ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 6 No. 1 (2026) 426-436
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Time Series Modelling of Daily Blood Donations in Malaysia

Nur Sakinah Azlin[1], Maria Elena Nor[1*]

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: maria@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.049

Article Info

Received: 1 January 2026
Accepted: 20 January 2026
Available online: 9 July 2026

Keywords

Blood Donation, Time Series Forecasting, SARIMA, Decomposition, Time Series Regression, MAE, MAPE, RMSE

Abstract

Blood is a vital medical resource, yet maintaining a stable blood supply remains challenging due to fluctuating donation patterns, seasonal effects, and the short shelf life of blood products. These challenges were further highlighted during the COVID-19 pandemic, emphasizing the need for accurate forecasting to support effective blood supply chain management. This study applies traditional time series forecasting techniques to analyse daily blood donation patterns in Malaysia using data from January 2019 to February 2025. The study aims to apply the decomposition method to identify trends, seasonal patterns, and irregular fluctuations in daily blood donation data, and to compare the forecast performance of the Additive Decomposition method, Seasonal Autoregressive Integrated Moving Average (SARIMA), and Time Series Regression. Forecast accuracy is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The results reveal clear weekly seasonality and moderate variability in donation patterns, with an overall increasing trend over time. Among the methods examined, Time Series Regression provides the most accurate forecasts with the lowest MAPE value of 36.69989092. The findings highlight the importance of reliable daily forecasting in supporting effective blood inventory planning and reducing shortages and wastage in Malaysia.

1. Introduction

Blood donation plays a vital role in supporting healthcare services, including surgeries, trauma care, maternal health, cancer treatment, and chronic disease management [1]. Blood must be readily available as an essential public health resource to meet unpredictable clinical demands. However, daily blood donation levels often fluctuate due to behavioural, seasonal, institutional, and environmental factors such as work schedules, holiday periods, community donation campaigns, and unexpected public health events [2]. These fluctuations, combined with the short shelf life of blood components, particularly platelets and red blood cells, pose significant challenges for hospitals and blood banks. Consequently, accurate forecasting is crucial to predict demand, reduce wastage, prevent shortages, and improve overall blood supply chain planning and management. Blood products have a limited shelf life, which necessitates short-term and high-frequency forecasting. Red blood cells can be stored for approximately 35–42 days, platelets for only 5–7 days, and plasma for up to one year if frozen. Due to the extreme perishability of platelets and the continuous daily demand from hospitals, forecasting at a daily level is more appropriate than monthly aggregation, as it enables timely inventory replenishment, minimizes wastage, and reduces the risk of shortages.[3].

Although numerous studies have applied time series and machine learning methods to forecast blood donations and demand in countries such as China, Canada, and Zimbabwe, most existing studies focus on monthly or weekly data [14]. Furthermore, empirical studies that specifically model daily blood donation patterns in Malaysia remain limited. Existing Malaysian studies largely emphasize descriptive statistics or aggregated trends rather than high-frequency daily forecasting. This highlights a clear gap in the literature concerning short-term, daily forecasting approaches tailored to the Malaysian context.

To address these challenges and research gaps, this study analyses daily blood donation patterns in Malaysia using three time series forecasting techniques: decomposition, Seasonal Autoregressive Integrated Moving Average (SARIMA), and time series regression. The primary objective of this study is to apply the decomposition method to identify underlying trends, seasonal patterns, and irregular fluctuations in daily blood donation data over a specified time period. In addition, the study aims to compare the forecast performance of the decomposition method, SARIMA, and time series regression in order to identify the most reliable forecasting technique for practical application. These methods provide valuable insights into donor behaviour by capturing weekly seasonal cycles and structural variations in the data [4,5]. The findings are expected to assist hospitals, blood banks, and policymakers in planning donation campaigns, managing blood inventory more efficiently, and ensuring a stable and responsive national blood supply system.

2. Literature Review

2.1 Blood Supply Chain and Forecasting Approach

The Blood Supply Chain (BSC) consists of interconnected activities including donor recruitment, blood collection, processing, testing, storage, distribution, and clinical use. The effective management of this supply chain is challenged by product perishability, supply instability, demand uncertainty, and strict safety regulations [6,7]. Donor turnout is influenced by multiple factors such as working days, seasonal cycles, sociodemographic characteristics, and unpredictable public health events, all of which contribute to significant variability in blood availability. Forecasting therefore plays a critical role in minimizing wastage and preventing shortages, particularly given the short shelf life of blood products, especially platelets, which can only be stored for five to seven days [1,6]. Poor forecasting and inventory planning can result in overstock or shortages, both of which negatively affect healthcare operations and patient outcomes.

To address these challenges, a variety of forecasting approaches have been applied in blood supply management. Recent studies have explored advanced machine learning techniques such as Long Short-Term Memory (LSTM), Temporal Convolutional Network (TCN), DeepAR, and Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) models, which are capable of capturing complex patterns and nonlinear relationships in blood demand data [14,15,16]. Despite their predictive power, these methods often require large datasets, substantial computational resources, and specialized expertise, which may limit their practical adoption in healthcare settings. Consequently, traditional forecasting techniques such as regression, exponential smoothing, and Autoregressive Integrated Moving Average (ARIMA) models remain widely used due to their simplicity, interpretability, and lower data requirements. Empirical evidence shows that ARIMA-based models perform reliably for short-term blood demand forecasting, particularly when strong seasonal patterns are present in historical data [17,18].

2.1.1 Decomposition Methods in Blood Donation Forecasting

Through organizing a time series into trend, seasonal, and irregular components, decomposition techniques make it easy to see the underlying patterns in the data [1,8]. Studies demonstrate the effectiveness of decomposition-based techniques, such as STL and classical decomposition, in detecting long-term behavior and seasonal donation patterns. Planning donor campaigns and foreseeing seasonal shortages are two operational decisions that are supported by these insights. Decomposition techniques are useful for strategic planning and interpretability even though they might not always produce the best predictive accuracy when compared to more advanced models.

2.1.2 Application of SARIMA in Blood Donation Forecasting

Strong seasonal patterns in forecasting datasets are a usual application for SARIMA models. SARIMA can accurately model weekly and monthly fluctuations in blood donations and clinical demand, according to studies done across several of different countries [9,10]. SARIMA is a useful tool for healthcare forecasting since it combines both autoregressive and moving-average behavior at seasonal and non-seasonal levels.

Previous research demonstrates that when seasonality is strong and consistent, SARIMA frequently performs better than simpler models, enhancing resource allocation and lowering blood bank shortages [11,12,13].

3. Research Methodology

3.1 Data Description

The Department of Statistics Malaysia (DOSM) provided daily blood donation data https://data.gov.my/data-catalogue/blood_donations_state, which involves the time duration from December 30, 2019 to February 17, 2025 [19]. There were 1,877 observations. For the purpose of evaluating the accuracy of out-of-sample forecasting, the dataset has been divided into 1,856 training observations and 21 testing observations. The information shows daily donation activity across the country, which is impacted by weekdays, weekends, and seasonal events.

3.2 Decomposition Methods

Decomposition is a method of forecasting that separates advanced structure or processes to the simpler part [1,7,8]. In the domain, it particularly involves dividing various types of factors that contribute to changing the behaviour of the dataset. By dividing a time series into few key components to better understand in the patterns and the behaviours of the data can improve the accuracy of the forecast. Time series contains four components which is trend, seasonal, cyclic and the residual. Additive Decomposition is applied when the magnitude of seasonal and residual fluctuations does not change into the trend level of the data. The formula for additive decomposition is shown as the Equation (1):

$$Y(t)=T(t)+S(t)+R(t) \tag{1}$$

where,

Trend $T(t)$*:* underlying pattern.

Seasonality $S(t)$*:* seasonal variations are proportional of the trend level.

Residual $R(t)$*:* random or irregular fluctuations.

3.3 Seasonal Autoregressive Integrated Moving Averages

Seasonal Autoregressive Integrated Moving Average (SARIMA) is the extension of ARIMA model by incorporating seasonal components to suit the time series data that express both trends and seasonality [10]. The general equation of SARIMA as Equation (2):

$$\phi_p\left(B^s\right)\phi_p\left(B^s\right)\nabla^d\nabla_s^D y_t = \vartheta_Q(B)\theta_Q\left(B^S\right)a_t \tag{2}$$

where,

$\boldsymbol{y_t}$: observation value at time *t*.

$\boldsymbol{a_t}$: error term at time *t* for assuming to be white noise.

$\boldsymbol{B}$: Backshift or lag, such $\boldsymbol{B^j y_t = y_{t-j}}$.

$\boldsymbol{\phi_p}$= non-seasonal AR polynomial of order *p*.

$\boldsymbol{\Phi_p}$=seasonal AR polynomial of order *p* with seasonality *S*.
$\boldsymbol{\nabla_s^D}$=seasonal differencing operator of order *D* with seasonal period *S*.

$\boldsymbol{\theta_Q(B)}$=non-seasonal MA polynomial of order *Q*.

$\boldsymbol{\Theta_Q(B^S)}$= seasonal MA polynomial of order *Q* with seasonality *S*.

3.3.1 Stationary Means and Variance Box-Cox Transformation

The data transformation in statistical work, in order to stabilize the variance of the data [11]. Box-Cox is a transformation of time series variable which is $\boldsymbol{yt}, \boldsymbol{t} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{n}$ that rely on the power parameter $\boldsymbol{\lambda}$ as shown in Equation (3):

$$w_t = \begin{cases} \log y_t, if\ \lambda = 0; \\ \frac{\left(y_t^{\lambda} - 1\right)}{\lambda} otherwise, \end{cases} \tag{3}$$

Based on the maximum likelihood estimation, the Box–Cox transformation parameter was estimated as λ= 1. Therefore, no transformation was applied, and the analysis was conducted using the original data scale.

where,

$\boldsymbol{w_t}$: observation at time t after applying the transformation.

$\boldsymbol{y_t}$: observation at time t.

$\boldsymbol{\lambda}$: estimated box-cox value.

$\boldsymbol{y_t^{\lambda}}$: observation at time t with power of $\boldsymbol{\lambda}$.

Table 1 shows the Box-Cox transformation that applies various power transformations depending on the value of lambda (λ) to stabilize variance and make data more normally distributed.

Table 1 Box-Cox Transformation [12]

λ	**Transformed Data**
-1	Inverse transformation
-0.5	Reciprocal square root transformation
0	Log transformation
0.5	Square root transformation
1	No transformation

3.3.2 Differencing

Differencing help in maintaining a time series mean. A time series is considered stationary if it is independent of the time at which it is observed. Time series that exhibit trend, seasonality, or both are referred to as non-stationary [7]. Differencing is the process of computing the differences between successive observations of a time series. Trend and seasonality are eliminated or minimized as a result of removing variations in a time series level. A stationary time series plots with a constant variance and an approximately horizontal appearance. As shown in Equation (4):

$$y_t^{'} = y_t - y_{t-1} = y_t - By_t = \left(1 - B\right) y \tag{4}$$

where,

$y_t^{'}$ = difference observation at time t

y_t = observation at time t

y_{t-1} = observation at time t-1

3.3.3 Autocorrelation Function and Partial Autocorrelation Function

The correlation between observations of a time series, *yt* and *yt-k*, separated by *k* time units is measured by the Autocorrelation Function (ACF), where *k* = 1, 2, 3…, and *yt* is the observation at time *t* and *yt-k* is the observation at time *t-k* [6]. ACF plots are a type of plot used to show ACF. After adjusting for the effects of various time lags, Partial Autocorrelation Function (PACF) is a measure of the correlation between observations of a time series, *yt* and *yt-k*, that are separated by *k* time units. A plot of PACF that is used to visualize PACF is known as a PACF plot. ACF plots are helpful in identifying nonstationary time series besides to time series plots. When a time series is stationary, the ACF will drop to zero on an ACF plot, however when a time series is nonstationary, the ACF will fall continuously. For a nonstationary time series, the autocorrelation value at lag 1 is often high and positive apart from that. ACF and PACF plots can also be used to determine whether the residuals are uncorrelated.

Table 2 lists the common patterns found in the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). The ACF decays gradually in an AR(*p)* model, while the PACF exhibits a notable peak at lag *p* and then stops. On the other hand, an MA(*q)* model exhibits an exponential decay in PACF and a notable spike in ACF at lag *q*. Both ACF and PACF exhibit exponential or decreased decay for ARIMA*(p,d,q)* or ARMA*(p,q)* models, suggesting a combination of autoregressive and moving average components. Under specific data patterns, the "AR(*p),* MA(*q)*" suggests that both ACF and PACF exhibit notable spikes at lag *q.* This could be a misunderstanding or a combined behaviour, but generally refers to mixed ARMA behaviour. These patterns help in identifying the suitable model based on the autocorrelation structure of the data.

Table 2 Behaviour of ACF and PACF [6]

Model	ACF	PACF
AR*(p)*	Exponentially decaying	There is significant spike at lag *p* but none beyond lag *q.*
MA *(q)*	There is significant spike at laq p but none beyond lag q.	Exponentially decaying
ARIMA*(p,d,q)* /ARMA *(p,q)*	Exponentially decaying	Exponentially decaying
ARIMA(p,d,q) *AR(p),MA(q)*	There is significant spike at lag p but none beyond lag q.	There is significant spike at lag p but none beyond lag q

The identification of the SARIMA model parameters $(\boldsymbol{p},\boldsymbol{d},\boldsymbol{q})\ (\boldsymbol{P},\boldsymbol{D},\boldsymbol{Q})_7$ followed the Box–Jenkins methodology. The orders of differencing *(d, D)* were determined based on visual inspection of the time series plot and stationarity considerations, while the autoregressive and moving average orders *(p, q, P, Q)* were identified using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. Candidate models were then compared using diagnostic checks and Mean Squared Error (MSE), with the final model selected based on residual independence and forecasting accuracy.

3.4 Time Series Regression

Time Series Regression is a method that is use to a time dependent model. It may include lagged predictors, seasonal patterns and intervention effects. Using time series regression, from the general regression context, time series analyst handles with stochastic relationship. Moreover, time series regression model and forecast the behaviour of dependent variable over time using time related to independent variables [15]. The general equation of time series regression model follows as Equation (5):

$$Y_t = \beta_\circ + \beta_1 t + \beta_{t+1} Day_1 \tag{5}$$

As for the seasonal time series regression or additive model is follow as Equation (6):

$$Y_t = \beta_\circ + \beta_1 Day_1 + \beta_2 Day_2 + \beta_3 Day_3 + \cdots + \beta_6 Day_6 + \varepsilon_t \tag{6}$$

where,

Yt: Observed value of the dependent variable at time t
$\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \ldots \boldsymbol{\beta_6 Day_6}$: Coefficients representing the effect of each dummy variable $\boldsymbol{Day_6}$ to $\boldsymbol{Day_6}$

$\boldsymbol{Day_1}, \boldsymbol{Day_2}, \ldots, \boldsymbol{Day_6}$: Dummy variables

$\boldsymbol{\varepsilon_t}$: error term

This research has a daily data, L= 6, as for the notation for the dummy variables by using $\boldsymbol{Day_1}$for Monday,

$\boldsymbol{Day_2}$ for Tuesday and continue until Saturday for $\boldsymbol{Day_6}$.

As for the multiplicative model is from the Equation (7) :

$$Y_t = \beta_\circ + \beta_1 Day_1 + \beta_2 Day_2 + \beta_3 Day_3 + \cdots + \beta_7 Day_7 + \beta_\circ + \beta_{11} t * Day_1 + \beta_{12} t * Day_2 + \beta_{13} t * Day_3 + \cdots \beta_{16} t * Day_6 + \varepsilon_t \quad (7)$$

Six dummy variables were constructed to represent the day-of-week effects, where Day 1 = Monday, Day 2 = Tuesday, Day 3 = Wednesday, Day 4 = Thursday, Day 5 = Friday, and Day 6 = Saturday. Sunday (Day 7) was excluded from the model and therefore serves as the baseline reference day. Each dummy variable takes the value of 1 if the observation corresponds to that specific day and 0 otherwise. The estimated coefficients thus measure deviations in daily blood donations relative to Sunday.

3.4.1 Forecast Accuracy Measures

The accuracy of the forecast is evaluated using the following metrics: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

MAE is a metric that calculates the average magnitude of the absolute errors between the predicted and actual values. MAE characterizes the alteration among the original and predictable values and is mined as the dataset's total alteration means [22]. MAE formula from Equation (8):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_t - \hat{Y}_t\right| \quad (8)$$

MAPE measures the average magnitude of error produced by a model, or how far off predictions are on average [23]. MAPE is the percentage equivalent of MAE. MAPE formula from Equation (9):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{Y_t} \times 100 \quad (9)$$

RMSE measures the average difference between a statistical model's predicted values and the actual values [24]. Mathematically, it is the standard deviation of the residuals. RMSE formula as in Equation (10):

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{\left(Y_t - \hat{Y}_t\right)^2}{n}} \quad (10)$$

where,

n: The total number of observations.

$\boldsymbol{Y_t}$: The actual (observed) value at time *t*.

$\hat{Y}_t$: The predicted (forecasted) value at time *t*.

$|\boldsymbol{Y_t} - \hat{Y}_t|$: The absolute error, which is the absolute difference between the observed and predicted values.

4. Result and Discussion

This section presents the results of the testing data for the Additive Decomposition, Seasonal Autoregressive Moving Averages (SARIMA), and Time Series Regressions models, evaluated using accuracy metrics such as MAE, MAPE, and RMSE.

4.1 Time Series Plot

The dataset consists of 1877 observations, with 1856 for training and the remaining 21 for testing. This allocation ensures a comprehensive assessment of the model performance on a large test set while retaining enough data points for training to identify the underlying patterns and seasonality in blood donation trends. Based on Fig. 1, the time series plot of daily blood donations from 30 December 2019 17 February 2025 reveals a seasonal pattern with recurring weekly variations higher donations during weekends and lower on weekdays, and a stable overall trend without notable long-term changes.

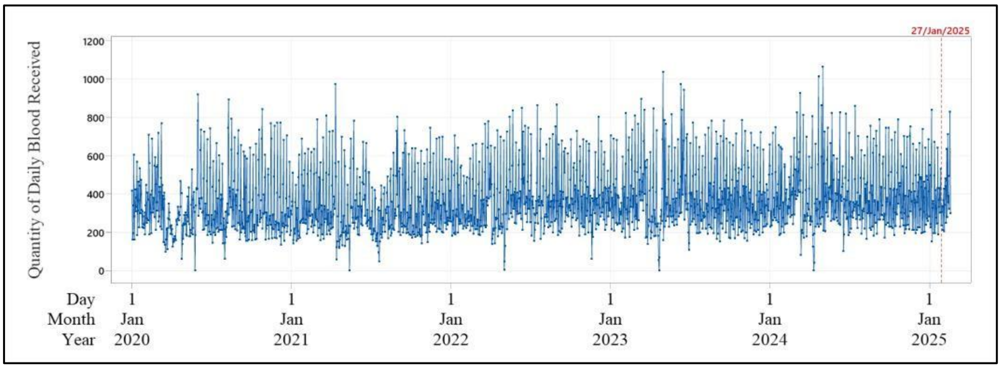

Fig. 1 Time Series Plot

4.2 Additive Decomposition

Fig 2 shows the trend component, the fitted values, and the actual daily blood received. The pattern shows different seasonal variations, which are probably linked to scheduled collection events or shifts in donor involvement. The trend line exhibits a slight upward movement despite these short-term fluctuations, indicating an overall increase in the average amount of blood obtained over time.

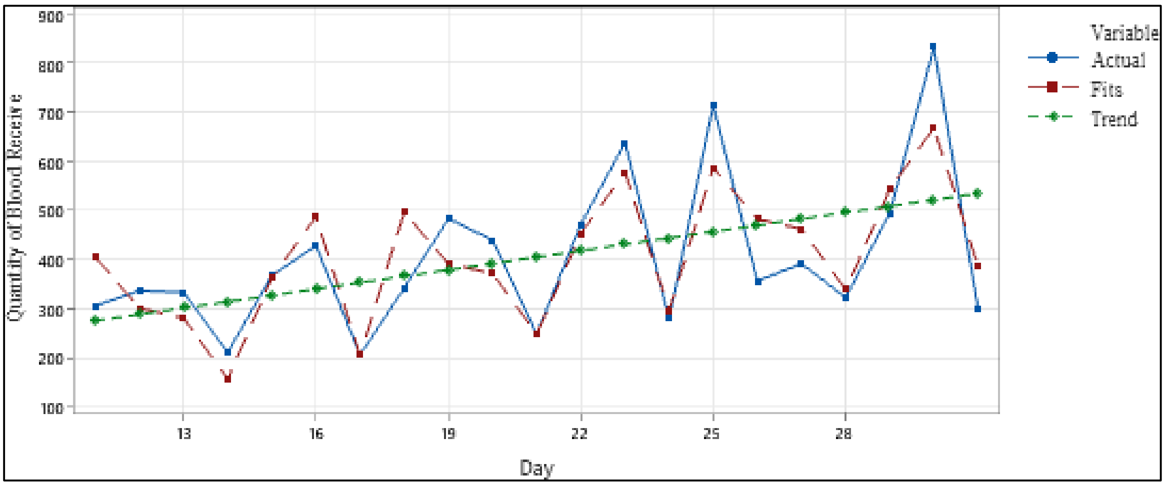

Fig. 2 Additive Decomposition plot for Daily Blood Donation in Malaysia

5. Seasonal Autoregressive Moving Averages

Based on Table 3, given the notable exception of a few cases, such as AR (1) in the first model p = 0.154 and SMA (7) in the second model p = 0.103, all of the candidate SARIMA models exhibit acceptable Ljung-Box p-values at lag 12, indicating no significant autocorrelation remaining in the residuals. The Mean Squared Error (MSE) is used to identify the optimal model; smaller values indicate better predictive performance. The $(0,1,1)(0,1,0)_7$ model has the lowest MSE of 10,433.2 of all the models, meaning that it fits the data the best and is therefore the model of choice for forecasting.

Table 3 Results for SARIMA tentative models

Model	*p*-value of estimated parameter	*p*-value of Ljung Box at lag 12	MSE
$(1,1,1)(1,1,1)_7$	AR 1 (0.154)	0.188,0.561,0.194	10581.8
	SAR 7 (0.000)		
	MA 1 (0.000)		
	SMA 7 (0.003)		
$(0,1,1)(0,1,0)_7$	MA 1 (0.00)	0.514,0.802,0.464	10433.2
	SMA 7 (0.103)		
$(1,1,1)(1,1,1)_7$	AR 1 (0.00)	0.118,0.312,0.061	12036.7
	SAR 7 (0.00)		

	SMA 7 (0.001)		
(1,1,0)(1,1,0)$_7$	AR 1 (0.00)	0.625,0.595,0.099	12647.6
	SAR (0.803)		
(0,1,1)(0,1,1)$_7$	MA 1 (0.00)	0.608,0.959,0.686	10949.5

In the SARIMA (0,1,1)(0,1,0)$_7$ model, the seasonal moving average term SMA (7) is statistically non-significant ($p = 0.103$). However, statistical significance of individual parameters is not the sole criterion for model selection in time series forecasting. After applying seasonal differencing, much of the weekly seasonal structure is already removed, resulting in a weaker residual seasonal effect at lag 7. Despite the non-significant SMA (7) term, the model demonstrates the lowest MAPE among the competing SARIMA specifications, indicating superior out-of-sample forecasting accuracy. In addition, residual diagnostics confirm that the model adequately captures the dependence structure of the data, with no remaining autocorrelation. Therefore, the model is retained based on its empirical forecasting performance and diagnostic adequacy rather than parameter significance alone.

5.1 Time Series Regression

For the objective to capture additive seasonality, daily data with 6 dummy variables (Monday to Saturday) and interaction terms between time t and each daily dummy were used to estimate the time series regression model. This enables the model to take into consideration both day-specific trend components and daily seasonal effects. As in Equation (11):

$$Y_t = 280.4 + 12.44t - 155.9Day1 - 191.9Day2 + 5.4Day3 + 189.9Day4 - 16.1Day5 + 74.4Day6 \tag{11}$$

Table 4 Results for Time Series Regression coefficients

Term	Coefficient	SE Coef	*p*-Value	VIF
t	0.04671	0.00441	0.00	1.00
Day 1	-402.31	8.83	-45.58	1.71
Day 2	-345.20	8.83	-39.07	1.71
Day 3	-299.13	8.83	-33.89	1.71
Day 4	-308.16	8.83	-34.91	1.71
Day 5	-402.61	8.83	-45.61	1.71
Day 6	-182.92	8.83	-20.72	1.71

Time series regression results on Table 4 show that both the trend component and the daily seasonal effects are important in explaining the series' behaviour. A consistent upward trend over time is indicated by the time variable t positive and highly significant coefficient (0.04671, $p < 0.05$). The dependent variable fluctuates significantly throughout the week in comparison to the reference day (Day 7), as evidenced by the statistical significance of all six dummy variables (Day 1 to Day 6). With Day 1 and Day 5 exhibiting the biggest drops of about 402 units, the negative coefficients for Days 1 through 6 suggest that these days continuously record lower values in comparison to the baseline.

Fig 3 presents the comparison between the actual daily blood donation amounts and the predicted values by the time series regression model. Overall, the forecasted values follow the general pattern of the observed data, indicating that the model is able to reflect the underlying trend in daily blood donations. Although there are several days where the predicted values differ from the actual amounts, particularly during sharp increases or decreases in donations, these differences are expected due to daily fluctuations in donor turnout. Despite this variability, the model provides a reasonable approximation of daily blood donation levels, suggesting that time series regression is appropriate for analysing and forecasting blood donation data over time.

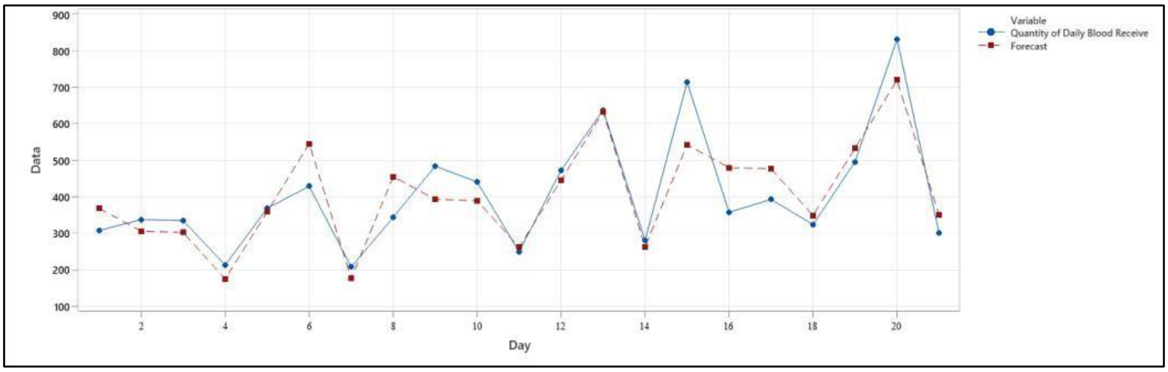

Fig. 3 Forecast Value for Time Series Regression

6. Model Evaluation

According to Table 5, Time Series Regression shows the highest level of accuracy with the lowest MAE, RMSE, and MAPE values among the three methods for the testing period, which is more significant in evaluating forecasting ability. While Additive Decomposition records the highest error values, indicating the worst out-of-sample performance, SARIMA performs moderately but still generates higher errors than Time Series Regression. Therefore, it can be concluded that Time Series Regression is the most suitable forecasting technique among the three for future forecasts and has the highest forecast accuracy in producing reliable predictions for this study.

Table 5 Comparison of model accuracy for all models

Method	MAE	MAPE	RMSE
Additive Decomposition	270.9877	78.43793	282.5248
SARIMA	263.5538	72.54102	290.9872
Time Series Regression	164.499825	36.69989092	173.4488958

The Time Series Regression model records a MAPE value of 36.70% on the testing dataset. This level of error reflects the inherent volatility of daily blood donation data, which is strongly influenced by public holidays, festive periods, and irregular events such as emergency donation campaigns. On days with unusually low donation counts, percentage-based accuracy measures such as MAPE tend to increase even when absolute forecast errors remain moderate. Therefore, the observed MAPE value does not indicate poor model performance but rather highlights the irregular and highly variable nature of daily blood donation behaviour.

7. Conclusion

In this study, the daily blood donation series in Malaysia was modelled using Additive Decomposition, Seasonal Autoregressive Integrated Moving Average (SARIMA), and Time Series Regression techniques. According to the analysis, there are distinct time series components in the blood donation data, such as a steady trend, significant weekly seasonality, and random variation. While a few observations show abnormally low donation levels likely as a result of public holidays or brief disruptions exploratory analysis shows that the series' general structure stays constant over the course of the study. According to residual plots and statistical tests, diagnostic checking of the chosen models shows well-behaved residuals with no notable problems with mean, variance, normality, or autocorrelation.

Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) were used to assess forecast accuracy for both training and testing datasets. In comparison to SARIMA and Additive Decomposition models, the Time Series Regression model consistently yields the lowest forecast errors, according to the results. This suggests that the dominant weekly seasonal pattern present in daily blood donation behaviour can be effectively captured by clearly modelling day-of-week effects. As a result, the chosen model can be used to produce accurate daily blood donation forecasts for upcoming planning periods in Malaysia.

As accurate forecasting is essential for effective blood supply management, this study provides valuable insights that may assist hospitals, blood banks, and policymakers in improving resource allocation and ensuring a stable and sufficient blood supply. Furthermore, the modelling framework adopted in this study may be extended to other regions or adapted to different temporal conditions, making it useful for future research in healthcare demand forecasting.

The National Blood Centre can integrate the proposed time series regression model into weekly operational planning by using the estimated regression equation to generate daily donation forecasts for the upcoming week. By inputting the time index and corresponding day-of-week dummy variables, planners can anticipate expected donation levels for each day. These forecasts can be used to schedule mobile donation drives, adjust staffing levels, plan inventory allocation, and proactively organize donor recruitment campaigns during predicted low-donation days, thereby improving supply stability and reducing wastage.

Acknowledgement

This author would like to thank to the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for the support.

Conflict of Interest

The author declares that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design****: Nur Sakinah Azlin, Maria Elena Nor;* ***data collection:*** *Nur Sakinah Azlin; analysis and interpretation of results: Nur Sakinah Azlin, Maria Elena Nor;* ***draft manuscript preparation:*** *Nur Sakinah Azlin. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Chaimae, M., & Adil, B. (2020). Blood collection supply chain management: A critical review and future perspective. *Proceedings of the 2020 IEEE 6th International Conference on Optimization and Applications (ICOA)*, 1–7.

[2] Ding, Y., Jiang, Y., Yu, H., & Xu, Y. (2020). Forecast and evaluation of blood donation using time series models: A case study in Sichuan, China. *PLoS ONE*, 15(9), e0238082. https://doi.org/10.1371/journal.pone.0238082

[3] Folléa, G., & Council of Europe Working Group and ISBT Working Party on Blood Supply Management. (2013). Blood supply management (RBC): Definitions, description as a process, tools for assessment and improvement. *ISBT Science Series*, 8(1), 37–40.

[4] Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting* (2nd ed.). John Wiley & Sons.

[5] Kumar, D., Malhotra, M., & Tiwari, S. (2019). Forecasting daily blood donations using time series analysis. *Asian Journal of Transfusion Science*, 13(2), 102–106. https://doi.org/10.4103/ajts.AJTS_165_18

[6] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. https://otexts.com/fpp2/

[7] Li, N., Chiang, F., Down, D. G., & Heddle, N. M. (2021). A decision integration strategy for short-term demand forecasting and ordering for red blood cell components. *Operations Research for Health Care*, 29, 100290.

[8] Das, P., & Barman, S. (2025). An overview of time series decomposition and its applications. In *Applied and Theoretical Econometrics* (pp. 1–15). IntechOpen.

[9] Motamedi, M., Dawson, J., Li, N., Down, D. G., & Heddle, N. M. (2024). *Demand forecasting for platelet usage: From univariate time series to multivariable models*. PLOS ONE, 19(4), e0297391. https://doi.org/10.1371/journal.pone.0297391

[10] Etuk, E. H., Nwosu, C. R., & Moffat, I. U. (2012). *Seasonal autoregressive integrated moving average (SARIMA) model for precipitation time series*. Journal of Mathematics and Statistics, 8(2), 215–221. https://www.researchgate.net/publication/289636031

[11] Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304), 993–1010.

[12] Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

[13] Mohammadi, R., Shabanpour, R., & Aghaei, J. (2022). Multivariate time-series blood donation/demand forecasting for healthcare supply chain management. *Computers, Materials & Continua*, 70(2), 2145–2162.

[14] Chideme, C., & Chikobvu, D. (2024). Application of time-series analysis and expert judgment in modeling and forecasting blood donation trends in Zimbabwe. *MDM Policy & Practice*, 9(1), 23814683231222483.
[15] Octavia, M., Setiawan, N. A., & Hidayat, R. (2023). Blood bags demand forecasting using ARIMA. *AIP Conference Proceedings*, 2727(1), 030006.
[16] Chideme, C., & Chikobvu, D. (2024). Application of time-series analysis and expert judgment in modeling and forecasting blood donation trends in Zimbabwe. *MDM Policy & Practice*, 9(1), 23814683231222483.
[17] Octavia, M., Setiawan, N. A., & Hidayat, R. (2023). Blood bags demand forecasting using ARIMA. *AIP Conference Proceedings*, 2727(1), 030006.
[18] Shokouhifar, M., & Ranjbarimesan, M. (2022). Multivariate time-series blood donation/demand forecasting for resilient supply chain management during COVID-19 pandemic. *Cleaner Logistics and Supply Chain*, 5, 100078.
[19] Turkulainen, E. (2020). Forecasting blood demand: Observations and an implementation.
[20] Thakur, S. K., Sinha, A. K., Negi, D. K., & Singh, S. (2024). Forecasting demand for blood products: Towards inventory management of a perishable product. *Bioinformation*, 20(1), 20.
[21] Government of Malaysia, Ministry of Health Malaysia, & National Blood Centre. (2025). *Daily blood donations by blood group & state* [Dataset]. data.gov.my.
[22] Rajawat, A. S., Ghosh, A., *et al.* (2022). Mean absolute error (MAE). In *Applications of AI and IoT in Renewable Energy* (Chapter 6.6.1). Elsevier.
[23] Roberts, A. (2023). Mean Absolute Percentage Error (MAPE): What you need to know. *Arize AI Blog*. https://arize.com/blog-course/mean-absolute-percentage-error-mape-what-you-need-to-know/
[24] Frost, J. (2023). Root mean square error (RMSE). *Statistics By Jim*. https://statisticsbyjim.com/regression/root-mean-square-error-rmse/