

Forecast of Clinical Waste using Trendline Linear and Moving Average

Syifa' Sufiah Mustakim, Nur Irdina Isham, Lau Xiu Lian, Jamilah Mohd Ghazali*

*Department of Science and Mathematics, Centre for Diploma Studies,
Universiti Tun Hussein Onn Malaysia, Pagoh Higher Education Hub, 84600 Pagoh, Johor, MALAYSIA*

*Corresponding Author: jamilahg@uthm.edu.my

DOI: <https://doi.org/10.30880/mari.2025.06.01.022>

Article Info

Received: 01 October 2024

Accepted: 30 November 2024

Available online: 15 January 2025

Keywords

Waste management, performance accuracy, Mean Absolute Error (MAE), public health

Abstract

The management of medical waste is increasingly critical due to environmental impact and economic considerations. The forecasting of medical waste generation is crucial for effective resource allocation and waste management strategies. However, the forecasting value of medical waste generated remains uncertain due to the complex and dynamic nature of healthcare facilities. By exploring the Trendline Linear method, which leverages linear regression to model the relationship between time and waste generation, and the Moving Average (MA) method, this study provides the forecasting value of clinical waste generated in 2023 and to address this gap by evaluating the effectiveness of two forecasting methods by using performance accuracy Mean Absolute Error (MAE) in predicting the generation of clinical waste. By doing so, it seeks to provide insights into the reliability and accuracy of forecasting methods in the context of clinical waste management, ultimately contributing to the development of more efficient and sustainable waste management practices in healthcare settings. Further research is encouraged to explore the applicability of these methods in diverse healthcare settings and to investigate additional factors that may influence medical waste generation patterns. The study seeks to provide insights into the reliability and accuracy of the two forecasting methods in the context of clinical waste management. The performance of the methods is evaluated using MAE metric to assess their ability to predict the generation of clinical waste. To conclude that the forecasting value for clinical waste generated in the year 2023 is 4584.57 tonnes/year by using Simple Linear Regression (SLR) and based on the Simple MA (SMA) calculations, the amount of clinical waste generated in the year 2023 is approximately 4737 tonnes per year. An evaluation of performance accuracy reveals that SLR outperformed SMA in this study, achieving an accuracy of 551.57 compared to SMA's accuracy of 756.89.

1. Introduction

In the context of waste management, an examination of the current state reveals the myriad challenges and issues faced by healthcare facilities in effectively handling clinical waste. Understanding the importance of forecasting in this domain becomes imperative, as it ensures meticulous planning and optimal resource allocation within the healthcare system. The significance of employing Trendline Linear and MA methods in forecasting for

This is an open access article under the CC BY-NC-SA 4.0 license.



clinical waste management cannot be overstated, emphasizing their applicability in addressing the complex challenges encountered by healthcare facilities. This comprehends sustainability and forecasting proves instrumental in realizing potential benefits, including cost reduction, environmental sustainability, and bolstered public health protection. By integrating accurate clinical waste forecasting, healthcare institutions not only achieve financial savings but also contribute to sustainable practices, thereby safeguarding both the environment and public health.

This study underscores the critical need for a strategic and well-informed approach to clinical waste management, highlighting the multifaceted advantages that accurate forecasting methods can bring to the forefront. According to [1], the researchers investigated a case study of the recent outbreak of the COVID-19 pandemic which resulted in the mass generation of medical waste which seemed to have further deteriorated the issue of environmental pollution. They proposed a novel application for the identification and classification of 10 hospitals in Iraq that generated more medical waste during the COVID-19 pandemic than others to address the issue more effectively. The the lack of preparedness in many countries to combat the challenge of COVID-19 resulted in numerous social, economic, environmental, and medical issues due to the limited recycling and inappropriate dumping facilities [1]. Therefore, the disposal of medical waste has become a strategic management issue across various countries, particularly in those countries where the medical infrastructure is not very developed and where there are no regulations for medical waste disposal.

Thus, this study proposed a novel intelligent medical waste management framework can help hospitals redefine their waste disposal strategies and design them within the perspective of a circular economy. This study focuses on the importance of accurate forecasting in managing medical waste, a growing concern in healthcare. While demand prediction and inventory control are crucial for patient care and cost efficiency, forecasting medical waste generation remains a challenge due to the complex nature of healthcare facilities[2][1]. This study aims to address this gap by evaluating the effectiveness of two forecasting methods, Trendline Linear and MA, in predicting the generation of clinical waste. By doing so, it seeks to provide insights into the reliability and accuracy of forecasting methods in the context of clinical waste management, ultimately contributing to the development of more efficient and sustainable waste management practices in healthcare settings. Inaccurate forecasting can lead to inadequate waste management, posing risks to public health. Hence, the research aims to determine the most accurate and applicable forecasting method for predicting clinical waste generation.

2. Materials and Method

2.1 Data Collection

17-years data provided by the Department of Statistics Malaysia were utilized to examine the forecasting accuracy of methods. The Department of Statistics Malaysia is responsible for collecting and interpreting statistics to formulate or implement government policies across various fields, including trade, commerce, industry, agriculture, and clinical waste. The data consists of information that includes historical data on which quantity of clinical waste is handled for destruction at an incinerator in Johor, Malaysia. Fig. 1 illustrates the quantity of clinical waste processed for destruction at an incinerator in Johor, as reported by the Department of Statistics Malaysia (DOSM). This figure provides a visual representation of the amount of clinical waste that was handled and incinerated in Johor over a specified period.

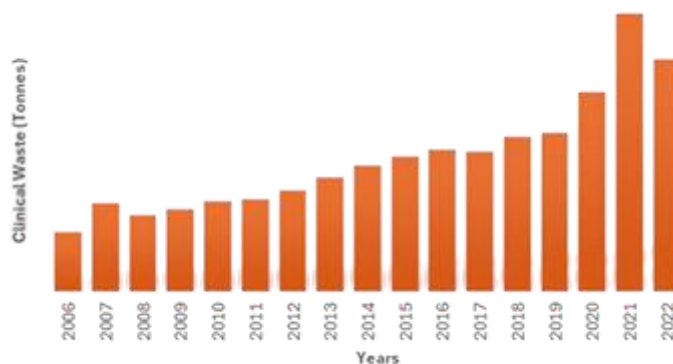


Fig. 1 Quantity of clinical waste handled for destruction at an incinerator in Johor from the Department of Statistics Malaysia (DOSM).

2.2 Trendline Linear

In this study, the Trendline Linear method used to forecast the quantity of clinical waste handled for destruction at an incinerator in Johor for 2023 was based on SLR. The SLR applies the least-squares method to search for the line of best fit for a set of data. It allows estimating the dependent variable, y with the aid of a given independent variable, x [3]. By plotting time on the x -axis (independent variable) and the quantity of clinical waste on the y -axis (dependent variable), SLR helps determine the linear relationship between them. In the context of clinical waste generation, linear regression can help identify and quantify the average rate of change in waste over time. The line of best fit can be defined by the equation $y = bx + a$ where b represents the slope of the line and a is the intercept. The values of a and b can be determined for a set of data comprising two variables and can be used to predict the value of y for any specified value of x [3]. The regression line summarizes the relationship between an independent variable and a dependent variable, characterized by its slope (rate of change) and intercept (initial value). This allows for projecting future values using the equation $y = bx + a$, facilitating informed predictions based on historical trends. The mathematical equation of the linear regression technique is very easy to understand, and its time complexity is much lower as compared to other machine learning algorithms [4]. Linear regression techniques can be used to model linearly separable data sets and can be used to find the nature of the relationship among variables [4]. The assumption here is that the relationship between time and clinical waste is best described by a straight line, making linear regression a simple yet effective method for forecasting [4]. Therefore, this study employed method for calculating the clinical waste generated utilizing Microsoft Excel spreadsheets.

2.2.1 Steps to forecast using Simple Linear Regression (SLR) in Excel Spreadsheet

To effectively forecast future values using SLR in an Excel spreadsheet, it's essential to follow a structured approach that involves several key steps:

- i. Step 1 (Insert Data): The data was organized in a spreadsheet with two columns:

Column 1: Year (Independent Variable) - Represents the year for which clinical waste data was recorded (2006-2022)

Column 2: Generation of Clinical Waste (Dependent Variable) - Represents the quantity of clinical waste handled for destruction at incinerators in Johor each year.

- ii. Step 2: Fitting the model by using historical data to estimate the regression coefficients.

The function of the line between the variables is to perform regression analysis. This function calculates the coefficients for the best-fit line by estimating the slope (a) and intercept (b) through the given data points. The fitted line's equation is shown in Fig. 2.

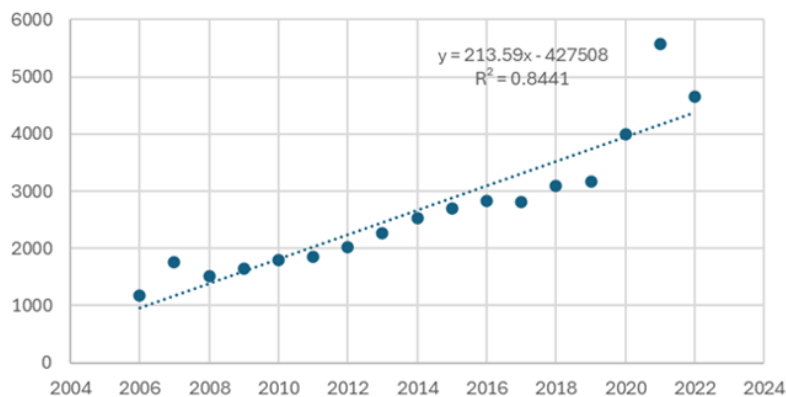


Fig. 2: Estimated regression coefficients line

- iii. Step 3: Testing the model

The objective of testing the model is to assess the model's performance in predicting clinical waste generation for unseen data in the year 2023 that needs to be forecast. The regression equation obtained in step 2 was used to predict clinical waste generation for the target year (2023). The equation was copied down throughout the spreadsheet in a new column (forecast column) to calculate the predicted value for 2023.

- iv. Step 4: Calculate Performance Evaluation

The performance evaluation used in this study was MAE. The calculation of the MAE will be further explained in the next discussion.

2.3 Moving Averages

This study incorporated the MA forecasting method as the SMA alongside the Trendline Linear method to enhance the accuracy of predicting the quantity of clinical waste handled for destruction at an incinerator by the state of Johor in 2023. The MA calculated the sequence time average, which was based on a certain number of items according to the time series data, to reflect the trend. Therefore, when the value of the time series significantly fluctuated owing to the influence of periodic variation and random fluctuation, and it was not easy to show the development trend of events, the moving average method can eliminate the influence of these factors and exhibit the trend of the time series [5]. This method forecasts value for the next period ($t+1$) by averaging a specified number of the most recent observations, each receiving equal weight. However, SMA was developed based on an average of weighted observations, which tends to smooth out short-term irregularity in the data series. It would be useful if the data series remained steady over time. SMA was a basic time-series analysis technique used to identify trends and smooth out fluctuations in data. In this paper, SMA was calculated for a yearly period to observe the clinical waste generation over time. The equation for predicting the value of F at time $t+1$ based on data up to time t in this study is shown as:

$$F_t = (\text{sum of actual values in previous } n \text{ periods})/n$$

$$F_{t+1} = \frac{(A_t + A_{t-1} + A_{t-2} + \dots + A_{t-n+1})}{n} \quad (1)$$

$$F_{t+1} = \frac{(A_t + A_{t-1} + A_{t-2})}{3}$$

The equation (1) is used to define forecasting data and actual data within the context of a moving average analysis. In this equation, F_{t+1} represents the forecasting data for the period, which is the predicted value for the next time period. A_{t-1} denotes the actual data for the previous period, which is the observed value from the most recent completed period. Additionally, n signifies the number of periods included in the MA calculation. This parameter determines the length of the data window used to compute the average, thereby influencing the smoothing effect and accuracy of the forecast [6]. By incorporating these elements, the equation provides a systematic approach to predicting future values based on historical data trends.

The SMA method calculates the arithmetic mean of the n most recent observations, incorporating the newest data and excluding the oldest for each forecasting period, which makes it a rational choice for practitioners. This approach is particularly beneficial in identifying clearer trends amidst variability in clinical waste data over time. By averaging the total amount of clinical waste over a specific time window, SMA minimizes the impact of short-term fluctuations, leading to a smoother curve that highlights the overall trend direction. The simplicity and ease of calculation of SMA make them a practical choice for forecasting when the focus is on identifying general patterns rather than complex relationships. The method's ability to filter out period-to-period noise and yield smoother forecasts is enhanced as the number of periods (n) in the moving average increases. This method's straightforward implementation and effectiveness in smoothing data fluctuations make it an ideal choice for forecasting clinical waste generation based on historical data trends [7].

2.3.1 Steps to forecast using Simple Linear Regression (SLR) in Excel Spreadsheet

In this paper, the steps outlined to calculate a SMA using Microsoft Excel. Here are the key steps involved in performing SMA calculations with an Excel Spreadsheet:

i. Step 1: Insert Data

The data was organized in a spreadsheet with two columns:

Column 1: Year (Independent Variable) – Represents the year for which clinical waste data was collected (2006-2022)

Column 2: Generation of Clinical Waste (Dependent Variable) – Represents the quantity of clinical waste handled for destruction at incinerators in Johor for each year

ii. Step 2: Fitting the model by using historical data

To calculate the SMA using Microsoft Excel, start by clicking on the 'Data' tab in the Excel ribbon. Then, select the 'Data Analysis' option, which opens the Data Analysis dialog box. In this dialog box, choose the 'Moving Average' option and click 'OK' to proceed to the 'Moving Average' dialog box. In the 'Input Range' field, select the data range for which the SMA needs to be calculated. Enter '3' in the 'Interval' option to set a three-point moving average. Next, specify the 'Output Range' where the results will appear. Additionally, check the options for 'Chart Output' and 'Standard Errors' to include these in the results. Finally, click 'OK' to generate the moving average, which will be displayed in the specified output range, providing a clearer view of the data trends over time.

iii. Step 3: Testing the model

The objective of testing the model is to forecast the value of clinical waste generation for the year 2023 and to evaluate the associated standard errors. The forecast value will be visualized through chart output. Using SMA calculations based on historical data, the model predicts the quantity of clinical waste for the upcoming year. Concurrently, it assesses forecast accuracy by examining standard errors. Visual representation via charts enhances understanding of both projected outcomes and the variability associated with forecasted clinical waste generation.

iv. Step 4: Calculate Performance Evaluation

The performance evaluation used in this study was MAE. The calculation of the MAE will be further explained in the next section.

2.4 Evaluation of Forecast Accuracy

In this study, MAE was applied to determine the accuracy of our two methods which are SLR and SMA values. MAE is a metric used to measure the average absolute differences between predicted values and actual values in a set of data. It provides a straightforward way to assess the accuracy of a forecasting or prediction model. In terms of quantifying average error, absolute-value-based measures such as the MAE are more interpretable than MSE or RMSE. Evaluations of model estimates and predictions should increasingly use absolute-value-based error measures such as MAE. These measures provide a more interpretable standard for evaluating model errors while also pointing to more specific types of error that may be reduced.

2.5 Steps to Evaluate Forecast Accuracy using Mean Absolute Error (MAE)

To calculate the mean absolute error of these two forecasting methods the steps below are the main steps to calculate the MAE.

i. Step 1: Calculate the difference between the actual and observed values for each record.

A dataset containing actual and predicted values in two separate Excel columns has been loaded. The data has 17 records of actual values (waste) and 18 predicted values (years) for each data point.

ii. Step 2: Calculate the absolute value of this difference.

In the calculation of MAE, we must find the absolute difference between the actual and observed values to obtain the value of MAE. The following formula = ABS (actual- observed) had been entered. This calculated the absolute difference between the actual and predicted values. The formula was copied down throughout the spreadsheet to calculate the absolute difference for each record. This method calculated the absolute difference for each record in the dataset, without needing to manually type the formula.

iii. Step 3: Calculate the average of these values.

Using the = AVERAGE A () function, the mean absolute error is calculated. The = AVERAGE A () function calculates the average value and includes logical and text values.

Table 1 Total Absolute Difference of SMA and SLR

Method	Year	Forecasted Value (tonnes)	Total Absolute Difference
SMA	2023	4737	11353.4
SLR	2023	4584.57	9928.25

Table 1 presents the following total absolute different absolute difference between SMA and SLR methods. The total absolute difference for the SMA method is 11353.4, and SLR for the SLR method is 9928.25. Manual computation of these two methods using mathematical formulas is consistent with those obtained from Excel calculations. To assess the forecasting accuracy of the two methods, MAE metric was employed. The results of this evaluation can be found in Table 2.

3. Result and Discussion

3.1 Forecast Simple Linear Regression (SLR)

The 17-years period of our data set from 2006 to 2022 gives the average annual clinical waste generated which Quantity of clinical waste handled for destruction at incinerator by the state, Johor, Malaysia was 2668.9765 tons with a standard deviation of 1174.002371 tons. This variation provides a general idea of the amount of clinical waste generated annually over this period and the standard deviation of the data indicates significant fluctuations around this average. In other words, the actual amount of waste generated in a given year could be 1174.002371 tons higher or lower than the average. A possible factor that affects the fluctuation of the variation is that the increase in the number of patients in Johor can lead to a rise in clinical waste generation and various hospital

activities or changes in procedure that may produce more waste. For example, a rise in surgery procedures may lead to an increase more waste compared to a period with mostly out-patient treatments.

This model is expressed by the following equation $y = 213.59x - 427,508$ which y represents the amount of clinical waste generated in tonnes and X represents the year. The model was constructed using a statistical technique called ordinary least squares regression. This method helps establish the best-fitting straight line that captures the relationship between the year (x) and the amount of clinical waste generated (y) within the data we analysed. Based on Fig. 3, it concludes that the forecasting value for clinical waste generated in the year 2023 is 4584.57 tonnes per year by using the regression equation of the model. The clinical waste is expected to increase from the year 2022 to 2023 based on forecasted value which is from 4370.98 tonnes to 4584.57 tonnes. In 2021, the amount of clinical waste rise from 3989.60 in 2020 to 55567.5 due to the widely known fact that the COVID-19 pandemic led to a major increase in healthcare services globally. This probably led to an increase in the production of clinical waste in Johor in 2021.

In regression analysis, the coefficient of determination, usually denoted by R -squared is also a useful statistic obtained from the regression model [8]. R -squared in regression acts as an evaluation metric to evaluate the scatter of the data points around the fitted regression line. It recognizes the percentage of variation of the dependent variable. R -squared is the proportion of variance in the dependent variable that can be explained by the independent variable[9]. The R -squared value obtained in this study is 0.8441, suggesting that 84% of the variability in clinical waste can be explained by years. A high value of R -squared that is close to 1 shows that the independent variables included in the regression model explained well the variation of the dependent variable. [8]. If the value of R is large, have a better chance of your regression model fitting the observations [9]. This signifies a strong relationship between x and y , making the model more reliable for forecasting future values of y based on x .

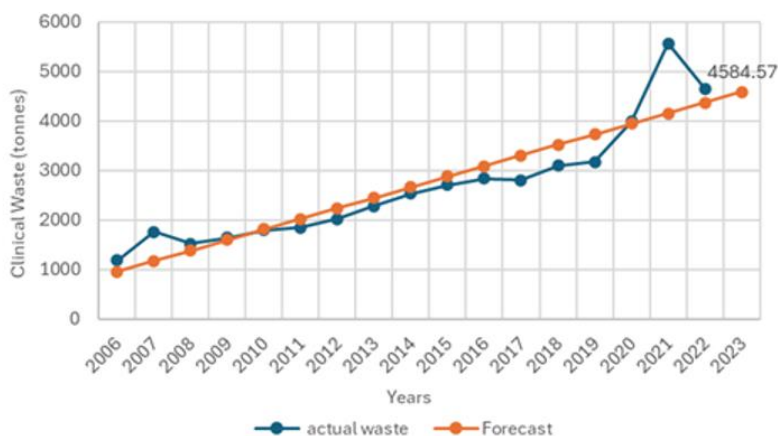


Fig. 3: Graph output SLR of clinical waste generation

On the other hand, the slope of observed clinical waste versus years alongside with fitted regression line clearly represents the rising trend in waste generation. The relationship between x and y is denoted using a line known as the regression line. A case study for megacity Istanbul by [3], aim to estimate the amount of waste to be generated in the coming years is critical for the evaluation of existing waste treatment service capacities. This study was conducted to evaluate the performance of various mathematical modelling methods such as using SVR, GM (1,1) and ARIMA models to forecast medical waste generation of Istanbul, the largest city in Turkey. This study analysis of observed trends suggests a steady increase in clinical waste generation. However, unlike their study which used ARIMA for forecasting, this study explores the effectiveness of simple linear regression for shorter-term predictions. This aligns with what the linear regression might forecast, assuming the chosen independent variable has a positive correlation with waste generation. The line may be in three ways, Positive slope, negative slope and zero slope. If the curve is positive slope, then the dataset is best suited for linear regression [10]. The analysis of this paper revealed an upward trend in clinical waste generation over the years. This is evident from the positive slope in the linear regression model. In simpler terms, this means that for every year that passes, an additional 213.59 tonnes of clinical waste is produced on average. Hence, the positive slope indicates that as the number of years increases, the amount of clinical waste generated also increases. Possible factors influencing the fluctuations and increasing trend in clinical waste generation like population growth and economic development [3].

Forecasting the future value of clinical waste generation with SLR provides strategies for improving waste management practices in many ways such as optimizing waste management. The expected rise in clinical waste can be adjusting waste collection schedules. Healthcare facilities expect large amounts of clinical waste can schedule frequent collection to avoid spilled over bins and possible dangers to health which may arise from

biohazards. Improper management of clinical waste poses a public health risk [11]. In addition, prior knowledge about the amount of medical waste generated can be used for both the planning and design of future facilities [3].

3.2 Forecast Simple Moving Average (SMA)

SMA is particularly useful when there is variability in the clinical waste data over time, and a clearer trend needs to be identified. As it stated in method, the data set spanning 17 years from 2006 to 2022 reveals the average annual clinical waste generated in Johor, Malaysia. By calculating the average of the total amount of clinical waste over a specific time window in the study, a moving average (over the past few years), the impact of short-term fluctuations can be minimized. Increasing the number of observations of actual values tends to smooth out fluctuation and easily obtain general information about fluctuation. This results in a smoother curve that highlights the overall direction of the trend.

Flood potential using SMA and Holt-Winters methods was investigated by [12]. The analysis was supported by collected data from observations, literature studies, and Landsat 8 images of land objects one of the analysis steps involved visualizing and forecasting the likelihood of high inundation or flooding based on NDWI accuracy using the SMA and Holt-Winters methods. Several new actual demand data were used to generate forecast values for future demand using the SMA algorithm. The study aimed to visualize flooding likelihood and trends based on NDWI values, which is similar to the approach of using SMA to identify trends in clinical waste generation in this study. Despite its simplicity, SMA effectively determines current trends by smoothing out short-term fluctuations.

To investigate clinical waste generation, this paper implemented a forecasting model based on the SMA model. This model smooths out the data by creating an average of any subset of the complete data set. A 3-period moving average was employed to analyse trends in clinical waste generation data. In this study the independent variable (time) and the dependent variable (clinical waste generated). This method helps to reduce the noise from random fluctuations in the data, providing a clearer view of the underlying trend.

According to [13], both the 3-period MA and the 5-period MA have a similar magnitude of error, but it is better to use the 3-period MA rather than the 5-period MA. Therefore, the 3-period MA is used in this study. The 3-period MA is calculated starting from the fourth actual data point, which is the data for the year 2009. Fig.4 presents the chart output SMA between the actual values and forecast values of clinical waste generation. Based on the MA calculations, the forecast value of clinical waste generated in the year 2023 is approximately 4737 tonnes per year. In SMA analysis, the window size of the average plays a crucial role in determining the smoothness and accuracy of the forecast. A smaller window size may capture more recent trends but can be more volatile, while a larger window size provides a more stable average but may lag in capturing recent changes. In this paper, a 3-period window was selected to balance responsiveness and stability.

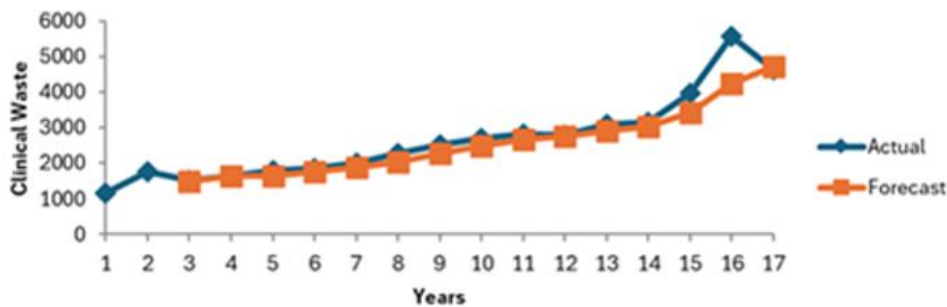


Fig. 4 Chart output SMA of clinical waste generation

The results from the SMA model indicated a rising trend in clinical waste generation over the years. This trend was evident from the increasing values of the moving averages over the observed period. In simpler terms, the data shows that the amount of clinical waste generated has been consistently increasing, reflecting a growing healthcare demand and corresponding waste production in Johor, Malaysia.

3.3 Accuracy Assessment

Table 2 presents the MAE for SMA and SLR.

Table 2 MAE for SMA and SLR

Method	Mean Absolute Error
SMA	756.8933
SLR	551.5694

Based on Table 2, the SLR model achieved a lower MAE of 551.5694 compared to the SMA model's MAE of 756.8933 indicating that the SLR model produced predictions that were closer to the actual values in the data. Since the MAE for SLR is lower than the MAE for SMA, the SLR model is performing better at predicting the actual values in our dataset compared to the SMA model [14].

This difference between predicted and actual values, captured by the MAE can be further analysed by examining the residuals. A positive residual corresponds to a point in the scatter plot that lies above the fitted line generated by the SLR model. Conversely, a negative residual indicates a data point falling below the line. [15]. These residuals, along with the predicted values (fits) are presented in the accompanying table for further exploration. This table allows us to not only see the overall average error (MAE) but also to identify potential patterns in the errors for specific data points.

Furthermore, the interpretability of MAE is bolstered by its range of 0 to positive infinity [16]. A perfect score of 0 indicates no absolute difference between predicted and actual values. On the other hand, large errors or outliers in the data can cause MAE to approach positive infinity. This happens because MAE sums up absolute differences between predictions and actual values. Even a single instance of a very large difference can significantly inflate the MAE, especially if the model performs poorly for outliers. While reaching the infinity is uncommon in practical applications, a very high MAE relative to the scale of our data indicates a poor model fit [14].

3.4 Limitations

The study was limited by data limitations of data historical on waste generation. To manage the scope of our study, we chose to focus on the data regarding the quantity of clinical waste handled for destruction at an incinerator in Johor. Although clinical waste in Malaysia is classified into subcategories by regulations, our ability to explore these details is constrained by limited data from the Department of Statistics Malaysia. Therefore, our data focusing on incinerated waste offered a readily available dataset that aligns with a common disposal method for clinical waste in the region.

The lack of 2023 data creates inaccuracies in your predictions of clinical waste generation. Lacking this new data could hinder the model's learning process, resulting in possible inaccuracies or being too focused on previous patterns. Underestimations may occur if waste reduction efforts become successful in 2023, or overestimation could happen if new protocols calling for increased use of disposables are introduced.

For the SLR model, only one independent variable is considered at a time and does not include many variables that are likely to influence clinical waste generation by multiple factors, such as bed occupancy rates, types of procedures, and so on. The study of [3] evaluates the performance of ARIMA, SVR, and Grey Modeling (1,1) on medical waste forecasting in Istanbul. While it does not use simple linear regression, it provides a benchmark for how well more sophisticated models perform compared to a simpler method. Therefore, it gives an idea of the potential limitations of using linear regression for forecasting clinical waste generation.

The SMA is a popular forecasting method due to its simplicity and ease of use. However, it has a significant limitation which is its reliance on a fixed window size that can impact its effectiveness in forecasting. In this paper using a 3-period moving average, this means that each forecast is based on the average of the three years of actual value for clinical waste generated in case to predict the forecasting value until year 2023. If the chosen window size does not adequately capture the underlying data pattern, the forecasts may either be too smooth or too noisy [13].

While MAE offers a robust and interpretable measure of model accuracy, it has limitations. MAE treats all errors equally, regardless of direction, which can be problematic when underestimations or overestimation have different consequences. Additionally, MAE may not adequately capture the impact of outliers due to its focus on absolute differences. Finally, the scale dependence of MAE hinders direct comparisons between models applied to data with different units. To address these limitations, consider using MAE in conjunction with other error metrics that offer complementary strengths [17].

3.5 Future Research

Future research holds immense potential to refine and enhance clinical waste generation forecasts. Expanding data sources beyond current limitations is crucial. Collaborations with hospitals, waste management companies, and regulatory bodies can unlock a wealth of information on waste composition, generation determinants like bed occupancy and procedure types, and even broader economic and social factors. Integrating this diverse data into a unified format will be essential.

The steps taken to mitigate this limitation lack of 2023 actual data on the forecasting accuracy is to focus on using the most recent data available and consider incorporating external information like industry reports or expert opinions. Short-term forecasts are less vulnerable to this data gap, and transparency is key – acknowledge the limitations and highlight any assumptions made with external data. Scenario planning with different

possibilities for 2023 can provide a range of potential outcomes. In extreme cases, explore alternative forecasting methods less reliant on historical data, like expert surveys or judgmental forecasting.

Second, the development of more sophisticated forecasting models is warranted. [3] demonstrated the effectiveness of ARIMA, SVR, and Grey Modeling for medical waste forecasting. These models hold promise for capturing the complex interplay between various factors and waste generation. Additionally, exploring hybrid models that combine the strengths of simpler methods like SLR with more advanced techniques could prove beneficial.

Finally, addressing the limitations of current forecasting methods is crucial. Employing dynamic window sizes for the SMA can improve its ability to adapt to changing trends in the data. Statistical methods can be used to determine the optimal window size for each forecast. Furthermore, supplementing Mean Absolute Error (MAE) with other error metrics like MSE or Root Mean Squared Error (RMSE) is recommended. This will provide a more comprehensive picture of forecasting accuracy, particularly regarding the influence of outliers. By pursuing these directions, future research can overcome the limitations of the current study and establish a more robust and informative model for clinical waste generation forecasting.

4. Conclusion

The study faced limitations primarily due to the historical data constraints on waste generation. To manage the scope, the research focused on clinical waste quantities handled for incineration in Johor, Malaysia. Although clinical waste in Malaysia is categorized into subcategories as per regulations, our study's depth was limited by the available data from the Department of Statistics Malaysia. The dataset on incinerated waste provided a consistent and relevant basis for analysis, reflecting a prevalent disposal method in the region.

The SLR model creates a best-fit line to model the relationship between years and waste generation. It predicts a steady increase of about 213.6 tonnes per year on average. The model explains 84% of the variation in waste generation ($R\text{-squared} = 0.8441$) and has a lower MAE compared to the SMA method, indicating better prediction accuracy. It employed in the study considered only one independent variable at a time.

For the SMA method, a 3-period moving average was used, meaning forecasts were based on the average of the past three years of actual clinical waste data. As a result, it forecasts a generation of 4737 tons in 2023. This approach can be problematic if the window size does not accurately capture the data's underlying patterns, potentially resulting in forecasts that are either overly smooth or excessively noisy.

While the MAE provides a robust and interpretable measure of model accuracy. The MAE ranges from 0 (perfect fit) to positive infinity, with a high MAE suggesting a poor model fit, especially relative to the data scale. To employ the MAE to assess the accuracy of the SMA and SLR methods in predicting the values of clinical waste generation, it is essential to recognize the MAE's strengths and limitations. The MAE averages the absolute errors and treats all errors equally, regardless of direction, which can be problematic when underestimations and overestimation have different consequences.

In summary, while the chosen methodologies, SMA and SLR provided valuable insights, the study highlights the importance of addressing data limitations and considering multiple variables and error metrics to enhance forecasting accuracy and reliability in clinical waste management.

Acknowledgment

The authors would like to thank the Centre for Diploma Studies, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

*The authors confirm contribution to the paper as follows: **study conception and design:** Syifa' Sufiah Mustakim, Nur Idina Isham, Lau Xiu Lian, Jamilah Mohd Ghazali; **data collection:** Syifa' Sufiah Mustakim, Nur Idina Isham, Lau Xiu Lian, Jamilah Mohd Ghazali; **analysis and interpretation of results:** Syifa' Sufiah Mustakim, Nur Idina Isham, Lau Xiu Lian, Jamilah Mohd Ghazali; **draft manuscript preparation:** Syifa' Sufiah Mustakim, Nur Idina Isham, Lau Xiu Lian, Jamilah Mohd Ghazali. All authors reviewed the results and approved the final version of the manuscript.*

Appendix A: Clinical Waste Forecast Analysis Simple Linear Regression

	A	B	C	D	E
1	Years(x)	Waste(y)	Forecast	error	abs error
2	2006	1176.87	953.54	223.33	223.33
3	2007	1754.83	1167.13	587.7	587.7
4	2008	1515.5	1380.72	134.78	134.78
5	2009	1639.9	1594.31	45.59	45.59
6	2010	1795.2	1807.9	-12.7	12.7
7	2011	1850.1	2021.49	-171.39	171.39
8	2012	2015.5	2235.08	-219.58	219.58
9	2013	2276.7	2448.67	-171.97	171.97
10	2014	2526.2	2662.26	-136.06	136.06
11	2015	2700.8	2875.85	-175.05	175.05
12	2016	2834.2	3089.44	-255.24	255.24
13	2017	2806	3303.03	-497.03	497.03
14	2018	3097.5	3516.62	-419.12	419.12
15	2019	3173.6	3730.21	-556.61	556.61
16	2020	3989.6	3943.8	45.8	45.8
17	2021	5567.5	4157.39	1410.11	1410.11
18	2022	4652.6	4370.98	281.62	281.62
19	2023		4584.57	-4584.57	4584.57

Appendix B: Clinical Waste Forecast Analysis Simple Moving Average

	A	B	C	D	E
1	Years (x)	Waste (y)	Forecast (formula)	Error	ABS Error
2	2006	1176.87			
3	2007	1754.83			
4	2008	1515.5			
5	2009	1639.9	1482	157.9	157.9
6	2010	1795.2	1637	158.2	158.2
7	2011	1850.1	1650	200.1	200.1
8	2012	2015.5	1762	253.5	253.5
9	2013	2276.7	1887	389.7	389.7
10	2014	2526.2	2047	479.2	479.2
11	2015	2700.8	2273	427.8	427.8
12	2016	2834.2	2501	333.2	333.2
13	2017	2806	2687	119	119
14	2018	3097.5	2780	317.5	317.5
15	2019	3173.6	2913	260.6	260.6
16	2020	3989.6	3026	963.6	963.6
17	2021	5567.5	3420	2147.5	2147.5
18	2022	4652.6	4244	408.6	408.6
19	2023		4737	-4737	4737

References

- [1] X. Chew, K. W. Khaw, A. Alnoor, M. Ferasso, H. A. Halbusi, and Y. R. Muhsen, "Circular economy of medical waste: novel intelligent medical waste management framework based on extension linear Diophantine fuzzy FDOSM and neural network approach," *Environmental Science and Pollution Research*, vol. 30, no. 21, pp. 60473–60499, Apr 2023, doi: 10.1007/s11356-023-26677-z.
- [2] Rajon Paul Pantha, "Demand Prediction and Inventory Management of Surgical Supplies," pp. 1–14, 2023, [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK558907/>
- [3] Z. Ceylan, S. Bulkan, and S. Elevli, "Prediction of medical waste generation using SVR, GM (1,1) and ARIMA models: a case study for megacity Istanbul," *Journal of Environmental Health Science and Engineering*, vol. 18, no. 2, pp. 687–697, Jun 2020, doi: 10.1007/s40201-020-00495-8.
- [4] Muhammad Ahmad Iqbal, "Application of regression techniques with their advantages and disadvantages," *Elektron Mag.*, vol. 4, no. September, pp. 11–17, 2021.
- [5] Y. Zhang, X. Deng, Y. Zhang, and Y. Zhang, "Generation of sub-item load profiles for public buildings based on the conditional generative adversarial network and moving average method," *Energy and Buildings*, vol. 268, p. 112185, Aug 2022, doi: 10.1016/j.enbuild.2022.112185.
- [6] A. S., "Sales predictor using simple linear regression and simple moving average model," *International Journal for Research in Applied Science and Engineering Technology*, vol. 11, no. 11, pp. 729–736, Nov 2023, doi: 10.22214/ijraset.2023.56609.
- [7] R. D. E. Renel and P. L. Samputra, "Strategy for Services anticipation at the Rehabilitation Center of BNN based on moving average and exponential smoothing Forecasting method," *Technium Social Sciences Journal*, vol. 42, pp. 264–282, Apr 2023, doi: 10.47577/tssj.v42i1.8700.
- [8] S. F. Ng, Y. M. Chew, P. E. Chng, and K. S. Ng, "An insight of linear regression analysis," *Scientific Research Journal*, vol. 15, no. 2, p. 1, Dec 2018, doi: 10.24191/srj.v15i2.9347.

- [9] M. Keer, H. Lohiya, and S. Chouhan, "Goodness of Fit for Linear Regression using R squared and Adjusted R-Squared," *Int. J. Res. Publ. Rev. J. homepage www.ijrpr.com*, vol. 4, no. 3, pp. 2431–2439, 2023, [Online]. Available: www.ijrpr.com
- [10] J. Jayasudha, "A Simple Linear Regression with Statistical Approach for Building Machine Learning Model and Prediction," vol. 9, no. 7, pp. 6038–6043, 2020.
- [11] R. Tabasi and G. Marthandan, "Clinical Waste Management : A Review on Important Factors in Clinical Waste Generation Rate," *Internatioanal J. Sci. Technol.*, vol. 3, no. 3, pp. 194–200, 2013.
- [12] R. H. Ramadhan, R. Yusman, and G. T. Pranoto, "Comparison of Holt winters and simple moving average models to identify the best model for predicting flood potential based on the normalized difference water index," *JISA(Jurnal Informatika Dan Sains)*, vol. 5, no. 2, pp. 99–105, Dec 2022, doi: 10.31326/jisa.v5i2.1316.
- [13] I. Otudi and A. Almaktoom, "The Impact of Forecasting Methods on Demand Projection in Fast Food Restaurant," *Palarch's J. Archaeol. Egypt/Egyptology*, vol. 18, no. 15, p. 334, May 2021..
- [14] Lin, H., and Li, M. *Practitioner's Guide to Data Science (1st ed.)*. Chapman and Hall/CRC. 2023. <https://doi.org/10.1201/9781351132916>
- [15] J. L. Devore, *Probability and Statistics for Engineering and the Sciences, Seventh Edition, Enhanced Edition*. 2009.
- [16] T. Iida, "Identifying causes of errors between two wave-related data using performance metrics," *Applied Ocean Research*, vol. 148, p. 104024, Jul 2024, doi: 10.1016/j.apor.2024.104024.
- [17] K. Parang, L. I. Wiebe, and E. E. Knaus, "Novel approaches for designing 5-O-Ester prodrugs of 3-Azido-2,3-dideoxythymidine (AZT).," *Current Medicinal Chemistry*, vol. 7, no. 10, pp. 995–1039, Oct 2000, doi: 10.2174/0929867003374372.