



Strathmore
UNIVERSITY

Strathmore University
SU+ @ Strathmore
University Library

Electronic Theses and Dissertations

2019

Dynamic vehicle routing model using geometric Brownian motion

Joseph M. Muroki
Faculty of Information Technology (FIT)
Strathmore University

Follow this and additional works at <https://su-plus.strathmore.edu/handle/11071/6779>

Recommended Citation

Muroki J. M. (2019). *Dynamic vehicle routing model using geometric Brownian motion* (Thesis).

Strathmore University. Retrieved from <https://su-plus.strathmore.edu/handle/11071/6779>

This Thesis - Open Access is brought to you for free and open access by DSpace @Strathmore University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DSpace @Strathmore University. For more information, please contact librarian@strathmore.edu

**DYNAMIC VEHICLE ROUTING MODEL USING GEOMETRIC BROWNIAN
MOTION**

Mungai, Joseph Muroki

Master of Science in Information Technology



2019

**DYNAMIC VEHICLE ROUTING MODEL USING GEOMETRIC BROWNIAN
MOTION**

Mungai, Joseph Muroki

**A research thesis submitted in partial fulfillment of the requirements for the Degree of
Master of Science in Information Technology at Strathmore University.**



Faculty of Information Technology

Strathmore University

Nairobi, Kenya

June, 2019

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

© No part of this thesis may be reproduced without the permission of the author and Strathmore University

Mungai Joseph Muroki

.....

June 2019

Approval

The thesis of Mungai Joseph Muroki was reviewed and approved by the following:

Prof. Ismail Ateya Lukandu, D. Sc.,
Associate Professor, Faculty of Information Technology,
Strathmore University

Dr. Joseph Orero, PhD,
Dean, Faculty of Information Technology,
Strathmore University

Prof. Ruth Kiraka,
Dean, School of Graduate Studies,
Strathmore University

Abstract

Traffic jams are one of the most common problems associated with urban areas worldwide. The effects of traffic congestion are usually felt during peak hours of the day and in cases of accidents or bad weather. Essentially, Traffic jams occur with or without the aid of traffic Marshalls or traffic lights. However, the traffic jams get worse with absence of these agents especially in developing countries where public transport vehicles have little regard to traffic rules and regulations. As such, other modes of traffic avoidance and mitigation are required to mitigate the effects of traffic jams. A majority of city dwellers have resulted to waking up early in order to beat the traffic or leaving work early or later than usual. Therefore, to mitigate traffic jams there must be the provision of an alternative solution to the traffic menace. Response time by members of the police force would play a key role in reducing traffic congestion during accidents or incidents. The use of applications such as Waze and Google maps have greatly helped in navigation as well as informing the public on traffic situations on the roads. Radio stations and other social platforms like Twitter and Facebook have become a source of information with regards to traffic situations. Essentially, to substantially help reduce the effects of traffic jams in urban areas, more still needs to be done in terms of providing information to both road users and authorities with regards to traffic situations. Therefore, the need for alternative sources of information that aims at providing relevant information in regards to the traffic congestions. Notably, proper information through social media and other channels has led to a reduced response time by traffic Marshalls as well as a way to warn motorists to avoid areas with accidents as well as high traffic volumes. This research work proposes a dynamic vehicle routing model which implements geometric Brownian motion to help divert traffic to routes less congested. Notably, the model requires a large amount of data in order to accurately predict the expected traffic volume at a particular time. The system calculates the expected traffic at a particular time using variables calculated from the historical data of the route. It is therefore important for historical volume data to exist for the model to work. Based on the provided data, the system then calculates expected traffic volume for the routes entered by the user at a particular time and displays the same to the user. The system is able to calculate the expected traffic volume at a time (t) and is able to find an expected average of traffic on a route.

Keywords: *Traffic congestion, Traffic jams, Traffic volume, divert.*

Table of Contents

Declaration.....	ii
Abstract.....	iii
List of Figures.....	vii
List of Tables.....	ix
List of Equations.....	x
List of Abbreviations/Acronyms.....	xi
Acknowledgements.....	xii
Dedication.....	xiii
Chapter 1: Introduction.....	1
1.1 Background of the Study.....	1
1.2 Problem Statement.....	2
1.3 Aim.....	2
1.4 Research Objectives.....	2
1.5 Research Questions.....	3
1.6 Justification.....	3
1.7 Scope and Limitations.....	3
Chapter 2: Literature Review.....	5
2.1 Introduction.....	5
2.2 Factors causing Increase in Traffic Volume.....	5
2.2.1 Population Growth.....	5
2.2.2 Improved Economic Status.....	6
2.3 Approaches used in Dynamic Vehicle Routing.....	7
2.3.1 Greedy Algorithm.....	7
2.3.2 The Multiple Plan Approach.....	8
2.3.3 Traffic Management Model.....	10
2.3.4 Pheromone-based Traffic Light Control.....	12
2.4 Mobile Traffic Applications.....	13
2.4.1 Waze.....	13
2.4.2 Google Maps.....	14
2.5 Algorithms Applicable to Traffic Routing.....	15
2.5.1 Complexity.....	16
2.5.2 Bellman-Ford Algorithm.....	17

2.5.3 Floyd-Warshall Algorithm	18
2.5.4 K-Shortest Paths (KSP) Algorithms	19
2.5.5 Yen's Algorithm.....	19
2.5.6 Eppstein's Algorithm	19
2.6 Summary	20
2.7 Brownian Motion	20
2.7.1 Why Brownian Motion.....	21
2.8 Stochastic Traffic Volume Prediction	23
2.9 Conceptual Framework	25
Chapter 3: Research Methodology.....	27
3.1 Introduction	27
3.2 Research Design.....	27
3.3 Location of the Study	27
3.4 Population and Sampling	28
3.4.1 Population.....	28
3.4.2 Sampling.....	28
3.5 Data Collection.....	28
3.6 Model Development.....	28
3.7 System Development Methodology	29
3.7.1 Agile System Development	29
3.7.2 Data Retrieval and Analysis	29
3.7.3 Development of the model	30
3.7.4 Validation of the model	30
3.8 System Design.....	30
3.6 Research Quality, Reliability and Validity	30
3.7 Ethical Considerations.....	31
Chapter 4: System Analysis and Design	32
4.1 Introduction	32
4.2 Requirement Analysis	32
4.2.1 Functional Requirements	32
4.2.2 Non-Functional Requirements.....	32
4.3 System Design and Architecture	33
4.3.1 System Model Architecture	33
4.3.2 Use Case Diagram	34

4.3.3 Data Flow Diagram	35
4.3.4 Sequence Diagram	38
4.3.5 Partial Domain Model	38
4.3.6 Database Schema	39
Chapter 5: System Implementation and Testing	40
5.1 Introduction	40
5.2 Hardware and Software Environment	40
5.3 Dataset Pre-processing	41
5.4 System Implementation	43
5.4.1 Storing System Details	43
5.4.2 System Testing	43
5.4.3 Brownian Motion Algorithm	44
5.4.4 Using Developed Model	45
Chapter 6: Discussions	46
6.1 Model Evaluation	46
6.1.1 Monte Carlo Simulation	46
6.1.2 Central limit Theorem	47
6.1.3 Brownian Motion	47
6.1.4 Normal Distribution	49
6.1.5 Fore casting Evaluation	49
6.2 Validity of the proposed solution	52
6.3 Limitations of the Application	52
Chapter 7: Conclusions and Recommendations	54
7.1 Conclusion	54
7.2 Recommendations	54
7.3 Suggestions and Future Research	55
References	56
Appendix A: Sample Data	58

List of Figures

Figure 2-1 : Kenya - Urban population as a share of total population.....	6
Figure 2-2 : Greedy Algorithm	7
Figure 2-3 : Dispatching situation involving two vehicles	8
Figure 2-4 : Vehicles depositing basic traffic pheromone	10
Figure 2-5 : Realization of pheromone system	12
Figure 2-6 : Sample Waze Route planning Use case	14
Figure 2-7 : Sample Google Maps Route planning Use case	15
Figure 2-8 : A graph or network with nodes and paths bearing the path costs.....	15
Figure 2-9 : Brownian Movement.....	21
Figure 2-10 : Markov Process.....	22
Figure 2-11 : Gaussian Process two sample set	23
Figure 2-12 : Conceptual Framework	26
Figure 3-1 : Agile Programming Process.....	29
Figure 4-1 : System Model	34
Figure 4-2 : Proposed Algorithm Use Case	35
Figure 4-3 : Context Diagram	36
Figure 4-4 : Level 0 Diagram.....	37
Figure 4-5 : Flow Chart.....	37
Figure 4-6 : Sequence Diagram	38
Figure 4-7 : Partial Domain Model.....	39
Figure 4-8 : Database schema	39
Figure 5-1 : Sample Dataset.....	41
Figure 5-2 : Test Run	42
Figure 5-3 : Brownian Algorithm in Pandas.....	44
Figure 5-4 : Mapbox	45
Figure 6-1 : Central Limit Theorem.....	47
Figure 6-2 : Probability one	48
Figure 6-3 : Probability two.....	48
Figure 6-4 : Normal Distribution	49
Figure 6-5 : Simulation one	50
Figure 6-6 : Simulation two	50

Figure 6-7 : Sample Comparison values51
Figure 6-8 : Expected future volume Test-run.....51
Figure 8-1 : Sample Data58



List of Tables

Table 2-1 : Results by Traffic Management Models	12
Table 6-1 : Hardware and Software Environment	40



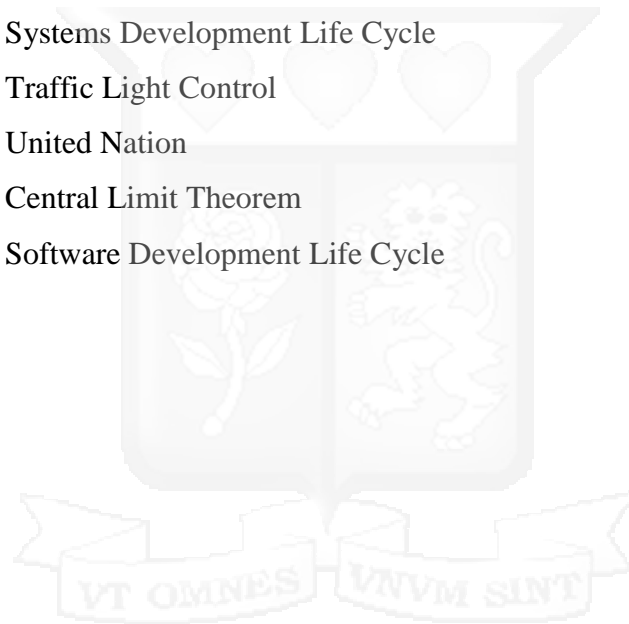
List of Equations

Equation 2.1 : Static Problem Equation.....	9
Equation 2.2 : Traffic Pheromone equation.....	11
Equation 2.3 : Intention pheromone equation one.....	11
Equation 2.4 : Intention pheromone equation two.....	11
Equation 2.5 : Evaporation and Propagation equation.....	11
Equation 2.6 : Diffusion equation.....	24
Equation 2.7 : Geometric Brownian Equation.....	24
Equation 6.1 : Monte Carlo Simulation.....	46



List of Abbreviations/Acronyms

CBD	-	Central Business District
DSDM	-	Dynamic Systems Development Method
DUA	-	Dynamic User Assignment
GBM	-	Geometric Brownian motion
GDP	-	Gross Domestic Product
IDE	-	Integrated Development Environment
OCR	-	Optical Character Recognition
PPP	-	Public-Private Partnership
SDLC	-	Systems Development Life Cycle
TLC	-	Traffic Light Control
UN	-	United Nation
CLT	-	Central Limit Theorem
SDLC	-	Software Development Life Cycle



Acknowledgements

I would like to express my deepest gratitude to God, who has made all this possible. I would also like to appreciate my supervisor, Prof. Ismail Ateya, who has been extremely helpful and welcoming throughout the duration of the Masters program.



Dedication

This research work is dedicated to my family who have been a continued source of motivation and encouragement all through the undertaking of the Masters program.



Chapter 1: Introduction

1.1 Background of the Study

According to a study conducted by the World Health Organization, (2018), the average person living in an urban area spends close to four hours in traffic every day. These do not include situations of bad weather or accidents, which are prone to considering the state of the majority of roads in developing countries. All this can be attributed to an increase in population in urban areas backed by poor infrastructural development and a rise in the Gross Domestic Product (GDP). Majority of persons living in urban areas are mostly the middle-class group where the average household has at least two vehicles (World Health Organization, 2018).

Vehicular traffic is the aggregation of vehicles coming and going in a particular locality("Dictionary, Encyclopedia & Thesaurus - The Free Dictionary," n.d.). The problem of vehicle routing is not new. Scholars and academicians have developed a myriad of solutions ranging from traffic flow systems and models to help ease if not eliminate this problem. The number of hours spent in traffic seems to be increasing by the day.

The increase in population and the high volume of vehicles in the cities are all straining for the same minimal resources that are available. With the population in urban areas growing steadily, there is a need to tackle the issue of traffic on our roads. Infrastructure development is one of the ways this can be handled. However, with policies and bureaucracies in government institutions this may take long to implement. Hartgen et al.(2009) did a study on the effect of traffic to the economy of urban cities. They found out that access to the CBD was a crucial aspect for most urban areas as majority of the population live within these areas. In the cities studied, reducing congestion would boost Gross Regional product by 6 to 30 percent if targeted at suburbs, malls, and universities. The economic gains would be 4 to 10 percent if targeted at CBDs, and just 2 to 8 percent if targeted at airports.

Currently, different applications enable vehicle routing and navigation through traffic with the use of various algorithms such as the A-star Algorithm used by Google maps or Dijkstra's Algorithm which calculates the shortest path between two points. The algorithms provide motorists and users with information on the shortest distance between two points. However, the algorithms do not give a value of the expected Vehicle traffic volume of the routes chosen by motorists. Such information would serve in the timely deployment of traffic Marshalls as well as enable motorists to make accurate and timely plans during travel.

1.2 Problem Statement

The pervasiveness of the traffic menace especially in urban areas presents a major problem to the society. The trend of urbanization has largely contributed to swarming of limited urban road networks with vehicles (Wang, Djahel, McManis, McKenna, & Murphy, 2013). According to a report by Schrank, Eisele, & Lomax, (2012) the economic loss in the United States in 2011 due to traffic delays and fuel consumption was over 121 billion dollars. Nairobi ranked was one of the second-worst cities around the globe in the year 2017 in terms of traffic congestion (Mutavi, 2017). According to a Serbia-based website *numbeo.com*, in the year 2018, Kenya was placed at number eight with a traffic index of 56.43 while cities such as Dhaka (Bangladesh), Kolkata (India) and Jakarta (Indonesia) are on top of the chart with time indexes of 63.94, 62.80 and 58.09 respectively (“Traffic Index by City 2018 Mid-Year,” n.d.). Traffic Index is a compound index of time spent in traffic due to travelling to work, approximation of time consumption displeasure, carbon dioxide consumption approximation in traffic and general inadequacies in the transportation system (Mutavi, 2017).

Notably, delays in traffic may result in loss of valuable time as well as affect the economic welfare of a country (Mutavi, 2017). A number of vehicle routing algorithms have been developed such as A-star algorithm which is utilized by google, Dijkstra’s algorithm which utilizes the shortest path mechanism to get to point as well as a host of other routing algorithms (Wang et al, 2013). However, the traffic menace continues to be a thorn in the flesh of the society. Therefore, more research needs to be carried out in the area of traffic routing. A probability based routing model that utilizes historical vehicle volume data to predict the expected traffic volume of a route at any specific time can help improve the current traffic situation.

1.3 Aim

The research aims at developing a web-based routing model that utilizes Geometric Brownian motion algorithm to inform motorists and users of the expected Vehicle Volume on routes they intend to use or ply.

1.4 Research Objectives

- (i) To investigate the existing algorithms used in traffic routing.
- (ii) To investigate the challenges associated with traffic routing.
- (iii) To appraise existing traffic flow systems, models and applications.

- (iv) To develop a Route suggestion model using Geometric Brownian Motion.
- (v) To test the developed model.

1.5 Research Questions

- (i) What are the major causes of traffic routing?
- (ii) What are the current challenges affecting motorists and commuters due to heavy traffic?
- (iii) How can the existing traffic flow systems, models and applications improve efficiency in the transport sector?
- (iv) How can the Geometric Brownian motion model if implemented, re-route traffic?
- (v) How valid is the Geometric Brownian motion model?

1.6 Justification

Firstly, the outcomes of the research will provide commuters and motorists with information on the expected vehicle volume on the chosen routes based on the time of day. This will assist commuters and motorists on decision making on which routes to take based on the expected vehicle volume. Secondly, the research provides a robust method of informing relevant government bodies on investment of road infrastructure. This is particularly useful where forecasting of expected vehicle volume data is required on roads. Consequently, government agencies such as the traffic police can use this data to know where best to deploy traffic Marshalls based on the expected volume of vehicles. This will in turn improve the planning process by government bodies and proper plans and preparations can be put in place. Thirdly, by modelling the necessary framework for expected vehicle volume, experts in the industry can implement the framework to other sectors of the economy. In conclusion, the study plays a key role in contributing to research and innovation for the academic professionals. It aims at providing insights currently not in the literature thereby important for future researchers in the field.

1.7 Scope and Limitations

The proposed model will only calculate the expected traffic vehicle volume of a route based on the historical vehicle volume of the same route and display the same to the user. The model will process historical vehicle volume data from a route and calculate the expected traffic volume for the route at different times of the day. The researcher has submitted that there might data limitations with regards to the study, as there might be occurrences where

primary and secondary research techniques might not be able to provide the necessary information in order to satisfy the research objectives.



Chapter 2: Literature Review

2.1 Introduction

The problem of vehicle routing is not new. Scholars and academicians have developed a myriad of solutions ranging from traffic flow systems and models to help ease if not eliminate this problem. We will analyze the various solutions to find out the gaps that they have not been able to fill. Finally, we will get to see how the Geometric Brown motion is used to solve the vehicle traffic routing dilemma.

2.2 Factors causing Increase in Traffic Volume

2.2.1 Population Growth

In the modern day world, more than fifty five percent of the world's populace lives in urban areas with an anticipated increase of more than sixty six percent by 2050 ("World's population increasingly urban with more than half living in urban areas | UN DESA | United Nations Department of Economic and Social Affairs," 2014). The UN report iterates that predictions collectively with increase of the world's populace may possibly lead to a further surge of two billion five hundred million persons to urban areas by the year 2050 with a majority of the upsurge focused in the continents of Asia and Africa. The urban populace of the world has grown at a very high rate from seven hundred and forty-six million in the year 1950 to three billion nine hundred million in the year 2014 and by the year 2045; the world's populace will be over six billion.

By the year 2050, Africa will have the highest rate of population growth and will account for more than half of the world's populace with majority of the population centered in urban areas("World population projected to reach 9.7 billion by 2050 | UN DESA | United Nations Department of Economic and Social Affairs," 2015).

Kenya, for example, has had a steady increase in urban population since 1968. The urban population rose from 9.5% in 1968 to 26.5% in 2017. This was a steady growth at an average annual rate of 2.12% ("Traffic Index by City 2018 Mid-Year," n.d.).

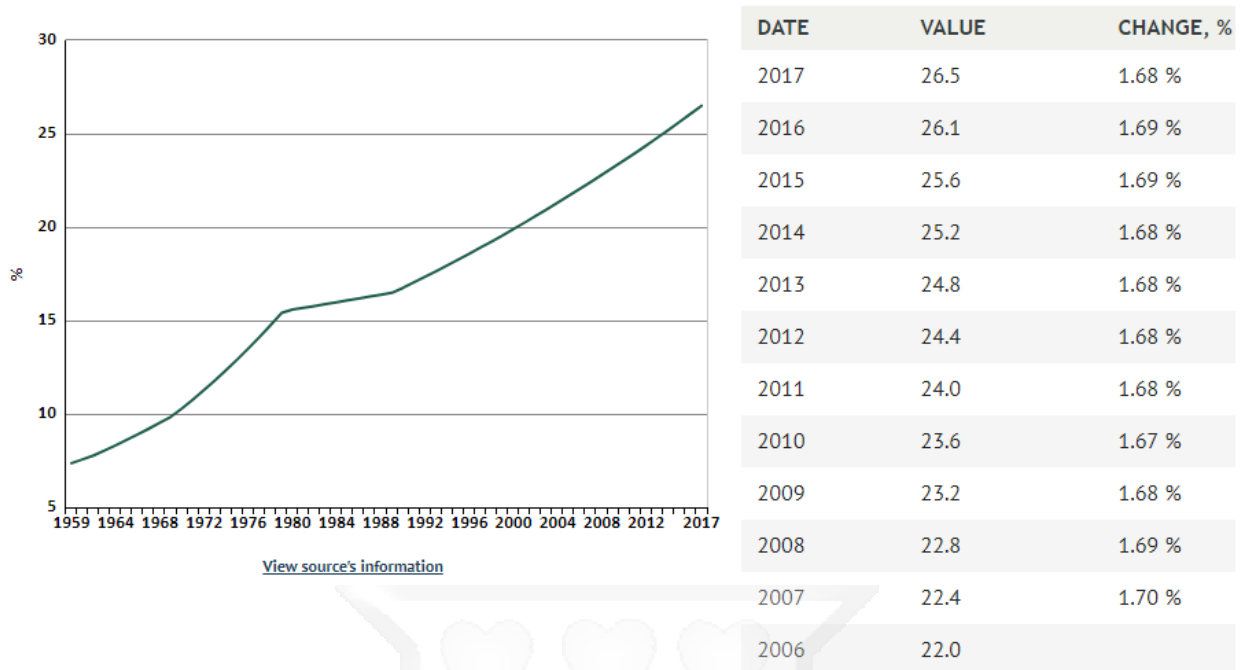


Figure 2-1 : Kenya - Urban population as a share of total population
 (“Urban Development,” 2018)

However, the rapid growth and scale of urbanization comes with challenges. This include meeting accelerated demand for reasonably priced housing, well-connected transport systems, and other infrastructure, basic services, as well as jobs, particularly for the nearly one billion urban poor who live in informal settlements to be near opportunities (“Urban Development,” 2018).

2.2.2 Improved Economic Status

Vehicular congestion is one of the most predominant issues in large urban settings with populations above one million inhabitants. Drivers in the 21st century spend more than 3 times the usual time in traffic as compared to the later era of the 20th century. More people are now able to afford vehicles especially due to improved lifestyles but unfortunately, the transport infrastructure has not grown with the same pace. This has seen highways reduced to temporary parking spaces due to traffic congestion especially during peak hours. The increase in the number of vehicles also means a need for parking space, which is taking a toll on the already strained transport infrastructures.

In the same breath, urban inhabitants are spending a lot of time commuting between the workplace and their residence. This time could have otherwise been used to do something constructive. The public transport system provided is also not reliable as most of them are

either under over used depending on the route. This has discouraged people from investing in the sector.

2.3 Approaches used in Dynamic Vehicle Routing

2.3.1 Greedy Algorithm

The greedy algorithm is the long-established approach to dynamic vehicle routing. The algorithm begins with a routing plan where it allows for multiple initial requests and attempts to inculcate new requests, as they are made available. The initial routing plan is generally obtained by an advanced local search, which is the only viable option for large-scale problems. Assuming a request for a customer arrives, the greedy algorithm dictates whether there is a reasonable point of introduction in to the plan. If none exists, the request is rejected (“Scenario-Based planning for Partially Dynamic Vehicle Routing With Stochastic Customers,” 2003.).

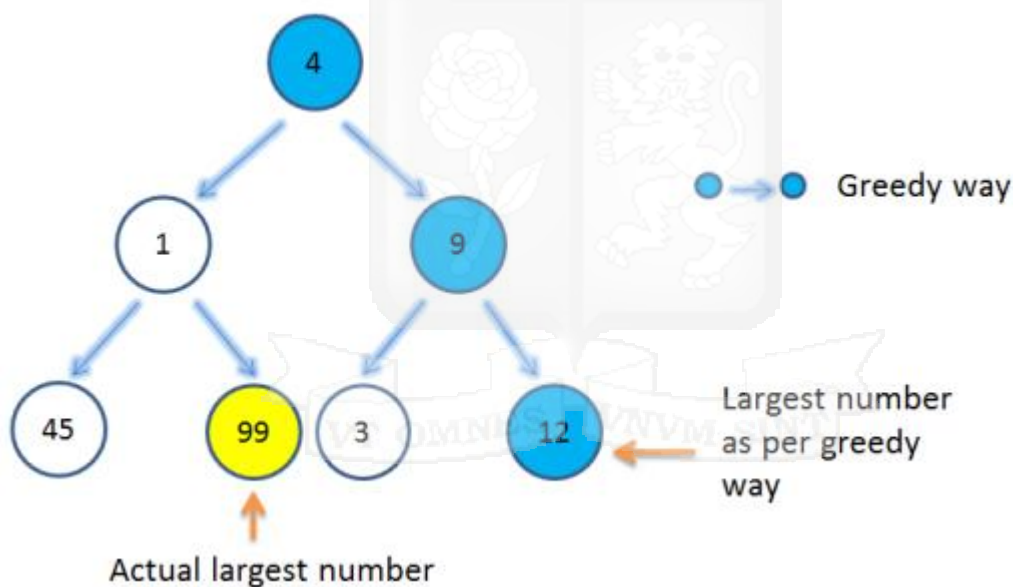


Figure 2-2 : Greedy Algorithm

(“Greedy Algorithm,” n.d.)

Figure 2-2 above is an illustration of how the Greedy algorithm works. It basically chooses the largest number from the starting point in this case 4 and then from 9 it chooses between 3 and 12 and chooses the largest in this case 12 and then stops. However, in the real sense the actual largest number is 99. Hence, the Greedy Algorithm is only suitable for some situations and not others especially the salesperson travel problem. Notably, the Greedy algorithm makes the choice, which appears as the best at that particular moment and

subsequently solves the sub-problems. It iteratively makes one greedy choice after another, reducing each given problem into a smaller one.

2.3.2 The Multiple Plan Approach

2.3.2.1 Parallel Tabu Search for Real-Time Vehicle Routing and Dispatching

The Parallel Tabu Search for Real-Time Vehicle Routing and Dispatching focuses on the dynamic problem inspired by messenger service applications established within the resident processes of international shipping services like DHL, where the dispatched rider is collected at dissimilar consumer localities and transported back to the main office for more processing and shipment. The ambiguity arises from the manifestation of fresh service requests that must be allocated in real time to a suitable vehicle. In the static scenario however, all the requests are known prior to the routes being constructed and each verdict must be taken based on the available requests (Gendreau, Guertin, Potvin, & Taillard, 1999).

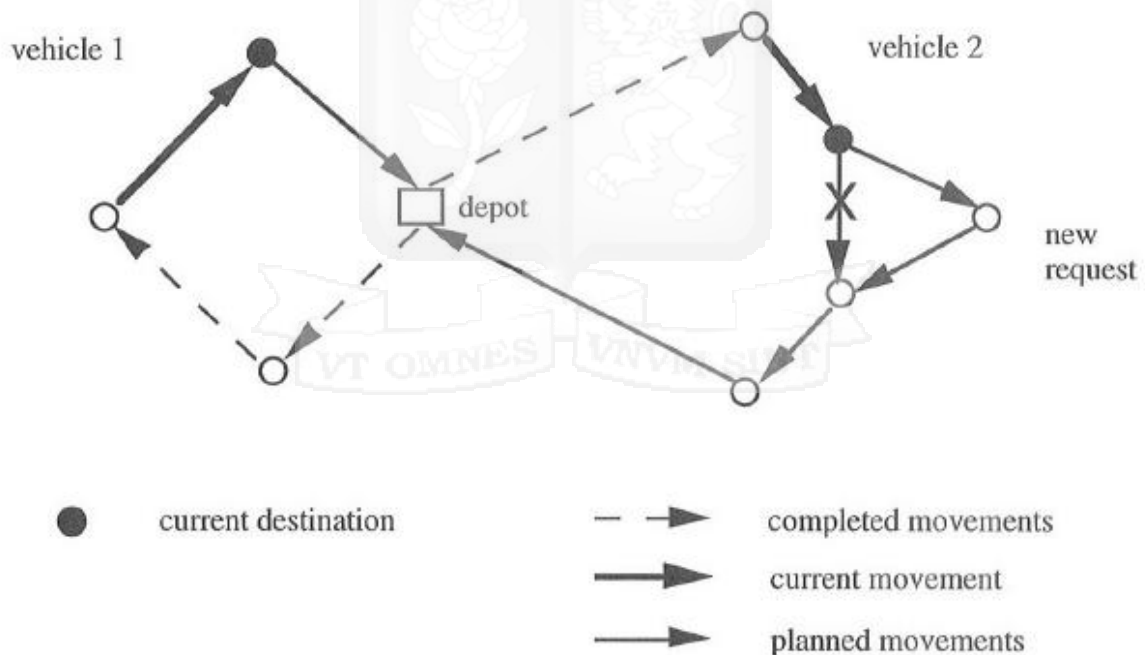


Figure 2-3 : Dispatching situation involving two vehicles

(Gendreau, Guertin, Potvin, & Taillard, 1999)

Figure 2-3, indicates a situation where two vehicles are involved and service requests are received from different customers. From the figure, the square is the depot and the little circles are the customer requests. The dotted lines correspond to completed customer requests hence are not visited again. The black circles are the current destinations of the vehicles. All the

vehicles have to start from a point and follow a path until they have serviced all the customers then circle back to the depot.

2.3.2.2 Static Problem

The static version of the problem as described by (Gendreau, Guertin, Potvin, & Taillard, 1999) is a representative of a class of problems known as the vehicle routing problem with time windows. Assuming you are given a fixed size fleet of similar vehicles, the aim is to pinpoint a set of minimum cost vehicle routes starting from and terminating at a depot such that:

- (i) Each vehicle services one route
- (ii) Each destination is visited exactly once
- (iii) Start time of each vehicle is greater than or equal to the earliest time
- (iv) End time of each vehicle is less than or equal to the latest end time of each route

The objective function f to be minimized over the set of feasible solutions S is

$$f(s) = \sum_{k=1}^m d_k + \sum_{i=1}^m \alpha_i (t_i - l_i)^+, s \in S$$

Equation 2-1 : Static Problem Equation

(Gendreau, Guertin, Potvin, & Taillard, 1999)

Where $y^+ = \max\{0, y\}$, d_k is the total distance traveled on a route k , $k=1, \dots, m$, and α_i is a lateness penalty coefficient associated with vertex v_i $i=1, \dots, n$.

2.3.2.3 Dynamic Problem

The dynamic version of the problem involves a number of service requests that are unknown at all before time, but instead are randomly revealed as time goes by. The dynamic operating scenarios are based on the following assumptions (Gendreau et al., 1999a).

- (i) Requests must be received before a certain fixed deadline so as to be serviced on the same day. Otherwise, those received after will have to wait until the next day.
- (ii) Uncertainty comes from the occurrence of new service requests.
- (iii) Communication between the central office and the drivers is conducted at the service stations where the main agenda is to identify their next stop.
- (iv) If for some reason a delay is expected at the next stop, then the drivers are required to wait at their current stop. This leaves room in case of any last minute changes.

The Tabu search heuristic allows for more customers to be attended to and reduces distance traveled plus delays at customer stops.

2.3.3 Traffic Management Model

Jiang, Zhang, & Ong. (n.d.) Developed a novel digital pheromone-based traffic management model that simultaneously improves vehicle routing and traffic light control. They argued that, every car agent leaves behind numerous digital pheromone when it uses a path. The road is layered with road infrastructure agents, which fuse the digital pheromone to project the current traffic situations on the particular road. The more the deposited pheromone the more the expected or forecasted traffic. When a road congestion is foreseen, a pre-emptive vehicle re-routing algorithm is adopted which then assigns alternate routes to other vehicles prior to them entering the congested road. Concurrently, the traffic light control agents then use an online method to allocate an extended duration of green traffic lights to the routes with very high amounts of pheromone deposits.

The *traffic pheromone* is defined to approximate the current traffic intensity:

$$T1(p, t) = \frac{N(p, t) \times L_{car}}{L_p \times lanes(p)}$$

Equation 2-2 : Traffic Pheromone equation
(Jiang, Zhang, & Ong, n.d.)

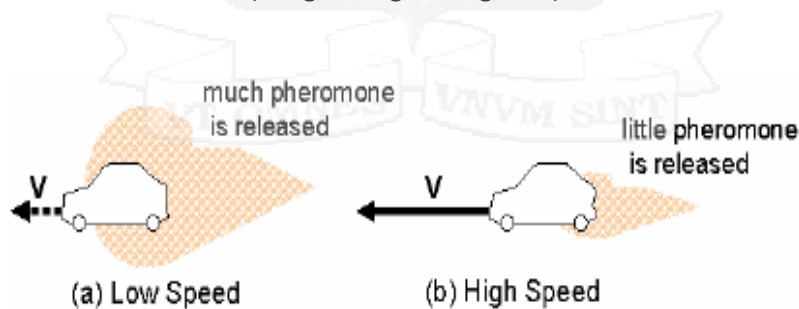


Figure 2-4 : Vehicles depositing basic traffic pheromone

(Jiang, Zhang, & Ong, n.d.)

$N(p, t)$ represents the number of vehicles on road p an time $(t-1, t]$, L_{car} represents the average length of cars, L_p represents the length of road p , and $lanes(p)$ is a representative of the number

of lanes on road p . Supposing that each vehicle makes a report of its expected intended path to the laid out roadside infrastructure (e.g., using smart phones). (Jiang et al., n.d.) Further, define an *intention pheromone* to approximate the future expected traffic density:

$$T2(p, t + 1) = \frac{I(p, t + 1) - O(p, t + 1) \times L_{car}}{L_p \times \text{lanes}(p)}$$

Equation 2-3 : Intention pheromone equation one
(Jiang, Zhang, & Ong, n.d.)

$I(p, t+1)$ and $O(p, t+1)$ represent the inbound and outbound car amounts on road p in time $(t, t+1]$, respectively. Both $I(p, t + 1)$ and $O(p, t + 1)$ take the same formulation. For example, $I(p, t + 1)$ is formulated as:

$$I(p, t + 1) = f(z) = \sum_{p' \in P_{nei}} Tg(p') Vf(p') T1(p', t) \times p \times \frac{\text{lanes}(p')}{L_{car}}$$

Equation 2-4 : Intention pheromone equation two
(Jiang, Zhang, & Ong, n.d.)

Where cars move from neighboring roads $p_- \in P_{nei}$ to road p , $Tg(p_-)$ is the time interval of green traffic lights on road p_- , $Vf(p_-)$ is the free speed of road p_- , and ρ is the proportion of vehicles that have aim of moving from road p_- to road p . The pheromone update includes evaporation and propagation. (Jiang et al., n.d.) Defined the evaporation rate as:

$$e(p, t) = \frac{V(p, t)}{V_f(p)} \times \frac{1}{1 + |\text{Halts}(p, t)|}$$

Equation 2-5 : Evaporation and Propagation equation
(Jiang, Zhang, & Ong, n.d.)

Where $\bar{V}(p, t)$ and $|\text{Halts}(p, t)|$ are the average speed and halting car counts on road p in time $(t - 1, t]$, respectively. Then, the laid roadside infrastructures combine the digital pheromone by:

$$\tau(p, t+1) = (1 - e(p, t)) \times \tau1(p, t) + e(p, t) \times \tau2(p, t+1).$$

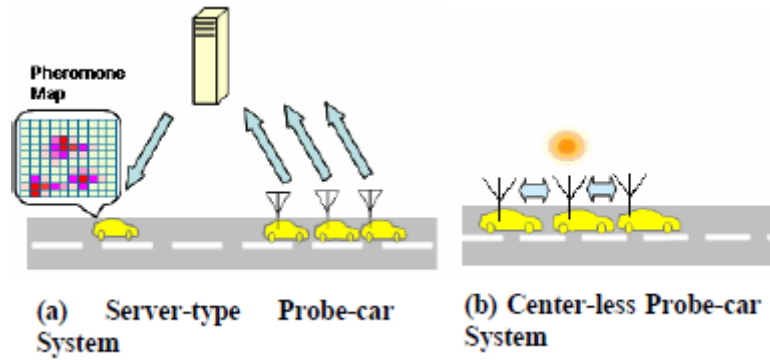


Figure 2-5 : Realization of pheromone system
(Jiang, Zhang, & Ong, n.d.)

2.3.4 Pheromone-based Traffic Light Control

The Traffic light control algorithm was developed by (Jiang, Zhang, & Ong, n.d.) to be used concurrently with the digital Pheromone-based traffic system and was named the pheromone-based traffic light control. The traffic light control automatically sets color stages using an algorithm and then calculates the time duration of these stages. They came up with the following results:

Table 2-1 : Results by Traffic Management Models

	Con. Roads	Traffic Density	Arr. Vehs	Travel Time
Grid Network				
Baseline	23.2	0.709 ± 0.173	339	314.5
TLC	10.0	0.355 ± 0.243	1177	152.8
Re-routing	4.9	0.256 ± 0.221	1800	170.4
DUA	3.2	0.218 ± 0.198	1770	144.7
Re-routing+TLC	0.3	0.102 ± 0.112	1896	94.2
Cityhall Network				
Baseline	21.4	0.340 ± 0.358	397	440.6
TLC	11.5	0.245 ± 0.325	555	434.1
Re-routing	9.5	0.209 ± 0.320	618	446.1
DUA	7.3	0.179 ± 0.296	696	378.8
Re-routing+TLC	3.4	0.099 ± 0.223	705	344.6

* Con. Roads: congested roads, Arr. Vehs: arrived vehicles

(Jiang, Zhang, & Ong, n.d.)

They verified their traffic management model on two road networks (i.e., Grid and City hall). They introduced two classical algorithms: “Baseline” which neither controls re-routing nor traffic lights, and “DUA” (dynamic user assignment). The pheromone-based traffic light control (TLC) was found to report better results as compared to Baseline, and pheromone-based vehicle re-routing algorithm (Re-routing) was found to be better than TLC. To explain, TLC was assigned the default traffic lights of two competing roads so as to find out if they have similar traffic pheromone.

They concluded that TLC is only appropriate for the unbalanced traffic scenarios. Re-routing, on the other hand explores substitute routes for eluding traffic congestion in comparison to TLC. Dynamic User Assignment (DUA), conversely, is far much better than Baseline, TLC and Re-routing. Essentially, DUA is an offline algorithm that leverages on the trial-and-error technique to calculate for potential alternative routes. DUA is therefore only best suited for simulation purposes since a vehicles movement in a real-life transportation environment is not reversible. According to the results in Table 2-1, the typical number of roads, facing congestion on Grid and Cityhall are 0.3 and 3.4, respectively. Consequently, the driver experience is also noted to increase through their model. The number of vehicles that have arrived on Grid and Cityhall are 1,896 and 705, respectively.

In conclusion, the digital pheromone-based traffic management model assimilates the benefits of both vehicle re-routing and traffic light control. The model is therefore better than Baseline, TLC, Rerouting, and DUA.

2.4 Mobile Traffic Applications

2.4.1 Waze

Waze is an application that is helping ease traffic congestion on roads. Waze helps to connect drivers to one another. Drivers share information about the traffic, road condition and if there are any accidents. This information is then used by other drivers to either warn them not to use a particular road or encourage them to use a route because of ease of traffic. The app works when the user types in a destination address and they just drive around with the app open to passively give information on traffic and other road data (Waze, 2018).

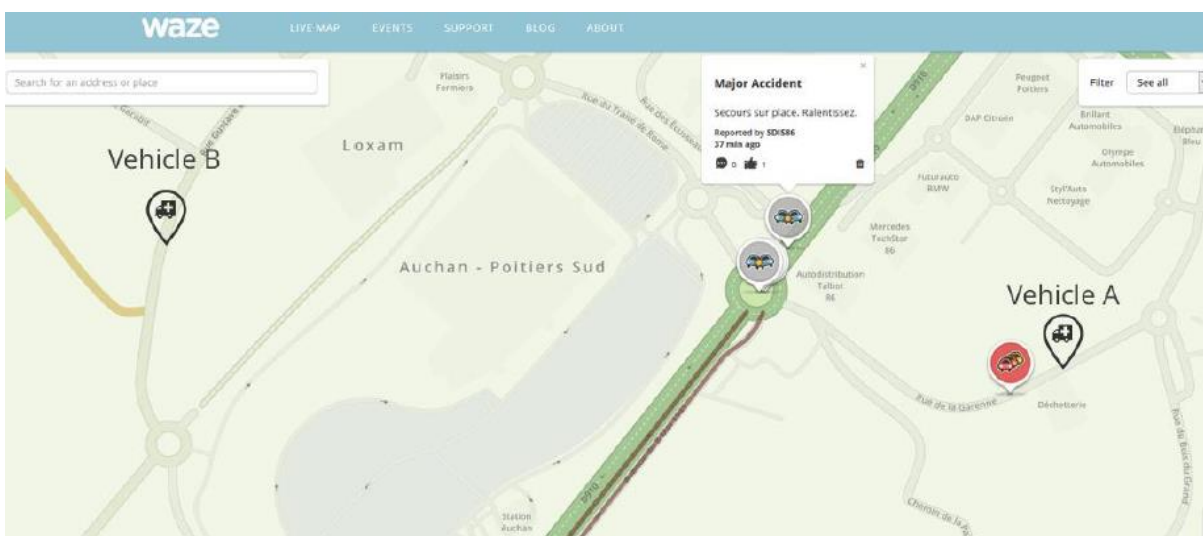


Figure 2-6 : Sample Waze Route planning Use case
(Waze, 2018)

2.4.2 Google Maps

Google Maps is a web mapping service developed by Google. It offers satellite imagery, street maps, 360-degree panoramic views of streets, real-time traffic situations, and route forecasting for traveling by foot, car, bicycle, or public transportation. Google Maps works by collecting data from various sources most of it being crowd-sourced data. None of the information provided by Google maps is over three years old. This is because Google is always constantly updating its data in case of new information or corrections.

To assist in the collection of data, Google has collaborated with a number of authoritative data sources through its Base Map Partner Program. These agencies include USDA Forest Service, the US National Park Service, the US Geological Survey, various city and county councils. To further assist with the data collection, Google has a massive squad of vehicles dispersed around the world with the aim of repeatedly driving around every accessible road they can locate all the while taking 360-degree pictures. The street views offer much more than just photos of the roads. They use their optical character recognition (OCR) which can recognize road signs, traffic signs, and business names. Another method that Google exploits to collect data is the use of its massive army of Google Map users who combined with the Local Guides contribute their own local knowledge to Google Maps. Local Guides take pictures of areas that they visit and answer any questions that may be brought up due to a place they visited.

With the vast database of knowledge, Google Maps is the best application to use in terms of determining the various available routes/roads to use. Though other maps like Bing Maps may exist, google maps is best suited as it is widely known and very detailed. The mapped routes on Google Maps will be used to route traffic in case of accidents, congestion or any other traffic incidents that may occur.

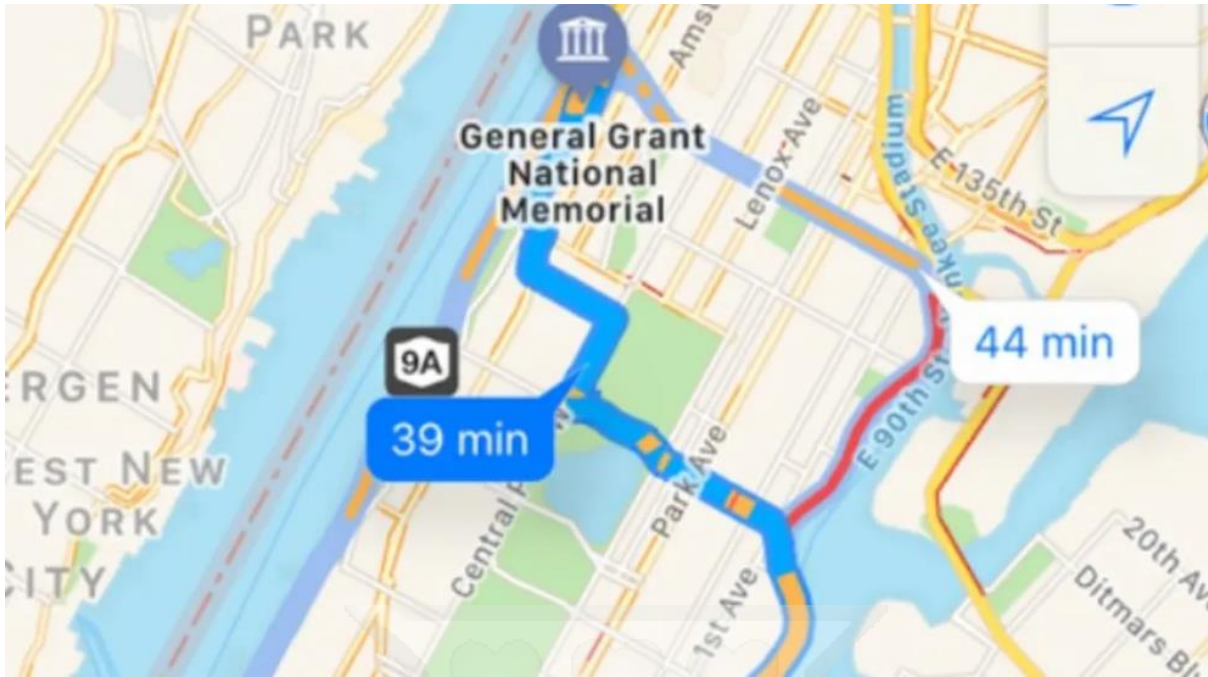


Figure 2-7 : Sample Google Maps Route planning Use case

2.5 Algorithms Applicable to Traffic Routing

Various routing algorithms have been developed with different merits and demerits. Many of the routing algorithms were developed for computer networks and data communications. These algorithms are also applicable in problems related to routing of road traffic. Concepts taken from graph theory can be employed in solving these problems. This is due to the fact that a road network can be represented as a graph with a number of nodes or vertices and a number of paths or edges, as is illustrated in Figure 2-8.

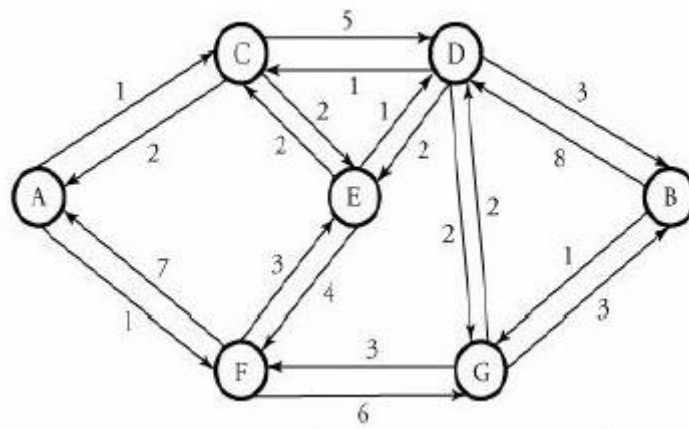


Figure 2-8 : A graph or network with nodes and paths bearing the path costs

(Mir, 2006)

The vertices can be used to represent road junctions and the edges can be used to represent the actual road segments. In the road network, the shortest or fastest route may at times be desired and at times the cheapest route may be desired. With the road network represented as a graph, these parameters can be determined by applying various algorithms related to graph theory.

2.5.1 Complexity

When analysing different algorithms, an important parameter useful for comparison purposes is the Complexity. Complexity theory investigates the interval and retention space an algorithm needs as a function of the extent of the input data (Jungnickel, 2008). Analysis of the time needed gives rise to a measure referred to as *time complexity* whereas analysis of the memory needed gives rise to *space complexity*. This approach is used to compare different algorithms to solve the same problem. The time complexity of an algorithm, A , is the function f , where $f(n)$ is the maximal number of steps A needs to solve a problem instance having input data of length n . The space complexity is defined analogously for the memory needed (Jungnickel, 2008). The complexity is usually measured taking into account the worst-case scenario for the length of the input data. The complexity is commonly measured using the Big-O notation. An explanation of the Big-O notation is given in Appendix A.

2.5.2 Dijkstra's Algorithm

This is a least-cost-path algorithm that was developed by the Dutch programmer Edsger Dijkstra. It finds the shortest path through a weighted graph. Dijkstra's algorithm can be stated as finding the shortest path from a given node to all other nodes by developing the paths in the order of increasing weight. By the k th stage, the shortest path to the k nodes from the source node have been determined. These are placed in the set M . At the $(k+1)$ th stage, the node not in M that has the shortest path from the source is added to M . As the nodes are added to M , their paths from the source are defined. Theoretically, the algorithm finds the shortest path for a single-source to another node for a directed graph with nonnegative path weights (Odhiambo, 2013). With V representing the number of Vertices or nodes and E the number of edges, the complexity of the Dijkstra Algorithm can be given as $O(V^2)$ (Jungnickel, 2008). If a fibonacci heap is employed the algorithm can be implemented with a complexity of $O(|E|+|V|\log|V|)$. The Dijkstra Algorithm can be represented as follows (Mir, 2006):

1. Define:

s = Source node

k = Set of visited nodes by the algorithm

α_{ij} = Cost of the link from node i to node j ($\alpha_{ii} = 0$ and $\alpha_{ij} = \infty$ if i is not connected to j)

β_{ij} = Cost of the least-cost path from node i to node j

2. Initialize:

$k = \{s\}$ $\beta_{sj} = \alpha_{sj}$ for $j \neq s$

3. Next node:

Find $x \in k$ that $\beta_{sx} = \min \beta_{sj}$ for $j \in k$.

Add x to k .

4. Least-cost paths:

$\beta_{sj} = \min(\beta_{sj}, \beta_{sx} + \alpha_{xj})$ for $j \in k$

Steps 2 and 3 are repeated until paths are assigned to all nodes. The algorithm terminates when all nodes have been visited and included in the algorithm.

2.5.2 Bellman-Ford Algorithm

This is also a least-cost-path algorithm. Given a weighted graph G and a source vertex s , the Bellman-Ford Algorithm finds the shortest (minimum cost) path from s to every other vertex in G (Tiang, 2010). Unlike the Dijkstra algorithm, the Bellman-Ford Algorithm is able to operate even when there are negative link weights. However, there must be no cycles of negative weight, a cycle being a closed path comprising a number of distinct vertices (Jungnickel, 2008). The Bellman-Ford Algorithm can be represented as follows (Mir, 2006):

1. Define:

s = Source node α_{ij} = Cost of the link from node i to node j ($\alpha_{ii} = 0$ and $\alpha_{ij} = \infty$ if i is not connected to j) $\beta_{ij}(l)$ = Cost of the least-cost path from i to j with no more than l links

2. Initialize:

$\beta_{sj}(0) = \infty$, for all $j \neq s$ $\beta_{ss}(l) = 0$, for all l

3. Least-cost path:

for any node $j \neq s$ with predecessor node i : $\beta_j(l+1) = \min_i[\beta_i(l) + \alpha_{ij}]$

The algorithm ends when all nodes have been visited and included in the algorithm. The time complexity of the Bellman-Ford Algorithms is $O(|V||E|)$. The Dijkstra Algorithm, therefore, has the advantage of having a lower time complexity (Tiang, 2010).

2.5.3 Floyd-Warshall Algorithm

This is a least-cost-path algorithm that allows for the calculation of costs between all pairs of vertices in a given network. It allows some of the edges to have negative weights but like the Bellman-Ford algorithm it does not allow for negative cycles. The Floyd-Warshall algorithm executes with a time complexity of $O(|V|^3)$, where V is the number of vertices (Jungnickel, 2008). In dynamic traffic routing, one may need to find the cost from a source vertex s , to a destination. Therefore, the feature of this algorithm of finding the distances between all pairs of vertices may not be very useful and applicable.

2.5.4 Algorithms of Prim, Kruskal and Boruvka

These are algorithms that are based on spanning trees. A spanning tree is a connected acyclic graph in which any two vertices of the graph are connected by a unique path. Network optimization helps in the process of finding a spanning tree in which the sum of all the edges is minimal. This procedure can then be utilized in road traffic applications involving the determination of optimal route from source to destination in a given network.

The three algorithms of Prim, Kruskal and Boruvka are specifications of a generic algorithm which is useful in determining a minimal spanning tree for a network (Jungnickel, 2008). The algorithm of Prim is a development of the generic algorithm and achieves a time complexity of $O(|V|^2)$. With regards to Kruskal's algorithm, if priority queues are used a time complexity of $O(|E|\log|E|)$ can be achieved. In the case of the implementation of Boruvka, a complexity of $O(|E|\log|V|)$ is achieved.

2.5.4 K-Shortest Paths (KSP) Algorithms

The algorithms presented in the previous section enable the identification of a path of least cost from the source node to a destination node. In terms of road traffic, this can translate identification of the a route with least traffic from a source to a destination, if a suitable metric for traffic can be obtained. The algorithms presented will generally obtain a single least-cost-path.

In a practical situation, it may be useful to identify a number of least cost paths, which are then presented in an ascending order of their costs. The problem of finding a number of shortest paths in a network is generally referred to as the "K-shortest-paths" problem, since the general goal is to find K shortest paths. In terms of road traffic routing, this could have application in providing a number of alternative routes and ranking them in terms of transport, as opposed to only giving the one best route.

2.5.5 Yen's Algorithm

This is an algorithm developed by Jim Y. Yen in 1971 that finds the K shortest paths from a single source vertex to a destination vertex for a graph with non-negative weights (Yen, 1971). The algorithm makes use of any shortest-path algorithm such as Dijkstra or Bellman-Ford to find the shortest path and then continues to find another K-1 shortest paths in ascending order. This algorithm works on the basis that there are no loops along the paths returned (Yen, 1971). This introduces a certain level of complexity. If the Dijkstra Algorithm is used for the implementation of Yen's Algorithm, the time complexity that is achieved is $O(K|V|(|E|+|V|\log|V|))$.

2.5.6 Eppstein's Algorithm

This algorithm was developed by David Eppstein and published in 1997 (Eppstein, 1997). It also finds the K shortest paths from a source vertex to a destination vertex in a digraph with non negative weights. The time complexity of Eppstein's Algorithm is $O(E + |V|\log|V| + K)$. This can be seen to be less complex than Yen's algorithm. However, this algorithm operates on the basis that there can be loops within the paths returned i.e. the paths are not required to be loopless (Eppstein, 1997).

2.6 Summary

The main methods used to gather and disseminate traffic information in Nairobi are radio stations, CCTV cameras placed alongside a few major roads, and social media applications. Applications that collect data, process it and disseminate it in an integrated manner are non-existent.

In order to conduct vehicle traffic routing, appropriate algorithms have to be utilized. Algorithms resulting from graph theory can be used. These have widely been implemented in computer communication networks. Some algorithms utilized include Dijkstra's Algorithm, Bellman-Ford Algorithm and the Floyd-Warshall Algorithm. Dijkstra's algorithm achieves a favorable time complexity. These Shortest-Path algorithms only give a single shortest path between a given source and a destination. If a number of best paths are required, K-shortest Path algorithms (KSP) can be used such as Yen's Algorithm and Eppstein's Algorithm. Notably, in as much as the mentioned algorithms work, they do not give information on the vehicle volume on the routes. This information can be used by motorists to make more informed decisions on which routes to use based on the traffic density.

2.7 Brownian Motion

Brownian motion is the distinctive (in distribution) process described as follows:

$\{B_t, t \geq 0\}$

It has the following properties:

- (i) No memory. If $0 \leq t_0 \leq t_1 \leq \dots$ then $B_{t_1} - B_{t_0}$, $B_{t_2} - B_{t_1}$, $B_{t_3} - B_{t_2}$ are autonomous
- (ii) Invariance. For $s, t \geq 0$, the distribution of $B_{s+t} - B_s$ depends only on t .
- (iii) Continuity. With probability 1, $t \rightarrow B_t$ is continuous.
- (iv) Normalization. With Probability 1, $B_0 = 0$. For all $t \geq 0$, $E B_t = 0$ and $E B_t^2 = t$



Figure 2-9 : Brownian Movement

As illustrated in figure 2-10 above, the Brownian motion movement is characterized by random unpredictable fluctuations. From the figure, you can see that the values keep changing either increasing or decreasing at every point. This random movement is known as the drift or the random step (Karatzas, & Shreve, 1998).

2.7.1 Why Brownian Motion

Brownian motion belongs to several families of well understood stochastic processes:

- (i) Markov processes
- (ii) Martingales
- (iii) Gaussian processes
- (iv) Lévy processes

2.7.1.1 Markov Process

A simple version of the Markov property says that for $s \geq 0$, the conditional distribution of $\{B_t, t \geq s\}$ given B_s is the same as the conditional distribution of $\{B_t, t \geq s\}$ given $\{B_u, 0 \leq u \leq s\}$.

Markov processes are considered as the processes with “No memory”. They are natural models for multiple natural phenomena, societal phenomena and synthetic systems.

On the mathematical side, the theory of Markov processes uses tools from numerous divisions of analysis (Burdzy, 2014):

- (i) Functional analysis (transition semigroups)
- (ii) Potential theory (harmonic functions, Green functions)

- (iii) Spectral theory (Eigen function expansion)
- (iv) Partial differential equations (heat equation)

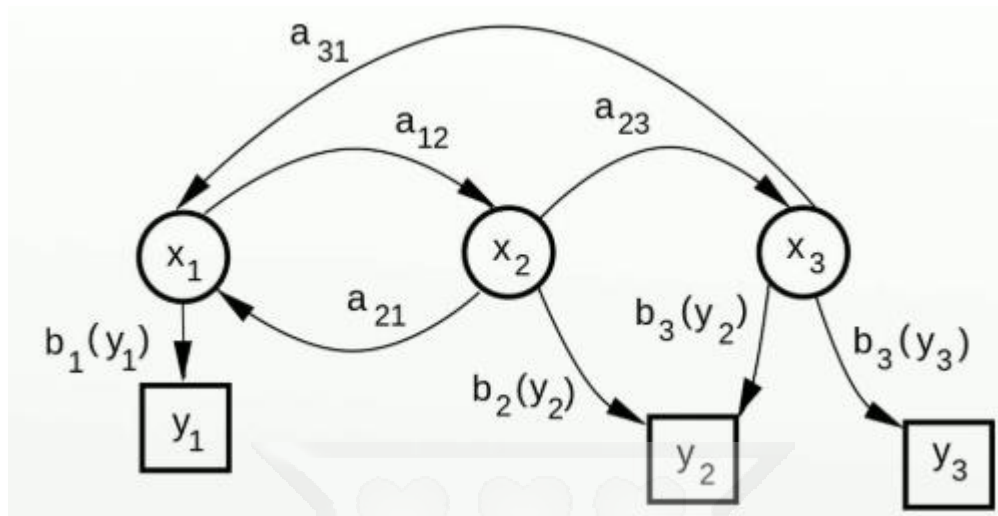


Figure 2-10 : Markov Process
(Burdzy, 2014)

A Markov process is generally a random process where the future is independent of the past given the present. Figure 2-11 illustrates a Markov chain, which is a stochastic model that describes a sequence of probable events in which the probability of each event (X_1, X_2 or X_3) depends on the state attained in the previous event.

2.7.1.2 Martingales

A simple version of the martingale property says that if 0 is less than or equal to s and s is Greater than or equal to t then, a.s.,

$$E(B_t | \mathcal{B}_u, 0 \leq u \leq s) = B_s.$$

Martingales are processes with “no drift.” The Martingales represents very few natural phenomena. Martingales are a synthetic and very successful mathematical construct.

Semi-martingales Y are the only family of processes for which the theory of stochastic integrals of the form $\int_0^t X_s dY_s$ is fully developed, successful and satisfactory(Burdzy, 2014).

2.7.1.3 Gaussian Process

If $0 \leq t_0 \leq t_1 \leq \dots$ then for any n , the $(B_{t_0}, B_{t_1}, B_{t_2}, \dots, B_{t_n})$ has a multidimensional normal (Gaussian) distribution.

The following are characteristics of Gaussian Processes (Burdzy, 2014):

- (i) Gaussian processes are excellent bounds for tails of Gaussian distributions.
- (ii) They are fully characterized by the mean and covariance structure.
- (iii) Gaussian processes can be easily defined on non-ordered (non-time-like) parameter spaces, unlike Markov processes and martingales.
- (iv) Gaussian processes arise naturally in limit theorems and are widely used in modeling of noise.

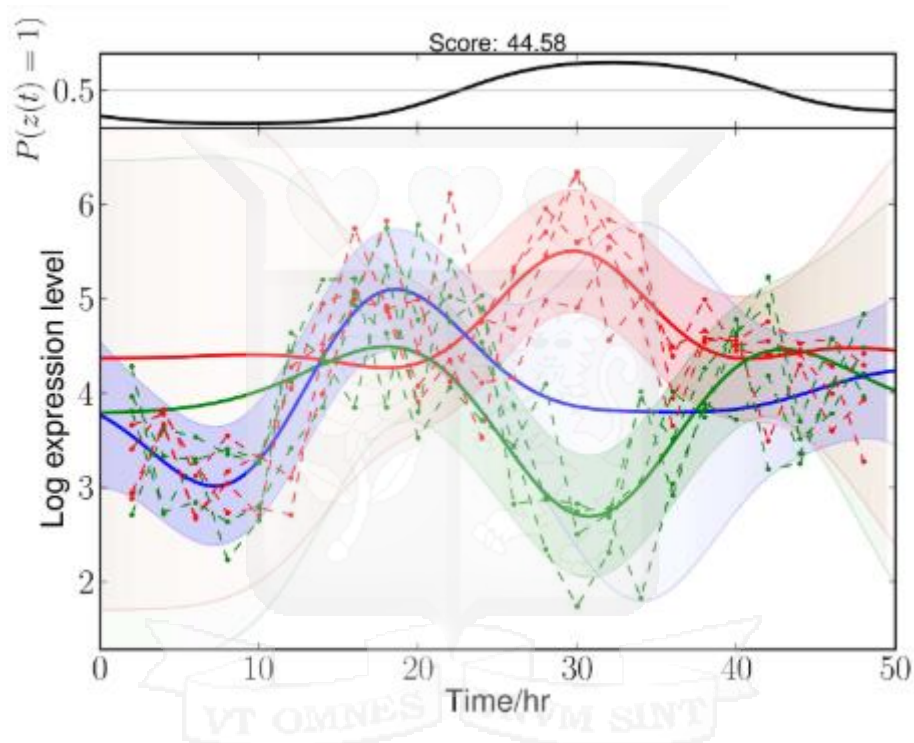


Figure 2-11 : Gaussian Process two sample set (Burdzy, 2014)

In this research, Geometric Brownian Motion (GMB) will be used to solve the problem of traffic routing. Geometric Brownian motion (GMB) falls in the family of the Martingale processes. This is because processes associated with a stochastic integral is always a Martingale.

2.8 Stochastic Traffic Volume Prediction

Transport plans and projects necessitate a massive amount of time, capital and investment to run. Due to these restrictions and many more, which include high risk and complexity, the private sector is called upon to take part in the manner of a Public-Private Partnership (PPP) scheme. Precise forecasting of traffic volume is a key factor in these projects. Traditional traffic volume forecasting methods are often centered on deterministic models,

which forecast only one value of traffic volume but do not take into consideration factors such as risk and uncertainty. To help address this issue, they applied the Geometric Brownian Motion method and the Monte-carlo simulation method to come up with a probabilistic traffic volume-forecasting model.

In the year 1905, Albert Einstein reiterated that Brownian motion atoms must satisfy the partial differential equation:

$$\frac{\partial P}{\partial T} = D \frac{\partial^2 P}{\partial x^2}$$

Equation 2-6 : Diffusion equation

Where

D is a physical constant and

p is the dissemination of the particles over space and time

The diffusion equation has two key qualities. Initially, the adjustment in the estimation of the variable pursues a normal distribution with the zero mean and the variance changing in time. Secondly, the adjustments in the estimation of the variable for any two non-overlapping intervals of time are autonomous. The standard Brownian motion process has a float rate of zero and a fluctuation of one demonstrating that the expected value of the variable at any time is equivalent to the present value and variance of the difference in the variable in a time interval of T is equivalent to T .

In this research, the researcher applied the probabilistic traffic-volume forecasting model to a data set of traffic volume in New York city area.

The variations of traffic volume in this research are assumed to follow a Geometric Brownian motion (GBM) process. That is, traffic volume incorporates a movement density plus a random step, which can be depicted as follows:

$$dq_t = \mu_q q_t dt + \sigma q_t dz$$

$$q_t = q_0 e^{\left(\mu_q - \frac{\sigma^2}{2}\right)t + \sigma \varepsilon \sqrt{t}}$$

Equation 2-7 : Geometric Brownian Equation

Where

dqt = differential increment of traffic

μq = growth rate of traffic volume

qt = traffic volume

dt = a differential period of time

σ = traffic volatility

dz = increment of a Wiener process

Δt = time step (in years)

$\varepsilon \sim N(0,1)$ = standardized distribution with the mean of 0 and the standard deviation of 1

The model indicates the relative adjustment in traffic volume over period of dt , which consists of two parts:

- (i) **Movement.** A anticipated fragment is $\mu q dt$. It is the anticipated traffic volume generated at the stage of dt . The positive value of this movement represents an increase in traffic volume, whereas the negative value indicates a decrease.
- (ii) **Variation.** An arbitrary and unpredictable fragment is $\sigma q dz$. This fragment echoes the arbitrary adjustment in traffic during dt . This arbitrary fragment is defined as a Brown motion process. The distinction represents the level of risk; a high disparity corresponds to the high level of risk.

2.9 Conceptual Framework

The proposed conceptual framework for a Geometric Brownian Motion model is as illustrated below. Vehicle volume data was retrieved from the Kaggle data sets platform and was used as a database to test the model. Firstly, a user makes a request through the web application on the best route to use. The web application then queries the database system on the number of vehicle volume on the specified routes. The Brownian motion model then calculates the expected vehicle volume of the specified route based on the historical data stored in the database. Further, once the vehicle volume value is generated, it is stored in temporary file from where the web application is able to retrieve it and display to the user. The user can then make a decision on the best route to use based on the expected vehicle volume on the different routes.

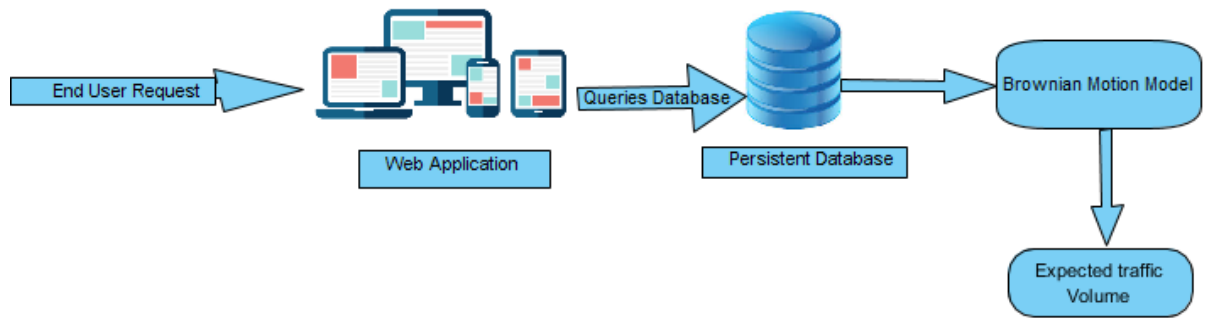


Figure 2-12 : Conceptual Framework



Chapter 3: Research Methodology

3.1 Introduction

A research method is a systematic plan for conducting research. This chapter outlines the methodology used in carrying out this research in consideration of the research objectives outlined in chapter one, the nature of this study as well as methodologies used in past-related works. The research design, which guides decision variables to be considered, data collection methods as well as data analysis techniques. Additionally, the research design highlights the target population of the study and the sample population used for the study. The algorithm development tools and modules will be outlined.

3.2 Research Design

The research aimed at developing a model that can assist commuters better navigate the city and its environs. The research was correlational in nature as it intended to show the correlation between the amount of traffic volume on a route and the traffic congestion. Basically, the developed model employs traffic volume data gathered over a period of time to estimate the expected traffic volume at a future date.

The research model used for the study was quantitative as the historical traffic volume data retrieved from Kaggle Data sets is characterized by numbers thereby quantified to present the output. Creswell and Clark (2007) affirms that quantitative research is based on the measurement of quantities.

3.3 Location of the Study

The data for the study was retrieved from Kaggle data sets. This is because of a shortage of availability of ready traffic volume data especially in Kenya. Conducting a traffic volume count of the different routes would have been a costly affair and would require various permits that would take long to process given the short time available to conduct the research. A New York City data set retrieved from Kaggle data set was used. Notably, the data set contained hourly traffic volume for different streets which was an advantage to the study. Lastly, the data set from Kaggle was freely available hence provided for a cheaper option in order for the necessary simulations to be conducted for the research.

3.4 Population and Sampling

3.4.1 Population

The target population for the study was the 2011- 2013 NYC traffic volume counts dataset. Notably, the researcher could not use a majority of the data as it did not have a large amount of data for specific streets. The data retrieved was only for a few days in the year for all the different streets. No single street in the data set was large enough to meet the required data of 365 days for better results. The dataset used was 24 rows of the 3395 rows that were available for the 2011-2012 dataset. Arguably, this presents about 0.7% of the total population. However, the rows used were very accurate in modelling the expected traffic volume. Notably, there is no readily available documented data in Kenya that relates traffic volume data though there are organizations that collect such data.

3.4.2 Sampling

The research focused on purposive or judgmental sampling when choosing the sample size of the data. This helped the researcher to choose from the population a sample that corresponds to objectiveness of the study. Arguably, this also ensured that the required sample population was chosen in a cost effective and timely manner. Notably, the sampling method was preferred due to limited primary and secondary sources of data required for the study.

3.5 Data Collection

The algorithm used in this study operates on structured historical traffic data. As primary data for analysis, vehicle volume data was retrieved from kaggle data sets and used to inspire the algorithm design as well as carrying out functionality testing. Among traffic parameters retrieved, vehicle volume values will be of use in this research because they conform to the rules of Brownian Motion. Brownian Motion simulations are used to model process that cannot easily be predicted due to the involvement of random variables (Kenton, 2019).

3.6 Model Development

The model development process involved a variety of steps which were targeted towards coming up with a suitable model. The steps can be defined as shown below:

- (i) Data retrieval and analysis.
- (ii) Development of the model.

- (iii) Validation of the model.

3.7 System Development Methodology

3.7.1 Agile System Development

Agile Development is a collection of programming-centric approaches that center on restructuring the SDLC. A project stresses on simple, iterative application development where every iteration is a comprehensive software project, including planning, requirements discovery, system design, system development, system testing and validation, and finally the documentation (Alan Dennis, 2012).

Agile Development was best suited for this project as the method allows for multiple iterations throughout the development life cycle. The model underwent four cycles as shown:

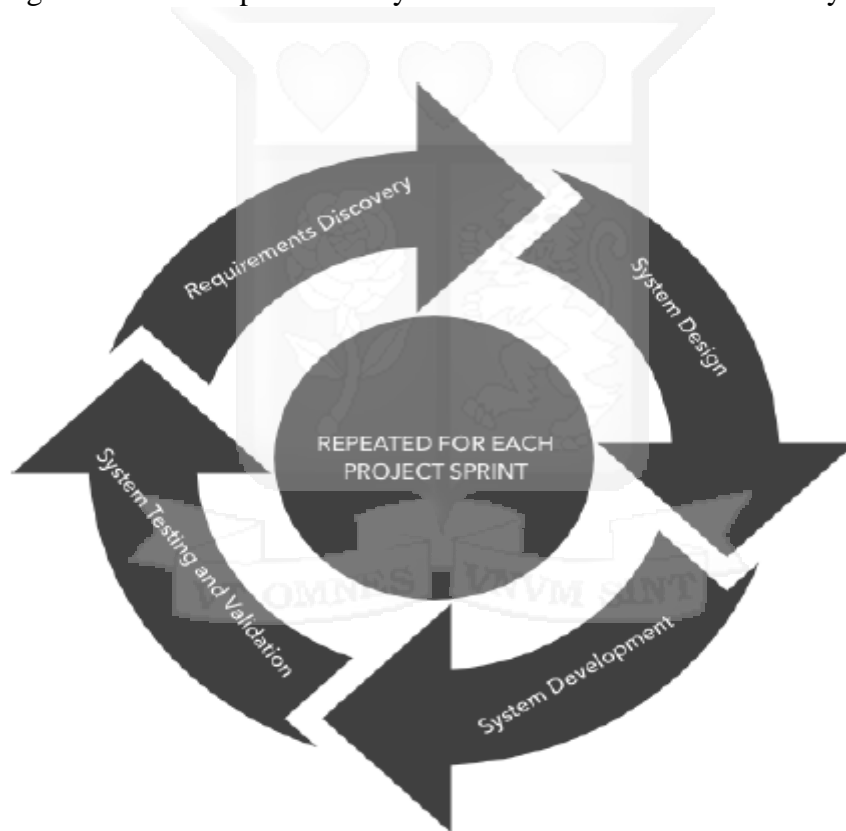


Figure 3-1 : Agile Programming Process
(Alan Dennis, 2012)

3.7.2 Data Retrieval and Analysis

A Data set comprising of Traffic volume data for New York City streets was retrieved from the Kaggle data sets website. The data set was analyzed and formatted so as to determine the quality of the data as well as to determine usability. The data comprised of traffic volume data for different streets but for only a few days over a period of one year. The best data would have

been to have traffic volume data for specific routes over a period of one year. The relevant data sampled was then be used to model the system.

3.7.3 Development of the model

The model was developed and simulated using Monte Carlo simulation technic. The Monte Carlo simulation technic is used to model variables that cannot be easily modelled due to the intervention of random variables (Kenton, 2019). Traffic volume conforms to the Monte Carlo model as it cannot be easily predicted and the involvement of random variables leads to fluctuations of the same. The periodic data rates when modelled formed a normal distribution with a Bell-shaped graph.

The model was then integrated with Open Street maps so as to give users information on the expected traffic volume of routes that they are using.

3.7.4 Validation of the model

The generated model was tested against the validation dataset. Essentially, the metric used to test the model was the Monte Carlo simulation technic. The metric served as the best indicator of the performance of the model on the provided task. The model outputs the expected traffic volume on the different routes at a future time.

3.8 System Design

The system architecture comprises of software components. The software components of the system include the Brownian motion model used to predict expected traffic volume. Consequently, the Brownian motion model was then integrated with Open Street Maps. Essentially, if a user is using the application to navigate, the application will alert the user on the expected traffic volume of the selected routes. Additionally, the system implemented a functional database to store Traffic volume data.

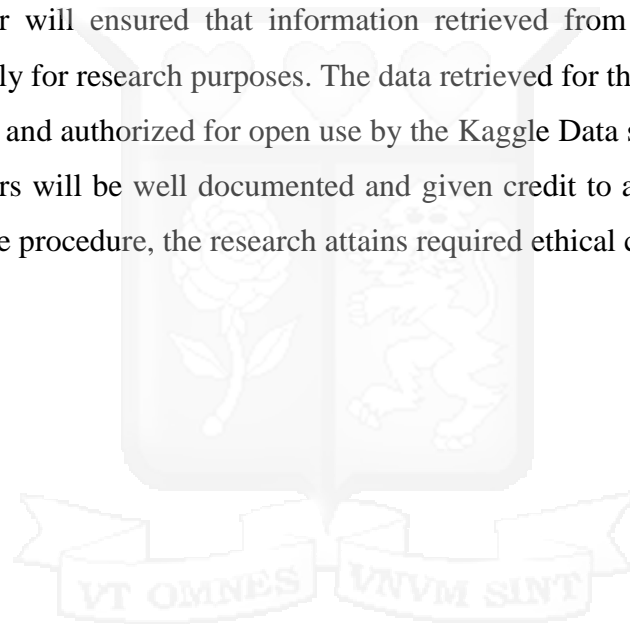
3.6 Research Quality, Reliability and Validity

Quality research refers to the scientific process incorporating all traits of a study design; in particular, it touches to the judgment regarding the match between the methods and questions, selection of subjects, measurement of outcomes, and protection against systematic bias, non-systematic bias, and inferential error (Boaz & Ashby, 2003; Lohr, 2004; Shavelson & Towne, 2002)

According to (Habib, Pathik, & Maryam, 2014), Reliability is an indicator of a measure's internal consistency. A measure is reliable when different attempts at measuring something converge on the same result while validity refers to the accuracy of a measure or the extent to which a score truthfully represents a concept. This research followed the required quality standards by ensuring proper procedure was used in data retrieval, accurate citation of resources works from other researchers and the outcomes achieved presented without influence. The retrieved data was a representative of the study and the use of counteractive mechanisms applied to enhance the quality of research. The model was tested against standard metric to ascertain its performance.

3.7 Ethical Considerations

The researcher will ensure that information retrieved from the Kaggle data sets website was used solely for research purposes. The data retrieved for the purpose of this study is publically available and authorized for open use by the Kaggle Data sets. Works from other authors and researchers will be well documented and given credit to avoid plagiarism. With the adherence to above procedure, the research attains required ethical considerations.



Chapter 4: System Analysis and Design

4.1 Introduction

This chapter describes the overall architecture and detailed design of the proposed prototype 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 users and various components of the system. To achieve this, various design diagrams were developed including: a depiction of the system architecture, use case diagram followed up with comprehensive use case descriptions, sequence diagrams, context diagrams and data flow diagrams.

4.2 Requirement Analysis

This research aimed at developing a model to aid motorists in navigating through traffic. Based on this objective, this section outlines the various requirements to be provided for by the proposed solution.

4.2.1 Functional Requirements

- (i) The application should allow a user to enter a current location and a destination.
- (ii) The application should calculate the expected number of vehicles on a route based on historical vehicle volume data of the particular route.
- (iii) The application should display to the user the expected number of vehicles on the particular route.

4.2.2 Non-Functional Requirements

4.2.2.1 Supportability Requirements

The proposed model should be able to be supported by a large amount of data to accurately predict the expected traffic volumes. Essentially, this allows the model to be utilized in the real world for any route.

4.2.2.2 Reliability Requirements

The system should provide the expected traffic volume information based on the probability with the greatest odds of occurring. In essence, this ensures a predicted result which conforms to the laws of Brownian motion and Monte Carlo simulation.

4.2.2.3 Security Requirements

The system uses traffic volume data and stores application user information. As such, the system needs to implement strong security measures to ensure its integrity and protect privacy. Furthermore, the alteration of the system configurations require authorized personnel to increase the security of the system.

4.2.2.4 Persistent Storage

The system should provide permanent storage for vehicle traffic volume both historical and predicted. Such data may be used to inform relevant government bodies on road expansion projects and usage.

4.3 System Design and Architecture

4.3.1 System Model Architecture

The system utilizes traffic volume of data as input. The traffic volume data is then passed through the Brownian motion model for expected traffic volume calculation. Essentially, to achieve high levels of accuracy expected volume calculations, the Brownian motion model requires a large data set of the specified route as shown in the figure below:

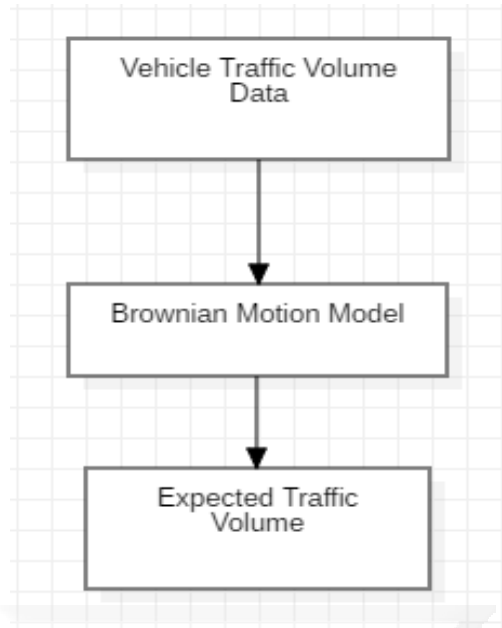


Figure 4-1 : System Model

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 hate speech detection prototype. The diagram also depicts the functionality that the proposed system should have.

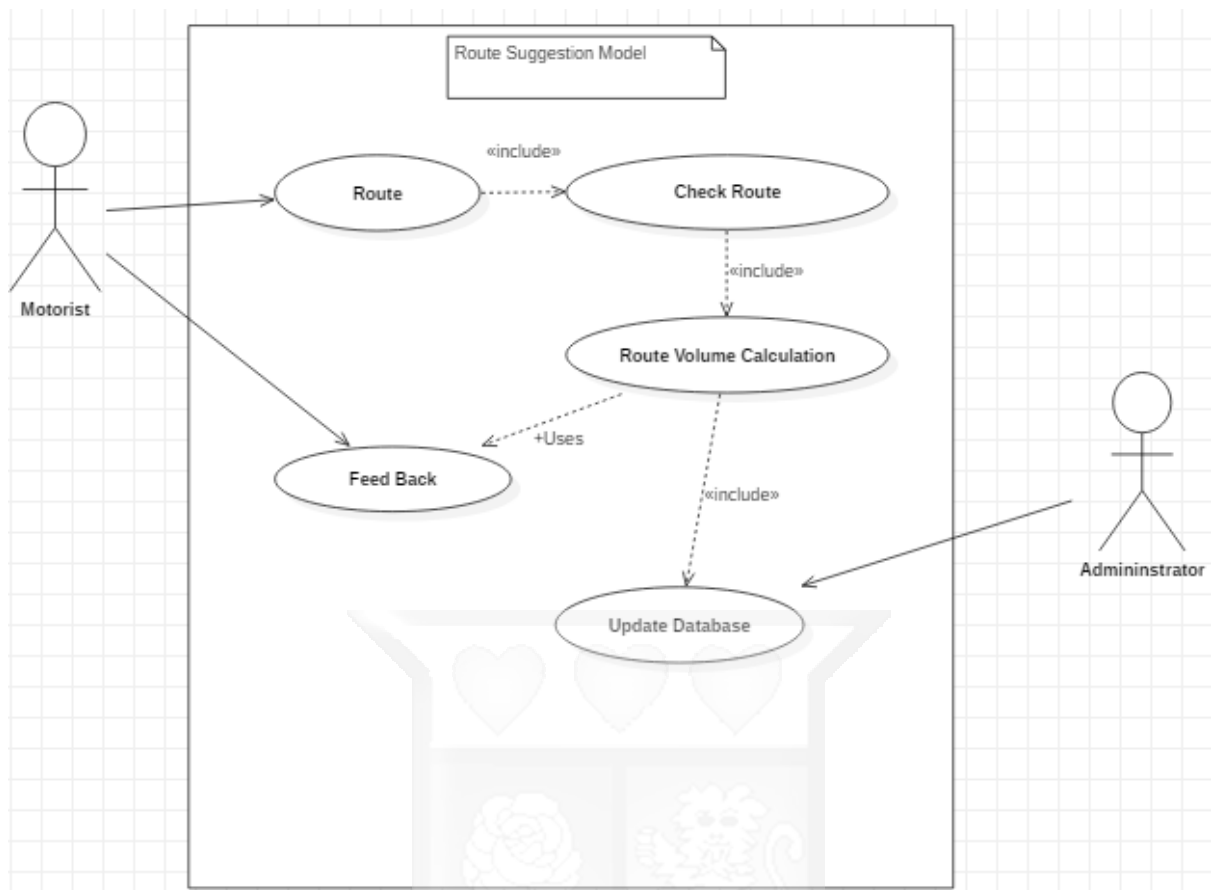


Figure 4-2 : Proposed Algorithm Use Case

This diagram illustrates the list of actions or use cases that the algorithm fulfils.

The use cases identified are:

Route – the user selects their current location and the destination that they wish to be travel to

Check route –the user initiates a search on the selected input.

Route Volume Calculation – the algorithm calculates traffic volume of the route from values extracted from the data set.

Update Database – The system Administrator updates Database with new volume values.

Feedback – System displays predicted volume of traffic to the user

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. Furthermore, it captures the storage of data from the processes. In essence, it basically shows the flow of data in the system serving in a better understanding of the system. In this system, the diagram will outline entities such as the users of the system.

4.3.3.1 Context Diagram

The context diagram as depicted in Figure 4-3 illustrates the boundary of the prototype, its environment and the entities that interact with it. It also shows the various inputs and outputs from the prototype to the entities. The main entities interacting with the proposed prototype are a user and the system administrator. The user enters their current location and their preferred destination. The model then calculates the expected vehicle traffic volume at that particular time. The model then provides an estimated value of traffic volume. The Administrator has to frequently update the model with vehicle traffic volume so as to improve accuracy.

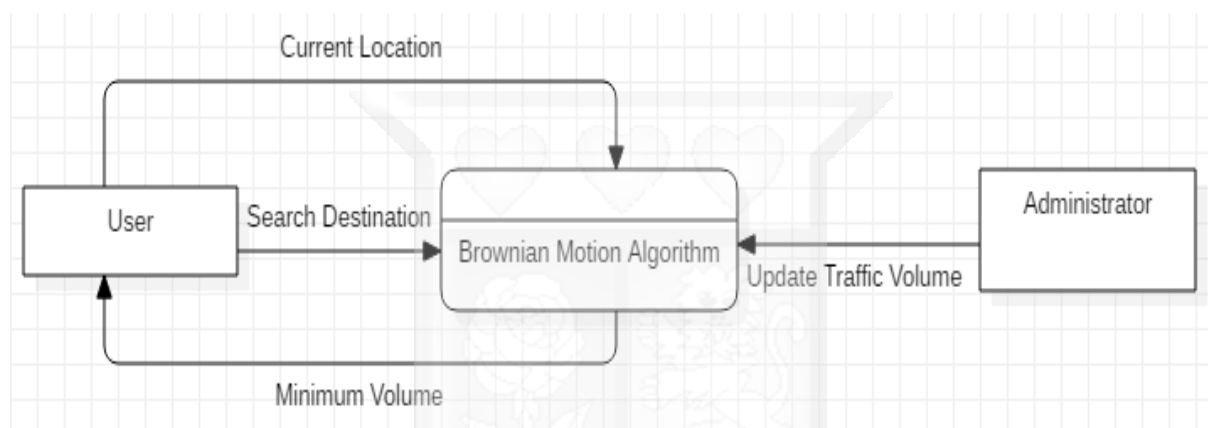


Figure 4-3 : Context Diagram

4.3.3.2 Level 0 Data Flow Diagram

The level 0 Data Flow Diagram (DFD) depicted in Figure 4-4 below gives a more detailed view of the prototype 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 called Navigation App receives a search request from the user. The Navigation upon successful validation of the user credentials queries the database which contains historical Traffic volume data. The Brownian Algorithm then calculates the estimated Traffic at that particular time and outputs a value based on the chosen route. The Navigation App displays the results to the user. Give Feedback process receives a feedback request from the user, reads the calculated value and outputs the expected vehicle volume to the user.

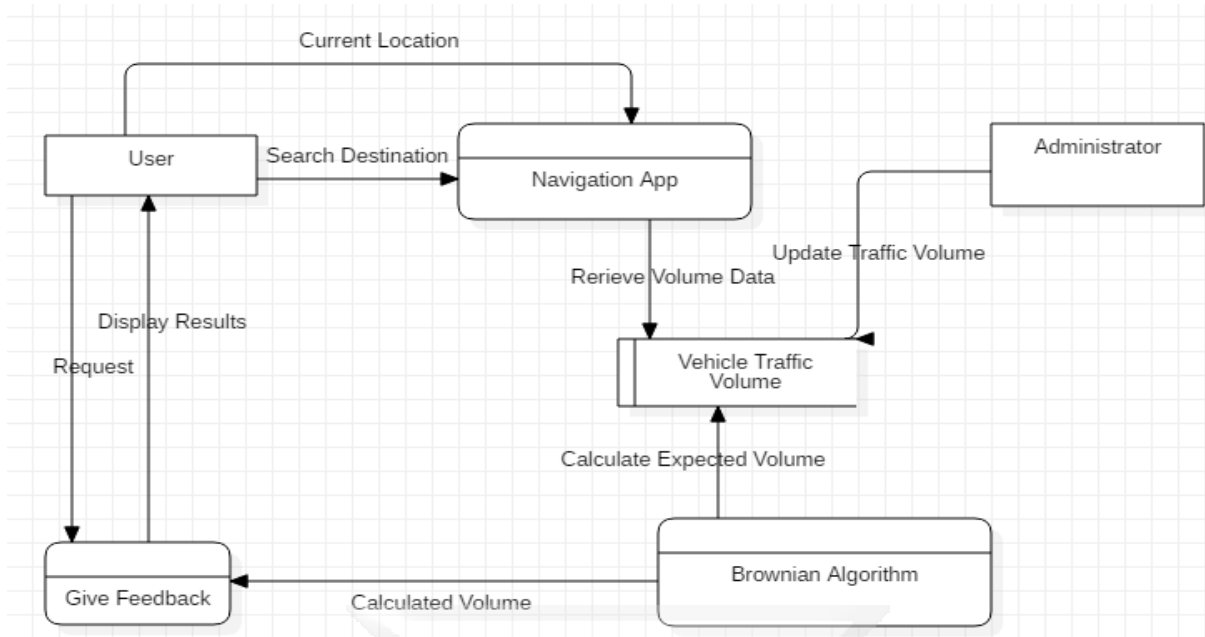


Figure 4-4 : Level 0 Diagram

4.3.3.3 Flow Chart

A flow chart diagram, figure 4-5, describes the sequence of operations the algorithm undertakes from start to completion.

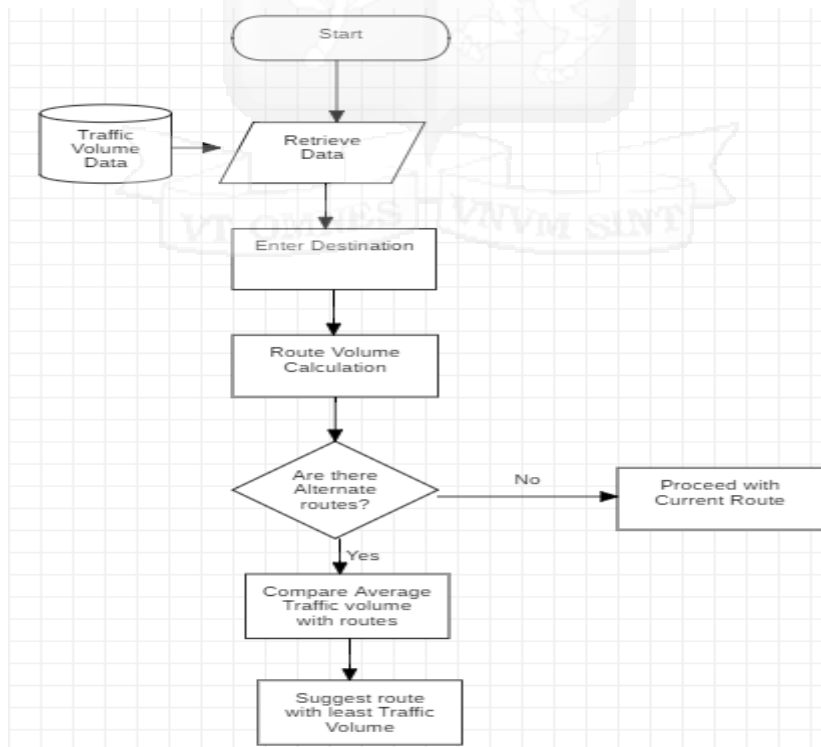


Figure 4-5 : Flow Chart

4.3.4 Sequence Diagram

The sequence diagram depicted in Figure 4-6 shows the sequence of interactions between the user and the proposed system as well as interactions between the various internal components of the system. The user enters their current location and their desired destination they wish to travel to through the Application. Once the current location and destination are obtained, they are passed on to the Application database where the calculation takes place. The algorithm calculates the traffic volume for the selected routes by the user and then displays the results to the user.

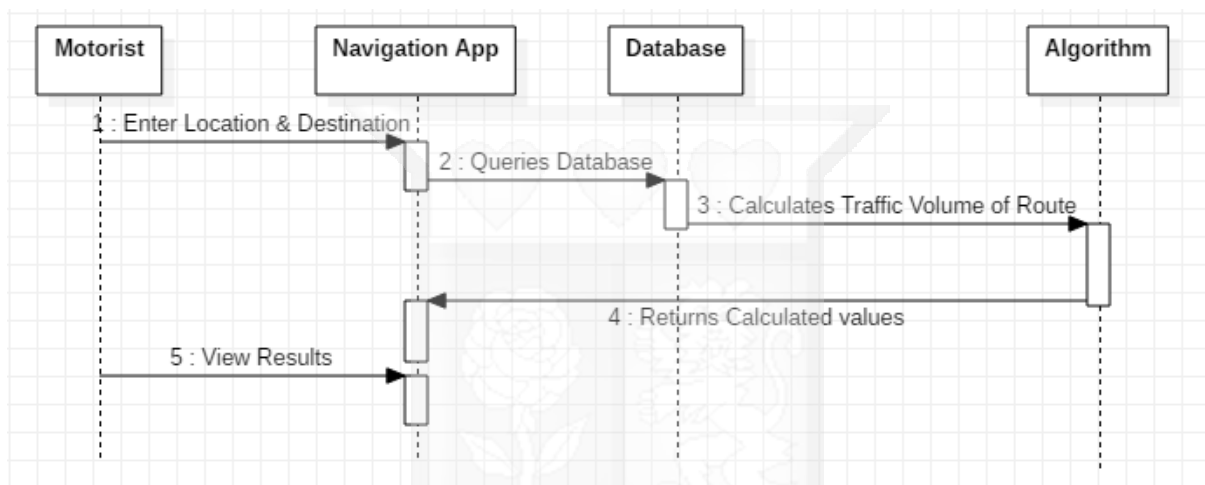


Figure 4-6 : Sequence Diagram

4.3.5 Partial Domain Model

The real-world objects which are represented using a domain model for the system is illustrated by the Figure 4-7 below. The diagram comprises conceptual classes, their associations between each other and their attributes. The end user, Road/Route, Brownian model, and the Web application forms the partial domain model classes of the system. The classes interact at different points.

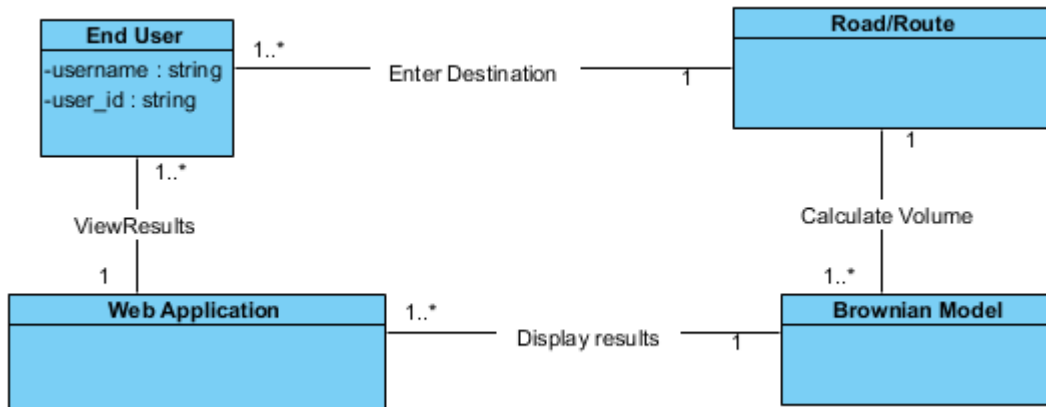


Figure 4-7 : Partial Domain Model

4.3.6 Database Schema

Figure 4-8 below shows the database that the system uses. The schema has been represented using an Entity Relationship Diagram (ERD). The schema shows the different entities in the system and their relations.

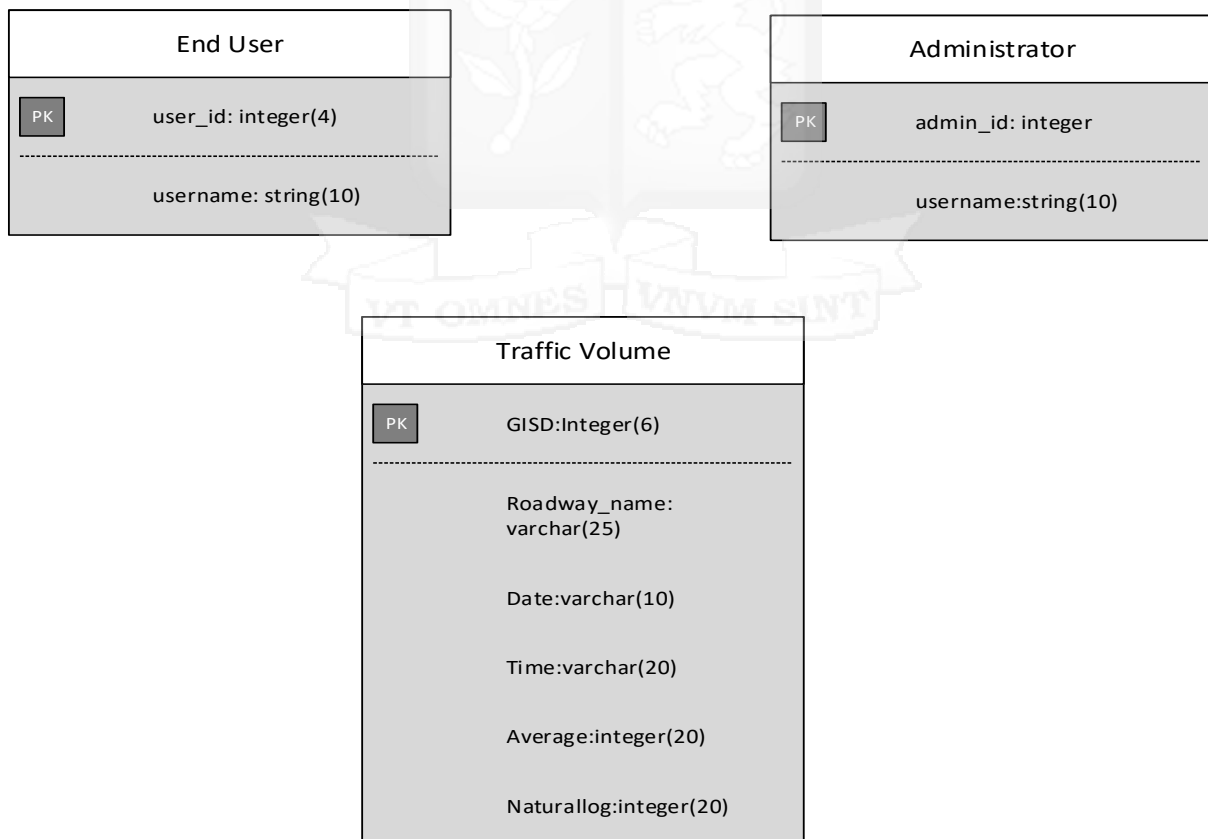


Figure 4-8 : Database schema

Chapter 5: System Implementation and Testing

5.1 Introduction

This chapter describes how the model was implemented, tested and validated. It begins by describing the process of coming up with the Brownian Algorithm. 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 feature types. The final section of this chapter describes the use of the model in predicting expected vehicle volume.

5.2 Hardware and Software Environment

The development of the model was undertaken on Jupyter notebook using Pandas. Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, 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 different specifications used for the model:

Table 5-1 :Hardware and Software Environment

Software	Specified Library	Version
Python (3.6)	Pandas	1.13.1
	Numpy	1.16.2
	Statistics	4.0.0
	Math	2.2.4
Hardware	Details	
	RAM	8GB
	CPU	Core i3
	HDD	500GB

Pandas Libray: Pandas as a software library was used data manipulation and analysis. In particular, it offered data structures and operations for manipulating the numerical tables containing the traffic volume data.

Numpy Library: NumPy is a library that helped to add support for a large collection of high-level mathematical functions to operate on help capture the various aspects of Brownian.

Statistics Library: This module provided functions for calculating mathematical statistics of numeric (Real-valued) data.

Maths Library: This module provided access to the mathematical functions defined by the C standard.

5.3 Dataset Pre-processing

The dataset comprised of Traffic volume data retrieved from Kaggle datasets. The dataset contained vehicle volume for different streets in the city of New York between 2011 and 2013. The retrieved dataset was capped into sections of one hour each covering the twenty four hours of the day. Such that, for each hour of the day they noted the number of vehicles using a particular route. Hence the data was spread over a period of twenty hours.

For the study, the sample used was of two streets for comparison purposes. Reason being that the dataset did not have sample data for a single street for the whole year. Hence, two streets were chosen as the sample data. A sample of the dataset is shown below:

GIS ID	Roadway	From	To	Direction	Date	12:00-1:00	1:00-2:00	2:00-3:00	3:00-4:00	4:00-5:00	5:00-6:00	6:00-7:00	7:00-8:00	8:00-9:00	9:00-10:00	10:00-11:00	11:00-12:00	12:00-1:00
15540	BEACH ST	UNION PL	VAN DUZE NB		01-09-12	20	10	11	14	13	20	34	66	100	52	68	85	85
15540	BEACH ST	UNION PL	VAN DUZE NB		01-10-12	21	16	8	6	13	13	31	70	67	45	57	67	73
15540	BEACH ST	UNION PL	VAN DUZE NB		01-11-12	27	14	6	5	12	16	34	75	69	71	67	70	90
15540	BEACH ST	UNION PL	VAN DUZE NB		01-12-12	22	7	7	8	11	12	33	75	89	66	70	60	105
15540	BEACH ST	UNION PL	VAN DUZE NB		1/13/2012	31	17	7	5	13	28	29	68	84	64	83	89	88
15540	BEACH ST	UNION PL	VAN DUZE NB		1/14/2012	42	27	21	18	21	13	17	18	46	53	29	0	
15540	BEACH ST	UNION PL	VAN DUZE SB		01-09-12	27	12	12	4	22	27	66	154	155	138	105	124	140
15540	BEACH ST	UNION PL	VAN DUZE SB		01-10-12	26	16	11	13	16	27	59	156	177	131	107	108	122
15540	BEACH ST	UNION PL	VAN DUZE SB		01-11-12	32	16	8	9	15	26	63	169	178	148	139	131	126
15540	BEACH ST	UNION PL	VAN DUZE SB		01-12-12	24	12	7	18	11	23	61	146	177	128	117	111	134
15540	BEACH ST	UNION PL	VAN DUZE SB		1/13/2012	39	22	8	6	16	30	77	147	187	155	156	159	155
15540	BEACH ST	UNION PL	VAN DUZE SB		1/14/2012	45	45	28	23	21	19	32	56	67	109	74	0	
12809	LITTLE CLC	NORTHERI	BRISTOL A EB		01-09-12	21	11	2	3	11	29	117	511	328	156	129	144	159
12809	LITTLE CLC	NORTHERI	BRISTOL A EB		01-10-12	20	14	5	2	13	24	104	477	322	148	124	142	157
12809	LITTLE CLC	NORTHERI	BRISTOL A EB		01-11-12	16	11	3	6	12	24	118	519	348	165	143	150	156
12809	LITTLE CLC	NORTHERI	BRISTOL A EB		01-12-12	26	14	7	1	6	27	128	483	390	195	144	136	150
12809	LITTLE CLC	NORTHERI	BRISTOL A EB		1/13/2012	29	11	10	2	9	30	119	401	301	182	106	144	161
12809	LITTLE CLC	NORTHERI	BRISTOL A EB		1/14/2012	32	31	17	20	9	14	23	50	70	105	127	97	
12809	LITTLE CLC	NORTHERI	BRISTOL A WB		01-09-12	18	7	7	4	3	23	28	179	177	136	126	204	173
12809	LITTLE CLC	NORTHERI	BRISTOL A WB		01-10-12	21	10	9	3	2	18	33	179	183	138	141	177	163
12809	LITTLE CLC	NORTHERI	BRISTOL A WB		01-11-12	19	12	6	9	5	18	33	185	173	146	139	160	158
12809	LITTLE CLC	NORTHERI	BRISTOL A WB		01-12-12	32	9	4	3	3	11	41	187	187	108	126	158	130

Figure 5-1 : Sample Dataset

The sample dataset in figure 5.1 was used to test the validity of the Brownian motion formula. For best results a larger dataset for a single street would ideally give the best results. But for this case due to insufficient data we opted to work with two streets.

The following steps were taken to compute the predicted value:

- (i) An average starting value (volume) of time zero in this case between 12:00-1:00 was taken for the two streets to give an initial value (S_0).
- (ii) Daily average volumes for the different days were calculated.
- (iii) An assumption was made that the number of days in a year are 365.
- (iv) The Natural logarithm of the Daily averages was calculated and then an average of the calculated Natural logarithms is calculated to give a single value which is now the Daily drift (μ). Since we are calculating traffic volume annually we annualize the daily drift by multiplying by 365 days.
- (v) The Daily volatility (σ) is calculated next by calculating the Standard deviation of the sample already calculated for Natural Logs then multiplying it by the square root of 365 to annualize it.
- (vi) The size of the step (t) equals current time divided by the number of hours in a day (24) multiplied by the number of days in a year (365).
- (vii) The value being predicted (S_t) is the current volume of traffic at a particular time which is calculated using the formula $S_t = S_0 * \text{Exponential}(\mu - 1/2\sigma^2)t + \sigma\text{sqrt}(t)$

Having calculated all the necessary variables a sample test run was done using the sample dataset in figure 5.1. The results are as shown below in figure 5.2:

```

step = x/(365*24)
step_root = math.sqrt(step)

final_value = roads_mean * math.exp(((annual_drift - volatility_halved)*step) + (annual_volatility * step_root))
print(final_value)
#print(annual_volatility)
#print(annual_drift)

```

Enter the streetBEACH STREET
Enter the time7
39.053201455565834

Figure 5-2 : Test Run

The Brownian model was able to predict that at 'Beach Street' at time seven (7) in this case (6:00-7:00AM) the number of vehicles expected on that route are Thirty nine. The average

number of vehicles plying the 'Beach Street' route at time seven are forty four for the indicated days. The model was able to calculate the expected volume of traffic on the route.

5.4 System Implementation

5.4.1 Storing System Details

The system works by providing Vehicle volume data to motorists and other road users. Therefore, to ensure that a proper database is maintained for the data, a Sqlite3 database was created to easily manage the dataset process.

5.4.2 System Testing

The system aims at providing information on the volume of traffic on the roads at particular times of the day. With the use of the Brownian motion formula using historical volume traffic data, the model calculates the expected average volume of a particular route and outputs to the user. Notably, for the testing of the system, a sample dataset of two routes is used as the input to the system for the generation of expected traffic volumes at a particular time. Essentially, this serves as an estimation of the real time traffic volume. The code below is a breakdown of the Brownian motion formula written in Pandas:

```
t = 0.002283105
mu = -12.85192017
sigma = 11.4425957
S0 = 29.66666667
W = np.cumsum(sigma)*np.sqrt(t) ### standard brownian motion ###
X = (mu-0.5*sigma**2)*t + sigma*np.sqrt(t)
S = S0*np.exp(X) ### geometric brownian motion ###
print (S)
```

t= Size of the Step

mu= Annual Drift

Sigma= Annual Volatility

S0= Initial starting Value

5.4.3 Brownian Motion Algorithm

The Brownian motion calculation is carried out from the database. When the user enters a route, the web application fetches vehicle volume data of the specified route from a temporary json file. The value from the temporary json file is retrieved from the overall calculation carried out by Brownian algorithm using the enter parameters. In order for the calculation to take place the user has to enter a street name and the time. The street name is the route the user intends to use and the time coincides with time of day the user intends to travel or navigate the keyed-in route. The model then calculates the expected vehicle traffic volume as per the time of day as the route to be used. The figure below is a breakdown of the Brownian motion algorithm as written in Pandas:

```
In [4]: import numpy as np
import json
import pandas as pd
import math
import statistics as stat
import matplotlib.pyplot as plt
def Average(lst):
    return sum(lst)/len(lst)
df = pd.read_csv('Traffic_Counts_2011-2012_1.csv')
street_name = input("Enter the street")
x= int(input("Enter the time"))
roads = df.loc[df['Roadway Name']==street_name]
roads_mean = roads['1'].mean(axis=0)
mu = df[['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19','20','21','22','23','24']].mean(a
divided = []
added=[]
for i in range(len(mu)):
    num = len(mu)
    if i == (num-1):
        break
    else:
        div = mu[i+1] / mu[i]
        divided.append(math.log(div))
num_days = 365
daily_drift = Average(divided)
daily_votality = stat.stdev(divided)
annual_votality = daily_votality * math.sqrt(num_days)
annual_drift = daily_drift * num_days
votality_halved= (annual_votality * annual_votality)/2

step = x/(365*24)
step_root = math.sqrt(step)

final_value = roads_mean * math.exp(((annual_drift - votality_halved)*step) + (annual_votality * step_root))
print(final_value)
#print(annual_votality)
#print(annual_drift)

Enter the streetBEACH STREET
Enter the time9
40.15305099433056
```

Figure 5-3 : Brownian Algorithm in Pandas

Essentially, the algorithm reads CSV file containing the vehicle traffic volume. The algorithm then prompts the user to enter the Street Name and the time. The algorithm then calculates expected vehicle traffic volume for the specified route using historical traffic volume data for that route. The calculated value is then stored in a json file where it is retrieved by the Web application and displayed to the user.

5.4.4 Using Developed Model

The Brownian motion Algorithm is integrated with a web Application which runs on the Django framework and Map box which is an Open Street Maps mapping tool, the result is as shown below. The web application requests the user to key-in a starting point and a destination. Mapbox then indicates the best navigation route to use based on the entered starting point and destination. Further, the web application indicates an approximation of the expected vehicle traffic volume as per the calculation of the Brownian Motion Algorithm. The users view is as shown in figure 5-4, below:

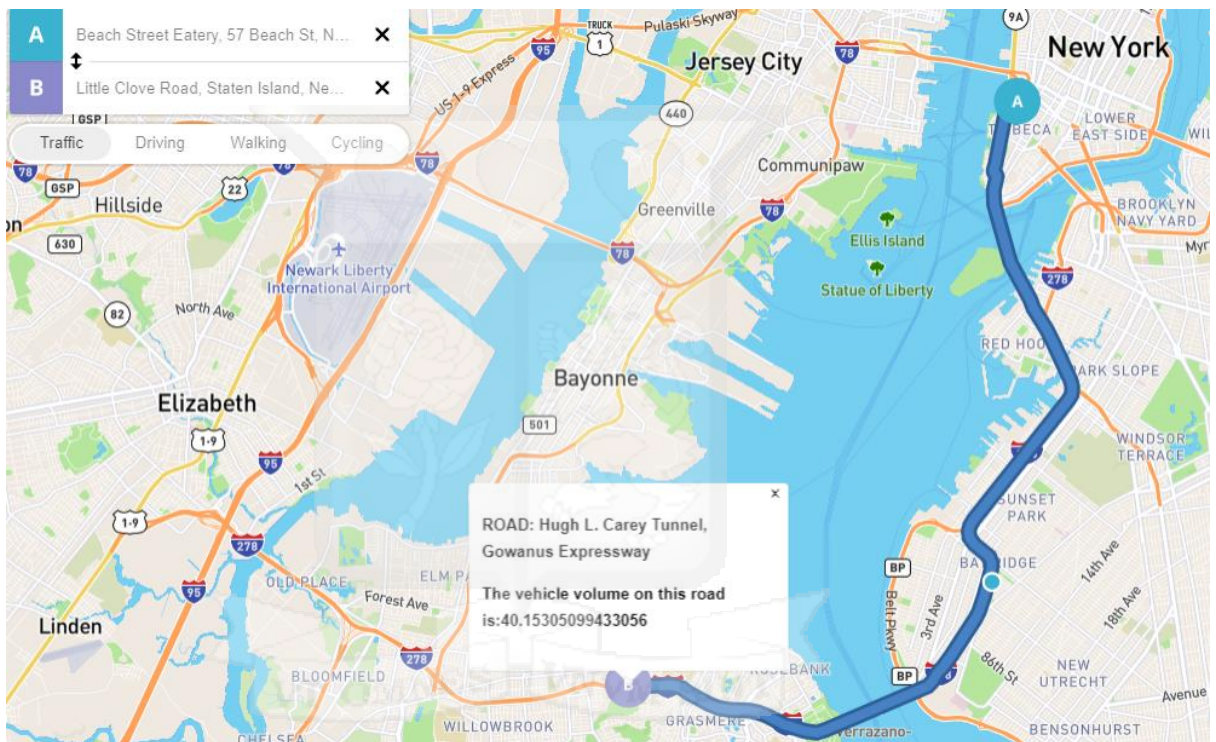


Figure 5-4 : Mapbox

Chapter 6: Discussions

This chapter discusses the results of the research in light of the objectives set out at the beginning. The objectives of this research were to develop a Route suggestion model using Geometric Brownian Motion (GMB) and to and validate it. A route suggestion model was created using Jupyter Notebook and Pandas. Its performance was tested against the test dataset.

6.1 Model Evaluation

Model evaluation entails the process of assessing the performance of a specified model against a specified metric. Essentially, this enables researchers to distinguish between the different specified model performances. The Brownian motion model was evaluated based on its accuracy to predict the traffic volume at a particular time given an initial starting value.

6.1.1 Monte Carlo Simulation

(Kenton, 2019) describes the Monte Carlo simulations as models used to model the probability of different outcomes in processes that cannot be easily predicted due to the intervention of different random variables. Monte Carlo simulations follows the equation below:

$$\ln\left(\frac{S_t}{S_0} - 1\right) = \alpha + Zt\sigma$$

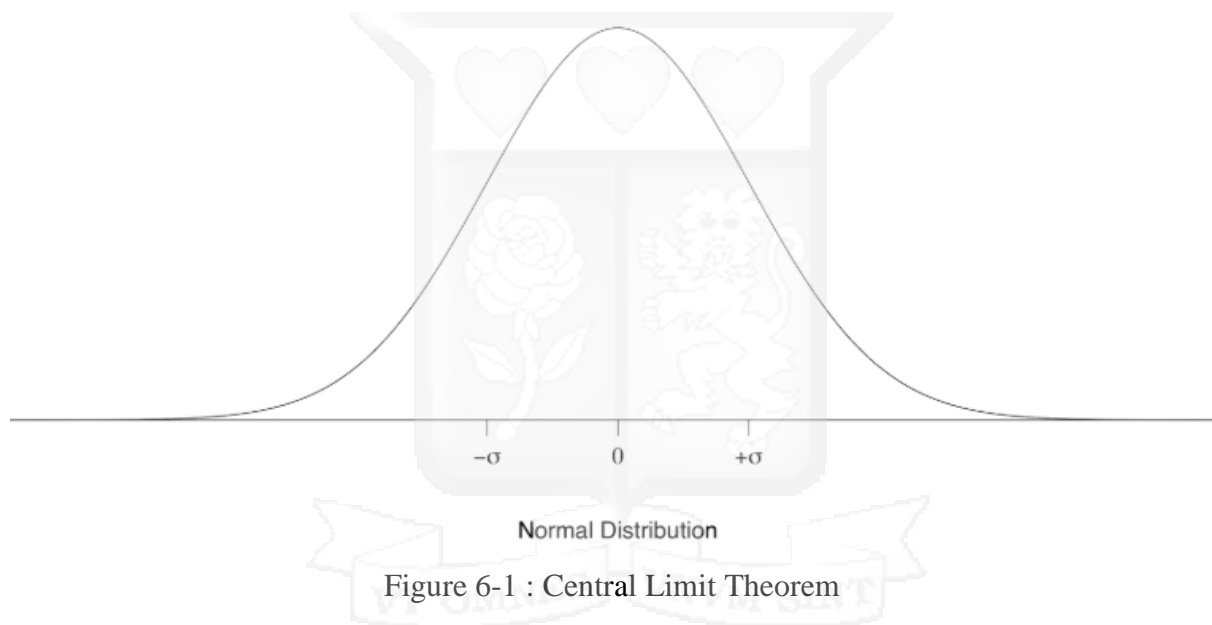
Equation 6-1 : Monte Carlo Simulation

The α represents the drift. This is the expected rate of change each day. In our case it is the expected rate of change for each hour of the day. These drift means that it is the rate with the greatest odds of occurring. For a standard a Monte Carlo, a volatility eroded historical mean of the periodic daily return is used. In other words, the historical traffic volume over a period of time such as one year is taken hourly. The periodic daily return is calculated for each and an average of the values is retrieved. However, the periodic daily return requires to be eroded based on volatility. The standard rate for return erosion is at the rate of half of the variance over time. These is basically the rate with the greatest odds of occurring which is $\mu - \sigma^2/2$ and represents the drift of traffic volume.

Additionally, the odds to be used can be determined using the expected rate of change (drift) and the historical volatility (Standard deviation). This is because for every hour of the day the traffic volume can increase or decrease at any random rate.

6.1.2 Central limit Theorem

Mishra (2018) describes the central limit theorem (CLT) as a statistical theory that “states that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population”. What this means is that if we graph enough of the periodic daily rates of return they will normally distributed forming a bell-shaped graph.



From the graph, at point 0 equals $\mu - \sigma^2/2$ where μ is equals to the average of the historical periodic daily returns.

6.1.3 Brownian Motion

Brownian motion means that if the future rates of returns are graphed, the graph will form a normal bell shaped curve using the drift as the mean and the historical standard deviation as the future standard deviation.

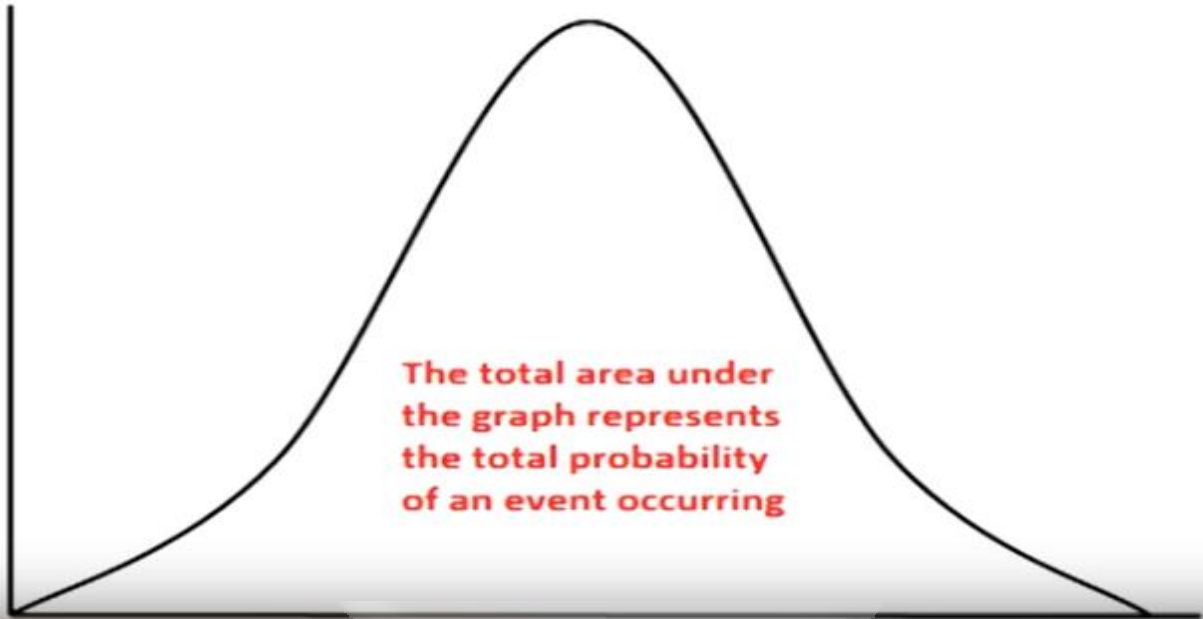


Figure 6-2 : Probability one

("Chegg Study ", n.d.)

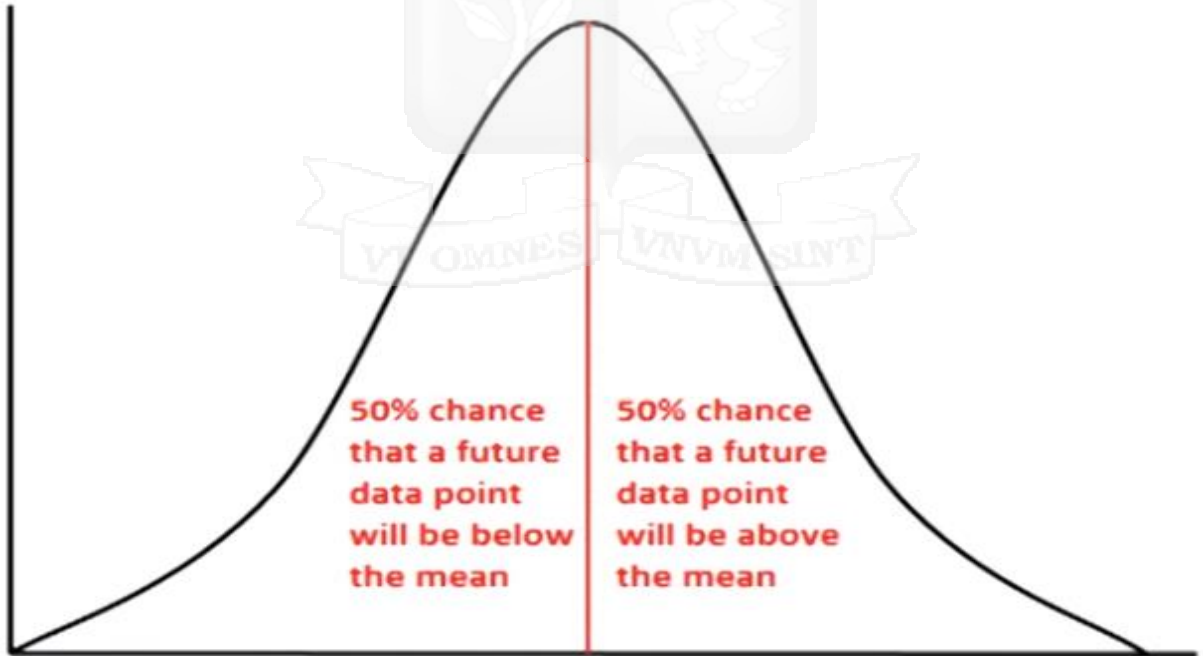


Figure 6-3 : Probability two

("Chegg Study ", n.d.)

6.1.4 Normal Distribution

A normal distribution follows the empirical rule. The area under the curve between one standard deviation below the mean and one standard deviation above the mean contains about 68% of the data. Consequently, in the past about 68% of the time, the periodic daily change of traffic volume was within one standard deviation above and below the mean.

Therefore, an assumption can be made that in the future, about 68% of the time, the rate that traffic volume will change each hour will be within one standard deviation of the rate that traffic volume is expected to change.

Additionally, in the past about 95% of the periodic daily traffic volume were within two standard deviations of the mean of all the periodic daily traffic volume. Therefore, an assumption can also be made that on a daily basis about 95% of the time in the future, the rate of traffic volume change of each day will be within two standard deviations of the expected future rate of change.

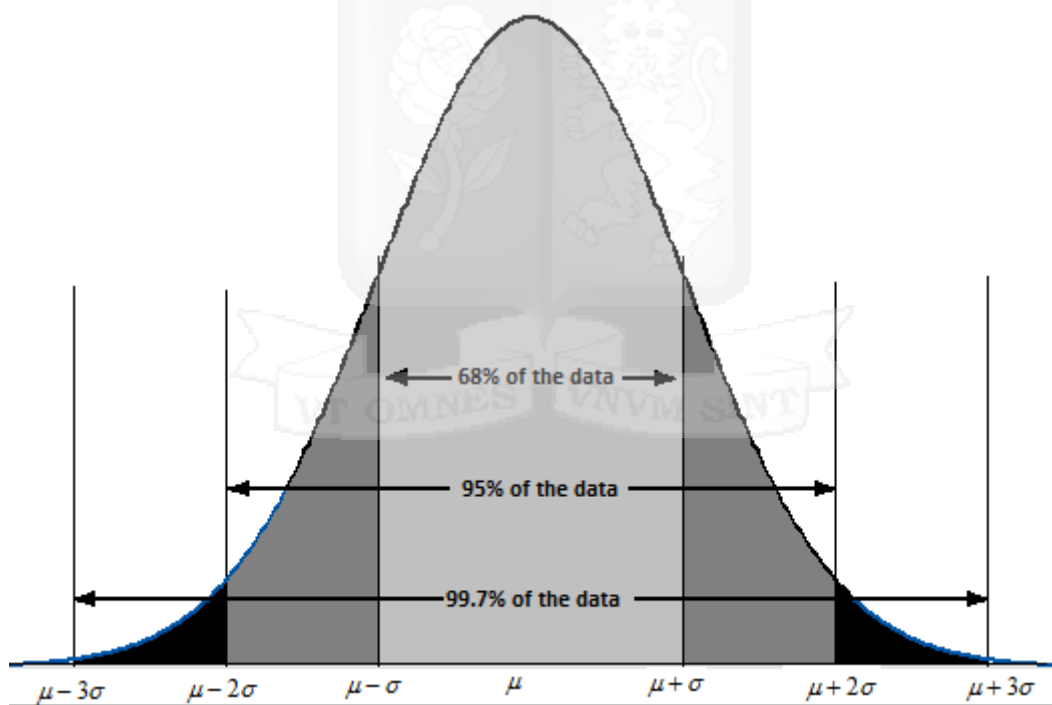


Figure 6-4 : Normal Distribution

(“Chegg Study”, n.d.)

6.1.5 Forecasting Evaluation

Essentially, for accurate traffic volume forecasting, the system ought to provide expected traffic volume for a route at any given hour and time of the day. As such, the system was evaluated on its capability to provide expected traffic volume values. This was achieved

with the use of Monte simulation which simulates the fluctuation of traffic volume for two different routes at different hours of the day. The graph below indicates the different values at the different times of the day:

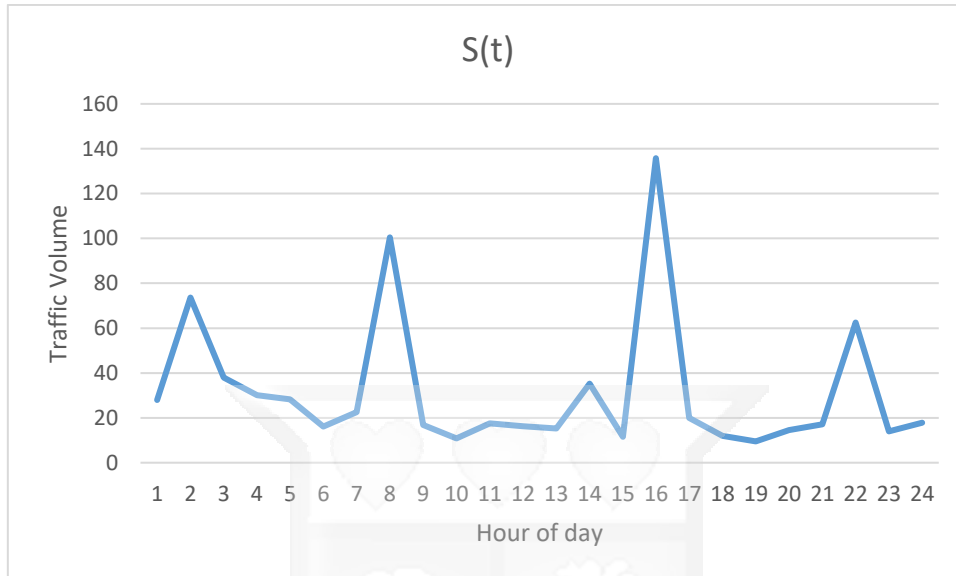


Figure 6-5 : Simulation one

In the figure 6-5 above and figure 6-6 below, the system generates traffic volume values for the different hours of the day. Hour one represents the period between zero hundred hours and one hundred hours while hour 24 represents the period between twenty three hundred hours and zero hundred.

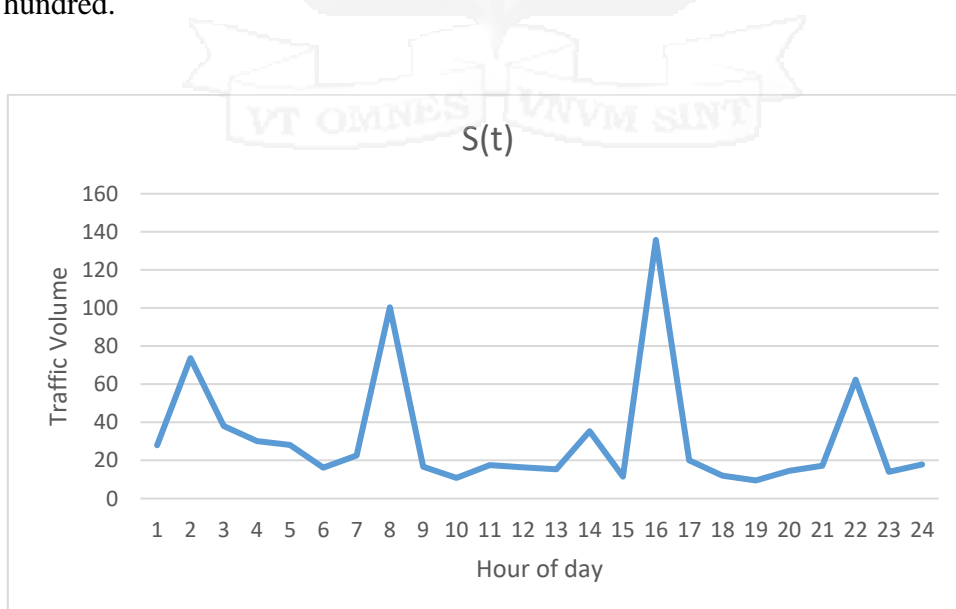


Figure 6-6 : Simulation two

Therefore, comparing the two graphs which are each generated separately, a change in the volume of traffic for the different times is noted. Notably, this means that for every hour of the day on the same routes, there is a fluctuation of traffic due to different unexpected variables. These variables can range anywhere from bad weather, road blocks or even accidents. Notably, this is acceptable since this model depicts what is most likely going to be the traffic volume at a particular point in time. The model therefore depicts results of traffic volume fluctuations with the greatest odds of happening.

The algorithm model however does not explicitly account for the various factors that affect traffic such as weather, accidents or any other incidents. The Brownian Algorithm therefore assumes that all factors remain constant. Notably, the Brownian motion algorithm includes a random step or drift that caters for fluctuations in traffic volume when affected by the different variables.

A test run was conducted using Pandas and the results were compared with the values that were in the csv as follows:

21	49	63	64	58	58	53	73	70	64	77	87	92	77
-----------	----	----	----	----	----	----	----	----	----	----	----	----	----

Figure 6-7 : Sample Comparison values

From the figure, at time 21 (between 9pm and 10pm) in bold, the traffic volumes for thirteen days at the same time are as indicated. The lowest figure is 49 vehicles and highest value is 92 vehicles. Having the historical data and running it through the Brownian motion algorithm, the following results as shown in figure were achieved:

```

step = x/(365*24)
step_root = math.sqrt(step)

final_value = roads_mean * math.exp(((annual_drift - volatility_halved)*step) + (annual_volatility * step_root))
print(final_value)
#print(annual_volatility)
#print(annual_drift)

```

```

Enter the streetLITTLE CLOVE ROAD
Enter the time21
37.32154775145295

```

Figure 6-8 : Expected future volume Test-run

The Model was able to predict that at the same time on a future date, the expected traffic volume will be an average of 37 vehicles on the street “Little Clove Road”. In conclusion, the Brownian motion model is probabilistic in nature. The outcome of the values is based on the normal distribution curve that assumes a fifty fifty chance of occurring. It assumes that in the future there is a 50% chance that the data will be above the mean and a 50% chance that the data will be below the mean with all factors remaining constant. Therefore, going by this assumption the value predicted by the model is within the range of probability of occurrence.

6.2 Validity of the proposed solution

Vehicle routing and control research has been undertaken for a period of time with the research mostly focused on least cost paths and fastest routes depending on time day. This has led to the development of a myriad of applications and algorithms that rely crowd-sourced data to help them in predicting current traffic scenarios. However, the systems do not cater for a scenario where users are able to view the number of expected vehicles at a particular route at a particular point in time. The proposed system caters for the option where users are able to view expected traffic volume on a route at a particular time, these is especially useful to government urgencies which can use the model to predict the growth in traffic and therefore help them make better and informed decisions with regards to planning especially in urban areas and cities. Additionally, the model can be used alongside other traffic navigation applications and algorithms to help better inform users on the expected traffic situations as they travel. Lastly, the model does not in any way require user data or input for it to run.

6.3 Limitations of the Application

The current model requires a very large dataset which may not be readily available as it requires that data is collected on continuous basis and over a long period of time so as to be able to make accurate predictions. In order to collect such kind of data, applications such as the video vehicle detection. One camera is used to detect vehicles passing at a particular point, counts the vehicles in real-time and uploads the data to server. The cameras are also able to work during any time of day or weather. The camera is able to identify the direction of movement of a vehicle. The data is uploaded directly to a server a cycles of fifteen minutes

where it can be analyzed. It does not accurately take in to consideration other variables that affect the movement of traffic.



Chapter 7: Conclusions and Recommendations

7.1 Conclusion

This research intended to develop a model to reroute traffic based on the volume of vehicles on a particular route. To enable successful execution of the research it was necessary to understand how traffic flow is and how it changes depending on time of day and even weather. To achieve this relevant literature was reviewed to determine the challenges encountered in the current processes of navigating traffic especially in urban areas. Further literature was reviewed to understand the application of various methods employed around the globe to help aid in combating this phenomenon.

The main objective of this research was to develop a Route suggestion model that leverages on Brownian motion to automatically predict expected traffic volume at a particular route and a particular time. Hence, the model should ideally give different values depending on the time of day.

Vehicle Traffic Volume data was retrieved from Kaggle Data sets and pre-processed, and further split into the different roads. The training set was such that we needed to use data for two different routes to be able to validate the model. The testing set was used to test the model and the output values were almost similar to the values that were initially in the dataset. Consequently, two sample routes were used to conduct the experiment in order to validate the approach taken by the researcher. The results of the experiment show that the predictions done by the model were close to the initial values of the dataset.

7.2 Recommendations

This work showed that the GMB model can be used to predict expected traffic volume at a future time and date instead of the frequent human analysis of vehicles currently employed by bodies such as NTSA officers; significantly improving the amount of time taken to analyse traffic volume changes and also assist governments in planning for road building projects. The researcher proposes several recommendations based on the findings made:

- (i) The research notes that better prediction results would have been obtained if a large data set had been used, that is more data on specific routes. A total number of 3395 rows of data were retrieved but only 24 rows were used which belonged to two different routes for the year 2011-2012. The rest of the data was grouped into different streets.

- (ii) The final training data was not large, the researcher therefore recommends that the size of the training data to be increased by collecting and retrieving more traffic volume for particular streets to improve the accuracy of the model. The best data set would require traffic volume of at least one year for an efficient and quality model.

7.3 Suggestions and Future Research

Traffic routing especially in urban areas continues to be a menace as the number of vehicles going into and out of the cities is growing. This is especially because of the growing middle class population. Some countries like Singapore have been able to curb the traffic menace by limiting the number of vehicles and improving public transport hence significantly reducing cost of travel and reducing number of vehicles within the city. For developing countries, the development and improvement of the public transport sector would help reduce the traffic menace. This coupled with increased security. The government also needs to invest in data especially primary traffic data. This would greatly assist researchers and the government on insights on areas to improve road infrastructure as well as reducing the amount of road accidents. Furthermore, the suggested model requires that data be updated on a regular for efficient traffic prediction, this acts a bottleneck to the current model as it has to be manually updated. An automation of the traffic data input would greatly increase the efficiency of the model. In addition, future researchers should consider the length of the stretch of road as well as which direction the vehicles are moving. This will contribute greatly to improving the accuracy of the current model. Consequently, the consideration of the various dynamic factors such as weather, holidays, and accidents would greatly improve accuracy of the model. This would help to better classify the outcomes of traffic volume based on the various dynamic factors. Lastly, collection of traffic volume data can be eased by using traffic sensors or video vehicle detection to count the number of vehicles passing at a particular point.

References

- Boaz, A., & Ashby, D. (2003). *Fit for purpose?: assessing research quality for evidence based policy and practice* (pp. 1-55). London: ESRC UK Centre for Evidence Based Policy and Practice.
- Burdzy, K. (2014). *Brownian Motion and its Applications to Mathematical Analysis* (Vol. 2106). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-04394-4>
- Dictionary, Encyclopedia and Thesaurus - The Free Dictionary. (n.d.). Retrieved October 25, 2018, from <https://www.thefreedictionary.com/>
- Eppstein, D. (1998). Finding the k shortest paths. *SIAM Journal on computing*, 28(2), 652-673.
- Gendreau, M., Guertin, F., Potvin, J.-Y., & Taillard, É. (1999). Parallel Tabu Search for Real-Time Vehicle Routing and Dispatching. *Transportation Science*, 33(4), 381–390. <https://doi.org/10.1287/trsc.33.4.381>
- Greedy Algorithm. (n.d.). Retrieved from <https://www.algoskills.com/greedy.php>
- Hartgen, D. T., Fields, M. G., & Moore, A. T. (2009). Gridlock and growth: the effect of traffic congestion on regional economic performance (No. 371). Los Angeles, CA: Reason Foundation.
- Jiang, S., Zhang, J., & Ong, Y.-S. (n.d.-a). A Pheromone-based Traffic Management Model for Vehicle Re-routing and Traffic Light Control (Extended Abstract), 2.
- Jiang, S., Zhang, J., & Ong, Y.-S. (n.d.-b). A Pheromone-based Traffic Management Model for Vehicle Re-routing and Traffic Light Control (Extended Abstract), 2.
- Karatzas, I., & Shreve, S. E. (1998). Brownian motion. In *Brownian Motion and Stochastic Calculus* (pp. 47-127). Springer, New York, NY.
- Kenya | Road Transport: No of Motor Vehicles: Registered | Economic Indicators. (n.d.). Retrieved February 1, 2019, from <https://www.ceicdata.com/en/kenya/road-transport-number-of-motor-vehicles-registered/road-transport-no-of-motor-vehicles-registered>
- Kenton.W.(2019, April 15). Monte Carlo Simulation Definition. Retrieved from <https://www.investopedia.com/terms/m/montecarlosimulation.asp>
- Lohr, K. N. (2004). Rating the strength of scientific evidence: relevance for quality improvement programs. *International Journal for Quality in Health Care*, 16(1), 9-18.
- Mishra, M. (2018, June 19). Understanding the Central Limit Theorem. Towards Data Science. Retrieved from <https://towardsdatascience.com/understanding-the-central-limit-theorem-642473c63ad8>

- Mutavi, L. (2017, April 17). Nairobi ranked world's "second-worst city." Retrieved October 22, 2018, from <https://www.nation.co.ke/news/Nairobi--is-world-s-second-worst-city-/1056-3892294-1208syiz/index.html>
- Overview. (2018, October 5). [Text/HTML]. Retrieved October 24, 2018, from <http://www.worldbank.org/en/topic/urbandevelopment/overview>
- Research Methodology – Contemporary Practices. (n.d.). Retrieved from <https://www.cambridgescholars.com/download/sample/61903>
- Scenario-Based planning for Partially Dynamic Vehicle Routing With Stochastic Customers.pdf. (n.d.).
- Schrank, D., Eisele, B., & Lomax, T. (2012). TTI's 2012 urban mobility report. Texas A&M Transportation Institute. The Texas A&M University System, 4.
- Sweet, M. (2011). Does traffic congestion slow the economy?. *Journal of Planning Literature*, 26(4), 391-404.
- Towne, L., & Shavelson, R. J. (2002). *Scientific research in education. Washington, DC: National Research Council.*
- Traffic Index by City 2018 Mid-Year. (n.d.). Retrieved October 22, 2018, from <https://www.numbeo.com/traffic/rankings.jsp>
- Wang, S., Djahel, S., McManis, J., McKenna, C., & Murphy, L. (2013, October). Comprehensive performance analysis and comparison of vehicles routing algorithms in smart cities. In *Global Information Infrastructure Symposium-GIIS 2013*(pp. 1-8). IEEE.
- World population projected to reach 9.7 billion by 2050 | UN DESA | United Nations Department of Economic and Social Affairs. (2015, July 29). Retrieved October 24, 2018, from [2015-report.html](#)
- World Health Organization. (2018). *Global status report on road safety 2018*. World Health Organization.
- World's population increasingly urban with more than half living in urban areas | UN DESA | United Nations Department of Economic and Social Affairs. (2014, July 10). Retrieved October 23, 2018, from [world-urbanization-prospects-2014.html](#)

Appendix A: Sample Data

ID	GIS ID	Roadway Name	From	To	Date	1	2	3	4	5	6
						12:00-1:00 AM	1:00-2:00AM	2:00-3:00A	3:00-4:00AM	4:00-5:00A	5:00-6:00A
1	15540	BEACH STREET	UNION PL	VAN DUZER STREET	9/1/2012	20	10	11	14	13	20
2	15540	BEACH STREET	UNION PL	VAN DUZER STREET	10/1/2012	21	16	8	6	13	13
3	15540	BEACH STREET	UNION PL	VAN DUZER STREET	11/1/2012	27	14	6	5	12	16
4	15540	BEACH STREET	UNION PL	VAN DUZER STREET	12/1/2012	22	7	7	8	11	12
5	15540	BEACH STREET	UNION PL	VAN DUZER STREET	1/13/2012	31	17	7	5	13	28
6	15540	BEACH STREET	UNION PL	VAN DUZER STREET	1/14/2012	42	27	21	18	21	13
7	15540	BEACH STREET	UNION PL	VAN DUZER STREET	9/1/2012	27	12	12	4	22	27
8	15540	BEACH STREET	UNION PL	VAN DUZER STREET	10/1/2012	26	16	11	13	16	27
9	15540	BEACH STREET	UNION PL	VAN DUZER STREET	11/1/2012	32	16	8	9	15	26
10	15540	BEACH STREET	UNION PL	VAN DUZER STREET	12/1/2012	24	12	7	18	11	23
11	15540	BEACH STREET	UNION PL	VAN DUZER STREET	1/13/2012	39	22	8	6	16	30
12	15540	BEACH STREET	UNION PL	VAN DUZER STREET	1/14/2012	45	45	28	23	21	19

Figure 8-1 : Sample Data

