SHIBAURA INSTITUTE OF TECHNOLOGY

Demand Modeling and Optimization Algorithms for Rebalancing Operations in Bike-Sharing Systems

by

Kanokporn Boonjubut

A thesis submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
in the Department of Functional Control Systems
Graduate School of Engineering and Science

March 2022

“Be the change that you wish to see in the world.”

Mahatma Gandhi

Acknowledgements

Foremost, the completion of this thesis would not be possible without expertise. I would like to express my deep and sincere gratitude to my advisor, Professor Hiroshi Hasegawa, who gave me an opportunity and provided guidance throughout this thesis. I would also like to thank my co-advisor, Professor Kayoko H. Murakami, and the rest of my thesis committee, including Associate Professor Dai Watanabe, Associate Professor Manabu Ichikawa, and Professor Keiji Suzuki, for their insightful comments and suggestions.

My special appreciation to the Royal Thai Government scholarship program for providing funds to study for a Ph.D. degree, the Office of Educational Affairs in Japan, the Office of the Civil Service Commission, Ministry of Higher Education, Science Research, and Innovation, and Nakhon Ratchasima Rajabhat University to provide an opportunity to enroll in higher education. All of them have encouraged me during the Ph.D. course.

Last but not least, I am grateful to all of my family and friends for all the support and encouragement during my Ph.D. studies. They are always by my side in all situations, and make me laugh even when I do not want to smile. It was a great comfort and a relief to be able to provide my work completely.

Kanokporn Boonjubut

Abstract

The number of bike-sharing services has rapidly increased in many cities worldwide. Bike-sharing schemes have become a popular and environmentally friendly transportation mode. They are an alternative to urban transport for connecting the first/last mile to main public transport modes. The bike-sharing system is a service that allows a customer to rent a bike from a bike-sharing station and return it to another bike-sharing station after they reach their destination in a short while. Thus, the impact of the bike distribution system based on the frequency of bike usage needs to be assessed.

One of the main challenges of the bike-sharing system operating costs is allocating enough bikes and parking space. There is a shortage of bikes due to an imbalance in bike distribution. Bike-sharing systems experience an imbalance in customer demand too at each station. The operation's primary focus has been on rebalancing the bikes among stations. To avoid customer dissatisfaction, the demand-supply imbalance of the bikes in the system must be minimized. The bike-sharing system operator needs to maintain a balance between available bikes and available docking.

This study was conducted to improve the efficiency of bike-sharing systems and to predict the demand accurately so that the planner knows how many bikes are needed at every station, which will aid the management process of the bike-sharing stations. First, a method was presented for predicting the demand for bikes. This study proposed an efficient and accurate model for predicting bike-sharing service usage using various features of a machine learning algorithm. This work compared the existing techniques for the sequential data prediction of artificial intelligence for time series data and analysis. It used the multivariate model with a recurrent neural network (RNN), a long short-term memory (LSTM), and a gated recurrent unit (GRU). In addition, it considered combining the LSTM and GRU methods to improve the model's effectiveness and accuracy.

Further, this work focused on the imbalances caused by problems with insufficient bikes or docking stations in such schemes, leading to operating costs for

relocating the bikes. This work presented a model for solving the bike-sharing relocation problem. Though the artificial bee colony (ABC) algorithm is an efficient approach, it is insufficient for the selection strategy. ABC has been adopted in various problems to improve the performance of various systems. This research proposed a modified ABC algorithm in a neighbor solution, namely guided local search (GLS), to enhance the solution performance. Computational experiments were performed to find out the best modeling solution in the case. The implementations were experimental for the same data instances, which made it possible to compare the performance algorithms to solve the bike-sharing relocation problem of pick-up and drop-off.

Recent studies have proposed reinforcement learning as a computation-based learning method that yields more accurate results. This study proposed Q-learning and SARSA of reinforcement learning as a fast convergence solution to a routing problem. We implemented this by determining the distance between each station and considering the capacity of the trucks. The model of this study introduced reinforcement learning, consisting of the Q – Learning and SARSA for finding the solution. Q-Learning and SARSA produce better results than GA and ABC. Consequently, our proposal can be applied to the CVRP problem.

An analysis has been carried out in this work using data mining to determine bike activity patterns at the station and gain insights into them. The activity model revealed an imbalance in the bike distribution. The data mining process supports operating decisions of bike-sharing systems to ascertain the critical point of the system, making a resolution easier. This case study used a simulation based on the arrival rate, which assists in managing a bike-sharing rebalancing system with the most profitable objective and meeting the users' needs.

One factor that should not be overlooked is the location of the depot. Here, a method was proposed, based on cluster analysis, for considering depot location in bike-sharing schemes. The main objective is to reduce operating costs by minimizing the total distance required in relocating bikes. The WK-means and Elbow Method were used to determine the number and location of depots; thereafter, the total distance required for different depot location options was considered. The results indicate that the proposed method performs well in terms of reducing the total distance required.

The components related to the development of bike-sharing systems consist of 1) the forecast model by combining the LSTM and GRU methods to improve effectiveness and accuracy of the predicted model, 2) modified ABC Algorithm, as the GLS-ABC can be a better solution than the original one, 3) application of reinforcement learning to use in routing to achieve the shortest distance (the impact of minimizing the route tour cost in solving the bike-sharing relocation problem), 4) model for maximizing profit by simulating and 5) locating the depot by the clustering method which found the results of the changes that occurred in the system development process. The methods presented in this research not only help to improve the quality of the bike-sharing system but also serve as a foundation for the development of methodologies related to industries with similar problems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Mobility services provide more mobility options capable of having unprecedented effects on the sustainable development of urban planning [1]. A transportation network facilitates shared mobility choices that enable new mobility services. Most notably, it employs ingenuity, which many see as the most important transportation newness recently. Moreover, they present the reality of road capacity, traffic conditions, and urban planning, especially in the absence of strict rules and domination. Mobility as a Service (MaaS) is a system that provides customers with a broad range of mobility services provided by a mobility operator. The service provider handles and delivers transportation demands through a single interface [2–4].

The major participants in any MaaS, the public transit providers, are critical. Many transportation companies have responded to customer demand by introducing new modes of transportation by partnering with complementing modes. MaaS idea [5] intends to integrate several means of transportation for instance public transportation, bike-sharing, car-sharing, and bus on demand into seamless trips through a single user interface. Combining public and private transportation allows for more flexible mobility options and travel possibilities. The fundamental idea behind MaaS is to meet mobility demands without the necessity for a private automobile or a variety of public transportation or mobility service provider travel cards.

One compelling argument in favor of MaaS is that it can support with the "first-mile/last-mile" problem. Because of the flexibility of numerous mobility services, this might be a concern. If they serve as feeders for permanent transit lines and rail services, they will be exploited Furthermore, MaaS has the potential to increase transportation accessibility because of its ability to fill spatiotemporal service coverage gaps.

In many countries, interest has increased in vehicle sharing systems (VSS) in cities, with policies intended to discourage citizens living or working in urban areas, from using private cars in cities by reducing the number of parking spots, the width of the road, etc. In this background, the vehicle sharing system seems to be a good solution to reduce traffic and parking congestion, noise, and air pollution [6].

Bike-sharing service is a kind of MaaS that has also been popular in the last few years, as the growing numbers of people and vehicles have increased traffic congestion and led to many environmental concerns in recent years. For instance, there has been significant growth in the number of customers of Citi bike in New York City (see Figure 1.1). This service can also serve as a connection between the primary transportation mode and the origin/destination. Bike-sharing services are gaining popularity worldwide. Numerous countries in Europe and Asia have been using the bike-sharing system (BSS) for decades. Thus, bike-sharing has seen growing demand as a transportation service [7]. Moreover, BSSs are sustainable transportation alternatives to private transport, as they do not lead to carbon emissions, traffic congestion, or the use of non-renewable resources [8]. However, the service quality has a great impact on customer satisfaction, which affects the increase in the client base, the service's popularity, and the overall economic performance of bike-sharing companies.

City residents often use bike sharing to commute to work and for trips, as bikes are convenient to use in cities and allow users to ride rented bikes from one bike station to another. Bike-sharing services are a good option for urban transportation in smart cities too [9]. Though BSS can use various technologies such as sensor devices to make them smart, a user simply uses a smartphone to assess the location of available bikes and docks that can be used for daily mobility. This system facilitates the use of rentals that can be picked up and dropped off at any station. Moreover, bike-sharing also reduces the need to travel by car or other modes

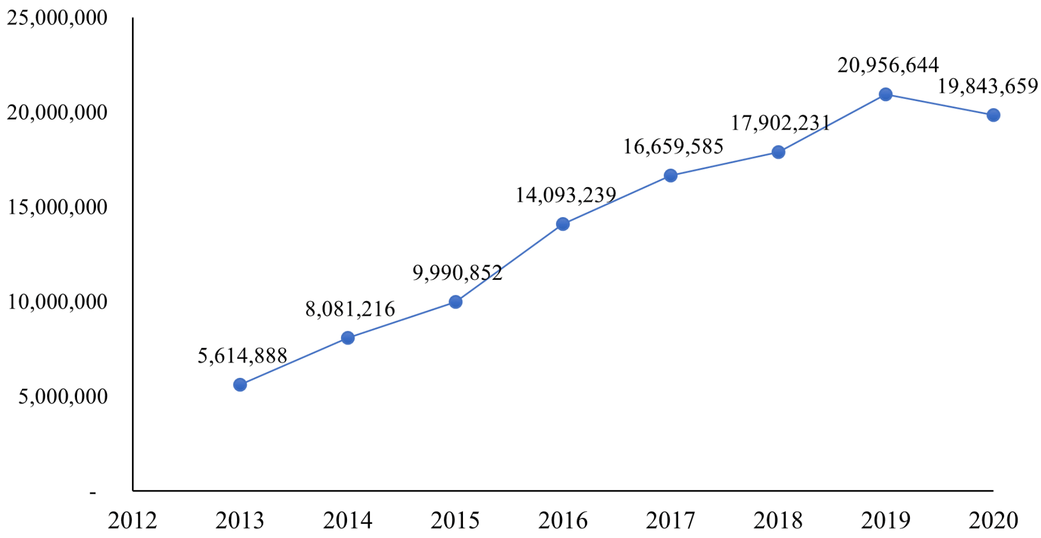


FIGURE 1.1: Number of customers use bike sharing in Citi bike.

of transport for shorter distances, thereby reducing congestion and can consequently assist in decreasing vehicle emissions, including carbon dioxide, apart from reducing fuel costs.

Bike-sharing services have also received much attention in metropolitan and tourist areas. In general, a BSS allows registered customers to request a ride after indicating the pick-up and drop-off times and locations. Thereafter, customers will be able to find an available bike and park it at any station after reaching their destination. Thus, BSSs need to maintain the optimal number of bikes and parking spots at each station.

Tourism has become an increasingly popular and essential industry, worldwide. According to statistics, the number of tourists tends to continuously increase, especially in Japan [10]. As tourism is related to travel, transportation is vital for moving passengers and goods from one point to another. Travelers have various expenses and satisfaction factors that impact their preferred modes of transportation, which include driving, walking, and using a taxi or bike-sharing service.

To save time, provide connectivity for short distances, and enable multimodal transportation connections, bike-sharing services have made bike rentals available to travelers and tourists, so that they can rent bikes and return them at any station.

Most bike-sharing systems provide automatic systems for users and operators so that customers can just use their smartphones to locate available bike stations, which makes it easier and more convenient to attract more customers. Recently, the frequency of using bike-sharing services has significantly increased due to the popularity of green travel, with people becoming increasingly aware of pollution and other common health issues. Moreover, due to the increase in carbon dioxide (CO_2) levels, many people are taking measures to reduce greenhouse gas emissions in every economic sector.

A well-designed bike-sharing system allows users to access it to travel easily, quickly, and cheaply, leading to a good image among service users, enhancing their satisfaction, and ultimately increasing the profits of the service providers. However, a system that lacks proper design can lead to a management failure, resulting in the collapse of the system [11]. One of the core problems of the bike-sharing system is that of users finding bikes and docking (for return) insufficient. Therefore, comprehensive planning and management are essential to maintain BSS as a component of the urban transportation system.

1.2 Objective

The majority of current bike-sharing systems are station-based. One-way rides are common in such systems, allowing customers to return bikes to locations other than where they were picked up. The station-based kind is a simple to manage renting and return system that is also well-organized in the service area. For managing to put enough bikes in the places where customers need to use them, and managing the parking slots to serve customers' needs to return the bikes at each station. The objective of this thesis is to manage bike-sharing systems to improve system performance and meet the needs of customers. To that end, some essential requirements are defined:

Research Question 1:

What specific dynamics may be found in the bike-sharing system?

The empirical data analysis provides a framework essential for identifying any existing imbalances. Demand predictions based on historical data may be used to identify supply and demand gaps for bike-sharing systems. The performance and operation of the system may be improved if the working area is adequately balanced. By means of effective forecasting, the process will be effective as well.

1) To implement a machine learning algorithm for the investigation of demand for bike-sharing systems, this thesis proposes a method for planning and management of bike-sharing systems, to improve the efficiency of the system, and to enhance user satisfaction for the survival and flourishing of the business. Firstly, for planning the systems, we need to estimate the additional demand to prepare the budget for manpower, tools, etc. Accurate demand forecasts allow timely estimates of the resources needed.

Research Question 2:

How to increase the efficiency of bike-sharing system?

This point entails a detailed examination of redistribution methods, also known as relocation strategies.This is in the process of relocating the bikes in the system to meet the needs of the customers and indicates the potential for system optimization. However, the operating cost is implemented as a result of the relocation procedure.Therefore, it is necessary to find a cost-effective way to carry out the relocation process.

2) To develop meta-heuristics for relocation of bike-sharing. The objective is to minimize the number of users who cannot be served, including those who try to take a bike from an empty station or to return it to a full station. The balancing problem requires the scheduling of truck routes to visit the stations performing pickup and delivery. Meta-heuristics are methods derived from the development and improvement of heuristic methods and can optimize the resolution of complex problems with many decision variables efficiently and timely. Therefore, this research recognizes the importance of applying the meta-heuristic method in the relocation of bike-sharing.

3) To implement a machine learning algorithm for relocation of bike-sharing. Machine learning is the application of advanced statistics to learn to identify patterns in the data. This is a model based on the existing algorithms and datasets to develop an appropriate problem-solving process and make predictions therefrom. This method is highly efficient and accurate, especially for complex problems or big data. Considering such importance, this research proposes a method of machine learning to solve problems for relocation of bike-sharing.

Research Question 3:

How should fleet imbalances be managed in such a way that there are limitations in terms of improving the capacity of individual stations or building new stations?

In a situation where bicycle rental business has already been carried out and unable to improve the efficiency of the system by improving the size of the bike rental station to meet the needs of customers.

4) To implement a simulation for investigating the number of bikes in systems. This thesis also proposes an effective method to reflect the overall picture of managing supply and demand imbalances through simulation methods, leading to profits.

Research Question 4:

Where should the relocation center or depot be located to minimize operating costs?

In the relocation process, trucks are taken from the hub or deppo to rebalance the bikes at each station in the system. In terms of operating costs, the location of the depot is critical, as is the immediate response to the needs of customers.

5) To implement the clustering method for determining the depot bike-sharing systems. A sustainable depot location and operation are required for distributing bikes at each station to meet the needs of users. An important factor that can manage the distribution of bikes at each station is to keep abreast of the needs of users. A central location, called the depot, can also be managed to reduce operating costs by choosing a distribution. The cluster analysis was used for finding a suitable

number of depots and locations to meet customer needs, i.e., to find the shortest total distance while relocating the bikes in the systems, which will reduce the operating cost.

1.3 Outline of this thesis

The structure of this thesis is organized as follows: Chapter 1 introduces the background of this study, to identify the area of study. Chapter 2 discusses the literature review of works related to the optimization of bike-sharing systems. Chapter 3 demonstrates the development of bike-sharing systems to improve efficiency. Chapter 4 presents the demand-prediction model for forecasting bike-sharing demand. Chapter 5 dwells on the development of meta-heuristic algorithms for truck routing to relocate bike-sharing systems. Chapter 6 discusses the reinforcement learning to implement truck routing in bike-sharing relocation. Chapter 7 talks about the simulation model for operation in bike-sharing rebalancing. Chapter 8 discusses the clustering approach for determining depot location for relocation of bike-sharing. Finally, Chapter 9 concludes this thesis and discusses the future work.

Chapter 2

Literature Reviews

2.1 Mobility as a Service

Mobility as a Service(MaaS) [12] provides customers personalised mobility solutions based on their specific demands, with easy access to the most appropriate mode or service combined in a bundle of flexible travel service possibilities, highlights the end-user perspective. Although this design satisfies the hype prerequisites of a paradigm shift, it only specifies theoretical goals rather than actual substance.

MaaS stands for the user's comfort was prioritized during the design process, and it is a new user-centered concept that makes use of digital platforms and real-time data. MaaS platforms can connect public transportation, ride-hailing, car-sharing, bike-sharing, walking, and other modes of transportation into one smooth journey.

MaaS is a platform that links public transportation, ride-hailing, car-sharing, bike-sharing, walking, and other modes of transportation to provide customers with a single, seamless trip. The MaaS platform allows user to plan vacation, book it, and pay for it all in one place. It enables users to specify their desired trip characteristics, such as modes of transportation or trade-offs between the number of transfers and journey duration.

Figure 2.1 displays one of the most widely used and arguably the first instances of the MaaS architecture [13, 14], which also depicts the MaaS ecosystem. The

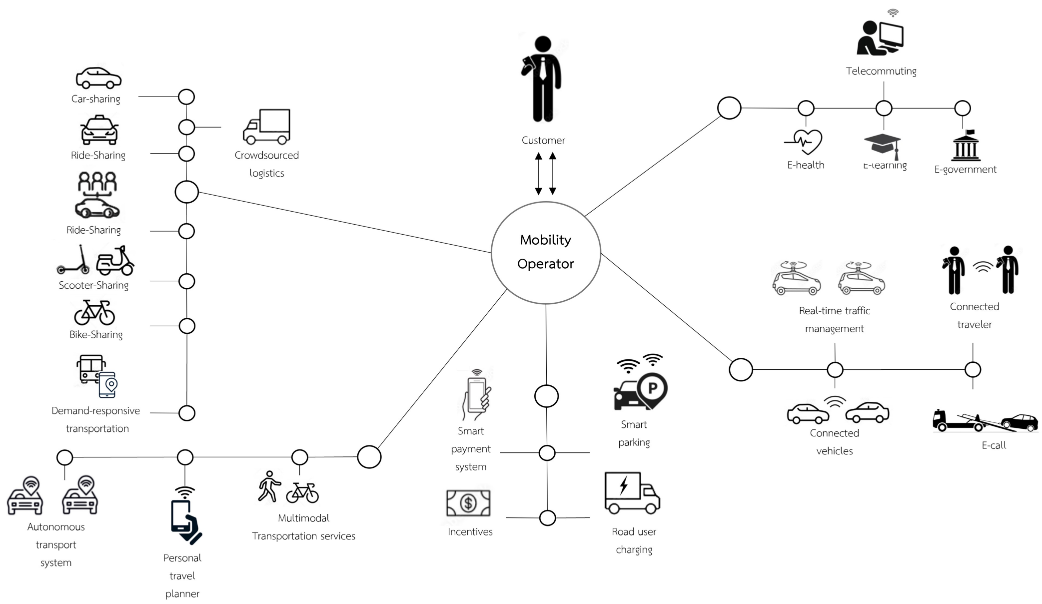


Figure 2.1: Framework of MaaS [13].

Finnish Ministry of Transport and Communications released the framework as a description of the envisioned scenario for MaaS operators.

Mobility service providers have comparable fundamental service supply procedures and operational elements [15]. In reality, due to the changing nature of mobility services, Sprei [16] indicates that conceptual differences between services might get obscured. Furthermore, examining mobility services gives a solid conceptual foundation as well as solutions to common operational issues. Despite the wide spectrum of existing mobility services, car-sharing, bike-sharing, ride-hailing, and demand-responsive transportation are considered the most important for generalization. These cover the majority of mobility services' operational responsibilities, as well as the three primary types of vehicles that are often represented in network models: cars, buses, and bikes. Current MaaS offerings are more specific and include the following:

1) Car-sharing is a type of car rental that allows users to rent cars. A car-sharing service can be provided in a variety of ways [17], according to mode.

- Two-way (station based): the available cars are parked at pick-up stations, which are established parking lots by the service provider or local government, and the journey must start and end in the same space under the two-way mode. As a result, this operational model ignores intermediate parking, which refers to the stops that a client may make for personal reasons. The parking lots are predetermined. The last one is identical to the last one, but in the event of a one-way trip, the parking lot where the journey ends may differ from the parking lot where it began. The parking lots are predetermined.

- One-way (station based): the customer must return the cars to a parking lot where the end of the journey may differ from the starting point. The customer must return the car to the designated parking place.
- Free-floating: this is the most recent variety to hit the market. Customers can return the car to the servicing area by parking in a public place. The travel can begin and terminate at any point within the service region.

2) Ride-hailing is a service that connects customers and drivers via an app-based platform such as Uber or Lyft, allowing for real-time transactions. The rise of ride-hailing services has been attributed to rising travel demand and technical advancement. Smartphone users can now readily use ride-hailing applications [18], and a market for transportation services based on sharing private cars has emerged as a result of technological advancements.

App-based ride-hailing, according to Tang, Bao-Jun, et al. [18], includes both hailing cabs via smartphone applications and sharing private cars using app-based ride hailing services. The way customer hail and share vehicle rides has evolved as a result of app-based ride hailing, which provides on-demand transportation services for passengers and allows private car owners to share their cars.

3) Ride-sharing is a service in which private car drivers share their vehicles with customers who have comparable destinations or itineraries on their own excursions. Ride-sharing is an innovative mode of transportation that expands mobility options by providing a variety of modes of transportation. Participants can benefit from pooled travel costs, reduced commuting stress, and reduced waiting and travel time by using high-occupancy vehicle lanes. The language used to describe new and growing transportation services may be complex [19], and it is occasionally employed erroneously, leading to erroneous public perception and usage of the services. For

example, when referring to ride-hailing firms in their original form, in the name of ride-sharing is frequently abused.

In principle, a ride-sharing journey should contain two or more passengers; nevertheless, most ride-hailing excursions only carry one passenger [20]. Transportation Network Companies (TNCs), ride-hailing, ridesourcing (supplying), ride-booking, ride-matching, on-demand-trips, and app-based rides are some of the other terms.

4) Demand-responsive transportation is a means of transportation in which passenger cars, vans, or small buses respond to calls from passengers or their agents to the transit operator, who then deploys a vehicle to pick up and carry the passengers to their destinations [21]. A demand-response (DR) operation is defined as follows: the vehicles do not operate on a fixed route or schedule, except perhaps on a temporary basis to meet a special need; and the vehicle may be dispatched to pick up several passengers at different pickup points before taking them to their respective destinations, and may even be interrupted en route to these destinations to pick up other passengers.

5) Scooter-sharing has gained in popularity as a novel vehicle-sharing idea in metropolitan areas [22], with numerous distinct advantages that make it an attractive alternative to travelling around town. Due to the lower speed of scooters and the smaller market segment compared to other shared modes, such as car-sharing, scooter-sharing systems are mainly limited to metropolitan areas. Compared to scooter sharing and bike sharing with cars, it is faster to commute in metropolitan areas during peak traffic times[23]. Existing systems are divided into two categories: station-based systems, which have a certain number of fixed rental locations, and free-floating systems, which allow customers to pick their own scooter drop-off places within a predetermined service area.

In scooter sharing models, a range of motorized and non-powered scooters can be used [24]. In the case of motorized scooters, the scooter service normally includes fuel or electric charging, maintenance, and parking. Scooter sharing services are divided into two categories: first, standing electric scooter sharing uses electric-powered shared scooters with a standing configuration that includes a handlebar, deck, and wheels. Aluminum, titanium, and steel are the most prevalent materials used. Second, Moped-style scooter sharing, which uses shared seated motorbikes that are either electric or gas-powered, has fewer licensing requirements than motorbikes built to operate on public roadways.

6) The bike-sharing is a service that allows users to rent on-demand and return bikes in public areas from one point to another or a round trip, mostly for short periods [25]. Bikes-sharing systems are divided into three categories according to the service provided [24] as follows.

- Station-based bike-sharing systems where users can rent and return bikes at any station, called one-way. At these docking stations, bike-sharing service is provided at a fixed station.
- The free-floating system, where users can rent, use and return bikes in public areas: This system allows renting and return not only at the stations but within the service area. The operator installs a GPS device on the bikes to track their location, and finds the location from the users.
- The hybrid system is a combination of station-based and free-floating systems, where the user can rent and return bikes at stations or in the public service area.

As mentioned above, there are many variations in shared mobility services. But one of the advantages of bike-sharing service is that it is a short-distance form of movement that solves the problem of missing links between the first/last miles and main transport, and most importantly, it is a small vehicle that can move anywhere even if there is congested traffic on the road. And for this type of service, users do not need a driver's license. It is also environmentally friendly and is considered a short-term exercise that is beneficial to the health of users.

2.2 Overview of bike-sharing systems

The bike-sharing system is an approach with increasing popularity worldwide. City residents tend to rent bicycles to ride to work and travel because the use of bike is convenient, saves time, reduces global warming by reducing the emission of toxic fumes, and reduces traffic congestion and fuel costs of vehicles. Biking also represents a good and healthy exercise and supports the connectivity of multimodal transport.

Currently, most existing bike-sharing systems are station-based, as it is easy to track the bikes' location and the number of available dockings. Bike rental systems use a connection with GPS and a mobile application where the users can access the destination information and estimated rent, including the return point. The user must log in to the mobile app to register a credit card and bank account or online payment system first. The app will show the map location of the nearest bike station; thereafter, the users can unlock the bike with the code sent by the system or automated smart cards or IC cards.

However, the main problem is that the demands for bike renting and bike parking in each area are different, causing insufficiency: when the station is empty, there are no bikes for rent and to those who want to return the bike in a full station, there is no docking available. Research has proposed a method for improving the service quality of bike-sharing systems. This can be classified into two aspects [26]. The first category is qualitative research, focusing on demand and supply perceptions, which interprets, understands, and analyzes the strengths and weaknesses of BSS and users' behavior, and evaluates the efficiency of infrastructure and characteristics of stations. The second category is quantitative research that focuses on the layout of the station and resolves the bike relocation problem for matching the supply in BSS to the demand.

2.3 Forecast theory

According to the study, bike-sharing systems are used worldwide [27, 28]as they aim to increase the convenience and flexibility of transport for individuals. Further, bikes are managed, and stations are located usually based on demand. There are many methods of forecasting demand, which are presented below:

2.3.1 Meaning of forecast

Forecasting is a qualitative and quantitative approach for estimating the equipment of the product and future services for customers in the short, medium, and long terms. Forecasting is useful for planning and decision-making in all components of the company organization, as indicated below:

- Finance department: Projected demand will be used for budgeting sales, which will be the starting point for formulating a financial budget for allocating resources to all parts of the organization thoroughly and appropriately.

- Marketing department: Estimated demand will be used for fixing targets for the salespersons or for creating a product-wise sales target for use in controlling sales and marketing.

- Production department: Estimated demand will be considered for information in various operations in the production department as follows: a) Inventory and procurement management, to have sufficient raw materials for production and sufficient finished products for sales, with reasonable inventory costs; b) Labor management by organizing manpower as per the amount of the production forecast for each period; c) Determination of production capacity to provide suitable factory size, have adequate machinery, equipment and production stations for production, to allocate labor and production capacity as per the procurement of raw materials and parts required for each period of production; d) Selection of location for warehouse and production or distribution center for goods; e) Production process planning and production scheduling to organize the production process suitably for the number of products to be produced, setting the production time according to the range of demand.

2.3.2 Components of demand forecast

Forecasting is based on time frames of demand behavior. There are 3 timeframes for forecasting, viz., a) short-term forecast for less than 3 months for use in inventory management and production scheduling during each week, each month, or each quarter. In other words, short-term forecasting is used for short-term planning. b)

Medium-term forecast, covering a timeframe of 3 months to 2 years. Such forecasts cover the entire group of products or the total sales of the organization for the planning of personnel, production, total production scheduling, purchasing, and distribution. The popular forecast period is 1 year because it is exactly one accounting period. c) Long-term forecast is for 2 years or more, to forecast the total sales of the organization, to select the location of the factory and its facilities for long-term capacity planning and production process management.

Forecasts are categorized by demand behavior, where trends are level indicators by long-term demand movement. At present, demand behavior is by norm of random nature, which is not normal behavior; there are many forms, namely, trend-cyclical behavior and seasons, detailed below.

a) Trend is a continuous line on the graph, reflecting continuous increase as a characteristic of future sales.

b) Cyclical pattern is used in graphs to reflect the features of periodical increase, decrease, and evenness. It is a product life cycle that depends on technology, competition, legal, political and economic systems, which are uncontrollable factors.

c) Seasonal: When the sales graph is drawn, it reflects short-lived increases and decreases at the same time of each year, for instance, the sales of coats in winter.

d) Trend and Season reflect mixed lines between trends and seasons.

2.3.3 Forecast method

Classical Forecasting Methods can be classified as below. The Judgment Method used when there is not enough historical data to forecast the sales of a new product or when technological advances occur; there are 4 methods of such forecasting:

- Sales Force Estimates use estimates of the salesperson who is most exposed to market conditions. The salesperson will collect the sales information in the area for which they are responsible and send it to the head office. This method can be erroneous, because some salespeople are too optimistic, while some others know that the sales forecast will be used to set sales targets and furnish values that enable them to easily exceed the targets. And some salespeople do not understand that demand refers to the" purchasing power" of the customer.

- Executive Opinion is used for forecast in the case of new products that have not been released to the market before. The opinion of one or more experienced executives is obtained to forecast and formulate strategies to suit the environment, such as marketing products in international markets. The disadvantage of this method is that significant time is spent in frequent meetings of executives to formulate forecasts.

- Market Research is a method that must be done systematically by creating assumptions and collecting data from users of the product to make predictions. Market research comprises questionnaire design and methods of data collection. Samples are taken for interviews, data is collected, processed, and analyzed. This method is used for short, medium, and long-term forecasting, but is costly and needs meticulous performance of its various steps.

- Delphi Method involves the meeting of a specific group of experts with knowledge of the product. It works best when no data is available to predict, and the organization's executives do not have enough experience with the product. This method starts with circulating the questionnaire to several experts, making a report, and submitting it to all experts for them to peruse, update with new ideas and return. The process is repeated many times until a conclusion emerges from everyone. The disadvantage of this method is that it is very time-consuming and may result in unanswered questions. It is used only for new products that cannot be subjected to the other methods.

Another method, named the Causal Method, is used when the data relates to a variable of sales. Variables are internal factors such as cost of sales or external factors like relationships and are characterized as a Linear Regression, with one variable being dependent and the other independent, related to each other such that when the independent variable changes, it results in the dependent variable changing as well.

- Measure the correlation of variables: In addition, the linear equation should be examined for the relationship between X and Y, to ensure that these two variables are indeed correlated and suitable for forecasting.

- The Coefficient of Correlation measures the direction and degree of the relationship between X and Y.

- The coefficient of Determination is used to measure the influence of independent variables on sales forecast. Time Series Method is used for forecasting future sales with the same expectation. As a current or future sales or demand are influenced by trends, cycles, and unusual events, there are 3 ways to use time series.

- A simple or Naive Forecast is the prediction that future sales will be equal to current sales. This method is simple and low-cost but works well in cases where various influences affecting sales results remain stable; if an abnormal event occurs, there will be wide discrepancies.

- Moving Average is to find the average of sales by using 3 or more data periods in the calculation. After the first period, new data is averaged instead of the furthest period, which is omitted. The moving average forecast requires data for at least 3 periods, so the first value is the 4th period. For example, if data collection starts in January, February, and March cannot be predicted. The forecast will begin at the end of March by calculating the forecast value for April. The amount of data used can be odd or even. If sales are relatively stable, it is better to use a large amount of data to average, to get a forecast value closer to the true value. However, data near the time of forecasting tend to have more influence on forecast values than data farther away. Therefore, a weighted moving average is found.

- Exponential Smoothing is a weighted moving average, the weights that organize the forecast values being in the form of a calculated equation. It uses a single default data value and weights. They are weighted using a smooth coefficient (α) that is between 0 and 1, to calculate the exponential. The first forecast value is determined to be equal to the actual value of the previous 1 period (that is, using the same principles as for simple forecasting). Exponential averaging uses less data and produces faster forecasts than moving averages but yields forecasts that are as accurate as the weighted moving average weight for values.

2.3.4 Artificial intelligence techniques

Bike-sharing demand prediction is a key component for solving the bike-sharing relocation problem, as important information for planning. The accuracy of predictions is also a factor in successfully solving problems. Regarding the demand prediction literature review, several researchers have studied bike-sharing demand, proposing several models for predicting certain station-level demand.

Predicting demand using analytical algorithms involves using historical data, which includes seasonal fluctuations and behavioral patterns. The most applicable time series models are the following:

1) The Automated Reverse Integrated Moving Average (ARIMA) model finds correlations in time series data. It tries to eliminate noise from time series in order to reduce error. ARIMA provides highly accurate projected values for user-specified time intervals for planning purposes.

2) The Seasonal Autoregressive Integrated Moving Average(SARIMA) model is an expansion of the ARIMA model, which adds seasonality to support time series data with related to seasonal component.

3) The Exponential Smoothing model (ETS) generates a forecast using the weighted average of past observations to predict future values. The essence of this models is to combine error, trends and seasonal components.

For instance, Kaltenbrunner et al.[29] applied a time series analysis technique which is an auto-regressive moving average (ARMA) model, using daily usage data to predict the future number of bikes at any station. Sathishkumar [30] adopted various statistical regression models for bike-sharing demand prediction each hour through historical usage data and weather data, finding that the Gradient Boosting Machine (GBM) model can yield results outperforming each regression model. Neural networks have been complemented to predict bike demand. Lin et al. [31] proposed a novel neural network labeled the Graph Convolutional Neural Network with a data-driven graph filter (GCNN-DDGF) for predicting the demand for bikes in large-scale bike-sharing systems. Sharma and Sikka [32] compared the linear model consisting of Auto-Regressive Integrated Moving Average (ARIMA) Exponential Smoothing (ETS), and Neural Network Auto Regression (NNAR) model for demand prediction in bikes-sharing systems, finding the ARIMA model producing the highest accuracy performances, compared to the other two models.

Neural Networks are commonly used to estimate functions with numerous unknown inputs. They may learn through supervised learning, unsupervised learning, and reinforcement learning. This method has several benefits, including the ability to be very resilient and to learn directly from observed data, eliminating the need for explicit functions. One of the downsides of this approach is that it requires a huge commitment of processing and storage resources to build big and effective neural networks, as well as a great diversity of training for real-world operations. Xu, C. et al. [33] forecasting the proposed deep learning method to create a dynamic demand forecasting model for free-floating bike-sharing systems. Geographical and temporal analyses were initially used to consider the free-floating bike-sharing movement pattern. The study showed that bike-sharing journeys have an unbalanced geographical and temporal demand. Long short-term memory neural networks (LSTM NN) were created to forecast the production and attractiveness of bike-sharing trips in traffic. According to the validation findings, the generated LSTM NN models have good prediction accuracy in trip productions and attractions for various periods. The ARIMA models were also created to test the LSTM NN's performance. For various periods, the comparative findings revealed that the LSTM NN models offer a prediction accuracy superior to the ARIMA model.

2.4 Rebalancing problem

Previous research has proposed a method for improving the service quality of the bike-sharing systems, which can be classified into two aspects [26]. The first category is qualitative research: a focus on the development of the bike-sharing systems by analyzing their strengths and weaknesses. Guo et al. [34] proposed the bivariate ordered probit (BOP) model to consider factors associated with bike-sharing usage and satisfaction. Moreover, Fishman et al. [28] analyzed factors influencing bike-share predictors to maximize and achieve the goal of quantifying the contribution of walking and cycling. Zhang et al. [35] proposed a second type, viz., quantitative research. This creates a layout of the station and resolves the bike relocation problem, which is like the strategy of Ban and Hyun [36] adopting the novel simulation method with the use of a three-dimensional (3D) approach to improve the truck route as a visual analysis for rebalancing the bike-sharing system. Furthermore, Chen et al.[37] proposed the mathematical programming model to maximize the interval between relocation activity and satisfaction of demands. Ghosh et al. [8] conducted a study to rebalance the bike-sharing system operators. A mixed linear model was used to estimate the influence of the infrastructure and the characteristics of land use. Faghih-Imani et al. [38] developed the binary logit model to identify rebalancing time and used the regression model to predict the number of bikes required for rebalancing. Alvarez-Valdes et al. [39] analyzed the unsatisfied demand and guided redistribution algorithm, Schuijbroek et al.[40] presented a new cluster-first route-second heuristic to consider the account service level and route cost.

However, for the related work of relocation of the bike-sharing solution, there are bike-different strategies [41] that can be classified into two categories. The first category focuses on the user-based approach, where the customers actively balance the bike-sharing. The second category adopts the operator-based approach, with the service provider operator relocating activity and satisfaction of customer needs by using optimal reallocation in the relocation process.

Operator-based strategies are well-known in vehicle routing problems, where, in the relocation process, they balance available bikes with the docking available in the systems to enhance customer satisfaction. Many researchers have demonstrated the vehicle routing problem (VRP) as a Mixed Integer Linear Program (MILP). Angeloudis, Hu and Bell [42] adopted the Travelling Salesman Problem (TSP) as a multiple Travelling Salesman Problem (mTSP) for optimization to minimize total distance while rebalancing bikes at all stations. The TSP approach can result in an exact and approximate algorithm solution [43, 44]. by proposing an objective function, which is to minimize the cost or maximize profit, under constraints of resource limitation in the systems, for instance, the load truck limit capacity for carrying bikes at each station. Likewise, the operator-based strategy enables the service provider operator to dispatch bikes through multiple trucks, in the static and dynamic problem versions. TSP is a mathematical problem in which the shortest path between two points must be discovered (bus stops in this case). Before returning to the starting point, the stated path must travel through each location just once. TSP determines the quickest and most efficient path for a person, given a list of particular destinations. In the discipline of operations research, it is a well-known algorithmic. Delivery cost reduction is fundamentally a Vehicle Routing Problem (VRP), which is an extended variant of the TSP and one of the most researched mathematical optimization problems. It is concerned with determining a set of routes or approaches that may be used to decrease delivery costs. There might be several depots, hundreds of delivery sites, and numerous vehicles in the problem area. VRP, like TSP, is NP-hard to solve; therefore, the number of problems that can be solved optimally through combinatorial optimization or mathematical programming may be restricted. Due to the frequency and magnitude of real-world VRP they must answer, commercial solvers typically utilize heuristics—which are like shortcuts for the human brain, removing a lot of arithmetic or computations for a quick and easy solution. TSP and VRP solvers in the real-world employ route optimization algorithms to discover near-optimal solutions in a fraction of the time, allowing delivery companies to design routes swiftly and effectively. Regarding the approach of the static version, Erdoğan et al.[45] presented an exact algorithm using a branch-and-cut algorithm that utilizes combinatorial Benders' cuts to separate infeasible solutions from the feasible region, to solve the static bicycle rebalancing problem by determining the minimum cost sequence of the stations visited by a single vehicle. Chemla et at.[46] presented the branch-and-cut algorithm by a tabu search for solving the static rebalancing problem as a capacitated single vehicle, which has many pickup and delivery problems. Yanfeng Li et.at. [47] presented the mixed integer programming problem and developed a combined hybrid GA to solve the static bicycle repositioning problem with multiple types of bikes. Mauro Dell et al. [48] proposed the metaheuristic algorithm that implemented a new constructive and a set of local searches to solve the bike-sharing relocation problem. Teobaldo Bulhoes et al.[49] proposed a branch-and-cut algorithm to solve the static bike relocation problem with multiple vehicles and visits and developed iterated local search based on heuristics, implemented for instance from 10 to 100 nodes. In the dynamic version, Leonardo Caggiani et al.[41] proposed a dynamic bike relocation that is a decision support system for a dynamic real-time. Federico et al. in 2018 [50] proposed a framework for a dynamic rebalancing of bike-sharing systems. The results show that a dynamic strategy able to adapt to the fluctuating nature of the network outperforms rebalancing schemes.

User-based strategies are one of the methods researchers used to improve the efficiency in the bike-sharing systems, focusing on customers actively balancing bike-sharing which incentivizes users to participate and encourages them to voluntarily relocate their rented bikes. Such strategies include static pricing and dynamic pricing. This strategy investigates the behavior of customers where the problem started for solving directly point of the problem and attracting customers by increasing the user numbers through policies or schemes. For instance, For instance, Singla et al.[51] presented a crowdsourcing mechanism for dynamic pricing, which enabled the calculation of each station's incentive values and the development of a dynamic incentive system by offering incentive amounts to users and utilizing smartphone applications. Pfrommer et al. [52] proposed using dynamic pricing strategies on real-time price incentives to entice users to return bikes to short-supply bike stations. Zhang et al. [53] presented a method for a user-based bike-sharing system as a dynamic pricing strategy with negative prices for improving the problem of bike imbalance with demand and supply in the system, using the user equilibrium dynamic traffic assignment model that was developed to capture the behavior of traveler's response for route mode selection.

This thesis mainly presented the strategy to be implemented for improving the efficiency of bike-sharing systems in the case of station-based type, which is currently the most widespread. According to the literature review, the existing solution can be improved to offer more efficiency. Nevertheless, the artificial approach is the one solution yielding high accuracy results. Thus, this work presented the suggestion of machine learning algorithms and developed the well-known algorithms for significant quality satisfaction.

2.4.1 Vehicle routing problem

Vehicle routing problem (VRP)is a generic term for a group of situations in which a fleet of vehicles must be calculated based on one or more depots for many cities or clients spread across a large geographic area. It solved the truck dispatching issue, which entails optimizing the fleet of gasoline delivery vehicle routes between a bulk terminal and the many service outlets supplied by the terminal. VRP's major goal is to cut the overall distance traveled, minimize the number of vehicles used, and meet service standards.

2.4.2 Solution for solving VRP

The VRP algorithms used most often are given below in Figure 2.2. When the number of cities is large, no exact solution algorithm can ensure that the best solutions will be found in a fair time. This is mostly due to the problem's NP-hardness [54]. Exact algorithms can only tackle small-scale routing issues and, depending on the magnitude of the problem, might take a long time. They were the first VRP solutions and were recognized to address the problem to the best of their abilities. One of the drawbacks of precise optimization techniques is that they are inefficient in providing a suitable solution in a high-dimensional search space.

Compared to exact methods, heuristics are more appropriate for very large-scale routing issues and require less time (not dependent on problem size) [55]. These heuristics analyze the search space in a restricted way and provide high-quality results in a reasonable time. Classic heuristics are both building and improvement strategies for developing routes, generally one at a time until a full route is built, as well as an attempt to enhance the answer for a more efficient solution.

Classic heuristics are naturally split into constructive and improvement heuristics. Descent heuristics constantly progress from a good solution to a better one in the neighborhood until there is no more improvement feasible.

Meta-heuristics will discover their first answer before moving on to find a more globally optimum solution while enabling non-improving and even infeasible intermediate solutions to be considered. Savings algorithms, route-first cluster second, cluster-first route-second, and insertion heuristics are among the most used construction heuristics. There are two sorts of improvement algorithms that may be used.

In the real world, meta-heuristics are utilized to solve routing and scheduling difficulties in routing issues, for instance, picking the quickest and highest-quality route to a destination. In software testing, the meta-heuristic method is shown in Figure 2.3. That is critical for obtaining the best collection of test data.

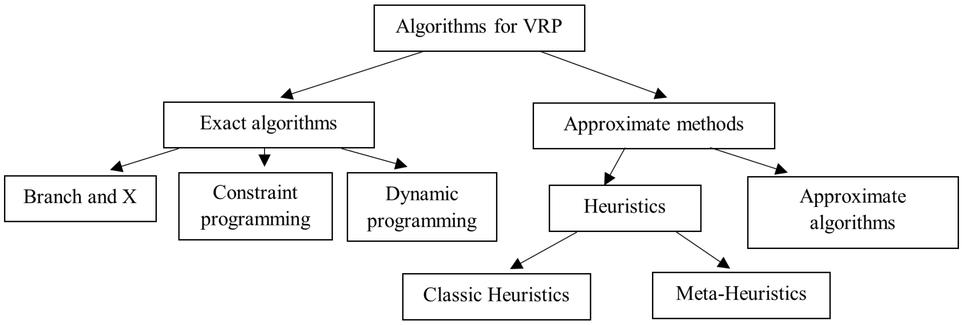


FIGURE 2.2: Classical algorithms for vehicle routing problem.

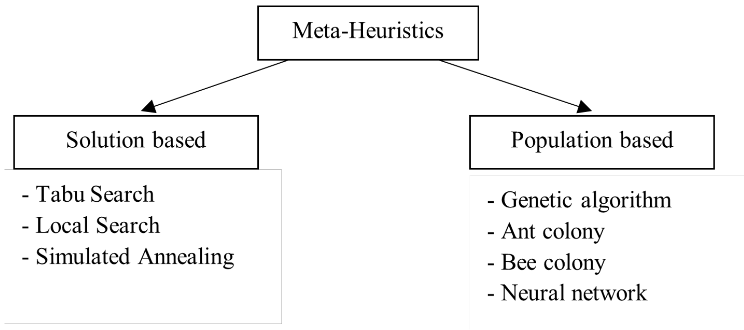


FIGURE 2.3: Meta-Heuristic Algorithms.

Also at present An intelligent approach has been applied to solve the routing problem that can improve the routing efficiency from the classic method. Especially in the matter of computational speed in case of a large number of problems.

2.5 Facility location

The objective of choosing a probable location is to try to find a location that creates the most suitable relationship between them. Capital combined with income, in other words, is the source of confinement. The capital is short-term, which means that it is effective. Internal affairs mites that will last the fastest.

The facility location problem aims to determine the least-cost allocation of customer demand to product distribution. Logistics and supply chain research started to focus attention on the problem of facility location in 1974 [56].

Determining the location of the warehouse business is necessary to have sufficient space. Due to clutching determining or choosing, the location of the warehouse must be efficient, available in both macro approaches, i.e., analysis to consider choosing the location of large products at the national level and the micro approaches considering specific locations from a region or country selected from the guidelines. Optimal location tools of location-allocation models have been created, which might be very useful for positioning bike stations in relation to the distribution of the prospective demand. This approach entails determining where and how many facilities of a particular kind should be situated to accomplish a pre-determined goal while meeting demand from a set number of centers [57]. Macroeconomic Location has established several macro concepts, proposing a guideline of on the strategy of choosing a location. There are 3 types of inventories, which are defined as follows.

1) Location strategy near the market (Market-positioned Strategy) involves setting up a warehouse as close to the final customer as possible, enabling efficient customer service. There are many important factors in choosing a location near the customer, such as shipping costs, waiting time for ordering, product safety, the volume of orders, the adequacy of local delivery vehicles, and the service the customer needs.

2) A strategic location near the production site (Production-positioned Strategy) places the location of the warehouse as close to the source of raw materials or the factory as possible. This will result in a lower level of service to customers than the first one but will save shipping costs. There are several important factors in choosing this type of location; for instance, whether the raw material is easily spoilable by nature. The number of ingredients is the product ingredient.

3) Strategic location between customers and production sites (Intermediately Positioned Strategy) is a strategy that requires the placement of products in between the production plant and the market. This type of location provides lower customer service than the first strategy, but higher than the second. This type of location is suitable for businesses that have a high demand to serve their customers and often have many manufacturing plants.

Different strategies have been suggested for the best placement of services and implemented in a variety of sectors from the earliest studies of location-allocation.

C. L. Stowers and U. S. Palekar [58] offer another example of precise algorithms for location routing issues. Their model is deterministic, with just one uncapitated facility and one uncapitated vehicle. In contrast to prior research, they used nonlinear programming approaches to tackle location routing issues. Their research provides a methodology for placing noxious facilities such as hazardous waste repositories, dump sites, and chemical incinerators in the most efficient way possible. The model varies from prior models in that it solves two concerns simultaneously: 1) The place is not limited to a known set of possible locations. 2) In addition to transportation hazards, the danger presented by the site's location is considered.

Although precise solution algorithms are beneficial in terms of understanding the problem's complexity, they can produce efficient solutions only for medium-sized issues. When time restrictions and route distance limits are imposed, the problems become considerably more difficult to solve. Approximate solution algorithms give near-to optimal answers in a short time for large-scale issues. Heuristics for location routing issues have been studied extensively for this purpose. Srivastava [59] examined the performance of three approximation solution approaches in terms of the optimal solution of location routing issues and the sequential solution of the classical location-allocation and vehicle routing problems. The first heuristic assumes that all facilities are open at the start and calculates the facility to be closed using approximate routing costs for open facilities. To approximate the routing costs, a modified version of Clark and Wright's [60] savings algorithm is employed for the multiple depot scenario.

The K-means algorithm has been extensively used because of its simplicity and because the performance of its clustering is within acceptable limits. K-means is used for finding the number of clusters from data that does not appear class or does not have a label, which is called unsupervised learning or learning without teaching used for unlabeled data. The main function of K-means is statistical clustering in the context of two or more data groups. Han et al. [61] used K-means to build a cluster location privacy for the Internet of Things (IOT). Razi [62] adopted the K-means algorithm for the clustering of maintenance stations and presented a new approach for selecting a portfolio of locations to solve the facility location problem. Luan et al. [63] used K-means to cluster data for the identification of the value anonymously to improve the anonymous accuracy position.

The weighted K-means (WK-means) clustering algorithm was proposed by Huang et al. [64], who modified the K-means method to consider the relative importance of attributes in exposing the cluster structure of a dataset within the cost function of a k-means procedure using feature weighting. It can be obtained by modifying the cost function of k-means. Chakraborty and Das [65] established the WK means-types algorithms for clustering; in general distance function, WK-means outperforms k-means. Therefore, the current study adopted a WK-means algorithm for optimizing depot location in the context of bike-sharing rebalancing.

2.6 Problem description

The main issue of the bike-sharing relocation problem for a network is the system used for relocation. However, in this work, the bike-sharing relocating problem is defined as follows. The network of the bike-sharing system consists of a depot for all trucks to start from and return to after visiting ‘n’ bike-sharing stations, and ‘v’ trucks are used for bike relocation. Each truck is assumed to be of the same capacity. The number of bikes for rebalancing is known before the start of the repositioning process. Further, the demand for the number of bikes needed for pick-up and drop-off at each bike station for relocation can be reduced or increased depending on the demand at each station during the relocation process.

Demand forecasting is the operation of predicting the future to offer a realistic picture of future products or sales for preparing the budget, tools, staff etc. The demand patterns can vary and evolve if there are no spatial restrictions by docking stations. For the target to reach further, the main approaches consider only booking data, with no consideration to invisible demand, i.e., potential unfulfilled demand. Concerning the level of demand at each station, El-Assi et al. [66] identified the factors influence bike-sharing demand as land use and weather variable with significant correlation with bike-sharing trip attraction. For the weather variable, temperature, wind, and humidity were used for the investigation. Thus, we concern ourselves with the factor that affects bike-sharing demand, similar to the previous work. This work considers the factor in Chapter 4, as the weather variable affects the behavior of bike-sharing demand.

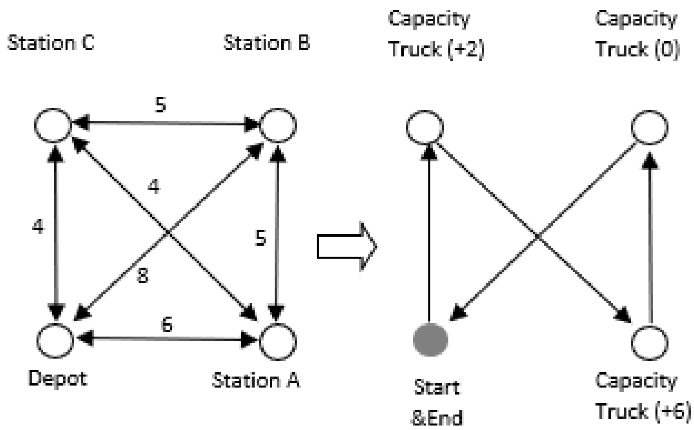


FIGURE 2.4: The tour graph example (left) and feasible tour graph (right).

The optimal routes determine all operating trucks and the number of pick-ups and drop-off bikes at each station to minimize the distance traveled by the operating trucks. The problem is considered on a complete directed graph $G = (N, A)$, where $N = \{1, \ldots, n\}$ is the set of bike-sharing stations (nodes), $A = \{i, j \epsilon V, i \neq j\}$ is the set of arcs, and the node number 0 represents the depot. The demand for the pick-up and drop-off of bikes is assumed to be known and can be estimated from historical usage records. The truck is needed to relocate bikes from stations with a surplus of bikes to those with a shortage, to allow more customers to use them. The truck is considered only a truck to relocate; it starts from the depot and returns to the depot after visiting some or all the bike-sharing stations, which is a selective pickup and drop-off (delivery) problem [67]. An example of a tour graph transportation network is shown in Figure 2.4.

Though many researchers have proposed solutions for the rebalancing problem, due to the dynamically changing environments, the systems need adaptability to perform their tasks effectively. Therefore, the rebalancing problem should adapt the intelligent agents to provide information and make decisions for making the routes the shortest or minimizing the total cost. This work proposes In Chapter 5-6 the intelligence such as ABC, and machine learning to an application for solving the rebalancing problem.

There are several methods such as truck routing or pricing strategies that can solve bike unbalancing problems. However, there is only one solution for the virtual bike balancing problem: the simulation method that is widely used in various fields. The simulator assesses the performance of a bike-sharing system before use or for those analyzing some aspects that are particularly difficult to measure; others can be analyzed from simulations. The simulator can also be used as a productivity tool in the planning process of the bike-sharing system. It can also solve complex problems. This research proposes to solve the problem of bike rebalancing using a simulation method that can meet customer needs and obtain the most profitable authorized provider. It is carried out to study the flow of various activities and analyze the appropriate model before it is put into practice. The simulation process consists of 3 steps, as follows:

Step 1: Define the problem or system of interest and create a mathematical model of the system. The simulator must have a good understanding of the system to create a model that must adhere to its definition. If the simulator does not own the problem but is no longer assigned as the problem solver, the owner of the problem must participate in the modeling and closely supervise the model creation.

Step 2: Simulate the system on the computer. The mathematical model of the system needs to be simulated on a computer. It will use mathematical knowledge, such as statistical probability, along with information technology, to create a computer model that can be efficiently processed (run). The advantage of a powerful computer model is that it allows the simulator to repeat an increased amount of processing in the same amount of time. Statistically, having a large amount of data permits a more accurate analysis of results.

Step 3: Analysis of the results. Since the simulation is random, the results obtained from the simulation are also random. Therefore, statistical techniques are needed for the correct interpretation of the results.

Few of research used the simulation modelling for solving rebalancing bike-sharing problem. Thus, we present the simulation method that efficiency tools to demonstrate the results that can be represented in the real situation and complex problem in Chapter 7.

The bike relocation is required for pick-up and drop-off, from a full station to an insufficient station. A depot can keep the bikes for repairing and relocating the bikes at each station. The trucks for relocation can visit the station only once and return to the depot after visiting each station under the constraint, viz., capacitated truck, and number of bikes required for rebalancing.

One of the factors influencing the operating cost is the location of the depot. The depot location and the number of depots directly affect transportation costs during operator-operated relocation of bikes in each station in the system. This helps the systems respond quickly to relocate the bike demand at each station. In this research, the method of analyzing the appropriate point has been presented. in order to be able to reduce the distance traveled. That is the cost of implementing the management of the bicycle system. That can make the bike-sharing system able to relocate the bikes that can meet the needs of customers as well.

Chapter 3

Overview of System Development

These existing systems may be found in both docking station-based and free-floating bike-sharing systems. For this study, focus on station-based bike-sharing systems, as this is still the case for most bike-sharing systems across the world. The great majority and many bike-sharing users commute during weekday morning peak hours. The number of users leaving residential areas is significant, perhaps resulting in a shortage of available bikes in certain regions [68]. Consequently, service dependability and user happiness suffer. For successful bike-sharing system management and fleeting, accurate, and up-to-date estimates of demands across the city over the day are critical. The bike-sharing system technology is continuously evolving to make management and user pleasure easier. The management and operation of these systems are not simple, as they necessitate guaranteeing that the quantity of bikes in the system and their distribution among the stations constantly meet the demand of the users. The operator required to move bikes from congested stations and redistribute them evenly among the stations with a lack of bikes accounts for most of the expense of operating present bike-sharing systems [69]. Those rebalancing procedures are generally carried out many times each day by specialized teams of field operators who use compact vehicles or carrier bikes. This work focuses on a framework to optimize the bike-sharing network balancing. Forecasts of bike utilization are used in the proposed system. Based on such estimates, a system has been created to forecast the number of bikes available at various stations in the future and to optimize the routing necessary to rebalance the network.

3.1 Demand

There are various processes in analyzing bike-sharing data. The initial stage is to gather the necessary information based on the trip creation and distribution model. The data from Citi bike are cleaned up according to the requirements based on the elements and variables that influence the trip. This pertains to the categorization of registered journeys based on a variety of variables, including date time, weekday, holiday, temperature, humidity, wind speed, and weather (as transparent, rain, and snow) type. After all the sets have been completed, when all the data is in, the demand analysis will be processed as shown in Figure 3.2 below.

Since an accurate forecast can have a big influence on the system's quality, the quantity and scheduling of these rushes may be used to alert users that the station they are biking to is filled-up, and to prevent depletion and saturation. It may also give system operators useful information that allows them to alter the system to completely prevent this rush.

3.2 Relocating bike-sharing

After using an intelligent approach to anticipate the future condition of the stations, an algorithm is required to show the operator which stations need to be rebalanced and how to operate them. The key distinctions between user-based and operator-based relocation strategies. Although some user stimulation will be used to decrease the costs of rebalancing the bikes, this project will employ an operator-based relocation approach owing to the dependability factor (precise description of past and future balancing) and the ease of integration of electric vehicles in the case of depots.

Imbalances lead to unfavorable situations, such as when one station has no bikes while another is full. Users are inconvenienced because of this scenario, as they are unable to utilize the system in some situations owing to a shortage of available bikes at the origin station, and in others must return the bike to a location

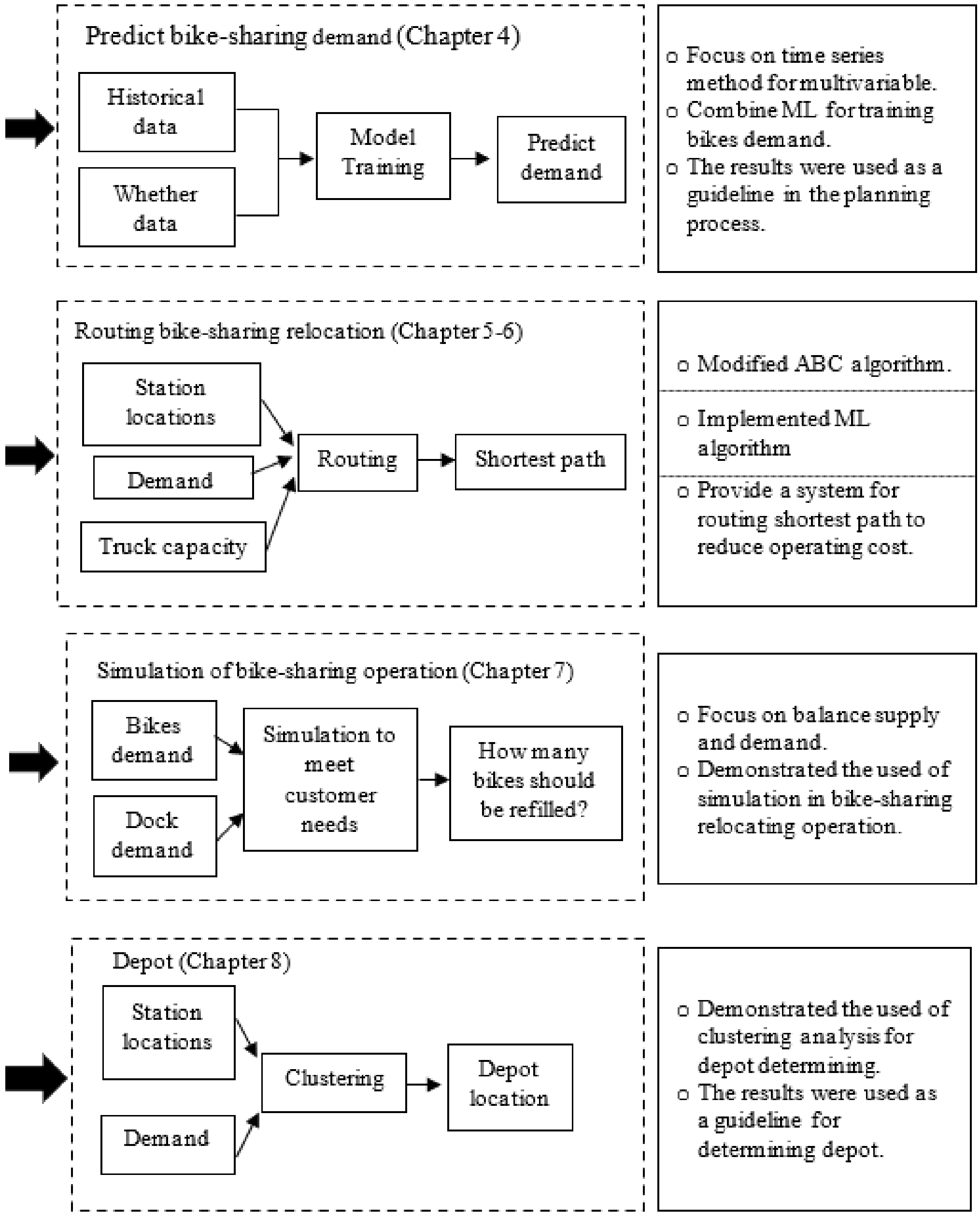


FIGURE 3.1: Overview of the systems in this study

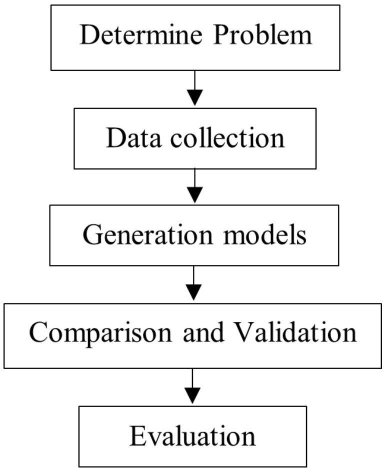


FIGURE 3.2: Workflow of demand predicting

distant from their planned destination station. Relocation operations are commonly acknowledged as a way to moderate these problems, and most systems feature vehicles that artificial relocation bikes among stations. During peak hours, the bikes are refilled from the full to the empty stations using tucks to redistribute the bikes among the stations in the system.

The best solution to the imbalance problem depends not only on the rebalancing strategy. It can be used for example relating to routes, and the number of rebalancing trucks, and on the bike-sharing system's strategic design such as fleet size, and station. While the imbalance problem will inevitably worsen as the number of stations increases, it can be mitigated by oversizing the bike fleet. Different linked trade-offs might be identified [70], implying the necessity for a holistic strategy that includes both strategic and operational levels. Depending on demand patterns and the geometry of the service zone, specific solutions may differ. Despite this, many introductory studies tend to provide only broad suggestions based on prior experience and trial and error. This can result in unexpected system performance, which might finally lead to failure.

When a firm launches a station-based bike-sharing system, it must decide on the location and number of station stands, the number of bikes in the system, the number of vehicles for re-balancing operations, and the shifts of the operators, as well as the pricing and payment method for the service. When a service is in place, it is frequently essential to change it, and in many situations, to extend it,

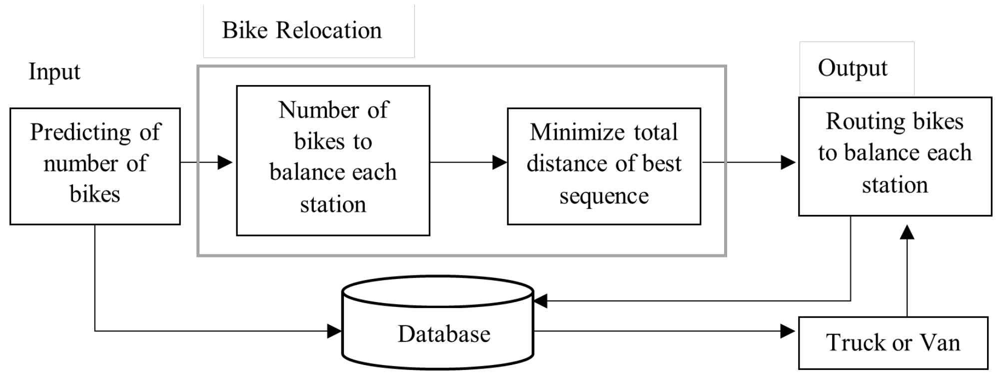


FIGURE 3.3: The framework of bike relocation systems

necessitating the revision of numerous choices. An overview of the strategic, tactical, and operational planning issues that arise in bike-sharing systems.

Trucks or vans are employed to rebalance the bikes among the stations, moving bikes from full stations to low-level bike stations each day. The vehicles begin and conclude their trips at a depot, which is also where spare bikes and bikes in need of maintenance are often stored. The company's information system has real-time access to the system's status, which includes the number of bikes and unoccupied stands at each station. This information is also made available to potential consumers over the internet. The information system also knows where each vehicle is when it arrives at and departs from a station, and how many bikes are loaded and unloaded.

The engineering portion of the internship is represented in this module. The first step is to obtain data from open data sources (Citi bike); the most troublesome stations are identified, followed by the stations that will be vacant or full in the future. Using the currently available distribution trucks, the selected problematic stations and the proposed number of bikes being transferred at each station are computed. This framework as shown in Figure 3.3. The framework is based on the VRP, which employs the most advanced algorithms to determine the optimal order for bike-sharing relocation.

3.3 Simulation of operating bike-sharing relocation

The capacity to simulate the design and operation of bike-sharing systems in advance would alert and avoid such problems. To imitate the performance of bike-sharing systems, this work presented the simulation method to solve the problem. This entails establishing a set of assumptions and techniques that can accurately mimic the real-world problem.

In the literature, the issue of replicating a bike-sharing system has been tackled in several ways. Romero et al. [71] proposed a modeling program that considers both personal vehicles and communal bikes. The tool is used to optimize the location of bike-sharing stations to increase the transportation system's efficiency and sustainability. A simulation optimization technique for large-scale bike-sharing systems was presented by Jian et al. [72].

The issue aims to minimize the predicted number of unsatisfied users, i.e., users who wish to rent a bike while a station is empty or return a bike when a station is full, given a set of stations, the size of the bike fleet, and the number of stands available, by determining the capacity of each station and the number of bikes available at the start of the day. The proposed techniques are validated using data from New York City's Citi Bike system. The following contributions are particularly important to the work described here. An agent-based simulation framework was presented by Soriguera et al. [70]. The simulator was used to evaluate both periodic and continuous rebalancing methods. Fern´andez et al. [73] also proposed an agent-based simulation environment for a station-based bike-sharing system. The ability to make a reservation is part of the user's behavior.

The policies for repositioning are straightforward. To minimize overcrowding or under-crowding, the simulation model dynamically redistributes bikes among stations. We will assume that there is a command center in charge of all the repositioning trucks. When a station runs out of racks for returning bikes, it will send a signal to the center or depot. If the depot has a vehicle available, it will dispatch it to the station to transport bikes back to the depot. Similarly, if a station runs out of bikes to rent, the station will send a signal to the depot. A vehicle will be dispatched by the depot. A van will be dispatched from the depot to transport bikes to the station in Figure 3.4. We assume that after completing a single job, each vehicle will

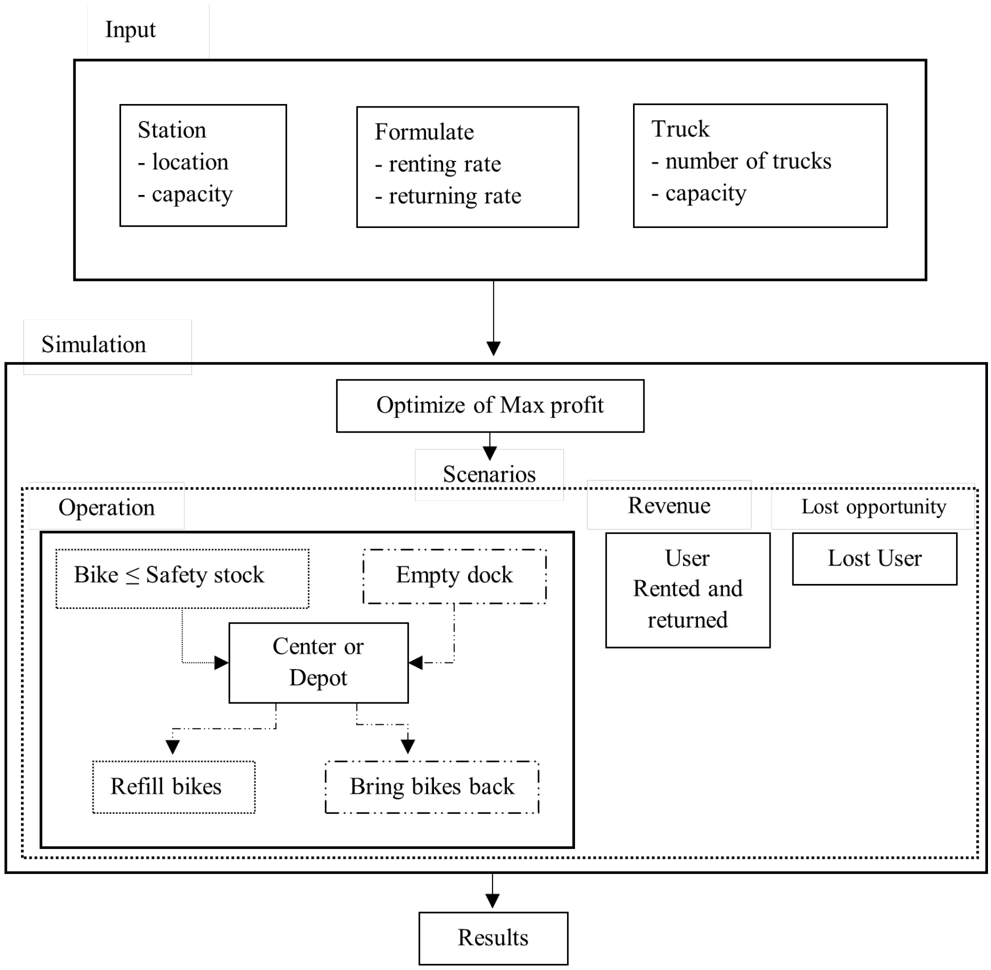


FIGURE 3.4: Framework of the simulation

return to the depot. If vehicles are unavailable, the "first come, first served" policy will apply, based on whatever event occurs first.

3.4 Depot

The objective of this work is to determine the number and locations of depots, the allocation of customers to these depots, the number of vehicles departing from each depot, and their distribution routes to the associated customers, all while keeping the system-wide expenses to a minimum. This work analyzed a distribution scenario in which transportation is provided between depots and the bike-sharing station supplied. Each station will be assigned to a specific depot and will be served by

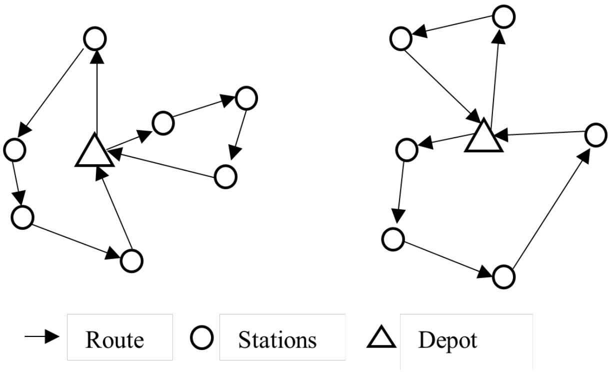


FIGURE 3.5: Distribution network

vehicles that depart that location in Figure 3.5. Vehicle capacities are specified, and the overall demand of users supplied by a vehicle cannot exceed that capacity. The coordinates of the places are given in a two-dimensional format. The number of trucks in a depot is a measure of its capacity. The decision variables determine the total bikes and docking variables are determined at the depot, as well as the total demand for that depot.

3.5 Conclusion

This work proposed a new aspect in bike-sharing systems for the relocation. It is important that the strategy should begin with the prediction of demand, and proceed to planning, with due consideration for the operating cost. The future value can lead the operator to plan the transportation cost. The new model, by combining machine learning as a LSTM and GRU for a time series method multivariate input consisting of the historical data and environment data, predicts bike-sharing demand, and also a method for selecting input variables was presented by a statistical method, which is an important part in leading to accurate forecasts. A modified ABC algorithm is proposed for routing of relocation of bike-sharing, which gives a shorter distance than the original one. Artificial intelligence as reinforcement learning is also implemented, which is the efficient method for routing to relocation bike-sharing. The simulation model for operating bike relocation focuses on balancing supply and demand for maximizing profit. The important factor affecting the operation of relocation is that as the depot location is the point for considering the sustainable location that impacts to, it also impacted the total distance during operator relocation bike in systems. All three factors affect the operating costs of the organization. It is also a guideline for implementing routing operations to move bikes to each station to respond to the needs of customers.

Chapter 4

Demand Analysis

4.1 Overview

The travel demand is importance component for planning and management in the transportation section. In the order to forecast bike rental demand its useful information for planning strategy to success bike-sharing system. To know the pattern usage for prepare and planning to balance the demand and supply. The prediction models of the reach (obtain) forecast demand can be categorized into two functions: (1) using the knowledge or experience of the expertise (2) using historical data consist of the traditional method and intelligent method

The bike-sharing service system is a service that allows a customer to rent a bike from a bike-sharing station and then return it to another bike-sharing station in a short time after they reach their destination. Thus, the impact of the bikes distribution system based on the frequency of bike usage needs to be assessed. The bike-sharing system operator needs to predict the demand to accurately know how many bikes are needed in every station so as to assist the planner in the management process of the bike-sharing stations.

Nevertheless, the technique was not appropriate for any problems, which is undesirable for operations. According to related work, Sharma and Sikka proposed that the ARIMA model produces high accuracy performance. On the other hand, Xu, Chengcheng et al. [74] demonstrated that the LSTM model provides better than

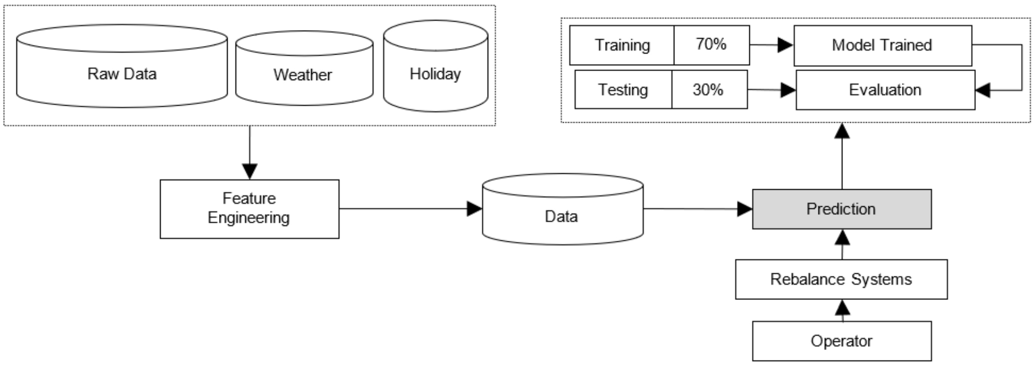


FIGURE 4.1: Overview of the framework for predicting.

the ARIMA model; additionally, Cho et al. [75] explain that the GRU is pretty similar to the LSTM network. On this note, the research question remains about which strategy is superior for predicting bike-sharing demand. Therefore, our research focused on the hourly bike-sharing demand to forecast future demand.

A comprehensive framework built to aid bike-sharing firms in managing their field operations [76]. Figure 4.1. describes an overview of the effective demand forecasting framework in a bike-sharing system. Since each of the monitors of the prediction system produces predictions on the operator's request, the rebalancing system creates a list of rebalancing tasks for operation. The rebalancing system is based on the forecasted information data. The operators might see the rebalancing missions that need to be completed as well as a list that must be completed according to the rebalancing algorithm's shortest paths.

This work proposes an efficient and accurate model for predicting the bike-sharing service usage using various features of a machine learning algorithm. by comparing the exiting techniques for the sequential data predicting of artificial intelligence for time series data and analysis. Then, we considered the use of the multivariate model with a recurrent neural network (RNN), a long short-term memory (LSTM), and a gated recurrent unit (GRU). In addition, we considered combining the LSTM and GRU methods together to improve the model's effectiveness and accuracy. The results showed that all the RNNs, including the LSTM, GRU, and the model combining the LSTM and GRU, are able to achieve high performance using the mean square mean absolute, mean squared error, and root mean square error.

TABLE 4.1: The dataset and used variables

Variable	Describe
Date Time	Indicates date including the date, month, year, and time information
Weekday	Indicates variables it is a weekday or weekend
Holiday	Indicates variable that government holiday isn't it
Season	Divided into four seasons: spring, summer, fall, and winter
Weather	Divided into four types: transparent, mist and cloudy, rain, and snow
Temperature	Temperature in Celsius
Humidity	The values of humidity
Wind speed	The values of wind speed
Count	The number of bikes used

However, the mixed LSTM–GRU model accurately predicted the demand in this case.

4.2 Methodology

4.2.1 Pilot experiment

First, pilot experiment of the proposed model as part of the research process, which is effort to learn from successes and failures. The data was downloaded the bike-sharing usage downloaded the bike-sharing usage data from January 1, 2011, to December 31, 2012, from UCI Machine Learning Repository official website and acquired the weather data from the UCI bike-sharing dataset [77]. Which is the dataset that has already been prepared. The dataset and variables are shown below in Table 4.1.

From the dataset, created a coefficient correlation diagram using Scikit-learn and Keras as shown in Figure 4.2 The, and the results showed the correlation of each variant (feature), where each feature depends on the temperature, season, humidity, wind speed, weekday, holiday, and count. This experiment, the data to predict bike rentals is divided into two sets: training and testing. The training set is used to evaluate the performance of the model, while the testing set is used to train the models. Data is used for testing in 70 percent (512 observations) and 30 percent (219 observations) of cases, respectively.

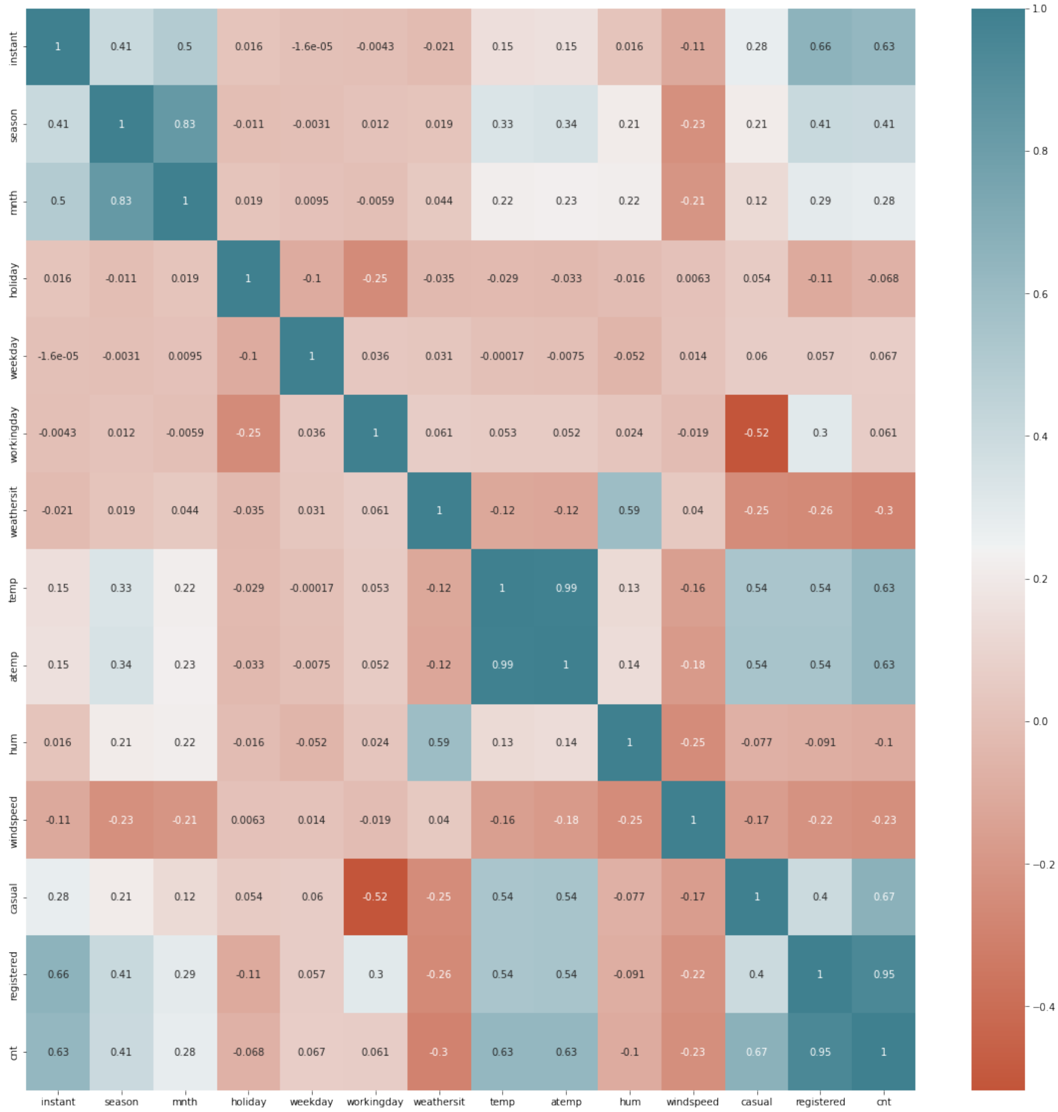


FIGURE 4.2: The correlation coefficient diagram.

Table 4.2 presents the performance of the models for pilot experiment. As can be seen the experimental results showed that the mixed LSTM–GRU model had the least error values, as its MAE, MSE, and RMSE are 0.0420, 0.0040, and 0.0630. This is the outcome that may be used as a reference point in the following experiment.

TABLE 4.2: Predicted errors of the bike-sharing usage.

Model	MAE	MSE	RMSE
RNN	0.0744	0.0055	0.0505
LSTM	0.051	0.051	0.0786
GRU	0.0494	0.0053	0.0735
LSTM–GRU	0.0420*	0.0040*	0.0630*
GRU–LSTM	0.047	0.005	0.0708

4.2.2 Data preprocessing

This work used data from a BSS in Jersey City of Citi bike. Data between January 1st, 2020 and December 31st, 2020 contains 333,802 records (see the Figure 2), were used in the following prediction. In this model, only used the numbers of rents and returns, information on the stations, and numbers of users, and the dataset and variables are shown below in Table 3. According to the raw data, it can be divided by at each hour is computed using the data and collected the dataset, which was ID instead of the station name and the number of users per hour added together. The date/time variable is used to obtain the weekend or weekday, and the day of the week (Official) holidays[ref]. The next stage is to extract some more characteristics from the date/time variable in order to improve the efficiency of the machine learning algorithms. This work also considered other various factors, such as historical weather data and days of the week. Then, Feature engineering entails the data's transformation that has been converted incorrectly might produce false findings. Afterward, this work converted the data into a training dataset and tested it with a ratio of 70:30, respectively (the dimensions of training and testing set is shown in Table 4), where one sequence contained 24 consecutive hourly data inputs. Then, the data were transformed using MinMax Scaler [78] by scaling each feature to a given range. This scaler works in scales across input variables that may have different units, which can be used to create the required model for predicting the future demand for bike-sharing usages.

From the dataset, we created a coefficient correlation diagram using Scikit-learn and Keras as shown in Figure 4.2, and the results showed the correlation of each variant (feature), where each feature depends on the temperature, season, humidity, wind speed, weekday, holiday, and count.

Table 4.3: Jersey City of Citi bike dataset used variables and descriptions.

Features	Acronym	Type	Measurement
Date	Date	Month-day-Year	1/1/2020 to 12/31/2020
Season	Season	Categorical	Spring =1, summer=2, fall=3, and winter=4
Hour	Hour	Continuous	0,1,2,...,23
Holiday	Holiday	Categorical	Holiday=1, workday =0
Weekday	Weekday	Categorical	Sunday=0, Monday=1,...,Saturday=6
Weather	Weather	Categorical	Transparent =1, Mist and cloudy = 2, Rain = 3, and Snow =4
Temperature	Temp	Continuous	Fahrenheit
Humidity	Hum	Continuous	%
Windspeed	Wind	Continuous	mph
Used bikes count	Count	Continuous	0,1,2,...,338

Table 4.4: Training and Testing dataset

Dataset	Measurement values
Training set	6149 datasets
Testing set	2635 datasets

Table 4.5: The table provides the R, R^2, adjusted R^2, and the standard error of the estimate, determined for a regression model fits the data

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.436^a	0.19	0.19	38.24	0.19	2058.14	1.00	8782	0
2	0.542^b	0.29	0.29	35.70	0.10	1295.94	1.00	8781	0
3	0.588^c	0.35	0.35	34.38	0.05	686.30	1.00	8780	0
4	0.619^d	0.38	0.38	33.37	0.04	542.17	1.00	8779	0
5	0.621^e	0.39	0.39	33.30	0.00	39.98	1.00	8778	0
6	0.623^f	0.39	0.39	33.24	0.00	28.68	1.00	8777	0

a. Predictors: (Constant), temperature
b. Predictors: (Constant), temperature, hour
c. Predictors: (Constant), temperature, hour, humidity
d. Predictors: (Constant), temperature, hour, humidity, season
e. Predictors: (Constant), temperature, hour, humidity, season, weekday
f. Predictors: (Constant), temperature, hour, humidity, season, weekday, weather

Table 4.6: Stepwise Multiple Regression Coefficient analysis of factors affecting bike use

Variables	**B**	**Std. Error**	**Beta**	**t**	**Sig**
(Constant)	-53.785	3.138		-17.138	0
temperature	1.346	0.026	0.534	51.117	0
hour	1.547	0.054	0.252	28.874	0
humidity	-0.456	0.019	-0.219	-24.223	0
season	11.463	0.489	0.242	23.444	0
weekday	1.139	0.178	0.054	6.39	0
weather	-4.024	0.752	-0.047	-5.355	0

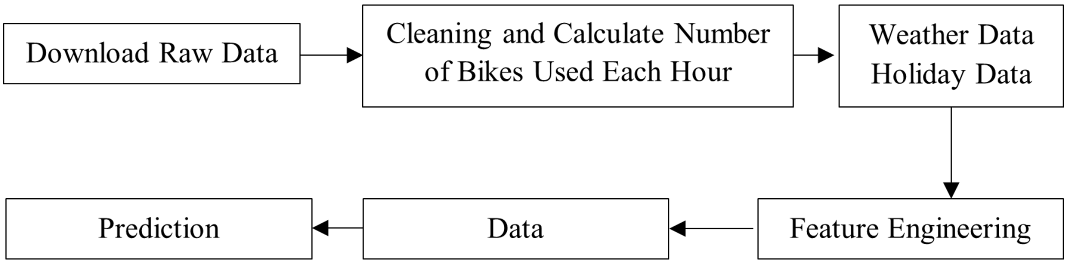
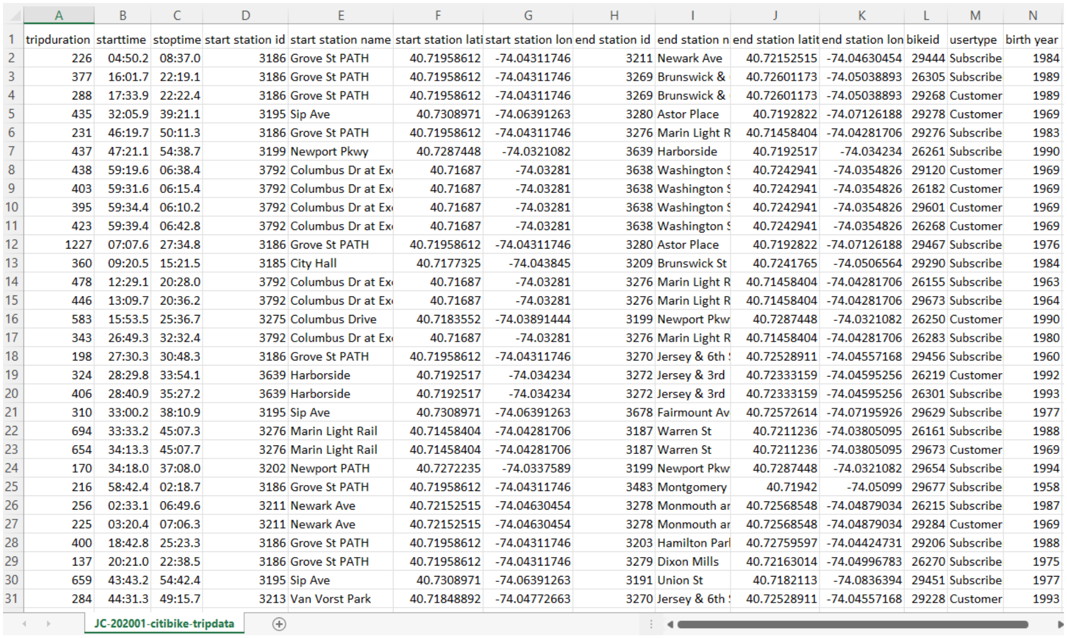


FIGURE 4.3: Dataset processing.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	tripduration	starttime	stoptime	start station id	start station name	start station lati	start station lon	end station id	end station n	end station latit	end station lon	bikeid	usertype	birth year
2	226	04:50.2	08:37.0	3186	Grove St PATH	40.71958612	-74.04311746	3211	Newark Ave	40.72152515	-74.04630454	29444	Subscribe	1984
3	377	16:01.7	22:19.1	3186	Grove St PATH	40.71958612	-74.04311746	3269	Brunswick &	40.72601173	-74.05038893	26305	Subscribe	1989
4	288	17:33.9	22:22.4	3186	Grove St PATH	40.71958612	-74.04311746	3269	Brunswick &	40.72601173	-74.05038893	29268	Customer	1989
5	435	32:05.9	39:21.1	3195	Sip Ave	40.7308971	-74.06391263	3280	Astor Place	40.7192822	-74.07126188	29278	Customer	1969
6	231	46:19.7	50:11.3	3186	Grove St PATH	40.71958612	-74.04311746	3276	Marin Light R	40.71458404	-74.04281706	29276	Subscribe	1983
7	437	47:21.1	54:38.7	3199	Newport Pkwy	40.7287448	-74.0321082	3639	Harborside	40.7192517	-74.034234	26261	Subscribe	1990
8	438	59:19.6	06:38.4	3792	Columbus Dr at Ex	40.71687	-74.03281	3638	Washington S	40.7242941	-74.0354826	29120	Customer	1969
9	403	59:31.6	06:15.4	3792	Columbus Dr at Ex	40.71687	-74.03281	3638	Washington S	40.7242941	-74.0354826	26182	Customer	1969
10	395	59:34.4	06:10.2	3792	Columbus Dr at Ex	40.71687	-74.03281	3638	Washington S	40.7242941	-74.0354826	29601	Customer	1969
11	423	59:39.4	06:42.8	3792	Columbus Dr at Ex	40.71687	-74.03281	3638	Washington S	40.7242941	-74.0354826	26268	Customer	1969
12	1227	07:07.6	27:34.8	3186	Grove St PATH	40.71958612	-74.04311746	3280	Astor Place	40.7192822	-74.07126188	29467	Subscribe	1976
13	360	09:20.5	15:21.5	3185	City Hall	40.7177325	-74.043845	3209	Brunswick St	40.7241765	-74.0506564	29290	Subscribe	1984
14	478	12:29.1	20:28.0	3792	Columbus Dr at Ex	40.71687	-74.03281	3276	Marin Light R	40.71458404	-74.04281706	26155	Subscribe	1963
15	446	13:09.7	20:36.2	3792	Columbus Dr at Ex	40.71687	-74.03281	3276	Marin Light R	40.71458404	-74.04281706	29673	Subscribe	1964
16	583	15:53.5	25:36.7	3275	Columbus Drive	40.7183552	-74.03891444	3199	Newport Pkw	40.7287448	-74.0321082	26250	Customer	1990
17	343	26:49.3	32:32.4	3792	Columbus Dr at Ex	40.71687	-74.03281	3276	Marin Light R	40.71458404	-74.04281706	26283	Subscribe	1980
18	198	27:30.3	30:48.3	3186	Grove St PATH	40.71958612	-74.04311746	3270	Jersey & 6th	40.72528911	-74.04557168	29456	Subscribe	1960
19	324	28:29.8	33:54.1	3639	Harborside	40.7192517	-74.034234	3272	Jersey & 3rd	40.72333159	-74.04595256	26219	Customer	1992
20	406	28:40.9	35:27.2	3639	Harborside	40.7192517	-74.034234	3272	Jersey & 3rd	40.72333159	-74.04595256	26301	Subscribe	1993
21	310	33:00.2	38:10.9	3195	Sip Ave	40.7308971	-74.06391263	3678	Fairmount Av	40.72572614	-74.07195926	29629	Subscribe	1977
22	694	33:33.2	45:07.3	3276	Marin Light Rail	40.71458404	-74.04281706	3187	Warren St	40.7211236	-74.03805095	26161	Subscribe	1988
23	654	34:13.3	45:07.7	3276	Marin Light Rail	40.71458404	-74.04281706	3187	Warren St	40.7211236	-74.03805095	29673	Customer	1969
24	170	34:18.0	37:08.0	3202	Newport PATH	40.7272235	-74.0337589	3199	Newport Pkw	40.7287448	-74.0321082	29654	Subscribe	1994
25	216	58:42.4	02:18.7	3186	Grove St PATH	40.71958612	-74.04311746	3483	Montgomery	40.71942	-74.05099	29677	Subscribe	1958
26	256	02:33.1	06:49.6	3211	Newark Ave	40.72152515	-74.04630454	3278	Monmouth ar	40.72568548	-74.04879034	26215	Subscribe	1987
27	225	03:20.4	07:06.3	3211	Newark Ave	40.72152515	-74.04630454	3278	Monmouth ar	40.72568548	-74.04879034	29284	Customer	1969
28	400	18:42.8	25:23.3	3186	Grove St PATH	40.71958612	-74.04311746	3203	Hamilton Parl	40.72759597	-74.04424731	29206	Subscribe	1988
29	137	20:21.0	22:38.5	3186	Grove St PATH	40.71958612	-74.04311746	3279	Dixon Mills	40.72163014	-74.04996783	26270	Subscribe	1975
30	659	43:43.2	54:42.4	3195	Sip Ave	40.7308971	-74.06391263	3191	Union St	40.7182113	-74.0836394	29451	Subscribe	1977
31	284	44:31.3	49:15.7	3213	Van Vorst Park	40.71848892	-74.04772663	3270	Jersey & 6th	40.72528911	-74.04557168	29228	Customer	1993

JC-202001-citibike-tripdata

FIGURE 4.4: The datasets are downloaded from Jersey of Citi bike.

Since this research, many variables were considered to use in experiment. Some factors may affect the forecasting process. Therefore, the factors affecting the dependent variables were analyzed. Regression analysis is an analytical statistic used to study the causal relationships resulting in the results of the data analysis used to compare the direction of the influence of each parent variable on the dependent variable. There is a method for selecting variables in the equation so that the equation can predict the criterion variables as much as possible. This research used Stepwise Multiple Regression Analysis to evaluate the input variables used in demand forecasting had significant effects on the dependent variables.

	A	B	C	D	E	F	G	H	I	J	K
1	dteday	season	hr	holiday	weekday	weather	temp	hum	windspeed	cnt	
2	1/1/2020	4	0	1	3	2	40	63	14	10	
3	1/1/2020	4	1	1	3	2	40	67	14	14	
4	1/1/2020	4	2	1	3	2	39	65	14	9	
5	1/1/2020	4	3	1	3	2	39	60	10	6	
6	1/1/2020	4	4	1	3	2	39	62	12	2	
7	1/1/2020	4	5	1	3	2	38	73	13	2	
8	1/1/2020	4	6	1	3	2	36	65	8	3	
9	1/1/2020	4	7	1	3	2	37	59	14	4	
10	1/1/2020	4	8	1	3	2	37	53	12	12	
11	1/1/2020	4	9	1	3	2	38	46	15	5	
12	1/1/2020	4	10	1	3	2	39	45	17	21	
13	1/1/2020	4	11	1	3	2	40	43	20	37	
14	1/1/2020	4	12	1	3	2	40	41	18	29	
15	1/1/2020	4	13	1	3	2	40	43	22	33	
16	1/1/2020	4	14	1	3	2	39	43	20	26	
17	1/1/2020	4	15	1	3	2	39	45	21	39	
18	1/1/2020	4	16	1	3	2	39	48	17	27	
19	1/1/2020	4	17	1	3	2	38	48	16	20	
20	1/1/2020	4	18	1	3	2	38	50	13	22	
21	1/1/2020	4	19	1	3	2	37	50	15	17	
22	1/1/2020	4	20	1	3	2	36	52	15	11	
23	1/1/2020	4	21	1	3	2	35	54	8	10	
24	1/1/2020	4	22	1	3	2	34	64	6	5	
25	1/1/2020	4	23	1	3	2	30	67	6	6	
26	1/2/2020	4	0	0	4	2	29	68	7	2	
27	1/2/2020	4	1	0	4	2	27	69	6	2	
28	1/2/2020	4	2	0	4	2	29	72	10	0	
29	1/2/2020	4	3	0	4	2	28	72	7	0	
30	1/2/2020	4	4	0	4	2	27	75	6	3	
31	1/2/2020	4	5	0	4	2	29	72	8	12	
32	1/2/2020	4	6	0	4	2	28	75	8	33	
33	1/2/2020	4	7	0	4	2	29	75	12	67	
34	1/2/2020	4	8	0	4	2	28	75	7	118	
35	1/2/2020	4	9	0	4	2	34	67	6	57	

data

FIGURE 4.5: The example of Jersey City dataset.

Therefore, this experiment was conducted to compare the efficiency of each model that differed in inputs by eliminating some insignificant variables. Data analytical of independent variables related to bike use by using The Statistical Package for the Social Science (SPSS). Stepwise Multiple Regression Analysis was used to find the factors affecting bike use. The seven factors, namely temperature, hour, humidity, season, weekday, weather, holidays, and wind speed, were used to find the factors affecting cycling, as shown in Table 4.6. Based on the Stepwise Multiple Regression analysis, the results of the analysis of the data revealed that there were factors related to the use of bikes as follows: temperature, hour, humidity, season,

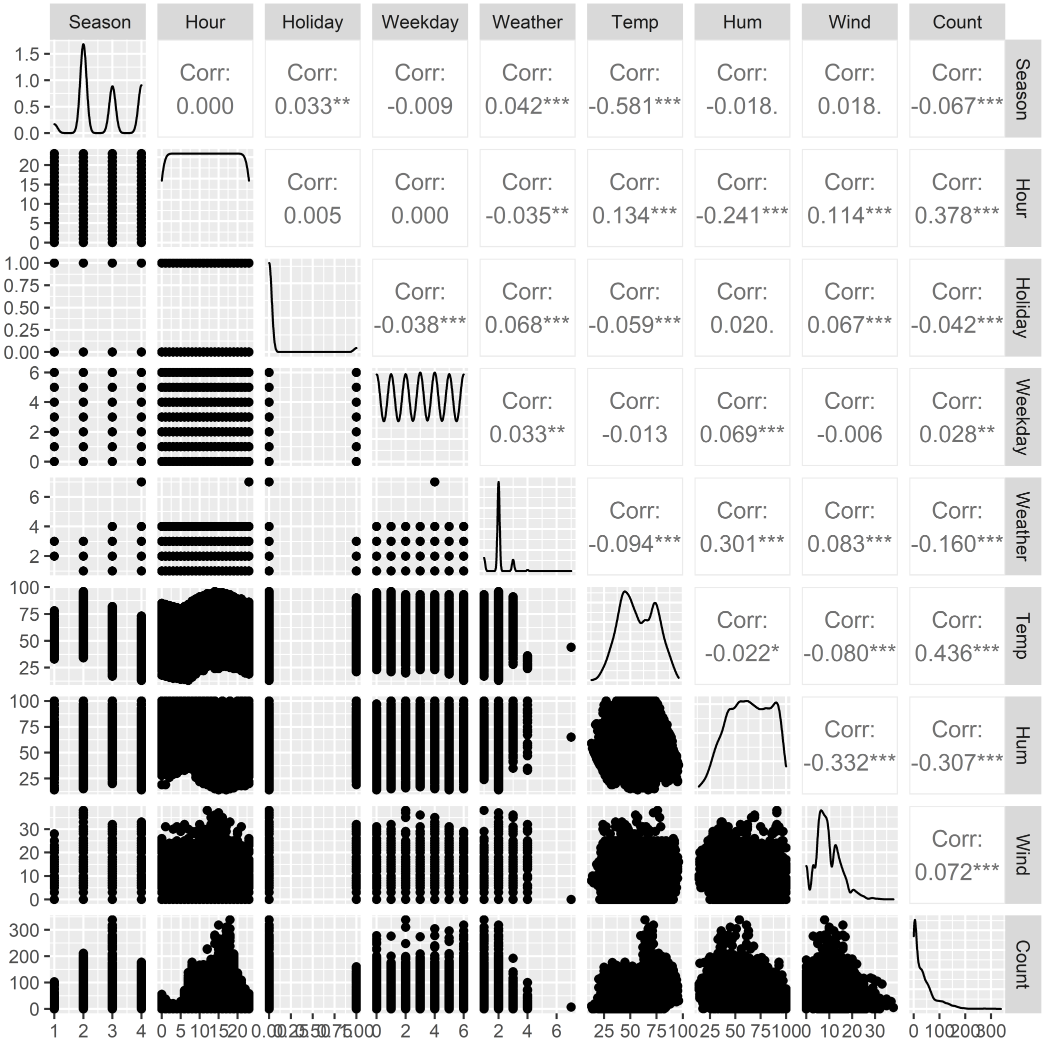


FIGURE 4.6: The correlation coefficient diagram of Jersey City data.

weekday, and weather. The factors that did not significantly affect the number of bikes used included holidays and wind speed. Therefore, the experiment was conducted by eliminating the aforementioned factors.

Several factors were considered in this study, that could be related and affect the use of the bike. On the other hand, taking factors into account in forecasting may cause the model used in forecasting to be affected internally, causing inaccuracies. To generate a forecasting model by verifying data assumptions. Normality, linearity, heteroscedasticity, and multicollinearity are the four assumptions that should be checked. The major goal of this section is to choose appropriate controlled variables for forecasting. The approach of statistics in regression known as multiple linear regressions (MLR) is used to examine the correlation between a single respond or dependent variable and two or more controlled variables or independent variables.

Stepwise regression is a popular approach [79, 80], and most statistical software programs offer it—which clearly shows demand and, ironically, may inspire researchers to adopt it. Efroymson [81] presented of automated stages to choose the explanatory factors for a multiple regression model from a set of candidate variables. The candidate variables are examined one by one at each stage, with the t statistics for the coefficients of the variables being considered being used in most cases.

Forward selection and backward elimination are combined in the stepwise regression approach. In stepwise regression formulas, controlled variables are added and removed as needed for each step. When all variables not in the model have p-values less than the provided Enter value and all variables in the model have p-values larger than or equal to the specified Remove value, the process has completed its operation.

The dependent variable with logarithmic, the correlation analysis allowed for the quick identification of dependencies. The values presented in Fig. 4.6 show the Pearson correlation coefficients. The variables statistically significant for p-0.05 are indicated in symbol as ***, and those statistically significant for p-0.01 are indicated in symbol as **.

The correlation between each feature is shown in Fig. 4.6, and the results show that count, hour, weekday, temperature, and wind exhibit a positive correlation. Furthermore, it shows that the two variables are related in the same direction. When the correlation between count, season, holiday, weather, and hum shows a negative correlation, it means that one variable is high and the other is low.

From the selection of input variables by stepwise regression method. As a result, these statistics show which control variables (temperature, hour, humidity, season, weekday, and weather) best fitting the data. Also, the R^2 coefficient determination results pertain to the model that was fitted to the data. The summary of model is based on the result in Table 4.5. The comparing each model, it was found that model 6, higher the R^2 and adjusted R^2 . Thus, the model 6 is the best of all model fits to the data. As shown in Table 4.6, the coefficient obtained from the input data analysis with stepwise regression method.

4.3 Models

In this section, present a basic introduction of the techniques for the sequential data forecasting/predicting for time series data, where the analysis is one in statistical. This work also proposes a model to predict the future usages of bike-sharing services. Many articles were presented with univariate datasets. However, the data of real systems do not only have one or two different variables (bivariable), but they have three or more variables, which is why multivariable analysis is needed, for which a multivariable dataset is required.

The data in a series of time periods that involve the trend, seasonality, and the cycle are called the time series data, and this method can be applied to accurately solve this case study. According to the various predicting/forecasting methods performed in the application, the advantages of artificial intelligence, including machine learning techniques, can decrease the errors and increase the chances of attaining higher accuracy. Generally, an RNN, an LSTM, a GRU, and a combination of each model can be used. However, a new architecture model combing LSTM and GRU is used in this research.

4.3.1 Recurrent neural network (RNN)

A model in an artificial neural network algorithm that can recognize patterns such as time series data, natural language processing, and video recognition in addition to other patterns is called the RNN [82], and it is the process of the sequential data, where the previous hidden state is used to calculate the current hidden state and the current hidden state is used to estimate the next period's state.

$$h_t = (x_t W + h_t - 1U) \tag{4.1}$$

$$o_t = (h_t V) \tag{4.2}$$

Where t is the current step, h is the hidden state, x is the input, o is the output, σ is the activation function, and W, U, V the weight matrix connecting x_t (input) to the previous hidden state h_{t-1} and the current hidden state h_t.

Therefore, the RNN is suitable for the sequence data or the time series data, where the RNN can be used with data that are not different in distance in a sequence. The problem with the transition function is the vanishing gradient influent difficulty learn [75, 83, 84].

4.3.2 Long short-term memory (LSTM)

A kind of an RNN is used to process the sequential data, where the LSTM is developed based on the RNN [85, 86] while using the cell state and the hidden stage to store and access data in the next stage so as to prevent the disappearance of the gradient problem. They are deciding for three gates to consist of the forgetting gate, which considers the unnecessary information, and the input gate is saved in the cell stage. The output gate is the data transmission into an output.

$$f_t = (W_f[h_{t-1}, x_t] + b_f) \tag{4.3}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{4.4}$$

$$\widetilde{C}_t = tanh(W_c[h_{t-1}, x_t] + b_c) \tag{4.5}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{4.6}$$

$$o_t = \sigma(W_0[h_{t-1}, x_t] + b_0]) \tag{4.7}$$

$$h_t = \ o_t * tanh(C_t), \tag{4.8}$$

where t is the current step, x is the input, o is the output, W is the weight matrix, and bis the bias. f_t is the forget gate at a time t, i_t is the input gate, C_t is the cell state of the internal memory at a time t, h_t is the hidden state, and $\widetilde{C}_t$ is the candidate hidden state that can be used in the next state, which decides to remember or forget the input data.

4.3.3 Gated recurrent unit (GRU)

GRU is a developed unit based on the LSTM in a recurrent network. The GRU process is a gate adjustment in the LSTM into the reset gate. Then, the gate is updated to consider how much enough data in the previous stage merges into the input and forget gate that can be saved more computation resources than LSTM. The hidden state of the previous stage is connected to the reset gate, and the GRU exposes all the memory, as it does not have the output gate. It consists of two gates: the reset gate and the update gate.

$$z_t = \sigma(W_z[h_{t-1},\ x_t]) \tag{4.9}$$

$$r_t = \sigma(W_r[h_{t-1},\ x_t]) \tag{4.10}$$

$$\widetilde{h}\ _t = tanh(W[r_t * h_{t-1}, x_t]) \tag{4.11}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{4.12}$$

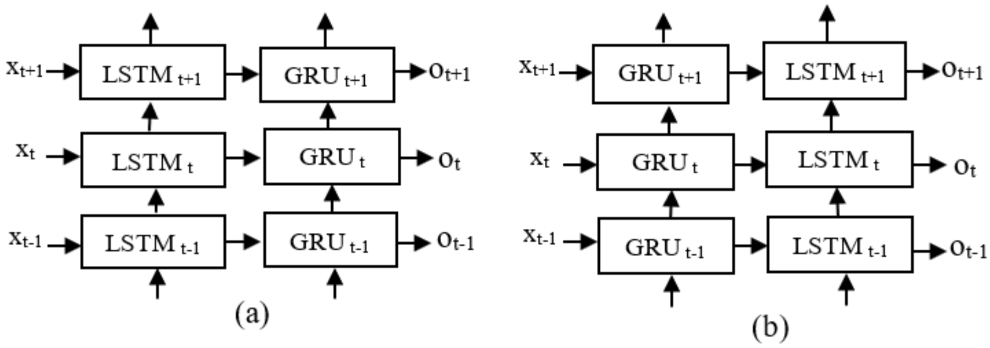


FIGURE 4.7: Structures of the mixed LSTM–GRU.

4.3.4 The combination of LSTM and GRU

A model combining an LSTM and a GRU is proposed [87], which are assembled and connected in time, and a parameters propagation method using the LSTM or the GRU sequence is considered. The cell and hidden state are converted to the next unit, where the long-term memory affects the current epoch output. As for the multiple layers between the LSTM and the GRU, the first LSTM or GRU layer calculates the hidden units using the inputs. Then, the second LSTM or GRU layer calculates the output of the hidden units, but the different parameters are independently calculated. At last, the neural network calculates the loss function and tries to optimize it. Figure 4.7 shows the structure of the combined LSTM and GRU.

4.3.5 Experimental

In this work, the dataset of the input matrix consists of temperature, season, humidity, wind speed, weekdays, holidays, and count to predict future bike-sharing usage. In the experiment, used the hyper-parameter as presented in Table 4.7. The performance of the results became stable after selecting each of the variables. The time step is 24 because the pattern of the data was acquired for every 24 hours together. Furthermore, we used the same hyper-parameter to build the proposed model. In the next step, each training step is processed. Then, the predicted results are estimated according to the bike-sharing usage of each model. In addition, from the comparison results, which indicate the model that gives the most accuracy. This

TABLE 4.7: Description of variable

Parameter	Value
Epoch	100
Batch size	100
Hidden node	100
Hidden layer	1-2
Time step	24

research examined the variables that did not affect the use of bikes, obtained from the correlation analysis.The selection results eliminated the input variables analyzed by the Stepwise Multiple Regression method as shown in Table 4.5-4.6.

4.3.6 Evaluation

This research used the mean square mean absolute (MAE), mean squared error (MSE), root mean square error (RMSE), and R^2 is the determination coefficient, which normally runs from 0 to 1, with a high value indicating that the predicted values completely match the actual ones that methods to measure the effectiveness and accuracy of the different models and to evaluate the bike-sharing usage prediction using the proposed model with multivariate time series analysis. The predicting error was calculated as follows, where n is the testing sample number, y is the real data of the bike-sharing usage, and $\widetilde{y}$ is the corresponding prediction of the bike-sharing usage.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widetilde{y_i})^2 \tag{4.13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widetilde{y_i}| \tag{4.14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widetilde{y_i})^2} \tag{4.15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\mathrm{y}_i - \tilde{y}_i)^2}{\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2} \tag{4.16}$$

4.4 Results

This work used a different model to predict the demand and compare the results. By used the R^2, RMSE, MAE, and MSE methods to measure the performance of the proposed model. Table 4.9 shows the predicted errors of the predictions of the bike-sharing usage and raw bike-sharing usage. which produces reliable outcomes based on pilot experiment. The experimental results showed that the mixed LSTM–GRU model had the least error values, as its R^2, RMSE, MAE, and MSE are 0.84, 16.02, 15.12, and 260.28, respectively.

As shown in the Table 4.5-4.6, The selection results eliminated the input variables analyzed by the stepwise multiple regression. By experimenting with entering and eliminating all six models of input variables, it was found that the sixth model had the highest R-squared value. Temperature, hour, humidity, season, weekday, and weather were discovered to be the variables influencing bike use.Then, in this research, the experiment was done to eliminate input variables consisting of wind speed and holiday. Therefore, the input variables mentioned above are used to predict the bike use in the LSTM-GRU model. The error value result was found to be less than the model without removing the input variables, as its R^2, RMSE, MAE, and MSE are 0.87, 12.80, 11.28, and 163.88, respectively

The approach is followed, the stepwise multiple regression models are substantially more successful. The results of this stepwise application show that employing theoretical expert opinion to select the initial set of input data for forecasting is beneficial.

Other approaches for selecting features have been presented [88]. In addition to procedural regression, the number of features used to cope with the error increases as the number of features increases. In practice, this may not be practicable. It also leads to oversampling and incorrect predictions owing to inconvenient variable combinations. When feature selection methods are evaluated using out-of-sample

Table 4.8: Result of prediction of the bike-sharing usage for Jersey City.

Order	Actual	RNN	LSTM	GRU	LSTM-GRU	GRU-LSTM	LSTM-GRU *
0	9	27.67	20.52	25.74	19.07	22.63	10.53
1	3	24.13	20.74	22.25	18.25	22.04	9.78
2	4	25.47	20.76	21.57	17.06	22.17	8.85
3	0	30.50	24.21	19.06	17.10	22.28	12.05
4	0	30.81	26.77	16.21	17.77	22.57	13.99
5	0	34.57	27.12	19.30	15.00	22.77	15.09
6	2	31.07	28.38	21.02	16.21	23.35	18.93
7	3	32.37	27.80	23.27	18.35	23.86	20.29
8	10	29.86	27.90	28.66	17.30	24.22	23.41
9	18	30.58	31.67	24.36	26.14	25.51	25.48
10	18	29.85	36.08	28.57	29.82	26.93	28.97
11	17	28.80	41.61	31.17	32.98	27.34	33.97
12	20	34.07	34.74	30.76	40.19	33.38	32.20
13	20	28.32	27.46	31.31	40.32	49.62	31.43
14	31	33.21	26.15	50.74	47.27	54.37	39.42
15	29	32.08	41.13	57.38	51.26	56.25	46.52
16	35	31.71	36.72	54.96	57.37	54.29	50.50
17	43	29.95	36.81	47.14	56.51	55.98	50.13
18	37	31.81	36.63	47.05	52.65	55.31	47.65
19	24	49.26	39.44	41.57	51.10	45.17	49.99
20	22	46.47	45.07	42.19	45.59	32.56	42.50
21	22	36.50	40.24	32.78	33.01	18.34	30.86
22	10	33.91	34.49	23.86	27.06	11.79	20.32
23	11	33.89	22.38	18.94	19.18	6.16	11.13

*Eliminate some input data

Table 4.9: Predicted errors of the bike-sharing usage for Jersey City.

Models	R^2	RMSE	MAE	MSE
RNN	0.43	19.83	17.52	393.05
LSTM	0.62	17.50	15.73	306.13
GRU	0.77	16.90	15.59	285.66
LSTM-GRU	0.84	16.02	15.12	260.28
GRU-LSTM	0.66	17.35	15.81	301.17
LSTM-GRU*	0.87	12.80	11.28	163.88

*Eliminate some input data

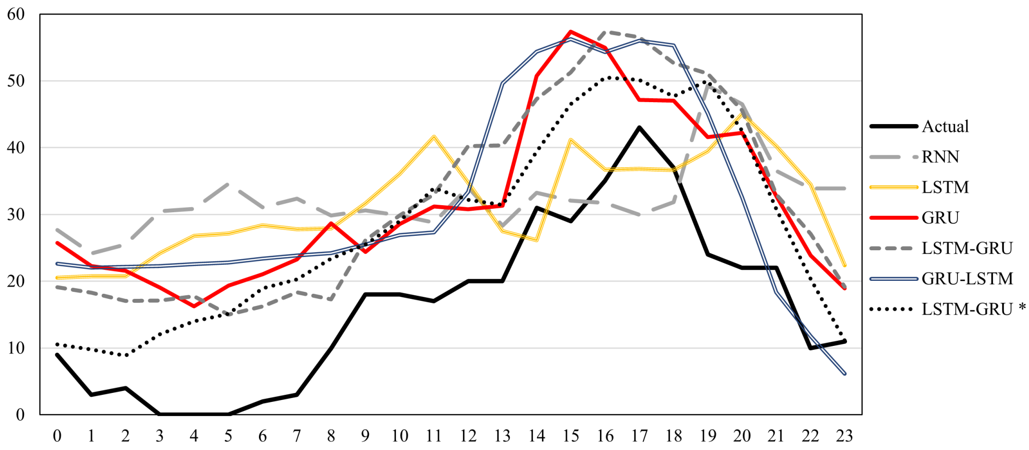


FIGURE 4.8: Comparison of forecast results between the actual value and predicted value of each model.

data, deleting recursive features with cross-validation is one of the benefits, will have a higher chance of being effective.

4.5 Conclusion

In this work, time series data and a machine learning technique were used to predict the demand for bike-sharing usage. Using historical data, including the season, weather, temperature, humidity, and wind speed with the time series analysis, this work compared the performance of the different prediction models that used an RNN, an LSTM, a GRU, a mixed LSTM–GRU, and a mixed GRU–LSTM. The results showed that the prediction of the demand for bike-sharing usage with the mixed LSTM–GRU model was the most accurate. The process of considering inputs used in forecasting is also important. It can also help increase the accuracy of the model used in forecasting. Thus, the findings of this study can assist the planners of bike-sharing companies in the distribution of bikes in bike stations. For future work, it should be considered the effects of each variable on the bike usage rate and will further develop the predicted results using other methods such as deep learning.However, future studies may require more information about the factors that influence bike-sharing demand or other statistical techniques to be used in the dataset.

Chapter 5

Bike-sharing relocation problem

5.1 Overview

Bike-sharing services have made bike rentals available for travelers and tourists so that they can rent bikes and return them at any station. Most bike-sharing systems provide automatic systems for users and operators so that customers can just use their smartphones to locate available bike stations, which makes it easier and more convenient to attract more customers. Recently, the frequency of using bike-sharing services has significantly increased due to the popularity of green travel, as many people have become more aware of pollution and other common health issues. Moreover, due to the increase in carbon dioxide (CO_2) levels, many people are taking more measures to reduce greenhouse gas emissions in every economic sector.

However, with the increasing popularity of bike-sharing services, many problems have arisen, including problems during peak hours, not enough bikes at some stations, etc. Also, sometimes, there are no available parking spaces to return bikes, especially in residential areas or near train stations. Thus, the operators of bike-sharing systems face many challenges when it comes to the allocation of enough bikes and parking spaces. This problem can be solved by determining the route of relocation bikes, which includes the picked up and returned bikes. Thus, cost-efficient operation can be used to guarantee profit maximization optimization. Moreover, customer satisfaction is important so customers can continue using the service, and it is also important for increasing the number of customers and improving service reliability.

Recently, bike-sharing systems have received a lot of attention from researchers. Operational relocation can be classified into two groups. The first group includes user-based strategies, which incentivize users to participate and encourage them to voluntarily relocate their rented bikes. Such strategies include static pricing and dynamic pricing strategies. Singla et al. [51] presented a crowdsourcing mechanism for dynamic pricing, which enabled the calculation of each station's incentive values and the development of a dynamic incentives system by offering incentive amounts to users and utilizing smartphone applications. The second group includes operator-based strategies, where relocation operators work on optimizing the pickup and delivery costs. Erdoğan et al. [45] presented an exact algorithm using a branch-and-cut algorithm which utilizes combinatorial Benders' cuts to separate infeasible solution from the feasible region to solve the static bicycle relocation problem by determining the minimum cost sequence of the stations to be visited by a single vehicle. Cruz et al. [89] presented an iterated local search heuristic for solving the bicycle relocation problem. Gajpal and Abad [90] proposed a construction rule in Ant Colony Optimization as two multi-route local searches to solve VRP with simultaneous delivery and pickup. Shui and Szeto [91] offered a dynamic green bike repositioning problem that simultaneously minimizes the total unmet demand of bike-sharing systems and the fuel and CO_2 emission costs of repositioning vehicles. The solution method was based on the ABC algorithm.

Numerous studies on bike-sharing systems have proposed solutions for the relocation problem for operating a vehicle routing. Consequently, the underlying vehicle routing problem has received the most attention. In our study, This work have proposed the effectiveness of the solution. This thesis also modified the efficiency of the swarm-based metaheuristic algorithm, ABC, and enhanced its search efficiency in solving the bike-sharing relocation problem.

5.2 Bike-Sharing relocation problem description

5.2.1 Problem description

Bike-sharing relocation is a part of VRP that can be applied to relocation. Battarra et al. [92] classified the VRP as widely treated pickup and delivery problems (PDP). PDP is a kind of VRP where goods must be transported from different origins to different destinations, and it is divided into three classes. The first class is one-to-one (1-1), where each good and request are provided with a pair of origin and destination. The second class is the One-to-Many-to-One (1-M-1), which represents how some goods must be delivered from a depot to customers and how other goods must be recollected from the customers and carried back to the depot. The third class is Many-to-Many (M-M), which represents how each good can have various origins and destinations and how any location can be the origin and destination. The problem of relocating bikes in a bike-sharing system lies in this class.

The model for solving vehicle routing problems for deliveries and pickups (VRPDP) aims at minimizing the cost or distance by providing customers with the allocation of vehicle routes for truck trips to service customers. One limitation that must be considered is the vehicle loading capacity. Although this is a significant problem, it is not extending VRP due to the lack of multiple travel plans. Customers receiving goods from a depot are called linehaul or deliveries. Customers who send goods back are called pickup or backhauls. It is possible that customers may want to both send and receive goods at the same time. This case is called combined demands. Also, in bike-sharing relocation, a customer at a station may need to pick up or drop off a bike [93]; hence, it is possible to adopt VRPDP with combined demand in solving this problem. This can be modeled using integer linear programming models, which involve minimizing the cost or distance. Starting from the depot, the truck drives to visit each station exactly once to drop off or pick up a bike for relocation. Then, the truck must be back to the depot. The problem can be defined on a graph $G = (V, A)$, where $V = 1....,n$ is a set of nodes (stations) in a depot located at node 0, and $A = \{(i, j) : i, j \in V_i \neq j\}$ is a set of arcs (distance between each pair of vertices). Each station i has a demand dd_i or pd_i, where pd_i denotes the pickup bikes that must be removed at station i, and dd_i denotes the drop off bikes that must be supplied at station i. The bikes removed from the pickup stations can either go to a drop off station or back to the depot. The bikes supplied to a drop off station can either come from the depot or from a pickup station. Also, the feet of m is the limitation of the capacity Q of each available vehicle at the depot. This problem is related to determining a relocation while minimizing the total cost of a fleet with a number of (m) vehicles through the graph.

There are various the method to solve bike-sharing relocation as vehicle routing problem. The Swarm Intelligence is an efficient method as the ABC algorithm [94], but it is still insufficient for the selection strategy. Karaboga, D., and Basturk, B. [95] and Pathak, N., Mishra, M., and Kushwah, S. P. S. [96] proposed the modified ABC by using Local Search (LS). The GLS is a way to improve the selection strategy in the ABC algorithm because the GLS gets the results better than LS [97]. To narrow the gaps of the previous research work, this work proposes the modified ABC in a neighbor solution to enhance the solution performance, namely GLS to improve the solution performance to apply in bike-sharing relocation problem.

5.2.2 Mathematical modeling for the bike-sharing problem

According to [98], with the objective of minimizing the total cost, a model was proposed for solving the VRPDP in order to make the mathematical model of the bike-sharing problem mimic VRPDP. The mathematical model for minimizing the total cost in this problem was defined as follows:

Parameter

Decision variables $x_{ij} = \begin{cases} 1, when\, vehicle\, m\, travels\, from\, station\, \text{ito}j \\ 0, otherwise \end{cases}$

$$\min \sum_{i=0}^{n} \sum_{j=0}^{n} D_{ij} x_{ij} \tag{5.1}$$

$$s.t. \sum_{i=0}^{n} x_{ij} = 1,\, j \in \{1, ..., n\} \tag{5.2}$$

$$\sum_{i=0}^{n} x_{ji} = 1,\, j \in \{1, ..., n\} \tag{5.3}$$

$$\sum_{i=0}^{n} Dr_{ij} - dd_i \ = \sum_{i=0}^{n} Dr_{ji}, j \in \{1, ..., n\} \tag{5.4}$$

$$\sum_{i=0}^{n} Pu_{ij} + pd_i = \sum_{i=0}^{n} Pu_{ji},\ j \in \{1, ..., n\} \tag{5.5}$$

$$\sum_{i=0}^{n} Dr_{i0} = 0 \tag{5.6}$$

$$\sum_{i=0}^{n} Pu_{0i} = 0 \tag{5.7}$$

$$Dr_{ij} + Pu_{ij} \leq Qx_{ij},\ i, j \in \{1, ..., n\} \tag{5.8}$$

$$\sum_{i=1}^{n} x_{0i} = m \tag{5.9}$$

$$x_{ij} = \{0, 1\}, i, j \in \{1, ..., n\} \tag{5.10}$$

The objective function is to minimize the total travel distances of all trucks drivers for bike relocation that define the feasible solutions of the routes given the constraints. Constraints 5.2 and 5.3 ensure that the vehicle must only visit stations. Constraint 5.4 and 5.5 guarantee that the flow conservation constraints are met. Constraints 5.6 and 5.7 confirm that the vehicle starts at the depot with zero pickup bikes and finishes with zero drop off bikes. Constraint 5.8 make sure that the vehicle picks up and drops off loads at any customer location within the vehicle load capability. Constraint 5.9 verify that the vehicle leaves from the depot, and Constraint 5.10 is a binary variable.

5.3 Trial and Error of the Algorithm for Solving Relocation Bike-Sharing

To solve the vehicle routing problem, researchers have proposed a variety of methods, such as the exact, heuristic, and metaheuristic methods. Mathematical models can be used to solve such problems and to explain various problem aspects. Likewise, metaheuristic methods, such as the genetic algorithm or ABC algorithm, have been applied to solve such problems.

The most common relocation operation problem that exists all day every day is the bike-sharing relocation operation problem, and it can either be static or dynamic. The static problem is when relocation is performed on a predetermined schedule when a system is closed or minimally operating at night. The dynamic problem is when relocation occurs in the daytime when the system rapidly changes and needs relocation. Essentially, for bike-sharing relocation systems in small or medium cities, the bikes are often carried out at night using a vehicle that visits each station exactly once. Many researchers proposed methods for solving the vehicle routing problem to find out the minimizing cost. Using the exact method (genetic algorithm (GA)) [99], that has been widely used in various real-life applications, This work used a trial-and-error experiment to compare the performances of the well-known in the literature and those of different kinds of optimization algorithms consisting of mixed integer programming problems (MIP). The representation of chromosomes is closely associated with real-life problems. The main advantages of GA are that it is robust, efficient, and accurate, and the artificial bee algorithm (ABC) [100, 101] has become one of the most common optimization methods in the field of artificial intelligence since it was first conceived in the early nineties. As a result, many research works elaborated on the value of using it in well placement optimization to solve the bike-sharing relocation problem. Gurobi solved such problems using MIP. GA and ABC were coded in Python.

In the trial-and-error experiments, the ABC algorithm showed the best performance, but it took a long time compared with the other methods. This work aimed at modifying the performance of the ABC algorithm in solving the bike-sharing relocation problem. The results were consistent with those of Rothlauf [102], who

TABLE 5.1: Comparison of the performances of the experimental results for each algorithm.

Number of Stations		Algorithm		
		MIP	GA	ABC
10	Avg.(km.)	348.96	291.57	282.84
	CPU time (S.)	9.12	1.96	8.29
20	Avg.(km.)	303.29	313.15	302.05
	CPU time (S.)	10.43	2.8	10.25
30	Avg.(km.)	n/s	567.7	423.9
	CPU time (S.)	>24hr.	2.98	21.3

proposed an exact optimization method that guaranteed finding an optimal solution. Also, using heuristic optimization methods, there were no guarantees that an optimal solution can be found. Usually, the exact optimization method is a choice method if it can solve an optimization problem with an effort that polynomially grows with the problem size. The situation is different if the problems are NP-hard, as exact optimization methods need exponential effort. Then, even medium-sized problem instances often become intractable and cannot be solved anymore using exact methods. To overcome these problems, heuristic optimization methods can be used. Binitha and Sathya [103] presented the GA algorithm to solve the convergence problem for local minima or maxima. Further, GA was unable to effectively solve constrained optimization problems.

5.4 Modified ABC algorithm

5.4.1 The original ABC

In 2005, Karaboga [104] developed the ABC algorithm [105], which is a method for solving optimization problems. The algorithm imitates the behaviors of bees when searching for food (Honeybees). The bee colony population is divided into three groups: employed bees, onlooker bees, and scouts. The employed bees search for food and then came back to share information about food sources to the onlooker bees, which search their nests, inspect the selected food sources, and compare them with the food sources nearby. This algorithm improves the ability of exploring and finding the best solution (optimization algorithm) using scout bees, which guide the mutation process to the algorithm by searching for new food sources in the previously unexplored areas of the survey area to increase the chances of finding better food sources. Then, the selected food sources are discarded and transformed from the employee bees to the scout bees so as to find new food sources. The location of a food source is the value of a possible answer. The number of employed bees and scout bees combined is the total number of possible solutions by finding food sources.

The ABC algorithm is an efficient approach for solving the vehicle routing problem. Nevertheless, this algorithm has not been confirmed as a global solution, but it provides optimal solutions for NP-hard problems [106].

Step 1: Initial population creation of all bees with location of food sources chosen by random selection (initial phase) based on Equation 5.11 where $i = 1, \ldots SN, SN$ is indicate the number of food sources, and $j = 1, \ldots, D$, and D is the dimension of the problem.

$$x_{ij} = x_{\min j} + rand\,[0,1]\,(x_{\max j} - x_{\min j}) \tag{5.11}$$

Step 2: Employed bee phase. The employed bees search for new food sources based on Eq. 5.12, where v_{ij} is the new solution in the next generation, $\varnothing$ is the value obtained from the randomness in the range of 0-1, $i \in \{1, 2, ..., SN\}, SN$ is the size of colony, $j \in \{1, 2, ..., D\}$, and D is the dimension of the problem. Then, the suitability is calculated based on Eq. 5.13 if the new position value is better than before to update the position to a new value.

$$v_{ij} = x_{ij} + \varnothing \cdot x_{ij} - x_{kj} \tag{5.12}$$

$$fit_{ij}\left(\vec{x}\right) = \begin{cases} \frac{1}{1+f_i\left(\vec{x}_i\right)}, & if\, f_i \geq 0, \\ 1 + \left|f_i\left(_i\right)\right|, & if\, f_i < 0 \end{cases} \tag{5.13}$$

where

x	denotes the vector answers from random,
i	is the population,
j	is the parameter value,
$f_i\left(x_i\right)$	is the x objective function value,
$fit_{ij}\left(x\right)$	is the fitness value solution of the food source.

Step 3: Onlooker bee phase. The onlooker bees consider the obtained food sources from the employed bees using a probability that can be obtained from equation 5.14, where pi is the selection probability of the current solution. If the food source has high probability, it is very likely to be chosen. Then, the onlooker bees send the selected data for calculation in order to find more suitable food sources, just as the employed bees.

$$P_i = \frac{fit_i(x_i)}{\sum_{i=1}^{\tau} fit_i(x_i)} \tag{5.14}$$

where fit_i is the fitness value of solution i, which is proportional to the nectar amount of the food source in position i, and τ is the number of food sources, which is equal to the number of employed bees or onlooker bees.

Step 4: Scout bee phase. When the original food sources of the employed bees are not selected by the onlooker bees, the scout bees calculate the new food sources by randomly replacing those that were not selected.

Step 5: Condition satisfied. The bees stop searching when finding the best food source. Otherwise, they go back to Step 2 again until max iterations.

The ABC is still a qualification of poor exploitation [107]. This causes capturing in the optima areas and results in slower convergence rates, thus tackling various problems. In the past, many researchers modified the basic ABC structure. The proposed variable was named LS-ABC (local search based ABC). The performance was tested on more than 12 standard functions [108]. Pathak [109] presented an enhanced ABC algorithm with local search using an incremental approach for the traveling salesman problem.

5.4.2 Local search

The local search strategy is a kind of constraint propagation. When solving constraint networks, local search strategies are often modified to discard the states that cannot be solutions and to rank the states that are still solution candidates. This idea has been applied to efficiently explore large neighborhoods [110]. The local search algorithm starts from a candidate solution, and iteratively moves to a neighbor solution. From the original ABC, in the onlooker bees' states, a neighbor solution is only used for replacement when the onlooker bees find the best neighbor solution [111].

The local baseline search algorithm starts with an arbitrary solution and ends with a local minimum, which cannot be further improved. During these steps, there are several local search ways. For the best improvement, such as greedy selection, the local search replaces the current solution with the most cost-improving solution after searching the entire neighborhood. The local search method can quickly resolve optimal routing [112]. The limitations of the problem variance were also found to be quite high. This thesis found the method to improve the operators in the steepest descent search strategy to avoid the local optimal for reaching global optima as with the guided local search (GLS) algorithm [97]. To probably improve the efficiency of the solutions in the search process.

Guided local search is an optimization technique which is an intelligent search algorithm that exploits information to guide the local search in avoiding the local optimum [113, 114]. The GLS solution modifies action from local search by augmented cost function of minimizing the problem objective function with the cost function to a penalty term that was applied by a penalty vector p, where pi is the penalty value of feature i. The GLS uses local search to minimize objective function by augmented objective function. Therefore, local search is performed via local search (S, p) function, staring from solution S and then returning to a new solution improved by the augmented objective $h(S)$which spread of penalties as follows:

$$h(S) = O(S) + \lambda \sum\nolimits_{i \in M} p_i l_i(S) C_i \tag{5.15}$$

where $O(S)$ is the original objective function problem. The C_i is a cost vector of feature i. p_i is the penalty parameter and if the feature is not exhibited in the local optimum, then the penalty value is 0, when the local search is tapped, where the penalty parameters are incremented by 1. Penalty is the indicator involving the feature i, which is the distance between customer location and other locations, and λ, a parameters to GLS, represents the relative value of penalties to control the information on the search process with respect to the actual solution cost. Arnold and Sörensen [114] found that $\lambda = 0.1$ works well. A l_i is a Boonlean indicator in the solution feature i. The essential effectiveness of GLS is the penalty parameters that are the costliest features in the current solution and are weighted by the number of times a feature has already been penalized. The penalties of the features are initialized to zero and are incremented for the features that maximize the utility formula. After the improvement method when local search settles, the penalty factor was used to penalize the bad features of i. If they keep a local search in local optimization, the current solution, which has the most cost, is penalized by weighing the number of times. They choose the features for which $C_i/(p_i + 1)$ is the largest among the features in S.

5.4.3 Proposed algorithm

This study proposed states for improving the performance of the ABC algorithm to solve the bike-sharing relocation problem, which is modified in the scout bee phase using local search based on the neighboring operators. The proposed algorithm is called the GLS-ABC algorithm, and it was demonstrated, as shown in Figure 5.1.

This work proposed a modified state so that x_i is replaced by the best neighbor solution, which is a guided local search algorithm that starts from a randomly selected complete instantiation and moves to the next instantiation. This idea may prevent the bad regions of the solution search.

The solution starts searching for the route tour that is closest to the depot. Next, searching for stations makes addition of more routes feasible based on the demand of each station and the capacity of the truck's delivery constraints.

```
Pseudocode of GLS-ABC
Begin:
  Initialization of generate a set of food sources xi, i=1,…,SN according to Eq. (4.11)
  Evaluate each xi, i=1,…,SN according to Eq.11
  Set v=0 and li = 0, i =1,2,…,Nb
  While (v<MaxInteration) do
  For i=1 to SN do (employed bees)
  Select a random solution according to Eq. (4.12)
  Calculate its fitness value of new food source according to Eq. (4.13)
  Apply random neighborhood operator, GLS selection by
  Penalize the worst distance (i, j) by incrementing p (i)
  Apply local search using Eq. (4.15)
      End for
        Calculate the probability each food source according to Eq. (4.14)
      For i=1: (onlooker bees)
        select a food source x_i using fitness-based roulette wheel selection method.
        Apply neighborhood operator on x_i → x̂, GLS selection by
        Penalize the worst distance (i, j) by incrementing p (i)
        Apply local search using Eq. (4.15)
    if f_i(x_i) < f_i(x̂_i),
    select x̂_j that is set of food sources with j is maximize of set of food sources,
    replace x̂_j with x̂,
    End if
    End for
    if l_i is limit, (determine abandoned solution for the scout)
          then using a neighborhood operator on x_i → x̂,  and replace x_i with x̂.
          End if
      Memorize the best solution.
      V=V+1
  End while
```

FIGURE 5.1: The correlation coefficient diagram.

The algorithm finds the best route by utilizing the cost function, which is determined from the distance between two stations and the demand of each station, under truck constraint considering the truck's capacity possible load and unload itself. For the fitness function, is equal to $1/Z(X_i)$, where $Z(X_i)$ is the sum of the route distances of the food source X_i. Thus, the fitness value inverses the total distance value; as such, the minimized total distance affects the fitness value. The problem was defined by N bike-sharing stations and a symmetric distance matrix, where $D = [d_{ij}]$ gives the distance between any two stations i and j. The goal is to find the minimum total distance for bike-sharing relocation based on the truck constraints; the truck starts from depot and visits each station exactly once, and after pickup or drop off, the truck must return to the depot.

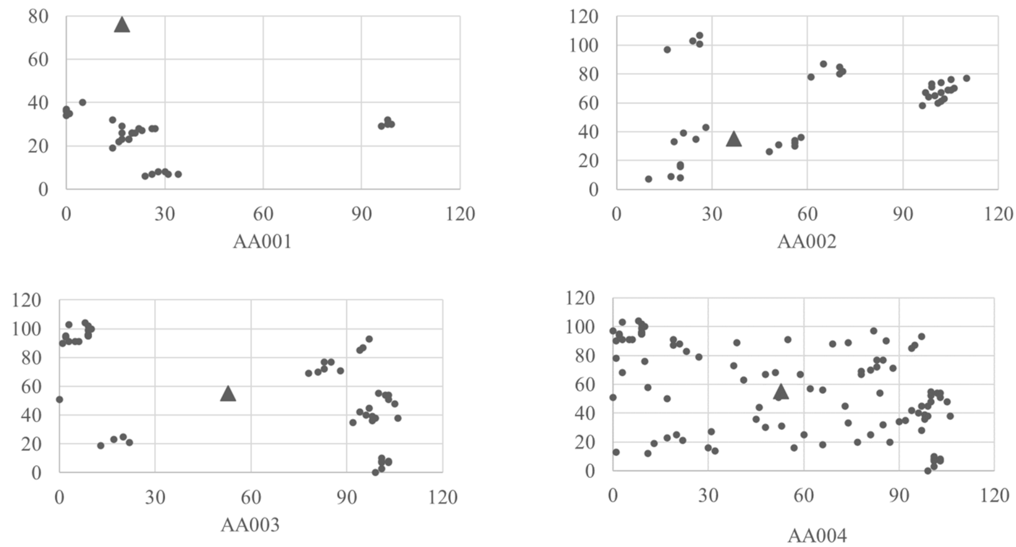


Figure 5.2: Bike-sharing station distribution at each instance using difference types (Solomon Problem) (▲ is depot, • is station).

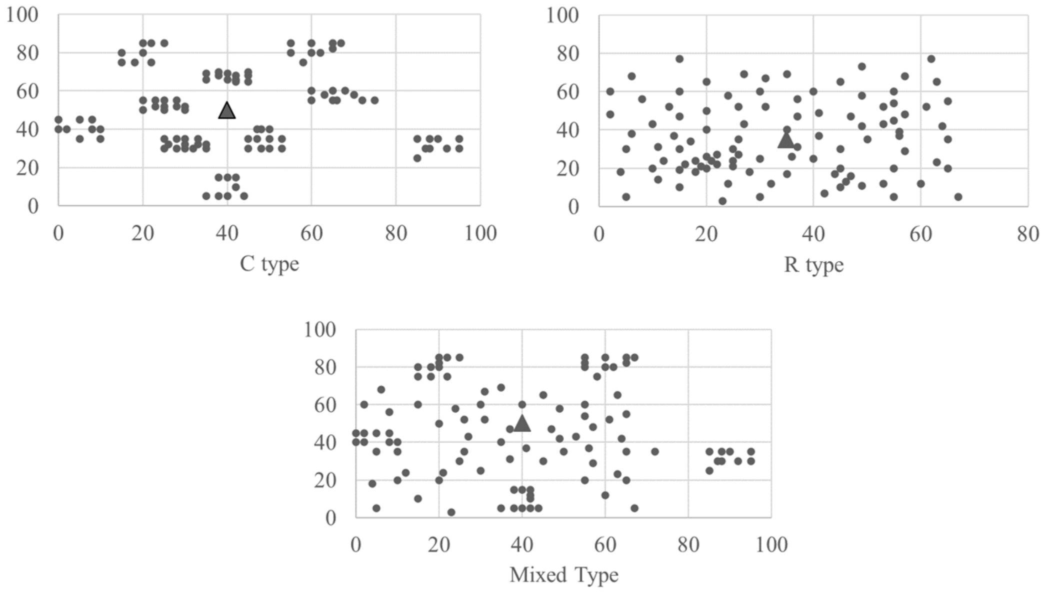


Figure 5.3: Bike-sharing station distribution at each instance using difference types (Solomon Problem) (▲ is depot, • is station).

5.5 Experimental

5.5.1 Data

The experiments revealed that each depot provided service using capacitated vehicles for relocation bikes. Each truck can carry a maximum of 20 bikes. The illustrated dataset was used; each instance consisted of the coordinate of location, pickup demand, and drop-off demand. The coordinates of the bike-sharing stations were randomly generated in a Euclidean plan. Also, the drop off and pick up demands of the bikes at each station were randomly generated to vary a dataset consisting of four instances as AA001-AA004.

This research evaluated the performance of the proposed algorithm based on the locations of the customers using the Solomon problem [115], which is classified into three types. First, the R data set having randomly distributed bike stations locations was generated in a problem consisting of R101 and R102. Second, the C data set have clustered bike stations location distribution in a problem consisting of C101 and C102. Last, the RC data set have a mix of randomly located and clustered bike stations distribution structures in a problem consisting of RC101, and RC201, where each problem has 100 bike stations. The drop off and pick up demands of the bikes at each station were randomly generated. The experiments considered how to operate a minimized total distance of the truck routing for bike-sharing relocation.

5.5.2 Evaluating the performance of the modified algorithm.

This work evaluated the performances of the original ABC algorithm (LS-ABC) and proposed ABC algorithm (GLS-ABC). The experimental design of each condition contains 20 replicates (run 20 times) to improve the accuracy and reduce experimental errors. This work set the number of employed bees and number of onlookers to be equal to the number of food sources ($\tau = 25$) based on Karaboga and Basturk [95], and each time the algorithm was run for 2,000 iterations [111]. The experiments were conducted in randomization, and the collected results consisted of the best total distance and average total distance. Then, the results were confirmed by statistically comparing the performances of the difference approaches. The paired

TABLE 5.2: Comparing ABC, LS-ABC, and GLS-ABC.

Instance	Num of stations	Average			Best		
		ABC	LS-ABC	GLS-ABC	ABC	LS-ABC	GLS-ABC
AA001	30	423.9	379.08	361.82	371.91	311.24	301.12
AA002	40	568.22	504.36	500.75	561.86	481.94	460.04
AA003	50	876.67	751.44	728.82	811.62	730.41	721.8
AA004	100	1907.87	1559.15	1530.12	1772.31	1469.9	1447.74

TABLE 5.3: Comparison of the improvements of the experiment results for LS-ABC and GLS-ABC.

Instance	Average				Best			
	LS-ABC		GLS-ABC		LS-ABC		GLS-ABC	
	%	P-Value	%	P-Value	%	P-Value	%	P-Value
AA001	10.57	0.0000	19.02	0.0000	16.3	0.0109	14.65	0.0101
AA002	3.87	0.0000	18.12	0.0001	14.2	0.0069	11.87	0.0005
AA003	14.29	0.0000	10.20	0.0000	10.00	0.0000	17.67	0.0000
AA004	18.28	0.0000	18.31	0.0000	17.10	0.0000	19.8	0.0000

TABLE 5.4: Comparison of the CPU times of the experiment results for ABC, LS-ABC, and GLS - ABC.

Instance	CPU time (in seconds)			P-Value
	ABC	LS-ABC	GLS-ABC	
AA001	20.56	26.73	57.75	2.29632E-11
AA002	20.67	28.13	64.31	3.900364E-18
AA003	17.84	27.75	80.45	2.4363E-10
AA004	30.70	47.02	120.14	8.1925E-19

t-test and analysis of variance method were used. The simple cost operation categories were divided into two groups: fixed cost groups may include the vehicle leasing cost and the diver's salary, and the variable cost group, which is the fuel per meter, and vehicle maintenance. For reduce cost that change directly as the fuel cost per kilometer. Thus, this work only presented the total travel distance. This work has raised the issue of improving the routing efficiency for the bike-sharing relocation problem. Therefore, the critical point of the change was compared, namely the cost of fuel per kilometer.

Table 5.5: The experiment results of the different types of data sets for ABC, LS-ABC, and GLS-ABC.

Instance	Average			Best			P-Value
	ABC	LS-ABC	GLS-ABC	ABC	LS-ABC	GLS-ABC	
C101	910.84	882.8	862.8	854.5	822.42	818.7	0.015409375
C202	1038.19	1012.44	1002.44	987.82	947.32	930.1	6.1318E-13
R101	1061.59	1023.04	1010.12	970.3	956.94	944.35	0.00276636
RC101	1174.39	1132.31	1110.71	1045.79	1027.4	980.5	0.016862495
RC201	1330.34	1073.17	1073.9	1058.67	999.9	960.84	3.6600E-06

5.6 Results and discussion

The model for finding out the best solution was coded using Python and run on a computer (Intel core i7 CPU3.80 GHz PC with 16 GB RAM, Windows 10). The optimal solution results of the experiments consisted of the best objective value of the minimum total distance and the average total distance. This work measured the performance using two methods: number of stations and data set type. The results of the various data of the number of stations were shown in Table 5.2, the CPU times were shown in Table 5.4, and the percentage improvement of the objective value (total distance) was compared between the original ABC and LS- ABC, and the original ABC and GLS- ABC. Furthermore, the p-value of the t-test of the comparative mean objective values between the algorithms based on the original ABC was shown in Table 5.3. From Tables 5.2 and 5.5, regarding the test results of the total distance and average tour distance of the bike-sharing relocation problem, the GLS-ABC algorithm improved the total distance by more than 3 percent at each instance and on average distance by more than 10 percent at each instance. The original ABC was statistically significant with the LS-ABC and the GLS-ABC via p-value¡0.05 at each instance for both total and average distance. Nevertheless, as shown in Table 5.4, the GLS-ABC took more time than LS-ABC and the original ABC. Overall, the test results show that the GLS-ABC algorithm can produce much better solutions than the original ABC and the LS-ABC in regard to solving bike-sharing relocation problems.

Table 5.3 shows that the GLS-ABC algorithm improved both the average and best total distances compared with the basic ABC algorithm and LS-ABC algorithm. The essential operation cost is a variable cost of the fuel cost per kilometer. The

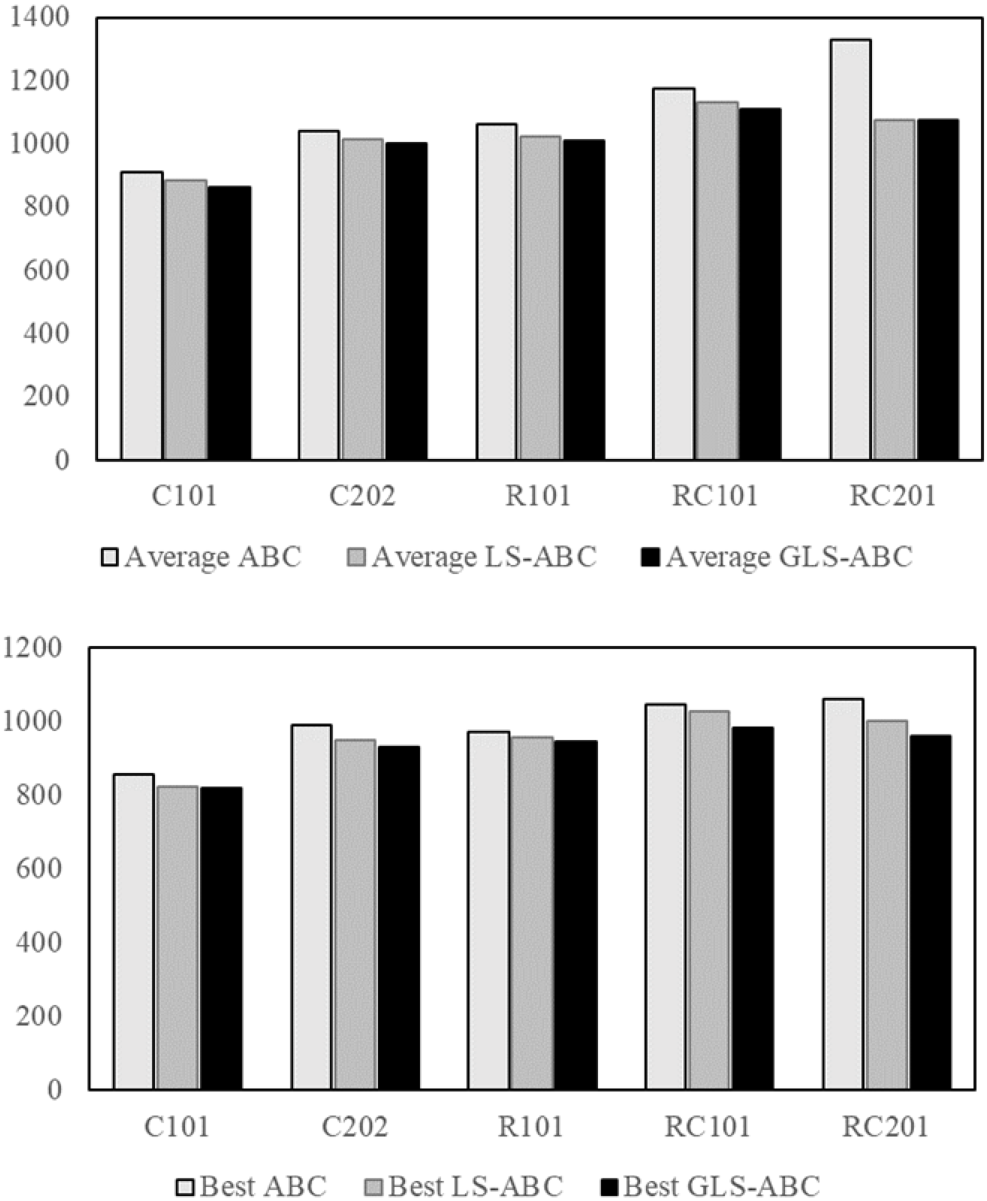


FIGURE 5.4: Comparisons of the different types of data sets of the experiment results by average value

results showed that the GLS-ABC algorithm is most likely better in reducing the bike-sharing relocation cost.

5.7 Conclusion

In conclusion, this chapter provide the following contributions: The bike-sharing relocation problem involves customer satisfaction and the benefits of bike-sharing service providers. Also, bike-sharing services are alternative transportation options for many travelers and locals, and they offer an environmentally friendly and healthy transportation method. In this work, an alternative algorithm was presented to solve the bike-sharing relocation problem. The ABC algorithm could effectively find a solution for the routing problem. This research modified the original artificial bee colony to improve the effective solution, and the results showed that the modified algorithm is better than the original ABC algorithm and GLS-ABC in regard to diminishing the bike-sharing relocation problem. The GLS-ABC algorithm could also offer better solutions than those of the original one. The operational costs could be reduced by reducing the vehicle fuel cost. However, the proposed algorithm took a longer time than the original one. In the future, This research aim to use the proposed algorithm to solve other real data set in real situations. Also, this work aim to breakeven the number of times for relocation a day and to develop a different algorithm. Then, one may consider being served with uncertain demand with unknown distributions, which is similar to the real situation.

Chapter 6

Reinforcement Learning for bike-sharing relocation problem

6.1 Overview

Several researchers have proposed a solution to the imbalance. While the balancing problem is divided into two categories, the user-based approach is the economical method of motivating the user. To encourage users to use bike sharing at stations with high inventory stations rather than stations with low inventory in the bike or parking slot available. To achieve relocation in all bike-sharing stations, the vehicle-based approach utilizes trucks or bike trailers to load or unload bikes. Almost all research on bike-sharing systems is still in the process of rebalancing which is, essentially the most crucial consideration in the vehicle routing problem. Thus, we propose that the vehicle-based approach be used in this study.

Existing artificial intelligence (AI) is thriving; for example, reinforcement learning (RL) is computationally cheaper than model-free learning and labeled data that are not available [116]. This method can also be used in complex problems to obtain the best solution with high accuracy. A type of RL is used to learn how to pay AlphaGo for becoming the professional player champion and possibly the strongest player. Generally, dynamic systems work with AI as well. We only found one related work that proposed a deep learning approach to forecasting user demand

and hierarchical reinforcement pricing for rebalancing dockless bike-sharing systems. That is not common; only a few studies propose that reinforcement learning can be used to rebalance bike-sharing in a vehicle-based approach.

This study proposes an optimal rebalancing model to determine the minimum cost for relocating bikes from high supply to high demand when empty stations prevent users from picking up bikes. Furthermore, it presents a vehicle-based approach for solving the relocation bike-sharing problem using the Q-learning algorithm and the SARSA algorithm from the reinforcement learning branch.

6.2 Related work

In recent years, bike-sharing systems solutions have been proposed to mitigate the system's demand and supply imbalance. The related work of relocating bike-sharing solutions can be divided into two categories. The first category focuses on the user-based approach; here, customers actively balance bike-sharing, Pfrommer et al. [52] proposed using dynamic pricing strategies on real-time price incentives to entice users to return bikes to short-supply bike stations. Zhang et al. [53] present a method on a user-based bike-sharing system as a dynamic pricing strategy with negative prices for improving the problem of bike imbalance with demand and supply in the system using user equilibrium dynamic traffic assignment model that was developed to capture the behavior of traveler's response for route mode selection. The second category focuses on the truck-based approach, in which the service provider operator dispatches bikes using multiple trucks with both static and dynamic problem versions. The static version approach, Chemla et al. [46] presents a method for solving the rebalancing issue that is considered to be a problem of many-to-many pickup and delivery problems by using tabu search on the branch-and-cut algorithm as a capacitated single vehicle. Yanfeng Li et al. [47] presented mixed-integer programming and a modified hybrid GA for solving the static bike-sharing relocation problem with various bike types. Dell et al. [48] proposed a metaheuristic algorithm for bike-sharing rebalancing problems that uses both constructive and local search. Bulhões et al. [49] presented a branch-and-cut algorithm for solving the multi-vehicle sharing relocation problem and developed iterated local search-based heuristic, which was implemented on instances ranging from 10 to 100 nodes. Caggiani et al. [41] propose a dynamic method in real-time as a decision support system for dynamic rebalancing problems in the dynamic version. On the dynamic rebalancing of bike-sharing systems, Chiariotti et al. [50] proposed a framework for dynamic strategies that can be better adapted to the volatile nature of the network than rebalancing models.

However, existing research has shown that the results of the user-based approach for bike-sharing relocation in the actual state are not significant compared to the truck-based approach [52, 117]. The user-based approach is applied, which may not be sufficient to persuade the customer to change their attitude. As a result, this study focuses on a truck-based approach to balancing bike-sharing relocation.

Artificial intelligence (AI) has grown in popularity in recent years. Machine Learning (ML) and Deep Learning (DL) are subsets of AI. According to Xin et al. [118], the difference between traditional ML and DL is the amount of data, as DL does not perform well with small datasets. Because DL must work with GPU, it relies on higher-performance GPU-enabled machines. However, DL works as well; rather than breaking down the problem into multiple sub-problems, DL contributes directly to end-to-end solutions.

Thus, this study focuses on machine learning to solve the problem because ML works equally well with small data sets. The previous study [119] described the overview of the ML method in particular. Reinforcement Learning (RL) can be divided into several types, the wild use as the on-policy and off-policy. Machine learning has been proposed to solve the vehicle routing problem (VRP) [120], adopting the RL that consists of greedy and RL-based for solving the VRP, with the results also showing the RL.

Many variants of the RL can solve the basic problem. Q-learning has evolved into one of the RL and has been extensively studied for optimization and operation research. The Q-learning algorithm, an off-policy algorithm that uses maximum action value for expected value, is the most applicable reinforcement learning approach. Bouhamed et al. [121] propose a reinforcement learning routing scheduling framework, using Q-learning to train the agent and a reward function to consider task times and delays. Furthermore, the SARSA outperforms the well-known RL. Both do not estimate the environmental model, which is an algorithm that estimates the function value directly from the simulation of the direct policy search method

[122]. We discovered that both Q-learning and SARSA [123] are well-known and have the advantage of using a good policy as a short-time convergence learning process [124]. Thus, this study assumes Q-learning and the SARSA branch of ML for problem-solving.

The preceding discussion focused on the reinforcement learning framework. This study will propose using Q-Learning and SARSA to solve the bike-sharing relocation problem as a Capacitated Vehicle Routing Problem (CVRP). As shown in next Sections, it may be able to be used in the relocation bike-sharing problem to improve efficiency systems.

6.3 Problem definition

This study investigates bike-sharing relocation for a network that the system has used for relocation. The bike-sharing system's network includes a depot where all trucks start and return after visiting bike-sharing stations that service n stations and v trucks used for relocation. Each truck is assumed the same to have the same capacity. The number of bikes to be rebalanced is known before the start of the relocation process.

Furthermore, the demand for the number of bikes necessitates repositioning for relocation. During the relocation process, the demand at each station can be changed to be smaller or larger. The bike relocation is necessary to balance supply from a full station to an insufficient station. A depot can store the bikes for repair before relocating them at each station. The trucks for relocation can only visit the station once and must return to the depot after visiting each station, subject to the truck's capacity and the number of bikes required for rebalancing.

The problem considers the routing problem for operating trucks and the number of pickups and drop-off bikes at each station in the system to optimize the shortest possible distance between operating trucks. This problem can be traced back to the CVRP related to operations research and manages logistics operations, vehicle routing, and delivery schedule. The problem is considered on a complete directed graph G = (N, A), where the nodes are the bike-sharing stations, and the set is N =1,. . . ,n, the set of arcs or distance between stations is $A = \{i, j \in V, i \neq j\}$, and the station number 0 is represented to the depot. The demand for bikes for relocation was assumed to be known before the operation, and it can also be estimated using historical record datasets [125, 126]. For bike relocation, each station in the system from the depot or stations with more bikes more than customer demand to shortage stations. The trucks are considered to relocate; the truck must depart from the depot and return to the depot after visiting some or all stations, which is the selective delivery problem [127].

Many researchers have proposed solutions to the problem of rebalancing. However, in order to perform their tasks effectively in dynamically changing environments, systems must be adaptable. As a result, the rebalancing problem should be adapting intelligent agents to provide information and make decisions based on the shortest path or the lowest total cost. This study proposes the intelligence to an application for solving the rebalancing problem as machine learning. Nazari et al. [120] use the RL to solve the VRP, and the results show that the RL also does not require a distance matrix, and only one feed-forward pass of the network is required to update the routes. Thus, in real-world situations, we may need to use a distance matrix to determine which truck route to take in order to minimize total route cost. This study employs the Q-learning and SARSA to determine the best route by calculating the distance matrix. The online system is based on real-time data.

6.4 Methodology

Machine Learning is a type of artificial intelligence that uses software to predict outcomes more accurately. It was developed via pattern recognition learning. They involve learning information and then using computer programming to create an algorithm to predict data. There are three types of "training data" or "input dataset," namely: supervised learning, unsupervised learning, and reinforcement learning.

Reinforcement Learning (RL) has been a decade-long AI trend for better integration with statistics, mathematics, and optimization. RL differs from both supervised and unsupervised. The supervised method is trained using the data label, whereas the unsupervised method does not use the data label to identify patterns in the data. In operation research, reinforcement learning has had a strong interaction with methods for learning with estimates from classic variables. The solution to the bike-sharing relocation problem is to use reinforcement learning, which allows agents to learn in an environment through trial and error by providing feedback on actions and experiments. The essential agent behavior is to learn and take the best course of action for the highest reward, thereby completing the purpose application. The core elements of reinforcement learning are the environment state, the agent's actions, and the environmental rewards. The majority of Reinforcement learning is Temporal-Difference (TD), which combines the Monte Carlo measurement method [123] with a model-free experience and the advantages of dynamic programming. Q-Learning is a popular reinforcement learning method that supports information flow. It is not necessary to have prior knowledge (the value function) to participate in Q-learning. Meanwhile, the basic function of Q-learning is that the agents learn by trial and error for the best value of reward by policy from interactions. Additionally, to solve dynamic problems, Jang et al. [128] describe a comprehensive classification and applications in Q learning algorithms using the Markov Decision Process (MDP).

6.4.1 Q-Learning algorithm

Q learning is defined as an off-policy temporal difference (TD) of reinforcement learning. Reinforcement learning is described by MDP. In MDP, there are optimal policies, which have been defined as a maximum reward, and which described the value function with sequential action by the mathematical approach to establish the Bellman equation. Thereafter, the bellman is used to solve the Q-learning, MDP component shown in Fig 6.1 as follow:

1) The state (S) is an agent's observation of the situation set.

2) The action: The possibility that an action (A) will occur in the state S.

3) The state transition probability value in the matrix in equation 6.1 : the agent moves from one state to each state by taking action is the numerical representation of the state transition probability, where $P^a_{ss'}$ is the probability to move to state s' in the matrix P :

$$P_{ss'}^{a} = P\left[S_{t+1} = s' \middle| S_t = s, A_t = a\right] \tag{6.1}$$

4) The reward: The environment can be learned the information from the agent, and it is called a reward when s state and action occur at time t, where $R_{ss'}^{a}$ is the reward function, t is the time, and E is the expected reward value in equation 6.2:

$$R_{ss'}^{a} = E\left[R_{t-1} \middle| S_t = s, A_t = a\right] \tag{6.2}$$

5) Discount factor: The agent acting in the state. The discount factor was defined as the responsible action from the compensation operation. During the learning process, the reward value decreases due to the depreciation concept, which is between 0 and 1. Thus the amount the agent receives over time is reduced.

6) Policy or Control program: The agent decides the action by using the policy, where π is the probability in equation 6.3 when the agent selects an action in s state of the at time t. Then RL learns until it finds the best policies that represent an optimal policy.

$$\pi\left(a \middle| s\right) = P\left[A_t = a \middle| S = s\right] \tag{6.3}$$

The Q-function value is integrated into the policy for all actions. This work applied the Q-learning algorithm to find the shortest path problem. We defined the current truck for bikes repositioning location as the state, and the action is to immediately select the next node. The proposed model is based on the developed model of Jeon et al. [129], and we consider truck capacity when relocating each station in advance. To solve the RL problem, the agent must learn to take the best action in each possible state that it encounters. To accomplish this, the Q-learning algorithm learns how much long-term reward it will receive for each state-action pair (s, a), which is known as the state action-value function, and this algorithm represents it as the function Q(s, a), which outputs the return the agent will receive when acting in state s, and according to the policy indicated by the function Q until

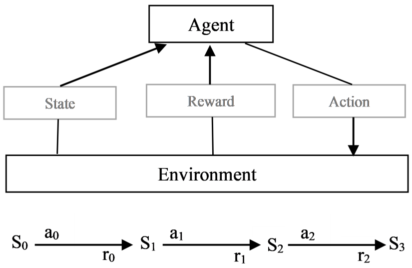


FIGURE 6.1: The flow of the Markov decision process (MDP)

Q-Learning (off-policy TD control)

Initialize $Q(s,a), \forall s \in S, a \in A(s)$, *arbitrarily, and* $Q(terminal - stage, \cdot) = 0$.

Repeat for each episode:

  Initialize s

  Choose a from s using policy derived from Q (e.g., $\varepsilon - greedy$)

  Loop for each step of episode:

    Act a, observe reward r, and next state s'

    $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma Q \max_a(s', a') - Q(s,a)]$

    $s \leftarrow s'$

  Until termination condition

FIGURE 6.2: The pseudocode of Q-Learning algorithm.

the episode ends. The pseudocode of the Q learning algorithm is shown in Figure 6.2 .

6.4.2 SARSA algorithm

The SARSA algorithm is an on-policy TD control that learns the policy value for decision making, similar to the Q learning algorithm. In contrast to Q-learning, which performs one policy and evaluates another, the agent learns from the value of state-action pairs on the current policy. The SARSA, except for action, comes from fact for updating the action value, and SARSA needs to update the Q values at previous state-action pairs at a time step, and the learning agent learns the value policy to the current action from the current policy, which is shown in Figure 6.3.

SARSA (on-policy TD control)

Initialize $Q(s,a), \forall s \in S, a \in A(s)$, *arbitrarily, and* $Q(terminal-stage, \cdot) = 0$
Repeat for each episode:
  Initialize s
  Choose a from s using policy derived from Q (e.g., $\varepsilon - greedy$)
  Loop for each step of episode:
    Act a, observe reward r, and next state s'
    $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma Q(s',a') - Q(s,a)]$
    $s \leftarrow s'$; $a \leftarrow a'$;
  Until termination condition

FIGURE 6.3: The pseudocode of SARSA algorithm.

6.4.3 Implementation

In this study, we defined the current truck for repositioning the bike's location as the state and action by selecting the next station immediately. The model was developed from Jeon et al.'s model [129]. When relocating each station, we plan ahead of time in terms of truck capacity. To solve the RL problem, the agent must learn to take the best action in each possible state that it encounters. The Q-learning algorithm learns how much long-term reward it will receive for each state-action pair (s, a) for this purpose. This is known as a state action-value function, and this algorithm represents it as the function Q(s, a), which outputs the return the agent will receive when acting in state s, and according to the policy indicated by the function Q through the terminal or ending the episode. The following is how the notation in bike-sharing relocation is defined in the learning process:

j	Current truck destination
k	Truck location of the current station
(k,j)	The tuck state when a k current station and t destination station
A(k,j)	The candidates set of action for the next station when a truck in state (k,j)
a	The action corresponds from select to next station in A(k,j)
γ	The discount factor ($0 \leq \gamma \leq 1$)
r[(k,j),a]	The penalty as a total distance at state (k,j) from the current station to the next station (a)
Q[(k,j),a]	The expected discounted total distance at state (k,j) when selecting action a from current station to destination station

The learning procedure typically consists of 4 steps, which are as follows:

1) Staring with the default element (k,j) value for the Q table, the set of truck locations for bike-sharing relocation at the bike-sharing station.

2) The number of bikes assigned to each station for relocation is determined by a pair of original and destination probability matrices. Then, under truck capacity, select the next bike-sharing station node.

3) Update the Q-table while the destination is at t and the next station is at k. Calculate the probability action (a) for station immediately selection at state (k,j) using equation 6.4:

$$p\left(a|\left(k,j\right)\right)=\frac{\rho^{\hat{Q}\left[(k,j),a\right]^{-1}}}{\sum Q\in A\left(k,j\right)\rho^{\hat{Q}\left[k,j,a\right]^{-1}}} \tag{6.4}$$

Where a $\in$ A(k,t) and ρ is a positive constant and $\hat{Q}_n\left[\left(k,j\right),a\right]$ is the estimated value of Q[(k,t), a].

4) Updating the Q-table in equation 6.5 when the current truck arrives at the next station in accordance with the equation. After the agent loop at the end of the episode, update the Q-table using Monte Carlo and select the using epsilon-Greedy ($\varepsilon-$ Greedy). The Q-table decides where the best direction should be. The probability that has never been used before must be used ε- Greedy for tolerating has occurred.

$$Q[(k,j),a](i)=Q[(k,j),a](i)+\alpha(r_{i+1}+\gamma\max Q[(k,j),a](i+1)-Q[(k,j),a](i)) \tag{6.5}$$

Where Q is the Q-factor, i is the current iteration. α is the learning rate used to update the current Q-factor. r is the one-step reward acting at state (k,j); in this work, r that is the distance between stations for all possible routes is preferred. γ is the discount factor that determines the weight of rewards ($0\leq\gamma\leq1$).

$a\max Q\left[\left(k,j\right),a\right]\left(i+1\right))$ is the maximized Q-factor when the system state has the highest Q value during all actions probability.

6.5 Experimental

6.5.1 Experiment setup

Implementation in python demonstrated the Q-learning and SARSA form of reinforcement learning. The model was run on an Intel core i7 CPU3.80GHz PC personal computer with 16 GB of RAM. We generate a numerical experiment data set based on set A of illustrated instances from Augerat [130] that includes the location of bike-sharing stations and depots and the demand for the number of bikes for rebalancing at each station. We create the distance matrix through the origin and destination pair probability in Euclidean distance as a reward in Q-learning by convert to negative value, while bike-sharing systems decide whether to improve the efficiency routing following the capacity truck constraint. The corresponding rewards and constraint capacitated trucks are used to evaluate the performance of those algorithms. The Q-Learning policy returns the highest total shortest path. We conduct a comprehensive investigation of the proposed strategy algorithm's solution quality, comparing the Q-Learning and SARSA algorithms to popular metaheuristics algorithms in vehicle routing problems such as Genetic Algorithm and ABC [131].

6.5.2 Results

The experiment's goal is to find a minimum truck routing in terms of total travel distance, taking into account the learning method. Also, to obtain the total shortest distance route of trucks for relocation bike-sharing from the learning method. We ran our tests on the same problem sizes but with a different algorithm that is a metaheuristic consisting of GA and ABC in comparison to ML such as Q-Learning and SARSA. That experiment was repeated ten times for the GA and ABC [111, 132, 133] The analysis of the exploration and exploitation trade-off for Q-Learning was sufficient state and action to find an efficient solution by using Euclidean reward.

The algorithm that presents the results that have the potential to be implemented is shown in Table 6.1- 6.2.

This study describes a reinforcement learning method for solving the bike-sharing problem that employs Q-learning and SARSA in the learning process. The optimization for minimizing the total distance of the route truck operation and the sequence of stations for rebalancing at each bike-sharing station. Also considered in this work are capacitated trucks for rebalancing bike-sharing systems.

To compare the performance of results with a short distance between the reinforcement, heuristic, and Mathematical programming models, only two instances of SARSA result in a shorter distance than the optimal value, while only one instance of Q-learning results in a shorter distance than the optimal value. On the other hand, Q-learning and SARSA used shorter processing times than GA and ABC. ML found that the results were far from optimal. However, the mathematical programming model's optimal value has the disadvantage of taking significantly longer than metaheuristics in terms of processing time or execution time [134, 135].

When compared to the Genetic Algorithm, the results showed that the proposed algorithm, as a Q-Learning algorithm, can develop outstanding bike-sharing relocation problems. When compared to GA and ABC, Q-learning and SARSA performed well, especially SARSA, which was successfully used. The proposed algorithm can serve as a guideline for a future related problem.

Table 6.1: The performance comparison of routing in bike-sharing relocation problem

Instances	**Optimal value**	**GA**	**ABC**	**Q-learning**	**SARSA**
B-n41-k6	829	1073.89	854.20	1276	1140
B-n56-k7	707	996.43	750.50	831	602
B-n63-k10	1496	1954.42	1703.52	1572	1441
B-n64-k9	861	1248.76	1122.36	1139	1103
B-n66-k9	1496	1722.7	1508.04	1296	1507
B-n67-k10	1032	1410.36	1194.36	1100	1205
B-n68-k9	1272	1541.94	1396.67	1313	1431
B-n78-k10	1221	1597.45	1533.75	1382	1346

Table 6.2: The performance comparison of CPU times (second)

Instances	**GA**	**ABC**	**Q-learning**	**SARSA**
B-n41-k6	44.65	27.098	1.77	1.855
B-n56-k7	42.82	46.11	1.439	1.808
B-n63-k10	45.49	60.748	2.194	2.829
B-n64-k9	48.06	56.821	2.058	2.631
B-n66-k9	51.76	61.259	2.382	2.65
B-n67-k10	47.8	67.727	2.12	2.784
B-n68-k9	48.57	65.335	2.406	3.164
B-n78-k10	53.29	80.890	2.252	2.905

6.6 Conclusion

This study demonstrates how to use Q-learning and SARSA in the learning process to solve the bike-sharing problem using reinforcement learning. Optimization for minimizing the total distance of the route truck operation and the sequence of stations for relocation at each bike-sharing station. Also considered in this work are capacitated trucks and feet of tucks for rebalancing bike-sharing systems. The results show that the proposed algorithms, Q-Learning and SARSA, can develop outstanding bike-sharing relocation problems compared to GA and ABC. Q-learning and SARSA outperform the best-known meta-heuristic methods such as GA, ABC, and they were successfully used in terms of routing and CPU times. The proposed algorithm can serve as a guide for future-related issues.

In addition, future work should implement real-time services, consider how many times the bike should be relocated in terms of the economy, and implement a large-scale bike-sharing system.

Chapter 7

Simulation for Operation and Cost Optimization

7.1 Overview

Bike-sharing relocation is a method that can improve bike-sharing systems. This method is commonly referred to as rebalancing. Owing to various demands, bike usage can suffer an imbalance. Some stations have a high demand for bike renting, which can result in bike shortage at such stations; whereas, at some stations, users like to return the bike, which might result in a space shortage for the users which are next in line. To balance the network, operators can plan truck routing to maintain the bike supply by refilling the bikes at each station and to manage bike availability in docks by picking them up at each station [50]. Here, the operation trucks can depart from the same station or different stations depending on the network and operating costs of the service of each provider.

There is a lot of competition in the current situation. In order to operate the business, profits and customer satisfaction are the main aspects to be considered. bike-sharing service business is one that, in addition to providing people with a mode of transport, affects the world in terms of reducing carbon dioxide emissions from non-motorized vehicles or electric power vehicles. To make this business sustainable, it is necessary to improve the quality of the service system that can meet the needs of customers while considering the operating costs that affect the profits for the organization incurred.

Several previous research works have conducted studies to solve bike-sharing rebalancing problems: Forma et al. [136] proposed a 3-step mathematical program based on a heuristic algorithm for solving large-scale instances of a static bike rebalancing problem, which aims to minimize the total traveling distance. Lin and Liang [137] presented a model on the Arena simulation software to obtain the optimal number of relocations for minimizing the waiting time by using the O-D probability matrix, arrival time, and rental time. Chen et al.[138] presented the pricing strategies of bike-sharing to adjust the price and demand based on users' travel behavior, which may increase their participation in using bikes. Soriguera et al.[70] proposed a tool to support the decision-making regarding daily operation of planning. They implemented a MATLAB programming code to access and relocate adjustments. The simulator assesses the performance of a bike-sharing system before use or for those in performing the analysis of some aspects that are particularly difficult to measure; others can be analyzed from simulations. The simulator can also be used as a productivity tool in the planning process of the bike-sharing system.

There are several methods such as truck routing or pricing strategies that can solve bike unbalancing problems. However, there is only one solution that can solve the virtual bike balancing problem: the simulation method that is widely used in various fields. It can also solve complex problems. This research proposes to solve the problem of bike rebalancing using a simulation method that can meet customer needs and obtain the most profitable authorized provider.

However, engineering design systems is an important part of making the system run more efficiently. This consists of the various models such as the mathematic model, diagram model and schematics, etc. The simulation model is also a type of the model that can be support of the complex systems [139] and help decision-making by experimenting with different policy approaches. The disadvantage of the simulation modeling is the cost may appear high when designing or planning among alternative with trial implementation for decision the outcomes. Simulation is the process of designing a situation or behavior of a real system by using computer programs.

Resource limitation is the area in urban affect that cannot improve the service by building new stations or redesigning stations for supporting demand. Increasing the number of bikes in the system to meet the demand could affect investment and maintenance costs. Therefore, this research proposed the management the currently utility. This work considers the sustainability of the proportion of the initial number of bike available in the system for balancing demand and supply to maximize profit. This is undertaken in order to satisfy customers, reduce the number of customers lost, return to the use of the bike-sharing service, and improve quality of service then became to mouth word. Thus, for cost optimization and improved resource allocation that is intended to balance demand and supply in the bike-sharing system, this work proposes a simulation that represents the system and support of complex problem.

7.2 Determining bike-sharing travel patterns

High dynamic movements of users cause an imbalance between demand and supply of bikes. The bike-sharing provider could operate the systems to satisfy users. Data mining is a decision-making method used for the operation of bike-sharing systems. It is the process of analyzing large amounts of data to discover hidden patterns and relationships. Several research works have applied data mining to many cases. For instance, it has been applied to the business that supports the decision-making of executives regarding this subject. Moreover, data mining is the process of defining patterns and correlation in a large dataset using statistical techniques and artificial intelligence; this helps to explore and analyze raw data and convert it into potential information. Citi Bike has been providing the historical data since 2013. This work uses trip data from Jersey City's Citi Bike in 2020. The datasets consist of the trip's start day and time, trip duration in seconds, trip's stop time and date, name of the station of departure, name of the station of arrival, station ID, station latitude and longitude, bike ID, user type, gender, and year of birth of the user.

This study explores various parameters for improving bike-sharing services. In particular, the information provided is essential in understanding critical points of the system and bike activities on the operation [140, 141]. The station activities were identified by analyzing the travel pattern on the system during rush hour. First, activity patterns of the dataset were determined. The Citi Bike data facilitated the in-depth analysis of trip patterns. The data helped investigate the changes in daily usage based on season and time. The usage of bike stations and the bike rental demand for each month were analyzed. Furthermore, we analyze different cycle lengths as a month and a day.

Figure 7.1 shows the pattern of daily bike usage in 2020. The difference in weekday and weekend patterns can be clearly seen from the figure. On weekdays, the demand is high during morning, and there is a lower demand pattern during the afternoon, which is accompanied by a higher demand during the rush hour in evening. On weekends, the trend shows a high demand at around 9 AM, and the demand continues to rise until 5 PM, and then gradually reduces until prime time. A monthly bike usage analysis was conducted to reflect the needs of customers' monthly bike use. This is beneficial for planning in accordance with demand and supply. In Figure 7.2, the results show the pattern of monthly bike usage in 2020. It can be seen that the period from June to October was highly active. This was found during summer, which may affect the demand for travelling [142] where the environment is suitable for the use of bikes. Winters and rainy seasons may not be convenient for a heavy bike use; this may reduce its use as compared with summer.

An analysis of the customer's bike usage characteristics in term of distance and time was made using a frequency distribution with a graph. Figure 7.4 shows that most of the users traveled short distances, within 2 miles and Figure 7.3 shows with the most ravel time of less than 20 minutes. This means that most customers use the bike for short distances and shorter periods of time.

To find out patterns in data, we used circles for each station on the map. The radius corresponds to the number of trips at a particular station. We applied logarithmic scaling to the total number of trips. Set the exact radius size color represents the ratio of inbound and outbound journeys per station is a gradient station with all incoming travel are stations with all incoming travel, as shown in Figure 7.5. Following that, we analyzed the number of users for each station in order to provide resource management to suit user needs. Figure 7.6 shows the stations that are frequently used.

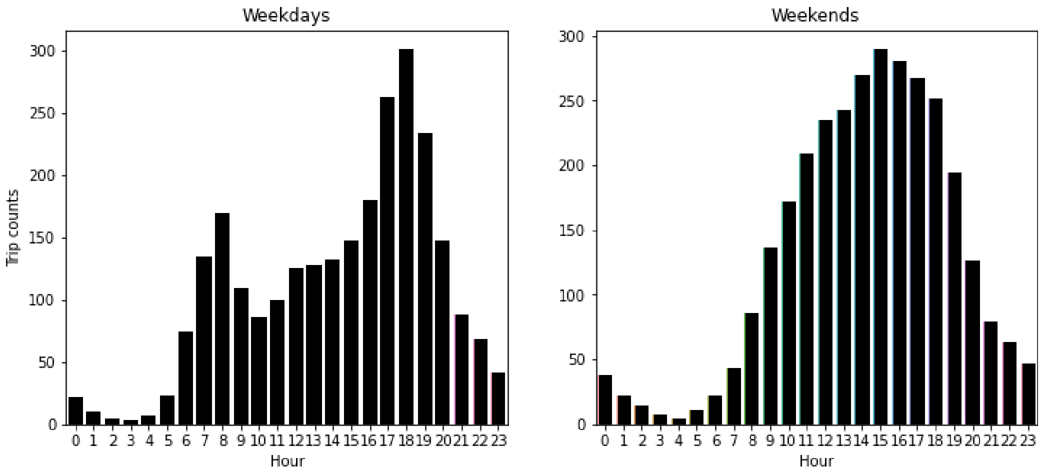


FIGURE 7.1: Trip count by hour.

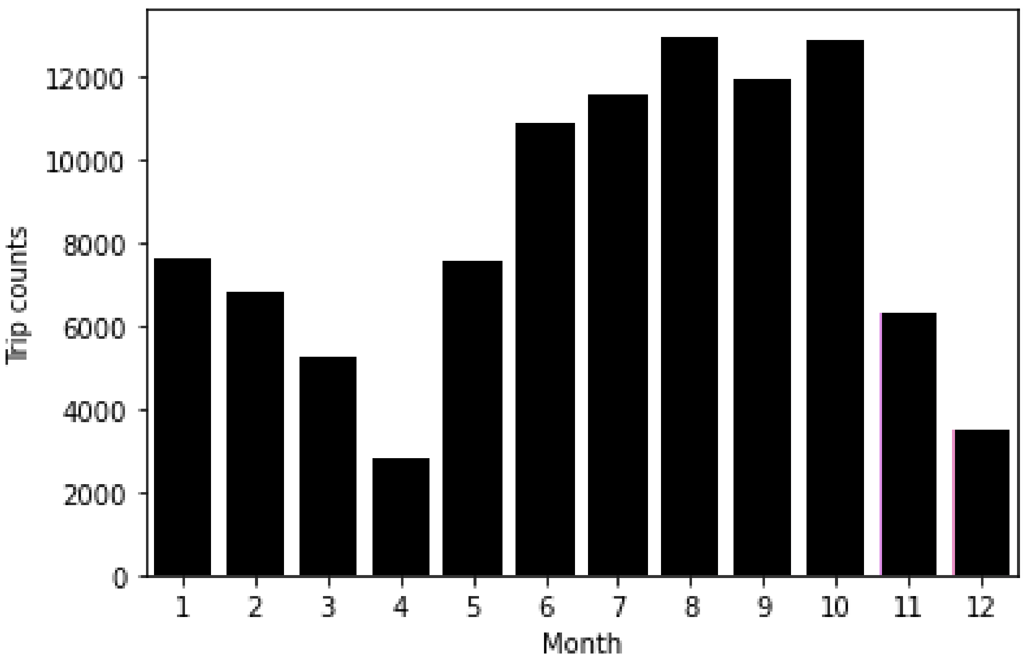


FIGURE 7.2: Trip count by month.

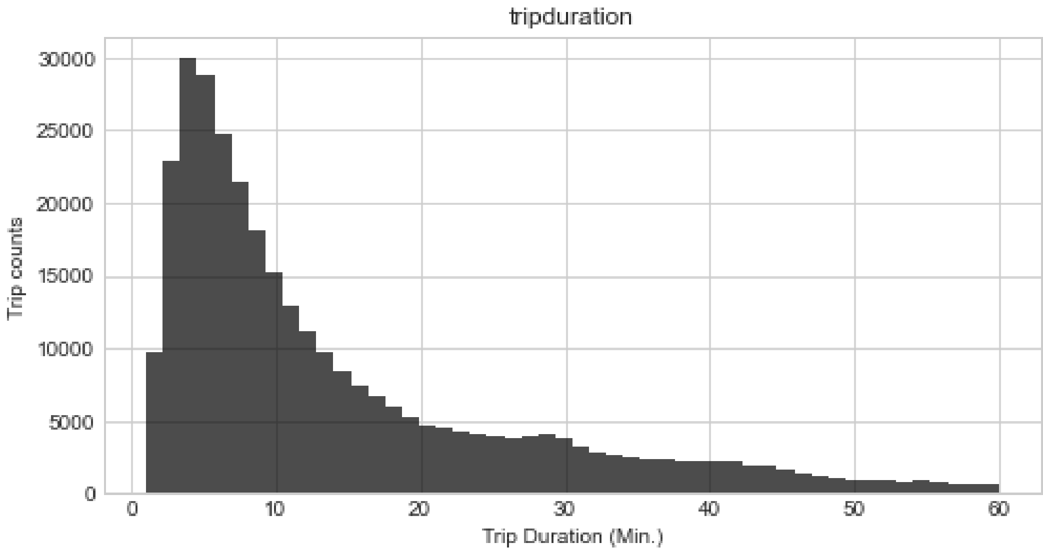


FIGURE 7.3: Customer usage cumulative frequency in trip duration (minutes).

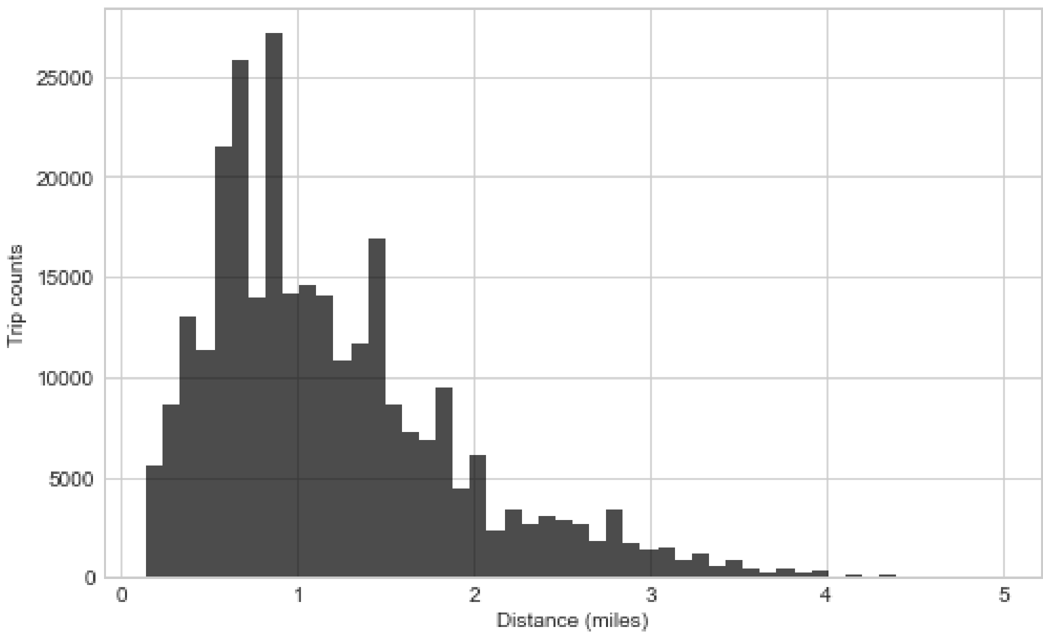


FIGURE 7.4: Customer usage frequency in distance trip (miles).

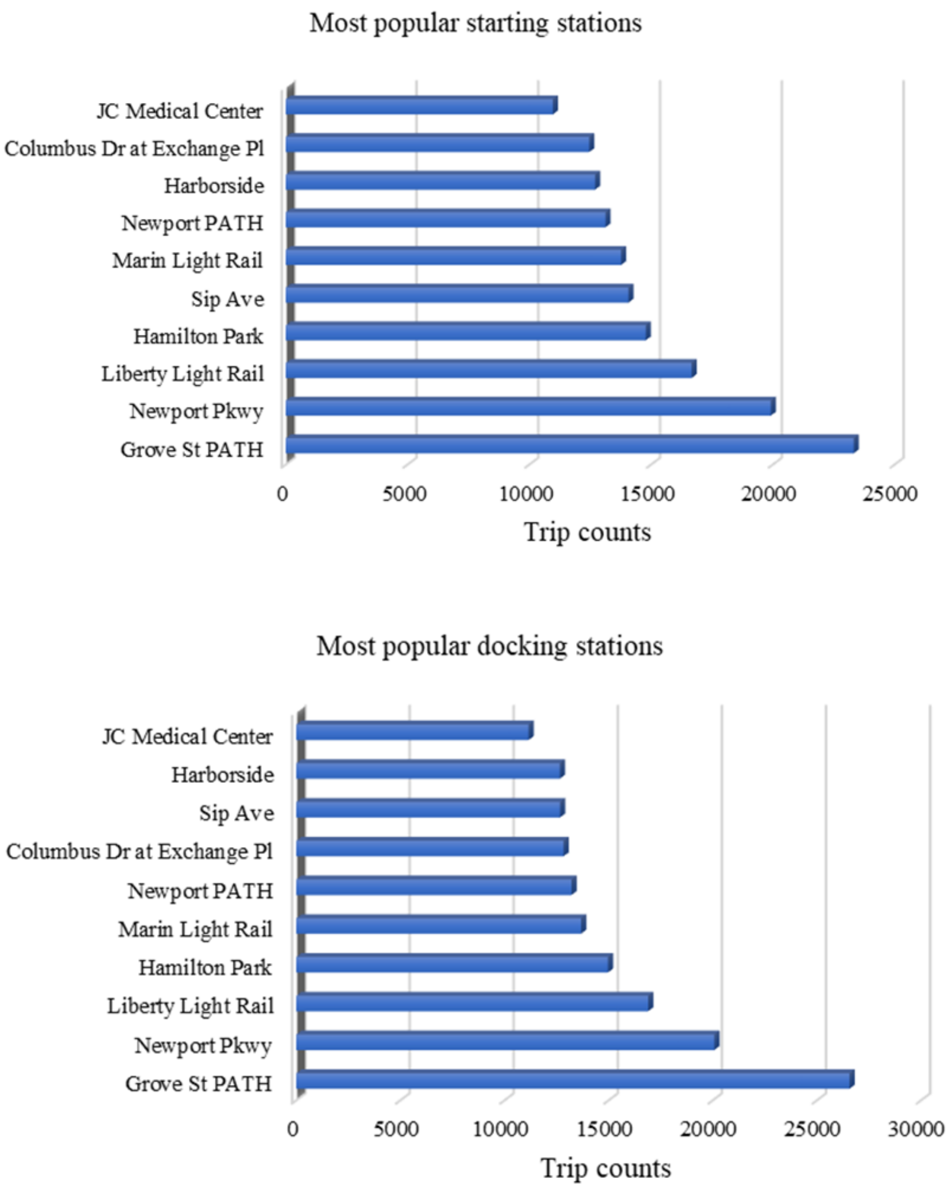


Figure 7.5: Most popular starting stations (above) and docking stations (bottom).

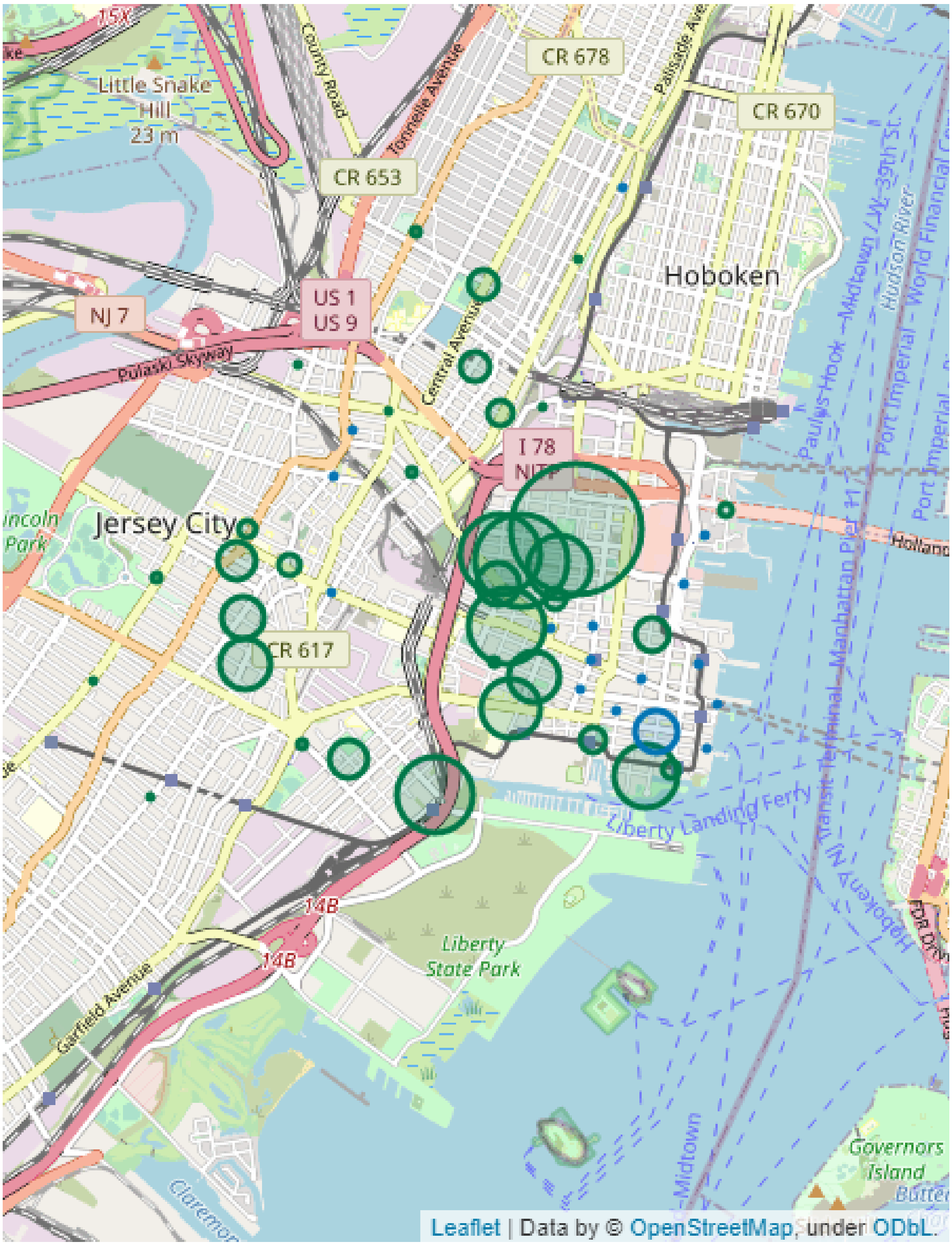


FIGURE 7.6: Map of bike-sharing station; green color shows high demand, and blue color shows low demand.

TABLE 7.1: Analysis of the data's distribution model.

<table>
<tr><td colspan="3"></td><td>Renting</td><td>Return</td></tr>
<tr><td>Exponential parameter.</td><td colspan="2">Mean</td><td>13.3006</td><td>12.6261</td></tr>
<tr><td rowspan="3">Most Extreme Differences</td><td colspan="2">Absolute</td><td>0.444</td><td>0.292</td></tr>
<tr><td colspan="2">Positive</td><td>0.23</td><td>0.157</td></tr>
<tr><td colspan="2">Negative</td><td>-0.444</td><td>-0.292</td></tr>
<tr><td>Test Statistic</td><td colspan="2"></td><td>0.444</td><td>0.292</td></tr>
<tr><td rowspan="3">Monte Carlo Sig. (2-tailed)</td><td>Sig.</td><td></td><td>0.000</td><td>0.000</td></tr>
<tr><td rowspan="2">99% Confidence Interval</td><td>Lower Bound</td><td>0.000</td><td>0.000</td></tr>
<tr><td>Upper Bound</td><td>0.000</td><td>0.000</td></tr>
</table>

7.3 Simulation

7.3.1 Data Set

We used the user trip data to simulate users and real user travel data based on the location information, from where the real user wants to start the trip or where the docking station is located. To best reflect the actual situation, we used station coordinates. By analyzing the previously mentioned data, we found 7 AM to 10 AM to be the peak hours. Owing to these reasons, this period was used for simulation. Using real data from the Jersey City's Citi Bike in 2020, we calculated the inter-arrival rate of demand at each station [143, 144], as shown in Figure 7.8, along with calculating the destination's demand rate and when the user wants to return the bikes, as shown in Figure 7.8.

7.3.2 Validate input data

The verification that ensures the implementation and model are correct [145]. The distribution fitting by Kolmogorov-Smirnov Goodness-of-fit Test was used to test for exponential distribution [146] by using the SPSS Program for testing. We choose the represented data from Grove St PATH station on 1st January 2020 to test whether the input data is correct. The results show that the analysis of the renting bike rate and the return rate of bikes seen in the probability distribution of the data shows that the data characteristics are exponentially distributed at a significance level of 0.01 as shown in Table 7.1.

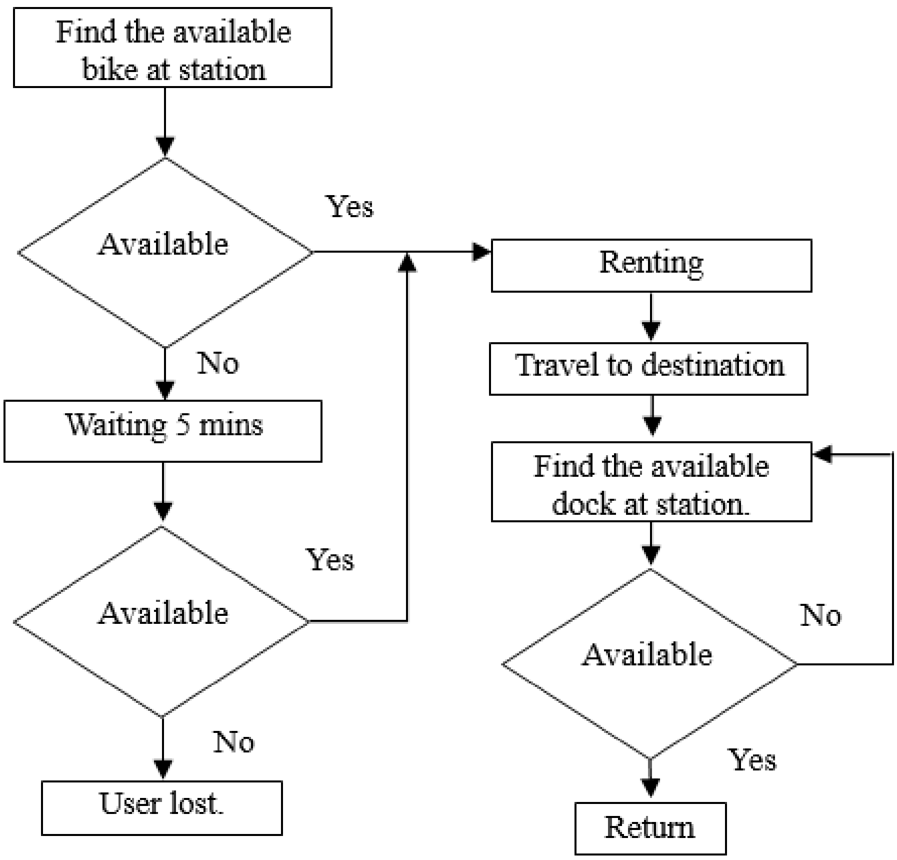


FIGURE 7.7: The flow of customers through the bike-sharing process for renting.

7.3.3 Simulation model

The simulation model's objective is to assist in making decisions regarding balancing bike-sharing, revenue, operation costs, and opportunity costs, as well as the key variables that need to be observed when calculating profitability. In this research, we used discrete-event simulation programming with Python. The balancing event has attributes that specify the queue station, response terminal, and number of bikes ordered for rebalancing. Conversely, it may be more than one station to track all of them without specifying which station is being simulated. The simulation began by recording the changes in the value of the interest variables.

Consider bike-sharing event for a station. A bike-sharing system needs to check if someone is present or not. If users have to wait too long to return their bikes, they get a money refund, but if no one is in the queue, the bike's presence or absence is checked. If there is an idle bike, the user pays to rent the bike. When the number of bikes is less than two, a balancing act is conducted to transport bikes from depot to station. The system calculates the wait time and checks if the user has left due to a high waiting time and then creates a time between the time the bike is rented and till it's returned. When the user finishes the trip and arrives at the station, the system needs to check if another user is waiting to rent the bike, as there are no bikes at the station, or the dock is available for the user to return. This is done in order to retain users since costs increase due to loss of business opportunities.

The rebalancing policy is simple. This model dynamically distributes the bike to avoid the station being full or empty. We assumed that there is a centralized control: When no port is available, a station sends a signal to transport the bike back to the center. In organizing the balancing event, operating and fuel cost are calculated.

The number of bikes that can be refilled at the station before the rush hour is a problem to optimize for each station. For this calculation, the expected number of bikes coming in and going out of the station during each minute of peak hours, capacity, and number of bikes can be used. The model for simulating bike rebalancing aims to maximize the profit and is defined as follows:

$$Max\,Z = \sum_{i,j \in N} R_i x_{ij} - \sum_{i,j \in N} C_i x_{ij} - \sum_{i \in N} pl.y_i \tag{7.1}$$

subject to

$$\sum_{i,j \in N} x_{ij} \leq Q \tag{7.2}$$

$$W_{ij} \leq T \tag{7.3}$$

$$x_i \leq q_i \tag{7.4}$$

$$x_{ij}\,,\ w_{ij} \geq 0\ for\, i,j \in \{1,...,N\} \tag{7.5}$$

where

R_i	:	revenue,
C_i	:	operating cost,
pl	:	loss-of-opportunity cost,
W_i	:	waiting time,
q_i	:	capacity of each station,
Q	:	bikes in system,
N	:	number of bike stations,
T	;	acceptable waiting times for the available bike or docking,
x_{ij}	:	when bikes travel at station i in the interval j, the value is 1; otherwise it is 0, and
y_i	:	when user cannot wait for an available bike or docking, the value is 1; otherwise it is 0.

We selected the five most popular stations for simulating bike pickups and returns for each station. Following that, the data was analyzed for distribution, which is scheduled to the arrival time and returns bike time rate, as shown in Figure 7.8. The nonstationary Poisson arrival process ($NSPP$) [147] was used for arrival rate by modeling real data. The interval time was calculated to be 20 minutes from the data analysis which highlights that the almost every trip duration is 20 minutes. This is shown in Figure 7.4.

This research work assumed that the revenue cost is \$3, operating cost is \$2, W 41, 13, 21, 26, 33, opportunity cost is \$3, and acceptable waiting time is 5 minutes. The first process began at 7:00 AM and ended at 10:00 AM, with the number of replications being 100 cycles. The result obtained is shown in Table 7.2-7.3, which indicates that refilling bikes before rush hour can maximize profit.

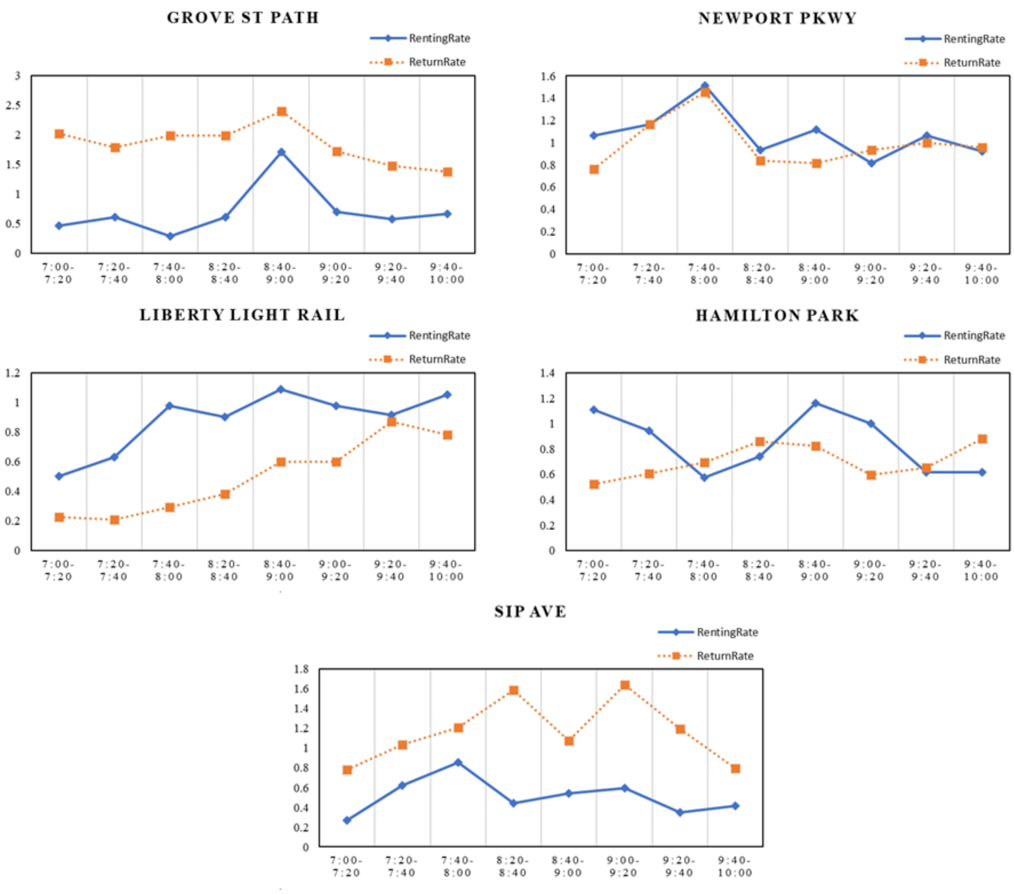


FIGURE 7.8: Arrival time rate and return bike time rate for each station.

7.4 Results and discussions

In this simulation, bike management options that suited the customer's needs were created. This was done by simulating the results with the help of the proportion of bikes in the system and the number of dock bike parking spaces that meet the needs of each station with different usage requirements. From the analysis of the data in Figure 7.8, it was found that the traffic of bike users renting the bike and the return rate of those who wish to return bikes at each station have different needs. Some stations require a lot of bikes for renting; however, some stations require more bikes to be returned at different times. This is an essential factor that impacts the imbalance including the available number of bikes available and the docks available which are supplied and demanded in the systems. This does not take into account the route allocation method or pricing strategies or analyzed factors influencing bike share systems. This work proposed the factors that influence the profit for operator rebalance bike-sharing systems and customer satisfaction. We simulated the trial and error by varying the proportion of bikes available per available dock bike in systems as shown in Table 7.2- 7.3.

Table 7.3 displays the scenario of bikes' proportion in systems for yielding the best profit with 70 percent of the bikes in the system and 30 percent of the vacant parking spaces in the system. A profit of 142.5 dollars can be found with only 19 customers lost, which is a small number. This implies that a loss of customers can be avoided without affecting customer satisfaction. The proportion of bikes available in systems and the available dock to return have also influenced the system profit. The amount of bikes available at each station is the high level that affects the current customer, rather than customers that are lost. Thus, the number of available bikes is low at each station that affects the current number of bikes is not enough for renting then lost customer. In addition, this simulation can also optimize the number of bikes at each station to be filled at each station for meeting the needs of customers, maximizing the profits and avoiding a loss of the customer. We propose that the providers should be concerned with the proportion between the number of bikes available and number of available docks in the systems. When there is a high amount of bikes or a smaller number of bikes, it affects the quality of bike-sharing systems for responsible customer and, consequently, total profit.

At times, the provider ignores to modify the solution for customer satisfaction because it increases the cost in the operation systems. However, customer satisfaction should be one of the top goals of an organization. This is because of the long-term benefits of having satisfied customers such as positive word-of-mouth reviews, customer loyalty, and sustainable profitability in the service sector [148, 149].

7.5 Conclusion

This research provides simulations to solve the balancing problem of bike-sharing in order to meet consumer needs and obtain more revenue. Based on data analysis for knowing critical points of the process in the system, it can be concluded that

Table 7.2: Results of simulation: Optimal number of bikes at each station

Proportion of bikes available /available dock bike in systems	**Grove St PATH**	**Newport Pkwy**	**Liberty Light Rail**	**Hamilton Park**	**Sip Ave**
90/10	35	12	19	25	30
85/15	28	12	19	25	30
80/20	36	10	12	25	24
75/25	34	12	18	20	17
70/30	32	9	14	11	28
65/35	29	6	13	18	21
60/40	24	12	14	15	15
55/45	21	5	11	17	20
50/50	16	11	12	13	15

Table 7.3: Results of simulation carried out to find the optimal the proportion of number of bikes and available dock in systems.

Proportion of bikes available /available dock bike in systems	**Cost ($)**	**Customer lost**	**Number of bikes to be moved**	**Profit ($)**
90/10	80	15	20	135
85/15	74	19	21	135
80/20	82	22	11	120
75/25	59	25	14	131.25
70/30	88	19	13	142.5
65/35	42	29	8	116.25
60/40	66	24	10	127.5
55/45	66	24	10	127.5
50/50	60	19	8	116.25

rental vehicles users' travel distances as short as 2 miles, travel time of less than 20 minutes, and the time pattern of use differs on weekdays and weekends. The imbalance problem between the demand and supply that arises in finding available bikes till renting, and finding available dock to return the bike can be solved by assessing user satisfaction based on arrival time rate and return bike time rate from historical data; this would also help to determine profitable benefits to the authorized provider by helping the provider to calculate the number of bikes to refill at each station before the rush hour. This research presents the solution to the problem by conducting simulation to maximize the profit. In this simulation, a reasonable number of bikes were presented that had to be refilled before the rush hour in terms of ratio of bike in the systems to allocation levels that might be required to find balance between demand and supplies of the bike-sharing systems.

Since this simulation is a short-term simulation and does not present the number of times bikes can be refilled at a station, in the future, we aim to analyze the number of appropriate bike or how many times should refill per day that would benefit both users and providers.

Chapter 8

Depot Location for a Bike-sharing Operation

8.1 Overview

According to research studies have used the relocation bike-share problem to improve the efficiency of bike-sharing schemes' relocation activity. The cost of routing vehicles represents the operating cost, with the objective being to relocate bikes in order to meet customer needs by optimizing the availability of bikes and docking slot at each station. Maximizing meeting the needs of customers can increase business operating costs. However, maintaining a high level of customer satisfaction can lead to an increase in customer loyalty as a result of word of mouth recommendations.

Generally, because vehicles must begin and end each trip at the depot, depot location needs to be considered as an element of the relocation problem.

Based on industrial location theory, proposed by Weber [150], location factors are labor costs, transportation costs and material costs, so that a facility should be located near customers, near the source of raw materials and in a central location.

The current work focuses on depot location and the routing of vehicles to distribute bikes among stations in a bike-sharing system. Cluster analysis was used to compare the total distances associated with a central depot location and total distances with a location determined by the clustering method.

8.2 Methodology

In this work, the detail of a bike-sharing demand forecast model is described and there is consideration of the optimal number of depots in terms of minimizing the total distance traveled, while operate bike relocation each station between only depot and the depots propose clusters using WK-means and the Elbow Method was compared.

The WK-means group of analysis techniques was used. This is because a cluster analysis method was required in which the number of clusters for analysis can be determined and the central position of each cluster can be precisely fixed with repeated iterations until the best position is obtained. According to the process of grouping with WK-means, an appropriate number of groups (K) must be determined. With WK-means, centroid points are calculated based on mean distances. Ultimately, for this research, the Elbow Method was used, which involves error measurement that focuses on the sum of the distances between the data and the centroid. The minimization of error smooths the slope of the curve to form an angle that resembles an elbow; the point of the elbow is known as the Elbow Point and is the point that indicates the optimal number of groups (K). Thus, the current study used the Elbow Method to find clusters in the dataset in order to calculate the optimal number of depots required for a bike-sharing operation. For clustering, the cluster analysis using K-means can be divided into four steps, as follows:

1) Group the data into k groups, with K random points as centroids.

2) Calculate the centroid of each group, so that the average of the centroids of group C is $\bar{x}_c$

3) Calculate the sum of distance of each unit to the center of the group, with the unit being located by assigning ESS (Error Sum Square) to the squared distance of each unit to the center of the group, as in Eq (8.1).

$$ESS = \sum_{i=1}^{k} \sum_{x_i \in C_i} (x_i - \bar{x}_{c(i)})^2 \tag{8.1}$$

$$\bar{x}_{c(i)} = \sum_{x_i \in C_i} \frac{x_i}{|C_i|} \tag{8.2}$$

where, C(i) is a group of units i and ESS is the sum of the distances of each unit in the group to the center of the group, which includes all groups. If any groups have a low ESS value, that means they are not difference.

To calculate the distance between the centroid and the data points (each station), the Haversine distance [151] was used to calculate the distance in Eq. (8.3).

$$D = 2.R.\arcsin\left(\sqrt{\begin{matrix}\sin^2\left(\frac{lat2-lat1}{2}\right)+\\ \cos\left(lat1\right)\cos\left(lat2\right)\cdot\\ \sin^2\left(\frac{lon2-lon1}{2}\right)\end{matrix}}\right) \tag{8.3}$$

where, D is distance, lat1 is the latitude of the first point, lat2 is the longitude of second point, and lon1 and lon2 are the longitude of first point and second point, respectively, and R is the mean radius of the earth (radius=6,371km).

4) Consider group transfer. Criteria for moving must be applied based on the values calculated in step 3. Unit i can be moved to the group that has the lowest ESS value. In the case group have no longer moved in the step 4, it means optimal. Nevertheless, in step 4 each unit has moved groups by moved in or moved out, in process will have to calculate the centroid each group that is to go back and do the second step again.

The Elbow Method is an error method for measuring the sum of distances between an object and a centroid [152], as described in Eq (8.4). This is known as the Within-Cluster Sum of Squares (WCSS), where C represents the cluster centroids and x is the data point in each cluster.

$$WCSS = \sum_{i=1}^{n} \sum_{P_i in C_i} distance(P_i, C_i)^2 \tag{8.4}$$

Wk-mean is a clustering technique that can compute variable weights automatically. Algorithm-generated variable weights indicate the relevance of variables in clustering. As a consequence, for appropriate clustering results, the procedure may be utilized as a weighted variable and a subtract weighted variable [**?**], as described in Eq (8.5)-(8.6).

$$ESS = \sum\nolimits_{i=1}^{n} W_i(x_i - \bar{x}_{c(i)})^2 \tag{8.5}$$

$$\bar{x}_{c(i)} = \sum_{x_i \in C_i} \frac{W_i x_i}{|C_i|} / \sum_{x_i \in C_i} W_i \tag{8.6}$$

The inventory of bikes to be relocated to each station is assessed by the initial number of bikes carried on a vehicle and starts from the depot. Routing is calculated simultaneously in order of the number of bikes to be moved. In other words, the depot is indirectly optimized by reducing the number of vehicles used and the total distance required for the relocation operation. The inventory and routing solutions are calculated simultaneously a sequence of inventory to determine the depot suitability.The determination of the location of the depot affects the total distance or handling time for relocating the bikes.

The inventory of bikes at the depot is affected by the initial number of bikes carried on a vehicle. In other words, the depot is indirectly optimized by reducing the number of vehicles used and the total distance required for the relocation operation. The inventory and routing solutions are calculated simultaneously a sequence of inventory to determine the depot suitability.

To assess the optimum distribution of bikes among the bike-sharing stations, SARSA Algorithms were used, and the results were compared truck routing for rebalancing. According to related research studies, it was found that SARSA is widely used in the variation of problems which solve a complex problem. Likewise, SARSA have generated high-quality solutions for optimization problems, as shown in Chapter 6. Thus, SARSA constitute another effective routing method, especially in the context of the vehicle routing problem.

8.3 Experimental results

This experiment, the first was to predict demand, the second was to determine depot location by comparison K-means and WK-means. Then the Elbow Method was used, and the final phase was to compare the total distance required for the distribution of bikes among stations calculated using K-means and Wk-means Method.

Trip history data from Citi Bike, in Jersey City, were used. This dataset consisted of user IDs, the time users started and ended the renting of a bike, the name of the start and end stations used and the latitude and longitude of the start and end station of each rental. In total, data from 75 stations in Figure 8.1 were used.

For determining depot location for a bike-sharing system, the key parameter was the level of demand for bike-sharing at each station. Latitude and longitude data were used to plot the requirement at each station on a map. A dataset for the period January 2020 to December 2020 was used for training, with the period January 2021 used for testing. It was to predict demand as a guideline for knowing the customer's needs in the future, which is also useful in future planning.

For calculating depot location and the number of depots required, the dataset was transformed so that it was appropriate for the clustering problem using the WK-means method; the data required was average demand at each station and stations' latitude and longitude coordinates. An iterative process was then followed until the location with the shortest distance was found. Allocating the different demands of each station representative with calculations by Haversine distance forgiven the weight of their distance. Calculating for all bike-sharing stations which demand pints serve each station based on total distance from each station in the system.

Determining optimal grouping is a fundamental problem addressed using the K-means method. No attempt was made to either determine the exact number of groups required, or to use trial and error to arrive at the optimum solution.

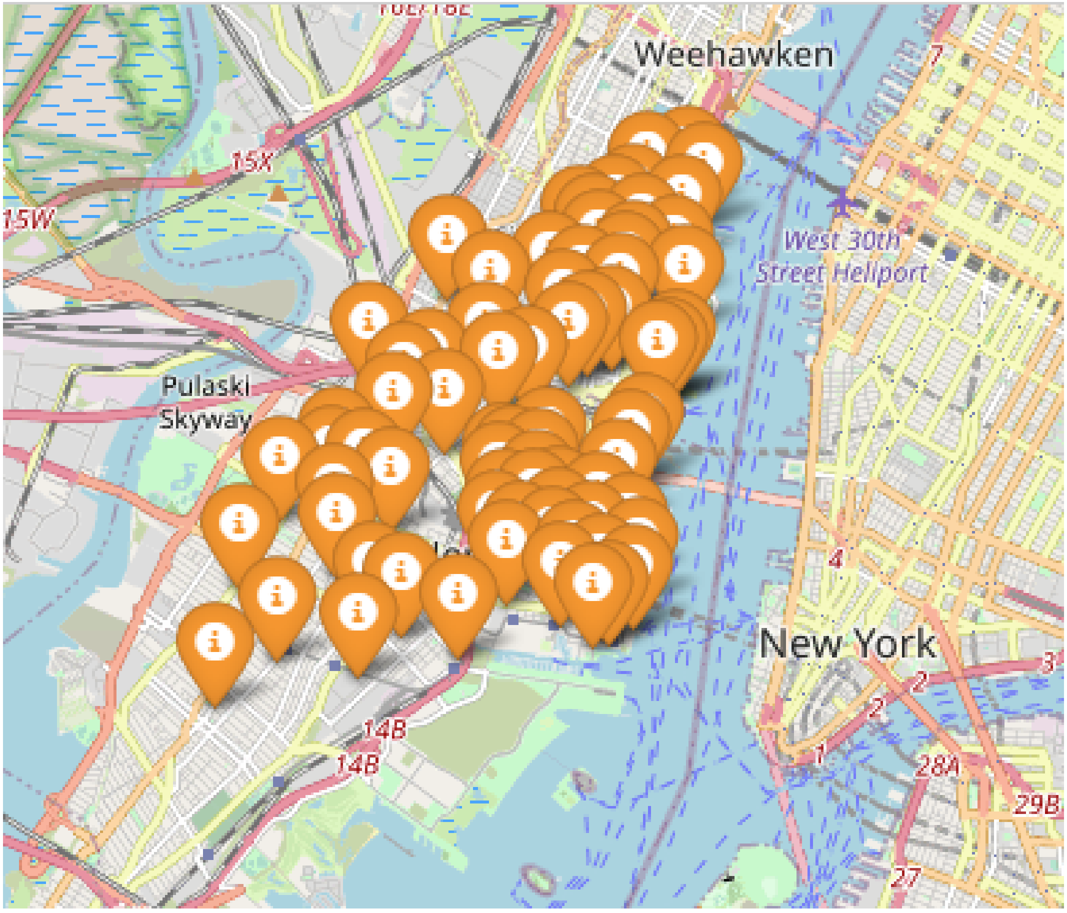


FIGURE 8.1: The coordinates of all bike stations.

This study used the Elbow Method to determine the optimal number of groups for analysis. This method calculates the sum of the distances between each object and the Centroid, as with the Within Groups Sum of Squares (WGSS), in order to minimize distance. As each iteration decreases the value of WGSS, the number of clusters increases and the number of members in each group decreases. The ESS value is the value that causes the curve, so identifying the appropriate number of groups (K).

The appropriate number of groups for a bike-sharing depot location using the Elbow Method, on the basis of the error value of the distance between locations and the Centroid or the Within Groups Sum of Squares (WGSS), was found to be four (k = 4), as shown in Figure 8.2 - 8.3. Historical data from Citi Bike in Jersey City,

TABLE 8.1: Comparison of routing with a different number of depots and methods.

Solutions	Distance (km)
one depot	262
4 depots using K-means	239
4 depots using WK-means	226

for the period January 2021, were used. These data featured average hourly usage at each station.

From the results of the analysis of the optimal position of the Centroid points in both groups, it was found that the coordinates by using K-means method of this point for the first group were 40.71933362, -74.07288881. Those of the position of the second group were 40.74480725, -74.03398164. For the third group they were 40.7364589, -74.0569353 and for the fourth group they were 40.72004832, -74.04254468. The coordinates for all groups are plotted on the map shown in Figure 8.4.

The coordinates of this point using the WK-means approach for the first group were 40.71967619, -74.04685883, according to the findings of the study of the optimal position of the Centroid points in both groups. The coordinates of the second group's location were 40.72074001, -74.0349121. They were 40.7255558, -74.06765504 for the third group and 40.74220135, -74.04334608 for the fourth. All of the groups' coordinates are plotted on the map in Figure 8.5.

This work aims to comparison the total distance during route for relocation bike-sharing between only a depot at the central of the whole station and the depot location from computation using K-means, WK-means, and Elbow method. A SARSA Algorithm was used for the routing of a vehicle used to distribute bikes among the stations. The procedure was carried out yielding results that varied between one depot and four depots.

Referring to the results in Table 8.1, shows the results the routing analysis, only one depot had an average total distance of 262 km. The optimum number of depots was found to be four, according to the results of clustering by K-means and the Elbow Method. This solution provided an average total distance of 239 km, the improve being 23 km. Therefore, the use of four depots would reduce total bike routing distance by 8.77 percent. Using the Wk-means technique, it was found that

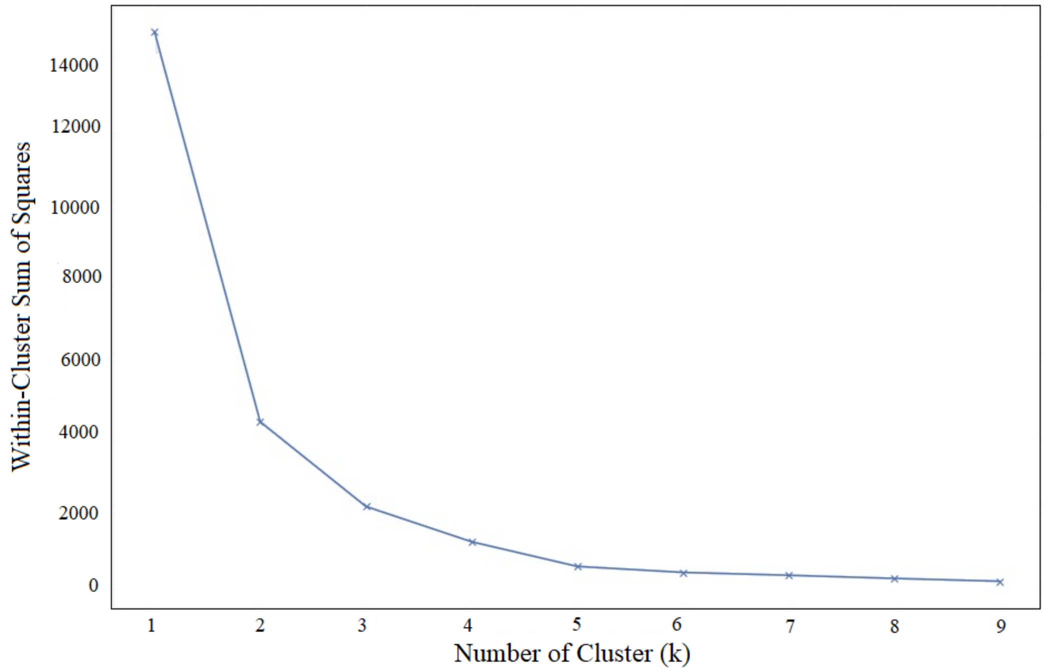


FIGURE 8.2: Selecting the number of groups for K-mean method using the Elbow Method.

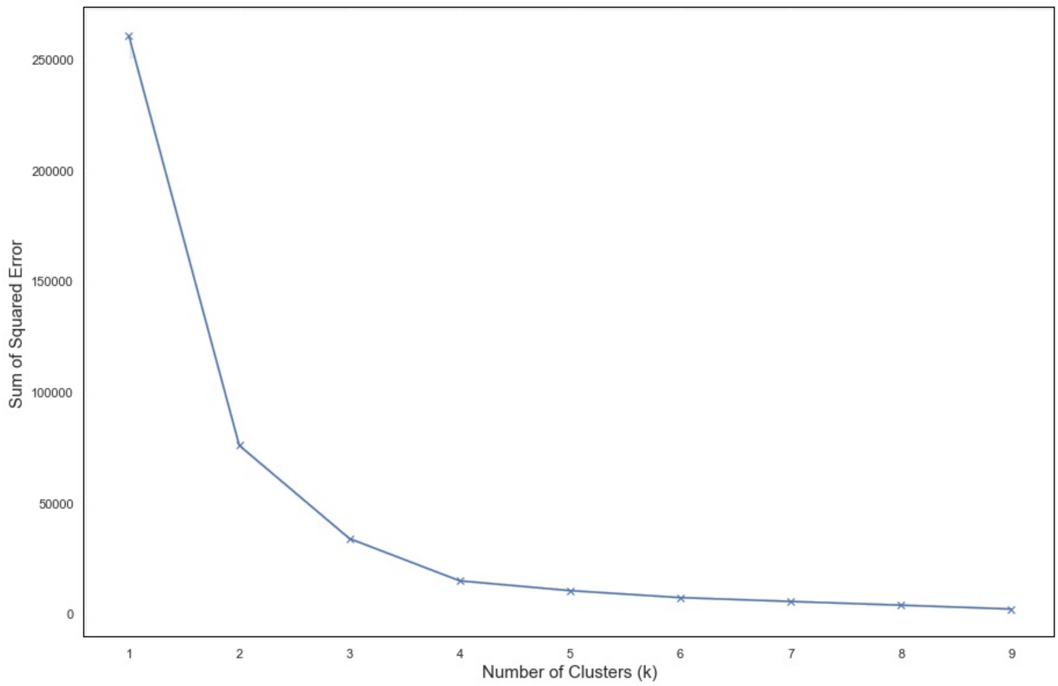


FIGURE 8.3: Selecting the number of groups for WK-mean method using the Elbow Method.

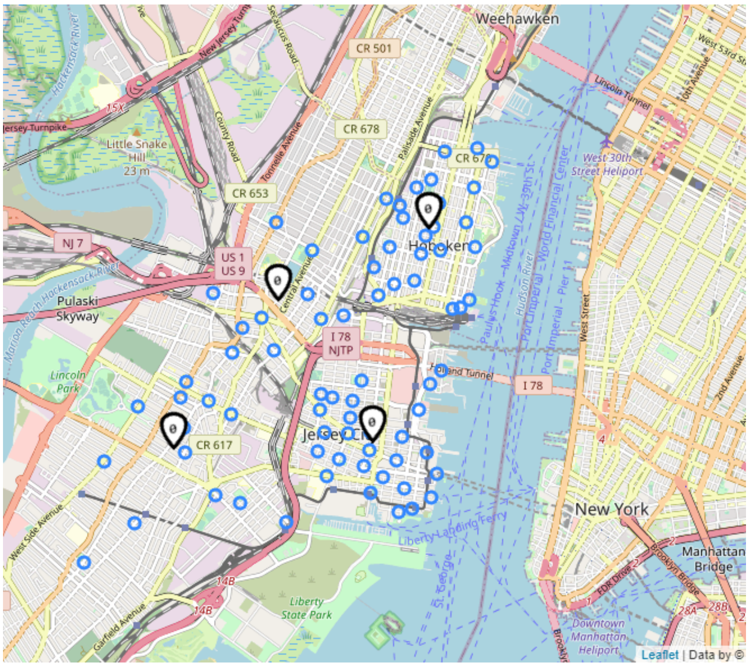


FIGURE 8.4: The result of depot location calculations using k-means.

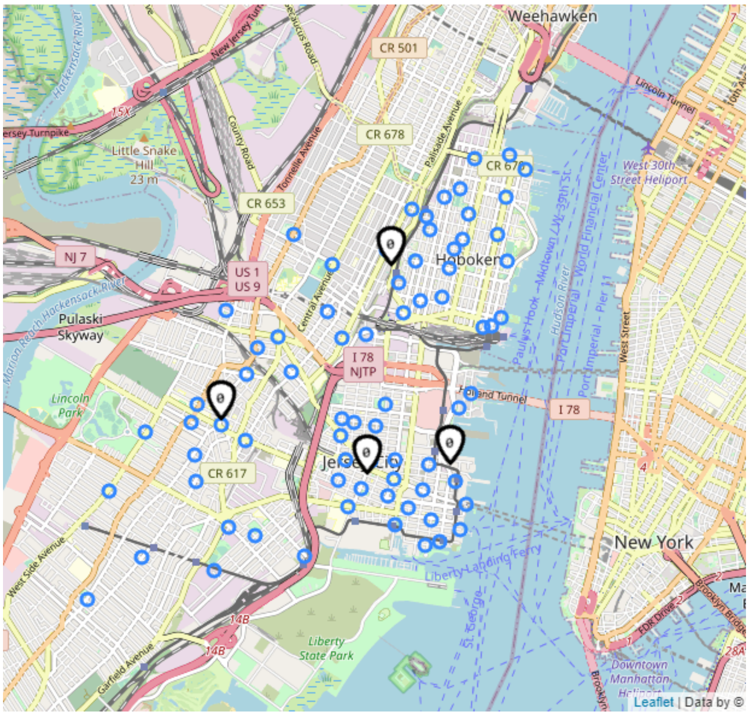


FIGURE 8.5: The result of depot location calculations using Wk-means.

the solution provided an average total distance of 226 km, with a 36 km improvement. As a result, the deployment of four depots would result in a 13.74 percent reduction in total bike routing distance.

8.4 Conclusions

This research aims to inform the operation of bike-sharing systems in terms of the relocation of bikes among stations. A factor that should not be overlooked is the optimal number and location of depots for this relocation activity. Which K-means, WK-means and the Elbow Method were used to ascertain the number and location of depots that minimized the distance required for distributing bikes among the stations. A SARSA algorithm was used for distance comparisons.

The proposed method aims to minimize operating cost by optimizing the total distance required for the bike distribution aspect of a bike-sharing system. Using four depots instead of one, WK-means gave a shorter total distance than K-means method.

Future research into the number and location of depots in bike-sharing systems, or related problems, might use total cost, including both fixed costs and variable cost of a whole scheme, as a performance measure.

Chapter 9

Conclusion and Recommendations

9.1 Conclusion

Bike-sharing systems have become very popular, due to the traffic congestion caused by the growing numbers of people and vehicles, leading to many environmental concerns. Bike-sharing also represents a growing demand for transportation services. Bike-sharing systems are sustainable transportation alternatives to private transport, as they do not lead to carbon emissions, traffic congestion, or the use of non-renewable energy resources. However, the service quality has a great impact on customer satisfaction, which affects the increase in the number of customers, the service's popularity, and the overall economic performance of the bike-sharing companies. In general, bike-sharing systems allow registered customers to request a ride after indicating the pick-up and drop-off times and locations. Thereafter, a customer will be able to find an available bike, use it and park it at any station after reaching their destination. However, the usage characteristics of bikes feature demands that vary rapidly throughout the day. There is also an imbalance between supply and demand. Thus, bike-sharing systems need to maintain the optimal number of bikes and parking spots at each station. This study aimed to improve the efficiency of bike-sharing systems by implementing and developing an algorithm for the imbalance problem in the systems. To sum up, this study covers four proposed methods, which are as follows:

First, the initial Research Question 1: What specific dynamics may be found in the bike-sharing system? which is the bike-sharing system, it was found that the bike demand was different in each area. Therefore, to meet customer needs, the system must be maintained responsibly to match supply and demand, and availability of bikes and space free for parking them in the systems for customer renting. A demand prediction model was proposed to assist the planner. This research explained how to forecast bike demand with multivariate as historical data and environmental factors for time series analysis using a machine learning algorithm. Also, presents a method for selecting input variables in forecasting that will result in high accuracy.

Second, in Research Question 2: How to increase the efficiency of bike-sharing system? In order to improve the efficiency of the bike-sharing system, it is important to meet the needs of customers, responsive bike and bike parking slots are provided to meet the needs of customers. Whereas this process incurs operating costs. An important way to reduce business operating costs is to reduce the total transportation cost. This study developed an algorithm to solve the vehicle routing problem to ensure the shortest total distance for rebalancing bike-sharing, modifying the ABC algorithm to avoid local optimal. The study found that the proposed modified ABC performs well. Further, implementation of machine learning, viz., the Q-learning and SARSA to solve the vehicle routing problem, is a method presented by many researchers as highly accurate.

Third, in Research Question 3: How should fleet imbalances be managed in such a way that there are limitations in terms of improving the capacity of individual stations or building new stations? when there is a limitation in terms of supply, how can operator manage the supply to meet the needs of customers as much as possible. The simulation model is an efficient one for solving the complex problem. This work implemented the simulation model to present the economic impact during the management of the bike-sharing rebalancing process at each station. This work tried to estimate the potential impact of the number of bikes and parking slots on operating costs.

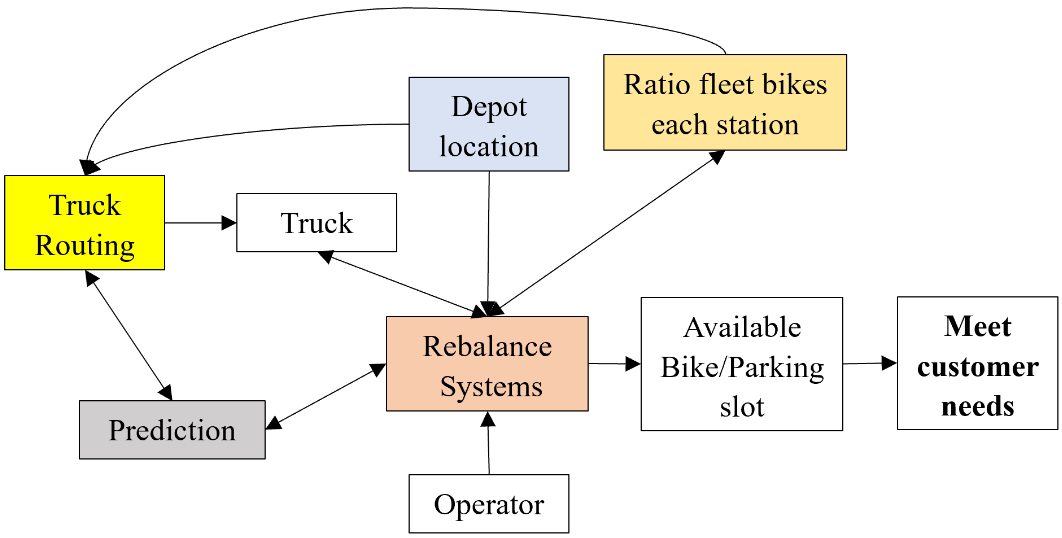


Figure 9.1: Framework of whole systems.

Finally, in Research Question 4: Where should the relocation center or depot be located to minimize operating costs? the location of the release center or depot is also important because the truck that distributes the bikes has to leave and return to the depot. This work presented a method for determining the depot distribution location of bike-sharing by using WK-means and Elbow method. This work demonstrated the influence of the depot location on the total distance during the bike-sharing rebalancing operation. Moreover, the total distance directly affects the transport cost, which is part of the business operating cost.

In all of the four main components of this research as shown in Figure 9.1, it is the part that supports the bike rebalancing system. That will be able to cause minimal operating costs and can also respond to the needs of customers. As a result, the efficiency of the bike system can be increased.

9.2 Contribution

The research has contributed in the strategic design of bike distribution in a bike-sharing system, taking into account factors such as alternative network architectures and possible facility locations. Furthermore, this approach is capable of dealing with dynamic demand. This model is recommended for use in scenarios when the goal is to optimize the whole network, from launch to operation. The model can determine the optimal number of bikes to service from each station in a system. Furthermore, the model was built with acceptable facility and truck capacity limits in consideration. These can help the planner estimate the basic resource requirements, such as the number of trucks, drivers, and personnel, as well as the estimated time.

The bike-sharing system is a type of service in the MaaS that is often used for a short time and for a short time to continue to the main transport. The problems encountered with this bike-sharing system are that it has dynamic demands and offers a one-way service: borrowing from one location and returning it to another. This causes the problem of unbalanced bikes and the customer's need to operate some full bike stations, and some stations do not have available bikes to use. The following conclusions can be contributed as below:

1) This study demonstrated that the proposed machine learning has been tried and tested on historical data and environmental influence data to predict the demand for bike-sharing. The accuracy of the model shown in Chapter 4 indicated its potential applicability to predict the demand for bike-sharing. It is useful for the planner to formulate daily plans for bike relocation involving trucks, drivers, staff, and tools in the systems. It is useful for scheduling the current resource usage, in budget planning, and as a tool to set goals in the operation.

2) This study proposed the algorithm for implementation in truck routing for rebalancing bike-sharing systems; the following findings support this contribution:

2-I) Chapter 5: The meta-heuristic algorithm ABC was developed to solve the truck routing problem for rebalancing bike-sharing. In the literature, the ABC algorithm is touted as an efficient approach that provides optimal solutions for NP-hard problems. However, it has not been confirmed given the results as a global optimal. Thus, we proposed a modified state replaced by the best neighbor solution, which is a guided local search algorithm that starts from a randomly selected complete instantiation and moves to the next instantiation. This idea may prevent the bad regions of the solution search. This study demonstrated that the proposed modified ABC algorithm performs well and indicates the result of shorter total distance than the original ABC, with potential use in the vehicle routing problem.

2-II) Chapter 6: The study implemented machine learning to solve the vehicle routing problem for rebalancing bike-sharing. The advantages of machine learning are its suitability for analyzing big data and solving complex problems. This study demonstrated the possibility of using machine learning where it trained the routing with Q-table.

3) Chapter 7: This study proposed the simulation model for maximization of profit and avoiding loss of customers. The problem of imbalance between demand and supply that arises in finding available bikes for renting and finding available docks for returning the bikes can be solved by assessing customer satisfaction based on arrival time rate and return bike time rate from historical data.

4) Chapter 8: The study adopted clustering analysis to minimize the operating cost. The relocation of the bike-sharing problem can be defined as the use of vehicles able to carry bikes between a depot and bike-sharing stations, where the depot represents the start and end of each trip. The vehicles pick up or drop off bikes to redistribute them among stations, but a supply of bikes must be maintained at the depot to enable replenishment for stations that have few bikes. The proposed method aims to minimize the operating cost by optimizing the total distance required for the bike distribution aspect of the bike-sharing system.

From the above conclusions, the results show that the proposed method can be applied appropriately bike-sharing systems. This research proposes a method for implementing a bike-sharing system with regard to customer needs by focusing on the total distance traveled as a key variable in terms of operating costs. The proposed method for improving the efficiency of the bike-sharing system is to increase the efficiency of MasS that it is responsive to short trips and short distances, especially connecting the missing link between the first/last mile and main transport. Bike-sharing service also supports a travel model that is good for the environment, with many sectors focusing on green travel that creates the least air pollution. In addition, it can be a model that reflects the development of transportation services in the form of MaaS.

9.3 Recommendations

The study of data and assumptions of the system was the study's limitation. The research and outcomes are quite pertinent to the management system. The system is concerned with the information required to administer the facility as well as the assumptions behind the supply (bike) and demand balancing system's operation. When the required prerequisites are satisfied, the methods employed for sequencing and routing are likely to be beneficial. These assumptions and behaviors, on the other hand, can be guaranteed to respond accordingly. Furthermore, AI is used to solve problems, according to the model. As a result, increasing the global optimization of the response is not assured. This work, on the other hand, adopts a method that is more efficient in terms of processing time. Furthermore, bike-sharing is considered a type of transportation service in the MaaS. This means that if part of the system is effective, it affects the efficiency of MaaS.

Moreover, there is a high probability that this approach can be applied to other MaaS transport services in existing mobility services such as car-sharing, and scooter-sharing. Because they are a shared transport service as well as a one-way service, there is a problem of an imbalance between the vehicle to serve the customer and the customer's needs. The vehicle-sharing system is absolutely necessary to relocation the vehicles in the system to meet the needs of customers. This includes having to take into account the operating costs for the survival of the organization.

In this research, there are various modifications and improvements in the methods for solving the bike-sharing imbalance problem, representing a new addition to the existing literature. This thesis contributes to both the planning and the implementation of bike-sharing systems and creates an important foundation for future system planning and maintenance. The prospects for better implementation of the approach are described below:

1) Forecasting customer demand is the key to ensuring that every part of the planning is correct. To do this, big data need to be used with another algorithm. The predicting model can be defined by using another algorithm, for instance deep learning, to improve the computing performance of the forecasting model.

2) This work implemented the meta-heuristic algorithm and machine learning algorithm for ascertaining routing for relocation of bikes at each station in bike-sharing systems. In the development of the ABC algorithm, it was found that there can be an improvement in computing time, along with the ability to offer the best solution. The way machine learning was presented found the ability to find answers is not globally optimal. We hence recommend conduct a study to ascertain the global optimal with another algorithm.

3) Deep analysis of customer usage behavior is a tool for better understanding the problem; hence this thesis's focus on basic analysis customer usage behavior. Therefore, implementation of deep analysis of customer behavior can be studied clearly to better understand solutions for the relevant problem.

4) In determining the depot location, fixed cost and variable cost received no consideration in this research. These important components of costs could be refined and tested, and comparison may provide a significant finding.

5) The bikes that are in service are often seen as electric bikes, which involve charging the battery. There is also a factor in tire inflation and the breakdown of the bike that makes it unusable. The bike-sharing maintenance base or depot must distribute a certain number of bikes to specific locations, plan a specific route for each maintenance truck, and meet the demands of all the locations. Therefore, it is imperative to bikes relocate that this important factor is taken into account.

Appendix A

List of Publications

A.1 International journal paper

[J.1] Kanokporn Boonjubut, Hiroshi Hasegawa, and Suphanut Kongwat. Bike-sharing rebalancing problem modified artificial bee colony algorithm modeling. *Journal of Global Tourism Research*, 6(1):21–30, 2021.

[J.2] Kanokporn Boonjubut, and Hiroshi Hasegawa. Accuracy of hourly demand forecasting of micro mobility for effective rebalancing strategies. *Management Systems in Production Engineering*(in press).

A.2 International conference papers (peer-reviewed)

[C.1] Kanokpron Boonjubut, and Hiroshi Hasegawa. A comparison of clustering method to determine depot location for a bike-sharing operation. *the 11th International Conference on Industrial Technology and Management*, 2022.

[C.2] Kanokpron Boonjubut, and Hiroshi Hasegawa. Simulation for operation and cost optimization in bike sharing rebalancing of user demand. In*Proceeding of the 14th International Conference of Eastern Asia Society for Transportation Studies (EASTS)*, Easter Asia Society for Transportation Studies, 2021.

[C.3] Kanokporn Boonjubut and Hiroshi Hasegawa. Multivariate time series analysisusing recurrent neural network to predict bike-sharing demand. In *Smart Transportation Systems 2020*, pages 69–77. Springer, 2020.

Bibliography

[1] Yale Z Wong, David A Hensher, and Corinne Mulley. Mobility as a service (maas): Charting a future context. *Transportation Research Part A: Policy and Practice*, 131:5–19, 2020.

[2] David A Hensher, Corinne Mulley, Chin Ho, Yale Wong, Goran Smith, and John D Nelson. *Understanding Mobility as a Service (MaaS): Past, present and future*. Elsevier, 2020.

[3] Sonja Heikkilä et al. Mobility as a service-a proposal for action for the public administration, case helsinki. 2014.

[4] I-F Sampo Hietanen CEO. Mobility as a service the new transport model? Technical report, tech. rep., MaaS Finland, 2014.

[5] Roni Utriainen and Markus Pöllänen. Review on mobility as a service in scientific publications. *Research in Transportation Business & Management*, 27:15–23, 2018.

[6] Ariel Waserhole. *Vehicle sharing systems pricing optimization*. PhD thesis, Université de Grenoble, 2013.

[7] Susan A Shaheen, Stacey Guzman, and Hua Zhang. Bikesharing in europe, the americas, and asia: past, present, and future. *Transportation research record*, 2143(1):159–167, 2010.

[8] Supriyo Ghosh, Pradeep Varakantham, Yossiri Adulyasak, and Patrick Jaillet. Dynamic repositioning to reduce lost demand in bike sharing systems. *Journal of Artificial Intelligence Research*, 58:387–430, 2017.

[9] Monika Rani and OP Vyas. Smart bike sharing system to make the city even smarter. In *Advances in Computer and Computational Sciences*, pages 43–55. Springer, 2017.

[10] Japan tourism statistics. `https://statistics.jnto.go.jp/en/graph/`, February 2020.

[11] Svenja Reiss. Demand modeling and relocation strategies for free-floating bicycle sharing systems. *Schriftenreihe des Instituts für Verkehrswesen und Raumplanung*, (64), 2019.

[12] Raphael Giesecke, Teemu Surakka, and Marko Hakonen. Conceptualising mobility as a service. In *2016 Eleventh International Conference on Ecological Vehicles and Renewable Energies (EVER)*, pages 1–11. IEEE, 2016.

[13] M Kivimäki. Maas-finland on the leading edge. In *Mobility as a Service seminar and networking event*, 2014.

[14] José Roberto Reyes García, Gadi Lenz, Steven P Haveman, and Gerrit Maarten Bonnema. State of the art of mobility as a service (maas) ecosystems and architectures—an overview of, and a definition, ecosystem and system architecture for electric mobility as a service (emaas). *World Electric Vehicle Journal*, 11(1):7, 2020.

[15] Francisco Calderón and Eric J Miller. A literature review of mobility services: definitions, modelling state-of-the-art, and key considerations for a conceptual modelling framework. *Transport Reviews*, 40(3):312–332, 2020.

[16] Frances Sprei. Disrupting mobility. *Energy Research & Social Science*, 37:238–242, 2018.

[17] Francesco Ferrero, Guido Perboli, Mariangela Rosano, and Andrea Vesco. Carsharing services: An annotated review. *Sustainable Cities and Society*, 37:501–518, 2018.

[18] Bao-Jun Tang, Xiao-Yi Li, Biying Yu, and Yi-Ming Wei. How app-based ride-hailing services influence travel behavior: an empirical study from china. *International Journal of Sustainable Transportation*, 14(7):554–568, 2020.

[19] P Goddin. Redefining uber: Why the term rideshare doesn't fit. *Retrieved from Mobility Lab website: http://mobilitylab. org/2014/04/17/redefining-uber-whythe-term-rideshare-doesnt-fit*, 2014.

[20] Alejandro Henao and Wesley E Marshall. The impact of ride-hailing on vehicle miles traveled. *Transportation*, 46(6):2173–2194, 2019.

[21] Nicole Ronald, Russell Thompson, and Stephan Winter. Simulating demand-responsive transportation: A review of agent-based approaches. *Transport Reviews*, 35(4):404–421, 2015.

[22] Álvaro Aguilera-García, Juan Gomez, and Natalia Sobrino. Exploring the adoption of moped scooter-sharing systems in spanish urban areas. *Cities*, 96:102424, 2020.

[23] Grant McKenzie. Urban mobility in the sharing economy: A spatiotemporal comparison of shared mobility services. *Computers, Environment and Urban Systems*, 79:101418, 2020.

[24] Susan Shaheen and Adam Cohen. Shared micromoblity policy toolkit: Docked and dockless bike and scooter sharing. 2019.

[25] Susan Shaheen and Nelson Chan. Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections. *Built Environment*, 42(4):573–588, 2016.

[26] Aijun Liu, Xiaohui Ji, Lei Xu, and Hui Lu. Research on the recycling of sharing bikes based on time dynamics series, individual regrets and group efficiency. *Journal of cleaner production*, 208:666–687, 2019.

[27] Paul DeMaio. Bike-sharing: History, impacts, models of provision, and future. *Journal of public transportation*, 12(4):3, 2009.

[28] Elliot Fishman, Simon Washington, Narelle Haworth, and Angela Watson. Factors influencing bike share membership: An analysis of melbourne and brisbane. *Transportation research part A: policy and practice*, 71:17–30, 2015.

[29] Andreas Kaltenbrunner, Rodrigo Meza, Jens Grivolla, Joan Codina, and Rafael Banchs. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4):455–466, 2010.

[30] VE Sathishkumar, Jangwoo Park, and Yongyun Cho. Using data mining techniques for bike sharing demand prediction in metropolitan city. *Computer Communications*, 153:353–366, 2020.

[31] Lei Lin, Zhengbing He, and Srinivas Peeta. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies*, 97:258–276, 2018.

[32] Nonita Sharma, Geeta Sikka, et al. Autoregressive techniques for forecasting applications. In *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*, pages 551–554. IEEE, 2021.

[33] Chengcheng Xu, Junyi Ji, Pan Liu, and Long Peng. Forecasting the travel demand of the station-free sharing bike using a deep learning approach. Technical report, 2018.

[34] Yanyong Guo, Jibiao Zhou, Yao Wu, and Zhibin Li. Identifying the factors affecting bike-sharing usage and degree of satisfaction in ningbo, china. *PloS one*, 12(9):e0185100, 2017.

[35] Lihong Zhang, Jun Zhang, Zheng-yu Duan, and David Bryde. Sustainable bike-sharing systems: characteristics and commonalities across cases in urban china. *Journal of Cleaner Production*, 97:124–133, 2015.

[36] Seonghoon Ban and Kyung Hoon Hyun. Curvature-based distribution algorithm: rebalancing bike sharing system with agent-based simulation. *Journal of Visualization*, 22(3):587–607, 2019.

[37] Qun Chen, Mei Liu, and Xinyu Liu. Bike fleet allocation models for repositioning in bike-sharing systems. *IEEE Intelligent Transportation Systems Magazine*, 10(1):19–29, 2018.

[38] Ahmadreza Faghih-Imani, Robert Hampshire, Lavanya Marla, and Naveen Eluru. An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville. *Transportation Research Part A: Policy and Practice*, 97:177–191, 2017.

[39] Ramon Alvarez-Valdes, Jose M Belenguer, Enrique Benavent, Jose D Bermudez, Facundo Muñoz, Enriqueta Vercher, and Francisco Verdejo. Optimizing the level of service quality of a bike-sharing system. *Omega*, 62:163–175, 2016.

[40] Jasper Schuijbroek, Robert C Hampshire, and W-J Van Hoeve. Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research*, 257(3):992–1004, 2017.

[41] Leonardo Caggiani, Rosalia Camporeale, Michele Ottomanelli, and Wai Yuen Szeto. A modeling framework for the dynamic management of free-floating bike-sharing systems. *Transportation Research Part C: Emerging Technologies*, 87:159–182, 2018.

[42] Panagiotis Angeloudis, Jun Hu, and Michael GH Bell. A strategic repositioning algorithm for bicycle-sharing schemes. *Transportmetrica A: Transport Science*, 10(8):759–774, 2014.

[43] Emine Es Yurek and H Cenk Ozmutlu. A decomposition-based iterative optimization algorithm for traveling salesman problem with drone. *Transportation Research Part C: Emerging Technologies*, 91:249–262, 2018.

[44] Shijin Wang, Ming Liu, and Feng Chu. Approximate and exact algorithms for an energy minimization traveling salesman problem. *Journal of Cleaner Production*, 249:119433, 2020.

[45] Güneş Erdoğan, Maria Battarra, and Roberto Wolfler Calvo. An exact algorithm for the static rebalancing problem arising in bicycle sharing systems. *European Journal of Operational Research*, 245(3):667–679, 2015.

[46] Daniel Chemla, Frédéric Meunier, and Roberto Wolfler Calvo. Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*, 10(2):120–146, 2013.

[47] Yanfeng Li, WY Szeto, Jiancheng Long, and Chin Sum Shui. A multiple type bike repositioning problem. *Transportation Research Part B: Methodological*, 90:263–278, 2016.

[48] Mauro Dell, Manuel Iori, Stefano Novellani, Thomas Stützle, et al. A destroy and repair algorithm for the bike sharing rebalancing problem. *Computers & Operations Research*, 71:149–162, 2016.

[49] Teobaldo Bulhões, Anand Subramanian, Güneş Erdoğan, and Gilbert Laporte. The static bike relocation problem with multiple vehicles and visits. *European Journal of Operational Research*, 264(2):508–523, 2018.

[50] Federico Chiariotti, Chiara Pielli, Andrea Zanella, and Michele Zorzi. A dynamic approach to rebalancing bike-sharing systems. *Sensors*, 18(2):512, 2018.

[51] Adish Singla, Marco Santoni, Gábor Bartók, Pratik Mukerji, Moritz Meenen, and Andreas Krause. Incentivizing users for balancing bike sharing systems. In *Twenty-Ninth AAAI conference on artificial intelligence*, 2015.

[52] Julius Pfrommer, Joseph Warrington, Georg Schildbach, and Manfred Morari. Dynamic vehicle redistribution and online price incentives in shared mobility systems. *IEEE Transactions on Intelligent Transportation Systems*, 15(4):1567–1578, 2014.

[53] Jie Zhang, M Meng, and ZW David. A dynamic pricing scheme with negative prices in dockless bike sharing systems. *Transportation Research Part B: Methodological*, 127:201–224, 2019.

[54] Julia Sidorova and Sai Chandana Kaja. A new approach for solving the disruption in vehicle routing problem during delivery. *Sovremennye informacionnye tehnologii i IT-obrazovanie*, 16(2):371–378, 2020.

[55] El-Ghazali Talbi et al. *Hybrid metaheuristics*, volume 166. Springer, 2013.

[56] Arthur M Geoffrion and Glenn W Graves. Multicommodity distribution system design by benders decomposition. *Management science*, 20(5):822–844, 1974.

[57] Juan Carlos García-Palomares, Javier Gutiérrez, and Marta Latorre. Optimizing the location of stations in bike-sharing programs: A gis approach. *Applied Geography*, 35(1-2):235–246, 2012.

[58] Curtis L Stowers and Udatta S Palekar. Location models with routing considerations for a single obnoxious facility. *Transportation Science*, 27(4):350–362, 1993.

[59] Rajesh Srivastava. Alternate solution procedures for the location-routing problem. *Omega*, 21(4):497–506, 1993.

[60] Geoff Clarke and John W Wright. Scheduling of vehicles from a central depot to a number of delivery points. *Operations research*, 12(4):568–581, 1964.

[61] Guangjie Han, Hao Wang, Mohsen Guizani, Sammy Chan, and Wenbo Zhang. Kclp: A k-means cluster-based location privacy protection scheme in wsns for iot. *IEEE Wireless Communications*, 25(6):84–90, 2018.

[62] Farshad Faezy Razi. A hybrid dea-based k-means and invasive weed optimization for facility location problem. *Journal of Industrial Engineering International*, 15(3):499–511, 2019.

[63] Jihai Luan, Yong Xia, Yuancheng Xie, Dong Zhao, Zhaoxin Zhang, and Ning Li. Anonymous node location based on clustering. In *Journal of Physics: Conference Series*, volume 1976, page 012018. IOP Publishing, 2021.

[64] Joshua Zhexue Huang, Michael K Ng, Hongqiang Rong, and Zichen Li. Automated variable weighting in k-means type clustering. *IEEE transactions on pattern analysis and machine intelligence*, 27(5):657–668, 2005.

[65] Saptarshi Chakraborty and Swagatam Das. On the strong consistency of feature-weighted k-means clustering in a nearmetric space. *Stat*, 8(1):e227, 2019.

[66] Wafic El-Assi, Mohamed Salah Mahmoud, and Khandker Nurul Habib. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in toronto. *Transportation*, 44(3):589–613, 2017.

[67] Sin C Ho and Wai Yuen Szeto. Grasp with path relinking for the selective pickup and delivery problem. *Expert Systems with Applications*, 51:14–25, 2016.

[68] Elliot Fishman. Bikeshare: A review of recent literature. *Transport Reviews*, 36(1):92–113, 2016.

[69] Janett Büttner, Hendrik Mlasowsky, T Birkholz, Dana Gröper, ACFG Emberger, T Petersen, M Robert, SSVP Reth, H Blumel, CR Rodriguez, et al. Optimising bike sharing in european cities. *Recuperado el*, 28:2017, 2011.

[70] Francesc Soriguera, Víctor Casado, and Enrique Jiménez. A simulation model for public bike-sharing systems. *Transportation Research Procedia*, 33:139–146, 2018.

[71] Juan P Romero, Angel Ibeas, Jose L Moura, Juan Benavente, and Borja Alonso. A simulation-optimization approach to design efficient systems of bike-sharing. *Procedia-Social and Behavioral Sciences*, 54:646–655, 2012.

[72] Nanjing Jian, Daniel Freund, Holly M Wiberg, and Shane G Henderson. Simulation optimization for a large-scale bike-sharing system. In *2016 Winter Simulation Conference (WSC)*, pages 602–613. IEEE, 2016.

[73] Alberto Fernández, Holger Billhardt, Sascha Ossowski, and Óscar Sánchez. Bike3s: A tool for bike sharing systems simulation. *Journal of Simulation*, 14(4):278–294, 2020.

[74] Chengcheng Xu, Junyi Ji, and Pan Liu. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation research part C: emerging technologies*, 95:47–60, 2018.

[75] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[76] Simon Ruffieux, Elena Mugellini, and Omar Abou Khaled. Bike usage forecasting for optimal rebalancing operations in bike-sharing systems. In *2018 IEEE*

30th International Conference on Tools with Artificial Intelligence (ICTAI), pages 854–858. IEEE, 2018.

[77] Inc. Motivate International. System data. `https://www.capitalbikeshare.com/system-data`, February 2020.

[78] Christopher M Bishop. Neural networks: a pattern recognition perspective. 1996.

[79] Salam Hussein Ewaid, Salwan Ali Abed, and Safaa A Kadhum. Predicting the tigris river water quality within baghdad, iraq by using water quality index and regression analysis. *Environmental Technology & Innovation*, 11:390–398, 2018.

[80] Pao-Feng Tsai, Chih-Hsuan Wang, Yang Zhou, Jiaxiang Ren, Alisha Jones, Sarah O Watts, Chiahung Chou, and Wei-Shinn Ku. A classification algorithm to predict chronic pain using both regression and machine learning–a stepwise approach. *Applied Nursing Research*, page 151504, 2021.

[81] MA Efroymson. Multiple regression analysis. *Mathematical methods for digital computers*, pages 191–203, 1960.

[82] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[83] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

[84] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[85] Hasim Sak, Andrew W Senior, and Françoise Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. 2014.

[86] Christopher Olah. Understanding lstm networks. 2015.

[87] Apeksha Shewalkar, Deepika Nyavanandi, and Simone A Ludwig. Performance evaluation of deep neural networks applied to speech recognition: Rnn,

lstm and gru. *Journal of Artificial Intelligence and Soft Computing Research*, 9(4):235–245, 2019.

[88] Yoan Fourcade, Aurélien G Besnard, and Jean Secondi. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2):245–256, 2018.

[89] Fábio Cruz, Anand Subramanian, Bruno P Bruck, and Manuel Iori. A heuristic algorithm for a single vehicle static bike sharing rebalancing problem. *Computers & Operations Research*, 79:19–33, 2017.

[90] Yuvraj Gajpal and Prakash Abad. An ant colony system (acs) for vehicle routing problem with simultaneous delivery and pickup. *Computers & Operations Research*, 36(12):3215–3223, 2009.

[91] Chin Sum Shui and WY Szeto. Dynamic green bike repositioning problem–a hybrid rolling horizon artificial bee colony algorithm approach. *Transportation Research Part D: Transport and Environment*, 60:119–136, 2018.

[92] Maria Battarra, Jean-François Cordeau, and Manuel Iori. Chapter 6: pickup-and-delivery problems for goods transportation. In *Vehicle Routing: Problems, Methods, and Applications, Second Edition*, pages 161–191. SIAM, 2014.

[93] Muhammad Usama, Yongjun Shen, Onaira Zahoor, and Qiong Bao. Dockless bike-sharing system. *Journal of Transport and Land Use*, 13(1):491–515, 2020.

[94] Dervis Karaboga and Bahriye Basturk. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *Journal of global optimization*, 39(3):459–471, 2007.

[95] Dervis Karaboga and Bahriye Basturk. On the performance of artificial bee colony (abc) algorithm. *Applied soft computing*, 8(1):687–697, 2008.

[96] Neha Pathak, Manuj Mishra, and Shiv Pratap Singh Kushwah. Improved local search based modified abc algorithm for tsp problem. In *2017 4th International Conference on Electronics and Communication Systems (ICECS)*, pages 173–178. IEEE, 2017.

[97] Philip Kilby, Patrick Prosser, and Paul Shaw. Guided local search for the vehicle routing problem with time windows. In *Meta-heuristics*, pages 473–486. Springer, 1999.

[98] Niaz A Wassan and Gábor Nagy. Vehicle routing problem with deliveries and pickups: modelling issues and meta-heuristics solution approaches. *International Journal of Transportation*, 2(1), 2014.

[99] Sourabh Katoch, Sumit Singh Chauhan, and Vijay Kumar. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, pages 1–36, 2020.

[100] Ajit Kumar, Dharmender Kumar, and SK Jarial. A review on artificial bee colony algorithms and their applications to data clustering. *Cybernetics and Information Technologies*, 17(3):3–28, 2017.

[101] Emmanuel Emeka Okoro, Okorie E Agwu, David Olatunji, and Oyinkepreye David Orodu. Artificial bee colony abc a potential for optimizing well placement–a review. In *SPE Nigeria Annual International Conference and Exhibition*. OnePetro, 2019.

[102] Franz Rothlauf. Optimization methods. In *Design of Modern Heuristics*, pages 45–102. Springer, 2011.

[103] S Binitha, S Siva Sathya, et al. A survey of bio inspired optimization algorithms. *International journal of soft computing and engineering*, 2(2):137–151, 2012.

[104] Dervis Karaboga. An idea based on honey bee swarm for numerical optimization. Technical report, Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, 2005.

[105] Bahriye Basturk. An artificial bee colony (abc) algorithm for numeric function optimization. In *IEEE Swarm Intelligence Symposium, Indianapolis, IN, USA, 2006*, 2006.

[106] Dervis Karaboga and Bahriye Akay. A comparative study of artificial bee colony algorithm. *Applied mathematics and computation*, 214(1):108–132, 2009.

[107] Weifeng Gao, Sanyang Liu, and Lingling Huang. A global best artificial bee colony algorithm for global optimization. *Journal of Computational and Applied Mathematics*, 236(11):2741–2753, 2012.

[108] Tarun Kumar Sharma and Millie Pant. Enhancing the food locations in an artificial bee colony algorithm. *Soft Computing*, 17(10):1939–1965, 2013.

[109] Neha Pathak, Manuj Mishra, and Shiv Pratap Singh Kushwah. Incremental enhanced artificial bee colony algorithm with local search. *International Journal of Computer Applications*, 116(13), 2015.

[110] Pedro Meseguer, Francesca Rossi, and Thomas Schiex. Soft constraints. In *Foundations of Artificial Intelligence*, volume 2, pages 281–328. Elsevier, 2006.

[111] Wai Yuen Szeto, Yongzhong Wu, and Sin C Ho. An artificial bee colony algorithm for the capacitated vehicle routing problem. *European Journal of Operational Research*, 215(1):126–135, 2011.

[112] Patrick Prosser and Paul Shaw. Study of greedy search with multiple improvement heuristics for vehicle routing problems, 1996.

[113] Christos Voudouris Edward PK Tsang. Guided local search", handbook of metaheuristics.

[114] Florian Arnold and Kenneth Sörensen. Knowledge-guided local search for the vehicle routing problem. *Computers & Operations Research*, 105:32–46, 2019.

[115] MM Solomon. Best known solutions identified by heuristics. *Northeastern University, Massachusetts, Boston*, 2005.

[116] Ngoc Duy Nguyen, Thanh Thi Nguyen, Hai Nguyen, and Saeid Nahavandi. Review, analyze, and design a comprehensive deep reinforcement learning framework. *arXiv e-prints*, pages arXiv–2002, 2020.

[117] Aritra Pal and Yu Zhang. Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C: Emerging Technologies*, 80:92–116, 2017.

[118] Yang Xin, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. Machine learning and deep learning methods for cybersecurity. *Ieee access*, 6:35365–35381, 2018.

[119] Thanh Thi Nguyen, Ngoc Duy Nguyen, and Saeid Nahavandi. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE transactions on cybernetics*, 50(9):3826–3839, 2020.

[120] Mohammadreza Nazari, Afshin Oroojlooy, Lawrence V Snyder, and Martin Takáč. Reinforcement learning for solving the vehicle routing problem. *arXiv preprint arXiv:1802.04240*, 2018.

[121] Omar Bouhamed, Hakim Ghazzai, Hichem Besbes, and Yehia Massoud. Q-learning based routing scheduling for a multi-task autonomous agent. In *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 634–637. IEEE, 2019.

[122] Marek Grzes. *Improving exploration in reinforcement learning through domain knowledge and parameter analysis*. PhD thesis, University of York, 2010.

[123] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[124] Xavier Dutreilh, Sergey Kirgizov, Olga Melekhova, Jacques Malenfant, Nicolas Rivierre, and Isis Truck. Using reinforcement learning for autonomic resource allocation in clouds: towards a fully automated workflow. In *ICAS 2011, The Seventh International Conference on Autonomic and Autonomous Systems*, pages 67–74, 2011.

[125] Xin Liu and Konstantinos Pelechrinis. Excess demand prediction for bike sharing systems. *Plos one*, 16(6):e0252894, 2021.

[126] Guangnian Xiao, Ruinan Wang, Chunqin Zhang, and Anning Ni. Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks. *Multimedia Tools and Applications*, 80(15):22907–22925, 2021.

[127] Waldy Joe and Hoong Chuin Lau. Deep reinforcement learning approach to solve dynamic vehicle routing problem with stochastic customers. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 30, pages 394–402, 2020.

[128] Beakcheol Jang, Myeonghwi Kim, Gaspard Harerimana, and Jong Wook Kim. Q-learning algorithms: A comprehensive classification and applications. *IEEE Access*, 7:133653–133667, 2019.

[129] Su Min Jeon, Kap Hwan Kim, and Herbert Kopfer. Routing automated guided vehicles in container terminals through the q-learning technique. *Logistics Research*, 3(1):19–27, 2011.

[130] Philippe Augerat. *Approche polyèdrale du problème de tournées de véhicules*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 1995.

[131] Essia Ben Alaia, Imen Harbaoui, Pierre Borne, and Hanen Bouchriha. A comparative study of the pso and ga for the m-mdpdptw. *International Journal of Computers Communications & Control*, 13(1):8–23, 2018.

[132] Smithin George and Sumitra Binu. Vehicle route optimisation using artificial bees colony algorithm and cuckoo search algorithm-a comparative study. *International Journal of Applied Engineering Research*, 13(2):953–959, 2018.

[133] Mazin Abed Mohammed, Mohd Sharifuddin Ahmad, and Salama A Mostafa. Using genetic algorithm in implementing capacitated vehicle routing problem. In *2012 International conference on computer & information science (ICCIS)*, volume 1, pages 257–262. IEEE, 2012.

[134] M Keshtzari, B Naderi, and Esmaeil Mehdizadeh. An improved mathematical model and a hybrid metaheuristic for truck scheduling in cross-dock problems. *Computers & Industrial Engineering*, 91:197–204, 2016.

[135] Wooyeon Yu and Pius J Egbelu. Scheduling of inbound and outbound trucks in cross docking systems with temporary storage. *European journal of operational research*, 184(1):377–396, 2008.

[136] Iris A Forma, Tal Raviv, and Michal Tzur. A 3-step math heuristic for the static repositioning problem in bike-sharing systems. *Transportation research part B: methodological*, 71:230–247, 2015.

[137] Yang-Kuei Lin and Francois Liang. Simulation for balancing bike-sharing systems. *International Journal of Modeling and Optimization*, 7(1):24, 2017.

[138] Yujing Chen, Yong Zha, Dong Wang, Hongping Li, and Gongbing Bi. Optimal pricing strategy of a bike-sharing firm in the presence of customers with convenience perceptions. *Journal of Cleaner Production*, 253:119905, 2020.

[139] Raymond J Madachy and Daniel X Houston. *What Every Engineer Should Know about Modeling and Simulation*. CRC Press, 2017.

[140] Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences*, 20:514–523, 2011.

[141] Xiaomei Xu, Zhirui Ye, Jin Li, and Mingtao Xu. Understanding the usage patterns of bicycle-sharing systems to predict users' demand: A case study in wenzhou, china. *Computational intelligence and neuroscience*, 2018, 2018.

[142] Jonathan Corcoran, Tiebei Li, David Rohde, Elin Charles-Edwards, and Derlie Mateo-Babiano. Spatio-temporal patterns of a public bicycle sharing program: the effect of weather and calendar events. *Journal of Transport Geography*, 41:292–305, 2014.

[143] Song-Hee Kim and Ward Whitt. Using simulation to study statistical tests for arrival process and service time models for service systems. In *2013 Winter Simulations Conference (WSC)*, pages 1223–1232. IEEE, 2013.

[144] Tu-Cheng Kuo, Wen-Chih Huang, Sheng-Chieh Wu, and Pei-Lun Cheng. A case study of inter-arrival time distributions of container ships. *Journal of Marine science and technology*, 14(3):155–164, 2006.

[145] Bill Roungas, SA Meijer, and Alexander Verbraeck. A framework for optimizing simulation model validation & verification. *International Journal on Advances in Systems and Measurements*, 11(1-2), 2018.

[146] C Maio and C Schexnayder. Simulation model probability distribution functions: Relationships between input data and goodness-of-fit tests. In *ISARC'99: international symposium on automation and robotics in construction (Madrid, 22-24 September 1999)*, pages 103–108, 1999.

[147] Alexander Goldenshluger and David T Koops. Nonparametric estimation of service time characteristics in infinite-server queues with nonstationary poisson input. *Stochastic Systems*, 9(3):183–207, 2019.

[148] T Christopher Greenwell, Janet S Fink, and Donna L Pastore. Assessing the influence of the physical sports facility on customer satisfaction within the context of the service experience. *Sport Management Review*, 5(2):129–148, 2002.

[149] Yinghua Liu and SooCheong Shawn Jang. Perceptions of chinese restaurants in the us: what affects customer satisfaction and behavioral intentions? *International Journal of Hospitality Management*, 28(3):338–348, 2009.

[150] Alfred Weber, Carl Joachim Friedrich, et al. Alfred weber's theory of the location of industries. 1929.

[151] Ganda Yoga Swara et al. Implementation of haversine formula and best first search method in searching of tsunami evacuation route. In *IOP Conference Series: Earth and Environmental Science*, volume 97, page 012004. IOP Publishing, 2017.

[152] Mengyao Cui et al. Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1(1):5–8, 2020.