PACF plot shows correlation after removing the influence of intervening values, indicating a potential AR(1) process.

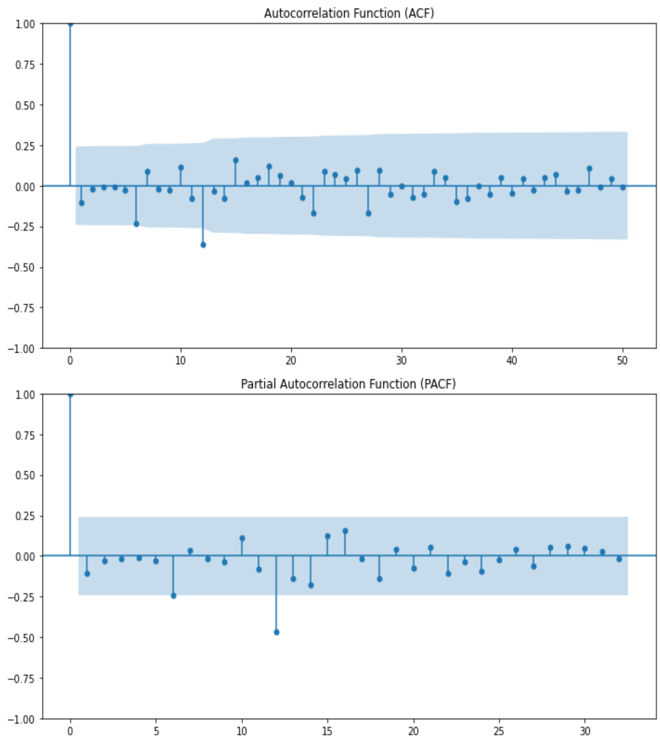

Figure 9. ACF and PACF plot after differencing

The ACF plot in Figure 9 shows strong positive correlation at lag 0 that then decreases, with insignificant correlations at other lags, indicating that differencing has removed repeating patterns. The PACF plot also shows an increase at lag 0 but insignificant decreases at higher lags, indicating that correlations can be explained by nearby observations.

In Figure 10 presents the long-term forecast generated using the SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) model. The prediction is based on historical data combined with relevant external variables, aiming to capture long-term trends, seasonal patterns, and the influence of exogenous factors. The model is expected to provide more accurate insights into future developments by incorporating these components

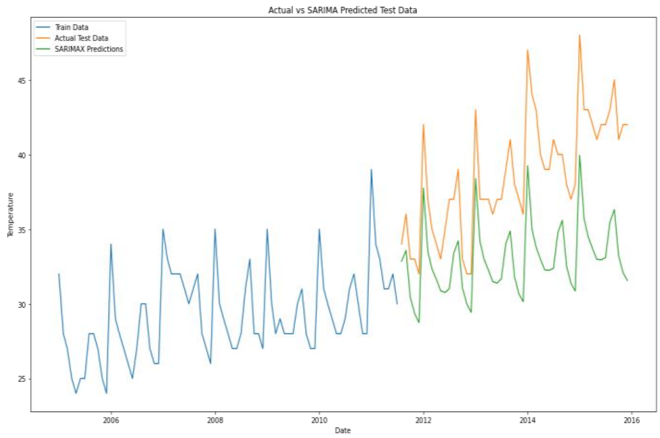

Figure 10. Comparison graph between actual test data and long-term SARIMAX model prediction results

Figure 10. Comparison graph between actual test data and long-term SARIMAX model prediction results This figure illustrates a comparison between the actual test data and the forecasted values produced by the long-term SARIMAX model. The graph helps evaluate the model's performance in capturing underlying patterns and trends over an extended prediction horizon, highlighting the alignment or deviation between observed and predicted value

4.3.2 Short-Term SARIMAX Forecast

For short-term analysis, the data was split into training and testing sets with two proportions. 60% for training and 40% for testing, and 85% for training and 15% for testing. Predictions were then made using OLS. The ADF test results showed sufficient evidence to reject the hypothesis of non-stationarity, with a p-value of 0.2014, which is greater than 0.05. The ADF test statistic of -2.2135 is also greater than the critical values at the 1%, 5%, and 10% significance levels. This indicates that the data is not stationary. Therefore, the data needs further attention and may need to be transformed to become stationary before being used for modeling or further analysis.

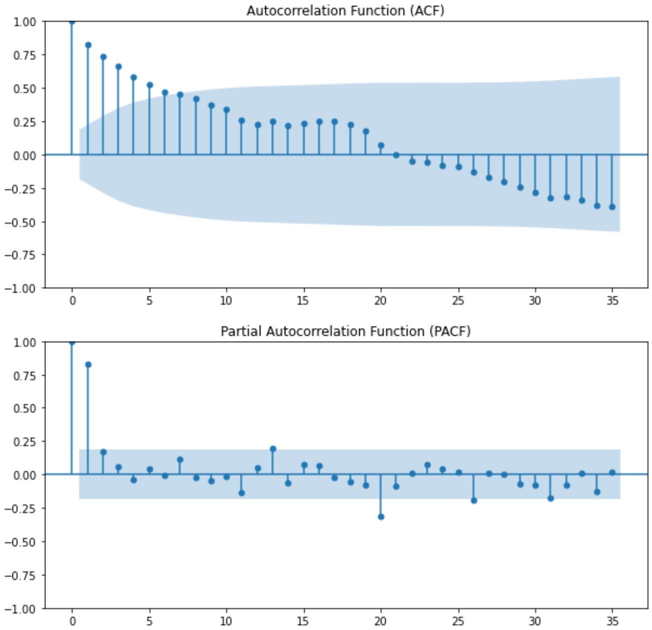

Figure 11. ACF and PACF plot before differencing

The ACF plot in Figure 11 shows strong positive correlation at lag 0 that decreases, with a large spike at the initial lags, indicating a seasonal or trend component. This helps determine the seasonal period or order of the MA component. The PACF plot shows a large spike at lag 1 and near-zero at subsequent lags, indicating an AR(1) process and helps determine the order of the AR component.

The PACF plot shows in Figure 12 an increase at lag 0 and near-zero at subsequent lags, but no significant increases outside the confidence interval, indicating that correlations may be due to nearby observations. The ACF plot shows strong positive correlation at lag 0 that gradually decreases, with values within the confidence interval, indicating no repeating patterns after differencing.

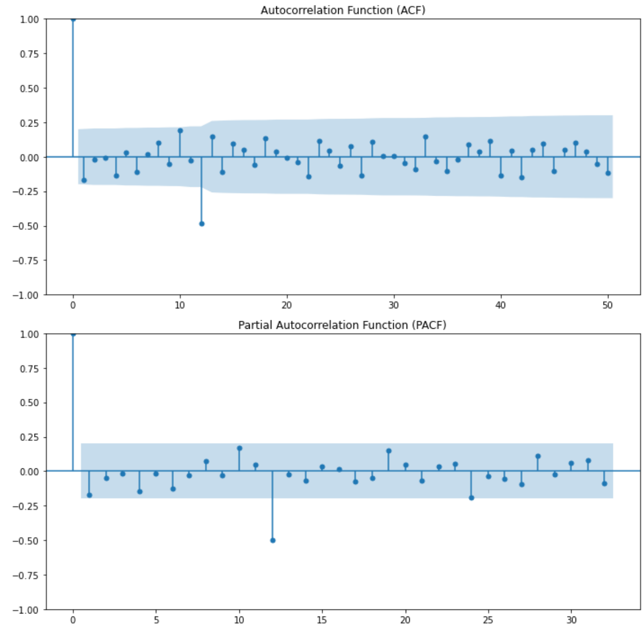

Figure 12. ACF and PACF plot after differencing

Figure 13. Comparison graph between actual test data and short-term SARIMAX model prediction results. This figure presents a comparison between the actual test data and the prediction results generated by the short-term SARIMAX model.

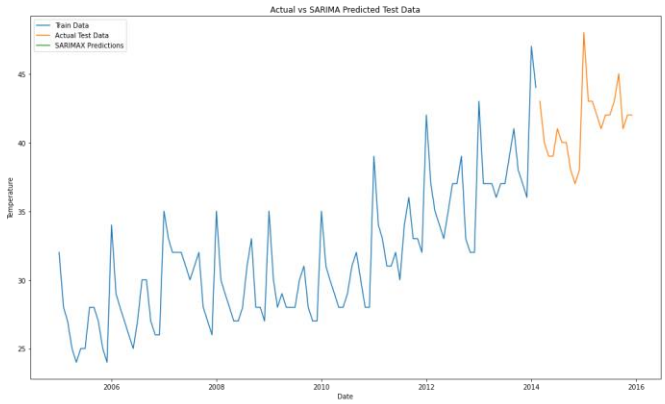

Figure 13. Comparison graph between actual test data and short-term SARIMAX model prediction results

The training data (blue), based on the SARIMA plot, shows the trend from 2005 to around 2012. The actual test data (green) and SARIMAX predictions (orange) start to appear after that period. Although there are differences between the two, the patterns are generally similar. This indicates that the

5.3.3 Evaluation of Short-Term and Long-Term SARIMAX Model

This indicates that the SARIMAX model can predict trends quite accurately despite some mismatches.

Table 3. SARIMA Model Evaluation

	Model	MSE	RMSE	MAE	MAPE
0	SARIMAX Short-Term	5.50	2.34	2.04	5.00
1	SARIMAX Long-Term	31.8	5.63	5.17	13.05

Based on the performance evaluation in Table 3, the short-term SARIMAX model is recommended for forecasting data because it provides better and more accurate results compared to the long-term SARIMAX model.

The results and discussion of this study indicate that short-term forecasting approaches consistently deliver more accurate performance compared to long-term models, across Naïve, SARIMA, and SARIMAX methods. This is evidenced by lower evaluation metrics such as MSE, RMSE, MAE, and MAPE in short-term models, reflecting their better ability to capture seasonal patterns and data fluctuations. The short-term SARIMAX model, in particular, demonstrates the highest accuracy due to the integration of exogenous variables that enhance prediction capabilities. In contrast, long-term models tend to be more stable but less responsive to recent data variations. Key steps such as differencing, ACF and PACF analysis, and stationarity testing play crucial roles in optimizing model performance. Therefore, the short-term SARIMAX model is recommended as the most reliable method for rainfall forecasting in this study.

5. Conclusion and Future Scope

This study evaluates the effectiveness of Naïve, SARIMA, and SARIMAX methods in forecasting Google search trends within the fitness industry, specifically for the keyword "Gym." The results indicate that SARIMAX, which incorporates external variables, outperforms other methods in terms of accuracy. The Naïve method serves as a simple benchmark, while SARIMA, with its integration of seasonal, autoregressive, and moving average components, offers improved trend detection. SARIMAX further enhances predictive accuracy by incorporating external factors, making it more adaptable to fluctuations influenced by marketing campaigns, seasonal patterns, and lifestyle changes. The evaluation, based on MAE, RMSE, and MAPE metrics, highlights that short-term forecasting models demonstrate greater responsiveness to seasonal variations compared to long-term models.

The findings of this study have significant implications for businesses and marketers in the fitness industry, as improved trend forecasting enables better decision-making for inventory management, advertising strategies, and resource allocation. Despite its advantages, SARIMAX requires careful selection of relevant external variables, and its complexity may limit accessibility for practitioners with limited expertise in statistical modeling. Additionally, variations in evaluation metrics across different studies make direct comparisons challenging, underscoring the need for standardized assessment approaches in forecasting research.

Future research can explore the integration of machine learning techniques with SARIMAX to further enhance predictive accuracy and adaptability to dynamic search trends. Additionally, investigating the impact of various external variables, such as economic indicators, social media trends, and global events, could provide deeper insights into search behavior patterns. Expanding the analysis to other industries with volatile search trends would further validate the applicability of these forecasting models in diverse domains.

Data Availability

The data supporting the findings of this study were obtained from Google Trends (https://trends.google.com/trends/), a publicly accessible platform. All datasets used in this research are available online and can be retrieved by replicating the keyword queries within the specified time range used in the study. No proprietary or restricted data were used.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper titled “Comparison of NAIVE, SARIMA, and SARIMAX Models in Short-Term and Long-Term Forecasting of Google Search Trends”.

Funding Source
This work was supported by the Research and Community Service Institution (LPPM), Telkom University.

Authors' Contributions
Author-2 conducted the literature review, designed the methodology, and performed the data analysis. Author-1 contributed to data interpretation and wrote the initial draft of the manuscript. Both authors reviewed, edited, and approved the final version of the manuscript.

Acknowledgements

The authors would like to express their gratitude to the Research and Community Service Institution (LPPM), Telkom University, for providing funding support for this study. We also thank our colleagues and collaborators who provided valuable insights and feedback throughout the research process.

References

[1] I. Majid, "Analisis Penerapan Google Custom Search API pada Sistem Pencarian Data," *Journal of Computer Information Systems (J-CIS)*, Vol.**1**, pp.**29-35, 2021**.

[2] L. H. Hasibuan, S. Musthofa, D. M. Putri, M. Jannah, "Comparison of Seasonal Time Series Forecasting Using SARIMA and Holt Winter's Exponential Smoothing (Case Study: West Sumatra Export Data)," *Barekeng Journal of Mathematics and Its Applications*, Vol.**17**, Issue.**3**, **pp.1773-1784, 2023.**

[4] N. Suhermi, Suhartono, R. P. Permata, S. P. Rahayu, "Forecasting the Search Trend of Muslim Clothing in Indonesia on Google Trends Data Using ARIMAX and Neural Network," in *Soft Computing in Data Science*, M. W. Berry, B. W. Yap, A. Mohamed, M. Köppen (Eds.), Springer Singapore, **pp.272-286, 2019.**

[5] W. Anggraeni, L. Aristiani, "Using Google Trend Data in Forecasting Number of Dengue Fever Cases with ARIMAX Method Case Study: Surabaya, Indonesia," *In the Proceedings of the 2016 International Conference on Information & Communication Technology and Systems (ICTS)*, Indonesia, **pp.114-118, 2016**.

[6] N. Newton, A. Kurnia, I. M. Sumertajaya, "Analisis Inflasi Menggunakan Data Google Trends dengan Model ARIMAX di DKI Jakarta," *Indonesian Journal of Statistics and Its Applications*, Vol.4, Issue.**3**, pp.**545-556, 2020**.

[7] V. Shah et al., "Forecasting Maximum Temperature Trends with SARIMAX: A Case Study from Ahmedabad, India," *In the Proceedings of Sustainability Conference*, **2024.**

[8] M. C. Medeiros, H. F. Pires, "The Proper Use of Google Trends in Forecasting Models," *Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)*, pp.**1-25, 2021**.

[9] D. S. Priyarsono, S. Magfirotul Laelia, "Studi Penggunaan Data Google Trends: Kasus Peramalan Tingkat Pengangguran Usia Muda," *Bina Ekonomi*, Vol.**27**, Issue.**2**, pp.**100-123, 2023**.

[10] N. N. Lingga, A. A. Rohmawati, "Pemodelan dan Peramalan Kedatangan Wisatawan ke Tempat Wisata dengan Google Trends Menggunakan Metode Variasi Kalender ARIMAX," *In the Proceedings of the eProceedings of Engineering, Telkom University*, Indonesia, **pp.1-5, 2023**.

[11] A. D. Riyanto, "Pemanfaatan Google Trends dalam Penentuan Kata Kunci Sebuah Produk untuk Meningkatkan Daya Saing Pelaku Bisnis di Dunia Internet," *In the Proceedings of the 2014 Conference on Internet Business Competitiveness*, Indonesia, **pp.1-5, 2014.**

[12] N. Thakur, S. Karmakar, "Deep Learning Approach Using Long Short Term Memory Technique for Monthly Rainfall Prediction in Chhattisgarh, India," *International Journal of Scientific Research in Computer Science and Engineering*, Vol.**9**, Issue**.1**, **2021**.

[13] R. P. Permata, D. D. Prastyo, Wibawati, "Hybrid Dynamic Harmonic Regression with Calendar Variation for Turkey Short-Term Electricity Load Forecasting," *Procedia Computer Science*, Elsevier, **pp.25-33, 2022**.

[14] A. A. Karim, E. Pardede, S. Mann, "A Model Selection Approach for Time Series Forecasting: Incorporating Google Trends Data in Australian Macro Indicators," *Entropy*, Vol.**2**5, Issue.**8, 2023**.

[15] S. Olukanmi, F. Nelwamondo, "Digital Influenza Surveillance: The Prospects of Google Trends Data for South Africa," *In the Proceedings of the 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, South Africa, pp.**1-6, 2020**.

[16] T. Boone, R. Ganeshan, R. L. Hicks, N. R. Sanders, "Can Google Trends Improve Your Sales Forecast?," *In the Proceedings of the Production and Operations Management Society Conference*, USA, **2018**.

[17] I. O. Muraina, O. M. Adesanya, M. A. Agoi, S. O. Abam, "The Necessity of Exploratory Data Analysis: How are preprocessing activities beneficial to Data Analysts and Professional Researchers in Academia?," *International Journal of Scientific Research in Computer Science and Engineering*, Vol.**11**, Issue.3, **2023**.

AUTHORS PROFILE

Author's 1 Regita Putri Permata received her Bachelor's and Master's degrees in Statistics from Sepuluh Nopember Institute of Technology (ITS), Surabaya, Indonesia. She is currently a Lecturer in the Data Science Study Program at Telkom University, Surabaya, since 2021. She is an active contributor in the field of data science, with involvement in several research and community service initiatives. She has authored more than 15 research publications in reputed international journals and conferences including Springer and IEEE. Her primary research interests include Time Series Forecasting, ARIMAX and SARIMA Modeling, Neural Networks, Google Trends Data Analysis, Education Data Mining, and Applied Statistical Modeling. She has over 5 years of teaching experience and a strong background in data analytics for solving real-world problems. She is committed to academic excellence and innovation in the field of data science.

Author's 2 Ananda Taqhsya Dwiyana is an undergraduate student in the Data Science Study Program at Telkom University Surabaya, Indonesia (TUS). She is actively involved in research activities and academic initiatives that focus on data science, education, information technology. Her academic interests include data mining, digital literacy, social media analysis, and modeling. During her studies, she has participated in various research projects, internships, and scientific competitions to deepen her understanding and application of science. She is committed to contributing to the development of data-based solutions for the world of education and society through research, collaboration, and innovation.

