Prediction Model for Determining Healthcare Facility Locations by the Kenyan County Government

By

Edward Isaboke Nyakeruma

073737

A Thesis Submitted to the Faculty of Information Technology in partial fulfillment of the requirements for the award of a degree in Master of Science in Information Technology

Masters of Science in Information Technology

Faculty of Information Technology

Strathmore University

March 2020

Declaration and Approval

I Edward Isaboke Nyakeruma declare that this work has not been previously submitted and approved for the award of a degree by this or any other university.

Student's Name: Edward Isaboke Nyakeruma

Sign: ___________________________ Date: __________________________

Supervisor's Name: Dr. Vitalis Ozianyi

Sign: ____________________________ Date: __________________________

Abstract

Health has been a key subject of interest to governments and non-governmental organizations. Health includes various building blocks with human resource being the core of health systems. Evidence globally shows a correlation between a country's healthcare work force and health. Infrastructure relating to health is crucial for efficient and effective healthcare systems. A strategic location for a healthcare facility can improve facility utilization and save costs. Knowledge of methods and techniques need to be regularly updated for the location and establishment of healthcare facilities. This research investigates the criteria for decision-making of locations of healthcare facilities to ensure the most strategic location for the facilities is opted for. This research looks at the challenges that affect the location of healthcare facilities in detail to model the most appropriate locations for the facilities. Existing location models developed are many and discussed in detail to cater for special cases. The two most important criteria for healthcare services are cost and efficiency. They enable reduced distance to travel by patients to the health facilities. Unpredictability is identified as an unavoidable element of healthcare facilities location problem. The purpose of this research is to develop a model for determining strategic locations to establish healthcare facilities. The model was developed using the rapid application development methodology. This research proposed a strategy which enabled predictions according to the location factors that are suitable for establishment of healthcare facilities. This is after a set of locations are identified to be the target locations by the set cover location model. The model calculates and identifies the strategic locations within an area entered by a user. It is therefore important for historical data to exist for the model to work. Based on the provided data, the model then calculates the priority of the strategic locations identified hence a sequence is known on which location to establish the facility. Datasets from an online source were used as data inputs into the prediction model. The research recommended the model as a base for decision making for establishment of a healthcare facilities.

Keywords: *Strategic locations, Set cover location model, Healthcare facilities, Location factors.*

Dedication

This research is dedicated to my loving family members, mentors and friends for the support they gave me during the research period and the duration of the Masters program.

Acknowledgments

I would like to express my thankfulness to God, who has provided inspiration, grace and determination of doing this research to the best of my ability. I would also like to appreciate my supervisor, Dr. Vitalis Ozianyi, who has been extremely helpful and welcoming throughout the duration of the Masters program.

Table of Contents

List of Tables

List of Figures

List of Equations

Abbreviations/Acronyms

CSV – Comma Separated Values

DACL - Dynamically Available Coverage Location

DFD – Data Flow Diagram

DSM - Double Standard Model

GIS - Geographic Information Systems

IDE - Integrated Development Environment

LSCP - Location Set Covering Problem

MCLP - Maximal Covering Location Problem

NGO – Non-Government Organizations

RAD - Rapid Application Development

SFP - Strategic Facility Plan

UML - Unified Modeling Language

Chapter 1: Introduction

1.1 Background

Health supports the social and economic aspects in a major way but is still a challenge globally which affects human development. The healthcare field is regularly changing hence it is important to plan for the future which is uncertain. This will help hospitals, clinics, private practices be better prepared for offering health services even in times of emergencies. Healthcare has become increasingly complex as decisions and planning are formulated in light of the current economy, an increased demand for services, new global competition, and impending legislation reform (Hoadley, Jorgensen, Masters, Tuma, & Wulff, 2010).

The research addresses strategic locations of healthcare facilities. The International Facility Management Association defines a strategic facility plan (SFP) as a two to five-year plan that comprehensively articulates the type, quantity and location of spaces required to support an organization's business objectives. Facility locations decisions perform a critical role in strategic design of systems for public and private institutions and in specifying the optimum location options for various types of facilities (Ahmadi-Javid, Seyedi & Syam, 2016). Due to the long-term obligation of human and capital resources, the location decision will have a significant effect on different aspects of an organization (Epping, 1982). The first research on locations of facilities was conducted by Weber in 1909 (Ahmadi-Javid et al., 2016). Several research and applications have been written and developed based on Weber's research.

Decisions on facility locations have always been based on economic aspects (Terouhid et al., 2012). Occasionally, the challenge facing facility locations was to find facilities regarding existing facilities and customers in order to optimize some economic measure (Dzator, M., 2016; Dzator, J., 2016). A significantly low number of facilities or facilities located in poorly chosen locations can deeply increase costs and expenses such as inventory and capital whereas a strategic and optimum location may offer a competitive advantage (Ahmadi-Javid et al., 2016).

Currently, there has been a shift in location decisions taking into consideration environmental and social consequences apart from the common economical consequence due to an increasing desire of sustainable development. Modeling for facility locations especially healthcare facilities has a greater impact when located in poor locations apart from increased costs and unreliable customer service. Healthcare facilities that are not located in strategic locations or underutilized will increase morbidity and death from diseases (Ahmadi-Javid et al., 2016).

Healthcare costs have been on the increase in developed countries and it is believed that this is partly caused by the inefficient running of healthcare facilities (Worthington, 2014). Global trends such as decreasing birth rates, increase in the average life expectancy and increasing environmental issues have showed that healthcare facility locations are important and critical to the society (Ahmadi-Javid et al., 2016). Facility location planning decisions have become increasingly important and aspects such as market analysis, financial feasibility analysis and operational review are now included in healthcare facility plans with the objective of getting a competitive advantage (Hoadley et al., 2010).

According to Cich (2017), the current criteria for choosing healthcare facilities includes four aspects which are the location, size, cost and speed to opening the healthcare facility. Location is one of the most important factors to consider. Aim is to maximize the coverage of people accessing the facility. Size is also a fundamental factor as organizations need to plan for their current and future needs. This includes the spaces of operation and parking infrastructure. Strategic locations will incur less operational costs as opposed to a poorly planned location. An organization may also decide on locations based on how quickly it desires to open its doors. This may result to the organization having to select locations themselves depending on their preferences. Time and distance are also other criteria to consider when choosing a healthcare facility. Reduced distance and time are ideal so as to reach the facilities as quickly as possible.

The best and most used practice globally is to align the location of healthcare facilities based on the coverage of the people on an area. Three models which have been in use to tackle coverage include the set covering location model, maximal covering model and P-median problem. All three models are in the class of discrete facility location models as opposed to the class of continuous location models (Daskin & Dean, 2004).

The challenges facing coverage in Kenya mostly include the lack of preserved data required such as longitude and latitude of locations which are readily available for use. Another challenge is the expertise required to implement the best practices. Models used in coverage need expertise not only to decide which model to use based on the problem at hand but also the process of implementation to come up with a system that the model is integrated into.

The proposed solution is to come up with a prediction model that will choose strategic locations for healthcare facilities which are prioritized. The set cover model was used and is based on minimizing the number of healthcare facilities required while increasing the coverage. Time is also a criterion used in the model where travel time to and from a location is determined. To determine the priority of strategic locations obtained from the set cover model, a prioritization algorithm was also used. It used the location factors which are water distribution, sewage disposal, electricity distribution, emergency exits, ample parking, internet connectivity legal license requirements, good road networks, age of building and whether the buildings are green buildings or not. The resultant locations are ordered from the highest priority to the lowest priority. This will provide a sequence of locations to follow as healthcare facilities are being established.

1.2 Problem Statement

Nelson (1976) emphasized the link between health needs and healthcare facility locations. He addressed some errors in optimizing resources for healthcare including errors in the definition, measurement, over distribution and under distribution of healthcare requirements so as to select a strategic location for healthcare facilities.

The Kenyan county government is expected to locate strategic locations to establish healthcare facilities. This is an issue as the increase in population makes it harder to establish healthcare facilities where everyone is catered for without traveling for long distances. The problem is identifying strategic locations where citizens can access the healthcare services provided at the facilities with minimal inconveniences as possible such as travel time to the facilities.

The problem is as a result of poor planning and decision making mostly based on costs without accounting for coverage of the healthcare facilities to serve the people. Citizens who cannot access these health facilities continue to suffer from their illnesses unless help is brought to their doorstep.

1.3 Aim

The purpose of this research is to develop a prediction model for determining healthcare facility locations by the Kenyan county government.

1.4 Research Objectives

i. To investigate factors that affect locations of healthcare facilities.
ii. To review existing techniques used to determine locations of healthcare facilities.

iii. To develop a prediction model for determining healthcare facility locations by the Kenyan county government.
iv. To test the developed prediction model with locations distribution data collected from the study.

1.5 Research Questions

i. What factors affect the locations of healthcare facilities?
ii. What techniques currently exist in determining the locations of healthcare facilities?
iii. How will the prediction model for determining healthcare facility locations by the Kenyan county government be developed using locations distribution?
iv. How will the prediction model be tested?

1.6 Justification

Importance of this research is to ensure that the Kenyan county government utilizes an efficient and strategic facility location identifier which can reduce the time taken for decision making on establishing a healthcare facility hence improving the facility accessibility. The distance required to travel to the healthcare facilities by the patients, workers and any other stakeholder will be greatly reduced. This will result in an increase in the ratio between healthcare delivery and patients receiving the services. The research can be used by Kenyan county government institutions to provide a strategic location for establishment of healthcare facilities as guided by the prediction model.

1.7 Scope and Limitation

The source of data for this research was limited to online datasets. The type of data includes areas, locations within the areas and the factors found within the locations which are water distribution, sewage disposal, electricity distribution, emergency exits, ample parking, internet connectivity legal license requirements, good road networks, age of building and whether the buildings are green buildings or not. The factors, within the locations, to consider while developing the prediction model covered what was recorded on the dataset and not actual findings physically collected from the locations. The research did not build the complete healthcare facilities locations system but rather demonstrated how the model that will be integrated into the system works. The challenge that affected the outcome of the research is that the data was biased due to judgment being required to determine relevant data from the dataset by the researcher.

Chapter 2: Literature Review

2.1 Introduction

The aim and intent described in this chapter is to investigate factors that affect the location of a healthcare facility and analyze methods and techniques in healthcare used to determine location of healthcare facilities. This chapter will detail research in healthcare strategic planning which involves various aspects such as location, infrastructure and capacity distribution. The subsequent sections of this chapter will detail and critique the current techniques so as to determine their weaknesses to build up on the proposed prediction model.

2.2 Global Standards Used in Determining the Appropriate Locations of Healthcare Facilities

Location theory and modeling has its roots in the revolutionary work of Weber who considered the problem as locating a single facility to minimize the total traveling distance between the site and a set of customers (Ahmadi-Javid et al., 2016).

This section begins with a review of the taxonomy of modern location models which are analytic, continuous, network and discrete models. Next, is a review of the three basic discrete facility location models from which most of the other models are derived from. They include the set covering model, the maximal covering model and the P-median model. Next, the three principles that apply to healthcare facilities are discussed. They include accessibility, adaptability and availability. Finally, the criteria for choosing healthcare facility locations are discussed. They include coverage, cost, distance, time and multiple objectives.

2.2.1 Taxonomy of Location Models

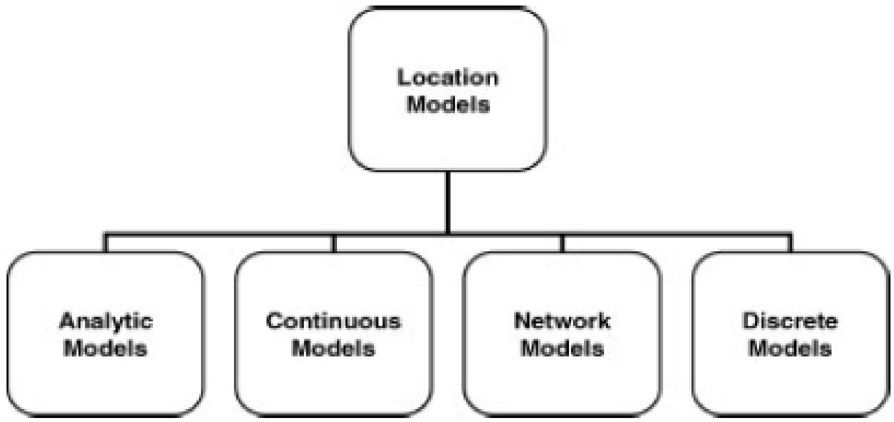

Figure 2.1: Taxonomy of location models (Daskin, 2008)

Analytic Models

They are the simplest models of all location models. These models normally assume that demand is distributed in a standard way such as uniformly over an area and facilities can be located anywhere within the area. Analytic models are solved using calculus or other simple calculation techniques (Daskin, 2008).

Continuous Models

These models normally assume that demands only arise at specific points. According to Drezner, Klamroth, Schobel and Wesolowsky (2001), this model is solved using iterative numerical procedures such as the Weiszfeld algorithm which was reviewed to ensure modifications to the algorithm did not distort the models.

Network Models

These models assume that demands and facilities can be located solely on a network composed of points and links. Frequently, demands occur only on the points, while facilities can be located anywhere on the network. The focus on the network location literature is finding polynomial time algorithms (Daskin, 2008).

Discrete Models

These models may use distance or cost metrics between any points in a random pattern. Normally, they follow rules such as the euclidean rule. Demands arise on the points and the facilities are restricted to a fixed set of potential locations (Daskin, 2008).

2.2.2 Discrete Facility Location Models

All the three models are in the class of discrete facility location models, as opposed to continuous location models. Discrete location models assume that demands can be aggregated to a fixed number of discrete points. Discrete location models assume that there is a fixed set of potential locations which facilities can be established. Continuous location models assume that demands are distributed continuously across an area. Continuous location models do not automatically assume that demands are uniformly distributed although this is a common assumption (Daskin & Dean, 2004).

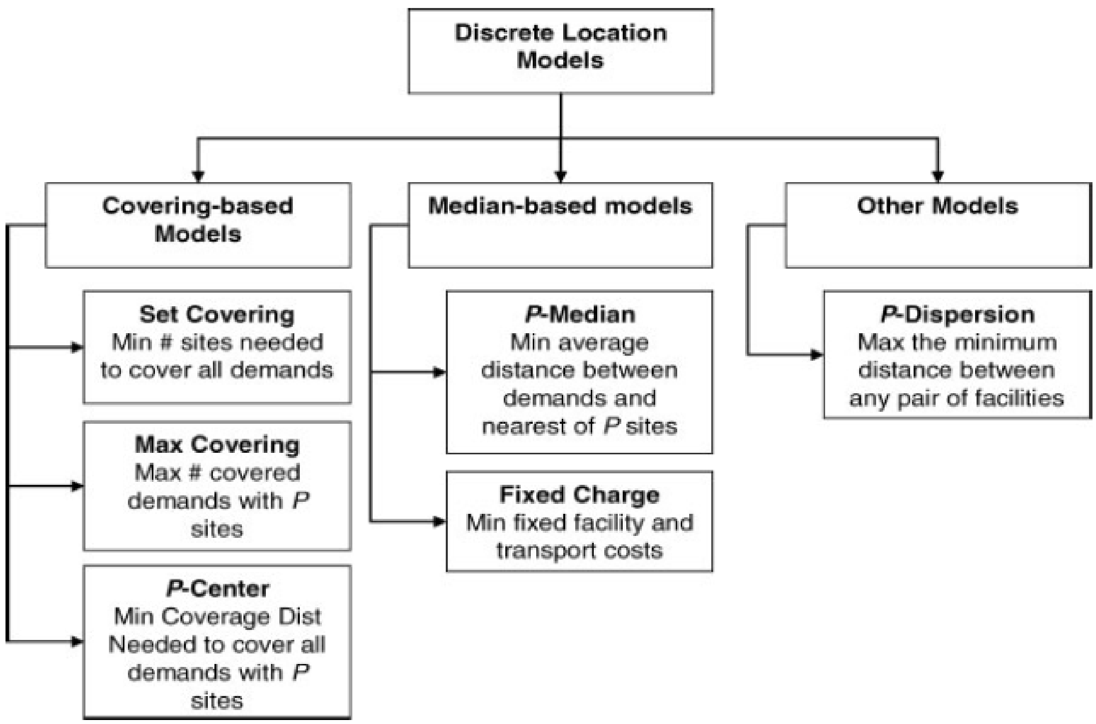

Figure 2.2: Breakdown of discrete location models (Daskin, 2008)

The discrete location models are discussed below.

i. Set Covering Model

This model is centrally based on coverage. Demands at a point are said to be covered by a facility located at some other point if the distance between the two points is less than or equal to some anonymous coverage distance. Normally, the coverage distance is the same for all demand points, though additional restrictions on the set of potential locations that can cover any particular demand point may be added. Such additional restrictions might reveal the ease of travel between population locations and a potential location for a healthcare facility (Daskin & Dean, 2004). An indicator variable is illustrated below.

$$a_{ij} = \begin{cases} 1 & \text{if demand node i can be covered by a facility at candidate site j} \\ 0 & \text{if not} \end{cases}$$

This model tries to minimize the cost of the facilities that are selected so that all demand points are covered. This model is formulated below together with the following sets and inputs.

I = set of demand nodes

J = set of candidate facility sites

f_j = fixed cost of locating a facility at candidate site j

The decision variable is illustrated below

$$X_j = \begin{cases} 1 & \text{if we locate at candidate site j} \\ 0 & \text{if not} \end{cases}$$

Taking into account the decision variable and the sets and inputs, the set covering problem is illustrated below.

$$\text{Minimize} \quad \sum_{j \in J} f_j X_j \qquad (1)$$

$$\text{Subject to} \quad \sum_{j \in J} a_{ij} X_j \geq 1 \qquad \forall i \in I \qquad (2)$$

$$X_j \in \{0,1\} \qquad \forall j \in J \qquad (3)$$

Equation 2.1: Set covering model formula (Daskin & Dean, 2004).

The objective function denoted by (1) minimizes the total cost of all selected facilities. It is referred to as the set covering problem. The right-hand side of equation (2) specifies that each demand point must be covered by at least one of the selected facilities. The left-hand side of equation (2) represents the total number of selected facilities that can cover demand point *i*. Finally, the constraints in equation (3) are standard integrality conditions (Daskin & Dean, 2004).

However, in location problems, we are often interested in minimizing the number of facilities in a location and not the cost of locating them (Daskin & Dean, 2004). In this case, the objective function which is known as the location set covering problem, becomes

$$\text{Minimize} \quad \sum_{j \in J} X_j \qquad (4)$$

Equation 2.2: Location set covering problem objective function (Daskin & Dean, 2004).

There are problems that occur with the set covering model. The first problem is related to the objective function. When the equation denoted by (1) is used as the objective function, the cost of covering all demand points is often restrictive. When the equation denoted by (4) is used as the objective function, the number of facilities required to cover all demand point becomes too large especially if the population location includes so many locations. This may be solved by increasing the coverage distance or minimizing the total coverage requirements (Daskin & Dean, 2004).

ii. Maximal Covering Model

This model was formulated because of the limitation of the set covering model in specific scenarios. If all demands cannot be covered because the cost of doing so is restrictive, we would prefer to cover those demand points that generate a lot of demand (Church and ReVelle, 1974). Additional inputs added onto the set covering model include

h_i = demand at node i
P = number of facilities to locate

The decision variable is illustrated below.

$$Z_i = \begin{cases} 1 & \text{if demand node i is covered} \\ 0 & \text{if not} \end{cases}$$

With the additional inputs to the set covering model and the decision variable, the formula for this model is illustrated below.

$$\text{Maximize} \quad \sum_{i \in I} h_i Z_i \qquad (5)$$

$$\text{Subject to} \quad Z_i - \sum_{j \in J} a_{ij} X_j \leq 0 \qquad \forall i \in I \qquad (6)$$

$$\sum_{j \in J} X_j = P \qquad (7)$$

$$X_j \in \{0,1\} \qquad \forall j \in J \qquad (8)$$

$$Z_i \in \{0,1\} \qquad \forall i \in I \qquad (9)$$

Equation 2.3: Maximal covering model formula (Daskin & Dean, 2004).

The objective function denoted by equation (5), maximizes the number of covered demands. The constraint denoted by equation (6) shows that demand node *i* should be covered by at least one facility. The constraint denoted by equation (7) states that exact number of facilities to be located is P. The constraints denoted by equations (8) and (9) are standard integrality constraints (Daskin & Dean, 2004).

Two problems may arise in this model if the constraint denoted by equation (6) is reduced. The problems are related to the coverage variables and the location variables. The coverage variables problem can be solved by inspection. The location variables problem can be solved by sorting. (Daskin & Dean, 2004).

iii. P-Median Covering Model

This model emerges due to the maximal covering model needing too many facilities to cover all demand points. This is tackled by reducing the coverage distance. The model finds the number of facilities denoted by P in certain locations to minimize the coverage distance ensuring that all demand points are covered. An innovative formulation of the problem was reviewed that exhibits improved computational characteristics when compared to the traditional formulation (Elloumi, Labbé & Pochet, 2001).

The P-Median problem is interested in interested in the average distance that a patient has to travel to a healthcare facility (Daskin & Dean, 2004). It minimizes the demand weighted total distance. This problem is formulated with the input below.

d_{ij} = distance from demand node i to candidate location j

The decision variable is shown below.

$$Y_{ij} = \begin{cases} 1 & \text{if demands at node i are assigned to a facility at candidate site j} \\ 0 & \text{if not} \end{cases}$$

The problem is therefore formulated below

$$\text{Minimize} \quad \sum_{j \in J} \sum_{i \in I} h_i d_{ij} Y_{ij} \qquad (10)$$

$$\text{Subject to} \quad \sum_{j \in J} Y_{ij} = 1 \qquad \forall i \in I \qquad (11)$$

$$Y_{ij} - X_j \leq 0 \qquad \forall i \in I; \forall j \in J \qquad (12)$$

$$\sum_{j \in J} X_j = P \qquad (13)$$

$$X_j \in \{0,1\} \qquad \forall j \in J \qquad (14)$$

$$Y_{ij} \in \{0,1\} \qquad \forall i \in I; \forall j \in J \qquad (15)$$

Equation 2.4: P-Median covering model formula (Daskin & Dean, 2004).

The objective function denoted by equation (10) minimizes the demand weighted total distance. The constraint denoted by equation (11) shows that each demand point must be assigned to exactly one facility. The constraint denoted by equation (12) specifies that demand points can only be assigned to open facilities. The constraint denoted by equation (13) shows that P facilities are to be located. The constraints denoted by equations (14) and (15) are standard integrality constraints (Daskin & Dean, 2004).

2.2.3 Principles for Determining Healthcare Facilities

Accessibility

Accessibility refers to the ability of patients being able to reach the healthcare facility or, in the case of emergency services, the ability of the healthcare providers being able to reach the patients (Daskin & Dean, 2004). The locations that have been identified to establish healthcare facilities should be accessible and this is facilitated by good road networks surrounding the locations and their proximity to patients seeking healthcare services should be as minimal as possible.

Adaptability

Decision regarding locations must be vigorous considering the long-term uncertainty of the future. This is especially crucial for healthcare facilities that cannot be relocated such as hospitals established in permanent sites. Various approaches have been designed to cater for uncertainties in the future. Scenario planning is regularly used to handle future uncertainty. Future conditions are identified and plans are developed that will prosper in the defined scenarios (Kouvelis & Yu, 1996). The locations that have been identified to establish healthcare facilities should take into consideration many decision factors so that the locations identified will be relevant in the long-term future.

Availability

Availability addresses short-term changes in the condition of the system that result from facilities being busy. It applies to emergency service systems such as ambulances in whereby the ambulance is busy serving one health facility at a different location and at the same time, it is needed to respond to another emergency in a different location (Daskin & Dean, 2004). The consideration of the location factors will bring about a clear picture to decide whether a certain location is most suitable to establish a healthcare facility there.

2.2.4 Criteria for Choosing Facility Location

A number of criteria have been used to tackle the aim of healthcare facilities location optimizations. The following are criteria used to choose healthcare facility locations.

i. Coverage

This is the most used feature to model healthcare facility. Various models have been tackled to increase coverage. Shariff, Moin and Omar (2012), presented an objective function which increased the population allocated to a facility within the stated coverage distance. Maximal Covering Network Improvement Problem (MC-NIP) was launched to maximize the number of people within the coverage area (Murawski & Church, 2009).

ii. Cost

Benneyan, Musdal, Ceyhan, Shiner and Watts (2012) calculated the total cost of technique, travel, and start-up for the facility with readily available constraints using the minimized cost as the objective. Mahar, Bretthauer and Salzarulo (2011) recommended a model to decrease expected total cost for the facility. Syam and Côté (2010) presented an objective function which minimized the sum of fixed costs, lost service costs, travel, variable patient treatment costs, labor costs, lodging and overhead costs.

iii. Distance

Decreased distance aspect is used in various methods. Huang, Kim and Menezes (2010), proposed a model to reduce the maximum weighted distance between one point and the nearest facility for emergency evacuation. Sahin, Süral and Meral (2007), presented a hierarchical model for designing a network of blood services that reduced the total referral demand-weighted distances from blood centers to regional blood centers.

iv. Time

Time measures includes factors such as traveling time and waiting time. Mestre, Oliveira and Barbosa-Póvoa (2012), applied a geographical objective function that is evaluated as the reduction of time used to travel.

v. Multiple objectives

These are used to improve coverage, efficiency and distance which are the mutual measures among others in healthcare facility locations. Mitropoulos, Mitropoulos and Giannikos (2013) suggested a method to deliberate the productivity of the healthcare provider based on the following objectives. Maximization of the mean efficiency of healthcare facilities. This ensures services offered are utilized by most if not all the patients and do not remain unused. Another objective is the minimization of under achievement of the number of patients served. This ensures that majority of the patients served should receive exemplary services. The last objective is the minimization of total distance travelled by patients. This ensures patients reach the healthcare facilities quickly without having to travel for long distances (Mitropoulos el al., 2013).

2.3 Challenges Faced by Kenya in Determining the Location of Healthcare Facilities

The challenges to be addressed that relate to modeling and determination of healthcare facility locations will determine the effective and efficient solution to be adopted. The challenges are discussed below.

i. Difficulties for patients to reach healthcare facilities

An essential task in healthcare facility location planning is reducing the distance between the patients and the healthcare facility. There are several planning methods to determine the distance required to travel to facilities. One of the methods is the P-median model. It seeks to reduce the distance required to travel by patients to healthcare facilities nearby (Afshari & Peng, 2014). It is amongst the common models used to tackle facility location problems. P-median models are comprehensively used in the interpretation of healthcare facility location problems. Church and ReVelle (1976), noted a significant method to gauge the effectiveness and efficiency of a facility location is to clarify the average distance (Afshari & Peng, 2014). P-median is a straightforward model with its objective being only one. A lot of effort and involvements have been made to manage problems in healthcare facility location based on P-median.

ii. The number of healthcare facilities required to cover all patients

The number of healthcare facilities required to fulfil requested healthcare facilities needs modeling for the location of the healthcare facilities. The aim is to maximize the patients covered by the facilities within a maximum reasonable distance. A frequently used model to determine the number of facilities required so as to maximize coverage is the Location Set Covering Problem (LSCP) (Afshari & Peng, 2014). The model assigns facilities by reducing the average distance. Full coverage is difficult to accomplish due to inadequate resources. Intensive facility location problems might need more attributes and goals.

iii. Difficulties to cover all patients' healthcare needs within the healthcare facilities

This difficulty begins with the inadequate availability of resources to satisfy needs of patients. The purpose is to increase healthcare service coverage for selected facilities. A maximum deterministic model was proposed as a solution for this challenge (Afshari & Peng, 2014). It was later extended and called MCLP. LSCP and MCLP are unable to provide dependable results when emergency services are required (Afshari & Peng, 2014).

Gendreau (1997), suggested the DSM model. This model wishes to assign facilities among probable locations to deliver full coverage over a long distance while increasing the coverage along a short distance. Doerner, Gutjah, Hartl, Karall, Reimann (2005), came up with a model that was borrowed from DSM to insert penalties into the objective function to avoid unsatisfied coverage of requirements and unequal capacity. Backup Coverage Problem models are used to maximize the population coverage with more than two facilities while covering all demand points once (Afshari & Peng, 2014).

iv. Uncertainties in providing patients with healthcare services

Doubtfulness is unavoidable when modeling a location that has no details about the possibilities for decision makers to improve the scenario of the system (Snyder, 2006). The doubtfulness can happen in stochastic problems or robust optimization problems. This doubtfulness has been modeled using queuing, simulation or mathematical programming formulation. The queuing theory and P-median model were combined to form the Stochastic Queue Median (SQM) (Afshari & Peng, 2014). It allocates mobile servers to the demand points to minimize the average cost of the response. A probabilistic covering location model called MALP which has two versions came about. MALP-I's main assumption is that that the facilities have the same busy fraction. MALP-II assumes the busy fraction associated with each demand point was computed as the ratio of the total duration of all calls generated (Afshari & Peng, 2014).

v. Varying demands for healthcare services

Some diseases, with an epidemic or seasonal nature, brings difficulties in the prediction of healthcare services. A proposed resolution for this task is to analyze the trends of diseases. Dynamic models can maximize the coverage goal efficiently.

Gendreau, Laporte and Semet (2001), presented variables to reveal the dynamic nature of the model. DDSMt was the name of the model. This relates to determining a facility location model for emergency medical services. Rajagopalan, Saydam and Xiao (2008), demonstrated the DACL model where time is grouped into clusters based on a substantial difference of demands.

vi. Determining the best location of healthcare facilities

Actual and authentic issues in healthcare facility locations is as a result of a combination of different challenges. A decision-making structure is required so as to tackle the challenges. An optimal solution is required to address healthcare facility location problems that takes into account metrics and methods which have been discussed earlier (Afshari & Peng, 2014).

2.4 Proposed Solutions for Determining Location of Healthcare Facilities

Proposed solution methods have been discussed with strengths and weaknesses of each solution identified. These methods include:

Exact Methods

They can ensure an improved solution. The techniques used are effective for small-sized problems which are usually in linear environments hence this is a strength for these methods. A weakness related to these methods is that they will lose their effectiveness and efficiency if a feature of the problem leads to more complexity or if problems are of a larger size (Afshari & Peng, 2014).

Exact methods have been applied by various researchers. Contreras et al. (2010), presented the branch-and-bound and lagrangian relaxation as the solution techniques. Lagrangian relaxation can be incorporated in a branch-and-bound algorithm in order to obtain the optimal solution. Hinojosa et al. (2008), presented the branch-and-bound, lagrangian relaxation and heuristic algorithm as solution techniques. They were used to develop an ascent procedure to generate a good solution for the relaxed problem.

Heuristic Methods

The strengths in these methods are that they can be applied for a case with time-consuming solution search process and where complexity of problems cannot be solved by exact methods to give an optimal solution (Afshari & Peng, 2014). A weakness to these methods is that an optimal solution is not a guarantee although their efficiency in complex situation is trusted.

Heuristic methods have been applied by various researchers. Romauch and Hartl (2005), presented the stochastic dynamic programming and heuristic algorithm as the solution techniques that were used to compare the heuristic results and the exact solution method results. Ghaderi and Jabalameli (2013), presented the branch and bound, hybrid greedy heuristic, fix-and-optimize heuristic and hybrid simulated annealing as the solution techniques. Fix-and-optimize heuristic is based on simulated annealing.

Meta-heuristic Methods

A strength associated with these methods is that they are used in applications to solve increasing optimization problems and their efficiency can be relied upon when models are complicated. These methods however cannot assure the realization of an optimal solution hence their weakness. Non-linear environments use these methods when a global optimal solution is difficult to find among local optimal solutions (Afshari & Peng, 2014).

Meta-heuristic methods have been applied by various researchers. Bozkaya et al. (2010), presented the genetic algorithm and tabu search as the solution technique. The genetic algorithm principles decide which locations to open. Gendreau et al. (2001), presented tabu search and heuristic algorithm as solution techniques that describes a sequential tabu search heuristic.

2.5 Research Gaps

A research gap is an area for which missing or insufficient information limits the ability of the researcher to reach a conclusion for a question (Enago Academy, 2019).

A research gap in the study is coming up with a realistic distance metric. Distance is one of the inputs in facility location problems. The distance is given without an explanation about how it is calculated but rather is given as an assumption or an example. In the literature of facility location, in order to calculate distance Euclidean and squared Euclidean distances are considered (Seyedhosseini, Makui, Shahanaghi, 2016).

Decision makers' risk attitude. In an uncertain situation, decision makers may make different decisions depending on their risk attitudes. Facility location models tend to find optimal solutions which are predicted. In a case where emergency facilities need to be established, the decision makers may not prefer the prediction of an optimal solution as a poorly selected location may cause delays in reaching the facilities and could lead to death of extremely sick patients. In business-oriented facilities such as clinics or pharmacies, the decision maker may take the risk as many people accessing the facility means profit is gained. Hence, the behavioral aspect of the decision maker is an important gap (Seyedhosseini, Makui, Shahanaghi, 2016).

Existing healthcare facilities. Healthcare facility location models will determine optimal locations for the facilities. The models have not accounted for an opportunity to include existing facilities so that they are not put together with other demand points. It is also unclear whether the introduction of existing healthcare facilities will lead to a decrease in the number of strategic locations identified.

2.6 Conceptual Framework.

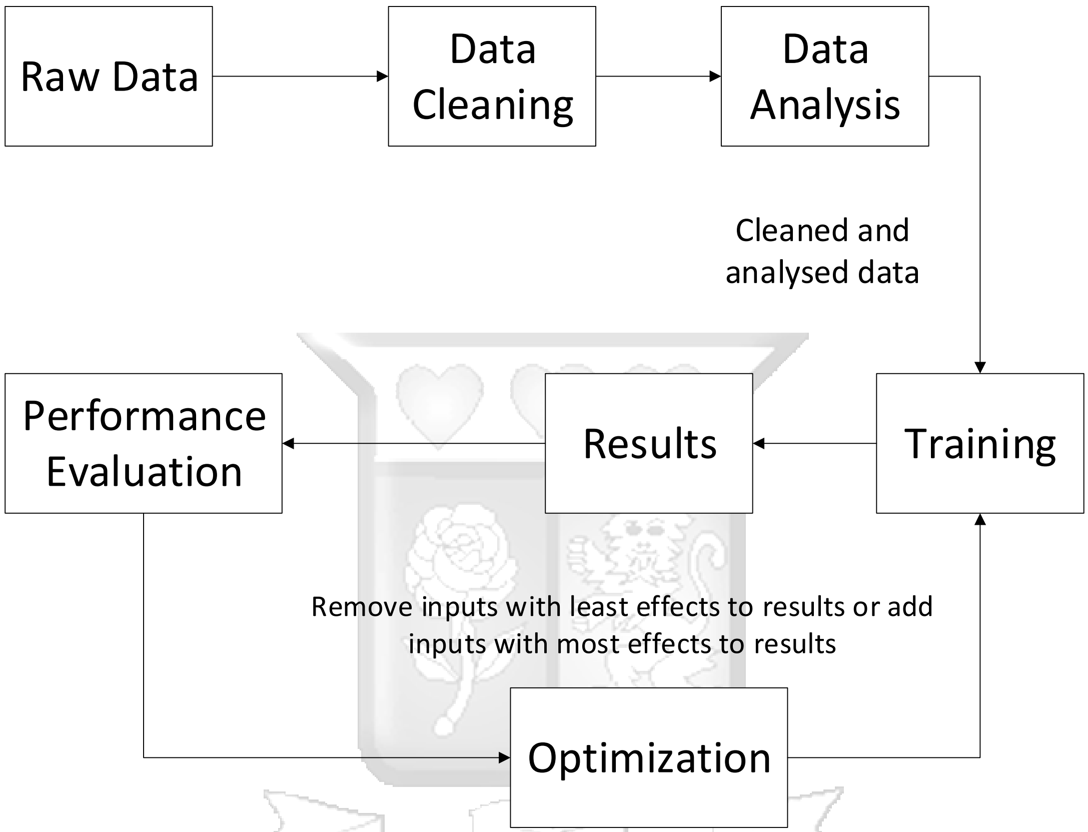

Figure 3.3: Conceptual framework

Raw data is collected from the dataset and undergoes pre-processing so as to clean the data. Data cleaning and analysis involves removing missing values and restricting location factors to those that are required. Training involves adding the inputs which are the time constraint, locations and location factors to get the results which are prioritized strategic locations. Performance evaluation is performed to ensure if the locations in an area are increased then the strategic locations should shift instead of remaining stagnant. If errors are experienced, corrections are made on the model and it is tested again to ensure determination of strategic locations. Therefore, this process is iterative.

Chapter 3: Research Methodology

3.1 Introduction

Research methodology outlines the procedures the researcher took in conducting the study. It is also termed as the study of gaining knowledge. Research methodology discusses the research variables used, the data collection methods used, the data collected, and how the data will be processed. It also details the sampling methods that the researcher used and the development approach used to come up with a solution. The suitability of the methodology chosen for the research problem is determined at this point (Rajasekar, Philominathan, & Chinnathambi, 2006).

3.2 Research Design

Degu and Yigzaw (2006) defined a research design as the process that guides a researcher on how to collect, analyze and interpret observations. This being an applied research, involved the use of quantitative approaches to examine, analyze and represent the relationship between variables. The research therefore adopted an experimental research design.

The researcher used a quantifiable dataset from the population sample on locations. Essentially, the research design involves running the prediction model with quantifiable data. This was therefore a good approach as it ensured objectivity was achieved and subjective was eliminated.

3.3 Target Population

Target population refers to an entire group of individuals or objects which researchers are interested in taking a broad view of the conclusions. In some cases, the researcher may require purposively selecting a study population, and in so doing, the researcher must consider questions of appropriateness and practicability (Degu & Yigzaw, 2006). The target population the researcher is interested in is the distribution of locations within an area.

3.4 Sampling

Sampling is the process of selecting units such as people or organizations from a population of interest for the purpose of research by the researcher (Trochim, 2020). A sample is a true representative of the entire population and the values driven out from that sample are held to true for the entire population.

Purposive sampling was used to conduct this research. This sampling method was used because the locations data obtained was from an online dataset hence deliberate selection of the data was required to fulfil the objectives of this research.

3.5 Data Collection

Secondary data was primarily used to obtain the data required for this proposed research. The greedy set cover algorithm used structured data of the target locations to identify locations to establish healthcare facilities as inputs.

3.5.1 Secondary Data

This data was collected from datasets obtained from OR-Library. The factors that were found within the locations that were of interest to the researcher included water distribution, sewage disposal, electricity distribution, emergency exits, ample parking, internet connectivity legal license requirements, good road networks, age of building and whether the buildings are green buildings or not.

3.6 Data Analysis

This is the organization of the data collected and summed up in a form that can be manipulated. It is also done to ensure that the data is complete and consistent hence preserve data quality (Onyango, 2018). The researcher populated the quantitative data such as locations and location factors into excel spreadsheets where pre-processing was done. The excel file was then converted to csv file format for use by the prediction model.

3.7 System Development Methodology

A system development methodology refers to the framework used to arrange, plan and control the process of developing an information system. The methodology shows how the research design is implemented. The locations data collected and feed into the prediction system is an iterative process hence the methodology that was chosen supported an iterative approach.

3.7.1 The Methodology

The development approach selected for this research was the Rapid Application Development (RAD). The aim was to produce high quality systems promptly through the use of iterative prototyping at any stage of development with the use of computerized development tools.

The aim at the end of the systems development phase was to come up with a prototype of a healthcare facility location system. The selected approach was arrived at on consideration of the time constraints involved in the project process. The RAD approach selected was combined with prototyping. The model went through the steps described by the RAD approach shown below during development. Requirements definition which includes the data collected, system design and coding were repetitive processes done before implementation and maintenance of the complete model.

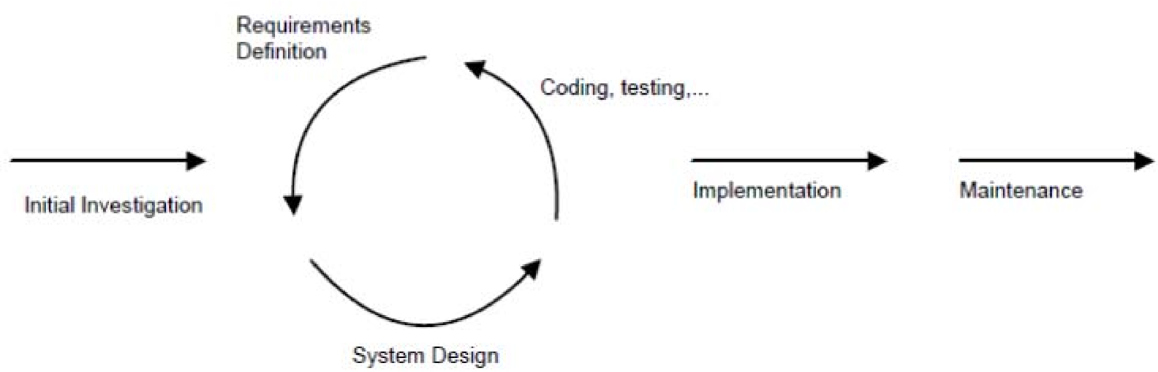

Figure 4.1: Rapid Application Development (Centers for Medicare & Medicaid Services, 2008)

3.7.2 Phases of the Methodology

System Analysis

This phase involves defining the problem and the inputs of the system. The requirement of this research is to predict suitable healthcare facility locations for ease of access by the people. The system will use the time constraint, locations and location factors as inputs and determine the locations of the healthcare facilities. The system will therefore enable the choosing of the most suitable locations to set up a healthcare facility.

System Design

This focuses on how to accomplish the objective of the system. The use of Unified Modelling Language diagrams was used in this phase. System designs used are use cases that showed how the user interacted with the system and how the processes interacted with each other. The steps were outlined of the interactions take place. Data flow diagram (DFD) that showed a graphical representation of the processes, which captured, manipulated, stored, and distributed the data between the components of the system. Database schemas that were used to store results from the prediction model. Sequence diagrams that showed the sequence of interactions between the user and the rest of the objects.

The tool that was used for creating the UML diagrams was Microsoft Visio and draw.io. The system architecture comprises of software components. The software components of the system will include the greedy set cover algorithm that determines the number of locations in an area such as a town that will ensure all the people in the area are covered.

Systems Development and Testing

This entails coming up with a workable prototype and various tests done on the prototype to determine its usability. The development phase involves coming up with the prototype, which is the healthcare facility location system.

Testing involves performing checks against the requirements to determine whether or not the requirements have been met. It will include identification of failures against expected results. Some of the tests that are anticipated include functional testing, which will be mapped to the requirements, and usability testing to determine whether the system is usable.

System Implementation

This entailed setting up the model in an environment where it can be operational. Tools used include:

i. Python which was used as the programming language because it is most suitable when creating prediction models. Advantages of using python include, more functions are built using fewer lines of code, it is perfect for building prototypes, it has plenty of resources and it is open source hence its use is free of charge.

ii. Jupyter Notebooks. It ships with the IPython kernel, which allowed the researcher to build the model in using python code. It was used for data cleaning, statistical modeling and building the model.

iii. Anaconda enterprise. This is a dynamic, modern software platform that enables organizations to develop, govern and automate data science and AI pipelines from local environment to production environment. It is part of a unified browser-based notebook collaboration with incorporated data science environments such as Jupyter Notebooks and JupyterLab.

iv. Microsoft word and Microsoft Excel. This software was used for system documentation and data analysis.

v. Windows operating system. This was the operating system that was used on the researcher's laptop.

System Maintenance

This phase involved changes that were made to the system once it was complete. The objective is to improve the user's experience hence the strategic locations that were identified by the model were arranged vertically, a change from being arranged horizontally, with the first location having the highest priority and the last location having the lowest priority to establish the healthcare facilities.

3.8 Research Quality

To test the research quality aspect, validity and reliability measures were used.

3.8.1 Validity

This approach measured the extent to which data accurately measured what they were intended to measure. Validity was accomplished by the researcher taking on multiple perspectives of the research. This took the form of using different locations from the dataset. This research followed the required quality standards by ensuring proper procedure was used in data retrieval and accurate citation of works from other researchers was performed.

3.8.2 Reliability

This approach is used to measure the extent to which data collection method yield consistent findings if replicated by others. Reliability was achieved by presenting the findings to an expert in the field and getting their views on the subject matter. This ensured relevance of the study and its findings. The research was thoroughly reviewed by the researcher's supervisor to ensure the research objectives were met.

3.9 Ethical Considerations

The researcher ensured that information retrieved from the OR-Library dataset was used solely for research purposes. The data collected for the purpose of this research is open source and accessible to the public. Relevant information from other authors and researchers were documented and given credit to avoid plagiarism. The research met the general ethical requirements as required and approved by an Independent Ethical Board. Adherence to above requirements ensured that the research attained the required ethical considerations.

Chapter 4: System Analysis and Design

4.1 Introduction

This chapter describes the overall architecture and detailed design of the system by incorporating various requirements. UML diagrams were used to describe the overall architecture of the system, give detailed descriptions of the various components of the system and illustrate interaction between the user and various components of the system.

To achieve this, various design diagrams were developed which included a depiction of the system architecture, use case diagram followed up with comprehensive use case descriptions, sequence diagram, context diagram, partial domain diagram and a database design schema.

The system works by taking into account the target locations from all the locations in an area and the location factors to obtain strategic prioritized locations. The set covering model will use the target locations to identify the strategic locations. The location factors which are, water distribution, sewage disposal, electricity distribution, emergency exits, ample parking, internet connectivity legal license requirements, good road networks, age of building and whether the buildings are green buildings or not, will be used by the priority algorithm to determine the priority of strategic locations hence the results are the strategic prioritized locations.

4.2 Requirement Analysis

This research aimed at developing a model to aid in locating strategic prioritized locations for the establishment of healthcare facilities. Based on this objective, this section outlines the various requirements to be provided for by the proposed system.

4.2.1 Functional Requirements

i. The system should allow a user to enter an area that contains locations.
ii. The system should query the database to search for the entered area.
iii. The system should identify locations within the area entered by the user if found.
iv. The system should identify target locations by filtering locations found in the entered area depending on a set condition such as time to traverse between locations or distance between locations.
v. The system should allow for the updating of required location factors that will determine the priority of the strategic locations identified to establish the healthcare facilities.

vi. The system should determine the strategic prioritized locations to set up healthcare facilities.

vii. The system should display to the user the strategic prioritized locations to set up the healthcare facilities.

viii. The system should have a database to save area information and the strategic prioritized locations identified to establish the healthcare facilities.

4.2.2 Non-Functional Requirements

i. Can run on a web browser.

ii. The user interface should be responsive such as adapting to various screen sizes.

iii. The system should be able to analyze and return strategic prioritized locations within milliseconds so as to enhance user experience.

iv. The system should be available whenever the user wants to use it and should not crash when in use.

v. The system should be easy to maintain and support.

4.3 System Design and Architecture

The literature review and the research findings guided the development of the system. The various system techniques reviewed assisted in selecting the appropriate platform for the system and the research findings enabled the collection of data and design of the system. System architecture, use case, context and level 0, database design schema, partial domain diagram and a sequence diagram were used.

4.3.1 System Architecture

Figure 4.1 illustrates the system architecture. Raw data is collected from the dataset in form of a csv file and undergoes pre-processing so as to clean the data. Data cleaning and analysis involves removing missing values and selecting only the location factors that are required. Data is split into training data and test data. The data, that is time constraint, locations and location factors, are entered into the set cover model to get the strategic locations. Locations priority is then identified using a priority algorithm that organizes the locations from highest priority to lowest priority. The result obtained is showed to the user. The user is then prompted to save the results in the database.

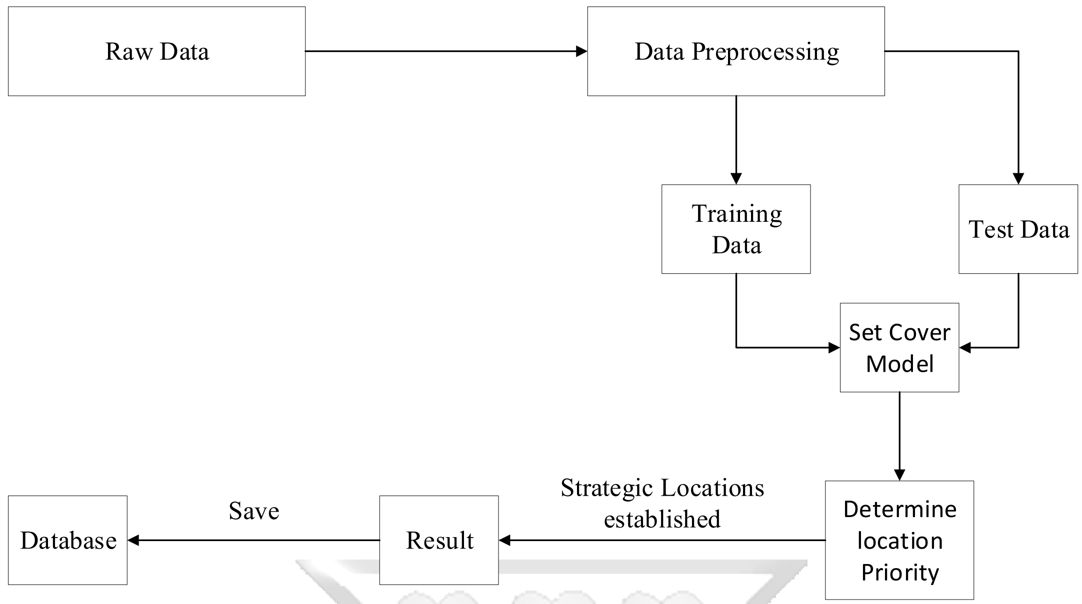

Figure 4.1: System Architecture

4.3.2 Use Case Diagram

Use case diagrams are used to illustrate interaction between actors and the system. Figure 4.2 illustrates these interactions between the various actors and the proposed healthcare facility location system. The diagram also shows the functionality that the proposed system should have.

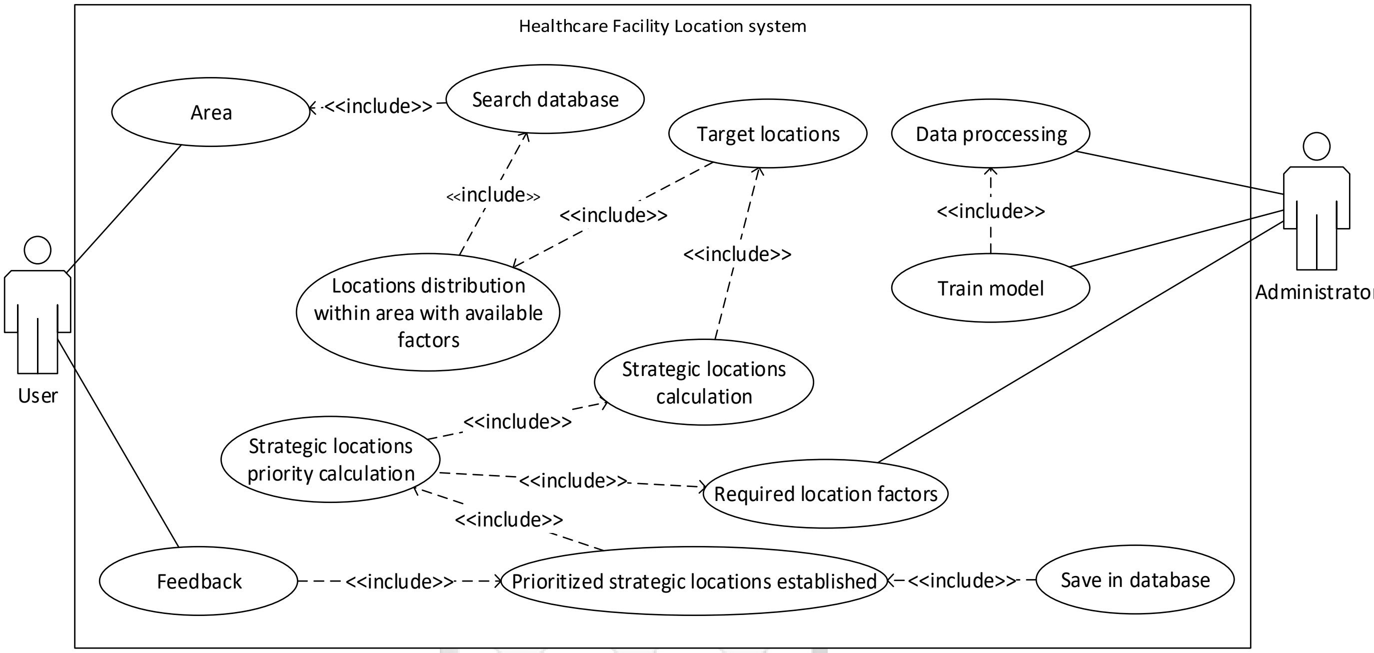

Figure 4.2: Use Case Diagram

This diagram illustrates the list of actions or use cases that the model fulfils. The use cases identified are:

i. Area - The user enters an area which they desire to establish healthcare facilities.
ii. Search database - Entered area is searched for in the database.
iii. Locations distribution within area with available factors - Locations found within an area together with available factors such as electricity and water supply in which the healthcare facilities will be established. They include buildings or sites.
iv. Target locations - Locations obtained after filtering all locations within the area based on set conditions such as time taken to traverse from one location to another or distance between locations.
v. Strategic locations calculation - Done using the set cover model and produces locations that ensure maximum coverage among the target locations.
vi. Required location factors - These are factors and their weights provided by the administrator that are necessary to determine priority of strategic locations identified.
vii. Locations priority calculation - Find the required factors and corresponding weights within the strategic locations and prioritize the locations depending on the total weights from highest to lowest.

viii. Prioritized locations established - The strategic locations are arranged from highest priority to lowest.
ix. Feedback - The strategic prioritized locations are shown to the user.
x. Save in database - Store the strategic prioritized locations in a database.
xi. Data processing - The administrator updates the areas, their locations and location factors. Required factors are also added or removed and their weights are also updated.
xii. Train model - The administrator uses training data to train the model.

4.3.3 Data Flow Diagram

A data flow diagram illustrates the processes and entities in a system outlining how data flows from each of the entities to the processes. It captures the storage of data from the processes. The flow of data in the system is shown, enabling a better understanding of the system.

4.3.3.1 Context Diagram

The context diagram in figure 4.3 illustrates the boundary of the system, its environment and the entities that interact with it. It also shows the various inputs and outputs from the system to the entities. The main entities interacting with the proposed system is a user and the administrator. The user enters the area to establish healthcare facilities. The model then calculates the strategic locations within the area to establish the healthcare facilities that are prioritized from highest to lowest. The Administrator has to frequently update the required factors that are required in determining priority of strategic locations to establish healthcare facilities. The required factors are determined by a separate team and are used to determine the quality of the strategic positions identified.

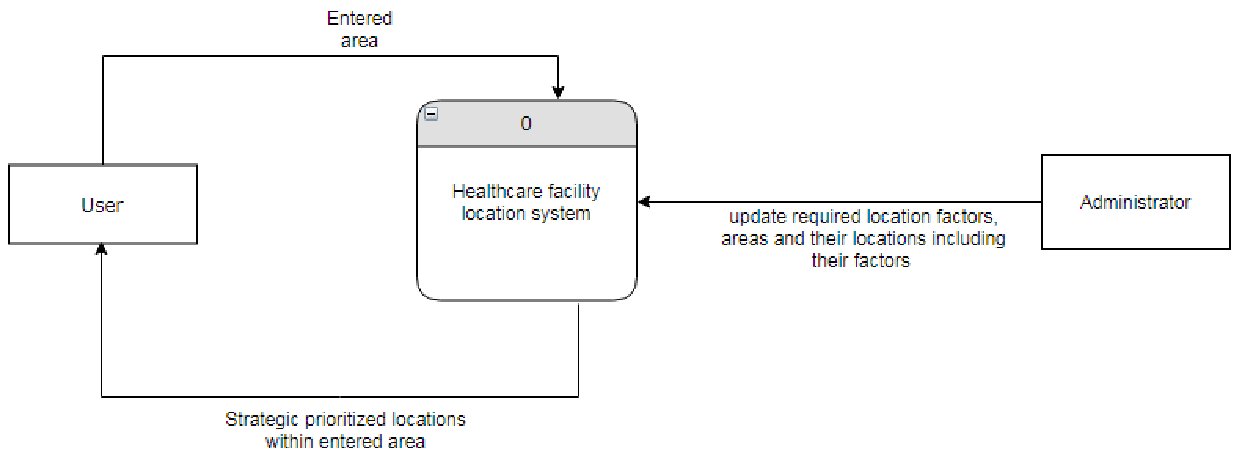

Figure 4.3: Context Diagram

4.3.3.2 Level 0 Data Flow Diagram

The level 0 Data Flow Diagram (DFD) shown in figure 4.4 gives a more detailed view by illustrating the various processes contained in the module, data stores and entities. Arrows depict the flow of data among various components of the DFD. Process 1 depicts an administrator adding areas and their information into the system which is saved in the database. The user enters an area which is searched among areas in the database as shown in process 2. The locations within the area entered by the user are identified together with their factors as shown in process 3. The administrator adds or updates required factors with their weights that will be used to determine priority of strategic locations obtained as shown in process 4. The available locations are then filtered depending on a set condition and target locations are identified as shown in process 5. Feedback is sent to the user with information concerning the strategic locations identified from the healthcare facility location model which are then prioritized.

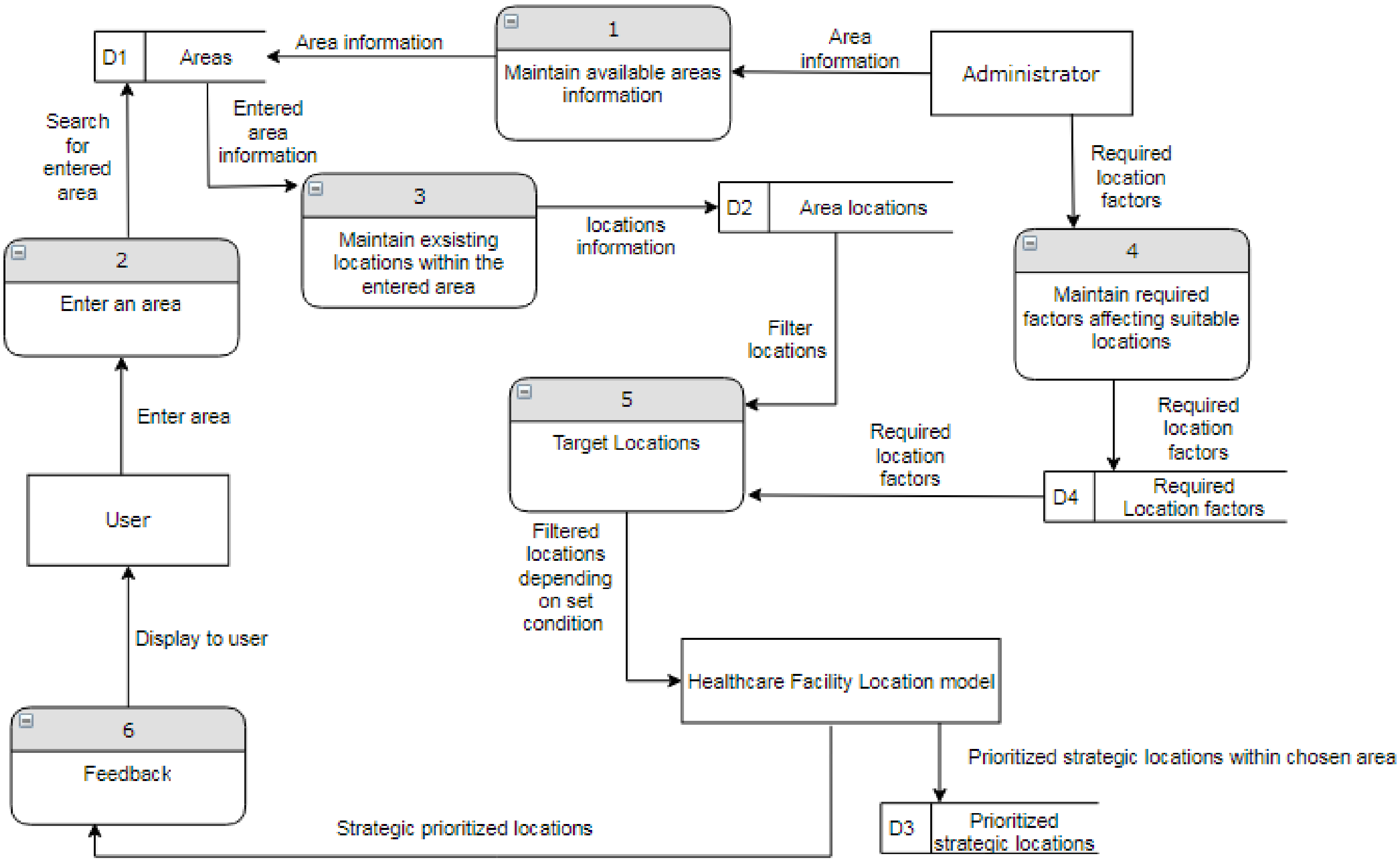

Figure 4.4: Level 0 Diagram

4.3.4 Sequence Diagram

A sequence diagram is an interaction diagram that represents the interaction between objects through messages. The objects communicate through messages, represented by horizontal arrows from the sender to a recipient of the message (Mukherjee, Upadhyay & Achariya, 2015). The sequence diagram in figure 4.5 shows the sequence of interactions between the user and the rest of the objects. The user begins by entering an area into the system. They system queries the database and returns the results back to the user. The results are either the area is found or not found. If the area is found, the locations within the area are identified. The required factors are then identified from all the location factors. A set condition such as distance or time is used to determine the target locations. The target locations are then entered into the healthcare facility location model to identify the strategic prioritized locations which are then displayed to the user, who then saves the results into the database.

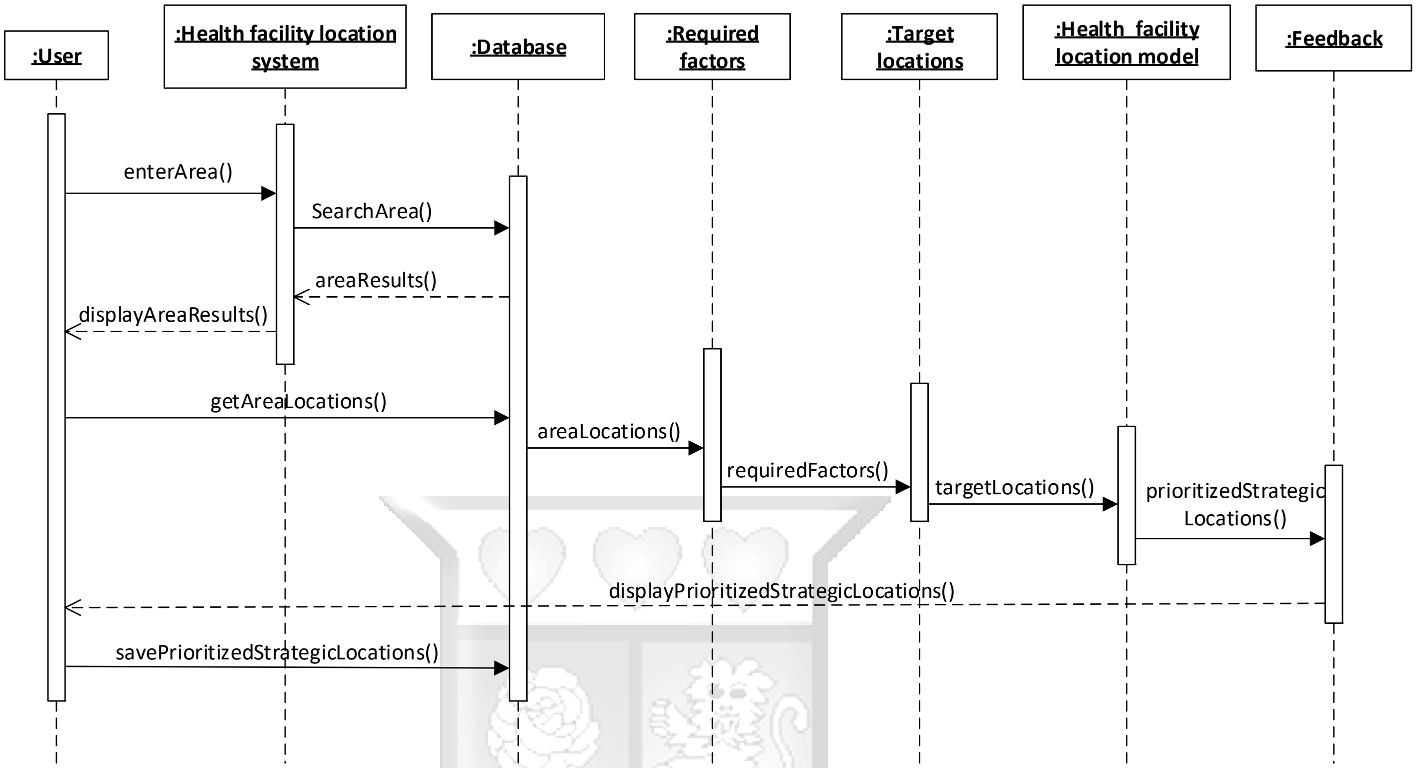

Figure 4.5: Sequence Diagram

4.3.5 Partial Domain Model

A domain model is a visual representation of conceptual classes or real-world objects in a domain of interest (Fowler, 1996). The domain model is explained with class diagrams that have got no operations defined. Associations between conceptual classes and attributes of conceptual classes are shown in this diagram. Figure 4.6 comprises of conceptual classes and the associations between each other. An area is entered by a user. An administrator adds required factors and area information. The area entered by the user is searched within the area information added by the administrator. Areas have locations within them. The locations within an area are filtered either by distance or time to get the target locations. The target locations are used to calculate the strategic locations. Priority of the strategic locations is finally obtained.

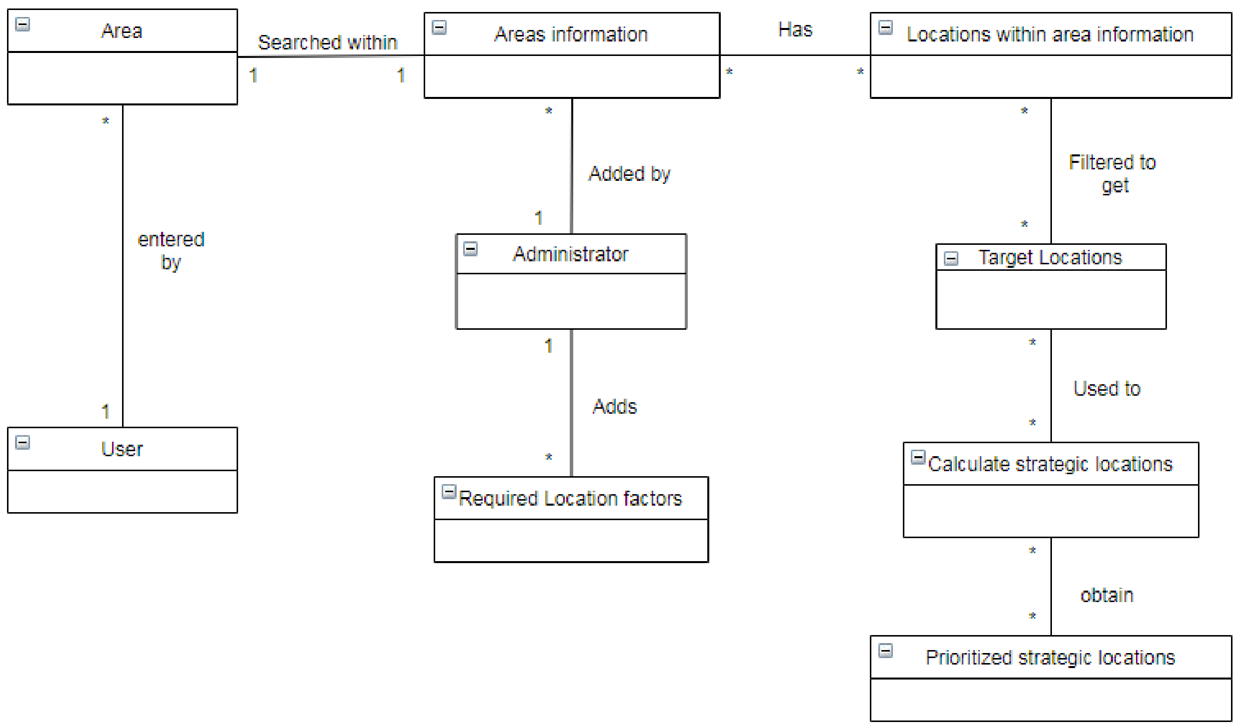

Figure 4.6: Partial Domain Model

4.3.6 Database Schema

A database schema defines its entities and the relationship among the entities. It contains an explanatory description of the database, which can be represented by schema diagrams. Figure 4.7 shows the database design that the system uses. The areas table contains area information together with the user who added the information and timestamps. The users table has a relationship with the areas table. Area locations table contains location information and factors within the location. This table has a relationship with the areas table. The required factors table contains the factors that need to be identified within a location's factors. The target locations table contains the required factors and area locations hence a relationship with both the area locations table and required factors table. The prioritized strategic locations table contains the strategic locations identified from the target locations and their priority.

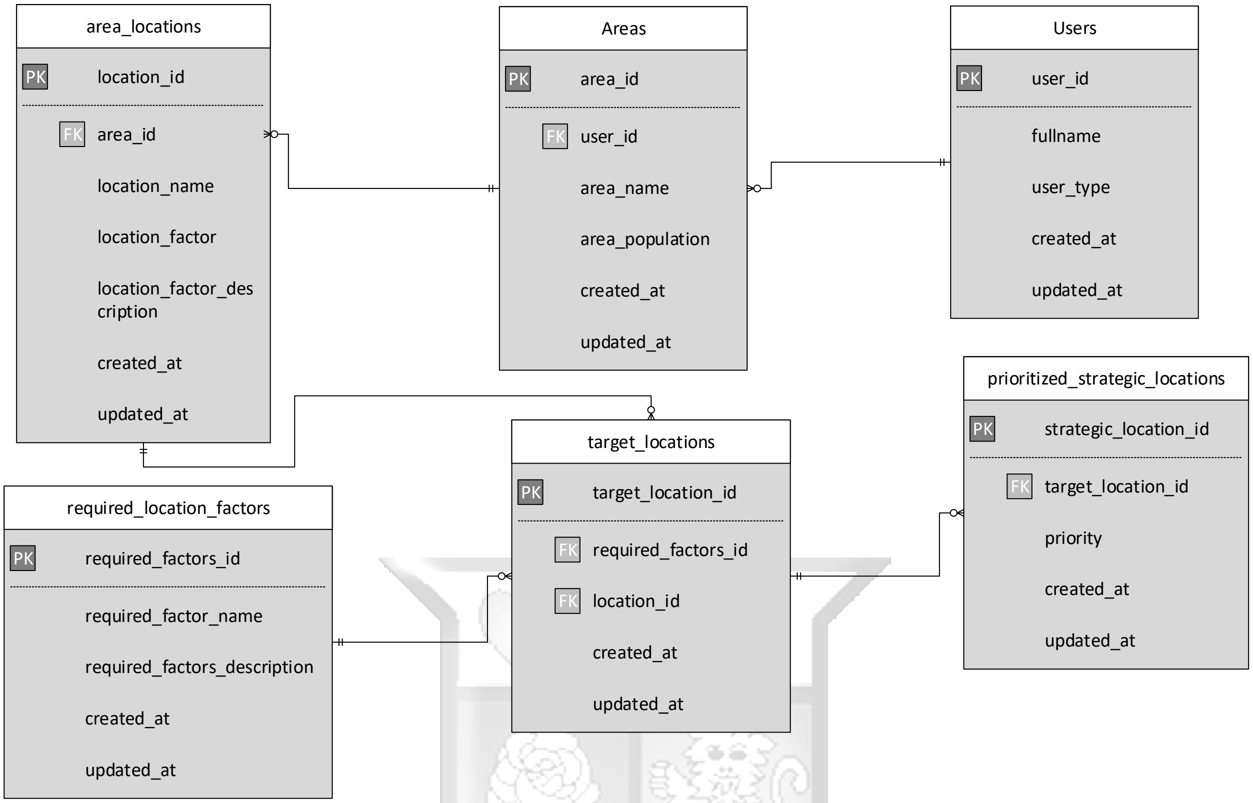

Figure 4.7: Database Schema

Chapter 5: System Implementation and Testing

5.1 Introduction

The implementation stage of a system comprises of developing and testing the system's software, documentation and operation procedures (Dennis, Wixom & Roth, 2012). This chapter describes how the model was implemented, tested and validated. It begins by describing the process of coming up with the set covering problem model. The model is then tested to confirm output values.

To further validate the researcher's approach experiments described in chapter 3 were implemented to determine the best configuration of features. The final section of this chapter describes the use of the model in predicting the strategic locations to establish healthcare facilities.

5.2 Hardware and Software Environment

The development of the model was undertaken on Jupyter notebook which is a data science environment, using Pandas and NumPy. Pandas was used for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series. The default programming language for the development of the model was Python due to the availability of the required libraries. The table below highlights the specifications used for the model.

Table 5.1: Hardware and Software Environment

Software	**Specified Library**
Python (3.7)	Pandas (1.0.1)
	NumPy (1.18.1)
Hardware	**Details**
RAM	4GB
CPU	i5
HDD	500GB

Pandas is a software library that was used for data manipulation and analysis. It offered, in particular, data structures and operations for manipulating the numerical figures used in the calculation of determining priority of select locations.

NumPy is a library that helped to add support for a large collection of high-level mathematical functions to operate and help capture the various aspects of the greedy set cover algorithm.

5.3 Dataset Overview

The figure below represents a map of RPI campus. Healthcare facilities need to be established within the campus to ensure that students get their treatment and prescriptions and go to class in time. A student should take a maximum of two minutes to get to the healthcare facility from any building within campus. In this case, the area is the campus and the locations are the buildings.

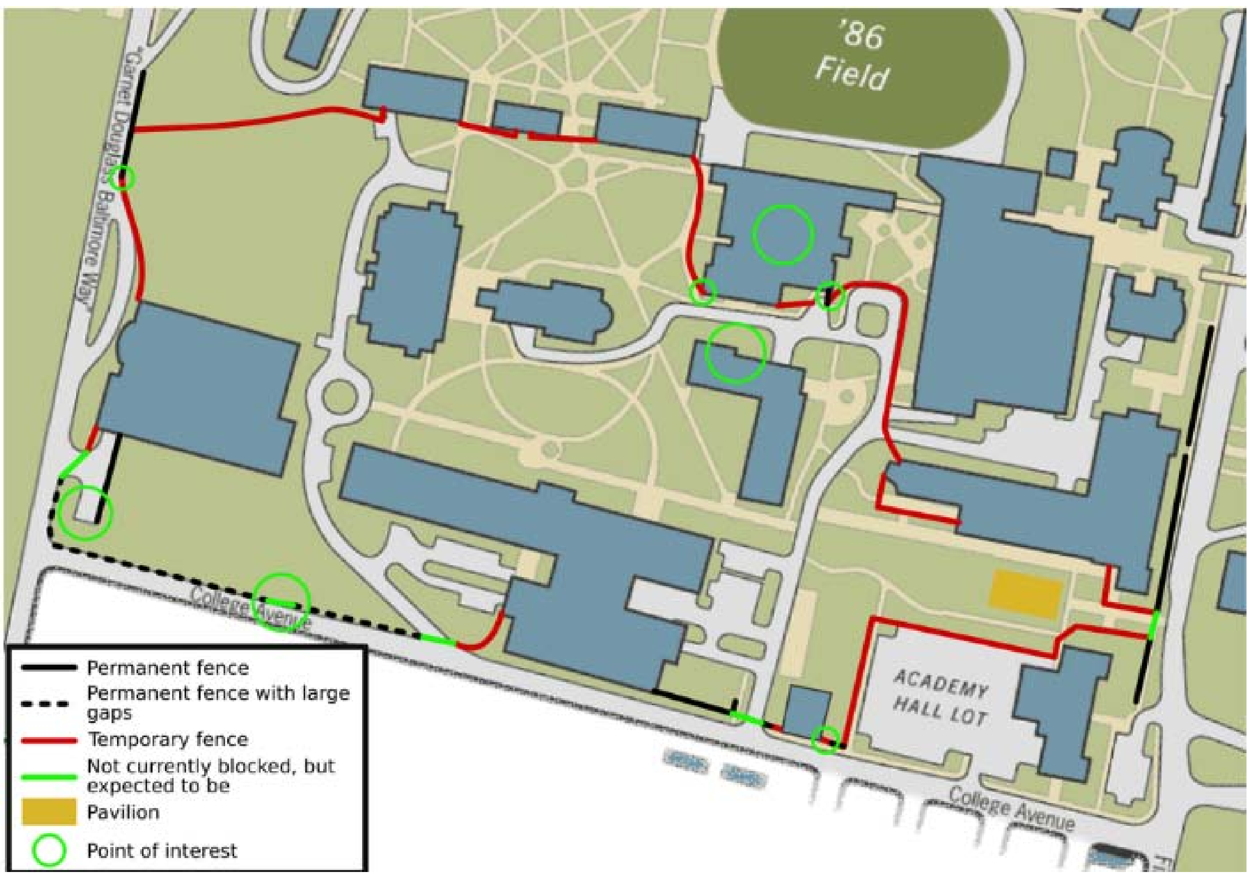

Figure 5.1: Locations Distribution (Deep Carbon Observatory, 2018)

Based on this information, figure 5.2 shows the connections between buildings that takes a maximum of two minutes to travel to and from. These are the target locations obtained out of all the building within the campus.

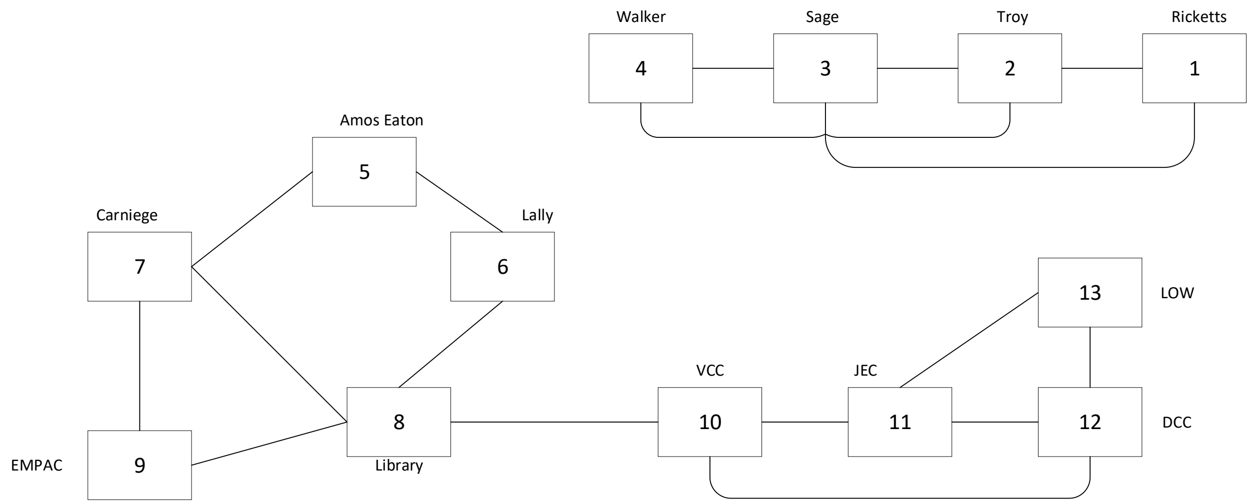

Figure 5.2: Target Locations

The sample dataset shown in figure 5.3 was obtained from the analysis of figure 5.2. The subsets identified are the target locations.

Subsets	Covers
1	1,2,3
2	1,2,3,4
3	1,2,3,4
4	2,3,4
5	5,6,7
6	5,6,8
7	5,7,8,9
8	6,7,8,9,10
9	7,8,9
10	8,10,11,12
11	10,11,12,13
12	10,11,12,13
13	11,12,13

Figure 5.3: Set Cover Dataset

5.4 Set Cover Problem

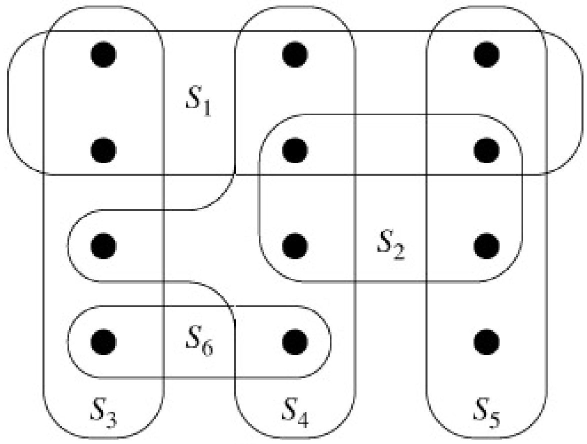

Figure 5.4: Set covering problem (Harman et al., 2020)

In the set cover problem, we are given a universal set U and a family of subsets $S_1 \ldots, S_k \subseteq U$. A set cover is a collection of subsets C from $S_1 \ldots, S_k$ whose union is the universal set U. Minimization of |C| is the primary goal. The problem of finding the optimum C is NP-Complete, but a greedy algorithm can give an $O(\log_e n)$ approximation to optimal solution (BroadHurst, 2017).

The greedy heuristic is the most appropriate approach for the set cover problem. The initial procedure is placing the largest subset in the set cover then indicate all its elements as covered. The subset is repeatedly added to the subset comprising of the largest number of uncovered elements until all the elements are covered. The greedy heuristic at all times, gives many sets as optimal and in practice, it regularly does a lot better than other approaches (Skiena, 1997).

5.5 Implementation Setup

The following steps were taken to compute the predicted value:

i. A location was entered and the buildings were identified. For a connection between a building and another to be valid, commuting from one building to another and then back to the initial building, a maximum of two minutes was set.

ii. Connections between the buildings were established and grouped into subsets.

iii. Constraints were identified. For a constraint to be valid, every building must be served by at least one healthcare facility. The constraints established were:

x1 + x2 + x3 >= 1 (Ricketts)

x1 + x2 + x3 + x4 >= 1 (Troy)

x1 + x2 + x3 + x4 >= 1 (Sage)

x2 + x3 + x4 >= 1 (Walker)

x5 + x6 + x7 >= 1 (Amos Eaton)

x5 + x6 + x8 >= 1 (Lally)

x5 + x7 + x8 + x9 >= 1 (Carniege)

x6 + x7 + x8 + x9 + x10 >= 1 (Library)

x7 + x8 + x9 >= 1 (EMPAC)

x8 + x10 + x11 + x12>= 1 (VCC)

x10 + x11 + x12 + x13 >= 1 (JEC)

x10 + x11 + x12 + x13 >= 1 (DCC)

x11 + x12 + x13 >= 1 (LOW)

Decision variable = $x_i \in \{0,1\} \forall i$

Equation 5.5: Decision Variable

iv. Calculation of the optimal solution. This will identify the buildings to set up the healthcare facilities. The general formulation of the set cover problem is

$$\text{Min} \sum i\, x_i$$

Subject to

$$\sum_j a_i j\, x_i \geqslant 1 \quad \forall i$$

$$x_i \in \{0,1\} \forall i$$

Equation 5.6: Set Cover Formula

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [20]:
```python
def set_cover(universe, subsets):
    # Find subsets that cover the universal set
    elements = set(e for s in subsets for e in s)
    # Check if the subsets cover the universe
    if elements != universe:
        return None
    covered = set()
    cover = []
    # Greedily add the subsets with the most uncovered points
    while covered != elements:
        subset = max(subsets, key=lambda s: len(s - covered))
        cover.append(subset)
        covered |= subset
    return cover

def main():
    universe = set(range(1, 14))
    s1 = set([1, 2, 3]); s2 = set([1, 2, 3, 4]); s3 = set([1, 2, 3, 4]); s4 = set([2, 3, 4]);
    s5 = set([5, 6, 7]); s6 = set([5, 6, 8]); s7 = set([5, 7, 8, 9]); s8 = set([6, 7, 8, 9, 10]);
    s9 = set([7, 8, 9]); s10 = set([8, 10, 11, 12]); s11 = set([10, 11, 12, 13]);
    s12 = set([10, 11, 12, 13]); s13 = set([11, 12, 13]);
    dicts = [
        {"name":"Ricketts", "data":s1}, {"name":"Troy", "data":s2}, {"name":"Sage", "data":s3}, {"name":"Walker", "data":s4},
        {"name":"Amos Eaton", "data":s5}, {"name":"Lally", "data":s6}, {"name":"Carniege", "data":s7},
        {"name":"Library", "data":s8}, {"name":"EMPAC", "data":s9}, {"name":"VCC", "data":s10}, {"name":"JEC", "data":s11},
        {"name":"DCC", "data":s12}, {"name":"LOW", "data":s13}
    ]
    subsets = [s1,s2,s3,s4,s5,s6,s7,s8,s9,s10,s11,s12,s13]
    cover = set_cover(universe, subsets)
    for c in cover:
        result = next(filter(lambda item: item["data"] == c, dicts))
        print(result["name"])

if __name__ == '__main__':
    main()
```

```
Library
Troy
JEC
Amos Eaton
```

Figure 5.5: Strategic Locations Calculation

v. Analysis of the optimal solution is undertaken. Factors required to set up a health facility are checked at this step. Priority is given to the buildings depending on the weight of the factors present. Highest priority is given to the building with the most weight and lowest priority is given to the building with the least weight. The dataset shown will be used to calculate the priority.

Code	Factor	Weight	Ricketts	Troy	Sage	Walker	Amos Eaton	Lally	Carniege	Library	EMPAC	VCC	JEC	DCC	LOW
1	Water distribution	5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	Sewage disposal	5	Yes	Yes	Yes	N/A	Yes		Yes	Yes		Yes	Yes	Yes	Yes
3	Electricity distribution	5	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
4	Emergency exits	4	No	No	Yes	Yes	No	No	Yes	No	No	No	No	No	No
5	Ample parking	3	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	Internet connectivity	4	Yes	No	No	No	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes
7	Legal license requirements	5	Yes	Yes	N/A	Yes	Yes	Yes	Yes		No	No	Yes		No
8	Good road networks	4	Yes	Yes	yes	No	Yes	Yes	Yes	No	Yes	No	No	Yes	No
9	Green building	4	No	Yes		Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
10	Building age[1-2 yrs]	5	yes	No	No	No	yes	No	Yes	No	No	No	No	No	No
11	Building age[3-4 yrs]	4	No	No	Yes	No	No	No	No	No	No	No	Yes	No	No
12	Building age[5-6 yrs]	3	No	Yes	No	No	No	Yes	No	No	No	No	No	No	No
13	Building age[7-8 yrs]	2	No	No	No	No	No	No	No	No	Yes	Yes	No	No	No
14	Building age[8 yrs and over]	1	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	Yes

Figure 5.6: Locations Factors

vi. The file is read and calculation of the priority is done resulting to an output of the buildings organized from top to bottom with priority decreasing from top to bottom.

```
# import libraries
import pandas as pd
import numpy as np

df = pd.read_csv("documents/priority model dataset.csv") # Read the csv file
df.replace(to_replace = np.nan, value = 0, inplace=True) # Replace nan values with a 0
df.replace(to_replace = ["No", "no"], value = 0, inplace=True) # Replace no values with a 0
df.replace(to_replace = ["Yes", "yes"], value = 1, inplace=True) # Replace yes values with a 1
location_columns = ["Library","Amos Eaton","JEC","Troy"] # initialise target locations/buildings as target columns
# Replace target column values where the value is 1 with the weight value in the same x axis
df[location_columns] = df[location_columns].mask(df[location_columns].apply(lambda x: x == 1), df['Weight'], axis=0)
df.loc['Total',:]= df.sum(axis=0) # Create a new row with totals of the target columns
new_df = df[location_columns].copy() # create a new dataframe with target columns
my_dict = [] # Initialise a new list
for (columnName, columnData) in new_df.iteritems():
    # append a dictionary with data and priority to the list
    my_dict.append({'name' : columnName, 'priority' : columnData.values[-1]})
print("The Locations/Buildings are sorted from highest to lowest priority: ")
my_sorted_dict = sorted(my_dict, key = lambda i: i['priority'], reverse=True) # Sorted data in descending order
for data in my_sorted_dict:
    print(data['name'])
```

```
The Locations/Buildings are sorted from highest to lowest priority:
Amos Eaton
JEC
Library
Troy
```

Figure 5.7: locations Priority Calculation

5.6 Model Testing

Testing was achieved by entering the target locations within the area and the location factors into the model and the results obtained were prioritized strategic locations to establish healthcare facilities. The target locations were identified by applying the time constraint. If it took a maximum of two minutes to go from one building to another and then back to the initial building, then the buildings are identified as target locations. The location factors to be identified within the buildings include water distribution, sewage disposal, electricity distribution, emergency exits, ample parking, internet connectivity legal license requirements, good road networks, age of building and whether the buildings are green buildings or not.

The tests done through the model's development and implementation processes include: unit testing to eliminate bugs and ensure the model functioned as expected, integration tests to verify that the algorithms used can be integrated to ensure that the functional and nonfunctional requirements of the model meet the user requirements provided. Usability test was also carried out to test the functionality of the model. This was carried out to verify and validate that the communication components function as expected, the model is easy to learn and use, the navigation flow is smooth and that the resultant system is user friendly and meets the user requirements and specifications.

Table 5.2: Test Cases

Test	Case	Outcome	Comment
1	Enter area to establish healthcare facilities.	Pass	No error occurred. Saved successfully
2	Establish building that take a maximum of 2 mins to traverse between.	Pass	Identified from the csv file
3	Identify buildings in the area with required information that meet the criteria set in test 2.	Pass	Successfully retrieved from a csv file with all location/building related information.
4	Determine the buildings to set up healthcare facilities that cover most of the people in terms of accessing the facilities.	Pass	The buildings were identified
5	Show priority of the buildings to set up the healthcare facilities.	Pass	Priority of buildings were shown from highest to lowest.

Chapter 6: Discussions

6.1 Introduction

This chapter discusses the results of the research in relation to the objectives. It outlines how this has been achieved through discussions in the literature review, to the design and implementation processes. It also discusses the parameters used as a measure for the results of the implementation. An outline on the resultant findings and suggestions on future research is also mentioned.

This research study analyzed the criteria for choosing facility location. After reviewing the methods, the researcher chose to focus on a heuristic method which is the greedy algorithm that is well suited to handle the set cover problem. The set cover model used identified the strategic healthcare locations which were then prioritized from highest to lowest using the priority algorithm.

6.2 Factors Affecting Locations of Healthcare Facilities

The first objective was used to determine the factors affecting location of a healthcare facility. Based on the literature reviewed, it was revealed that the major factors affecting the location of a healthcare facility are ease of access by the people, good road networks, water distribution, electricity distribution, sewage disposal, parking space, internet connectivity, legal license requirements, green building design and age of existing buildings to set up the healthcare facilities.

6.3 Existing Techniques Used to Determine the Locations of Healthcare Facilities

The second objective was to analyze the existing techniques used to determine the location of a healthcare facility. From the literature review, it was found out that organizations use cost effective measures as the sole and main objective to establish healthcare facilities.

Findings from the literature review also showed that multiple objectives were not majorly considered hence factors such as time taken to reach the healthcare facilities and coverage by the healthcare facilities was not determined. Factors used to determine the location of a healthcare facility was also analyzed by the study. These factors generally affect existing locations to establish healthcare facilities such as a clinic or a new hospital.

Literature review also revealed that three models are used to determine the location of healthcare facilities. They include set covering model, the maximal covering model and the P-median model.

The set covering model was used in this research due to the model focusing on coverage hence finding optimal locations where the entire population in an area is covered.

6.4 Prediction Model for Determining Healthcare Facility locations by the Kenyan County Government

The third objective was to develop a prediction model for determining healthcare facility locations by the Kenyan county government. Research findings reveal that the model developed not only identified locations to set up healthcare facilities but also determined the locations priority depending on the factors of the location. This provides a sequence to establish the healthcare facilities. If the factors of the locations are updated, the model is run and gives a new sequence of locations with priority from the highest to the lowest. This is particularly important for future development of the facilities.

The literature reviewed discusses criteria for choosing facility location and solution methods which guided the development of the prediction model for determining locations to set up healthcare facilities.

6.5 Testing of the Developed Model

The last objective was to carry out testing of the developed model. Testing was achieved by entering the target locations within the area and the location factors into the model and the results obtained were prioritized strategic locations to establish healthcare facilities. The target locations were identified by applying the time constraint. If it took a maximum of two minutes to go from one building to another and then back to the initial building, then the buildings are identified as target locations.

6.6 Validity of the Developed System

Locations for establishing facilities are always considered in research undertaking with a sole purpose of reducing costs while offering high quality health services. This has led to the development of many applications and algorithms. However, the applications do not cater for scenarios where patients cover least amount to reach the healthcare facilities or the coverage of the healthcare facilities depending on their location. The proposed system accounts for least time to reach these facilities and maximum coverage of these facilities. The proposed system is useful to government agencies that can use the system to predict the locations to set up healthcare facilities now and in future, therefore, help them make better and informed decisions with regards to planning in developed and developing areas. The model can be used alongside other applications and algorithms to help inform users on the locations to establish the facilities. Finally, the model does not in any way require user data for it to run.

6.7 Advantages of the Developed System to Current Systems

The priority algorithm used during this research study is the most significant advantage of the developed system to current systems. The developed system provided a solution to prioritize the strategic locations identified depending on the strategic locations factors to establish a healthcare facility. Human intervention is only required when entering the area where healthcare facilities are required or need to be established. The user enters an area and the system returns back a result which are specific locations where healthcare facilities can be established. This improves user experience especially for computer illiterate users. The system can be accessed through the web hence is accessible to the user and is not dependent on mobile phones or operating systems on laptops or desktops. Implementation of the model was done through the web, hence it is evident that the model can be integrated with a system that runs on the web.

6.8 Research Contributions

The model developed provided a solution to locating various locations that are easily accessible by the people in that area. The model further identifies the priority in which to establish the healthcare facilities from highest to lowest. The model will help organizations, especially the Kenyan county government, to establish healthcare facilities in a planned manner accounting for future locations to establish the facilities. This enables patients get to the facilities in the least amount of time while also reducing the decision-making process to determine next location to establish a healthcare facility.

6.9 Challenges Encountered

It was a challenge obtaining data regarding location factors. The data obtained had some buildings which completely lacked factors recorded. It was time consuming going through the subset covers as the data was not formatted in an easily comprehensible manner.

Chapter 7: Conclusion and Recommendations

7.1 Conclusion

The objective of this research study was to develop a model that can predict the most suitable locations to establish healthcare facilities. It is easy to determine that healthcare facilities are needed but the problem is identifying the strategic locations. Hence, this study began by reviewing factors that affect healthcare facilities. Existing methods and techniques for identifying healthcare facilities were reviewed and the research chose to focus on set cover problem to ensure maximum coverage of all locations within an area. This research showed that significant progress in predictions can effectively and efficiently contribute to decision making hence, reducing the time taken to deliberate on where healthcare facilities should be established. This research then presented a structured approach on how to solve the problem of healthcare facilities locations by using the set cover problem and a prioritization algorithm to come up with a model for identifying strategic locations to establish healthcare facilities.

The developed model performed as required and was able to identify strategic locations to establish healthcare facilities in an area provided by a user and determine priority of strategic locations identified. The user could then save the strategic locations results into a database and later on retrieve the results when required. Prioritizing the determined strategic locations is a unique approach introduced in this research.

7.2 Recommendations

The healthcare facility location model for determining strategic locations for the facilities was suitable in solving the problem identified in this study. However, the researcher noted that there was more that could be done in this area of research and therefore recommends the following.

i. More data could be included in datasets especially the location factors. It was difficult to find location factors in the dataset used for this research.
ii. Increasing availability of the model by developing a mobile application.
iii. The model can be extended and be used in determining strategic locations for schools, restaurants and hotels.

7.3 Suggestions for Future Research

More research could be done on how to estimate the number of patients being served at the healthcare facilities in each of the strategic locations identified. This data would then be used to restock supplies with a priority set on the healthcare facility that receives the most patients. Future research can also explore identification of strategic locations in multiple areas. The use of visual data can be used especially in the collection of location factors as datasets do not have comprehensive data on the location factors. The system on which the model runs on should be able to provide real time data on the areas, locations within the areas and the location factors. This is only possible if the data is available and maintained on a regular basis.

References

Afshari, H. & Peng, Q. (2014). Challenges and Solutions for Location of Healthcare Facilities. *Ind Eng Manage,* 3(2), 1 - 12.

Ahmadi-javid, A., Seyedi, P., & Syam, S. S. (2016). A survey of healthcare facilities. *Computers and Operations Research*, 79, 223-263.

Araz, C., Selim, H., & Ozkarahan, I. (2007). A fuzzy multi-objective covering-based vehicle location model for emergency services. *Computers and Operations Research*, 34, 705-726.

Baron, O., Berman, O., & Krass, D. (2008). Facility location with stochastic demand and constraints on waiting time. *Manufacturing and Service Operations Management*, 10, 484-505.

Benneyan, J. C., Musdal, H., Ceyhan, M. E., Shiner, B., & Watts, B. V. (2012). Specialty care single and multi-period location-allocation models within the Veterans Health Administration. *Socio-economic planning sciences*, 46, 136-148.

Blackwell, T. H., & Kaufman, J. S. (2002). Response time effectiveness: comparison of response time and survival in an urban emergency medical services system. *Acad Emerg Med*, 9, 288-295.

Bozkaya, B., Yanik, S. & Balcisoy, S. (2010). A GIS-based optimization framework for competitive multi-facility location-routing problem. *Netw Spat Econ*, 10, 297–320.

BroadHurst, M. (2017). *Greedy Set Cover in Python*. Retrieved February 27, 2020, from http://www.martinbroadhurst.com/greedy-set-cover-in-python.html

Cameron, R. (2011). Mixed Methods Research: The Five Ps Framework. *The Electronic Journal of Business Research Methods*, 9(2), 96-108.

Canadian Institute for Health Information (2012). *National Health Expenditure Trends, 1975 to 2012*. Canada: The Canadian Institute for Health Information.

Caunhye, A. M., Nie, X., & Pokharel, S. (2012). Optimization models in emergency logistics: A literature review. *Socio-economic planning sciences*, 46, 4-13.

Centers for Medicare & Medicaid Services. (2008). Selecting a Development Approach. *United States Department of Health and Human Services*

Chu, S. C. K., & Chu, L. (2000). A modeling framework for hospital location and service allocation. *International Transactions in Operational Research*, 7, 539-568.

Cich, B. C. (2017). *Four criteria for selecting health care sites*. Retrieved March 10, 2020, from https://www.hfmmagazine.com/articles/3116-four-criteria-for-selecting-health-care-sites

Contreras, I., Cordeau, J. F. & Laporte, G. (2010). The dynamic uncapacitated hub location problem. *Transp Sci,* 45, 18–32.

Daskin, M. S. (2008). What You Should Know About Location Modeling. *Naval Research Logistics*, 55, 283 - 294.

Daskin, M. S. & Dean L. K. (2004). Location of health care facilities. In Brandeau, M. L., Sainfort, F., Pierskalla, W. P. (eds) Operations research and health care. *A Handbook of Methods and Applications*, 43-76.

De Aguiar Vallim Fo, A. R., & da Silva Mota, I. (2012). Optimization models in the location of healthcare facilities: a real case in Brazil. *Journal of Applied Operational Research*, 4(1), 37-50.

Deep Carbon Observatory. (2018). *RPI campus Map*. Retrieved February 11, 2020, from https://www.deepcarbon.net/sites/default/files/2018-04/rpi_campus_map_2010.pdf

Degu, G., & Yigzaw, T. (2006). *Research Methodology*. Gondor: University of Gondor.

Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Karall, M., & Reimann, M. (2005). Heuristic solution of an extended double-coverage ambulance location problem for Austria. *Central European Journal of Operations Research*, 13, 325-340.

Dzator, M & Dzator, J. (2016). Health, emergency facilities and development: locating facilities to serve people and development better. *The Journal of Developing Areas*, 5(5), 31-142.

Epping, G. M. (1982). Important factors in plant location in 1980. *Growth and Change*, 13(2), 47-51.

Enago Academy (2019). *Identifying Research Gaps to Pursue Innovative Research*. Retrieved Apr 5, 2020 from https://www.enago.com/academy/identifying-research-gaps-to-pursue-innovative-research/

Fowler, M. (1996). *Analysis patterns: reusable objects models*: USA Addison-Wesley Longman Publishing Co., Inc.

Gendreau, M. (1997). Solving an ambulance location model by tabu search. *Location Science*, 5, 75-88.

Gendreau, M., Laporte, G., & Semet, F., (2001) A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, 27, 1641-1653.

Ghaderi, A. & Jabalameli, M. S. (2013) Modeling the budget-constrained dynamic uncapacitated facility location-network design problem and solving it via two efficient heuristics: a case study of health care. *Math Comput Model,* 57, 382–400.

Gu, W., Wang, X., & McGregor, S. E. (2010). Optimization of preventive health care facility locations. *Int J Health Geogr*, 9, 17.

Harman, M., Kim, S., Lakhotia, K., Mcminn, P., & Yoo, S. (2020). *Reducing Oracle Cost in Search Based Test Data Generation.* Retrieved Apr 21, 2019 from https://www.researchgate.net/publication/265799358_Reducing_Oracle_Cost_in_Search_Based_Test_Data_Generation

Hinojosa, Y., Kalcsics, J., Nickel, S., Puerto, J. & Velten, S. (2008). Dynamic supply chain design with inventory. *Comput Oper Res,* 35, 373–391.

Hoadley, D. E., Jorgensen, B., Masters, C., Tuma, N., & Wulff, S. (2010). Strategic Facilities Planning: A Focus On Health Care. *Journal of Service Science (JSS).* 3. 10.19030/jss.v3i1.797.

Huang, R., Kim, S., & Menezes, M. B. C. (2010). Facility location for large-scale emergencies. *Annals of Operations Research*, 181, 271-286.

Kouvelis, P. and G. Yu (1996). *Robust Discrete Optimization and its Applications.* Kluwer Academic Publishers, Dordrecht.

Li, X., Zhao, Z., Zhu, X., & Wyatt, T. (2011). Covering models and optimization techniques for emergency response facility location and planning: A review. *Mathematical Methods of Operations Research*, 74, 281-310.

Mahar, S., Bretthauer, K. M., & Salzarulo, P. A. (2011). Locating specialized service capacity in a multi-hospital network. *European Journal of Operational Research*, 212, 596-605.

Mestre, A. M., Oliveira, M. D., & Barbosa-Póvoa, A. (2012). Organizing hospitals into networks: A hierarchical and multiservice model to define location, supply and referrals in planned hospital systems. *OR Spectrum*, 34, 319-348.

Ministry of Health Kenya. (2014). *Health Sector Human Resources Strategy 2014 – 2018*. Nairobi: Ministry of Health.

Mitropoulos, P., Mitropoulos, I., & Giannikos, I. (2013). Combining DEA with location analysis for the effective consolidation of services in the health sector. *Computers and Operations Research*, 40, 2241-2250.

Murawski, L., & Church, R. L. (2009). Improving accessibility to rural health services: The maximal covering network improvement problem. *Socio-economic planning sciences*, 43, 102-110.

Narad, R. A., & Driesbock, K. R. (1999). Regulation of ambulance response times in California. *Prehosp Emerg Care*, 3, 131-135.

Onyango, M. M. (2018). *A Prototype to forecast trends in rainfall amounts in Kenya*. Nairobi, Kenya: Strathmore University. Retrieved from https://suplus. strathmore.edu/handle/11071/6000

Peng, Q. & Afshari, H. (2014). Challenges and Solutions for Location of Healthcare Facilities. *Industrial Engineering & Management*. 03. 10.4172/2169-0316.1000127.

Pons, P. T., Haukoos, J. S., Bludworth, W., Cribley, T., Pons, K. A., et al. (2005). Paramedic response time: does it affect patient survival? *Acad Emerg Med*, 12, 594-600.

Rahman, S., & Smith, D. K. (1999). Deployment of rural health facilities in a developing country. *Journal of the Operational Research Society*, 50, 892-902.

Rajagopalan, H. K., Saydam, C., & Xiao, J. (2008). A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers and Operations Research*, 35, 814-826.

Romauch, M. & Hartl, R. F. (2005). Dynamic facility location with stochastic demands. *Stoch Algorithm Found Appl 3777*, 180–189.

Sahin, G., Süral, H., & Meral, S. (2007). Locational analysis for regionalization of Turkish Red Crescent blood services. *Computers and Operations Research*, 34, 692-704.

Seyedhosseini, S.M., Makui, A. & Shahanaghi, K. (2016). Models, solution, methods and their applicability of dynamic location problems (DLPs) (a gap analysis for further research). *J Ind Eng Int,* 12, 311–341.

Shariff, S. S. R., Moin, N. H., & Omar, M. (2012). Location allocation modeling for healthcare facility planning in Malaysia. *Computers and Industrial Engineering*, 62, 1000-1010.

Skiena, S, S. (1997). *The Stony Brook Algorithm Repository*. Retrieved Apr 21, 2019 from https://www8.cs.umu.se/kurser/TDBAfl/VT06/algorithms/BOOK/BOOK5/NODE201.HTM

Snyder, L. V. (2006). Facility location under uncertainty: a review. *IIE Transactions Taylor & Francis*, 38, 547-564.

Sonnen, C. (2008). *Infrasrtucture as Economic Stimulus*. Canada: Informetrica Limited.

Syam, S. S., & Côté, M. J. (2010). A location-allocation model for service providers with application to not-for-profit health care organizations. *Omega*, 38, 157-166.

The Value Group. (2004). *The Value of Investment in Health Care-Resources*. Washington, USA: Federation of American Hospitals.

Terouhid, S. A., Ries, R., & fard, M. M. (2012). Towards sustainable facility location: A literature review. *Journal of Sustainable Development*, 5(7), 18-34.

Trochim, M. K. W. (2020). *Research Methods Knowledge Base*. Retrieved June 22, 2020, from https://conjointly.com/kb/sampling-in-research/

Wang, L., Shi, H., & Gan, L. (2018). Healthcare Facility Location-Allocation Optimization for China's Developing Cities *Utilizing a Multi-Objective Decision Support Approach. Sustainability*. 10. 4580. 10.3390/su10124580.

Worthington, A. (2004). Frontier efficiency measurement in healthcare: A review of empirical techniques and selected applications. *Medical Care Research and Review*, 61(2), 1-36.

Yantzi, N, Rosenberg, M. W., Burke, S. O., & Harrison, M. B. (2001). The impacts of distance to hospital on families with a child with a chronic condition. *Soc Sci Med*, 52, 1777-1791.

Zhang, Y., Berman, O., & Verter, V. (2009). Incorporating congestion in preventive healthcare facility network design. *European Journal of Operational Research*, 198, 922-935.

Zimmerman, S. M. (1997). Factors influencing Hispanic participation in prostate cancer screening. *Oncol Nurs Forum*, 24, 499-504.

Appendix

Appendix A: Strathmore University Ethical Approval Certificate

Strathmore
UNIVERSITY

4th February 2020

Mr Isaboke, Edward
edward.nyakeruma@strathmore.edu

Dear Mr Isaboke,

RE: Prediction Model for Determining Health Care Facility Locations by the Government

This is to inform you that SU-IERC has reviewed and **approved** your above research proposal. Your application approval number is **SU-IERC0577/19.** The approval period is **4th February, 2020 to 3rd February, 2021.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-IERC.
iii. Death and life threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-IERC within 72 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-IERC within 72 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to SU-IERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology and Innovation (NACOSTI) https://oris.nacosti.go.ke and also obtain other clearances needed.

Yours sincerely,

for. Dr Virginia Gichuru,
Secretary; SU-IERC

Cc: Prof Fred Were,
Chairperson; SU-IERC

STRATHMORE UNIVERSITY INSTITUTIONAL ETHICS REVIEW COMMITTEE (SU-IERC)
04 FEB 2020
TEL: +254 (0)703 034 000
P. O. Box 59857 - 00200
NAIROBI - KENYA

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email info@strathmore.edu www.strathmore.edu

Appendix B: Originality Report