University of Bahrain

College of Science

Department of
Postgraduate Programs

Prediction of Domestic Solid Waste Generation During Ramadan Using Deep Machine Learning Approaches

Thesis submitted inpartial fulfilment of the requirements for the degree of M.Sc. In Big Data Science and Analytics

Prepared by:

Husain Yusuf Hani

20166225

Supervised by:

Dr. Majeed Safar Jasim

Associate Professor

University of Bahrain

Kingdom of Bahrain

March 2023

University of Bahrain

College of Science

Department of Postgraduate Programs



Prediction of Domestic Solid Waste Generation During Ramadan Using Deep Machine Learning Approaches

Thesis submitted in partial fulfilment of the requirements for the degree of M.Sc. In Big Data Science and Analytics

Prepared by:

Husain Yusuf Hani

20166225

Supervised by:

Dr. Majeed Safar Jasim

Associate Professor

University of Bahrain

Kingdom of Bahrain

March 2023

Defense Committee's Approval

University of Bahrain
Deanship of Graduate Studies
And scientific Research

جامعة البحرين
عمادة الدراسات العليا والبحث العلمي

Thesis Discussion Committee Decision

Student Name: Husain Yusuf Hani	Student ID: 20166225
Department: Choose an item.	College: **Science**
Program: **M.Sc. in Big Data Science and Analytics**	
Thesis Title: Prediction of Domestic Solid Waste Generation During Ramadan Using Deep Machine Learning Approaches	
Supervisor Name: **Dr. Majeed Safar**	Student GPA: **3.89**

The Discussion Committee met on 24-May-23 In S41-1002. Upon discussing the student whose details are stated above in public, the discussion committee recommends the following:

The student is awarded a grade Choose an item. in the thesis subject.

☐ Without making any changes to the thesis.

☒ After making the modifications attached with this decision.

Discussion Committee Members Approval:

Name		Signature
1. **Dr. Majeed Safar**	(Advisor)	Majeed
2. **Prof. Nabil Hewahi**	(Internal Examiner)	Nabil Hewahi
3. **Dr. Saeed Sharif**	(External Examiner)	SaeedSharif

Approved at the University Council 6th meeting, 2013
In accordance with Decision No. 736/2013 dated 14/3/2013
Modified by University Council Decision No. 2043/2013 dated 20/11/2013.

Abstract

Excess domestic waste generation during religious festivals has adverse environmental impacts. The holy month of Ramadan provides opportunity to study relationship between food consumption pattern and waste generation. Therefore, the objective of this research is to investigate the relationship between Ramadan fasting and the generation of domestic solid waste in Bahrain for the period between 2019-2022. Furthermore, this study develops accurate time-series models using deep machine learning approaches to predict domestic waste generation for pre-, during, and post-Ramadan periods. The research findings revealed that during the month of Ramadan, the Bahraini's generates the highest amount of waste, with an average of 1,570 tons per day, compared to 1,450 tons per day during the rest of the year. The study conducted a comparative analysis of three machine learning models: Long Short-Term Memory (LSTM) Recurrent Neural Networks, Bidirectional LSTM (BILSTM) Recurrent Neural Networks, and Gated Recurrent Unit (GRU), along with three statistical models: Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Prophet. The prediction of BILSTM model outperformed the others with MAE value of 28.120, RMSE value of 36.879, MAPE value of 0.020, and R^2 value of 0.78.

Acknowledgements

Thank God for guiding me to complete this project successfully. I would like to express my sincere appreciation to the many people who have supported me on this journey.

First and foremost, I am grateful to my supervisor, Dr. Majeed Safar, for his invaluable guidance, support, and encouragement throughout this research. I feel privileged and fortunate to finally have Dr. Majeed as my project supervisor. His vast experience and inspiring personality always motivated me to do my best. I would also like to express my gratitude to Dr. Gulnur Coskuner for her guidance and great support and encouragement during this research. I would also like to express my gratitude to Dr. Sawsan Hilal for her encouragement and unwavering support.

I would like to thank my supervisor and co-workers for their continued support and motivation throughout this journey. Additionally, I would like to thank the Ministry of Municipal Affairs and Agriculture for their generosity in sharing their data, allowing me to conduct this study. Their contribution was crucial to the success of my research, and I am deeply grateful for their willingness to collaborate.

I would also like to express my sincere appreciation to my parents for their unwavering love, prayers, and patience. Finally, I would like to thank my friends for their continued encouragement and support.

Table of Contents

Defense Committee's Approval ..... A

Abstract ..... B

Acknowledgements C

Table of Contents ..... D

List of Tables ..... G

List of Figures ..... H

List of Abbreviations ..... J

Chapter One Introduction 2

1.1. Holy Month of Ramadan 2

1.1.1. Food Waste and Hunger 3

1.1.2. Environmental Impacts of Domestic Waste 4

.1.1.3 Economic impacts of Domestic Waste 6

1.2. Research Problem 6

1.3. Research Questions 7

1.4. Research Objectives 8

1.5. Research Significance 8

1.6. Theses Structure 10

Chapter Two Literature Review 12

2.1. Introduction 12

2.1.1. Waste Generation 12

2.2. Effects of Religious Events on Waste Generation 14

2.3. Application Domains in Machine Learning 15

2.4. Solid Waste Generation Prediction Models 17

2.4.1. Artificial Neural Networks (ANN) 17

2.4.2. Support Vector Machines (SVM) 18

2.4.3. Linear Regression Analysis (LR) 19
2.4.4. Decision Trees (DT) 19
2.4.5. Genetic Algorithm (GA) 20
2.5. Solid Wastes Generation Time-series Forecasting Models 25
2.6. Implemented Methods 27
2.6.1. Autoregressive Integrated Moving Average (ARIMA) 28
2.6.2. Seasonal Autoregressive Integrated Moving Average (SARIMA) 31
2.6.3. Prophet Algorithm 32
2.6.4. Long Short-Term Memory (LSTM) Recurrent Neural Networks 33
2.6.5. Bidirectional LSTM (BILSTM) Recurrent Neural Networks 36
2.6.6. Gated Recurrent Unit (GRU) 37
2.7. Research Gaps 38
Chapter Three Methods and Procedures 41
3.1. Research Design 41
3.2. Research Data Sources 42
3.3. Data Description 43
3.4. Data Pre-Processing 43
3.5. Training Approaches 45
3.5.1. ARIMA and SARIMA Training Approach 45
3.5.2. Prophet Training Approach 46
3.5.3. LSTM, BILSTM, and GRU Training Approach 47
3.6. Evaluation Metrics 48
3.6.1. Mean Absolute Error (MAE) 49
3.6.2. Root Mean Square Error (RMSE) 49
3.6.3. Mean Absolute Percentage Error (MAPE) 50
3.6.4. Coefficient of Determination (R^2) 50
Chapter Four Results, Discussion and Conclusion 53

4.1. Results 53
4.1.1. Descriptive Statistic 53
4.1.2. Statistical Analysis 56
4.1.3. ARIMA Model Performance 61
4.1.4. SARIMA Model Performance 67
4.1.5. Prophet Model Performance 70
4.1.6. LSTM Model Performance 72
4.1.7. BILSTM Model Performance 74
4.1.8. GRU Model Performance 76
4.1.9. Overall Models Performance Results 78
4.2. Discussion 80
Chapter Five Conclusion, Recommendations and Future Work 86
5.1. Conclusion 86
5.2. Recommendations 88
5.3. Future Work 89
References 90
Appendices 99
الملخص 101

List of Tables

Table (1): Selected food waste generation from Arabic countries 3
Table (2): Analyses of ML models in the SWM literature 21
Table (3): Domestic waste descriptive statistics from 2019 to 2022 53
Table (4): Domestic waste descriptive statistics for the dataset 55
Table (5): Results of the Dickey-Fuller test 61
Table (6): Results of the AUTOARIMA function 62
Table (7): ARIMA model summary 63
Table (8): ARIMA model prediction performance 65
Table (9): SARIMA model prediction performance 69
Table (10): MAE for different changepoint prior scale using Prophet 70
Table (11): Prophet model prediction performance 71
Table (12): LSTM hyperparameters 72
Table (13): LSTM model prediction performance 73
Table (14): BILSTM hyperparameters 74
Table (15): BILSTM model prediction performance 75
Table (16): GRU hyperparameters 76
Table (17): GRU model prediction performance 77
Table (18): Overall models performance results 78
Table (19): Waste type quantity (tons) 99

List of Figures

Fig. 1: Projected waste generation, by region 13
Fig. 2: The percentage of global waste composition 13
Fig. 3: Number of studies according to publication, year, and application fields 15
Fig. 4: Percentage of publications by ML model type 17
Fig. 5: Flow chart for ARIMA model selection 30
Fig. 6: The addition of seasonal orders 31
Fig. 7: illustration of the architecture of an LSTM 33
Fig. 8: LSTM hyperparameter tuning 35
Fig. 9: The process of bidirectional LSTM 36
Fig. 10: The difference between GRU and LSTM 38
Fig. 11: Illustration of research methodology 41
Fig. 12: Aggregation of the daily generation of waste 44
Fig. 13: The sliding windows algorithm 45
Fig. 14: Domestic waste frequency distribution for each individual year 54
Fig. 15: Domestic waste frequency distribution for the dataset 55
Fig. 16: Waste type quantity (percent) 56
Fig. 17: The trend of domestic waste for each individual year 57
Fig. 18: Average domestic waste per day in each month 57
Fig. 19: Average quantities of waste generated during the period pre, during, and post-Ramadan 58
Fig. 20: The daily waste generation pre, during, and post Ramadan 60
Fig. 21: Autocorrelation function (ACF), and partial autocorrelation function (PACF) plots 62
Fig. 22: Prediction of daily domestic waste generation using ARIMA 64
Fig. 23: Prediction of daily domestic waste generation pre, during and post-Ramadan using ARIMA 64
Fig. 24: ARIMA diagnostics plot 66
Fig. 25: Seasonal decomposition plots 68
Fig. 26: Prediction of daily domestic waste generation using SARIMA 69
Fig. 27: Prediction of daily domestic waste generation pre, during and post-Ramadan using SARIMA 69

Fig. 28: Prediction of daily domestic waste generation using Prophet 71
Fig. 29: Prediction of daily domestic waste generation pre, during and post-Ramadan using Prophet 71
Fig. 30: LSTM model fitting and Test performance 72
Fig. 31: Prediction of daily domestic waste generation using LSTM 73
Fig. 32: Prediction of daily domestic waste generation pre, during and post-Ramadan using LSTM 73
Fig. 33: BILSTM model fitting and test performance 74
Fig. 34: Prediction of daily domestic waste generation using BILSTM 75
Fig. 35: Prediction of daily domestic waste generation pre, during and post-Ramadan using BILSTM 75
Fig. 36: GRU model fitting and validation performance 76
Fig. 37: Prediction of daily domestic waste generation using GRU 77
Fig. 38: Prediction of daily domestic waste generation pre, during and post-Ramadan using GRU 77
Fig. 39: Results of Dickey-Fuller test 99
Fig. 40: ARIMA results for order (7,0,7) 100

List of Abbreviations

Abbreviation	Definition
ARIMA	Autoregressive Integrated Moving Average
BILSTM	Bidirectional Long Short-Term Memory
FAO	Food And Agriculture Organization
GHG	Greenhouses Gasses
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MSW	Municipal Solid Waste
SARIMA	Seasonal Autoregressive Integrated Moving Average
SCE	Supreme Council for The Environment
SWG	Solid Waste Generation

Chapter One

Introduction

1.1. Holy Month of Ramadan

1.2. Research Problem

1.3. Research Questions

1.4. Research Objectives

1.5. Research Significance

1.6. Theses Structure

Chapter One
Introduction

1.1. Holy Month of Ramadan

Ramadan (رمضان) is the ninth month in the Hijri calendar, after the month of Shaban (شعبان). It holds a special significance for the Muslims and is considered distinct from other months in the Hijri calendar. Ramadan is observed as the month of fasting, which is one of the essential practices in Islam and one of the five pillars of Islam. During this month, the Muslims abstain from consuming food and drinks, as well as refrain from certain activities that can break their fast, from dawn until sunset (SAKK, 1975).

Unfortunately, some people view Ramadan as a time for feasting rather than fasting. As a result, they tend to overcompensate for their daytime fasts by indulging in excessive eating and drinking after sunset. This shift in behavior can detract from the spiritual and physical benefits of the holy month, turning it into a month centered on gluttony rather than worship and well-being (Al Sheikh, 2018).

Regrettably, it's unfortunate that a significant amount is wasted due to social activities such as the high number of invitations and the increase in feasts during this holy month. This monthly annual event has significant implications for SWG, and hence managing solid waste has become a crucial environmental challenge (Alzamora et al., 2022).

1.1.1. Food Waste and Hunger

According to the Food and Agriculture Organization of the United Nations (FAO) (2018), food waste and hunger are interconnected issues that majorly impact the nexus of the environment, economy, and human health. When food is wasted, it not only results in environmental degradation but also exacerbates hunger and malnutrition by diverting resources away from food-insecure populations. According to the World Bank (2018), approximately one-third of worldwide food is lost or wasted on annual basis. This represents a significant amount of food that could have been used to feed millions of people who suffer from hunger and malnutrition. Table 1 shows the amount of domestic waste generated per person in the Arab countries. The city of Karbala and Bahrain exhibit the highest food waste generation.

Table (1): Selected food waste generation from Arabic countries

Country	Reference	Study area	Food waste (kg/capita)
Bahrain	(Alayam, 2018)	Nationwide	132
Saudi Arabia	(SAGO, 2019)	Nationwide	105
Lebanon	(Chalak et al., 2019)	Beirut	105
Iraq	(Al-Rawi & Al-Tayyar, 2013)	Mosul	85
	(Al-Mas'udi & Al-Haydari, 2015)	Karbala	142
	(Al-Maliky & ElKhayat, 2012)	Baghdad	75

Food waste not only results in the loss of food but also leads to the squandering of resources that were utilized in producing, transporting, and storing processes. As a result, food waste exacerbates people's difficulties in accessing the necessary sustenance to survive (Gustavsson et al., 2011). This is a crucial issue, especially in developing countries where food security is already a major challenge. The impact of food waste perpetuates the cycle of poverty and hunger, which creates

inequality and deprivation in these countries. Therefore, reducing food waste is essential to tackle the challenge of hunger and malnutrition, providing nutritious food to those who urgently need it, and creating a more sustainable and equitable world (FAO, 2019).

In addition to taking practical steps to reduce food waste and increase access to nutritious foods, raising awareness of the relationship between food waste and hunger is essential, hence promoting policies and initiatives that comprehensively address these issues is also necessary (Gustavsson et al., 2011).

1.1.2. Environmental Impacts of Domestic Waste

Domestic waste, consisting mainly of food waste, has a considerable detrimental effect on the environment. Food waste is a significant contributor to climate change, land degradation, and water scarcity, which is a cause for concern. Wasting food leads to squandering all the resources invested in its production, such as water, land, energy, and labor. Food production requires a significant amount of water, and when food is wasted, it exacerbates the problem of water scarcity. Studies indicate that a third of the water utilized in agriculture is lost due to food loss (NRDC, 2017).

The Ministry of Climate Change and Environment (MOCCAE) in the UAE has reported that during Ramadan, there is a 50 percent surge in food waste compared to other months. Based on the findings of a landfill site survey, it was revealed that food waste makes up 38 percent of the total waste generated in the country, which amounts to nearly two million tons annually (MOCCAE, 2014).

Once food waste is disposed of in landfills, it undergoes anaerobic decomposition, breaking down in an environment without oxygen and generating methane. This potent Greenhouse Gas (GHG) has a global

warming potential 28 times higher than carbon dioxide. Food waste is responsible for around 8% of the world's GHG emissions. Aside from its impact on climate change, food waste also contributes to land degradation by occupying landfill space and releasing harmful chemicals into the soil (FAO, 2013).

Plastic waste is a common component of domestic waste and contributes to several environmental problems. Due to its slow decomposition rate, plastic waste accumulates in landfills and natural surroundings, leading to visual pollution and endangering wildlife. The environmental impact of plastic waste is a multifaceted issue that demands immediate attention and action to be resolved. When plastic waste is not disposed of properly, it can find its way into waterways and oceans, causing harm to marine life and disrupting ecosystems. The contamination of seafood by plastic particles also poses a risk to human health. Additionally, the disposal of plastic waste consumes substantial resources and energy, leading to further environmental impacts (Lebreton, et al., 2017).

Reducing domestic waste is crucial to minimize its adverse environmental effects. Effective strategies such as composting, recycling, and improving food management practices can aid in decreasing the quantity of food waste sent to landfills. This approach not only minimizes GHG emissions and preserves natural resources but also contributes to building a sustainable and robust environment for both present and future generations (UNEP, 2021).

1.1.3. Economic impacts of Domestic Waste

Food waste has a significant economic impact on different levels, from households to national economies. The World Bank report (2018) reveals that the costs of managing wasted food are equal to the value of the food itself. Food waste greatly influences household budgets, as the money spent on purchasing and preparing is wasted. For low-income families, food waste makes up a large portion of their food budget. Additionally, retail businesses lose profits when stockpiling perishable and expired items (Buzby et al., 2014). The World Bank report estimates that food waste costs the global economy $1.2 trillion each year.

Food waste can have a wider economic impact, affecting employment and economic growth. Waste food decreases the demand for labour in agriculture and food processing, leading to possible job losses. As a result, the food supply decreases, which could lead to price increases and impact economic growth in countries that rely heavily on agriculture. Hence, reducing food waste has not only environmental and social benefits but also significant economic advantages (WRAP, 2019).

1.2. Research Problem

The Supreme Council for the Environment (SCE) in Bahrain has reported a significant issue of food waste, with approximately 400 tons of organic food waste generated daily, increasing to nearly 600 tons during the month of Ramadan (ClimateFaith, 2017). This puts a strain on sanitation services and has negative environmental impacts. The SCE estimates that during Ramadan, between 25-30 percent of purchased food is wasted, amounting to roughly 0.3 kg per person per day, a 50 percent increase compared to non-Ramadan periods.

Food waste in Bahrain contributes to approximately 2,700 tons of greenhouse gas emissions per day and affects various aspects, including social, economic, cultural, humanitarian, and religious matters (CFB, 2017). The generation of solid waste, particularly during festivals and religious months, poses a significant environmental challenge. Accurate prediction of domestic solid waste generation is crucial for effective waste management systems.

The aim of this research is to investigate the relationship between Ramadan fasting and domestic solid waste generation in Bahrain from 2019 to 2022 and develop accurate time-series models using deep machine learning approaches to predict domestic solid waste generation during holy month of Ramadan. To address the complex nature of waste generation, the following models will be used: Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet algorithm, Long short-term memory (LSTM), Bidirectional Long short-term memory (BILSTM), and Gated Recurrent Unit (GRU).

1.3. Research Questions

This study aims to address the following research questions:

- Can machine learning models accurately and efficiently predict domestic waste generation in Bahrain during different periods, including pre, during, and post-Ramadan?
- Which machine learning model demonstrates the highest effectiveness and optimal performance in predicting domestic waste generation in Bahrain?
- What is the extent of the relationship between Ramadan fasting and the generation of domestic waste in Bahrain?

1.4. Research Objectives

The main objectives of this research are as follows:

- To investigate the relationship between the fasting of Ramadan and the generation of domestic solid waste in Bahrain.
- To develop time-series models using deep machine learning approaches to accurately predict domestic solid waste generation (SWG) pre, during, and post-Ramadan periods.
- To compare the performance of the developed models in predicting the generation of domestic solid waste and to know which models are the most efficient.
- To analyze daily domestic SWG in the Kingdom of Bahrain during the specified three-month period in the Islamic calendar for the period 2019-2022.
- To provide recommendations for waste-related institutes to operate effective domestic SWM systems during the fasting period of Ramadan and other religious festivals.

1.5. Research Significance

The importance of this study significance is due to the use of deep machine learning approaches in Bahrain as a case study because there are a limited number of investigations about waste management. By applying these advanced techniques to model SWG during the holy month of Ramadan, this study can lead to the development of innovative and effective approaches to SWM in Bahrain. The outcomes of this study can also provide valuable insights to policymakers and waste-related organizations in Bahrain on how to design and operate effective waste management systems during religious festivals.

The research's impact on sustainable waste management practices extends beyond Bahrain and is applicable to other regions grappling with similar challenges. This is particularly relevant for Arab countries that consistently experience an increase in waste generation during the holy month of Ramadan. The designed predictive model can assist these countries in forecasting waste generation, enabling them to proactively plan and develop effective waste management systems. By leveraging the model's predictions, they can adequately prepare for the surge in waste and implement efficient waste management strategies tailored to their specific needs.

The environmental significance of this study is rooted in the fact that SWG is a major environmental challenge, and religious practices can exacerbate this challenge. The improper management of domestic waste has a significant negative impact on the environment, contributing to climate change, land degradation, and water scarcity. Wasting food squanders the resources invested in its production, exacerbates water scarcity issues, and accounts for a substantial portion of greenhouse gas emissions. Disposing of food waste in landfills leads to the production of methane, a potent greenhouse gas. Furthermore, plastic waste, another component of domestic waste, causes visual pollution, endangers wildlife, and poses risks to marine ecosystems and human health.

The study investigates waste generation patterns during religious festivals, providing valuable insights about behaviour patterns. By identifying the specific factors that contribute to increased waste production, such as cultural practices, consumption behaviour, it becomes possible to design waste management practices that cater to these unique circumstances. The study's findings can guide the development of targeted strategies for waste reduction, recycling, and proper disposal during religious celebrations, with the goal of mitigating the

environmental impact associated with waste generation. These strategies may involve implementing awareness campaigns, encouraging responsible consumption, and enhancing waste infrastructure and collection systems. By adopting sustainable waste management practices, it might be possible to minimize the environmental consequences of waste generation during religious festivals.

1.6. Theses Structure

The thesis includes a comprehensive review of the existing literature about time-series forecasting in the second chapter. The third chapter outlines the methodology used to conduct various experiments, while the fourth chapter provides a detailed discussion of the results obtained, along with conclusions and recommendations.

Chapter Two

Literature Review

2.1. Introduction

2.2. Effects of Religious Events on Waste Generation

2.3. Application Domains in Machine Learning

2.4. Solid Waste Generation Prediction Models

2.5. Solid Waste Generation Time-series Forecasting Models

2.6. Implemented Methods

2.7. Research Gaps

Chapter Two
Literature Review

2.1. Introduction

This chapter in detail reviews the literature related to this research. First, the concept of SWM, its impact and its challenges are explained. After that, the focus was on studies of SWM in religious events, especially during Ramadan, and their impact on waste generation. Next, the applications of different machine learning methods applied in SWM are discussed. Finally, literature involving prediction methods and time-series forecasting methods for SWG are discussed.

2.1.1. Waste Generation

Waste Generation and identification are essential aspects of the SWM system. At the beginning of each process, waste is generated as raw materials are converted into commodities for consumption. Based on the source of waste generation, the quantity and characteristics of the waste can be determined, as shown in Table 2. According to the World Bank (2018), global waste is expected to grow to 3.40 billion tons by 2050, as shown in Fig. 1.

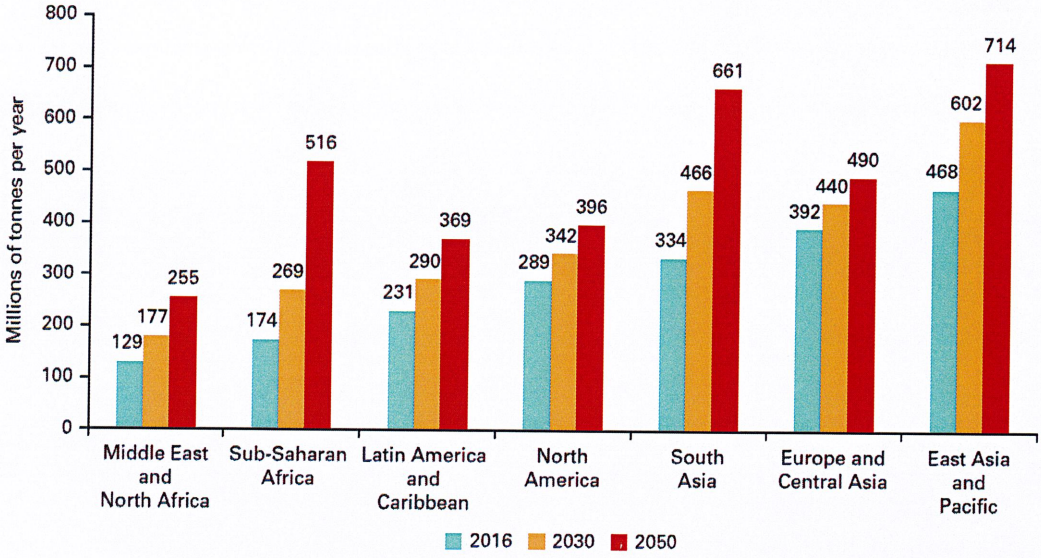


Fig. 1: Projected waste generation, by region (World Bank, 2018)

Reports and studies indicate that the generation of domestic solid waste globally exceeds two billion tons annually due to many factors, such as the rapid growth of population, development of the economy, and improvement of the standard of living, which may represent a threat to environmental destruction. According to World Bank (2018), all regions generate about 50 percent or more organic waste, as shown in Fig. 2. In addition, it was found that the average per capita waste generation in developed countries is usually 521.95 - 759.2 kg per year and 109.5 - 525.6 kg per year by developing countries (Karak et al., 2012).

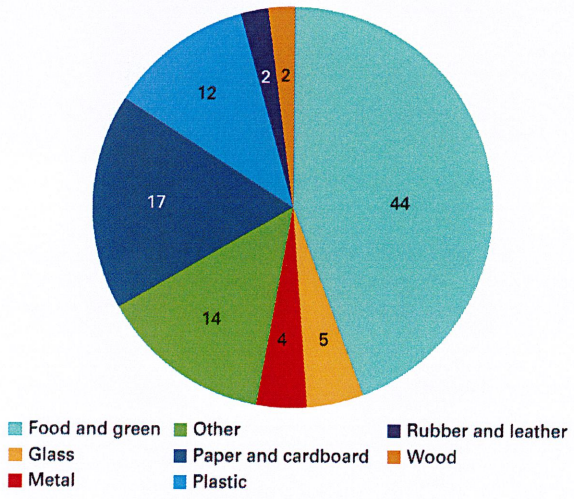


Fig. 2: The percentage of global waste composition (World Bank, 2018)

Moreover, the impact of tourism activities on SWG in Madeira was studied in the period from 1996-2018, where it was found that tourism activities are responsible for 41.9% to 46.6% of SWG per resident (Martins & Cró, 2021).

2.2. Effects of Religious Events on Waste Generation

Religious events of spiritual significance are spread worldwide, such as Hajj, Ramadan and Arbaeen. These events are attended by approximately 300 million individuals in religious events around the world on an annual basis. Therefore, huge amounts of MSW are generated, which increases the potential for serious negative environmental impacts (Abdulredha et al., 2020).

Large religious events significantly increase the amount of solid waste generated. According to the authority's estimates, 48 kilotons of MSW were generated during Arbaeen in 2015, and 3.69 million pilgrims generated 17 kilotons over five days during the Hajj event (Alsebaei, 2014). It was found that the highest percentage of waste was for organic waste with 56%, followed by plastic waste with 15%.

In Saudi Arabia, the holy city of Mecca hosts millions of Muslim pilgrims every year. As a result, huge amounts of MSW are generated. The city averages about 2.4 thousand tons of solid waste per day. It was found that the quantities during Ramadan and Hajj are higher than the average and reach 3.1 and 4.6 thousand tons per day, respectively (Nizami, et al., 2017).

Food consumption and waste during the Ramadan were studied and it was found that in some Arab countries such as the United Arab Emirates, Saudi Arabia, and Egypt, spending on food and groceries increased by 56%. In addition, the most critical wasted food products are

cooked meals (66.6%), bread (46.0%) and fruits (31.7%) (Ben Ismail et al., 2022). Ali-Arous et al. (2017) evaluated the effects of food-related behaviour, food management, barriers, and willingness to change behaviour. It was found that 88% of people change their attitudes towards wasting food, especially during Ramadan. The highest wasted food category was also found to be fruits, vegetables, grains and bakery products. In addition, about 15% of households throw away at least 250 grams of food that is still edible each week.

2.3. Application Domains in Machine Learning

Analysis of studies in SWM by Abdullah et al. (2020) revealed six areas in which artificial intelligence is applied. These applications include the areas of (1) waste bin level detection, (2) prediction of waste characteristics, (3) prediction of process parameters, (4) prediction of the process output, (5) vehicle routing, and (6) SWM planning. Fig. 3 shows the number of publications conducted in the period (2004-2019) for the six fields. Recently, it has been noted that artificial intelligence research in SWM has increased; Because of its effectiveness in solving problems, interest in the field of artificial intelligence has increased.

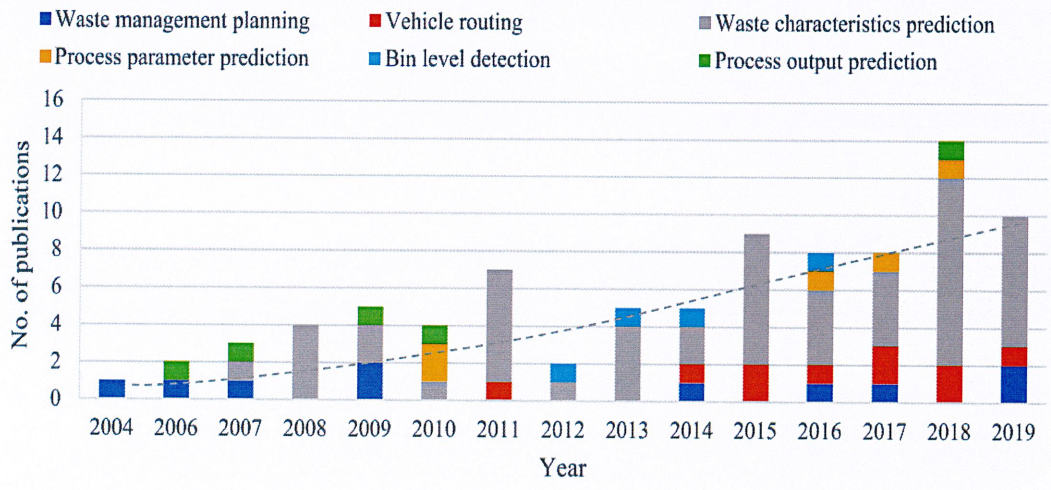


Fig. 3: Number of studies according to publication, year, and application fields (Abdullah et al., 2020)

Bin level detection models are employed to predict waste container fill levels, preventing overflow and ensuring timely emptying. According to Hannan et al. (2012), MLP classifier achieved a high accuracy of 98.98% for bin-level classification and 90.19% for class and grade classification.

Accurate prediction of waste characteristics is crucial for Municipal Solid Waste Management (MSWM) and is influenced by various social, economic, and environmental factors. AI studies in SWM predominantly focus on predicting solid waste characteristics, with ANNs being the most used approach, followed by SVMs (Abdallah et al., 2020).

ANN models have proven effective in predicting various aspects of waste management. For instance, Ozkaya et al. (2007) utilized ANN to predict biogas generation from bio-waste landfills, achieving an impressive 99.8% accuracy with an MSE of 0.0025. Qdais et al. (2010) applied ANN and GA to improve biogas production in a Jordanian biogas plant, resulting in a correlation coefficient of 0.87 and an MSE of $6x10^{-5}$.

Optimizing waste collection routing models has been effective in reducing vehicle emissions and traffic congestion. The collection process accounts for 85% of total SWM costs (Bi et al., 2022). Studies utilizing Genetic Algorithms (GA) have demonstrated success in this area. For example, Kroll et al. (2016) integrated GA with modern communication channels to optimize route length and vehicle count, resulting in reduced collection costs but longer average service times. These findings highlight the effectiveness of GA-based approaches in optimizing waste collection routing.

2.4. Solid Waste Generation Prediction Models

The literature review indicated that the most frequently used machine learning systems for waste generation prediction are ANN, SVM, LR, DT and GA as shown in Fig. 4 (Abdullah et al., 2020).

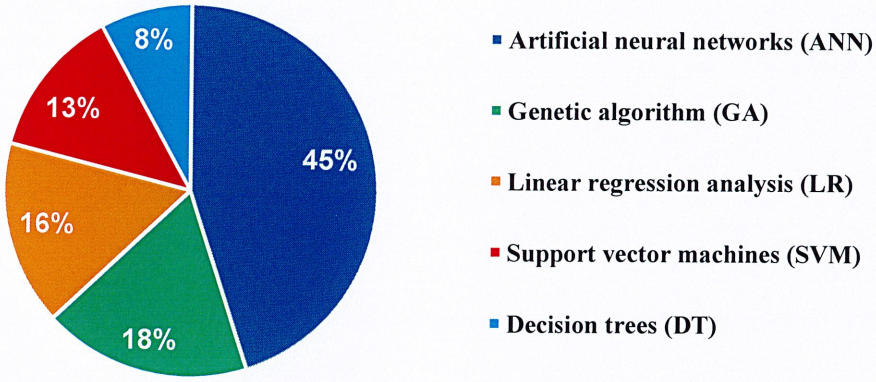


Fig. 4: Percentage of publications by ML model type (Abdullah et al., 2020)

ANN was the most commonly repeated machine learning method, with 45%. GA is another widely used algorithm, followed by LR, SVM, and DT.

2.4.1. Artificial Neural Networks (ANN)

Predictions of SWG processes with a nonlinear behaviour require tackling complex and imprecise tasks requiring human intuition (Jassim et al., 2022). In addition, it cannot be accessible due to the presence of multiple variables. ANN are effective in simulating the interactions of the biological nervous system. The essential components of neural networks are an input layer, hidden layers, and an output layer. In predicting SWG, ANNs have been widely used for their ability to depict multiple relationships and their tolerance for errors (Noori et al., 2010).

Neural network models often provide higher results and better performance compared to other models, as shown in Table 2. Kumar et al.

(2018) studied the relationship between plastic waste generation rate and specific socioeconomic groups using ANN, SVM and RF. The prediction accuracy of the ANN (R^2 = 0.75) was better than that of the SVM (R^2 = 0.74) and the RF model (R^2 = 0.66). In addition, a study to estimate the annual amount of household packaging waste showed the ability of ANN to overcome MLR. The Predicted (R^2 = 0.98) value is higher for the ANN model than for the MLR (R^2 = 0.73).

On the other hand, ANN lacks advantages, such as determining the relative importance of multiple factors since it is impossible to know which input characteristics have the most significant effect on the output. In addition, ANN faces challenges in calculations requiring high accuracy (Kannangara et al., 2018).

2.4.2. Support Vector Machines (SVM)

SVM is a practical supervised machine learning algorithm for data analysis used for both classification and regression. The SVM algorithm aims to find a hyperplane in N dimensions to classify the data points. In addition, it relies on kernel functions to solve regression problems. Moreover, it has been found to outperform many classic regression techniques. Furthermore, it has been widely applied in container filling level prediction, waste generation, and waste classification processes (Noori et al., 2009).

Jasim et al. (2022) developed a model using ANN and SVR to predict annual MSW generation. The results showed that the RBF-SVR model achieved the best results with an R^2 value of 0.97% and AFE values of 4.82% compared to the MLP-ANN model with an R^2 value of 0.94. Another study predicted SWG using the BPNN and SVR algorithms. The BPNN method showed relative errors between 6% and

16%, while the relative error for SVR ranges between 14% and 21% (Oguz-Ekim, 2021).

2.4.3. Linear Regression Analysis (LR)

Linear regression (LR), or multiple linear regression (MLR), is a supervised learning method that uses variables to predict the outcome of a response. SWM models and prediction of SWG contain many parameters; therefore, MLR is more suitable for simulating these processes. LR is characterized by its simplicity, easy-to-interpret results, and low computational costs. However, this method is generally not suitable for modelling nonlinear data.

Chhay et al. (2018) investigated the socioeconomic factors influencing the MSWG using LR, ANN and GM. The ranking of the MSWG prediction model is ANN, followed by GM, followed by LR. In addition, MLR showed poor prediction performance. On the other hand, the ANN model has a higher predictive accuracy when predicting the average rate of MSWG (Azadi & Karimi-Jashni, 2016). However, this does not mean that LR does not give good results. Popli et al. (2021) showed the effects of reliable urban solid waste statistics using LR.

2.4.4. Decision Trees (DT)

DTs have been used to predict waste generation, compaction, and waste classification for their effectiveness in extracting rules from unfamiliar data. This technique is effective in extracting a set of rules from unknown data. Then, the values of the responses are predicted by learning the decision rules derived from the features. DT is characterized by its ability to process missing data and low computational costs. In addition, the results and patterns can be easily interpreted and detected, respectively. However, they suffer from overfitting data, as shown in Table 2. The prediction accuracy for RF was the worst, with RMSE =

11.22 g/c/d, MAPE = 43.57%, and R^2 = 0.66 (Kumar, Samadder, Kumar, & Singh, 2018).

A new study by Dunkel et al. (2022) used SVM, GB, and RF to determine the quantities of solid waste types. RF shows high accuracy in predicting municipal waste, with R^2 = 1 and MAPE = 4.31 and household waste, with R^2 = 1 and MAPE = 3.03.

2.4.5. Genetic Algorithm (GA)

GAs have been widely applied in SWG prediction, waste accumulation prediction, and facility location determination. A genetic algorithm provides solutions to research and optimization problems in machine learning by improving a set of explanations or hypotheses rather than enhancing a single solution. It is a methodology that solves unconstrained and constrained improvement problems based on natural selection. There are three main components of GA: interception, mutation, and selection of the fittest. The algorithm uses arrays of bits or characters (binary strings) to represent the chromosomes. Then, information is exchanged between the arrays in search of optimized solutions. Finally, the solutions are evaluated to determine the overall suitability of each solution. Its efficiency and ease of implementation characterize GA. However, GAs requires careful construction to get good results or else it will take its toll (Wu et al., 2015).

Shamshiry et al. (2014) used GA with ANN to predict SWG, and the results showed that the ANN validation rate for testing and training was 94%. Another Study by Oliveira et al. (2019) developed an ANN model using GA algorithms to estimate the annual amount of household packaging waste. The results showed higher values for the GA-ANN model with R^2=0.98.

Table (2): Analyses of ML models in the SWM literature

Study	Model	Objective	Key finding
(Dunkel et al., 2022)	SVM, GB, RF	Finding links between the social, economic and demographic variables that determine the quantities of SW types.	▪ RF are effective in predicting the rate of produced wastes. ▪ RF shows high accuracy in predicting municipal waste, with $R^2 = 1$ and MAPE = 4.31 and household waste, with $R^2 = 1$ and MAPE = 3.03
(Jassim et al., 2022)	ANN, SVR	Implementation of ML algorithms to predict annual MSWG.	▪ Results showed the RBF-SVR model with an R^2 value of 0.97% and AFE values of 4.82%. ▪ The efficacy of the MLP-ANN model was successful, with an R^2 value of 0.94.
(Oguz-Ekim, 2021)	BPNN, SVR	Forecasting SWG using ML algorithms for different countries.	▪ The BPNN method shows relative errors between 6% and 16%. ▪ The relative error of SVR is between 14% and 21%.
(Popli et al., 2021)	MLR	Develop a predictive model to generate reliable statistics for urban SW.	▪ The best model was achieved with an R^2 value of 0.999, an MAE value of 1886.40, and a MAPE value of 0.67.
(Milojkovic & Litovski, 2008)	ANN	Forecasting waste management of electrical and electronic products.	▪ For their architecture, they got $f(9) = 18.2274$ for a network with four hidden and output neurons.
(Golbaz et al., 2019)	MLR, ANN	Predicting hospital SWG.	▪ For MLR, average $R^2 = 0.68$–0.74, and average MSE = 0.009–0.023. ▪ For ANN, the average R^2 value was 0.82–0.86, and the average MSE value was 0.003–0.008.
(Jalili Ghazi Zade & Noori, 2008)	ANN	Prediction of MSWG.	▪ Results of R^2 and MARE in the neural network for the test were equal to 0.746 and 3.18%, respectively.

(Noori et al., 2009)	SVM, PCA	Prediction of MSW for design and programming MSWM system.	▪ SVM with 13 inputs have a better result than 6 PCs-SVM, with R^2 0.78 and 0.75, MAE 356.02 and 362.15, and MARE% 3.31 and 3.35, respectively.
(Noori et al., 2009)	MLR, ANN, PCA	Predict the SWG in Tehran.	▪ ANN model has better results in comparison with the PCA-MLR. ▪ R and AARE in the ANN model were equal to 0.837 and 4.4%, respectively. ▪ PCA-MLR model (R= 0.445, MARE= 6.6%)
(Noori et al., 2010)	ANN, PCA-ANN, GT-ANN	Develop an ANN model with a simple structure and sample accuracy for weekly SW prediction.	▪ Suitable ANN structures are founded with the first seven PCs (7PCs-ANN) and the first five important input variables based on GT results (5GT-ANN).
(Shamshiry et al., 2011)	ANN	Predicting the amount of SWG using ANN.	▪ R^2 and MARE in the neural network for the test have been achieved equal to 0.989 and 0.06, respectively.
(Shamshiry et al., 2014)	GA, ANN, RSM	Predict SWG and optimize waste collection and transportation costs.	▪ There is an increase in SWG from 7825.7 tons in 2009 to 8030.68 tons in 2011; the cost reduction amount is 10.64%. ▪ The validations of ANN for tests and training are 94%.
(Abbasi et al., 2014)	SVM, WT-SVM	Examine the uncertainty of models for the prediction of MSWG.	▪ WT-SVM had lower uncertainty during the training and test stages. ▪ WT-SVM shows better results for both training and testing for RME, RMSE, MARE and R2.
(Abbasi et al., 2013)	SVM, PLS-SVM	Predicting MSWG using SVM as an intelligence tool combined with partial least squares (PLS).	▪ PLS-SVM is superior to the SVM model in predictive ability and calculation time-saving. ▪ PLS successfully identifies the complex nonlinearity and correlations among input variables and minimizes them.

(Abbasi & El Hanandeh, 2016)	SVM, ANN, KNN, ANFIS	Develop a model for accurate predicting of MSW generation that helps waste-related organizations design and operate effective MSWM systems.	▪ ANFIS systems are the most accurate forecasts, while KNN is the most accurate in predicting waste quantities' monthly averages. ▪ Annual MSW generated in Logan City will reach $9.4x10^7$ kg by 2020.
(Abbasi et al., 2019)	ANFIS, ANN, RBF	predict monthly and seasonal MSWG using radial basis function (RBF) neural network and assess the effect of many variables on waste generation.	▪ RBF achieved the highest R^2 and lowest RMSE values among all the models. ▪ High efficiency of RBF model in the domain of waste generation prediction. ▪ GDP, income, and household as the most significant factor influencing MSW generation.
(Abunama et al., 2016)	ANN, MLP, SVM	Identifying the key elements contributing to leachate production and developing various AI-based models to predict leachate generation rates.	▪ MLP showed the best-fit model with R^2 values (0.956) and the lowest RMSE (398.5). ▪ It was found that the main factors contributing to the production of leachate are the amount of waste dumped, the level of precipitation, and the gases emitted.
(Kumar et al., 2018)	ANN, SVM, RF	Establish a relationship between the plastic waste generation rate and the identified socioeconomic groups.	▪ The prediction accuracy of ANN (R^2 = 0.75) was better than SVM (R^2 = 0.74) and RF model (R^2 = 0.66).
(Azadi et al., 2016)	ANN, MLR	Examining the performance of ANN and MLR for predicting the average seasonal MSWG rate.	▪ MLR showed poor prediction performance. ▪ On the other hand, the ANN model has a higher predictive accuracy when predicting the average rate of MSWG.
(Wei et al., 2013)	MLR	Predict the future MSWG in China using MLR	▪ This MLR model's reasonable accuracy is 94.93%, and its prediction accuracy is 97.26%.

(Chhay et al., 2018)	LR, ANN, GM	Investigation of the influencing socioeconomic factors in the MSWG using the fuzzy logic method and short-range forecasting for the MSWG using a multi-model approach.	▪ The ANN model predicted that China's MSWG will be 24,666.65 (10^4 tons) by 2030. ▪ The higher R^2 and lower errors founded for the ANN model. ▪ The ranking of the MSW generation forecasting model is ANN>GM (1, 1)>LR.
(Cha et al., 2017)	CHAID (DT)	Studying a classification method for demolition waste and developing a prediction model for the rate of demolition waste generation using CHAID method.	▪ It was found that the CHAID concrete model classified 98.9% of concrete generation correctly. ▪ DT training samples accurately predicted the yields of 91.3% to 99.5%.
(Wu et al., 2015)	GA	Introducing and testing a new computer-based prediction model, Gene Expression Programming (GEP).	▪ The R^2 for training data is 0.81, and the average relative error for validation data is 11.70%.
(Kannangara et al., 2018)	ANN, DT	Develop models for accurate prediction of MSWG and diversion based on demographic and socioeconomic variables.	▪ Neural network models had the best performance, describing 72%of variation in the data.
(Oliveira et al., 2019)	GA, ANN, MLR	develops an ANN model using GA algorithms to estimate the annual amount of household packaging waste.	▪ Predicted values is higher for the ANN model (R^2 = 0.98) than for the MLR (R^2 = 0.73).

2.5. Solid Wastes Generation Time-series Forecasting Models

Time series forecasting is based on historical data using statistics and modelling to make predictions and strategic decision-making. The accurate forecast depends on what kind of variables they are and on factors beyond our control. In addition, it depends on the nature of the available data. The more comprehensive it is, the more accurate the predictions will be (Younes et al., 2015). Time series have been applied in the SWM process in various fields.

Marandi et al. (2016) analyzed, compared and selected the best time-series model to predict the amount of SWG using monthly data for 2015 and 2016. ARIMA (2, 1, 0) was compared with ARIMA (1, 1, 1). The results showed that ARIMA (2, 1, 0) excelled in terms of MAPE, MAE and RMSE measures. (Marandi & Fatemi Ghomi, 2016). Another predictive study by Mohamed et al. (2011) to predict monthly SWG using data from the years 2005 to 2010. The results showed that ARIMA (1, 1, 1) was the best predictive model.

Navarro-Esbrí et al. (2002) compared the nonlinear dynamics technique with its performance against the automatic regression and seasonal moving average (SARIMA) methodology. The results showed better annual prediction performance for SARIMA with an accuracy level of mean relative error of less than 5%. For the weekly forecast, the relative error was less than 10%. For time series analysis, it was found that the combination of ARIMA and seasonal exponential smoothing (SES) techniques showed more accurate results with a mean absolute ratio error equal to 6.5 and a significant value of predicting variance for waste generation data R^2 above 0.87 (Rimaityte, Ruzgas, Denafas, Račys, & Martuzevicius, 2012).

Literature showed that LSTM has been used to solve SWM problems. Niu et al. (2020) used LSTM, ARIMA and ANN models to predict MSWG. The superiority of the LSTM neural network in MSW predictions is proven. The results showed an accuracy of 0.92, 935.08, and 114.36 in R^2, RMSE, and MAPE, respectively. Another study by Teng et al. (2022) studied an industrial case for power generation from a waste-to-energy cogeneration system. HTM, LSTM, ARIMA, and Fourier transform extrapolation (FTE) were used. The results showed that the best model is HTM, followed by LSTM then, ARIMA. The best MSE for the HTM model was 0.08466%.

Moreover, the quantity and type of MSWG were investigated using LSTM and a deep learning time series prediction network (DLTSF). The dataset includes the amount of solid waste in each area, its type (e.g., plastic, glass, paper, cardboard, organic waste) and level patterns (e.g., poor, social, privileged areas). The models were built based on level patterns, and it was found that the average solid waste amounted to 0.42, 0.65 and 0.86 kg/person/day for the poor, social and privileged areas, respectively. In addition, the results showed a mean RMSE of 0.03371 for predicting total MSW using DLTSF (Abdella et al., 2022).

2.6. Implemented Methods

In this research, time-series forecasting methods were utilized. Time series is defined as a series of data recorded at different intervals of equal chronological order that may be daily, monthly, or yearly. Through the previously observed values, it analyzes the series and predicts future values. Time series prediction is generally divided into two types: univariate time series prediction and multivariate time series prediction. If the past values are used to predict future values, then it is called univariate time series prediction. On the other hand, if the other external variables are used for forecasting, then it is called multivariate time series forecasting. Analyzing time series data helps plan for the future and extract statistics that help understand past behaviour (Beyaztas & Shang, 2022).

In our research on univariate time series prediction, we implemented several algorithms that are widely used in this domain. Two traditional statistical models, ARIMA and SARIMA, were employed to capture the linear dependencies and seasonality present in the data. Additionally, Prophet, a robust forecasting model developed by Facebook, which incorporates both trend and seasonality components. To harness the power of deep learning in sequence data analysis, LSTM, BILSTM, and GRU, which are advanced models specifically designed for handling sequential data were employed.

On other hand, in multivariate time series prediction, the implemented algorithms primarily centered around three widely used models: Artificial neural Network (ANN), Support Vector Machine (SVM), and Decision Trees (DT). These algorithms were chosen due to their ability to effectively handle datasets with multiple variables and capture the interactions between them within the time series data. Each

algorithm brings its own distinct characteristics and strengths to the table, enabling a thorough examination of the multivariate time series and facilitating accurate predictions.

2.6.1. Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a widely used statistical method for predicting future values in a time series based on past data. The method assumes that only historical time series data can be used to make predictions. It consists of three main components: automatic regression (AR), integration (I), and moving average (MA), which work together to capture patterns in time series data. These components are used to determine the order of the pattern represented by the coefficients (p, d, q).

Autoregression AR (p) uses past time steps to predict future values. p denotes the number of delays in Y to be used as predictors. In addition, it searches for a relationship or dependency between the current observation and observations from previous periods (Box et al., 2015). The regression model for AR is a function of the lags of Y_t, is given by the following equation:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + e_t$$
$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, e_t) \quad (1)$$

Y_t is a function representing the lags of the series $(Y_{t-1}, Y_{t-2}, Y_{t-n}, \ldots)$. The model estimates the coefficient β_p for each lag Y_{t-p}. Additionally, the model estimates both the intercept α and the error term e_t.

Moving Average MA (q) uses the differences between observed and predicted values in an AR model applied to delayed observations to investigate the presence of dependency. The model represents errors or deviations as a collection of past errors. The number of prior errors to be

considered is determined by the value q, which specifies the number of terms included in the model. Likewise, the MA model relies on delayed prediction errors (Contreras et al., 2003). The typical equation is given below:

$$Y_t = \alpha + e_t + \varphi_1 e_{t-1} + \varphi_2 e_{t-2} + \cdots + \varphi_q e_{t-q}$$
$$Y_t = f(e_t, e_{t-1}, e_{t-2}, \ldots, e_{t-q}) \quad (2)$$

Where Y_t depends only on the lagged forecast errors $(e_t, e_{t-1}, e_{t-2}, \ldots, e_{t-q})$ and φ_q is the coefficient for each error e_{t-q}. Integration I (d) remove trends and patterns that recur over time (known as seasonality) by subtracting the current values of a series from its previous values "d". This results in a series with a constant mean and variance over time, making it easier for the model to predict (Marandi & Fatemi Ghomi, 2016).

The order of differencing (d) is determined by examining the stability of the time series data. Stability is tested using statistical tests, such as the Augmented Dickey-Fuller or KPSS tests. Depending on the time series trends, the order of divergence is determined. For example, if the time series shows a complex pattern or has a high degree of seasonality, more diversions may be required to make it stable. In simpler terms, the order of difference is the number of times the data needs to be stabilized (Box et al., 2015). Fig.5 shows the flow chart for ARIMA model selection.

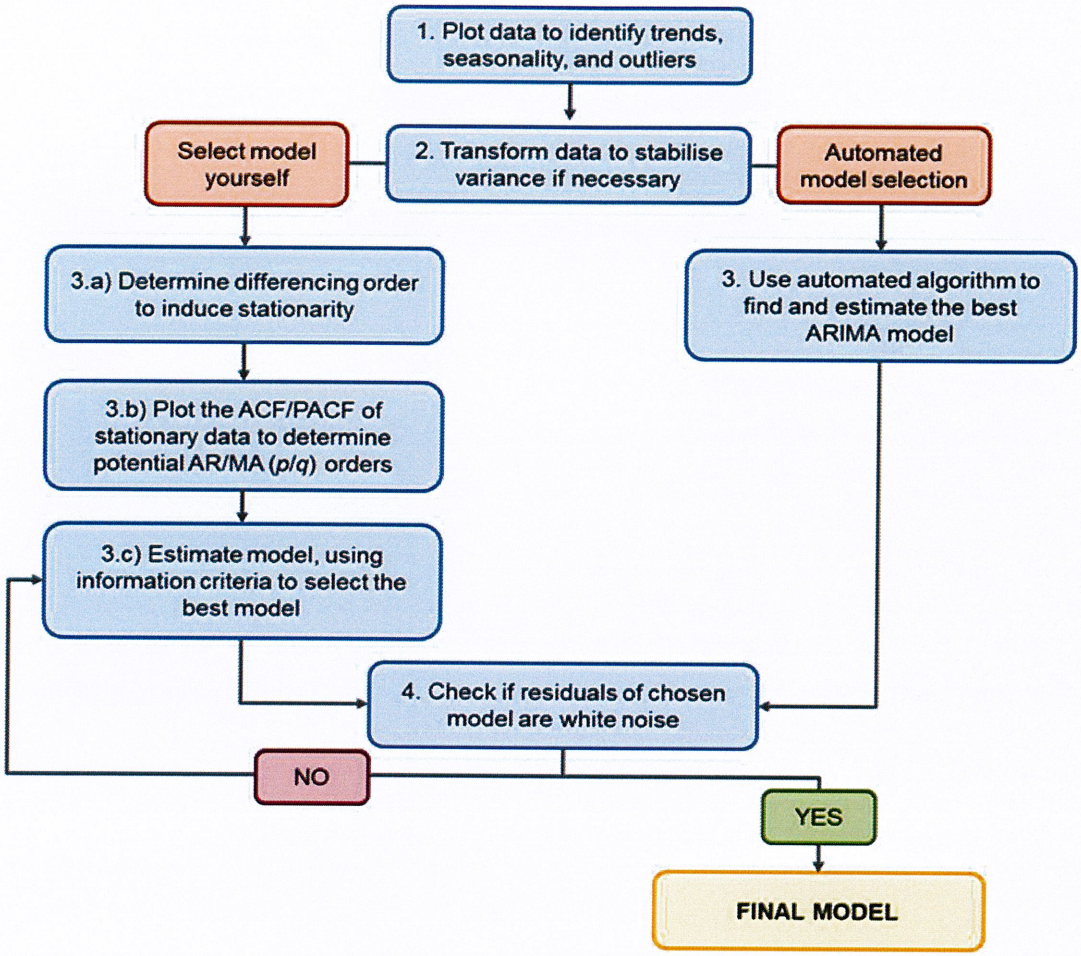


Fig. 5: Flow chart for ARIMA model selection

In time-series data, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used to identify potential orders for AR and MA models. They help catch basic patterns and make accurate predictions. The ACF represents the lag in the time series, while the PACF represents the lag after controlling for the mean lags (Benvenuto et al., 2020).

2.6.2. Seasonal Autoregressive Integrated Moving Average (SARIMA)

SARIMA is a model that builds upon the ARIMA model. It has the same order parameters as ARIMA (AR, I, and MA) but with the addition of seasonal orders. This addition makes the SARIMA model more robust for handling data with seasonal patterns (Navarro-Esbrí et al., 2002). It represented as shown in Fig.6.

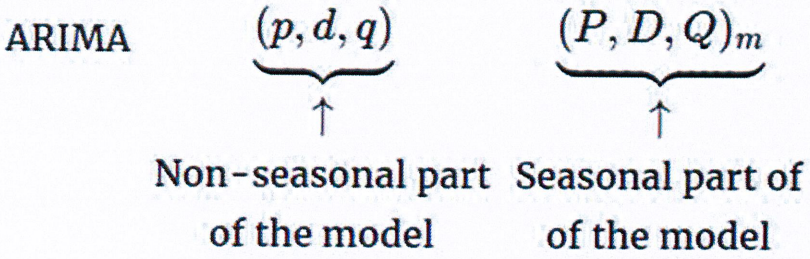


Fig. 6: The addition of seasonal orders

Where m is the number of observations per year or the period, and P, D and Q are values for the seasonal portions of the model, can be inferred from the ACF and PACF plots.

ARIMA models are easy to understand and implement for modelling time series data. It can predict non-stationary time series data after undergoing a differentiation process. In addition, it can capture various patterns, such as linear trends, seasonal patterns, and irregular patterns. On the other hand, the model follows some assumptions, such as the time series data must be homogeneous, which means that the variance of errors is constant over time (Box et al., 2015). Therefore, predictions from ARIMA may be inaccurate if this assumption is violated.

ARIMA models have difficulty modelling time series data with strong seasonal patterns. Therefore, SARIMA models are specifically designed to deal with them as they provide more accurate predictions by interpreting the correlation between observations during the season

(Rimaityte et al., 2012). SARIMA models can be challenging to understand and use, as they require a higher level of expertise in time series analysis. Additionally, they need a more significant amount of data to calculate the parameters, which can make the process more time-consuming. Moreover, SARIMA models are not effective for analyzing time series data that do not have seasonal patterns. (Korstanje, 2021)

2.6.3. Prophet Algorithm

Prophet is a time series prediction algorithm built by Facebook's Core Data Science team using machine learning techniques. It has the ability to handle trend changes and model seasonality. Additionally, its user-friendly nature, flexibility, and power in handling missing data and outliers make it a popular choice for data scientists and analysts working on time-series forecasting tasks (Prophet, 2018).

Prophet is based on an additive model that fits nonlinear trends to daily, weekly, and yearly seasonality, as well as holiday effects. The model is highly flexible and can be customized with various parameters to optimize it for specific time series forecasting tasks. When it comes to forecasting time series data, an additive model is utilized. This model type shows the observed data as a sum of various components. The components may consist of trends, seasonality, and other factors that help explain the data (Taylor & Letham, 2017). To express this concept in terms of the Prophet, the equation used is as follows:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (3)$$

In this equation, $y(t)$represents the time series data at time t, $g(t)$ represents the trend component of the model, $s(t)$ represents the seasonal component of the model, $h(t)$ describes the effects of vacations and other events in the time series, and $\varepsilon(t)$ represents the error or the remaining

component of the model. The model assumes that each component's effects are separate and can be combined to make a final forecast. This simplifies the interpretation and comprehension of each component's contributions to the overall behaviour of the time series (Hyndman & Athanasopoulos, 2021).

2.6.4. Long Short-Term Memory (LSTM) Recurrent Neural Networks

LSTM is a type of NN that is particularly useful for predicting future values in a time series. It is part of a larger class of networks called RNNs, but LSTMs are unique because they can hold important information for longer than regular RNNs. This is important because traditional RNNs suffer from a problem called vanishing gradients, in which information from steps prior to time is lost or forgotten. LSTM is designed to overcome this problem, making it a powerful tool for time series forecasting (Niu et al., 2021). Fig.7 illustrates the architecture of an LSTM technique.

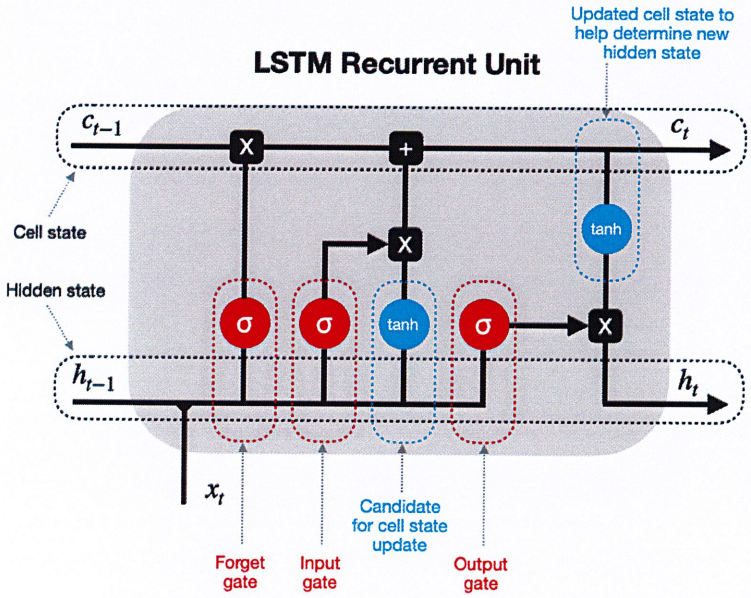


Fig. 7: illustration of the architecture of an LSTM

A typical LSTM module has four main components: a cell, an input gate, an output gate, and a forget gate. The cell is what allows the LSTM to store and remember the information at different points in time. Each of the three gates acts like an artificial neuron, calculating activation based on a weighted sum of inputs. In addition, they control the flow of information through the LSTM by regulating the values that pass through, which is why they are called "gates".

The input gate (i_t) specifies what information is allowed into the LSTM unit at the current point in time. Forgetting Gate (f_t) specifies information from the previous point in time that should be forgotten. The output gate (o_t) specifies what information from the current unit state will be used as the final output (Teng, et al., 2022). The functions of LSTM gates are described by the following equations:

$$i_t = \sigma\left(W_i[x_t, h_t] + b_i\right) \tag{4}$$

$$f_t = \sigma\left(W_f[x_t, h_{t-1}] + b_f\right) \tag{5}$$

$$o_t = \sigma\left(W_o[x_t, h_{t-1}] + b_o\right) \tag{6}$$

$$\hat{C}_t = tanh(W_c x_t + W_c h_{t-1} + b_c) \tag{7}$$

$$\mathrm{C}_t = f_t \times \mathrm{C}_{t-1} + i_t \times \hat{C}_t \tag{8}$$

$$\mathrm{h}_t = tanh\left(\mathrm{C}_t\right) \times \mathrm{o}_t \tag{9}$$

The sigmoid activation function (σ) is applied to regulate the flow of information. The weight matrix (W) is used to determine the strength of the connections between the different components of the unit. The input data at time t (x_t) is used to determine the information that needs to be processed by the unit. The hidden state at time t (h_t) is the unit output,

which represents the filtered information. Bias (b) is added to the calculations to adjust the overall behaviour of the unit. The previously hidden layer (h_{t-1}) is used to pass information from one point in time to another. The filter cell state ($\hat{C}_t$) is the intermediate state used to calculate the final cell memory state (C_{t-1}). The activation ($tanh$) function is used to introduce nonlinearity into the calculations.

Hyperparameter tuning of LSTM networks involves selecting the best set of network parameters to predict time series data accurately. The process typically involves specifying a range of values for each hyperparameter, such as the number of layers, the number of internal units per layer, batch size, dropout rate, and type of optimizer. Fig.8 shows the LSTM architecture design.

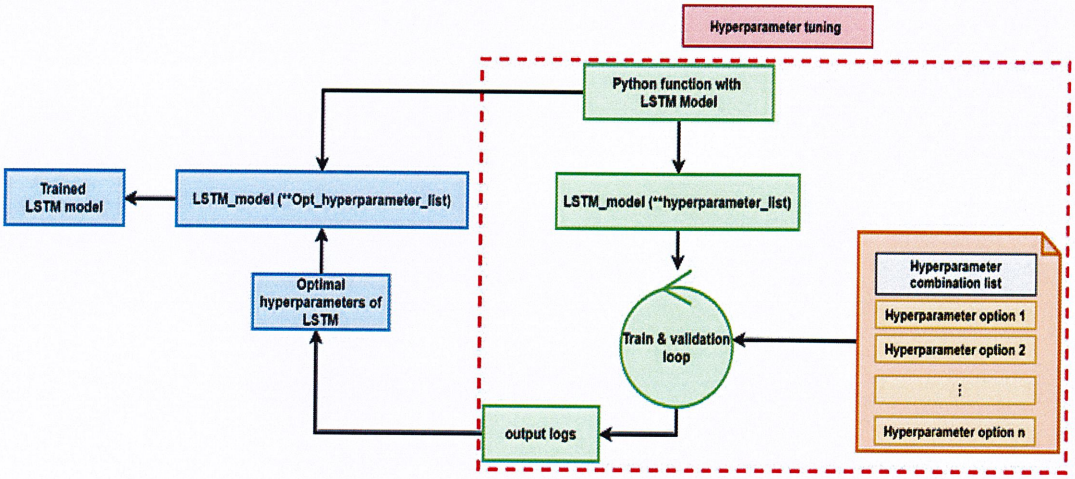


Fig. 8: LSTM hyperparameter tuning

The Keras tuner function is used to build a series of LSTM networks with different combinations of hyperparameters (TensorFlow, 2022). Each network is trained and evaluated using a training and validation set, and the results are recorded in a log. An optimal LSTM model is then determined based on performance metrics, such as loss and MAE.

2.6.5. Bidirectional LSTM (BILSTM) Recurrent Neural Networks

Bidirectional LSTMs are an evolution of traditional LSTMs that consider the past and future context of the input data. In traditional LSTMs, the hidden state is only updated based on the input data from the previous time step. Either way, a bi-directional LSTM updates the hidden state based on information from both previous and future steps. This approach can improve the performance of time series forecasting tasks by allowing the network to take advantage of past and future contexts in making predictions (Siami-Namini et al., 2019). The illustration in Fig.9 depicts the process of Bidirectional LSTM.

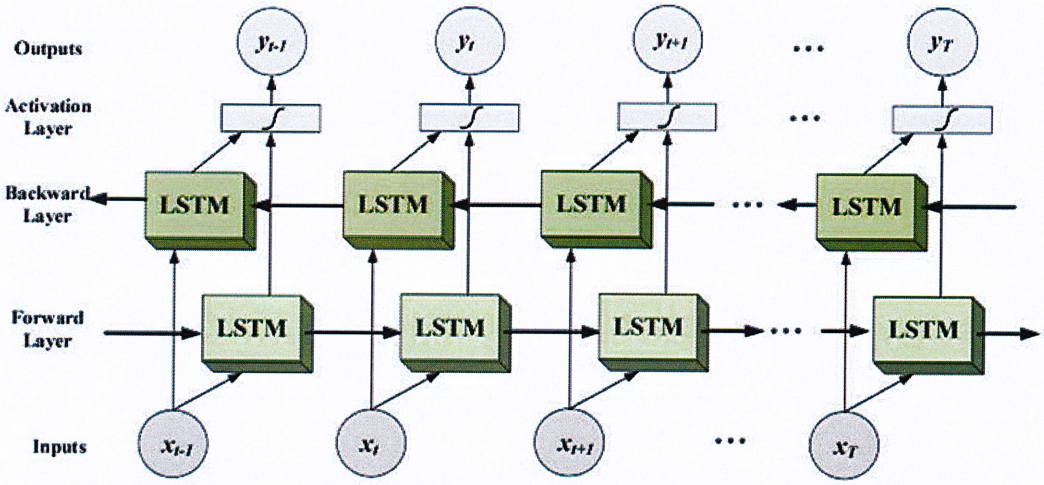


Fig. 9: The process of bidirectional LSTM

Two LSTMs are trained in parallel, one in the forward direction and the other in the backward direction. The final hidden state is formed by tracking the hidden state of each LSTM. The final prediction is then made by combining the outputs from both LSTMs.

2.6.6. Gated Recurrent Unit (GRU)

The GRU is a type of recurrent neural network (RNN) that's used to analyze time-series data. The GRU is similar to LSTM in that it addresses the issue of vanishing gradients, which is common in traditional RNNs, by introducing two gate functions: the update gate and the reset gate. These gates control the flow of information in the hidden state. The update gate decides which information should be retained and which should be discarded, similar to the forget-and-enter gate in LSTMs. The reset gate determines the extent to which information should be forgotten (Goodfellow et al., 2016). These functions of the GRU gates are represented mathematically in the following equations:

$$z_t = \sigma\left(W_z x_t + U_z\, h_{t-1} + b_z\right) \tag{10}$$

$$r_t = \sigma\left(W_r x_t + U_r\, h_{t-1} + b_r\right) \tag{11}$$

$$\hat{h}_t = tanh(W_h x_t + U_h(r_t h_{t-1}) + b_h) \tag{12}$$

$$h_t = z_t h_{t-1} + (1 - z_t)\,\hat{h}_t \tag{13}$$

Where (h_t) is the output vector and $(\hat{h}_t)$ is the candidate activation vector, which is controlled by two gate vectors, the update gate vector (z_t) and the reset gate vector (r_t). Fig.10 shows the difference between GRU and LSTM. LSTM has three gates (input, forget and output), while GRU has only two (update and reset). Additionally, LSTM contains a memory cell, while GRU does not have a separate memory cell. Moreover, LSTM is computationally more expensive than GRU because it requires more computations to update its gates and memory cell (Raschka & Mirjalili, 2017).

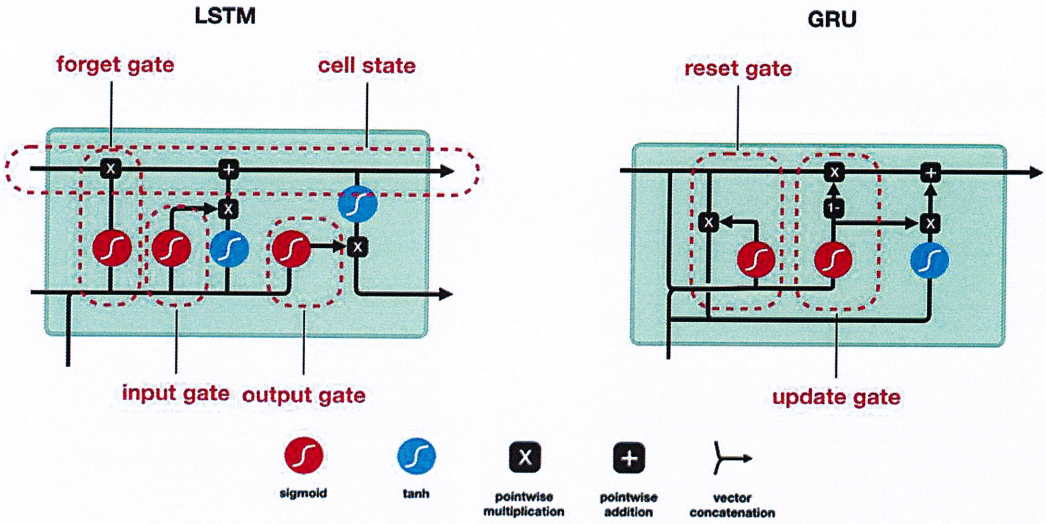


Fig. 10: The difference between GRU and LSTM

2.7. Research Gaps

A limited number of studies applied time-series models to predict SWG. It was found that the application of the ARIMA regression model and LSTM was limited to only 2% of the studies. In addition, studies related to SWM in Bahrain are limited. Time series models showed effective performance in predicting domestic waste generation rates. However, the models are highly dependent on their hyperparameters, including the number of delay feedback in the model, the number of layers, the number of units in each layer, and the initial learning rate. Therefore, all these hyperparameters must be tuned to develop the optimal model.

Regarding Ramadan, statistics in Arab countries confirm that the volume of spending during Ramadan, especially in the Gulf countries, increases by 50 percent compared to the rest of the year. Food turns into a treasure that must be obtained, and leftover bags and leftovers of various types of food are spread after breakfast in huge quantities. There are only a few studies that discuss Ramadan and waste generation, and these studies only provide a general overview without delving into specifics.

To address the environmental challenges related to waste in Bahrain, this research will focus on developing time-series models to predict SWG during Ramadan. Additionally, the study will investigate the relationship between Ramadan and waste generation in greater detail.

Chapter Three

Methods and Procedures

3.1. Research Design

3.2. Research Data Sources

3.3. Data Description

3.4. Data Pre-Processing

3.5. Training Approaches

3.6. Evaluation Metrics

Chapter Three
Methods and Procedures

3.1. Research Design

This chapter presents a high-level overview of the methodology applied in this research as shown in Fig. 11, starting with the planning, and understanding phase. Then, moves on to the data collection and processing phase, followed by the statistical data analysis stage. Finally, the model building and evaluation phase.

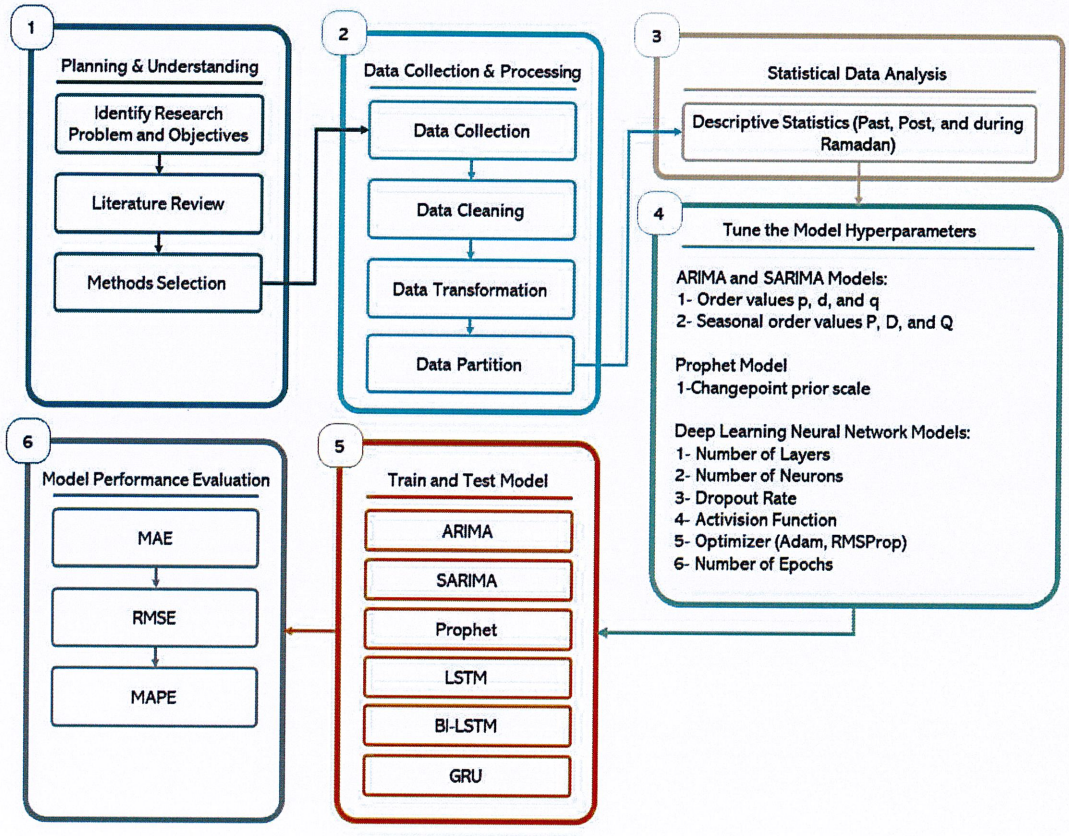


Fig. 11: Illustration of research methodology

The first stage in this study involves planning and understanding the objectives, needs, and challenges to be addressed in the research. A comprehensive literature review is conducted for the previous studies in the same field. Based on the results of this review, the appropriate methods were determined and selected.

The data collection and processing stage involves collecting and cleaning the data for the research. This stage ensures that missing or null values are handled, and outliers are identified. Furthermore, it includes verifying the data format and converting it into a usable format suitable for analysis.

The third stage involves analyzing the collected data to uncover its main characteristics and patterns of behaviour. This stage focuses explicitly on understanding domestic waste generation behaviour during pre-Ramadan, Ramadan and post-Ramadan, and whether there is a relationship between Ramadan and waste generation.

The last step in the research process involves model building and performance evaluation. The first step in this phase involves identifying the optimal hyperparameters. Then, the algorithms are applied, and their performance is evaluated to determine the most effective model.

3.2. Research Data Sources

The current study analyzed daily domestic waste deposited in the Askar landfill in Bahrain from 2019 to 2022. The data was obtained from the Ministry of Municipal Affairs and Agriculture, which manages the Askar landfill, the sole landfill currently receiving municipal waste, agricultural waste, and other types of waste (MUN, 2022). The study datasets represent a univariate time series of the amount of waste generated per day during the period of interest.

3.3. Data Description

This dataset includes 725,095 daily observations of various waste types. The dataset consists of time-series data related to waste generation in Bahrain. The data includes information on daily waste generation and the type of waste generated. There are 11 categories of waste which are domestic waste, garden waste, animal waste, building waste, electronic waste, blood waste, building recycling waste, commercial waste, liquid waste, and tyres waste.

The dataset covers multiple years (2019 to 2020), allowing for a comprehensive analysis of waste generation patterns and trends over time. It includes data from both the Ramadan period and the rest of the year, enabling a comparison of waste generation during these different periods.

3.4. Data Pre-Processing

After the data collection process, the cleanliness of the data was ensured by ensuring that there were no empty, zero, or negative values. In addition, it was ensured that there were no outliers, and that all data were in a normal state. This was done to ensure the data was accurate and reliable for analysis. Four missing values were filled with the average data value, and three negative values were discovered and corrected by taking their absolute values to make them positive. Finally, the daily waste readings for each type of waste were aggregated based on date and type, as shown in Fig.12.

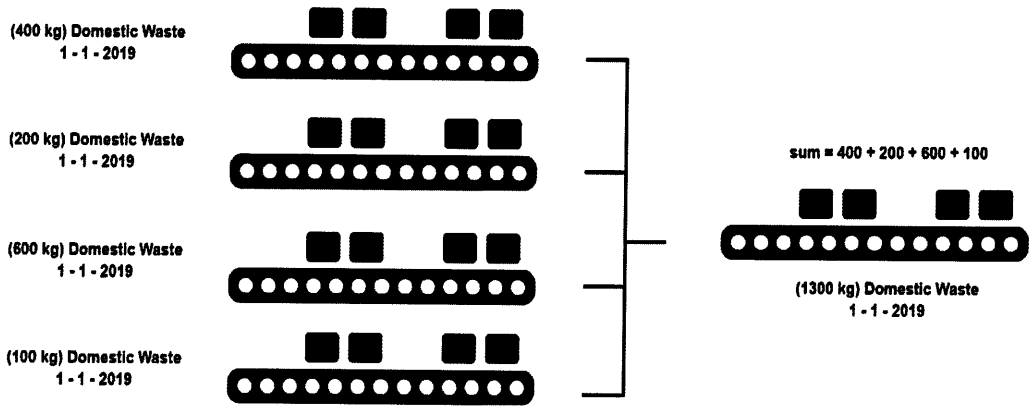


Fig. 12: Aggregation of the daily generation of waste

The waste amount dataset was scaled using the MinMax Scaler before training models. MinMax Scaler is a popular data scaling method that transforms each feature by subtracting the minimum value and dividing it by the range (maximum value - minimum value) for that feature. This measurement method standardizes the data and helps the model converge faster, making it useful for time series forecasting (Ahsan et al., 2021).

The sliding windows algorithm was utilized in the time series forecasting process. This algorithm determines the number of past time steps (look_back) used as input to make predictions for the subsequent time step (Hota et al., 2017). For instance, if look_back is set to 4, the model will use information from the preceding 4-time points to predict the next time point, as depicted in Fig. 13. The value of look_back is determined based on the size of the data set and the unique characteristics of the problem being addressed.

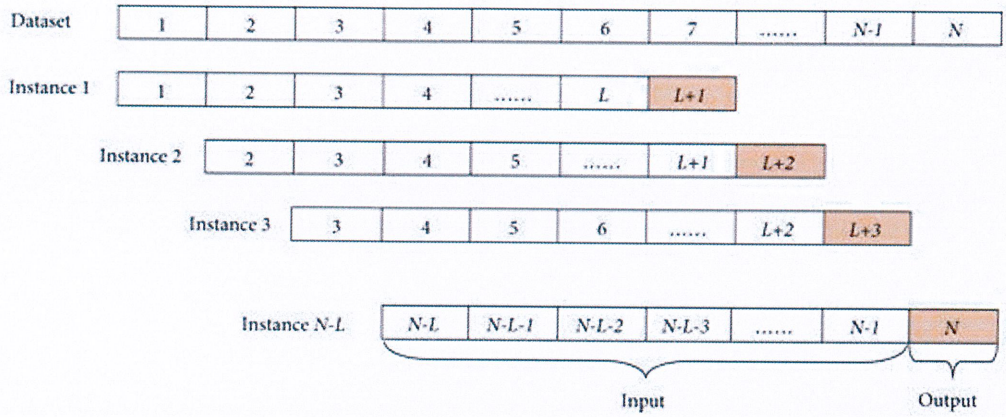


Fig. 13: The sliding windows algorithm

The data was split into two parts after pre-processing. The first part consisted of 75% of the observations, ranging from 2019 to 2021, which was used as the training set for the developed model. The other 25%, from 2022, served as the test set to evaluate the model's performance.

3.5. Training Approaches

This section explores various training approaches employed in the study, providing detailed insights into the selection rationale and methodology.

3.5.1. ARIMA and SARIMA Training Approach

The training approach utilized for ARIMA and SARIMA in this study involves three main components: automatic regression (AR), integration (I), and moving average (MA). These components work together to capture patterns within the time series data by determining the order of the coefficients (p, d, q) that represent the patterns.

To identify potential orders for the AR and MA models, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are employed. These plots help identify and capture the underlying patterns and enable accurate predictions. The ACF represents

the lag in the time series, while the PACF represents the lag after controlling for mean lags.

In the case of the ARIMA model, two approaches are used to determine the orders (p, d, q). The first approach involves analyzing graphs such as ACF and PACF plots. The second approach utilizes the auto ARIMA function. Additionally, the stationarity of the data is confirmed through the Dickey-Fuller test, where the p-value is evaluated to be less than 0.05.

For the SARIMA model, the auto ARIMA function is used to determine the seasonal order. Furthermore, seasonal decompositions are employed to confirm the presence of seasonal trends in the data.

These training approaches provide a robust framework for modeling and analyzing time series data, enabling accurate predictions and capturing underlying patterns.

3.5.2. Prophet Training Approach

The training approach of the Prophet model is known for its user-friendly and flexible nature, making it accessible to users with varying levels of expertise. The model requires minimal parameter tuning, simplifying the forecasting process for time series data.

Prophet utilizes an additive model for time series forecasting. This model represents the observed data as a combination of several components, including trend, seasonality, holidays, and error. By decomposing the time series into these components, Prophet captures the underlying patterns and fluctuations, enabling accurate predictions.

In the training phase of the Prophet algorithm, one important parameter to consider is the changepoint prior scale. This parameter controls the flexibility of the model in capturing potential changes or breakpoints in the time series. Different values of the changepoint prior

scale, such as [0.0001, 0.01, 0.05, 0.1, 0.5, 1.0], can be tested to explore their impact on the forecasting performance. This allows users to optimize the model for their specific time series forecasting tasks, providing flexibility and adaptability.

3.5.3. LSTM, BILSTM, and GRU Training Approach

The training approach for deep machine learning models involves hyperparameter tuning to optimize their performance in predicting time series data accurately. Hyperparameter tuning entails selecting the best combination of network parameters to achieve optimal results. This process includes specifying a range of values for each hyperparameter, such as the number of layers, the number of neurons per layer, batch size, dropout rate, and optimizer type.

Various hyperparameters were tested. These included the number of hidden layers (1, 2, 3), the number of neurons (50, 100, 150), forecasted windows (7, 14, 30, 40, 60, 90), and the type of optimizer (Adam, RMSprop).

To explore different combinations of hyperparameters, the Keras tuner function was employed. This function builds a series of LSTM networks with varying hyperparameter configurations. Each network is trained and evaluated using a training and validation set, and the results are recorded in a log. Performance metrics such as loss and MAE are used to assess the models' performance.

By iteratively testing different hyperparameter configurations and evaluating their performance, an optimal LSTM model can be determined. This approach allows for fine-tuning the models to achieve the best possible accuracy in predicting time series data.

3.6. Evaluation Metrics

In the study, multiple rating scales were used to assess the accuracy of the models. This was done to ensure a comprehensive analysis of the performance of the models. Therefore, four different measures of performance evaluation were used: MAE, RMSE and MAPE and R^2.

MAE and RMSE are widely used metrics for measuring the accuracy and precision of predictions. They provide insights into the average and overall magnitude of the errors between predicted values and actual values. These metrics are valuable for assessing the model's ability to minimize errors and evaluate its predictive performance.

MAPE is a valuable metric for understanding the relative error in percentage terms. It makes the evaluation of the model's performance in a more interpretable and meaningful way. In addition, it is particularly useful when comparing the accuracy of different models across different datasets or time periods.

R-squared, also known as the coefficient of determination, measures the proportion of variance in the dependent variable (actual values) that is explained by the independent variable (predicted values). It provides an indication of how well the model fits the data and captures the variability in the response variable.

By employing these performance metrics, the proposed system can be thoroughly validated and compared with other models. The combination of the metrics offers a comprehensive evaluation of accuracy, precision, relative error, and explained variance, allowing for a robust comparison of different models and enabling informed decision-making in selecting the best-performing model for the specific application.

3.6.1. Mean Absolute Error (MAE)

MAE is a common evaluation metric used in regression and time series forecasting problems. It measures the absolute difference between the actual and predicted values in data. Then, finding the average of all these differences set as shown in the following equation:

$$\text{MAE} = \frac{1}{\text{n}} \sum_{i=1}^{n} |y_i - \hat{y}| \tag{14}$$

Where (y_i) is the actual values and $(\hat{y})$ is the predicted value for (n) the number of observations. A low MAE value indicates a high prediction accuracy, as the difference between the actual and predicted values is small.

3.6.2. Root Mean Square Error (RMSE)

RMSE is a measure of the average deviation between expected values and actual values. It is calculated by taking the difference between expected and actual values, squaring them to make them positive, finding the mean squared differences, and taking the square root of the mean. RMSE is calculated using the following formula:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n}} \tag{15}$$

The resulting value indicates the average size of the differences, and smaller RMSE values indicate better performance of the model. The RMSE is suitable for comparing multiple models in a single dataset. However, it is sensitive to outliers (Casella & L. Berger, 2002).

3.6.3. Mean Absolute Percentage Error (MAPE)

MAPE is a commonly used method for evaluating prediction accuracy as it can be easily interpreted. It is measured by taking the absolute difference between the actual and predicted values, dividing it by the actual value, and multiplying by 100. The lower the MAPE, the better the predicted performance because it indicates a smaller difference between the actual and predicted values (Hyndman & Athanasopoulos, 2021). The following equation is used to determine MAPE:

$$\text{MAPE} = [\frac{1}{\text{n}} \sum_{i=1}^{n} \frac{(y_i - \hat{y})}{y_i}] \times 100 \quad (16)$$

MAPE is a useful method for comparing models and checking prediction accuracy in absolute terms. However, MAPE is not defined when the actual value equals zero and can be very sensitive to outliers. Despite these limitations, it is still widely used due to its ease of interpretation and size independence.

3.6.4. Coefficient of Determination (R^2)

The coefficient of determination, known as R-squared (R^2), is a statistical metric that shows how much of the variance in the dependent variable is explained by the independent variable in a regression model. It is used to evaluate whether the model is a good fit for the observed data. The R^2 value is a number that falls between 0 and 1, where higher values indicate a better fit of the model to the data. It can be calculated using the following equation:

$$\text{R}^2 = 1 - \frac{RSS}{TSS} \quad (18)$$

Where (RSS) is the sum of the squared residuals, which is the differences between the predicted values and the actual values. (TSS) is the total sum of squares which is the difference between the actual values and the mean value of the dependent variable. However, it is essential to understand that R^2 only measures the percentage of variation in the dependent variable and does not establish a cause-and-effect relationship between the variables. Furthermore, R^2 does not determine the accuracy or reliability of the model's predictions. Therefore, it is necessary to use additional statistical measures in combination with R^2 to obtain a comprehensive interpretation of the regression model.

Chapter Four

Results, Discussion and Conclusion

4.1. Results

4.2. Discussion

4.3. Conclusion

Chapter Four

Results, Discussion and Conclusion

4.1. Results

This chapter presents comprehensive descriptive statistics on domestic waste for the period between 2019 to 2022. In the statistical analysis chapter, the study examines patterns and behaviours related to domestic waste. The subsequent chapters report the performance outcomes of the models and compare them to determine which one achieved the highest level of accuracy.

4.1.1. Descriptive Statistic

Table 3 presents a statistical summary of domestic waste for the dataset for each individual year from 2019 to 2022, while Figure 16 illustrates the distribution of the data.

Table (3): Domestic waste descriptive statistics from 2019 to 2022

	Domestic Waste (tons)			
	2019	**2020**	**2021**	**2022**
count	365	366	365	365
mean	1435.88	1422.72	1445.13	1449.11
std	93.60	70.32	76.21	85.59
min	1236.05	961.95	1104.41	1293.01
max	1920.43	1634.14	1717.85	1992.97

It presents the following information for each year: the count of days, the mean value of waste generation, and the minimum and maximum waste generation per day. It provides an overview of the

distribution of waste generation across different years, allowing for a comparison of the average waste generation, as well as the range of variation for daily waste generation. Throughout the observed period, the number of days per year was 365, with the exception of 2021, which featured 366 days. The mean varied slightly over the four years, the difference was relatively minor, ranging from 1422.72 to 1449.11, approximately a variance of 27 tons. In 2022, the quantity of domestic waste generated per day reached the highest value of 1992.97 tons/d, whereas the lowest value was recorded in 2020 at 961.95 tons/d.

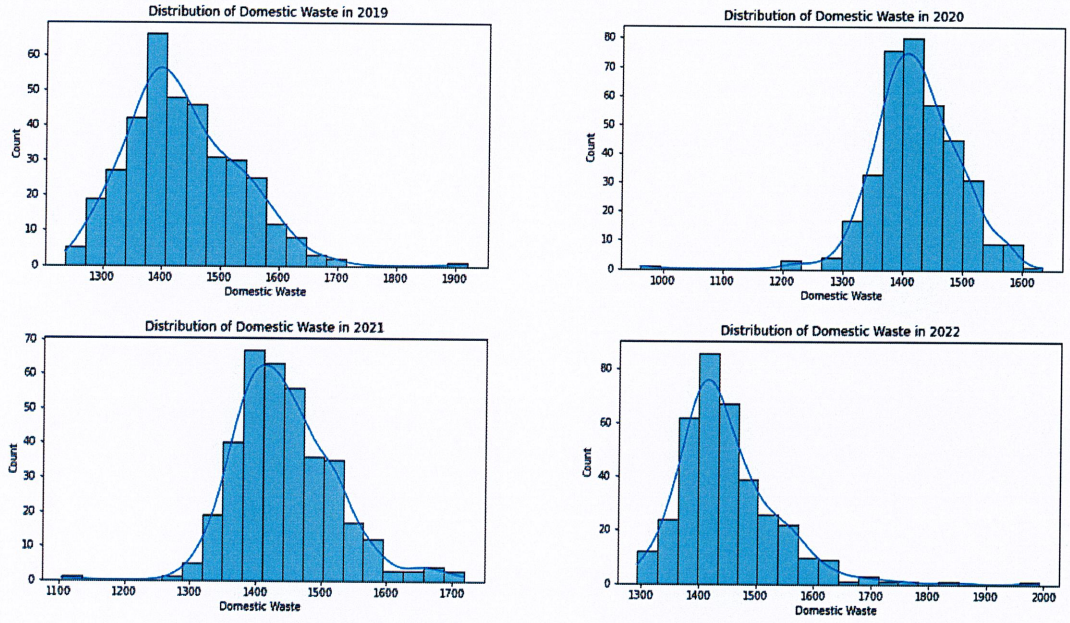


Fig. 14: Domestic waste frequency distribution for each individual year

Comprehending skewness in a given dataset can prove beneficial in the interpretation and analysis of the data. Such an understanding can offer insights into the presence of any potential outliers that may impact the statistical analyses validity. Moreover, knowing the skewness of a dataset can aid in selecting suitable statistical models and tests for analysis. Fig. 14 indicates that the data for 2019 and 2022 follow a positively skewed distribution, whereas the data for 2020 has a negatively skewed distribution. In contrast, the data for 2021 appears to be more closely aligned with a normal distribution. Table 4 presents a statistical

summary of domestic waste for the complete dataset, while Fig. 15 shows the distribution of the data.

Table (4): Domestic waste descriptive statistics for the dataset

	Domestic Waste (tons) 2019 -2022
count	1461
mean	1438.20
std	82.45
min	961.95
25% percentile	1384.46
50% percentile	1427.25
75% percentile	1483.76
max	1992.97

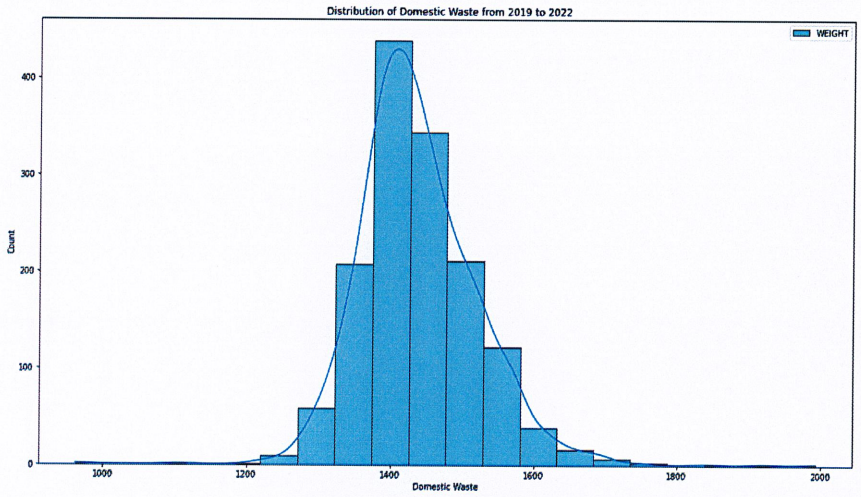


Fig. 15: Domestic waste frequency distribution for the dataset

The dataset consists of 1461 instances, with a mean value of 1438.20. Furthermore, the data exhibits a normal distribution. The standard deviation (SD) of the data is 82.45, indicating that the data points are relatively close to the mean value. The interquartile range (IQR) of the data, which represents the range between the 25th and 75th percentile, spans from 1384.46 to 1483.76, indicating a relatively narrow range of values.

4.1.2. Statistical Analysis

The dataset provided by the Ministry of Municipal Affairs and Agriculture was utilized to study waste classifications and determine which waste type constitutes the highest percentage. The following figure illustrates the percentage and quantity of waste for each type from 2019 to 2020.

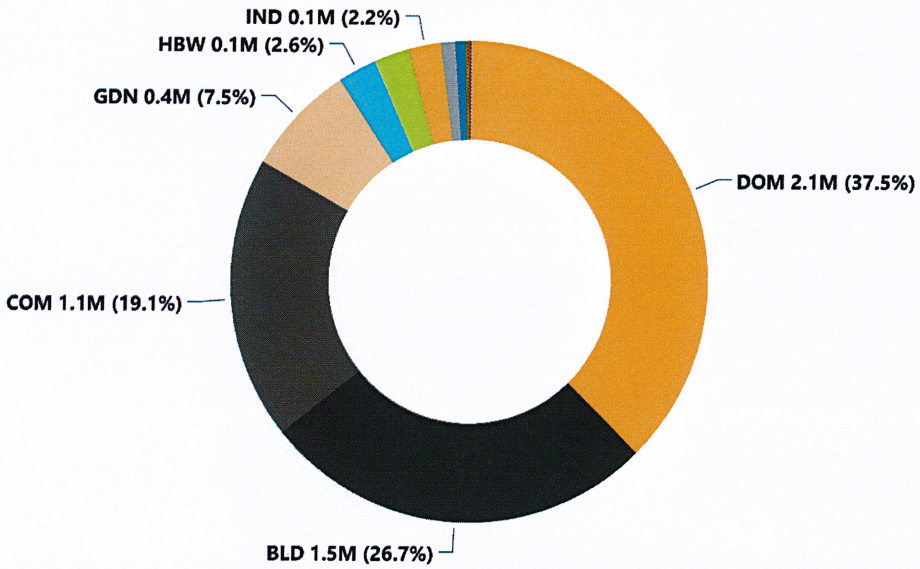


Fig. 16: Waste type quantity (percent)

The largest proportion of waste, amounting to 37.5% or 2.1 million tons, is attributed to domestic waste, followed by construction waste, which represents 26.7% or 1.5 million tons of the total waste, followed by commercial waste represents 19.1% or, 1.1 million tons. Concerning other waste classifications, they do not exceed 10% and amount to less than 1 million tons.

The trend of yearly domestic waste is depicted in Fig. 17. It is observed that the data fluctuates around the mean value for all years. Throughout the four years, only four data points were identified as being significantly different from the mean value. Fig. 18 illustrates the average domestic waste per day in each month where the orange color represents the month of Ramadan.

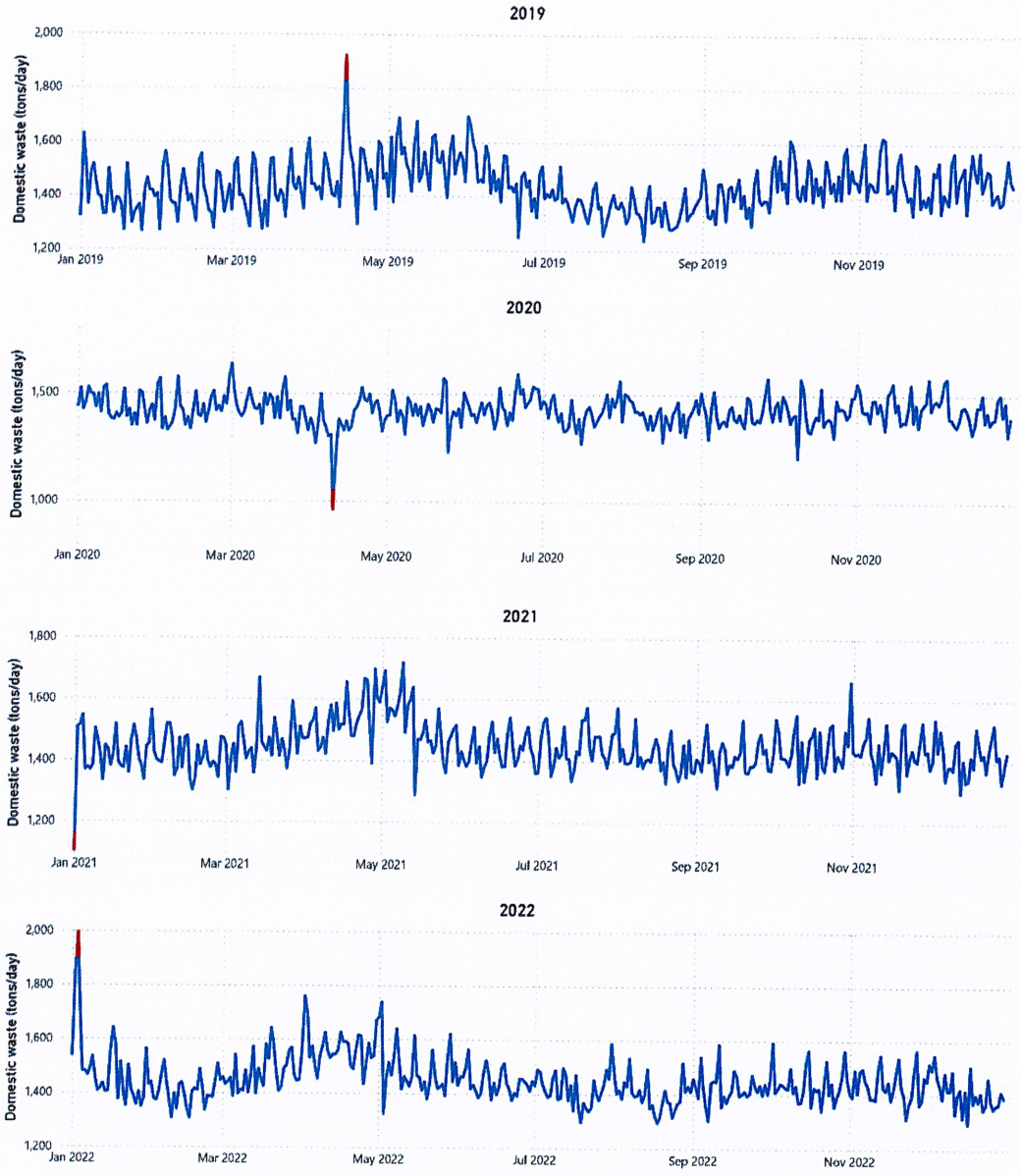


Fig. 17: The trend of domestic waste for each individual year

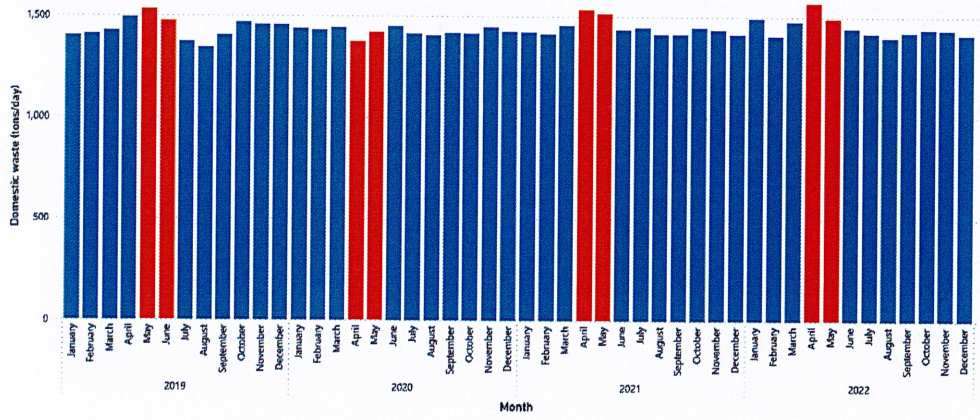


Fig. 18: Average domestic waste per day in each month

During the four years examined, there was a notable increase in the daily amount of domestic waste generated, particularly during Ramadan, except for the year 2020. This year was distinguished by a decrease in food waste, especially during Ramadan where the waste was significantly lower compared to other months. The decrease in food waste can be attributed to the impact of the COVID-19 pandemic. The pandemic resulted in various restrictions and changes in societal behavior, including altered consumption patterns and reduced social gatherings. With fewer large-scale Ramadan events, communal iftars, and gatherings, there was likely a decrease in food waste generation during this period.

The study focused on waste trends pre, during, and post-Ramadan, with the aim of identifying any fluctuations in the daily amount of waste generated, as well as any changes that occurred annually during this period. Fig. 19 presents the average quantities of waste generated during the period pre, during, and post-Ramadan.

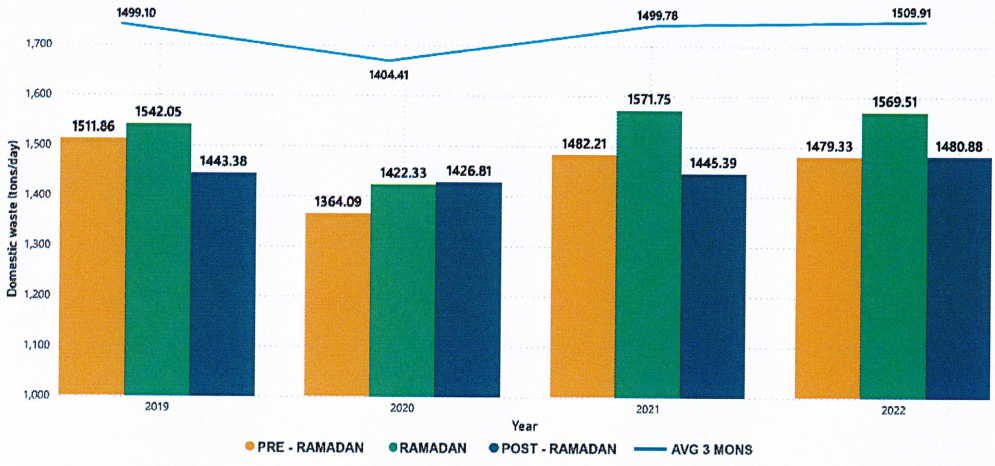


Fig. 19: Average quantities of waste generated during the period pre, during, and post-Ramadan

The data reveals a clear increase in the amount of domestic waste generated during the month of Ramadan, in comparison to the periods before and after it. For instance, in 2022, the waste generated during this month increased from approximately 1480 to 1570 tons per day, representing an increase of around 90 tons. Furthermore, when comparing the four years, it can be observed that the amount of waste generated during this period has been increasing annually. For instance, in 2021, the average waste per day during this period was around 1500 tons, while in 2020, it was approximately 1510 tons, indicating an increase of 10 tons.

To investigate the daily variation in domestic waste generation during the month of Ramadan, Fig. 20 displays the amount of waste generated on a daily basis for each year. The black dots indicate the weekend days, which are Friday and Saturday. It can be observed that, in all years, waste generation is higher on weekends compared to other days.

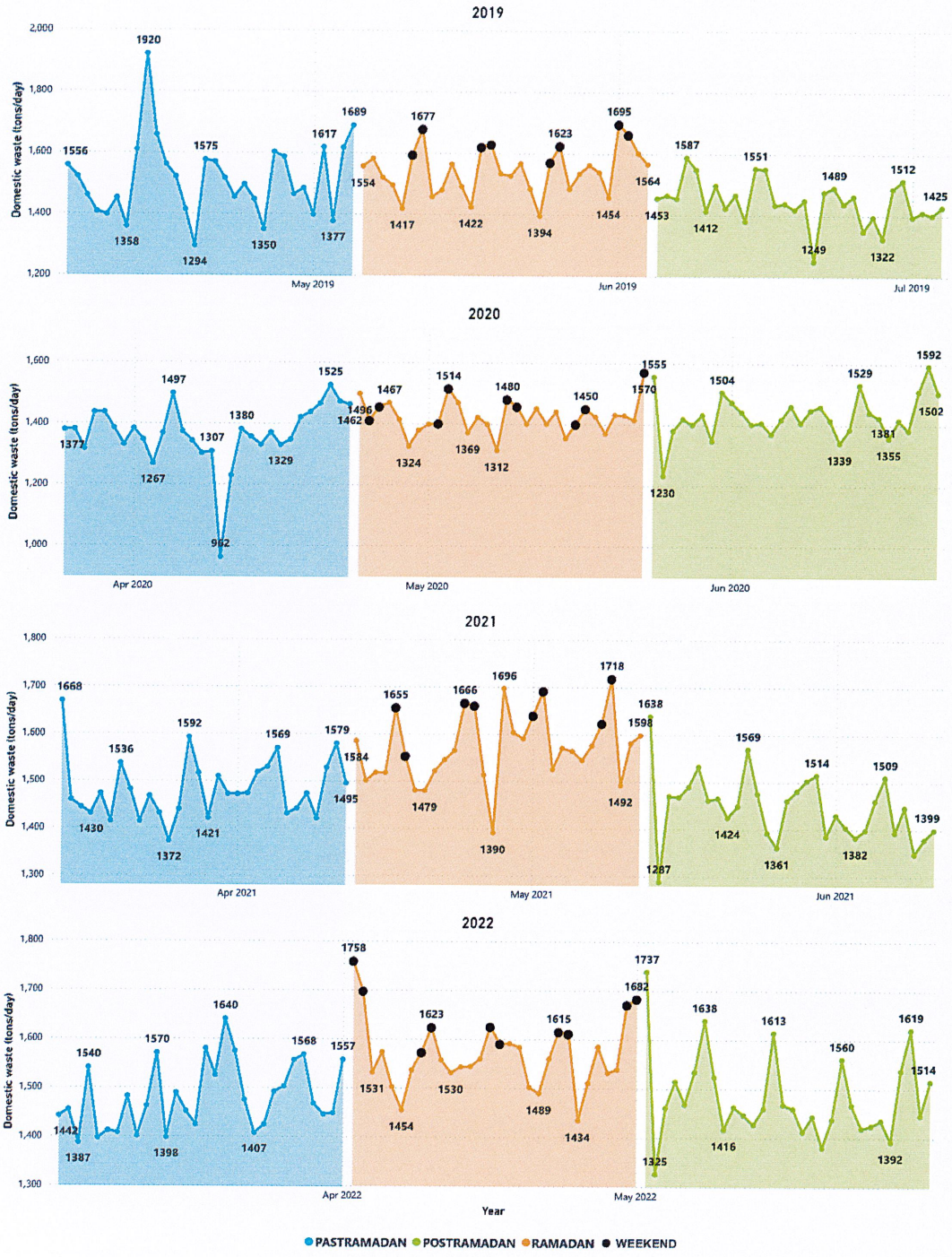


Fig. 20: The daily waste generation pre, during, and post Ramadan

4.1.3. ARIMA Model Performance

For the ARIMA model to work efficiently, it requires stationarity data. Therefore, before utilizing the ARIMA model, it is essential to confirm the stationarity of the data by conducting the Dickey-Fuller test. Table 5 displays the test outcomes.

Table (5): Results of the Dickey-Fuller test

Test Statistic	-4.94
p-value	0.00138
#Lags Used	22.00
Number of Observations Used	1438.00
Critical Value (1%)	-3.43
Critical Value (5%)	-2.86
Critical Value (10%)	-2.57

Based on the results presented in the table, there is strong evidence to suggest that the time series being tested is stationary. This is indicated by the test statistic being more extreme than any of the critical values and the p-value being very small (i.e., less than 0.05). These findings provide support against the null hypothesis of a non-stationary series.

To determine the appropriate order of the ARIMA model (p,q), the ACF and PACF are utilized, as depicted in Fig. 21. The PACF at lag 7 was observed to be significantly higher than the significance line, hence the value of p was fixed at 7. Similarly, the ACF at lag 7 was found to be significantly higher than the significance line, thus the value of q was fixed at 7. To ensure the validity of the results, another test was conducted using the AUTOARIMA function, which selects the order of the ARIMA model based on the lowest AIC value. Table 6 shows the results of the AUTOARIMA function.

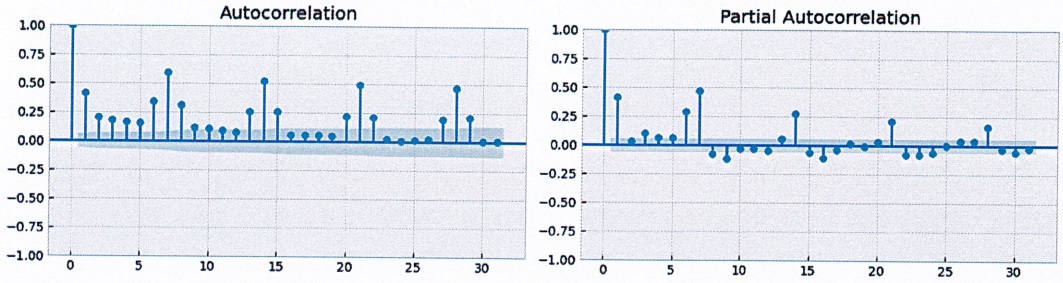


Fig. 21: Autocorrelation function (ACF), and partial autocorrelation function (PACF) plots

Table (6): Results of the AUTOARIMA function

Order	AIC
ARIMA (0,0,0)	25399.218
ARIMA (1,0,1)	16716.979
ARIMA (0,0,1)	23498.299
ARIMA (6,0,1)	16488.399
ARIMA (5,0,5)	16365.964
ARIMA (6,0,7)	16055.263
ARIMA (7,0,6)	16028.379
ARIMA (7,0,7)	16027.072

The lowest AIC value was found to be 16027.072 with an ARIMA order of (7,0,7). Notably, this order corresponds to the one obtained from the ACF and PACF plots, signifying a suitable selection of the order.

Table 7 shows the summary of the ARIMA model with the order of (7,0,7). The "coef" column shows the estimated value of each coefficient. The "std err" column shows the standard error of each coefficient estimate. The "z" column shows the value of the test statistic for the hypothesis that the true value of the coefficient is zero. The "P>|z|" column shows the corresponding p-value for this test. Typically, our interest lies in coefficients with p-values less than 0.05, as they imply statistical significance and make it improbable for the coefficient to be zero in the population. The confidence interval provides an estimate of

the likely range of values that the genuine value of the coefficient may encompass.

Table (7): ARIMA model summary

	Value	Std Err	Z	P>|Z|
const	1438.20	46.303	31.061	0.000
AR (1)	-0.1001	0.047	-2.145	0.032
AR (2)	0.2251	0.042	5.298	0.000
AR (3)	-0.3294	0.034	-9.701	0.000
AR (4)	0.3436	0.027	12.734	0.000
AR (5)	-0.3106	0.028	-11.210	0.000
AR (6)	0.1982	0.037	5.397	0.000
AR (7)	0.9261	0.043	21.367	0.000
MA (1)	0.3567	0.049	7.294	0.000
MA (2)	-0.0121	0.033	-0.362	0.717
MA (3)	0.5194	0.034	15.415	0.000
MA (4)	-0.1006	0.030	-3.320	0.001
MA (5)	0.4330	0.025	17.624	0.000
MA (6)	0.0520	0.041	1.262	0.207
MA (7)	-0.7017	0.032	-21.712	0.000
sigma2	3416.75	77.252	44.229	0.000

AIC: 16027.072 BIC: 16111.662

The statistical analysis shows that all the AR and MA coefficients are statistically significant, with a T-statistic greater than 2 in magnitude for most of the coefficients. Additionally, the P-values for most coefficients are less than 0.05, indicating their significance in the model. However, two MA coefficients, namely MA (2) and MA (6), have P-values of 0.717 and 0.207 respectively.

This study presents a performance comparison between the actual and predicted data of daily domestic waste generation using the ARIMA model as shown in Fig. 22. In addition, Fig. 23 shows the performance comparison of the prediction during the period pre, during and post-Ramadan.

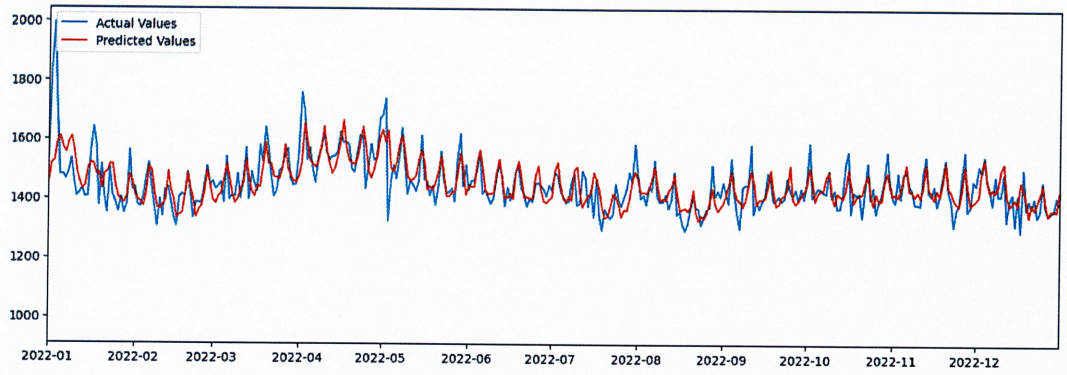


Fig. 22: Prediction of daily domestic waste generation using ARIMA

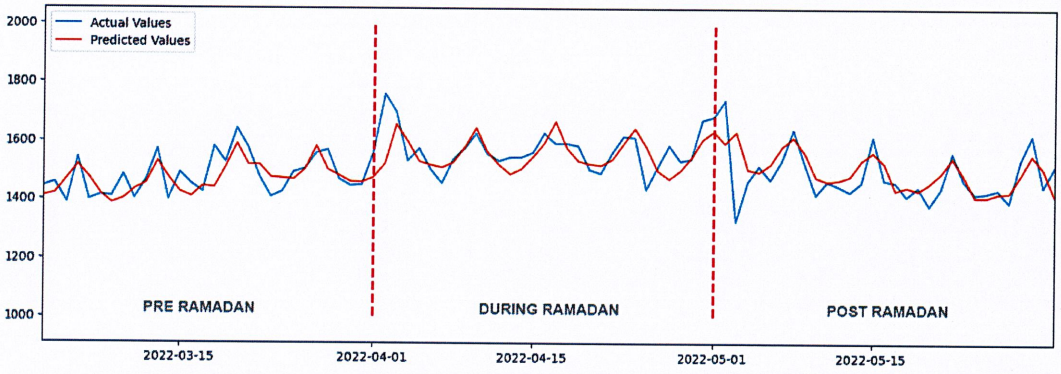


Fig. 23: Prediction of daily domestic waste generation pre, during and post-Ramadan using ARIMA

Fig.22 and Fig.23 demonstrate that the model exhibits a good fit to the daily data, as it tracks the fluctuations in the data accurately, both in terms of the upward and downward trends. The model exhibits good fitting for most of the peak values, but it is less accurate in predicting the trough values. Furthermore, the model does not accurately predict values that deviate significantly from the mean.

To gain a more comprehensive understanding of the model's accuracy, Table 8 provides an overview of the prediction performance

metrics for the ARIMA model in predicting daily domestic waste generation.

Table (8): ARIMA model prediction performance

ARIMA Model	MAE	RMSE	MAPE	R^2
Training	42.776	62.094	0.0293	0.474
Testing	41.763	55.973	0.0291	0.526

The MAE and RMSE values were 42.776 and 62.094 for the training set, and 41.763 and 55.973 for the testing set, respectively. The MAPE was 0.0293 for the training set and 0.0291 for the testing set. The R^2 was 0.474 for the training set and 0.526 for the testing set. Close training and testing performance helps prevent overfitting and underfitting of the data by the model. Overfitting occurs when a model becomes overly specialized to the training data, performing well on it but failing to generalize to new, unseen data. Underfitting, on the other hand, happens when a model is too simplistic and cannot capture the underlying patterns in the data during training, resulting in poor performance for both the training and testing data.

The closer training and testing performance enables it to better capture any temporal patterns in the time series. This familiarity with similar patterns during training enhances the model's ability to make more accurate predictions during the testing phase. Therefore, using a close training and testing performance ensures that the model's performance evaluation reflects its capability to handle real-world scenarios and make precise predictions on unseen data that closely resembles the training data.

Fig. 24 displays an ARIMA diagnostics plot that is utilized to evaluate the goodness of fit of an ARIMA model.

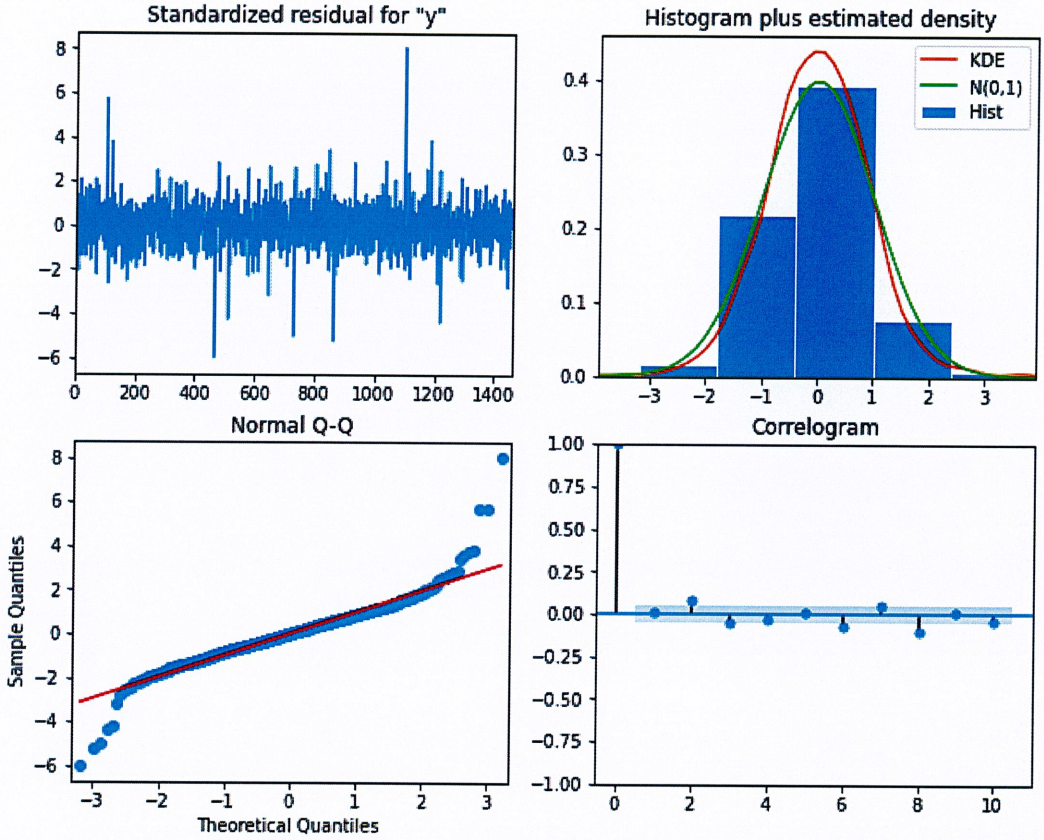


Fig. 24: ARIMA diagnostics plot

The residuals over time graph shows whether there are any patterns in the residual values over time. The more the random scatter of the residues is around zero, the better the model. The residual errors seem to fluctuate around an average of zero and have a uniform variance for most points where several points have irregular variance.

The histogram graph shows that the distribution of residual values has a normal residual distribution, which means that the model captures basic patterns in the data.

The QQ plot is commonly used to determine the goodness of fit of a model. A straight line of points on the plot indicates a good fit. As

observed in the figure, the points in the middle of the line follow a straight pattern, but on the sides, they deviate from the straight line.

Based on the Correlogram plot, it appears that the residual errors of the ARIMA model are not autocorrelated. Additionally, there is no observable pattern in the residual errors that can be attributed to the model's inadequacy. Therefore, the plot suggests that the ARIMA model is a good fit for the data.

4.1.4. SARIMA Model Performance

Understanding the underlying patterns and fluctuations in a time series is crucial for time forecasting using seasonal methods. Such an analysis can help in forecasting, detecting deviations, and performing other types of analyses. Fig. 25 presents the seasonal decomposition of the daily domestic waste generated. This plot breaks down a time series into its underlying components: trend, seasonal, and residual. It is noted that the trend of the data is not fixed as the values are increasing or decreasing in the series. As for seasonality, there is no clear recurrence in the series or a specific pattern for the data. The plot line of the residual component of a time series is flat and centered around the value of 1, which means that the residual variation is consistent and does not show any significant trends or patterns.

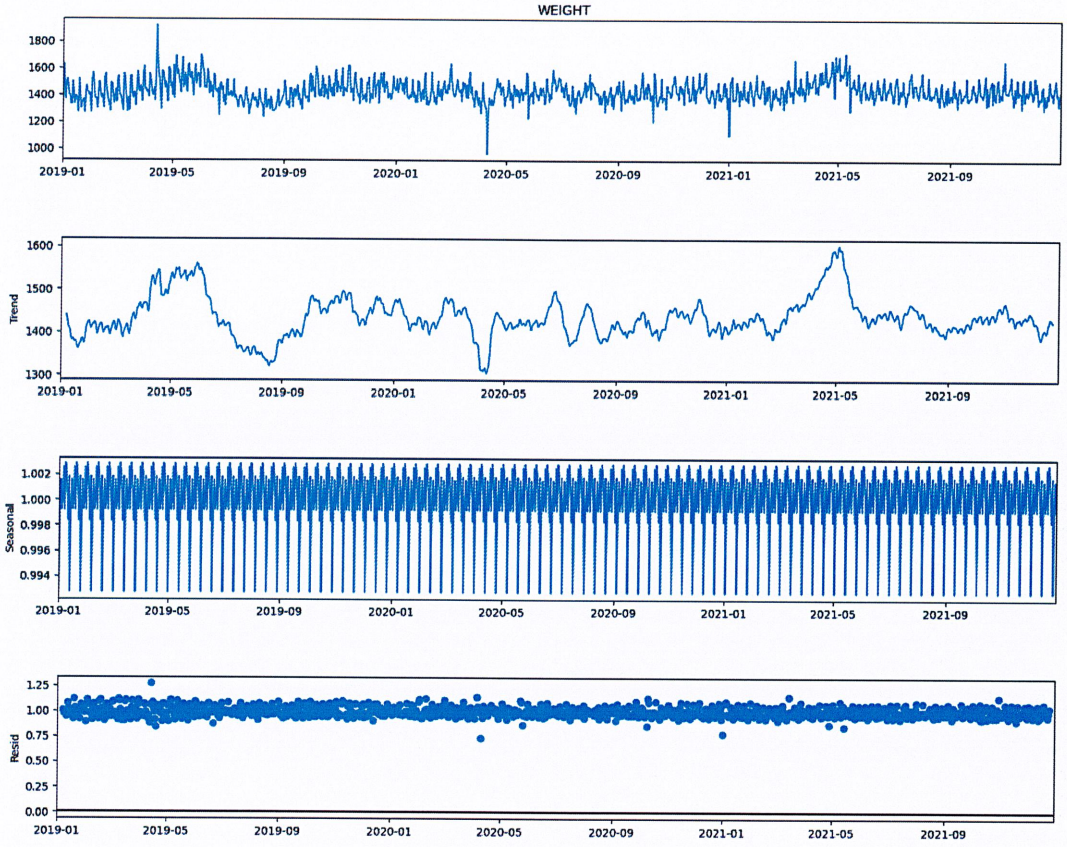


Fig. 25: Seasonal decomposition plots

The study attempted to determine the seasonal order for the data using the AUTOARIMA function. However, the results consistently returned an order of (0,0,0), indicating no seasonality in the data. To confirm this result, the traditional checking method was employed, which also led to a linear prediction without any fluctuations. Various seasonal orders were experimented with at different frequencies including monthly, quarterly, and annually.

The resulting SARIMA model is shown in Fig. 26, depicting the final prediction. The figure illustrates that the prediction produced by the SARIMA model is a straight line that does not exhibit any fluctuations or changes in the data. The line remains around the average of the data throughout the prediction period as shown in Fig. 27.

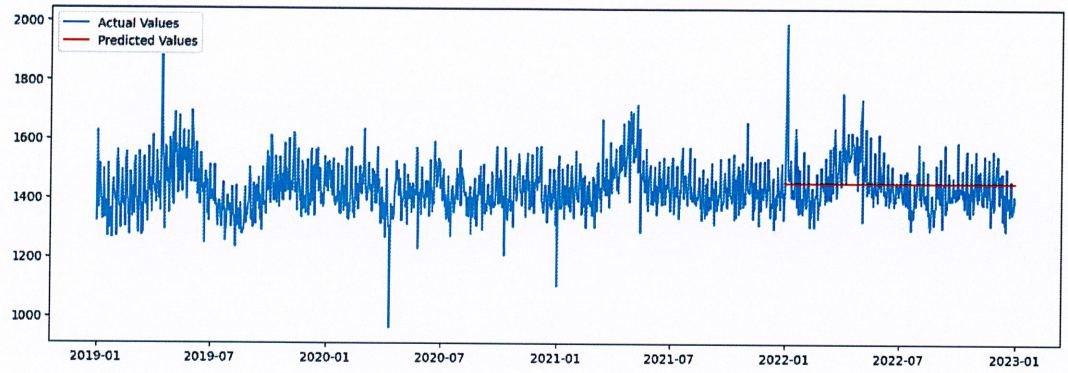


Fig. 26: Prediction of daily domestic waste generation using SARIMA

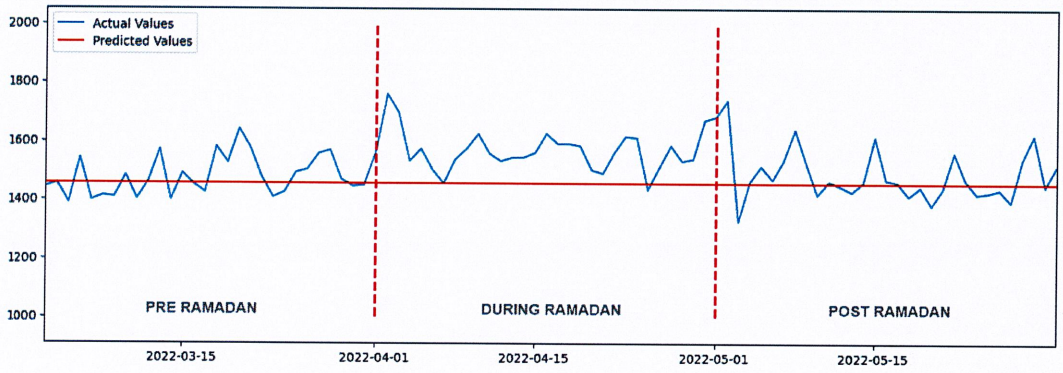


Fig. 27: Prediction of daily domestic waste generation pre, during and post-Ramadan using SARIMA

Table 9 presents the performance evaluation of the SARIMA model for predicting the daily domestic waste generation.

Table (9): SARIMA model prediction performance

SARIMA Model	MAE	RMSE	MAPE	R^2
Training	64.051	85.573	0.0441	0.002
Testing	77.988	102.331	0.0536	-0.0024

The evaluation of the model's performance on both training and testing data indicates that it has high MAE and RMSE values. Specifically, the testing data shows even higher MAE and RMSE values, which are 77.988 and 102.331 respectively. In this case, the SARIMA model has an R^2 of 0.002 for training and -0.0024 for testing. A low R^2 value suggests that the SARIMA model is not a good fit for the data.

4.1.5. Prophet Model Performance

The Prophet model offers a user-friendly implementation and does not demand numerous parameters. To explore the effect of changepoint prior scale parameters, several values [0.0001, 0.01, 0.05, 0.1, 0.5, 1.0] were tested, as presented in Table 10.

Table (10): MAE for different changepoint prior scale using Prophet

Changepoint_prior_scale	MAE
0.0001	55.8279
0.01	48.073
0.05	59.107
0.1	51.921
0.5	77.263
1.0	86.537

It shows that the best performance was achieved with a change point prior scale of 0.01, which resulted in MAE value of 48.073. Increasing or decreasing the change point prior scale from this optimal value led to a degradation in performance, with larger MAE values. This highlights the importance of selecting an appropriate change point prior scale for accurate forecasting using the Prophet model.

Fig. 28 and Fig. 29 depict the Prophet models' predictions for the daily domestic waste generation and daily domestic waste generation pre, during, and post-Ramadan, respectively.

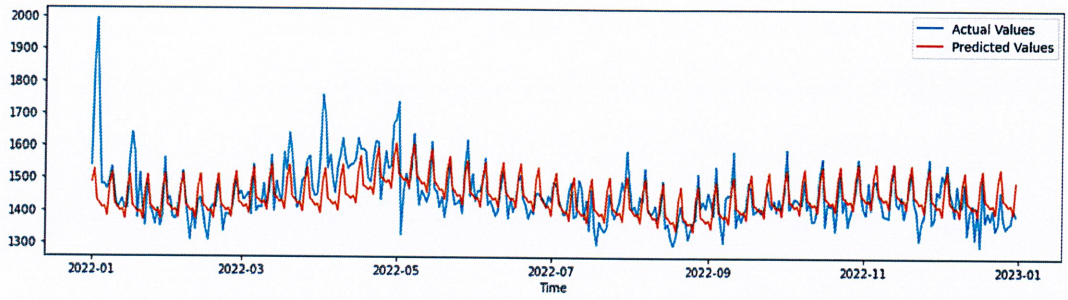


Fig. 28: Prediction of daily domestic waste generation using Prophet

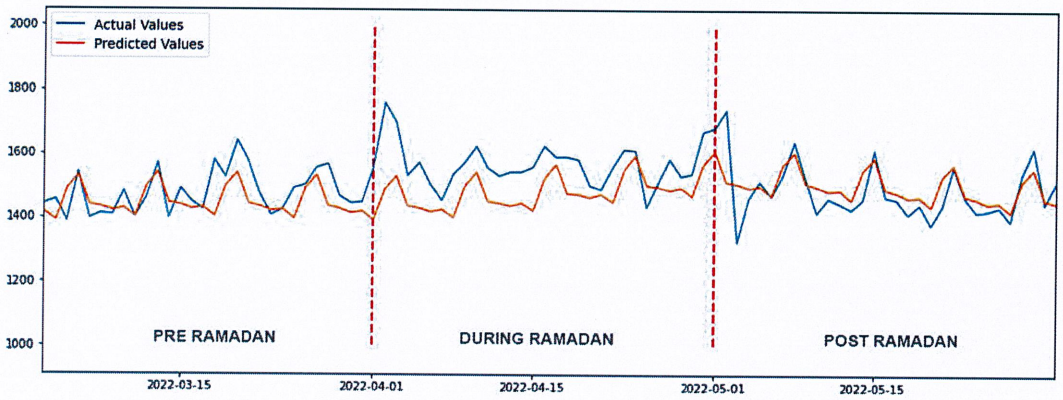


Fig. 29: Prediction of daily domestic waste generation pre, during and post-Ramadan using Prophet

The model exhibits a distinct pattern and endeavours to monitor the fluctuation of the data. However, during the month of Ramadan, when the trend of waste generation ascends, the model fails to accurately track the data. Moreover, this model exhibits poor performance in accurately forecasting values that deviate from the average by 150 tons. Table 11 presents the performance evaluation of the Prophet model for predicting the daily domestic waste generation.

Table (11): Prophet model prediction performance

Prophet Model	MAE	RMSE	MAPE	R^2
Training	44.763	61.146	0.032	0.431
Testing	47.505	70.569	0.031	0.318

On the testing data, the Prophet model achieved a higher MAE and RMSE values of 47.505 and 70.569, respectively, indicating less accurate

predictions compared to the training data. The MAPE value for testing data was 0.031, and the R^2 value was 0.318, suggesting that the model explains only 31.8% of the variability in the target variable in the testing data.

4.1.6. LSTM Model Performance

For the NN models (LSTM, BILSTM and GRU) different set of hyperparameters was tested for the number of hidden layers [1, 2, 3], the number of neurons [50, 100, 150] and forecasted windows [7, 14, 30, 40, 60, 90]. Table 12 shows the best hyperparameters achieved the best accuracy for the LSTM model.

Table (12): LSTM hyperparameters

Number of hidden layers	2
Number of neurons	100
Forecasted windows	40
Number of Epochs	50
Type of optimizer	Adam

Fig. 30 depicts the LSTM model training and test loss function on the domestic waste dataset.

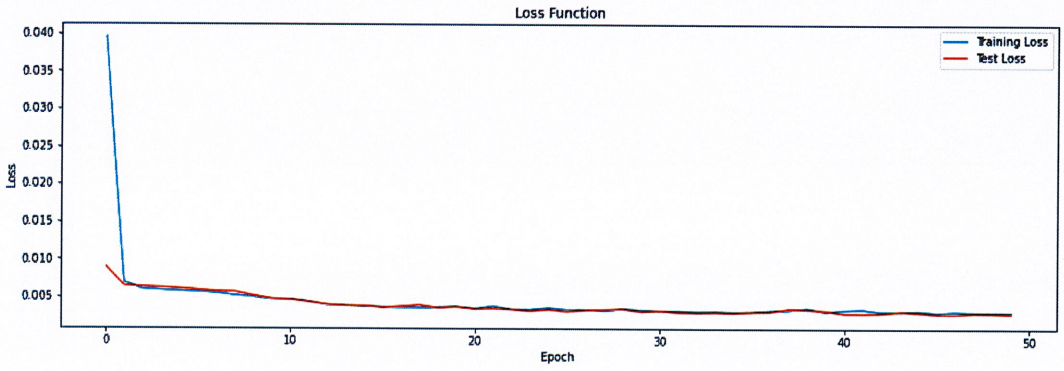


Fig. 30: LSTM model fitting and Test performance

Fig. 31 and Fig. 32 illustrate the LSTM model predictions for the daily domestic waste generation and daily domestic waste generation pre, during, and post-Ramadan, respectively.

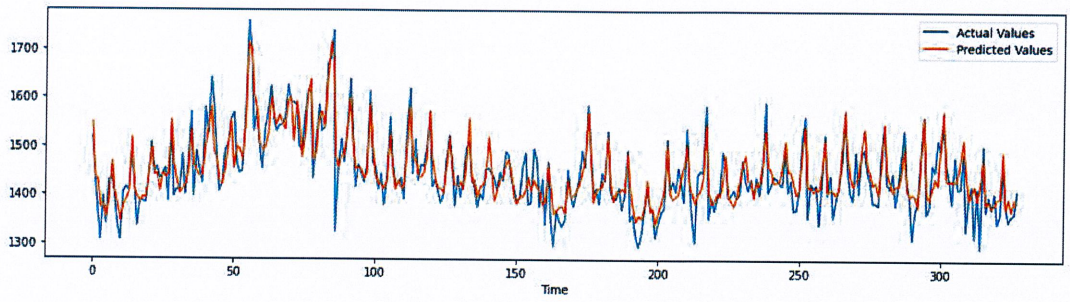


Fig. 31: Prediction of daily domestic waste generation using LSTM

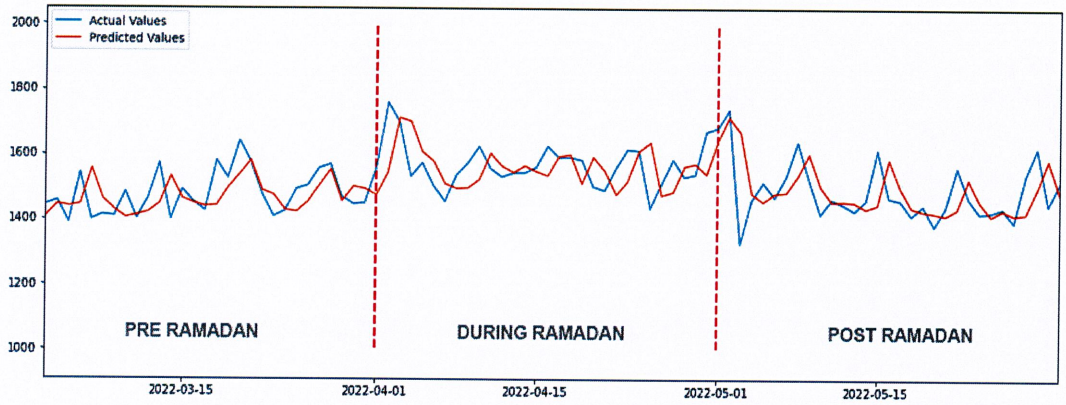


Fig. 32: Prediction of daily domestic waste generation pre, during and post-Ramadan using LSTM

As shown in the figure, the LSTM model exhibits a good level of performance in terms of peak values. In addition, the predicted trend closely follows the actual trend. However, its accuracy is comparatively lower for trough values. Concerning the Ramadan period, it is noticed that the trend of predicted values exhibited a slight shift towards the right-hand side.

Table 13 presents the performance evaluation of the LSTM model for predicting the daily domestic waste generation.

Table (13): LSTM model prediction performance

LSTM Model	MAE	RMSE	MAPE	R^2
Training	40.19	54.90	0.027	0.55
Testing	36.76	49.89	0.025	0.59

The LSTM model achieved low MAE and RMSE values of 36.8 and 49.89, respectively, compared with the previous models. The MAPE

value for testing data was 0.025, and the R^2 value was 0.59, suggesting that the model explains only 59% of the variability in the target variable in the testing data.

4.1.7. BILSTM Model Performance

Table 14 shows the best hyperparameters achieved the best accuracy for the BILSTM model.

Table (14): BILSTM hyperparameters

Number of hidden layers	2
Number of neurons	100
Forecasted windows	90
Number of Epochs	80
Type of optimizer	Adam

Fig. 33 depicts the BILSTM model training and testing loss function on the domestic waste dataset.

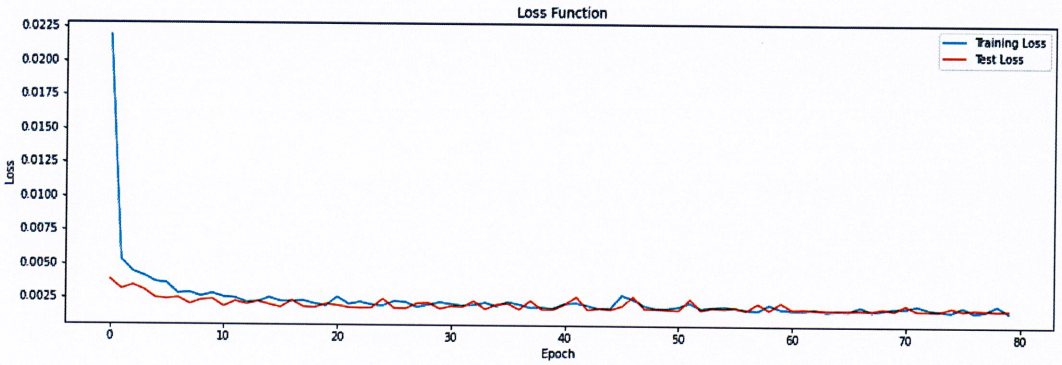


Fig. 33: BILSTM model fitting and test performance

Fig. 34 and Fig. 35 illustrate the BILSTM model predictions for the daily domestic waste generation and daily domestic waste generation pre, during, and post-Ramadan, respectively.

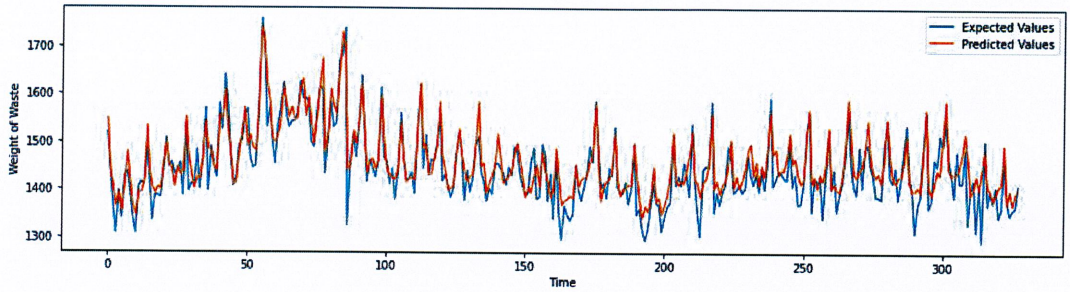


Fig. 34: Prediction of daily domestic waste generation using BILSTM

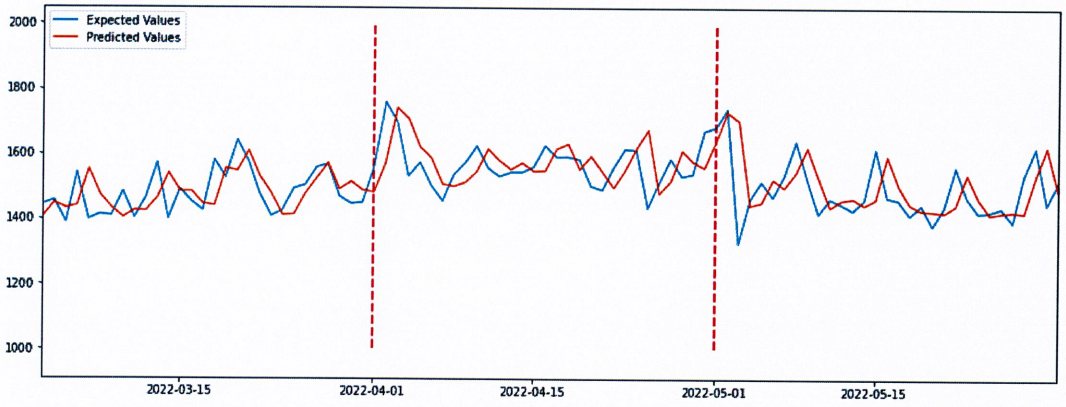


Fig. 35: Prediction of daily domestic waste generation pre, during and post-Ramadan using BILSTM

The figure demonstrates that the LSTM model performs exceptionally well, as the predicted trend aligns closely with the actual trend. Furthermore, the analysis of the Ramadan period indicates a minor rightward shift in the trend of the predicted values.

Table 15 presents the performance evaluation of the BILSTM model for predicting the daily domestic waste generation.

Table (15): BILSTM model prediction performance

BILSTM Model	MAE	RMSE	MAPE	R^2
Training	30.693	39.647	0.022	0.746
Testing	28.120	36.879	0.020	0.78

The model achieved MAE and RMSE values of 28.1 and 36.9, respectively, indicating that the predicted values are on average 28.1 units away from the actual values, and the model has a deviation of 36.9 units from the actual values. The MAPE value of 0.02 indicates that the

model's predictions have a low relative error. Additionally, the R^2 value of 0.78 suggests that the model explains 78% of the variation in the given dataset. Overall, the results of the BILSTM model perform better than the LSTM model in predicting values for the given dataset.

4.1.8. GRU Model Performance

Table 16 shows the best hyperparameters achieved the best accuracy for the GRU model.

Table (16): GRU hyperparameters

Number of hidden layers	3
Number of neurons	200
Forecasted windows	60
Number of Epochs	50
Type of optimizer	Adam

Fig. 36 depicts the GRU model training and testing loss function on the domestic waste dataset.

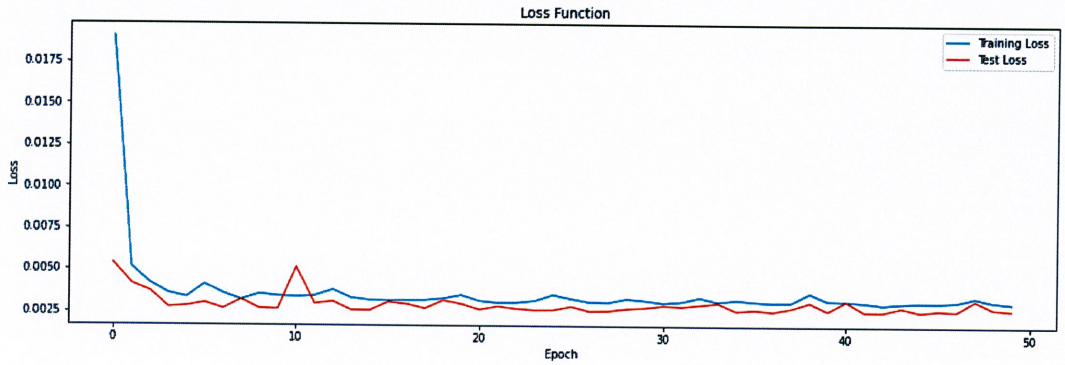


Fig. 36: GRU model fitting and validation performance

Fig. 37 and Fig. 38 shows the GRU model predictions for the daily domestic waste generation and daily domestic waste generation pre, during, and post-Ramadan, respectively.

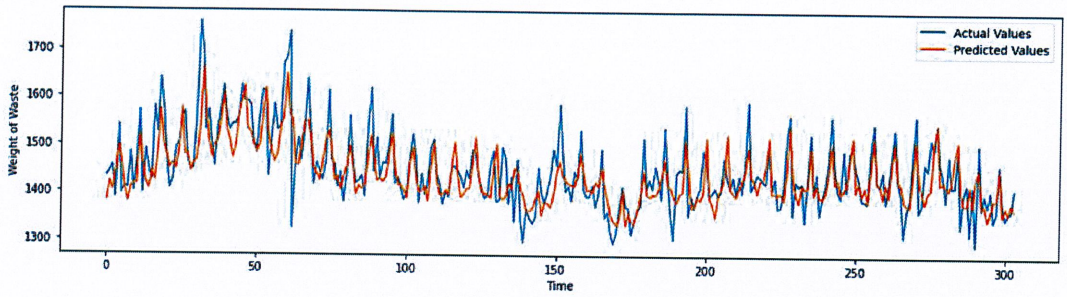


Fig. 37: Prediction of daily domestic waste generation using GRU

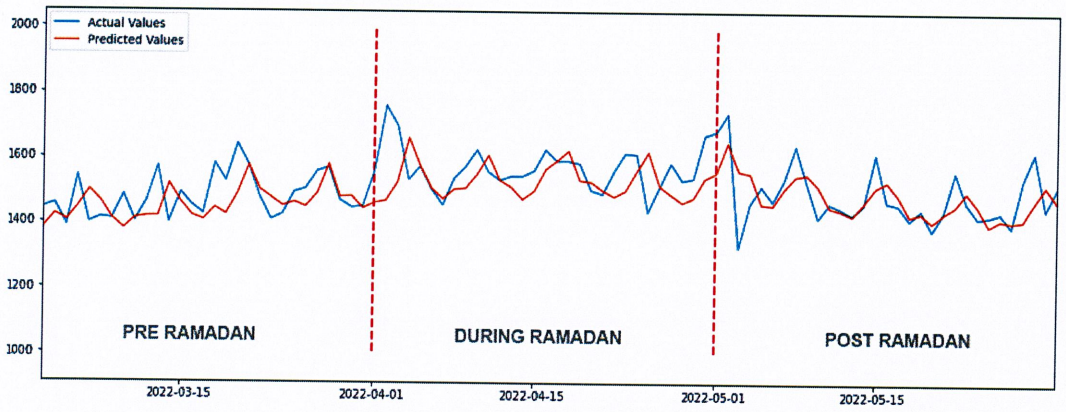


Fig. 38: Prediction of daily domestic waste generation pre, during and post-Ramadan using GRU

The GRU model exhibits a satisfactory performance in tracking the fluctuations of the actual data, although it shows relatively lower accuracy in predicting the peak values. Table 17 presents the performance evaluation of the GRU model for predicting the daily domestic waste generation.

Table (17): GRU model prediction performance

GRU Model	MAE	RMSE	MAPE	R^2
Training	40.164	55.484	0.028	0.545
Testing	38.452	51.462	0.026	0.568

The testing results of the GRU model show that the MAE value is 38.452 and the RMSE value is 35.11. The MAPE value is 0.026 and the R^2 value is 0.568.

4.1.9. Overall Models Performance Results

In this chapter, the outcomes of the experiments conducted on six different models for domestic waste datasets will be showcased. Table 18 compares the performance evaluation metrics of the six models.

Table (18): Overall models performance results

Rank	Model	MAE	RMSE	MAPE	R^2
1	BILSTM	28.120	36.879	0.020	0.78
2	LSTM	36.76	49.89	0.025	0.59
3	GRU	38.452	51.462	0.026	0.568
4	ARIMA	41.763	55.973	0.0291	0.526
5	Prophet	47.505	70.569	0.031	0.318
6	SARIMA	77.988	102.331	0.0536	-0.0024

According to the results, BILSTM achieved the best performance, while SARIMA obtained the lowest results among the six developed models. Overall, the results suggest that BILSTM is a promising model for forecasting domestic waste data and could be useful for waste management and planning purposes.

The superior performance of the three machine learning models (LSTM, BILSTM, and GRU) over the three statistical models (ARIMA, SARIMA, and Prophet) can be attributed to several factors. Firstly, ML models excel at capturing complex nonlinear relationships within the data. Unlike traditional statistical models that make linear assumptions, ML models can identify complex patterns and dependencies that may not be easily detected by statistical methods.

Secondly, ML models have the ability to automatically extract relevant features from the data. By leveraging advanced algorithms, ML models can reveal accurate information that enhances prediction accuracy. ML models, especially LSTM, BILSTM, and GRU, are specifically designed to deal with time dependencies effectively. They excel at analyzing time series data by capturing long-term dependencies and learning from sequences of data points. This is particularly important for accurate prediction of waste generation over time. In contrast, statistical models may not explicitly account for time dependencies, which can hinder their performance in capturing the dynamics of waste generation.

When comparing non-seasonal data, ARIMA models often outperform Prophet and SARIMA models for several reasons. Prophet is specifically designed to handle seasonal patterns and can automatically detect and model various seasonal components without manual specification. SARIMA also operates in a similar manner, making it less suitable.

ARIMA models capture trend information by incorporating autoregressive and moving average components. They are capable of capturing linear or polynomial trends based on the order of differencing. ARIMA models offer flexibility in capturing complex patterns and dependencies in the data, allowing them to handle different types of trends. Additionally, ARIMA models have a clear and interpretable structure, making it easier to understand the underlying dynamics. On the other hand, Prophet utilizes a piecewise linear model to capture abrupt and smooth changes in trends. While it automates many steps in the forecasting process, making it user-friendly for non-experts, the structure of the underlying model may be less interpretable compared to ARIMA.

4.2. Discussion

According to the data provided by the Askar landfill, domestic waste constitutes the highest percentage of total waste with 37%. This type of waste predominantly comprises food items that require safe disposal to prevent health complications. These findings emphasize the significance of researching domestic waste and its implications for designing and implementing effective waste management systems.

The results indicate an overall increase in the mean waste generated per year, except for the year 2020. The mean waste generated in 2020 was lower than in the subsequent years, which can be attributed to the outbreak of the COVID-19 pandemic. The pandemic led to widespread lockdowns and reduced economic activity, resulting in a decrease in waste generation. However, in the subsequent years, waste generation increased, as economic activity resumed, and people returned to their regular consumption patterns.

It has been observed that the quantity of domestic waste generated on a daily basis shows an increase, particularly during Ramadan, except for the year 2020. During Ramadan, people tend to prepare more food than they can consume, particularly during the iftar and suhoor meals. This is because Ramadan is considered a time of sharing meals with loved ones, and inviting guests and serving food is customary. However, creating awareness about reducing food waste and adopting proper storage methods can aid in minimizing food waste during Ramadan. Furthermore, it has been observed that food waste increases even further during weekends, as people are on vacation, and social events like invitations and banquets are more prevalent. Interestingly, in 2020, there was a decrease in food waste compared to other months due to the

cessation of these activities due to the closures and the prevention of gatherings due to the COVID-19 epidemic.

To address the limited research on the application of deep learning in predicting daily waste generation, particularly in Bahrain, a comprehensive study was conducted. The study aimed to evaluate and compare the efficacy of three machine learning models, namely LSTM, BILSTM, and GRU, along with three statistical models, ARIMA, SARIMA, and Prophet.

The results of the experiment indicate that the ARIMA model had MAE value of 41.763, RMSE value of 55.973, MAPE value of 0.0291, and an R^2 value of 0.526. In this analysis, coefficients with p-values of 0.717 and 0.207 are considered statistically insignificant at the 5% significance level. This implies that the null hypothesis (that the coefficient is equal to zero) cannot be rejected, and it cannot be concluded that these coefficients have a significant effect on the model. Conversely, a high z-statistic indicates that a coefficient is significantly different from zero. In this case, the constant term is found to be statistically significant due to its low p-value of 0.000, which confirms that it has a significant effect on the model. To summarize, the coefficients in the model are statistically significant, as indicated by T-statistics greater than 2 in magnitude and most P-values less than 0.05. These findings suggest that the hyperparameters selected for the model are appropriate. These results indicate that the model performs reasonably well in predicting daily domestic waste generation, with a low MAPE indicating accurate predictions on average. However, there is still room for improvement, as the MAE and RMSE values are relatively high, indicating that the model still makes some significant errors in predicting daily waste generation.

For the SARIMA model, the results of the experiment indicate MAE value of 77.988, RMSE value of 102.331, and MAPE value of 0.0536. In

this case, the model has an R^2 value of 0.002 for training and -0.0024 for testing. A low R^2 value for both training and testing suggests that the SARIMA model is not a good fit for the data. It means that the independent variables (in this case, the time series data) are not able to explain much of the variance in the dependent variable (the forecasted values). Additionally, the negative R^2 value for testing indicates that the model may be performing worse on new, unseen data than on the data it was trained on, which is not desirable for a predictive model. Therefore, the SARIMA model may need to be re-evaluated or revised to improve its performance on the given data.

The results of the experiment indicate that the Prophet model had MAE value of 47.505, RMSE value of 70.569, MAPE value of 0.031, and an R^2 value of 0.318. Prophet assumes that the relationship between the predictor variables and the target variable is linear or can be well approximated by a linear function. This assumption may not hold for some datasets, where more complex relationships exist between the predictor and target variables. Furthermore, it assumes that the underlying time series has a seasonal pattern that can be captured by Fourier terms. This assumption may not hold for all-time series, particularly those with irregular or non-periodic patterns, leading to inaccurate forecasts.

For the LSTM model, the results of the experiment indicate MAE value of 36.76, RMSE value of 49.89, MAPE value of 0.025, and an R^2 value of 0.59. LSTM models have a significant advantage in capturing long-term dependencies and temporal relationships in sequential data. This is due to their unique ability to maintain an internal state that retains critical information, enabling them to identify patterns and trends over extended time intervals. Additionally, LSTM models excel in handling irregular and missing data. Their ability to learn and remember dependencies over extended periods of time, coupled with their capacity

to handle complex sequential data, leads to more accurate predictions compared to other models.

The results of the experiment indicate that the BILSTM model had MAE value of 28.120, RMSE value of 36.879, a MAPE of 0.020, and R^2 value of 0.78. BILSTM is a model that enhances the performance of traditional LSTM models by incorporating an additional layer of hidden states in both forward and backward directions. This allows BILSTM to capture both past and future context and identify long-term dependencies in the data more effectively. By doing so, BILSTM can better integrate past and future contexts and produce more accurate results compared to traditional LSTM models. Additionally, BILSTM can help to reduce overfitting, which further improves its accuracy. Therefore, the experimental results suggest that BILSTM outperforms LSTM in terms of accuracy.

For the GRU model, the results of the experiment indicate MAE value of 38.452, RMSE value of 51.462, MAPE value of 0.026, and an R^2 value of 0.57. In this particular case, LSTM outperforms GRU due to its ability to maintain a dedicated memory cell that can store information for extended periods. Additionally, LSTM has gateways that can selectively add or remove information, which enhances its capacity to capture complex long-term dependencies. On the other hand, GRU lacks a dedicated memory cell and only has two gates, which limits its capacity to handle complex long-term dependencies. Therefore, LSTM is better suited for this task compared to GRU.

In comparison to other studies, Marandi et al. (2016) used monthly data for 2015 and 2016 to find the ARIMA (2, 1, 1), while Mohamed et al. (2011) used monthly data from 2005 to 2010 to determine the ARIMA (1, 1, 1). However, our study achieved the best results using ARIMA (7, 0, 7). In terms of LSTM, Niu et al. (2020) reported R^2 value of 0.92, and

RMSE and MAPE values of 935.08 and 114.36, respectively. Although their R^2 accuracy was good, their RMSE and MAPE were very high. On the other hand, our study achieved an accuracy of 49.89, with an RMSE and MAPE value of 0.025, which indicates better performance. There is a lack of research on time series analysis for predicting daily generated waste, as most existing studies have focused on monthly or annual data.

Chapter Five

Conclusion, Recommendations and Future Work

5.1. Conclusion

5.2. Recommendations

5.3. Future Work

Chapter Five
Conclusion, Recommendations and Future Work

5.1. Conclusion

This study examined the link between Ramadan fasting and the amount of solid waste produced by households in the Kingdom of Bahrain. The results show a strong relationship between waste generation and the month of Ramadan, as waste generation significantly increased during this period. On average, the study found that Ramadan witnessed the highest waste generation, with an average of 1570 tons per day, in contrast to the average of 1450 tons per day observed during the rest of the year. This finding emphasizes the impact of cultural and behavioral factors associated with Ramadan, such as increased consumption and food waste, on the overall waste generation patterns in Bahrain. This increase in waste is attributed to various factors, primarily excessive food preparations and social activities such as Ramadan Ghabga, particularly during the weekends when food banquets are prevalent, resulting in a daily waste amount ranging from 1600 to 1700 tons. Understanding and addressing this relationship is crucial for developing targeted waste management strategies and initiatives during Ramadan to effectively manage and mitigate the environmental impact of waste generation during this period.

In 2020, there was a decrease in the amount of waste generated compared to previous and following years (2019, 2021, and 2022). This decrease is possibly due to the COVID-19 pandemic, which caused

lockdowns and less economic activity. However, in 2021, people resumed their usual consumption habits, resulting in higher waste generation. Furthermore, the waste generation rate rose significantly on weekends, reaching around 1650 tons per day and this could be attributed to numerous social gatherings and events.

Results from the experiments demonstrated that machine learning models can accurately and efficiently predict domestic waste generation in Bahrain. Among the models tested, the BILSTM model exhibited superior performance compared to ARIMA, SARIMA, Prophet, LSTM, and GRU. The BILSTM model achieved MAE value of 28.120, RMSE value of 36.879, MAPE value of 0.020, and an R^2 value of 0.78. Furthermore, the SARIMA model confirmed the absence of seasonal patterns in the dataset. To ensure the BiLSTM model's generalization, several considerations need to be considered. Firstly, training the model on a diverse, representative, and larger dataset is crucial, encompassing various patterns, trends, and variations encountered in real-world scenarios. Secondly, adequate training data should cover different trends, and other influences on the time series behavior. Thirdly, regularization techniques like dropout and regularization penalties can be applied to avoid overfitting and promote more robust representations. Fourthly, optimizing hyperparameters through techniques like cross-validation can strike a balance between the model's capacity to capture complexity and avoiding overfitting.

By following these practices, the BiLSTM model can effectively generalize for accurate predictions on unseen data. The diverse training dataset, randomization, regularization techniques, and careful hyperparameter tuning all contribute to the model's ability to capture and understand underlying patterns and dynamics in time series data. This

enables the model to perform well in real-world scenarios, making it a reliable tool for time series prediction and analysis.

5.2. Recommendations

During peak times, effective waste management is crucial for maintaining a clean and healthy environment. To achieve this, the frequency of waste collection should be increased to prevent overflowing bins and rubbish. In addition, it is important to schedule collection times, ensuring timely waste collection and reducing the risk of littering and attracting pests. Additional containers should be provided, particularly in areas with high-density waste generation, to distribute waste and prevent overflowing.

Educating the population on proper waste management practices, including sorting waste and disposing of it correctly, is also essential in reducing pollution and improving waste collection efficiency. It is recommended to utilize smart waste management systems to ensure timely waste collection through sensors that detect full bins and alert waste management teams to empty them. The implementation of such recommendations during peak times can effectively improve the management of waste and maintain a clean and healthy environment.

In addition to landfill disposal, alternative waste management methods, such as composting or recycling, should be considered for food waste. Composting is an efficient method that can provide significant environmental benefits, such as reducing GHG emissions and preserving soil quality. By diverting food waste away from landfills and towards these alternative methods, we can reduce our environmental impact and create a more sustainable waste management system.

5.3. Future Work

In future work, it is recommended that additional explanatory variables be included in testing new algorithms. These variables can include population, gross domestic product (GDP), and carbon dioxide emissions. By considering these factors, a more comprehensive analysis and evaluation of algorithms can be performed, which may ultimately improve accuracy and effectiveness.

Population is an important variable to consider because it provides insights into the size and distribution of the population in a given region or country. Population can have a significant impact on various aspects of a system or phenomenon under study. GDP, another important variable, reflects the economic performance of a country or region. It measures the total value of goods and services produced within a specified time frame. GDP provides valuable information about the economic well-being, productivity, and development of a nation. Carbon dioxide emissions have a direct relationship with waste generation, and are a related environmental impact variable, also worth considering. With the growing concern about climate change and sustainability, understanding the relationship between algorithms and carbon emissions can be beneficial.

References

Abbasi, M., & El Hanandeh, A. (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management, 56*, 13–22. doi:10.1016/j.wasman.2016.05.018

Abbasi, M., Abduli, M. A., Omidvar, B., & Baghvand, A. (2013). Forecasting municipal solid waste generation by hybrid support vector machine and partial least square model. *International Journal of Environmental Research, 7*(1), 27-38. doi: 10.22059/IJER.2012.583

Abbasi, M., Abduli, M. A., Omidvar, B., & Baghvand, A. (2014). Results uncertainty of support vector machine and hybrid of wavelet transform-support vector machine models for solid waste generation forecasting. *Environmental Progress and Sustainable Energy, 33*(1). doi:10.1002/ep.11747

Abbasi, M., Rastgoo, M. N., & Nakisa, B. (2019). Monthly and seasonal modeling of municipal waste generation using radial basis function neural network. *Environmental Progress and Sustainable Energy, 38*(3), 1-10. doi:10.1002/ep.13033

Abdallah, M., Abu Talib, M., Feroz, S., Nasir, Q., Abdalla, H., & Mahfood, B. (2020). Artificial intelligence applications in solid waste management: A systematic research review. *Waste Management, 109*, 231-246. doi:10.1016/j.wasman.2020.04.057

Abdallah, M., Warith, M., Narbaitz, R., Petriu, E., & Kennedy, K. (2011). Combining fuzzy logic and neural networks in modeling landfill gas production. *Engineering and Technology, 78*, 559-565. doi:10.1.1.1017.8362.

Abdella Ahmed, A. K., Ibraheem, A. M., & Abd-Ellah, M. K. (2022). Forecasting of municipal solid waste multi-classification by using time-series deep learning depending on the living standard. *Results in Engineering, 16*, 1-8.

Abdulredha, M., Kot, P., Al Khaddar, R., Jordan, D., & Abdulridha, A. (2020). Investigating municipal solid waste management system performance during the Arba'een event in the city of Kerbala, Iraq. *Environment, Development and Sustainability, 22*(2), 1431-1454. doi:10.1007/s10668-018-0256-2

Abu Qdais, H., Bani Hani, K., & Shatnawi, N. (2010). Modeling and optimization of biogas production from a waste digester using artificial neural network and

genetic algorithm. *Resources, Conservation and Recycling, 54*(6), 359-363. doi:10.1016/j.resconrec.2009.08.012

Abunama, T., Othman, F., Ansari, M., & El-Shafie, A. (2016). Leachate generation rate modeling using artificial intelligence algorithms aided by input optimization method for an MSW landfill. *Environmental Science and Pollution Research, 26*(4), 3368-3381. doi:10.1007/s11356-018-3749-5

Ahsan, M., Mahmud, M., Saha, P., Gupta, K., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *International Journal of Emerging Technologies in Learning, 9*(3), 1-17. doi:10.3390/technologies9030052

Al Sheikh, F. (2018). The month of fasting.. not the month of food. *Alwatan News.* Retrieved from https://alwatannews.net/Opinion/article/772389/

Alayam. (2018). *Minister of works: 195 thousand tons of food waste annually.* Alayam.

Ali-Arous, S., Capone, R., Debs, P., Haddadi, Y., El Bilali, H., Bottalico, F., & Mohamed, H. (2017). Exploring household food waste issue in Algeria. *AGROFOR International Journal, 2*(1), 55-67. doi: 10.7251/AGRENG1701055A

Alsebaei, A. F. (2014). *Solid waste management and recycling during Hajj pilgrimage in Mina.* The University of Leeds.

Alzamora, B. R., Barros, R. T., de Oliveira, L. K., & Gonçalves, S. S. (2022). Forecasting and the influence of socioeconomic factors on municipal solid waste generation: A literature review. *Environmental Development, 44*, 100734. doi:10.1016/j.envdev.2022.100734.

Asare, W., Oduro-Kwarteng, S., Donkor, E. A., & Rockson, M. A. (2020). Recovery of municipal solid waste recyclables under different incentive schemes in Tamale, Ghana. *Sustainability (Switzerland), 12*(23), 1-19. doi:10.3390/su12239869

Azadi, S., & Karimi-Jashni, A. (2016). Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran. *Waste Management, 48*, 14-23. doi:10.1016/j.wasman.2015.09.034

Ben Ismail, H., Jribi, S., Doggui, D., & Debbabi, H. (2022). COVID-19 lockdown impact on Tunisian consumer behaviors and awareness related to food

consumption and wastage during celebrations: Case of study "Ramadan month". *The North African Journal of Food and Nutrition Research, 6*, 46-54. doi:10.51745/najfnr.6.13.46-54

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief, 29*, 1-4. doi:10.1016/j.dib.2020.105340

Beyaztas, U., & Shang, H. L. (2022). Robust bootstrap prediction intervals for univariate and multivariate autoregressive time series models. *Journal of Applied Statistics, 49*(5), 1-25. doi:10.1080/02664763.2020.1856351

Bi, W., Lu, W., Zhao, Z., & Webster, C. J. (2022). Combinatorial optimization of construction waste collection and transportation: A case study of Hong Kong. *Resources, Conservation and Recycling, 179*. doi:10.1016/j.resconrec.2021.106043

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control.* New Jersey, USA: John Wiley & Sons.

Buzby, J. C., Hyman, J., & Stewart, H. (2014). *The estimated amount, value, and calories of postharvest food losses at the retail and consumer levels in the United States.* United States: Economic Research Service, U.S. Department of Agriculture.

Casella, G., & L. Berger, R. (2002). *Statistical Inference.* Duxbury Press, 2nd edition.

Cha, G. W., Kim, Y. C., Moon, H. J., & Hong, W. H. (2017). New approach for forecasting demolition waste generation using chi-squared automatic interaction detection (CHAID) method. *Journal of Cleaner Production, 168*, 375–385. doi:10.1016/j.jclepro.2017.09.025

Chhay, L., Reyad, M. A., Suy, R., Islam, M. R., & Mian, M. M. (2018). Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting. *Journal of Material Cycles and Waste Management, 20*(3), 1761–1770. doi:10.1007/s10163-018-0743-4

ClimateFaith. (2017). *Green Ramadan Guide Bahrain.* Manama: The Global Shapers Manama Hub.

Contreras, J., Espínola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems, 18*(3), 1014-1020. doi:10.1109/TPWRS.2002.804943

Dunkel, J., Dominguez, D., Borzdynski, Ó. G., & Sánchez, Á. (2022). Solid Waste Analysis Using Open-Access Socio-Economic Data. *Sustainability (Switzerland), 14*(3), 1-24. doi:10.3390/su14031233

FAO. (2013). *Food Wastage Footprint: Impacts on Natural Resources.* Food and Agriculture Organization of the United Nations (FAO). Retrieved from https://www.fao.org/3/i3347e/i3347e.pdf

FAO. (2018). *FAO Publications Catalogue 2018.* Italy: Food and Agriculture Organization of the United Nations (FAO). Retrieved from https://wedocs.unep.org/20.500.11822/36567

FAO. (2019). *The state of food and agriculture 2019. Moving forward on food loss and waste reduction.* Rome: Food and Agriculture Organization of the United Nations.

Golbaz, S., Nabizadeh, R., & Sajadi, H. S. (2019). Comparative study of predicting hospital solid waste generation using multiple linear regression and artificial intelligence. *Journal of Environmental Health Science and Engineering, 17*(1), 1-11. doi:10.1007/s40201-018-00324-z

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., & Meybeck, A. (2011). *Global food losses and food waste: extent, causes and prevention.* Food and Agriculture Organization (FAO).

Hannan, M. A., Arebey, M., Begum, R. A., & Basri, H. (2012). An automated solid waste bin level detection system using a gray level aura matrix. *Waste Management, 32*(12), 2229-2238. doi:10.1016/j.wasman.2012.06.002

Hota, H., Handa, R., & Shrivas, A. (2017). Time Series Data Prediction Using Sliding Window Based RBF Neural Network. *International Journal of Computational Intelligence Research, 13*(5), 1145-1156.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice.* Otexts; 3rd ed. edition .

Jalili Ghazi Zade, M., & Noori, R. (2008). Prediction of municipal solid waste generation by use of artificial neural network: A case study of Mashhad. *International Journal of Environmental Research, 2*(1), 13-22. doi:10.22059/IJER.2010.170

Jassim, M. S., Coskuner, G., & Zontul, M. (2022). Comparative performance analysis of support vector regression and artificial neural network for prediction of

municipal solid waste generation. *Waste Management and Research, 40*(2), 1-10. doi:10.1177/0734242X211008526

Kannangara, M., Dua, R., Ahmadi, L., & Bensebaa, F. (2018). Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Management, 74*, 3-15. doi:10.1016/j.wasman.2017.11.057

Karak, T., Bhagat, R. M., & Bhattacharyya, P. (2012). Municipal Solid Waste Generation, Composition, and Management: The World Scenario. *Critical Reviews in Environmental Science and Technology, 42*(15), 1509-1620. doi:10.1080/10643389.2011.569871

Korstanje, J. (2021). *Advanced Forecasting with Python.* Apress.

Król, A., Nowakowski, P., & Mrówczyńska, B. (2016). How to improve WEEE management? Novel approach in mobile collection with application of artificial intelligence. *Waste Management, 50*, 222-233. doi:10.1016/j.wasman.2016.02.033

Kumar, A., & Agrawal, A. (2020). Recent trends in solid waste management status, challenges, and potential for the future Indian cities – A review. *Current Research in Environmental Sustainability, 2*, 1-19. doi:10.1016/j.crsust.2020.100011

Kumar, A., Samadder, S. R., Kumar, N., & Singh, C. (2018). Estimation of the generation rate of different types of plastic wastes and possible revenue recovery from informal recycling. *Waste Management, 79*, 781–790. doi:10.1016/j.wasman.2018.08.045

Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., . . . Cunsolo, S. (2017). Plastic waste inputs from land into the ocean. *Science, 347*(6223), 768-771. doi:10.1126/science.1260352

Liang, S., & Gu, Y. (2021). A deep convolutional neural network to simultaneously localize and recognize waste types in images. *Waste Management, 126*, 247–257. doi:10.1016/j.wasman.2021.03.017

Marandi, F., & Fatemi Ghomi, S. (2016). Time series forecasting and analysis of municipal solid waste generation in Tehran city., (pp. 1-5). doi:10.1109/INDUSENG.2016.7519343

Martins, A. M., & Cró, S. (2021). The impact of tourism on solid waste generation and management cost in madeira island for the period 1996–2018. *Sustainability (Switzerland), 13*(9), 1-16. doi:10.3390/su13095238

Milojkovic, J., & Litovski, V. (2008). Comparison of Some ANN Based Forecasting Methods Implemented on Short Time Series. *9th Symposium on Neural Network Applications in Electrical Engineering, 9*, 1-4. doi:10.1109/NEUREL.2008.4685606

MOCCAE. (2021, 4 14). Ministry of Climate Change and Environment Launches Sustainable Ramadan Campaign. *Ministry of Climate Change and Environment* .

MUN. (2022). Retrieved from Ministry of Municipal Affairs and Agriculture: https://www.mun.gov.bh/

Navarro-Esbrí, J., Diamadopoulos, E., & Ginestar, D. (2002). Time series analysis and forecasting techniques for municipal solid waste management. *Resources, Conservation and Recycling, 35*(3), 1-14. doi:10.1016/S0921-3449(02)00002-2

Niu, D., Wu, F., Dai, S., He, S., & Wu, B. (2021). Detection of long-term effect in forecasting municipal solid waste using a long short-term memory neural network. *Journal of Cleaner Production, 290*, 1-42. doi:10.1016/j.jclepro.2020.125187

Nizami, A. S., Shahzad, K., Rehan, M., Ouda, O. K., Khan, M. Z., Ismail, I. M., . . . Demirbas, A. (2017). Developing waste biorefinery in Makkah: A way forward to convert urban waste into renewable energy. *Applied Energy, 186*(1), 1-8. doi:10.1016/j.apenergy.2016.04.116

Noori, R., Abdoli, M. A., Ameri Ghasrodashti, A., & Jalili Ghazizade, M. (2009). Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of mashhad. *Environmental Progress and Sustainable Energy, 28*(2), 249-258. doi:10.1002/ep.10317

Noori, R., Abdoli, M. A., Ghazizade, M. J., & Samieifard, R. (2009). Comparison of neural network and principal component-regression analysis to predict the solid waste generation in Tehran. *Iranian Journal of Public Health, 38*(1), 74-84.

Noori, R., Karbassi, A., & Salman Sabahi, M. (2010). Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management, 91*(3), 767-771. doi:10.1016/j.jenvman.2009.%2010.007.

NRDC. (2017). *How America is losing up to 40 percent of its food from farm to fork to landfill.* National Resources Defense Council . Retrieved from https://www.nrdc.org/sites/default/files/wasted-food-IP.pdf

Oguz-Ekim, P. (2021). Machine Learning Approaches for Municipal Solid Waste Generation Forecasting. *Environmental Engineering Science, 38*(6), 489-499. doi:10.1089/ees.2020.0232

Oliveira, V., Sousa, V., & Dias-Ferreira, C. (2019). Artificial neural network modelling of the amount of separately-collected household packaging waste. *Journal of Cleaner Production, 210*, 401-409. doi:10.1016/j.jclepro.2018.11.063

Owusu-Sekyere, E., Harris, E., & Bonyah, E. (2013). Forecasting and Planning for Solid Waste Generation in the Kumasi Metropolitan Area of Ghana: An ARIMA Time Series Approach. *International Journal of Sciences, 2*(4), 71-83.

Ozkaya, B., Demir, A., & Bilgili, M. S. (2007). Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. *Environmental Modelling and Software, 22*(6), 815-822. doi:10.1016/j.envsoft.2006.03.004

Pandey, D. S., Das, S., Pan, I., Leahy, J. J., & Kwapinski, W. (2016). Artificial neural network based modelling approach for municipal solid waste gasification in a fluidized bed reactor. *Waste Management, 58*, 202-213. doi:10.1016/j.wasman.2016.08.023

Popli, K., Park, C., Han, S. M., & Kim, S. (2021). Prediction of solid waste generation rates in urban region of Laos using socio-demographic and economic parameters with a multi linear regression approacha. *Sustainability (Switzerland), 13*(6), 1-15. doi:10.3390/su13063038

Prophet. (2018). Retrieved from facebook: https://facebook.github.io/prophet/docs/quick_start.html

Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning.* Packt Publishing, 2nd edition.

Rimaityte, I., Ruzgas, T., Denafas, G., Račys, V., & Martuzevicius, D. (2012). Application and evaluation of forecasting methods for municipal solid waste generation in an Eastern-European city. *Waste Management and Research, 30*(1), 89-98. doi:10.1177/0734242X10396754

SAGO. (2019). *Saudi FLW Baseline: Food Loss & Waste Index in Kingdom of Saudi Arabia.* Saudi Grains Organization.

SAKK, A. H. (1975). Fasting in Islam. *Journal of the American Dietetic Association, 67*(1), 17-21. doi:10.1016/S0002-8223(21)14634-6

Shamshiry, E., Mokhtar, M., Abdulai, A. M., Komoo, I., & Yahaya, N. (2014). Combining artificial neural network-genetic algorithm and response surface method to predict waste generation and optimize cost of solid waste collection and transportation process in Langkawi Island, Malaysia. *Malaysian Journal of Science, 33*(2), 118-140. doi:10.22452/mjs.vol33no2.1

Shamshiry, E., Nadi, B., Bin Mokhtar, M., Komoo, I., Hashim, H. S., & Ahya, N. Y. (2011). Forecasting generation waste using artificial neural networks. *Proceedings of the 2011 International Conference on Artificial Intelligence, ICAI 2011*, (pp. 1-8).

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *IEEE International Conference on Big Data, Big Data* , (pp. 3285-3292). doi:10.1109/BigData47090.2019.9005997

Taylor, S. J., & Letham, B. (2017). Forecasting at Scale. *PeerJ Preprints*, 1-25. doi:10.7287/peerj.preprints.3190v2

Teng, S. Y., Máša, V., Touš, M., Vondra, M., Lam, H. L., & Stehlík, P. (2022). Waste-to-energy forecasting and real-time optimization: An anomaly-aware approach. *Renewable Energy, 181*, 142-155. doi:10.1016/j.renene.2021.09.026

tensorflow. (2022). *Tuning Hyperparameters with Keras Tuner and TensorFlow 2.0.* Retrieved from tensorflow: https://www.tensorflow.org/tutorials/keras/keras_tuner

UNEP. (2021). *Food waste index report 2021.* United Nations Environment Programme (UNEP).

Wei, Y., Xue, Y., Yin, J., & Ni, W. (2013). Prediction of municipal solid waste generation in china by multiple linear regression method. *International*

Journal of Computers and Applications, 35(3), 136–140. doi:10.2316/Journal.202.2013.3.202-3898

Wilson, D., Rodic, L., Scheinberg, A., Velis, C., & Alabaster, G. (2012). Comparative analysis of solid waste management in 20 cities. *Waste Management & Research, 30*(3), 237-254. doi:10.1177/0734242X12437569

WorldBank. (2018). *The global problem of food waste: Recent research and policy approaches.* The World Bank.

WorldBank. (2018). *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050.* The World Bank. doi:10.1596/978-1-4648-1329-0.

WRAP. (2019). *The food we waste.* Waste & Resources Action Programme. Retrieved from https://wrap.s3.amazonaws.com/the-food-we-waste-executive-summary.pdf

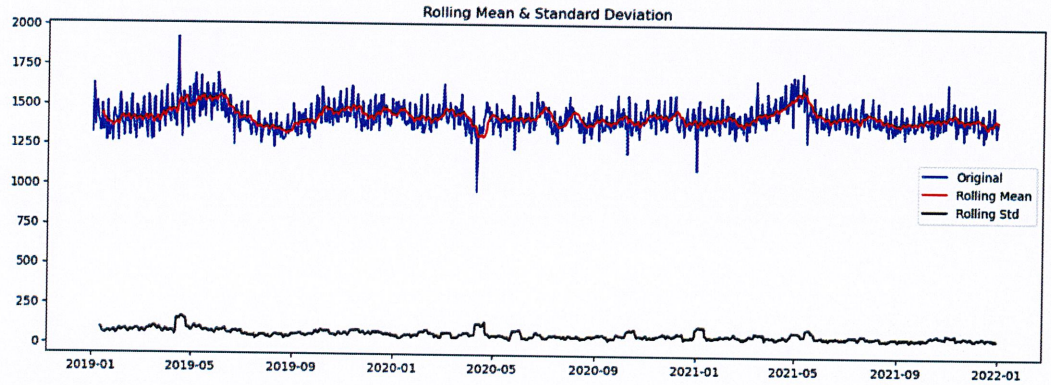
Wu, Z., Fan, H., & Liu, G. (2015). Forecasting Construction and Demolition Waste Using Gene Expression Programming. *Journal of Computing in Civil Engineering, 29*(5), 1-8. doi:10.1061/(asce)cp.1943-5487.0000362

Younes, M. K., Basri, N. E., Nopiaha, Z. M., Basri, H., Abushammala, M. F., & Maulud, K. N. (2015). Integrating approach to size and site at a sanitary landfill in Selangor state, Malaysia. *Environmental Engineering Research, 20*(3), 268-276. doi:10.4491/eer.2015.0054

Appendices

Table (19): Waste type quantity (tons)

WASTE_TYPE	Sum of tons
Domestic (DOM)	2101207
Building waste (BLD)	1497728
Commercial (COM)	1068017
Garden (GDN)	417941.9
Household bulky (HBW)	148122.6
Liquid waste (LIQU)	135326.3
Industrial waste (IND)	120906.6
Animal waste (ANM)	53677.6
TYRES	37178.22
SWAT	10072.59
WATER	8412.618
Expired food (FOD)	4687.57
BLOOD	2062.286




```
Results of Dickey-Fuller Test:
Test Statistic                  -4.00
p-value                          0.00
#Lags Used                      22.00
Number of Observations Used   1073.00
Critical Value (1%)             -3.44
Critical Value (5%)             -2.86
Critical Value (10%)            -2.57
dtype: float64
```

Fig. 39: Results of Dickey-Fuller test

Dep. Variable:	WEIGHT	No. Observations:	1461
Model:	ARIMA(7, 0, 7)	Log Likelihood	-7997.535
Date:	Tue, 07 Feb 2023	AIC	16027.070
Time:	12:44:32	BIC	16111.660
Sample:	01-01-2019 - 12-31-2022	HQIC	16058.625
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	1438.1978	46.321	31.048	0.000	1347.410	1528.986
ar.L1	-0.1001	0.047	-2.145	0.032	-0.192	-0.009
ar.L2	0.2251	0.042	5.297	0.000	0.142	0.308
ar.L3	-0.3295	0.034	-9.700	0.000	-0.396	-0.263
ar.L4	0.3437	0.027	12.734	0.000	0.291	0.397
ar.L5	-0.3106	0.028	-11.208	0.000	-0.365	-0.256
ar.L6	0.1982	0.037	5.396	0.000	0.126	0.270
ar.L7	0.9261	0.043	21.363	0.000	0.841	1.011
ma.L1	0.3567	0.049	7.294	0.000	0.261	0.453
ma.L2	-0.0121	0.033	-0.362	0.717	-0.077	0.053
ma.L3	0.5194	0.034	15.414	0.000	0.453	0.585
ma.L4	-0.1006	0.030	-3.320	0.001	-0.160	-0.041
ma.L5	0.4330	0.025	17.625	0.000	0.385	0.481
ma.L6	0.0520	0.041	1.262	0.207	-0.029	0.133
ma.L7	-0.7017	0.032	-21.710	0.000	-0.765	-0.638
sigma2	3416.7516	77.251	44.229	0.000	3265.342	3568.161

Ljung-Box (L1) (Q):	0.15	Jarque-Bera (JB):	2714.17
Prob(Q):	0.70	Prob(JB):	0.00
Heteroskedasticity (H):	1.03	Skew:	0.32
Prob(H) (two-sided):	0.73	Kurtosis:	9.65

Fig. 40: ARIMA results for order (7,0,7)

الملخص

إنتاج النفايات المنزلية الزائدة خلال الاحتفالات الدينية له آثار بيئية سلبية. يوفر شهر رمضان المبارك فرصة لدراسة العلاقة بين نمط استهلاك الغذاء وتوليد النفايات. لذلك فإن الهدف من هذا البحث هو معرفة العلاقة بين صيام رمضان وتوليد النفايات الصلبة المنزلية في البحرين للفترة بين 2019-2022. علاوة على ذلك ، تطور هذه الدراسة نماذج متسلسلة زمنية دقيقة باستخدام مناهج التعلم الآلي العميقة للتنبؤ بتوليد النفايات المنزلية لفترات ما قبل وأثناء وبعد رمضان. كشفت نتائج البحث أن البحرينيين يولدون خلال شهر رمضان أكبر كمية نفايات بمتوسط 1،570 طنًا يوميًا مقابل 1450 طنًا يوميًا خلال بقية العام. حللت الدراسة ثلاثة نماذج للتعلم الآلي: الذاكرة طويلة المدى للشبكات العصبية المتكررة ، الشبكات العصبية المتكررة ثنائية الاتجاه ، وحدة البوابات المتكررة، إلى جانب ثلاثة نماذج إحصائية: الانحدار الذاتي المتكامل المتوسط ، المتوسط المتحرك المتكامل الموسمي ، وبروفت. تفوق نموذج الشبكات العصبية المتكررة للذاكرة طويلة المدى للذاكرة ثنائية الاتجاه في الأداء على الآخرين في التنبؤ بمتوسط قيمة الخطأ المطلق 28.120 ، وقيمة الخطأ التربيعي لمتوسط الجذر 36.879 ، ومتوسط خطأ النسبة المئوية المطلقة 0.020 ، وقيمة تحديد القيمة 0.78.

جامعة البحرين

كلية العلوم

قسم برامج الدراسات العليا



التنبؤ بتوليد النفايات المنزلية الصلبة خلال شهر رمضان باستخدام مناهج التعلم الآلي العميقة

إطروحة مقدمة كجزء من متطلبات الحصول على درجة الماجستير
في علوم وتحليل البيانات الضخمه

إعداد

حسين يوسف هاني

٢٠١٦٢٢٥

إشراف

د . مجيد صفر جاسم
أستاذ مشارك

جامعة البحرين

مملكة البحرين
مارس 2023