JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 10, No. 1, JANUARY, 2025, Pp. 16 – 32
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Predicting Olympic Medal Trends for Southeast Asian Countries Using the Facebook Prophet Model

Bagus Al Qohar (1)*, Yulizchia Malica Pinkan Tanga (2), Putri Utami (3), Maylinna Rahayu Ningsih (4), Much Aziz Muslim (5)
[1,2,3,4] Department of Computer Science, Universitas Negeri Semarang, Semarang, Indonesia
[5] Faculty of Technology Management, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia
e-mail :
{bagusximipa6,yulizchiamalica,utamiputri575,maylinnarahayuningsih}@students.unnes.ac.id, gp200017@student.uthm.edu.my.
* Corresponding author.
This article was submitted on 18 October 2024, revised on 5 November 2024, accepted on 6 November 2024, and published on 31 January 2025.

Abstract

The Olympics are a world-class sporting event held every four years, serving as a meeting place for all athletes worldwide. The Olympics are held alternately in different countries. The Olympics were first held in Athens in 1896 and have now reached the 33rd Olympics, which will be held in Paris in 2024. Significant work has been conducted to develop prediction models, with a primary focus on enhancing the accuracy of predicting Olympic outcomes. However, low-performance regression algorithms are the main problem with prediction. By integrating custom seasonality with the Facebook Prophet prediction model, this study aims to enhance the accuracy of Olympic predictions. The proposed new model involves several steps, including preparing the data and initializing and fitting the Facebook-Prophet model with several parameters such as seasonal mode, annual seasonality, and prior scale. The model is tested using the Olympic dataset (1994–2024). The evaluation results indicate that this prediction model provides a reliable estimate of the total medals earned. On the Olympic Games (1994-2024) dataset, the model exhibits a very low error, as indicated by its MAE, MSE, and RMSE, and achieves an R^2 score of 0.99, which is close to perfect. This research shows that the model is effective in improving prediction accuracy.

Keywords: Custom Seasonality, Facebook-Prophet, Forecasting, Olympic Medals, Time Series

Abstrak

Olimpiade adalah acara olahraga dunia yang diadakan setiap 4 tahun sekali dan merupakan tempat pertemuan bagi semua atlet di seluruh dunia. Olimpiade diadakan secara bergantian di berbagai negara. Olimpiade pertama kali diadakan di Athena pada tahun 1896 dan sekarang telah mencapai Olimpiade ke-33, yang akan diadakan di Paris pada tahun 2024. Untuk memprediksi hasil Olimpiade, banyak upaya telah dilakukan untuk mengembangkan model prediksi dengan penekanan pada peningkatan akurasi. Namun, algoritma regresi berkinerja rendah adalah masalah utama dalam prediksi. Dengan mengintegrasikan musiman khusus dengan model prediksi Facebook-Prophet, penelitian ini bertujuan untuk meningkatkan akurasi prediksi Olimpiade. Model baru yang diusulkan melibatkan beberapa langkah, termasuk menyiapkan data, inisialisasi, dan menyesuaikan model Facebook-Prophet dengan beberapa parameter seperti mode musiman, musiman tahunan, dan skala sebelumnya. Model ini diuji dengan menggunakan dataset Olimpiade (1994-2024). Hasil evaluasi menunjukkan bahwa model prediksi ini dapat memberikan nilai yang baik dalam memprediksi total medali yang diperoleh. Pada dataset Olimpiade (1994-2024), model ini memiliki error MAE, MSE, dan RMSE yang sangat rendah serta memiliki nilai R^2 sebesar 0,99, yang mendekati sempurna. Penelitian ini menunjukkan bahwa model efektif dalam meningkatkan akurasi prediksi.

Kata Kunci: Musiman Khusus, Facebook-Prophet, Peramalan, Medali Olimpiade, Deret Waktu

1. INTRODUCTION

Almost a century ago, the Olympics brought athletes from all over the world to compete for medals. Many nations have made significant progress in the Olympics, earning respect and recognition for their achievements (Herzog, 2024; James, 2023; Theodorakis et al., 2024). Southeast Asia currently receives up to 104 medals. Not only will strategic planning and resource allocation benefit from an analysis of the various elements that influence the success of the Olympics (Badoni et al., 2023), but future achievements will also be enhanced. This paper clarifies the distribution of medals in individual Olympic sports. The aim is to understand how country variables, including population and economic size, influence the share of medals in various individual sports.

Another study examines the impact of public sports expenditure on the number of Olympic medals won (Wu et al., 2023). This study aims to investigate whether public sports investment increases in proportion to a nation's Olympic success rate. It also investigates Wen & Wang (2020), who suggest that a nation's climate significantly determines the success of the Olympic Games. The aim is to determine whether there is a correlation between a nation's climate origins and its success or specialization in a particular type of sport across the six editions of the Olympic Games reviewed between 1996 and 2016 (Scelles et al., 2020). To improve current approaches by considering economic, demographic, and historical factors, this paper reevaluates the estimate of the number of medals countries are expected to win at the Summer Olympics. Another study examines the factors that support or hinder a nation's Olympic performance (Rewilak, 2021). The objective is to identify the primary and secondary elements that influence a country's medal count, as well as to investigate the causes of varying degrees of success among different countries. Predictive modeling methods, including Prophet, have recently been utilized to identify and project various outcomes, ranging from sports results to Bitcoin projections (Cheng et al., 2024). Predictive modeling provides valuable information that enables numerous stakeholders, including countries, to make informed decisions.

In line with this, Asha et al. (2023) examined Olympic Games performance using data analytics, spotting public investment and economic power as the main determinants of a nation's medal count. Similarly, Badoni et al. (2023) compared machine learning algorithms to forecast Olympic medal counts and identified random forest and gradient boost as quite successful models. With XGBoost turning out to be the most accurate, Sagala & Ibrahim (2022) investigated the efficacy of several boosting techniques for estimating Olympic medals. Xinyi & Chenglong (2022) also examine visual analytics trends in Olympic medal distribution, highlighting the relevance of geographic and population elements in medal success. Recent Jia et al. (2022) analysis of public opinion worldwide during four Olympic Games (2008–2022). The research revealed that geopolitical and social elements significantly affect public opinion and sentiment about the Olympics, underscoring how perspective can help shape global expectations for national Olympic performances.

Furthermore, as underlined in the research by Agyemang et al. (2023), Predictive analytics are needed to improve national Olympic readiness. This study employs predictive models to examine how nations can optimize their resources and enhance their chances of success. Furthermore, research on predictive modeling has demonstrated the success of several strategies, including time-series analysis, in projecting various results (Satrio et al., 2021; Wulandari et al., 2021). For example, the adaptability of ARIMA and Prophet was demonstrated by their ability to forecast COVID-19 cases in Indonesia, thereby proving their effectiveness in this context. Further underscoring the resilience of these forecasting methods, Angelo et al. (2023) offered a comparison of the ARIMA and Prophet algorithms in estimating Bitcoin prices. Li et al. (2021) also evaluated two-stage network structures with the 2018 Winter Olympic Games, helping to clarify how various modeling techniques evaluate Olympic performance.

Thus, by guiding nations in the wise use of limited resources, predictive models such as Facebook Prophet help them better prepare for challenges (Agyemang et al., 2023). Predictive analysis is today a useful tool for evaluating past performance and projecting future outcomes. Although building the predictive model is easy, the complexity of the data makes achieving dependability in results difficult (Arteaga et al., 2024). From statistical approaches to machine learning, the literature Chowdary et al. (2024) and Lei et al. (2024) have investigated many ways to project sports performance. Low-quality data could potentially lead to errors due to the poor performance of the categorization algorithm. Therefore, we must improve the models to increase the accuracy of the prediction.

In this paper, we investigate the optimization of a predictive model by analyzing historical performance of several Southeast Asian countries and making future projections using the Prophet model. The choice of the Prophet model is justified due to its ability to effectively capture patterns in historical data, provide highly accurate predictions with low error values, and explain variations in medal acquisition almost perfectly, making it a valuable tool for policymakers and coaches to enhance future performance in Olympic events. Unlike ARIMA and other traditional time series models, Prophet can account for seasonality and external variables that influence athletes' performances, resulting in more accurate forecasts. Seasonality in Olympic forecasting refers to the recurring patterns and trends that occur in the context of the Olympic Games. Seasonality has a significant influence on medal predictions, enabling the model to capture recurring patterns, enhance forecast accuracy, and provide valuable insights into the factors that affect Olympic performance. This understanding is vital for countries planning their strategies for future Olympic events. The research contributes by providing more accurate predictions of Southeast Asian countries. Olympic medal counts and effectively guiding strategic planning and resource allocation for future Olympic games based on accurate projections for countries like Thailand, Indonesia, Malaysia, Singapore, and the Philippines, highlighting improvement areas and ongoing support for sports programs.

2. METHODS

Using the Prophet model, Figure 1 illustrates the approach for estimating the total number of medals Southeast Asia is expected to score at the Olympics. The data is then compiled to ensure they are in a suitable format for study (Guo et al., 2021). To increase the accuracy of the forecasts, the Prophet model is then started with custom parameters that consider both additional seasonal components and annual seasonality.

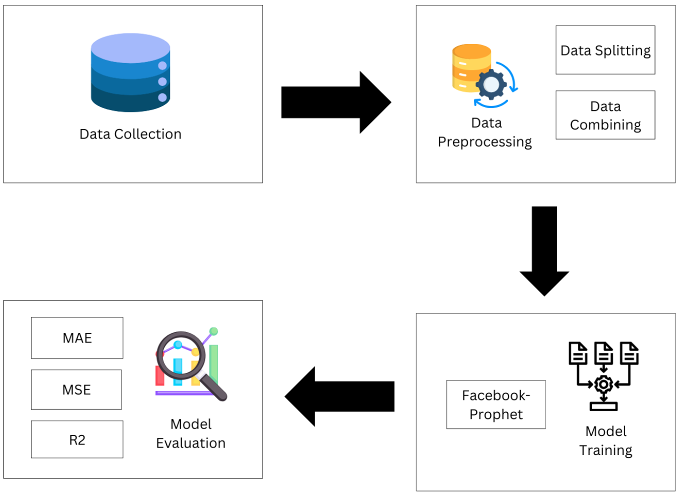

Figure 1 Research Method

2.1 Data Collection

Data collection for this study involved gathering Olympic medal counts from various Olympic Games, spanning both summer and winter events, from 1994 to 2024 (Ismail, 2024). Medal information was compiled from multiple CSV files, including data from the Olympic Games in Atlanta (1996), Beijing (2008), Athens (2004), Torino (2006), and Paris (2024), as well as other locations. The files contain the total number of gold, silver, and bronze medals won by the National Olympic Committee (NOC) of each nation. There have been a total of 879 entries, representing the nation's medal totals over several years and various sports. This dataset serves as the foundation for examining patterns and trends in Olympic achievements.

For accurate forecasting results, the quality of historical data is crucial. However, the period covered by this data is diverse, which poses various challenges, including the impact of climate factors. This climate factor is an external factor that can affect the predicted medal results because it impacts athlete performance, such as providing an advantage for local athletes or those accustomed to similar climates, or causing shifts in sleep and training patterns, among others. Table 1 summarizes the data collection for each Olympic event, showing the number of entries in each dataset.

Table 1 Table Data Collection

Year	Olympic event	Number of Entries
1994	Lillehammer (Winter)	22
1996	Atlanta (Summer)	78
1998	Nagano (Winter)	24
2000	Sydney (Summer)	79
2002	Salt Lake City (Winter)	24
2004	Athens (Summer)	74
2006	Torino (Winter)	26
2008	Beijing (Summer)	87
2010	Vancouver (Winter)	26
2012	London (Summer)	86
2014	Sochi (Winter)	26
2016	Rio (Summer)	85
2018	PyeongChang (Winter)	30
2020	Tokyo (Summer)	92
2022	Beijing (Winter)	29
2024	Paris (Summer)	91
	Total	**879**

2.2 Data Preprocessing

To manage data from many sources, we have created several significant data preprocessing methods (Tawakuli et al., 2024). This process begins by gathering CSV data from a specific directory containing Olympic records. We filter these CSV files depending on their .csv extension to handle the pertinent ones. Automated file loading ensures the extraction of all relevant files; hence, it is one of the most important techniques for handling large databases. Moreover, this script is dynamic since it accesses and lists every file in the specified directory using the OS.listdir tool. As noted by Phan et al. (2021), who emphasized the importance of dynamic and automated file-handling techniques in wind power forecasting, this approach provides greater flexibility in managing variable datasets over time.

Then, the code takes care of the important chore of deleting the year from every one of these file names. The code cleverly reads the filenames for the year. It adds them as a new column in every matching data frame, considering that the dataset comprises several files, each corresponding to a different year of Olympic data. This will maintain the data chronologically and simplify time-based analyses, including trend tracking and outcome prediction. Once we include the year column, downstream analyses or modeling jobs will find the dataset more consistent and manageable. Similar to the year-specific feature engineering used by Ding et al. (2024), our approach helps ensure the integrity and accuracy of time-based data studies for carbon emissions forecasting.

Using *pandas.concat* function (Dong et al., 2024), the last phase of this preprocessing process concatenates the individual data frames into a single, complete data frame. Similarly, the code concatenates data frames, aggregating information from multiple sources into a single dataset suitable for extensive study (Yin et al., 2024). It has developed a three-stage data preprocessing plan that aligns with this method, demonstrating the efficacy of combining multiple datasets for long-term freight market prediction. Figure 2 illustrates the merging of multiple datasets into a single combined data frame.

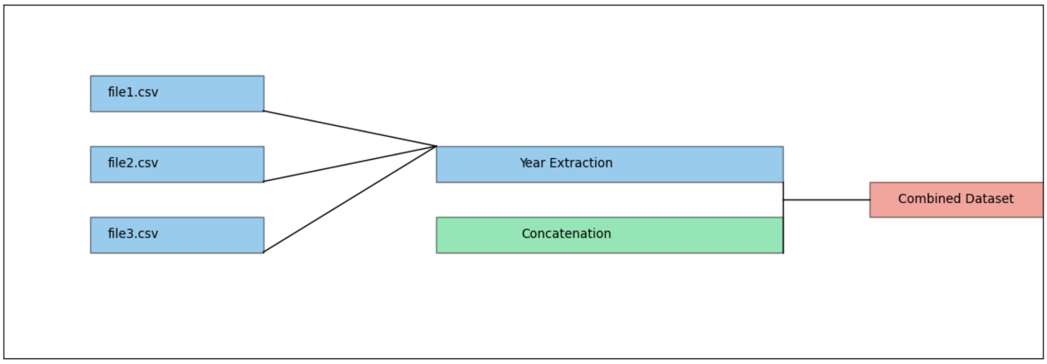

Figure 2 Process of Merging Multiple Datasets Into a Combined Data Frame

2.3 Model Training

We design the key steps in model training using the Prophet algorithm to capture trends and seasonal patterns forecasting Southeast Asia's total Olympic medals. The prepare_prophet_data custom function first structures the Olympic dataset in a format compatible with Prophet, using the 'total' medal column, which includes gold, silver, and bronze medals. Prophet is particularly suited for handling time-series data with irregular spacing and missing values, making it ideal for long-term forecasting tasks, as demonstrated in various forecasting domains (Annapoorna et al., 2024; Gautam et al., 2023; Mousa et al., 2023).

After preparing the data, we initialize the Prophet model with specific parameters to enhance forecast quality. For example, yearly_seasonality = True allows the model to capture recurring annual patterns, such as those seen in the Olympics. We also set changepoint_prior_scale = 0.8 so that the model can respond more flexibly to trends, such as significant performance spikes in countries like Indonesia. Additionally, we incorporate custom quadrennial seasonality to reflect the Olympic cycle, which occurs every four years, allowing the model to account for the unique periodicity of this event. This method utilizes the custom_seasonality parameter, which enables the model to recognize and capture specific quadrennial patterns more effectively.

The four-year seasonality in Olympic forecasting refers to the recurring patterns and trends that occur every four years in the context of the Olympic Games. This cycle is crucial for understanding and predicting medal achievements over time, considering the quadrennial nature of the event. (Gong et al., 2020; Verghese et al., 2021). In the Prophet model, the custom four-year seasonality is implemented by incorporating a specific parameter that accounts for the unique periodicity of the Olympic cycle. This feature enables the model to adjust its forecasts in response to the cyclic nature of the Olympics, thereby capturing the long-term trends and patterns in the medal data. By considering the four-year seasonality, the forecasting model can better predict medal counts by accounting for the historical patterns that repeat every Olympic cycle. This approach enhances the accuracy and reliability of the forecasts, providing valuable insights for strategic planning and resource allocation in the context of the Olympics. In summary, seasonality has a significant influence on medal predictions, enabling the model to capture recurring patterns, enhance forecast accuracy, and provide valuable insights into the factors that affect Olympic performance. This understanding is vital for countries planning their strategies for future Olympic events.

This training aims to forecast the total medal count for the next four Olympic cycles. The code also includes a visualization step that plots the actual and predicted total medals, allowing for a visual comparison of the forecasted outcomes. The plotting process, as demonstrated in Figure 3, results in the successful adaptation of Prophet's forecasting techniques for this task. This plot is useful for evaluating the model's performance in forecasting. Plotting predictions against actual results allows one to quickly assess how well the model captures historical trends and whether the future forecasted trend follows reasonable patterns. Moreover, including visual grids and labels enhances the plot's clarity and informational value, making it suitable for presentation.

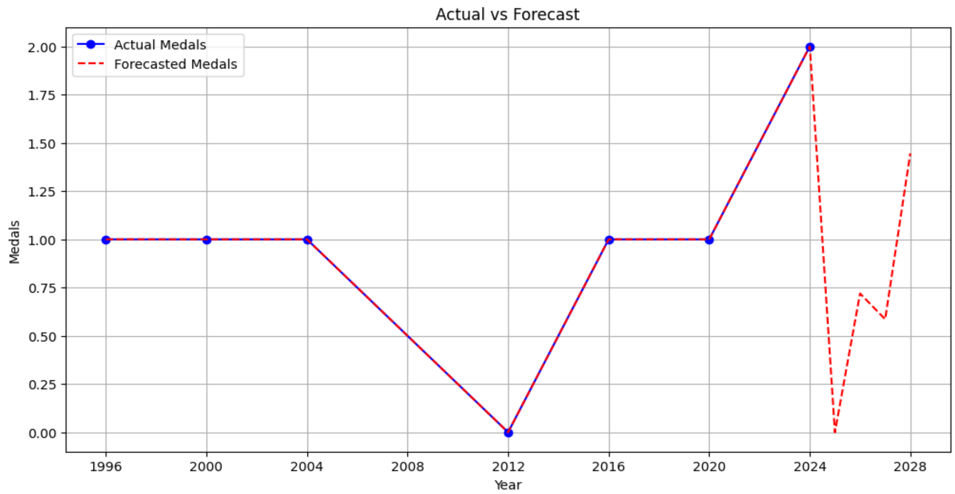

Figure 3 Plot of The Model Training Result

2.4 Model Evaluation

Confirming that the forecasts are accurate and reliable depends on assessing the model's performance first. The measures provide an insightful analysis of the Prophet model's performance in predicting the total Olympic medal count for Southeast Asian countries. Ignoring its direction, the mean absolute error (MAE) only considers the average size of the errors, allowing for a clear awareness of the difference between the predictions and the actual value (Mahajan & Shrivastav, 2023; Rajesh & Saravanan, 2022). Due to the square difference between predicted and actual values, MSE and RMSE highlight larger errors (Karunasingha, 2022; Qi et al., 2020). Therefore, these measures are especially sensitive to significant forecast deviations.

In the fields of predictive modeling and data analysis, the MAE is a metric frequently utilized. This equation provides a straightforward method for determining the accuracy of predictions by calculating the average magnitude of errors that occur in a set of predictions by using the equation. MAE is represented by Eq. (1).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (1)$$

The total of these absolute errors across all data points is calculated using this equation, which sums them up. Obtaining the mean average absolute error can be accomplished by dividing this total by n, which is the number of measurements. The value predicted to be the number of observations in the dataset is denoted by the symbol y_i. The symbolic value x_i denotes the corresponding actual or observed value for the same observation. The total number of data points contained in the dataset is denoted by the letter n. Absolute difference, denoted by the notation $|y_i - x_i|$, is a measurement that determines the degree to which each prediction deviates from its actual value. This ensures that all errors are regarded as positive values.

The mean squared error (MSE) is a metric frequently used to evaluate the precision of predictions made by regression and forecasting models. Therefore, it is sensitive to larger errors because it quantifies the average squared difference between the observed values and the predicted values. The equation for MAE is shown in Eq. (2).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2}$$

To measure the overall error in the predictions, this equation can be used to aggregate the squared deviations across all data points. The mean or average squared error can be calculated by dividing this total by the number of observations. Symbol Y_i is the value that has been observed, or the actual value for the number of data points that are contained in the dataset. For the same data point, the predicted value is denoted by the symbol $\hat{Y}_i$. Symbol n represents the total number of data points that are contained within the dataset. The squared difference, denoted as $(Y_i - \hat{Y}_i)^2$, is a statistical measure that determines the squared deviation for every prediction. This technique amplifies the impact of larger errors.

R^2 is a statistical metric used to evaluate the goodness of fit for regression models. It quantifies the proportion of variation in the dependent variable that is explained by the independent variables in the model. Values closer to 1 indicate that the model is more fitted. The equation for R^2 is shown in Eq. (3).

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3}$$

The Residual Sum of Squares, also known as RSS, is a statistical technique that allows for the quantification of the model's unexplained variance by measuring the total squared differences between the predicted and observed values. The RSS equation is shown in Eq. (4).

$$RSS = \sum_{i=1}^{n}(y_i - f(x_i))^2 \tag{4}$$

Using this equation, the squared residuals are accumulated across all data points included in the dataset. The RSS that is produced is a reflection of the overall magnitude of the prediction errors. RSS values that are smaller indicate that the model is a better fit for the data, whereas larger RSS values suggest greater deviations between the observed and predicted values. The value of y_i is the number of values that have been observed in the dataset that the model is attempting to predict. The predicted value for y_i is denoted by the symbol $f(x_i)$, which the regression model produces through the utilization of the input x_i. Symbol n is the total number of data points, which serves as the limit of the summation from the highest possible value. The squared term, denoted as $(y_i - f(x_i))^2$, is used to compute the square of the residual, which is the difference between the values that were observed and those that were predicted, for every single data point.

TSS is a statistical measure used to quantify the total sum of squares within a dataset. As a baseline for determining how well a regression model fits the data, it represents the overall dispersion of observed values around their mean. The equation for RSS is shown in Eq. (5).

$$TSS = \sum_{i=1}^{n}(y_i - \underline{\bar{y}})^2 \tag{5}$$

Using this equation, the total squared standard deviation (TSS) is calculated by averaging the squared deviations of all the observations. A regression model can explain the total variability in the dataset, as reflected by this metric. Symbol y_i represents the number of values observed in the dataset. It is commonly known as the sample mean, denoted by $\bar{y}$ represents the average of all the values observed in the dataset. In this dataset, the total number of observations is denoted by the letter n. To determine the distance between each point and the average, the squared term $(y_i - \bar{y})^2$ is utilized to calculate the squared deviation of each observed value from the mean.

This method extends the mean squared error (MSE) by further converting the units back to the original scale of the data, thus improving the interpretability of the results. The lower the RMSE value, the better the forecast value. In contrast, the R-squared value indicates how well the model predictions align with the actual data. Values closer to 1 indicate that the model is more fitted. When comparing different models or evaluating a model's ability to explain data variation, this metric proves particularly useful. These metrics combined provide complete insight into overall accuracy and areas where the model may need improvement.

Once the evaluation is complete, the results provide a clear understanding of the model's performance regarding Indonesia's total Olympic medals. Furthermore, the process enlightens advanced fine-tuning or adjustments in the parameter optimization or the data preprocessing step. High readings of the RMSE or MSE indicate that the model likely struggles to predict several medal trends, which may require revisiting the settings concerning seasonality or incorporating more relevant features into the data.

3. RESULTS AND DISCUSSION

This research focuses on forecasting trends in Olympic medal achievements for Southeast Asian countries comprising Indonesia (INA), Thailand (THA), Malaysia (MAS), Singapore (SIN), and the Philippines (PHI). We utilize historical data from medal records to identify patterns and make informed forecasts, which can help these nations plan effectively for future Olympic events. The insights derived from these predictions will help highlight potential areas for improvement and investment in your sports programs. In analysing these predictions, we use a robust prediction model known as the Facebook Prophet. It is pretty flexible for handling most types of time series data. We configured the model to annually, weekly, seasonal, and daily, to highlight the inherent patterns within the series. More importantly, a unique seasonal component was introduced to capture those specific variations within the Olympic cycle. This allows us to produce more customized and meaningful forecasts that align with the trend of medal achievements at the Olympics.

3.1 Indonesia (INA) Olympic Medal Trends

The model highlights the total number of medals Indonesia has won across various Olympic eras, based on the country's medal prediction results. Prophet's capacity to capture seasonal patterns and long-term trends, which thus improves Indonesia's future medal projections, was the main factor in selecting him. By comparing the actual medal data collected by Indonesia with the total expected medal results from the model, Figure 4 provides a clearer picture of the outcome. The solid blue line on the graph represents Indonesia's actual medal count across every Olympic cycle. In contrast, the red dotted line shows the expected model results. This visualization helps you see how the forecasts align with reality and identify areas where reality and the predictions diverge.

Visualization indicates that Indonesia has developed a trend pattern in its medal tally, with significant changes occurring in certain years. For example, the effectiveness of the training program or approach during those years, compared to previous periods, could explain the significant increase in medals between 2000 and 2020. To understand this trend, a detailed examination could be provided by analyzing research on different variables that may have contributed to the differences in these medal changes, such as changes in policy and sporting rules, athlete preparation, and investment in sports.

It also involves identifying years of depreciation, such as Indonesia's performance at the 2012 and 2016 Olympics. For example, the reasons could range from a lack of government support to ineffective strategy changes. Understanding the ups and downs of these trends is crucial for designing effective future strategies. The model in Table 2 was evaluated by comparing its performance with other models using four evaluation metrics: mean absolute error, mean squared error, root mean squared error, and R^2.

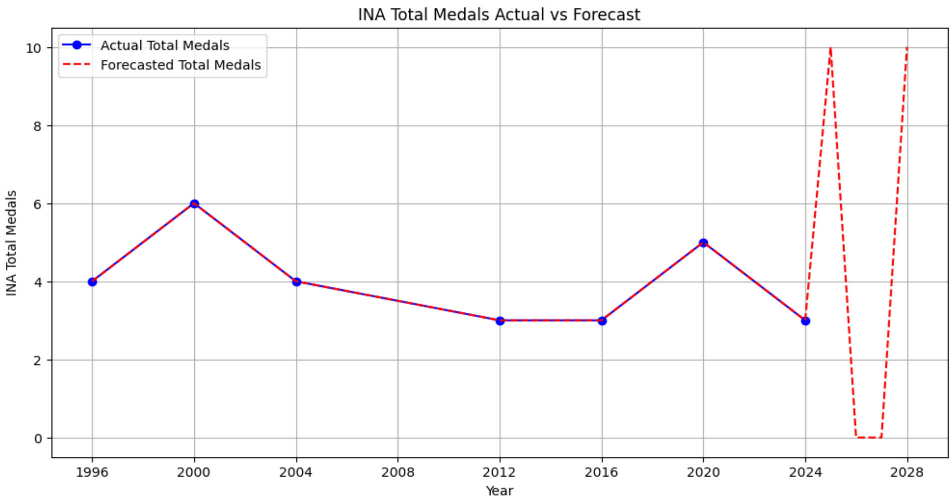

Figure 4 Indonesia Total Medals Actual vs. Predicted

Table 2 Metric Evaluation of Indonesian Total Medals: Actual vs. Predicted

Metric	Result
MAE	0.0001
MSE	0.0
RMSE	0.0001
R^2	0.9999999951

The same period is represented by the evaluation results, where the Prophet model yields very accurate predictions with low values for MAE, MSE, and RMSE, and high values close to 1 for R^2. This indicates that the model almost perfectly explains Indonesia's variation in medal acquisition. This assessment concludes that the Prophet model has successfully and accurately predicted Indonesia's medal wins. The accuracy of this model confirms that the Prophet model's use of season and trend components effectively captures patterns in historical data. It can also serve as a valuable tool for policymakers and coaches to suggest strategies that can help enhance Indonesia's future performance in Olympic events.

3.2 Thailand (THA) Olympic Medal Trends

The results of Thailand's medal prediction show notable changes in the number of gold, silver, bronze, and total medals won over multiple editions of the Olympic Games. These predictive data provide important new perspectives on Thailand's performance in international events and help identify trends in changes over time. The visualization compares the projected and actual medal counts for Thailand. Figure 5 shows the degree of mimicability of the medal trend of the Prophet model. The model's prediction is shown on the dashed line of the graph, while the solid line represents the actual medal count. This visualization shows that, although the model reasonably captures some variance, the pattern of the actual data and its projections generally match.

According to trend analysis, Thailand's medal count peaked in 2004 with eight medals: three gold, one silver, and four bronze. Developing athlete training schedules or more government support in that specific year could have helped explain this notable increase. However, the overall medal count dropped in 2020; it rebounded to six in 2024. This trend demonstrates the adaptability and resilience of Thailand's sports policy in the face of challenging conditions. We evaluated the performance of the prediction model using R², mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), as shown in Table 3.

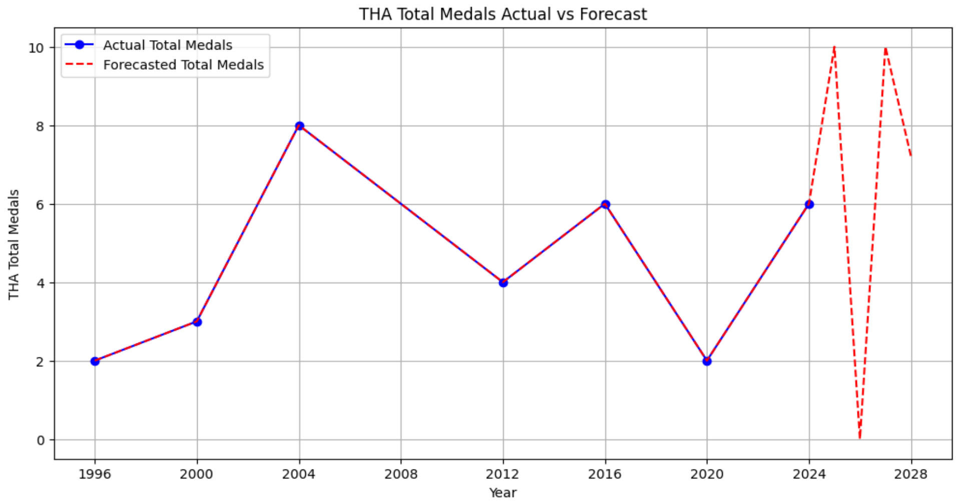

Figure 5 Thailand Total Medals Actual vs. Predicted

Table 3 Table Metric Evaluation for Thailand Total Medals Actual vs. Predicted

Metric	Result
MAE	0.0
MSE	0.0
RMSE	0.0001
R^2	0.9999999993

The results of the evaluation metrics show accurate predictions with very low values in estimating the total medals. With MAE and MSE values of 0.0 each, the average prediction error is almost non-existent, so the difference between the actual and predicted medal counts is very small. In addition, the very low RMSE of 0.0001 indicates that the prediction error is also very small, and the R^2 value of 0.9999999993 is close to 1, indicating that the model can almost perfectly explain the variation in Thailand's medal tally. This indicates that the Prophet model successfully and accurately predicted Thailand's medal tally.

3.3 Malaysia (MAS) Olympic Medal Trends

Malaysian medal prediction results (MAS) reveal notable patterns in the number of silver, bronze, and overall medals won by Malaysian athletes across several editions of the Olympic Games. These predictive data help identify changes in Malaysia's performance over time and provide an insightful analysis of its success in international sports events. It also enables one to spot changes in Malaysia's performance over time. Regarding Malaysia's medal counts, the visualization compares the actual data and the anticipated values. Figure 6, for example, shows the precision of replicating the medal trend of the Prophet model. Whereas the solid line represents the actual medal count, the dashed line on the graph illustrates the model forecast. This visualization enables us to observe that, although the model effectively captures some fluctuations, the general patterns of the actual data and its forecasts are remarkably similar to each other.

Over the years, Malaysia's overall medal count has fluctuated; it peaked in 2016 with five medals, four silver and one bronze. The increase could be attributed to improved athlete training schedules or more government support. The medals dropped to two annually in 2020 and 2024. This trend suggests that, despite difficulties, Malaysia's sports performance has shown some consistency. Using four criteria—mean absolute error, mean squared error, root mean squared error, and R^2—the model shown in Table 4 was evaluated against others.

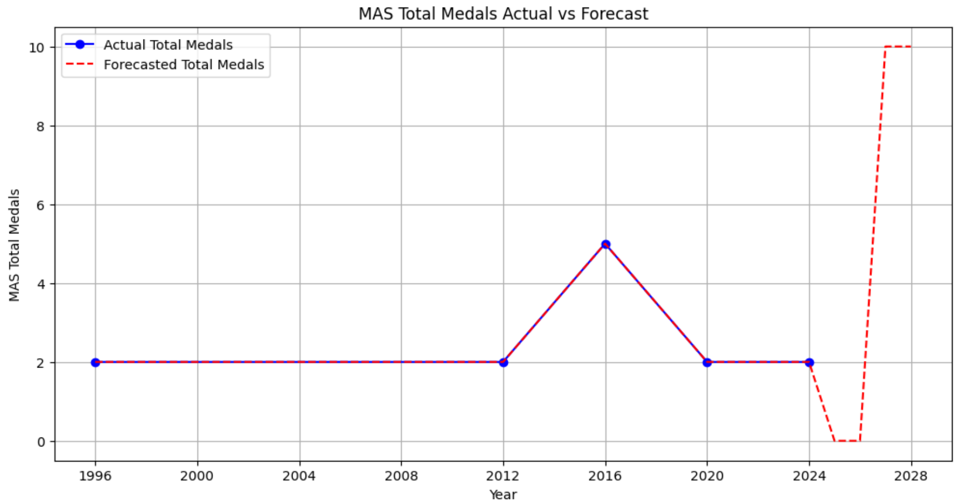

Figure 6 Malaysia Total Medals Actual vs. Predicted

Table 4 Table Metric Evaluation for Malaysia Total Medals Actual vs. Predicted

Metric	Result
MAE	0.0001
MSE	0.0
RMSE	0.0001
R^2	0.9999999974

The results of the evaluation metrics show accurate predictions with very low values in estimating the total medals. With MAE and MSE values of 0.0 each, the average prediction error is almost non-existent, so the difference between the actual and predicted medal counts is very small. In addition, the very low RMSE of 0.0001 indicates that the prediction error is also extremely small, and the R^2 value of 0.9999999974, which is close to 1, indicates that the model almost perfectly explains the variation in Indonesia's medal tally. This shows that the Prophet model successfully and accurately predicted Malaysia's medal tally. The model's accuracy shows that the seasonal and trend components of the Prophet model can successfully capture the patterns in the historical data. Moreover, this model can be a handy tool for coaches and policymakers to make plans that can help improve Malaysia's performance in future Olympics.

3.4 Singapore (SIN) Olympic Medal Trends

Singapore's (SIN) medal distribution at different Olympic Games tells an intriguing story about performance swings. Singapore had little success in 2012, winning two bronze medals. The nation won its first gold medal in 2016, a notable achievement on the international sports scene. However, Singapore's medal total dropped to just one bronze in 2024. These findings suggest that Singapore's overall Olympic performance was not entirely consistent. Figure 7 compares the total actual and predicted medal counts for Singapore at multiple Olympic Games, demonstrating the precision of the Prophet model in tracking Singapore's performance trends.

In 2016, Singapore achieved a significant turning point by winning its first and only gold medal. We can attribute this achievement to better athlete preparation, improved sports initiatives, or targeted government funding to promote sports nationwide. The lack of medals in the other categories (bronze or silver) during this period may indicate a focused strategy that produced a win in a single event but had no wider effects. This gold medal win remains a high point in Singapore's Olympic history, despite the decline in subsequent years.

By 2024, Singapore had won only one bronze medal, indicating that its performance had deteriorated again. This decrease in the number of medals could indicate difficulties maintaining momentum for 2016. The relatively low number of medals won by Singapore in these three Olympic cycles may suggest that the country's programs for developing athletes, sports infrastructure, or international competitiveness are still in their infancy. Although there is no doubt that success is possible in certain situations, consistency needs work. Table 5 presents the evaluation metrics, including MAE, MSE, RMSE, and R^2, which highlight the high precision of the Prophet model in predicting Singapore's total Olympic medal counts.

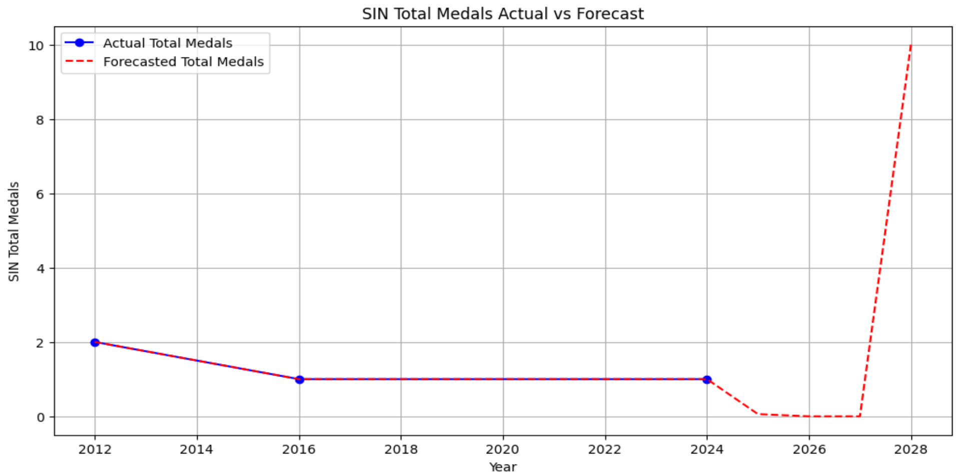

Figure 7 Singapore Total Medals Actual vs. Predicted

Table 5 Table Metric Evaluation for Singapore Total Medals Actual vs. Predicted

Metric	Result
MAE	0.0
MSE	0.0
RMSE	0.0
R^2	0.9999999973

The evaluation results show that the Prophet model accurately predicts the number of medals won by Singapore, with very low MAE, MSE, and RMSE values, and R^2 values approaching 1. The MAE, MSE, and RMSE values of 0.0 each indicate that the average prediction error is almost non-existent, so the difference between the actual and predicted number of medals is very small. The R^2 value of 0.9999999973, close to 1, indicates that the model explains the variation in Singapore's medal tally almost perfectly. This shows that the Prophet model successfully and accurately predicted Singapore's medal tally. The model's accuracy is demonstrated by the seasonal and trend components of the Prophet model successfully capturing the patterns in the historical data. Moreover, this model can be a valuable tool for coaches and policymakers to make plans that can help improve Singapore's performance in future Olympics.

3.5 Philippines (PHI) Olympic Medal Trends

The Philippines has shown a clear increasing trend in Olympic medal performance over several years; more recent Games have shown notable gains. From 1996 to 2016, the nation managed to win only one silver medal in each of these years, reflecting its relatively low degree of involvement in the international athletic scene. However, these outcomes laid the groundwork for a notable shift in the Olympic cycles that followed. Figure 8 shows a comparison between the actual total medal counts for the Philippines and the predicted total medal counts for the nation, highlighting the precision of the Prophet model in capturing the upward trend in Olympic performances. This comparison shows how faithfully the Prophet model captures the growing trend in the Philippines' Olympic performances.

With four overall medals, including two silver and one bronze, and a first gold medal, the Philippines achieved a notable milestone in 2020. Improvements in training programs, infrastructure, and increased support for sports development help explain the increase in performance, highlighting the improved capacities of national athletes. Multiple medals in various categories indicate that the country's competitiveness in different sports disciplines has improved.

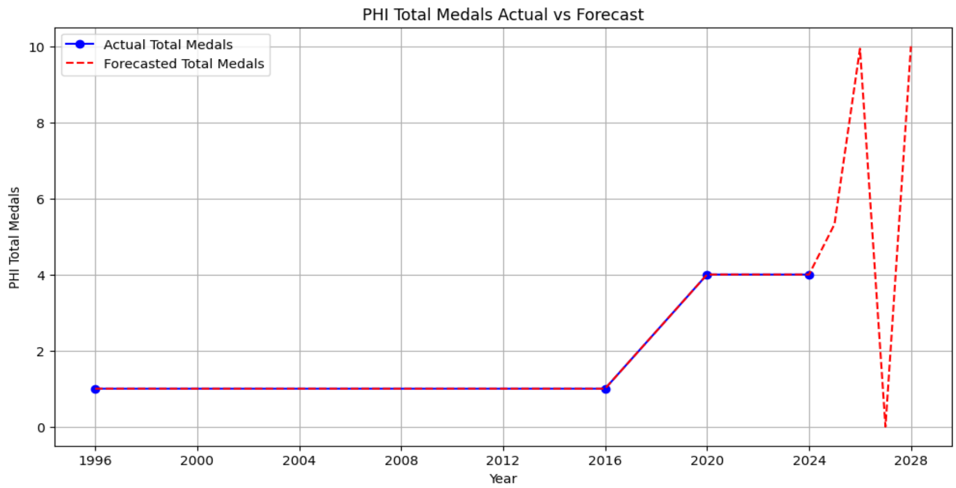

Figure 8 Total Philippines Medals Actual vs. Predicted

In 2024, the Philippines achieved continued success, securing two gold and two bronze medals, thus maintaining a total medal count of four. The rise in gold medals reinforces the country's growing dominance in specific events, while the stable medal count reflects ongoing momentum. The observed improvement may indicate the effectiveness of long-term strategies designed to develop elite athletes, as well as a sustained emphasis on elevating the nation's performance in international sports events. Table 6 presents the evaluation metrics, including MAE, MSE, RMSE, and R^2, which demonstrate the high precision of the Prophet model in forecasting the Philippines' Olympic medal totals.

Table 6 Table Metric Evaluation for Philippines Total Medals Actual vs. Predicted

Metric	Result
MAE	0.0
MSE	0.0
RMSE	0.0
R^2	0.9999999998

The results show that the Prophet model can accurately predict the number of medals the Philippines will receive, with very low MAE, MSE, and RMSE values, and R^2 values almost equal to 1. The MAE, MSE, and RMSE values are each 0.0, indicating that the average prediction error is almost non-existent, so the difference between the actual and predicted number of medals is very small. The R^2 value of 0.9999999998, close to 1, indicates that the model explains the variation in the Philippines' medal tally almost perfectly. This shows that the Prophet model accurately predicted the Philippines' medal tally. The model's accuracy is demonstrated by the seasonal and trend components of the Prophet model successfully capturing the patterns in the historical data. In addition, this model can be a valuable tool for coaches and policymakers to make plans that can help improve the Philippines' performance in future Olympics.

3.6 Discussion

Tables 2 to 6 show the total medal prediction performance of Southeast Asian countries based on evaluation metrics, including MAE, MSE, RMSE, and R^2. Overall, the Prophet model yields almost perfect prediction results for each country's analysis: Indonesia, Malaysia, Singapore, Thailand, and the Philippines, with very low MAE, MSE, and RMSE values, reaching 0.0 in some countries. The R^2 values for all five countries were also close to 1, indicating that the model could explain almost all of the variation in medal tally. This study shows that the Prophet model can capture patterns in the medal data of Southeast Asian countries. The model is so accurate that it can help coaches and policymakers improve athletes' performance in future Olympics.

Table 7 Comparison of the Table with Previous Research

Authors	Data Sample	Summer/Winter	Method	Result
Scelles et al. (2020)	1992 - 2016	Summer	Tobit and Hurdle Econometric Models	The hurdle model yields more accurate predictions for the 2016 and 2020 Olympics, particularly highlighting the significant impact of socio-economic factors and regional variables on medal outcomes.
Rewilak (2021)	1996 - 2016	Summer	Key and less influential factors using the Tobit and hurdle models.	Population size and host effect are significant determinants of Olympic success.
Jia et al. (2022)	2008 - 2022	Summer	International Public Opinion Analysis Using LDA, TF-IDF, Naive Bayes	The opinions of sports events were more positive in Chinese than in English.
Asha et al. (2023)	2000 - 2020	Summer	Data Analysis for Olympic Performance	The analysis revealed that the United States produced the highest number of athletes for the Olympics, followed by Germany. In contrast, Canada had the lowest number of athletes represented.
Badoni et al. (2023)	2000 - 2020	Summer	Comparative analysis of machine learning algorithms like linear regression and decision trees.	The research compares machine learning algorithms, including linear regression, decision trees, and support vector machines (SVM), to determine which model yields the most accurate predictions for Olympic medal counts. Decision Tree handles categorical and numerical data efficiently
Proposed Method	1994 - 2024	Summer / Winter	Prediction with the Facebook-Prophet Model	The research forecasts Olympic medal trends for Southeast Asian countries. The Facebook Prophet model effectively predicts medal achievements.

Examining the performance of the Prophet model in terms of predicting Olympic medal counts requires a careful comparison with previous sports analytics studies. Many earlier studies have used conventional statistical approaches or machine learning algorithms to project medal results, including linear regression and decision trees. However, these methods sometimes struggled to adequately depict the intricate trends and anomalies in the medal data over time. The Prophet model enhances accuracy by accurately adjusting for seasonality and external variables that affect athletes' performances. Based on its evaluation measures, this study demonstrates how effectively the model can provide more consistent forecasts. Table 7 presents the performance measures of the Prophet model in comparison to those of previous studies, highlighting the higher precision achieved in this work.

4. CONCLUSIONS

The study of Olympic medal counts in Indonesia, Thailand, Malaysia, Singapore, and the Philippines generally exposes distinctive trends in each nation's performance in recent years. The Prophet model captures these trends, as seen by the close alignment between expected and actual medal totals. The evaluation criteria of the Prophet model in all countries demonstrate its effectiveness in capturing historical trends and informing future strategies. Future research on predicting Olympic medal trends could explore various methods to enhance the understanding and accuracy of forecasts. One possible approach is to broaden the dataset to incorporate additional Southeast Asian countries, enabling a more comprehensive analysis of regional trends and patterns in Olympic performance. Additionally, researchers may investigate how socio-economic factors, such as government funding and athlete support systems, influence medal outcomes. This could give a better understanding of what influences success at the Olympics.

One possible direction for future research could be to explore the combination of machine learning techniques with the Prophet model. This could help compare predictive accuracy and determine which methods are most effective for forecasting. Furthermore, studies examining the long-term effects of training programs and policy changes on medal counts could provide valuable insights for countries seeking to enhance their Olympic strategies. By focusing on these areas, future research can help us better understand the dynamics of Olympic sports and improve the predictive models used in this area.

REFERENCES

Agyemang, E. F., Mensah, J. A., Ocran, E., Opoku, E., & Nortey, E. N. N. (2023). Time Series Based Road Traffic Accidents Forecasting via SARIMA and Facebook Prophet Model with Potential Changepoints. *Heliyon*, *9*(12), e22544. https://doi.org/10.1016/j.heliyon.2023.e22544

Angelo, M. D., Fadhiilrahman, I., & Purnama, Y. (2023). Comparative Analysis of ARIMA and Prophet Algorithms in Bitcoin Price Forecasting. *Procedia Computer Science*, *227*, 490–499. https://doi.org/10.1016/j.procs.2023.10.550

Annapoorna, E., Sujil, S. V, S, S., Abhishek, S., & T, A. (2024). Revolutionizing Stock Price Prediction with Automated Facebook Prophet Analysis. *2024 International Conference on Inventive Computation Technologies (ICICT)*, 1307–1314. https://doi.org/10.1109/ICICT60155.2024.10544766

Asha, V., Sreeja, S. P., Saju, B., C S, N., N, P. G., & Prasad, A. (2023). Performance Analysis of Olympic Games Using Data Analytics. *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, 1436–1443. https://doi.org/10.1109/ICEARS56392.2023.10084943

Badoni, P., Choudhary, P., Rudesh, C. P., & Singh, N. T. (2023). Predicting Medal Counts in Olympics Using Machine Learning Algorithms: A Comparative Analysis. *2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech)*, 116–121. https://doi.org/10.1109/ICACCTech61146.2023.00027

Cheng, J., Tiwari, S., Khaled, D., Mahendru, M., & Shahzad, U. (2024). Forecasting Bitcoin Prices Using Artificial Intelligence: Combination of ML, SARIMA, and Facebook Prophet Models. *Technological Forecasting and Social Change*, *198*, 122938. https://doi.org/10.1016/j.techfore.2023.122938

Chowdary, P. H., Kaur, V., Nandeesh, T., Krishan, K., & Kaur, A. (2024). From Athens to Rio: A Comprehensive Data Analysis and Visualization of 120 Years of Olympic History. *2024 11th*

International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 1–6. https://doi.org/10.1109/ICRITO61523.2024.10522373

Ding, S., Ye, J., & Cai, Z. (2024). Multi-Step Carbon Emissions Forecasting Using an Interpretable Framework of New Data Preprocessing Techniques and Improved Grey Multivariable Convolution Model. *Technological Forecasting and Social Change*, *208*, 123720. https://doi.org/10.1016/j.techfore.2024.123720

Dong, X., Guo, W., Zhou, C., Luo, Y., Tian, Z., Zhang, L., Wu, X., & Liu, B. (2024). Hybrid Model for Robust and Accurate Forecasting Building Electricity Demand Combining Physical and Data-Driven Methods. *Energy*, *311*, 133309. https://doi.org/10.1016/j.energy.2024.133309

Gautam, V., Yadav, V., & Kumar, S. (2023). Diagnosis and Forecast of Murder Rates in India Using Random Forest and Prophet Algorithm. *2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, 173–177. https://doi.org/10.1109/CICTN57981.2023.10141293

Gong, F., Han, N., Li, D., & Tian, S. (2020). Trend Analysis of Building Power Consumption Based on Prophet Algorithm. *2020 Asia Energy and Electrical Engineering Symposium (AEEES)*, 1002–1006. https://doi.org/10.1109/AEEES48850.2020.9121548

Guo, L., Fang, W., Zhao, Q., & Wang, X. (2021). The Hybrid PROPHET-SVR Approach for Forecasting Product Time Series Demand with Seasonality. *Computers & Industrial Engineering*, *161*, 107598. https://doi.org/10.1016/j.cie.2021.107598

Herzog, W. (2024). The Paris 2024 Olympic and Paralympic Games. *Journal of Sport and Health Science*, *13*(6), 717–718. https://doi.org/10.1016/j.jshs.2024.06.003

Ismail, Y. (2024). *Olympic Games (1994-2024)*. https://www.kaggle.com/datasets/youssefismail20/olympic-games-1994-2024

James, M. (2023). Human Rights and the Olympic Charter. *The International Sports Law Journal*, *23*(3), 267–270. https://doi.org/10.1007/s40318-023-00254-5

Jia, K., Zhu, Y., Zhang, Y., Liu, F., & Qi, J. (2022). International Public Opinion Analysis of Four Olympic Games: From 2008 to 2022. *Journal of Safety Science and Resilience*, *3*(3), 252–262. https://doi.org/10.1016/j.jnlssr.2022.03.002

Karunasingha, D. S. K. (2022). Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well. *Information Sciences*, *585*, 609–629. https://doi.org/10.1016/j.ins.2021.11.036

Lei, Y., Lin, A., & Cao, J. (2024). Rhythms of Victory: Predicting Professional Tennis Matches Using Machine Learning. *IEEE Access*, *12*, 113608–113617. https://doi.org/10.1109/ACCESS.2024.3444031

Li, Y., Liu, J., Ang, S., & Yang, F. (2021). Performance Evaluation of Two-Stage Network Structures with Fixed-Sum Outputs: An Application to the 2018 Winter Olympic Games. *Omega*, *102*, 102342. https://doi.org/10.1016/j.omega.2020.102342

Mahajan, A. S., & Shrivastav, A. (2023). Short Term Load Forecasting Based on Regression Models. *2023 International Conference for Advancement in Technology (ICONAT)*, 1–8. https://doi.org/10.1109/ICONAT57137.2023.10080359

Mousa, M. A., AlMansoori, A. N., & AlAjami, F. A. (2023). A Hybrid PV Power Forecasting Model Implementing Emerging Machine Learning Algorithms: Prophet and Neural Prophet. *2023 Middle East and North Africa Solar Conference (MENA-SC)*, 1–8. https://doi.org/10.1109/MENA-SC54044.2023.10374512

Phan, Q.-T., Wu, Y.-K., & Phan, Q.-D. (2021). An Overview of Data Preprocessing for Short-Term Wind Power Forecasting. *2021 7th International Conference on Applied System Innovation (ICASI)*, 121–125. https://doi.org/10.1109/ICASI52993.2021.9568453

Qi, J., Du, J., Siniscalchi, S. M., Ma, X., & Lee, C.-H. (2020). On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. *IEEE Signal Processing Letters*, *27*, 1485–1489. https://doi.org/10.1109/LSP.2020.3016837

Rajesh, K., & Saravanan, M. S. (2022). Prediction of Customer Spending Score for the Shopping Mall Using Gaussian Mixture Model Comparing with Linear Spline Regression Algorithm to Reduce Root Mean Square Error. *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 335–341. https://doi.org/10.1109/ICICCS53718.2022.9788162

Rewilak, J. (2021). The (Non) Determinants of Olympic Success. *Journal of Sports Economics*, *22*(5), 546–570. https://doi.org/10.1177/1527002521992833

Sagala, N. T. M., & Ibrahim, M. A. (2022). A Comparative Study of Different Boosting Algorithms for Predicting Olympic Medal. *2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED)*, 1–4. https://doi.org/10.1109/ICCED56140.2022.10010351

Arteaga, F. J. S., DiCaprio, D., Tavana, M., Cucchiari, D., Campistol, J. M., Oppenheimer, F., Diekmann, F., & Revuelta, I. (2024). On the Capacity of Artificial Intelligence Techniques and Statistical Methods to Deal with Low-Quality Data in Medical Supply Chain Environments. *Engineering Applications of Artificial Intelligence*, *133*, 108610. https://doi.org/10.1016/j.engappai.2024.108610

Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time Series Analysis and Forecasting of Coronavirus Disease in Indonesia Using ARIMA Model and PROPHET. *Procedia Computer Science*, *179*, 524–532. https://doi.org/10.1016/j.procs.2021.01.036

Scelles, N., Andreff, W., Bonnal, L., Andreff, M., & Favard, P. (2020). Forecasting National Medal Totals at the Summer Olympic Games Reconsidered. *Social Science Quarterly*, *101*(2), 697–711. https://doi.org/10.1111/ssqu.12782

Tawakuli, A., Havers, B., Gulisano, V., Kaiser, D., & Engel, T. (2024). Survey: Time-Series Data Preprocessing: A Survey and an Empirical Analysis. *Journal of Engineering Research*. https://doi.org/10.1016/j.jer.2024.02.018

Theodorakis, Y., Georgiadis, K., & Hassandra, M. (2024). Evolution of the Olympic Movement: Adapting to Contemporary Global Challenges. *Social Sciences*, *13*(7), 326. https://doi.org/10.3390/socsci13070326

Verghese, A., T, S., & S, V. (2021). Analysis and Forecasting Covid-19 Spread in India Using Logistic Regression and Prophet Time Series. *2021 International Conference on Computational Performance Evaluation (ComPE)*, 928–932. https://doi.org/10.1109/ComPE53109.2021.9752218

Wen, J., & Wang, X. (2020). Study of the Visualization and Interaction of Data: Take the Historical Data of the Winter Olympics as an Example. *2020 International Conference on Innovation Design and Digital Technology (ICIDDT)*, 78–82. https://doi.org/10.1109/ICIDDT52279.2020.00022

Wu, P., Zhu, X., Yang, S., & Huang, J. (2023). The Influence of the Beijing Winter Olympic Games on the Demand for Winter Sports: An Empirical Analysis Based on the Baidu Index. *Heliyon*, *9*(10), e20426. https://doi.org/10.1016/j.heliyon.2023.e20426

Wulandari, R., Surarso, B., Irawanto, B., & Farikhin, F. (2021). The Forecasting of Palm Oil Based on Fuzzy Time Series-Two Factor. *Journal of Soft Computing Exploration*, *2*(1), 11–16. https://shmpublisher.com/index.php/joscex/article/view/14

Xinyi, S., & Chenglong, X. (2022). Visual Analysis of the Distribution Characteristics and Influencing Factors of Olympic Medals. *2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 1633–1637. https://doi.org/10.1109/IMCEC55388.2022.10019867

Yin, K., Guo, H., & Yang, W. (2024). A Novel Real-Time Multi-Step Forecasting System with a Three-Stage Data Preprocessing Strategy for Containerized Freight Market. *Expert Systems with Applications*, *246*, 123141. https://doi.org/10.1016/j.eswa.2024.123141