

**ROUTE OPTIMIZATION MODEL FOR SOLID WASTE COLLECTION USING
MACHINE LEARNING AND GIS TECHNIQUES**

TITUS JEFWA CHARO

**A THESIS SUBMITTED TO THE INSTITUTE OF COMPUTING AND
INFORMATICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE AWARD OF A DEGREE OF MASTER OF SCIENCE IN INFORMATION
TECHNOLOGY OF TECHNICAL UNIVERSITY OF MOMBASA.**

2023

DECLARATION

This thesis is my original work and has not been presented for a degree award in any other university.

Student Name: Titus Jefwa Charo

Admission No.: MSIT/0001/2019

Sign _____ Date _____

This thesis has been submitted with our approval as the University supervisors.

Signature Date

Name: Dr. Kennedy Hadullo

Technical University of Mombasa (TUM)

Signature Date

Name: Dr. Kennedy Ondimu

Technical University of Mombasa (TUM)

DEDICATION

I would like to dedicate this work to my family, whose unwavering support, love, and encouragement have been instrumental throughout my academic journey. Your belief in me and your constant motivation have fueled my determination to pursue excellence and overcome challenges. This achievement would not have been possible without your understanding, patience, and sacrifices. I am forever grateful for your presence in my life and the inspiration you provide. This work is a tribute to your boundless love and the values you have instilled in me. Thank you for always being my pillar of strength.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have supported me during the course of this research project.

First and foremost, I am deeply grateful to my supervisors, for their guidance, expertise, and invaluable insights throughout this journey. Their mentorship and encouragement have been pivotal in shaping the direction of this study.

I would also like to extend my appreciation to the members of my examination committee for their time, feedback, and constructive criticism, which have greatly contributed to the refinement of this work.

My heartfelt thanks go to my family and friends for their unwavering support, understanding, and encouragement during both the highs and lows of this research endeavor. Your belief in my abilities has been a constant source of motivation.

I am indebted to the participants of this study for their willingness to share their insights and experiences. Without their contributions, this research would not have been possible.

Lastly, I would like to acknowledge the various academic resources, research facilities, and libraries that have provided access to the necessary materials and tools for this study.

To everyone who has played a part in this research, no matter how big or small, I offer my deepest appreciation. Your contributions have been invaluable, and I am truly grateful for your presence in my academic journey.

TABLE OF CONTENTS

DECLARATION	ii
DEDICATION.....	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS	v
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ACRONYMS AND ABBREVIATIONS.....	xiii
DEFINITION OF KEY TERMS.....	xv
ABSTRACT	xvi
CHAPTER ONE:.....	1
INTRODUCTION	1
1.1 Introduction.....	1
1.2 Background of the Study	2
1.3 Problem Statement	5
1.4 Objectives.....	6
1.4.1 Main Objective.....	6
1.4.2 Specific Objective	6
1.5 Research Questions	6
1.6 Significance of the Study.....	7
1.7 Limitations of the Study.....	7
1.8 Scope of the Study	8
1.9 Organization of the Study.....	8

CHAPTER TWO:	10
LITERATURE REVIEW	10
2.1 Introduction	10
2.2 Status of Mombasa Solid Waste Management	10
2.2.1 Challenges of SWM in Mombasa County	13
2.2.2 Interventions in SWM by the County Government of Mombasa	14
2.3 Optimization Theory	14
2.4 Related Work	19
2.5 Summary of Literature Review	21
2.6 Conceptual Framework	26
2.7 Chapter Summary	30
CHAPTER THREE:	31
METHODOLOGY	31
3.1 Introduction	31
3.2 Research Roadmap	31
3.3 Research Design	32
3.3.1 Study Site	37
3.3.2 Target Population	38
3.3.3 Sample Size	39
3.3.4 Data Collection Instruments	41
3.3.5 Data Collection Procedure	42
3.3.5.1 Data Cleaning	42
3.3.6 Geo-Spatial Analysis	43
3.3.7 Modeling	47

3.3.8	Model Testing	47
3.4	Ethical Considerations	48
3.5	Chapter Summary	49
CHAPTER FOUR:	51
ROUTE OPTIMIZATION MODEL FOR WASTE COLLECTION (ROMWC)	51
4.1	Introduction.....	51
4.2	Features of the Optimization Model.....	51
4.3	Machine Learning and GIS Based Optimization Model.....	52
4.4	Data Sources	53
4.4.1	Transfer Station Spatial Data	53
4.4.2	Waste Collection Truck Movement Data	54
4.4.3	Location Maps	54
4.5	Data Analysis.....	55
4.5.1	Data Cleaning.....	55
4.5.2	Normality Test.....	56
4.6	Descriptive Statistics	57
4.7	Supervised Machine Learning	60
4.7.1	Multiple Linear Regression Algorithm	61
4.7.2	Random Forest Algorithm	68
4.7.3	Ant Colony Optimization (ACO)	76
4.7.4	Integer Linear Programming (ILP).....	82
4.7.5	Simulated Annealing Algorithm	88
4.8	Summary Model Performance Analysis	97
4.9	Discussion.....	99

4.10 Chapter Summary	101
CHAPTER FIVE:.....	102
VALIDATION OF THE MODEL.....	102
5.1 Introduction.....	102
5.2 Model Validation.....	103
5.2.1 Methodology for Developing Three Optimization Models	103
5.2.2 Comparison of Model Performance	105
5.2.3 Statistical Significance Testing.....	106
5.3 Chapter Summary	110
CHAPTER SIX:	112
CONCLUSIONS AND RECOMMENDATIONS	112
6.1 Introduction.....	112
6.2 Answering Research Questions: An Exploration of Methodology	112
6.3 Contributions of the Study.....	115
6.4 Limitations of the Study.....	116
6.5 Recommendations for Future Research	116
6.6 Conclusion.....	117
REFERENCES	119
APPENDICES:	126
Appendix I. Field Survey Questionnaire for Mapping Transfer Station in Mombasa	126
Appendix II: Budget.....	129
Appendix III: Timeframe	130
Appendix IV: Publication	131

Appendix V: Data Source 1	132
Appendix VI: Data Source 2	134
Appendix VII: Simulated Annealing Pseudocode.....	138
Appendix VIII: County Government of Mombasa Letter.....	140
Appendix VIII: NACOSTI Research License.....	142
Appendix X: Scientific Ethics Review Committee.....	144

LIST OF TABLES

Table 2.1 Summary of Literature	22
Table 2.2: Steps for Conceptual Model	28
Table 4.1: Descriptive Analysis	60
Table 4.2: Multiple Linear Regression Model Performance.....	65
Table 4.3: Summary Results of Important Features for Multiple Linear Regression	66
Table 4.4: Random Forest Model Performance	72
Table 4.5: Results of Ant Colony Optimization Model Performance.....	79
Table 4.6: Ant Colony Optimization (ACO) Model performance	86
Table 4.7: Simulated Annealing Model Performance	91
Table 4.8: Number of Iteration for SA Model.....	92
Table 4.9: Results for Multiple Linear Regression Model.....	97
Table 4.10: Results for Random Forest Model.....	98
Table 4.11: Results for All Three Optimization Models.....	98
Table 5.1: Model Performance Results for Optimization Models	106

LIST OF FIGURES

Figure 2.1:Map for Mombasa County (Source: Author created using QGIS Software) ..	11
Figure 2.2: Waste Flow in Mombasa. Courtesy of UN Habitat 2019.....	12
Figure 2.3: Waste Characteristics (Source: Author created MS Excel)	12
Figure 2.4: Waste Composition (Source: Author created MS Excel)	13
Figure 2.5: Conceptual Framework	26
Figure 3.7: Distance Matrix on Spatial Scale (Source: Author).....	45
Figure 3.8: Heat Map of Distance Matrix of Transfer Stations (Source: Author).....	46
Figure 4.1: Transfer Station Data for Mombasa	53
Figure 4.2: Waste Collection and Truck Movement Data.....	54
Figure 4.3: Map Showing Location of Transfer Stations in Mombasa	55
Figure 4.4: Linear Test Using Q-Q Plot.....	56
Figure 4.5: Data Distribution Showing a Measure Of Central Tendency	58
Figure 4.6: A Plot Linearity Test of Data	63
Figure 4.7: Scatter Plot For Tested Variables.....	64
Figure 4.8: Normality Test Q-Q plot.....	65
Figure 4.9: Summary Results of Important Features Multiple Linear Regression	66
Figure 4.10: Relationship of Multiple Independent To Dependent Variables	68
Figure 4.11: Bar Plot for Extracted Feature Importance.....	69
Figure 4.12: Testing Random Forest Algorithm.....	71
Figure 4.13: Plot for Decision Tree For Prediction	74
Figure 4.14: Decision Tree Prediction Pseudocode and Scatter Plot Model Predications	76
Figure 4.15: Ant Colony Optimization (ACO) Results for Optimal Route and Best Cost	80

Figure 4.16: Spatial Location of the Optimal Route in Mombasa.....	81
Figure 4.17: Inter Linear Programming (ILP) of True vs Predicted Values	85
Figure 4.18: Integer Linear Programming Results as Optimal Route	87
Figure 4.19: Spatial Map Showing Results of Optimal Route.....	88
Figure 4.20: Results from Simulated Annealing Route Optimization for 100 Iteration ..	95
Figure 4.21: Results from Simulated Annealing With 1000 Iteration for Optimal Route	95
Figure 4.22: Results from Simulated Annealing Model for Optimal Route.....	96
Figure 4.23: Spatial map for Mombasa Showing the Optimal Route.....	96
Figure 5.1: Box Plot Showing a Comparison of Model Performance.....	109
Figure 6.1: Developed Conceptual Model	113

LIST OF ACRONYMS AND ABBREVIATIONS

ACO - Ant-Colony Algorithm

AI - Artificial Intelligence

CBO - Community Based Organization

CGM - County Government of Mombasa

CO₂ - Carbon Dioxide

DEM - Digital Elevation Model

DSR - Design Science Research

DSS - Decision Support System

GCP - Ground Control Point

GDP - Gross domestic product

GIS - Geographical Information Systems

GPS - Geographical Positioning System

ILP - Integer Linear Programming

IoT - Internet of Things

ISWM - Integrated Solid Waste Management

JOM - Java Optimization Modular

ML - Machine Learning

MSW - Municipal Solid Waste

NA - Network Analyst

QGIS - Quantum Geographical Information System

RL - Reduced Level

ROMWC - Route Optimization Model for Waste Collection

RQ - Research Question

RS - Remote Sensing

SWM - Solid Waste Management

TS - Transfer Station

UNEP - United Nations Environmental Program

DEFINITION OF KEY TERMS

Algorithm - a sequence of instructions or a set of logical steps that a computer or any problem-solving system follows to accomplish a specific task or calculation.

Arc - often called an edge.

Distance - the length of the space between two points.

Dumpsite - a piece of land where waste materials are dumped.

Model - represents what was learned by a machine learning algorithm.

Node - are collection points represented by circle.

NP Hard problem - is a problem that is at least as hard as the hardest problems in NP (nondeterministic polynomial time)

Optimization - to utilize a situation or resource in the most efficient or effective way possible.

Optimization technique - refers to the mathematical method or algorithm used to determine the most efficient routes for waste collection vehicles to follow.

Performance metrics - are measurements used to evaluate the performance of a model.

Route - a path or course that is followed in order to travel from a starting location to a particular destination

Shapefile - a digital format that stores geographical location and its related attributes using vector storage.

Solid Waste Management - refers to the collection, transportation, processing, and disposal of solid waste in an environmentally sound and sustainable manner.

Spatial - relating to or occupying space.

Transfer Station - facilities where solid waste, mainly municipal solid waste (MSW), is unloaded from collection vehicles or containers for reloading into larger, long-distance vehicles for transport to landfills/dumpsite.

ABSTRACT

The County Government of Mombasa (CGM) faces a major challenge in solid waste management (SWM), which involves collecting and removing waste. The study aimed to improve the waste transportation and logistics operations from collection points to final disposal by computing an algorithm for the optimal waste collection route in Mombasa. The ultimate goal of this study is to develop a model that optimizes the waste collection system using machine learning and geographical information system (GIS) technologies. The study aimed to determine the optimal waste collection route using ant colony, integer linear programming, and simulated annealing algorithms. These algorithms were used to analyze the data collected on the waste collection process, which included locations of transfer stations, time taken, distance, and cost for fuel. The data was first cleaned and tested for distribution, and the optimal factors were then identified. The supervised machine learning process was then applied to identify the relationships between the variables, make predictions, and find the optimization route. Finally, the integration of geographical information system (GIS) process was used to determine the final optimal route for the waste collection process. The results showed that the use of ant colony, integer linear programming, and simulated annealing algorithms was effective in finding the optimal waste collection route, resulting in a more efficient, cost-effective, and sustainable waste management system. This study is crucial and greatly benefit the department responsible for SWM in the CGM by improving the efficiency and effectiveness of waste collection. It primarily use quantitative research methods and intended to evaluate the current waste collection network. The study developed a map of the existing waste collection network using quantum geographical information system (QGIS) software and use a supervised machine learning model to find the best solution to the route optimization problem, determining the shortest path with the minimum cost.