Journal of Engineering Science and Technology
21st EURECA 2024 Special Issue August (2025) 1 - 10


# OPTIMIZING WASTE COLLECTION ROUTE WITH MULTI-LINEAR REGRESSION

FELIX VINCENT*, PATRICK THEN,
SIM KWAN HUA, HUDYJAYA SISWOYO JO

Swinburne University of Technology, Sarawak, Malaysia
*Corresponding Author: fvincent@swinburne.edu.my

## Abstract

The increasing volume of waste necessitates innovative approaches to waste management. This study presents a novel methodology for optimizing waste collection routes by integrating predictive modelling and route optimization techniques. Specifically, we propose a multi-linear regression model combined with Dijkstra's algorithm to optimize routes considering distance, congestion, and waste weight. Unlike traditional methods, the approach prioritizes fuel efficiency and sustainability, aiming to minimize costs and environmental impact. By leveraging historical data from waste management systems, a predictive model is developed to estimate fuel consumption based on route characteristics. This model informs route optimization decisions, with stops prioritized based on predicted fuel consumption. Through experimentation, the research demonstrates the effectiveness of the approach in reducing fuel consumption and improving route efficiency. The research contributes to the field of waste management by offering a comprehensive framework for optimizing waste collection routes.

Keywords: Dijkstra's algorithm, Fuel efficiency, Multi-linear regression, Route optimization, Sustainability, Waste collection.

# 1. Introduction and Literature Review

Waste management is essential for a country's economy and the well-being of its people. Unfortunately, waste production is rising worldwide. In Malaysia alone, over 23,000 tonnes of waste are produced every day, as noted by Sreenivasan et al. [1]. In Kuching, Malysia, each person generates about 0.7 kg of waste daily, according to Trienekens. This surge is mainly due to growing populations and expanding cities, leading to overflowing garbage bins in many public areas. Sadly, Malaysia's waste management practices are still lacking, hindered by outdated documentation and collection systems.

Managing waste effectively incurs significant expenses, covering collection, transportation, processing, and disposal. Researchers aim to minimize these costs to ensure the financial sustainability of waste management systems and allocate resources to other societal needs. Minimizing costs also promotes efficient waste management practices, including optimizing collection and transportation routes.

However, existing waste management systems face challenges such as fixed routes, inadequate responsiveness to demand changes, and manual stop allocation, leading to inefficient resource use. This paper proposes innovative strategies, focusing on optimization techniques to address these challenges. By integrating advanced algorithms considering factors like distance, congestion, and weight, the aim is to minimize costs while enhancing overall efficiency and sustainability in waste management systems.

## 1.1. Problem statement and significance

This study focuses on using a regression model to address municipal solid waste collection problems. By developing a methodology that integrates predictive modelling with route optimization, it aims to overcome the inefficiencies and high operational costs of existing waste management systems. The significance of this study lies in its potential to enhance the sustainability and efficiency of waste collection operations through the introduction of innovative techniques.

## 1.2. Literature review

He [2] compared the improved Dijkstra algorithm with the prim algorithm and Huffman coding algorithm, noticing that it preserves capacity space and progresses execution speed. Pyramid trees produced by the Lumetta 2.3 Dijkstra algorithm and the Pile in Dijkstra algorithm recognize all prompt focuses reachable from the beginning point and put these focuses into a pile at the side their separations. The Dijkstra algorithm is more productive than other calculations since it calculates the shortest way to each point. This effectiveness is illustrated within the illustration underneath, displaying its viable appropriateness.

Gunawan et al. [3] developed a search application designed to assist the public in locating specialist medical practices at the nearest hospitals in Bandar Lampung. This application aims to streamline the process of finding needed specialist doctors by providing a user-friendly interface. To achieve this, the application employs Dijkstra's algorithm to identify the shortest route to the hospital where the nearest specialist practices.

Hossain et al. [4] implemented route optimization using Dijkstra's algorithm for garbage truck operations. The trucks follow the optimized route to collect waste from bins based on the most efficient path. After each collection, the bins are removed from the graph, and the link cost is recalculated before applying Dijkstra's algorithm again. This iterative process continues until all nodes are visited. In this implementation, values for distance, bin status, and congestion were randomly assigned. The system can automatically determine the best route and the quantity of waste in the bins, ensuring that the appropriate size of garbage truck is dispatched. This fully automated process enhances efficiency in waste collection operations.

Hannan et al. [5] describes a model on optimizing cost saving, and emission reduction of waste collection vehicles by reducing fuel consumption. The model considers several phases of waste collection, including driving from the waste terminal to the collection zone, collecting trash, driving to a treatment office, and emptying the waste. The fuel consumption during the collection stage is calculated using an equation that considers the overall diesel utilization for one ton of trash collected, the whole fuel consumption when the vehicle is driving empty or full, and the full tons of collected trash.

Wang et al. [6] discusses the effect of waste weight on fuel consumption and efficiency in freight transportation. Vehicle weight influences the power needed for acceleration, overcoming rolling resistance, and climbing hills. The heavier the vehicle, the more energy is required. For instance, a complete loaded tractor-trailer can weigh up to 80,000 lbs.

Reducing waste weight has the potential to improve freight transportation efficiency by increasing the freight delivered per ton-mile. A referenced simulation demonstrates the effects of weight reduction on various driving cycles, including Heavy-Duty Diesel Truck (HHDDT), Composite International Truck Local and Commuter Cycle (CILCC), and West Virginia University City (WVU City). It reveals that weight reduction provides benefits on these cycles, particularly on the high transient speed cycle (WVU City), where a significant 12% reduction in fuel consumption was achieved when decreasing the vehicle weight from 36,000 lbs. to 28,800 lbs.

Çapraz et al. [7] discusses research that focuses on enhancing fuel consumption predictions for long-distance travel by evaluating different statistical models. The models assessed include Multiple Linear Regression, Artificial Neural Networks (ANN), and Support Vector Regression (SVR). These models are applied to data generated from three types of automobiles. The evaluation criteria encompass accuracy and consistency in predicting fuel consumption. Notably, SVR consistently outperforms the other models, demonstrating its efficacy in forecasting fuel consumption more accurately.

While MLR is a useful and commonly used statistical technique, it may not capture complex relationships as effectively as SVR, especially when dealing with nonlinear or intricate patterns in the data. SVR, with its ability to handle high-dimensional data and nonlinear relationships, demonstrated better performance in the context of predicting fuel consumption for long-distance travel in the study.

In broader terms, the study aims to identify the optimal method for estimating fuel consumption during extended travel, particularly capitalizing on the inclusion of road slope as a crucial parameter. The choice of the most suitable model often depends on the nature of the data and the underlying relationships being modelled, and SVR proved to be more effective in this particular scenario.

Ozili [8] examined in social science research the acceptable range for R-squared values. He noted that the general acceptable range is at an R-squared value between 0.51 and 0, particularly when most explanatory factors are statistically significant. However, he cautioned that a high R-squared ought to not result from spurious causation or multicollinearity among the informative factors. The paper also argued that a low R-squared value (at slightest 0.10) can be accepted in social science experimental modelling, given a few or most explanatory factors are statistically significant. The main takeaway is that the evaluation of model goodness should prioritize the statistical significance of explanatory factors rather than solely relying on the R-squared value. This commentary aims to guide young researchers new to empirical research in social sciences.

Uyanik and Güler [9] conducted a study using multiple linear regression analysis to explore the predictive capabilities of five independent factors: end-of-term scores in estimations and assessments, educational psychology, curriculum development, guidance, and teaching methods. These variables were analysed within a standard model to predict the KPSS score, which served as the dependent variable. The analysis, involving ANOVA statistics, included 240 undergraduate students from the Psychological Counselling and Guidance, Turkish Education, and Science Education departments at Sakarya University during the 2011-2012 academic year.

Following an examination of the assumptions required for multiple linear regression, the model was found to significantly predict the dependent variable, with a correlation coefficient of R=0.932 and an $R^2$ value of 0.87, showing that the model clarifies 87% of the fluctuation within the dependent variable. The highest contribution came from measurement and evaluation scores ($\beta$=1.421), followed by scores in teaching methods, guidance, curriculum development, and educational psychology. The study recommends conducting further research with more diverse variables and larger sample sizes to better understand the factors that predict scores in the KPSS, a crucial assessment for aspiring teachers' post-graduation.

Kushwaha, M.; and Abirami [10] conducted a comprehensive study to compare the efficacy of various regression models, including Multiple Linear Regression (MLR), Support Vector Machines (SVM), and Random Forests, in predicting the duration of road accidents. The researchers utilized a dataset containing detailed information about road accidents and applied each regression model to predict accident durations. Their findings indicated that MLR performed comparably, and in some cases better, than SVM and Random Forest models. The key reason for this was the linear nature of the relationships between the predictors and the target variable in the dataset. Additionally, MLR offered the advantages of simplicity, interpretability, and computational efficiency, which are significant in practical applications where understanding the model is crucial for decision-making processes.

### 1.3. Significance of using regression models

Hossain et al. [4] indicates that optimization techniques and algorithms have been applied to improve waste management systems. However, the specific application of regression models to predict fuel consumption in waste collection routes has not been thoroughly explored. This study fills this gap by integrating a multi-linear regression model with Dijkstra's algorithm, considering distance, congestion, and weight to predict fuel consumption and optimize routes. This approach not only aims to improve fuel efficiency but also addresses the economic and natural impacts of waste management operations.

## 2. Methods

This study aims to develop a methodology for improving route planning in waste collection by considering factors such as distance, road congestion, and waste weight. Data on these variables, along with fuel consumption, will be gathered from historical records of waste collection operations from Trienekens Sarawak (Waste Management Company). The predictive model will utilize this data to estimate fuel consumption, which will then be used to prioritize stops in route planning. Specifically, stops with lower predicted fuel consumption will be given priority. Dijkstra's algorithm will subsequently be employed for route optimization. The linear regression model will be formulated to predict fuel consumption:

$$Fuel\ Consumption\ = \beta 0 \times\ \beta 1\ \times Distance\ \times \beta 2 \times Road\ Congestion\ \times\ \beta 3 \times Weight + \epsilon$$

The coefficients (β1, β2, β3) will be estimated using statistical software such as Python. This model will serve as the basis for predicting fuel consumption based on route characteristics. For route optimization, an algorithm will be utilized to find the most efficient paths for waste collection. This algorithm considers factors such as distance between stops and prioritizes routes with minimal fuel consumption. Integration with fuel consumption predictions involves incorporating estimated fuel consumption values into the route optimization algorithm. Stops with lower predicted fuel consumption are given higher priority during route planning.

The decision criteria for prioritizing stops during route optimization are based on minimizing fuel consumption, thereby enhancing overall fuel efficiency and cost savings. Validation of the predictive model and route optimization algorithm will be conducted using a separate dataset. Sensitivity analysis will assess how variations in input variables affect fuel consumption predictions and route optimization outcomes. This methodology combines predictive modelling for fuel consumption with an optimization algorithm for waste collection routes. By prioritizing stops with lower predicted fuel consumption, the aim is to improve the productivity and sustainability of waste collection operations. Overall research modules as shown in Fig. 1.

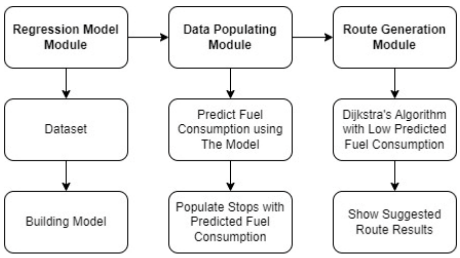


**Fig. 1. Overall research modules.**

The research methodology comprises three main modules, as illustrated in Fig. 2. Firstly, the regression model module involves gathering a dataset consisting of independent variables such as distance, congestion, and weight, along with the dependent variable, fuel consumption. This dataset is used to calculate coefficient correlations, employ the multiple regression equation, and assess the model's goodness through the R-squared value. Acceptable R-squared values, as noted by Ozili [8], range between 0.51 and 0.99.

Secondly, the data preparation module facilitates predicting fuel consumption based on new data for distance, road congestion, and waste weight in route collection. Fuel consumption data is then assigned to each data point (stop), representing the collection area's smallest unit.

Lastly, the routing algorithm determines the foremost efficient route by using Dijkstra's algorithm. Unlike traditional methods that consider only distance, this algorithm prioritizes fuel consumption as the key weight factor, thereby aligning with the research goal of minimizing waste management costs

## 3. Experiment and Results

The experiment aims to analyse the relationship between distance, congestion, waste weight, and fuel consumption using multiple linear regression (MLR) analysis. The data collected from waste collection operations in Trienekens Sarawak includes variables such as distance travelled, time of road congestion encountered, waste weight carried, and fuel consumption recorded. At least of 1000 data points were used for the analysis.

The Ordinary Least Squares (OLS) regression resulted in the experiment with an R-squared value of 0.998. This high R-squared value indicates that the independent variables (distance, congestion, and waste weight) collectively explain approximately 99.8% of the variation in fuel consumption. The percentage of error in predictions was calculated to validate the model, showing minimal differences between predicted and actual fuel consumption values.

To further analyse and validate the regression model, the prediction is tested with a few rows in the dataset. As a result, the predicted fuel consumption does not have huge differences with the real dataset, indicating that the model is performing well and well trusted.

The experiment demonstrates a strong relationship between the independent variables (distance, congestion, and waste weight) and fuel consumption, as indicated by the high R-squared value of 0.998. This suggests that these variables collectively provide a highly accurate prediction of fuel consumption. However, it is essential to validate the assumptions underlying the regression model and ensure the absence of multicollinearity or other potential issues.

As a result, the model is used to predict new data of distance, congestion, and weight, it should be able to predict fuel consumption. Below are example of new data and prediction results:

```
Predicted fuel consumption for data point 1:
    Distance: 70 km
    Congestion: 185 minutes
    Waste Weight: 8270 kg
    Predicted Fuel Consumption: 87.48640882795452

Predicted fuel consumption for data point 2:
    Distance: 81 km
    Congestion: 190 minutes
    Waste Weight: 9420 kg
    Predicted Fuel Consumption: 97.05538633345307
```

**Fig. 2. Model prediction validation.**

To validate the model, the percentage of error will be calculated using the formula:

$$Percentage\ of\ Error = \frac{Actual\ Value - Predicted\ Value}{Actual\ Value} \times 100$$

The percentage of error for the model ranges from 3.33% to 6%. The prediction model was validated using a separate dataset of 20% of total data points. The percentage of error was calculated for each prediction, showing an average error of less than 6%. This shows that the model is exceedingly accurate in predicting fuel consumption based on the given variables. The graph comparing the actual fuel consumption to the predicted fuel consumption as shown in Fig. 3.

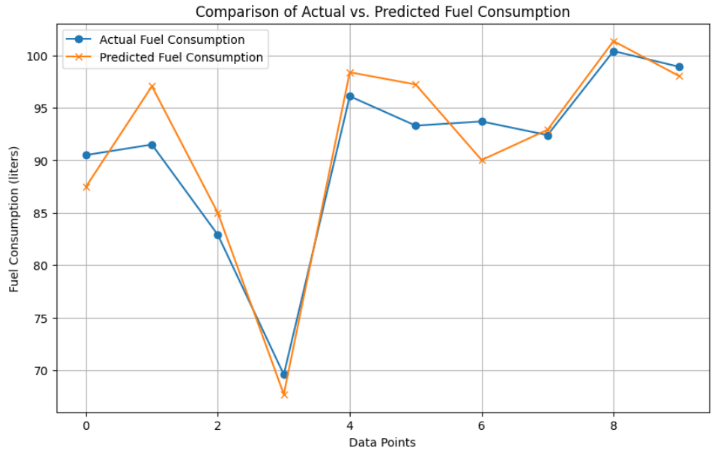


**Fig. 3. Graph comparing the actual fuel consumption vs. predicted fuel consumption.**

In Fig. 4, only partial connections between nodes are shown. However, in the actual implementation, there should be connections in all possible directions between all nodes. The route nodes for waste collection can be visualized in Fig. 4 below:

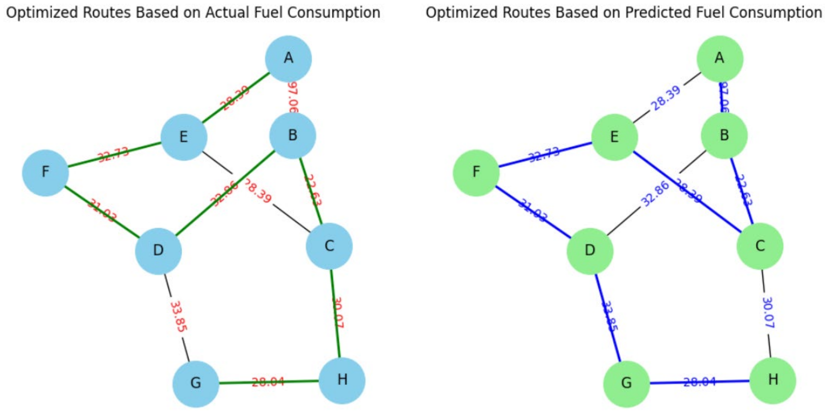


**Fig. 4. Graph comparing optimized routes based on predicted fuel consumption.**

The graph compares optimized waste collection routes based on actual and predicted fuel consumption. On the left, the route optimized using actual fuel data is shown with green-highlighted edges, resulting in a total fuel consumption of 209.98 litters. On the right, the route optimized using predicted fuel data is depicted with blue-highlighted edges, achieving a lower total fuel consumption of 205.75 litters. The experiment’s comparison shows there is significant impact on the result path depending on the variables considered in the algorithm and has saved fuel of 4.23 litters using fuel prediction with consideration of distance, congestion, and waste weight.

This comparison demonstrates the predictive model's effectiveness in closely approximating actual fuel usage, highlighting its potential for improving route optimization and enhancing fuel efficiency in waste collection operations. The ability of the predictive model to recommend a more fuel-efficient route underscores its practical value in reducing operational costs and environmental impact.

## 4. Discussion

This research highlights the significance of this study in the context of waste management. By integrating a multi-linear regression model with Dijkstra's algorithm, the research addresses the inefficiencies in current waste collection systems. The predictive model accurately estimates fuel consumption, enabling the optimization of waste collection routes based on fuel efficiency. The validation of the model through minimal error percentages underscores its reliability.

This study's contribution lies in its innovative approach to waste management, combining statistical modelling with route optimization to achieve cost savings and environmental benefits. The findings demonstrate a potential reduction in fuel consumption, emphasizing the study's practical implications for waste management operations.

## 5. Conclusion

In conclusion, this study presents a comprehensive methodology for predicting fuel consumption in waste collection vehicles and integrating this prediction into a route optimization algorithm. The predictive model, based on a multiple linear regression equation, uses distance, road congestion, and waste weight as coefficient variables. The collection of data, involving historical information from waste collection systems and various operational scenarios, ensures the model's accuracy and reliability.

Route optimization is achieved through Dijkstra's algorithm, prioritizing stops with lower predicted fuel consumption to enhance overall fuel efficiency and cost savings. The integration of fuel consumption predictions into the algorithm contributes to the sustainability and efficiency of waste collection operations.

In essence, the methodology presented in this study gives a strong framework for enhancing the efficiency and sustainability of waste collection operations by leveraging predictive modelling and route optimization techniques. The combination of linear regression and Dijkstra's algorithm offers a powerful approach to prioritize stops and minimize fuel consumption, contributing to overall operational effectiveness and environmental sustainability.

Future work will focus on incorporating real-time traffic data and expanding the model to other types of waste collection scenarios. The integration of predictive modelling and route optimization offers a robust framework for enhancing the efficiency and sustainability of waste collection operations, addressing current challenges, and contributing to overall operational effectiveness and environmental sustainability.

## Nomenclatures

| | |
|---|---|
| $R$ | Correlation Coefficient |
| $R^2$ | Coefficient of Determination (R-squared) |

***Greek Symbols***

| | |
|---|---|
| $\beta$ | Regression Coefficients for the independent variables in the MLR model |
| $\epsilon$ | Error term in the regression model |

**Abbreviations**

| | |
|---|---|
| GST | Goods and Services Tax |
| KPSS | Public Personnel Selection Exam |
| MLR | Multiple Linear Regression |
| OLS | Ordinary Least Squares |
| SVM | Support Vector Machines |
| SVR | Support Vector Regression |

## Acknowledgments

This research was funded by SDEC's Translational Research Grant. The authors would like to thank SDEC for their support and contribution to this study.

## References

1. Sreenivasan, J.; Govindan, M.; Chinnasami, M.; and Kadiresu, I. (2012). *Solid waste management in Malaysia-A move towards sustainability*. In Marmolejo Rebellon, L.F. (Ed.), *Waste management-An integrated vision*, InTech.
2. He, B. (2022). Application of Dijkstra algorithm in finding the shortest path. *Journal of Physics*: *Conference Series*, 2181(1), 012005.
3. Gunawan, R.D.; Napianto, R.; Borman, R.I.; and Hanifah, I. (2022). Implementation of Dijkstra's algorithm in determining the shortest path (Case study: Specialist doctor search In Bandar Lampung). *International Journal Information System and Computer Science* (*IJISCS*), 3(3), 98-106.
4. Hossain, M.A.; Ahmedy, I.; Harith, M.Z.M.Z.; Idris, M.Y.I.; Soon, T.K.; Noor, R.M.; and Yusoff, S.B. (2020). Route optimization by using Dijkstra's algorithm for the waste management system. *Proceedings of the* 2020 *The* 3rd *International Conference on Information Science and System*, Cambridge United Kingdom, 110-114.
5. Hannan, M.A.; Begum, R.A.; Al-Shetwi, A.Q.; Ker, P.J.; Al-Mamun, M.A.; Hussain, A.; Basri, H.; and Mahlia, T.M.I. (2020). Waste collection route optimisation model for linking cost saving and emission reduction to achieve sustainable development goals. *Sustainable Cities and Society*, 62, 102393.

6. Wang, L.; Kelly, K.; Walkowicz, K.; and Duran, A. (2015). Quantitative effects of vehicle parameters on fuel consumption for heavy-duty vehicle. *Proceedings of the SAE* 2015 *Commercial Vehicle Engineering Congress*, Rosemont, Illinois.
7. Çapraz, A.G.; Özel, P.; Şevkli, M.; and Beyca, Ö.F. (2016). Fuel consumption models applied to automobiles using real-time data: A comparison of statistical models. *Procedia Computer Science*, 83, 774-781.
8. Ozili, P.K. (2022). The acceptable R-square in empirical modelling for social science research. *SSRN Electronic Journal*, MPRA Paper No. 115769.
9. Uyanik, G.K.; and Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, 106, 234-240.
10. Kushwaha, M.; and Abirami, M.S. (2022). *Comparative analysis on the prediction of road accident severity using machine learning algorithms*. In: Sharma, D.K. et al. (Eds.). *Micro-Electronics and Telecommunication Engineering*. Springer Nature Singapore, 373, 269-280.