INTELLIGENT COMPLEMENTARY RIDE-SHARING SYSTEM

Project ID: 19-055

Project Proposal Report

V.A.Wickramasinghe
A.E.Edirisinghe
G.L.S.R.Gunawardena
R.M.A.N.Gunathilake

Bachelor of Science Special (Honors) in Information Technology
Specializing in Software Engineering / Information Technology

Department of Software Engineering / Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

March 2019

INTELLIGENT COMPLEMENTARY RIDE-SHARING SYSTEM

Project ID: 19-055

Project Proposal Report

Bachelor of Science Special (Honors) in Information Technology
Specializing in Software Engineering / Information Technology

Department of Software Engineering / Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

March 2019

DECLARATION

We declare that this is our own work and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
V.A.Wickramasinghe	IT16030190	
A.E.Edirisinghe	IT16025936	
G.L.S.R.Gunawardena	IT16011380	
R.M.A.N.Gunathilake	IT16033474	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Dr. Janaka Wijekoon Date:

Signature of the co-supervisor: Dr. Dharshana Kasthurirathna Date:

Abstract

Traffic congestion is a major concern, which has drawn the attention of the society. The major problem is the number of vehicles daily entering the urban areas a high. Most of the time, many vehicles are coming from the same area to defined destination resulting the vehicles are underutilized. To this end, this study proposes a solution to minimize this problem by implementing a ride-sharing platform: +Go. An initial study of the +Go platform is basically targeting the office staff in Sri Lanka because the majority of the offices are operating in highly congested urban areas. +Go ride-sharing platform matches the profiles of the passengers to the drivers, and vice versa, then suggests the rides using trajectory details, gender preference, personal interests, profession, rating, time and date as the parameters. Furthermore, the +Go platform uses computer vision methodologies to avoid fake registrations, and nonetheless, it uses crowdsourcing platforms to increase the accuracy of the suggested routes. The passengers, including the driver, have to share the cost for the trip and that is intelligently calculated according to fuel consumption of the vehicle, distance, the time taken to arrive at the destination, are taken as the parameters for it. Further, to improve the experience of the user, the +Go system maintains a rating system which allows the driver to rate the ride with the passenger, and the passenger to rate on the driver, vehicle and the co-passengers.

Keywords-Ride Sharing, Carpooling, Machine Learning, Image Processing, Crowdsourcing

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Background

Department of motor vehicle shows that, in 2015, Over 500000 vehicles arrived in Colombo and more than 1.8 million people arrived in rush hours [1]. In fact, 443,586 private vehicles entered into the Colombo area and those are privately owned vehicles [1]. Such amount of privately owned vehicles are used by the office crowd as Sri Lankan public transportation is still evolving and the comfortability level of the transportation system is not adequate. Moreover, the existing means of the public transportation system can carry approximately 33.6 people as an average, especially in buses. Yet, [1] incurs that during the morning and evening time periods, a single car usually carries only 1.87; which proves that most of the people drive almost alone (single-occupant) to work. If single occupants can share their trip among other occupants who travel to the same destination, which will reduce the numbers of vehicles on the road which results in reducing the traffic congestion and minimize the emission of carbon dioxide (CO2) to the environment. Hence, CO2 emission (metric tons per capita) in Sri Lanka was reported at 0.88555 in 2014 can be drastically reduced from this [22].

According to the Report of World Population Prospects 2017 in the United Nations, by 2030, the Sri Lankan population forecast will be around 21,474,701 [2]. Hence, the demand for private transportation has also increased over the last few years [2]. Therefore, huge traffic has occurred in urban areas, especially in Colombo area. According to the survey conducted by CoMTrans study team, which was mainly focused on Borella, Maradana, Dematagoda, Town Hall, and Nugegoda areas have identified that average travel speed is 17-18 $Kmph^{-1}$ in peak hours of both mornings and evenings [3]

When the statics on number of trips the users have during the rush hours, based on western province statistics, CoMTrans incurred that 10.0 million trips were recorded in

the year 2012. They predict that total trips in 2035 will be 17.8 million, whereas Sri Lanka expects 1.8 times high [4]. The survey identified different types of trips were used by the Sri Lankans such as home to work, work to home, home to school, school to home, home to other, other to home and non-home based trip. Non-home based trip highly participate in the past few years as well as such trips will be increased in future as [4] summarized. Therefore, the speed of vehicles on the road will effectively be reduced to 10 $Kmph^{-1}$ during morning peak hours by 2035. Moreover, as per [4], it was identified that Sri Lanka experienced 471 billion economic loss last several years because of the traffic congestions [4].

We came up with a concept of ridesharing application as a solution for the above context and we have done a survey to confirm our hypothesis. From the survey results, we identified that over 82% people up-voted for ridesharing as a good option for Sri Lanka, also over 72% of the people believe traffic will reduce from this proposed solution and over 61% of people stated that they like to collaborate in the car-pooling platform [5].
In addition, we conducted a survey to get the real-time fuel consumption of the vehicles in the urban areas that collected from this survey [5], we will predict the cost of the trip to the passenger before the trip starts.

1.2 Literature Survey

Ridesharing is considered as a possible solution to reduce traffic congestion [13]. Factors like cost of travelling, time, distance, ownership of vehicle, personal attitudes had an influence in choosing the individual transportation mode [9, 10]. Certain studies have focused on the factors like fuel price, demography and safety measures in selecting the transport mode [11, 12]. However, in Sri Lanka, over 443000 private vehicles enter the Colombo area in the rush hours. Out of them, 170000 vehicles are considered to be cars [1]. This is drastic increase of vehicles entering the city limits. The daily loss is calculated as Rs.500m due to the traffic congestion causes by the high number of vehicles [14]. Manzini and Pareschi [15] introduced a new Decision Support System

(DSS) for ridesharing systems. Passengers were given support to select the best car to be traveled. With our proposed system, we optimize the solutions provided by the previous studies by proposing only the drivers, which match with the status of the passenger. Another research was done by Swati R., Neha B., and Ajita A [16] to make the ridesharing application user-friendly of both the driver and the passenger. At the same time, they focused on providing a more reliable system, which provides security, especially to female passengers. Hence, their empirical study proved to be a great idea but unfortunately, it was a fail in the practical scenario. The major concern for many high-end companies was to provide security for women travelers. For that, company like "blablacars" and "Tripda.com" took several dynamic solutions for that. We took our consideration into the gender preference of the passenger, where the passenger is given the right to select the gender he/she wants to travel with. In the literature, Shang et [17] uses Global Positioning Systems (GPS) to predict the user-oriented trajectory for the passenger and the driver. In our proposed solution, we use advanced algorithms with the combination of GPS to recommend the best route, which has less traffic. From the point of algorithms, Fagin and Williams (1983) and later Ajtai et al. (1998) [18] proposed the algorithm to get a portion of users that willing to share their cars and from that study, they introduced a new scheduling algorithm. For the process of grouping individuals who matches with each other, was first done through using K-Means Clustering. Abdul N. and Sebastian P. [19] proved that the mean accuracy of using K-Means clustering gives you over 78% accuracy depending on the dataset, we are using. They also enhanced some features in the K-Means and proved that the new Enhanced version of the K-Means yield accuracy over 89%. From the study done by Chih-Ping Wei, Yen-Hsien Lee and Che-Ming Hsu to compare the most accurate clustering algorithm, they have proved that K-Means clustering yields comparatively accurate results than the other clustering algorithms they used in the study [20]. From the literature, number of clusters in a dataset is easily calculated using the elbow method. From the study of Ketchen Jr. and Shook CL. [21], they have analyzed Elbow

method returns accurate value depending on the dataset used. In this study, we will be using Elbow method to calculate the number of clusters and K-Means clustering to cluster the drivers and passengers.

Identification documents are one of the main sources for verifying the identity of citizens. To proceed with the verification of National Identity Card (NIC) and license card, we decided to use a smartphone to capture the picture, which is sent to a cloud server. Then the image is processed by server and results are sent back to the smartphone. As mentioned by Valiente, Sadaike, Gutiérrez, Soriano, Bressan, and Ruggiero (2016), by using this methodology, a cloud could be used for intensive processing by the devices, which has low computational power [22].

The process could be done in the following steps

1. **Load Image**
   As the way, Parwar, Goverdhan, Gajbhiye, Deshbhratar, Zamare, and Lohe propose, the user is allowed to capture an image from the camera. If the Image is blurred, the user can take a new image again [23].
2. **Crop Image**
   Eliminate the unnecessary parts in the image
3. **Process Image**
   Image processing can be used to observe objects, which are not visible, by the naked eye, sharpening or restoring images into better quality, measuring patterns in various objects, or distinguishing objects in an image [24]. Ohlsson (2016) states that the common way to start the preprocessing of an image is by converting it as a gray-scale image before continued preprocessing [24]. In the way, Chakraborty and Mallik (2013) explain, the color image has to be converted to grayscale for more accurate recognition. Next, the grayscale image is converted to binary using an Adaptive threshold. The adaptive threshold is necessary to convert the grayscale image to a binary image because it is difficult to convert some images to binary by applying a constant threshold and in order to simplify the extraction process [25, 26]. As mentioned in literature, this

process plays a significant role in extracting text from the image, as RGB images contain noise at most times, and are not perfect in identifying text and non-text objects of the images [26]. According to Mordvintsev and Abid (2017) in [27], image processing could be easily performed with the help of OpenCV. Further, this provides the capability for face detection using Haar Cascades. With reference to Clark (2018) in [28], the Python Imaging Library (PIL) has the ability to add image processing functionalities to the Python interpreter and it could be used in image archives, image display, and image processing.

4. Extract Image

The process by which image text converted into plain text is Text Extraction. Text Extraction is quite helpful in information retrieving, editing, documenting, searching, etc. Yet the need for a tool for text extraction has always been there [25, 26]. OCR is a technology for converting text on images into data strings. The strings can be used for many things but some examples are to digitize old documents, translate into other languages or to test and verify text positions [24]. This has also been widely used in various fields such as cheque processing, digital libraries, recognizing text from natural landscape, understanding handwritten text, etc.[29]. According to Chakraborty and Mallik (2013), the processed image can be sent for recognition using Tesseract recognition engine which was developed at HP between 1984-1994, then released for open source in late 2005, and currently owned by Google and is considered to be the best highly portable open source OCR engine currently available[23, 25, 30, 31]. It has been indicated by Ohlsson (2016) that Tesseract supports UTF 8, and has the ability to recognize 100+ languages, and the support for more languages is continuously increasing. The engine is trainable, meaning that a new language or font, which is not normally supported, can be trained and recognized. Tesseract is identified as the most accurate open source engine found on the market [24]. According to text extraction done from images of vehicle number plates by C. Patel, A. Patel and D. Patel (2012) in [31], Tesseract has proven the accuracy of 61% and 70%

with the color and grayscale images respectively. It proves the fact that Tesseract performs better in grayscale images as compared to color images. Further, it has mentioned that in some color images with text extraction accuracy of 100% or near to 100%, and when converted to grayscale, remained with the same extraction accuracy. In addition, it was observed that the processing time of character extraction from grayscale images is reduced by 10% to 50%. So it implies the fact that Tesseract works fast with better text extraction accuracy in when it comes to grayscale images.

Therefore, by considering all the facts regarding text extraction from images, we decided to proceed with Tesseract for text extraction and OpenCV for image processing functions.

To analyze the reviews given by the users, the sentiment analysis approach can be used. Sentiment analysis aims to determine the polarity of emotions like happiness, sorrow, grief, hatred, anger and affection and opinions from the text, reviews, posts which are available online on these platforms [32],[33]. Naive Bayes classifier is a simple method based on the Bayes rule which assumes that the presence of a particular feature independent to the presence of any other feature and contributes independently to the final probability[34],[33]. Llombart and Romero (2017) states that in a real scenario, this independence can hardly be found. As a precaution, they have used Multinomial Naive Bayes, which was provided by SciKit-Learn. This leads to model the same probability but with multinomial distribution [34]. The literature says that an advantage of Naïve Bayes' is that it only needs a little amount of training data for the estimation of the parameters required in classification. [32].To avoid the problem of memorizing data and performing poor with new data; which is mostly happened with machine learning algorithms, Llombart and Romero (2017) propose to work with the test-driven methodology as follows.

• Train (60%): Used in the learning process of the machine learning algorithm

• Test (20%): To verify the algorithm is overfitting or not.

• Validation (20%): To evaluate the accuracy of the machine-learning algorithm [34]

Ramya and Rao (2018) suggest using 80% for training and 20% for testing. Further, this paper comes into the conclusion that the Naïve Bayes algorithm performs better in text analysis when it is compared with algorithms such as Support Vector Machine (SVM) and Multinomial Logistic Regression [35]. The paper of Dey, Chakraborty, Biswas, Bose, and Tiwari (2016) elaborately compares overall accuracy, precisions as well as recall values of K-Nearest Neighbour(K-NN) and Naïve Bayes' and it was obvious that when it comes to movie reviews Naïve Bayes' gave far better results than K-NN, but with hotel reviews, both produced lesser nearly same accuracies [32]. According to the comparison between Naive Bayes Algorithm, K-Nearest Neighbour Algorithm and Random Forest Algorithms done by Baid, Gupta, and Chaplot (2017), the paper suggests that the Naïve Bayes algorithm gave the best accuracy with the accuracy of 81.4%, while others give the accuracy of 55.3% and 78.65% respectively [33]. With respect to the opinions of Vidushi and Sodhi (2017), it concludes that the results are found to be satisfactory and when the comparative analysis is done, Naïve Bayes algorithm outperforms KNN algorithm [36]. By considering all the facts in literature, we decided to use Naive Bayes in analyzing the reviews given by the users in our ride-sharing platform.

Since the general objective of this research is to minimize traffic congestion in urban areas, the identification of the optimum path is very significant. Dijkstra algorithm is known as the shortest path algorithm. Dijkstra Algorithm is used to identify the shortest path of the tree by considering the root of the tree as the starting point and then expand the tree node-by-node [39]. According to the weighted directed graph, there is a shortest path node. For the matter of fact, that node is starting from the starting point and extend to the earliest smallest point. In here, point where entire nodes are adjacent to is known

as the smallest point and the length of the arc is called as the chord length [41]. According to this weighted directed graph, vertices of the graph emphasize the cities while edges of the graph show the distances between two cities which adhere by a road. Therefore, Dijkstra's algorithm can be used to identify the closest path between a city and other entire cities [39, 40].

Similarly, Travelling Salesperson Problem (TSP) is another algorithm to identify the least weight tour, which cover all the nodes of the graph. According to the algorithm, it will identify a certain unique cycle in that graph and travel through each node of the graph for once with a minimal price [42]. Some algorithm have been developed some algorithm to solve TSP. Adewole (2011) have been proposed Genetic Algorithm for solving Traveling Salesman Problem [43]. From the study of Dweepna Garg and Saurabh Shah (2011), they used some other method to solve Travelling Salesperson Problem which is known as Ant Colony Optimization [44]. Tunon and Lopez (2005) proposed the algorithm of Branch and Bound to solve Travelling Salesman Problems [45]. Later on, Pragya, Dutta and Pratyush (2015) succeeded to identify another method to solve the Travelling Salesperson Problems, which is known as Dimensional Ant Colony Optimization (DACO) [46]. Another research used Dijkstra algorithm. It was performed by [47, 48].

The cost calculation is considered the business logic of this system. In a previous study done by Zoepf, Chen, Adu, and Pozo (2018), the fare has to be calculated according to the vehicle type along with a minimum cost. Furthermore based on time and distance added to the cost as dynamic factors. The driver has to bear all expenses associated with vehicle operation including depreciation, insurance, maintenance, repairs, and fuel which varies from the driver to driver, or from vehicle to vehicle [49]. From the study of Santos and Xavier (2013), riders can decide how much the passenger is willing to pay for the trip. Then the system computes the cost based on the current fuel price and fuel consumption for the ride. Next, the system suggests a driver according to the willingness to pay the amount [50]. According to a study of Riquelme, Banerjee, and Johari, they

identified two methodology for pricing .those methodologies are called as static and dynamic price. The static pricing method assigns a fixed price for all drivers who are on the platform. Hence, the price does not change based on instantaneously available service capacity. These parameters slightly different across of the day (Even most taxicab services price evenings differently from daytime hours). Importance of static price is not affecting to the instantaneous state, but it will only react to the course changes. They have derived an equation for dynamic pricing where their equation enters the list of available drivers, base price, distance and time and as the result, the dynamic price is given [51]. Considering the above literature regarding the cost calculation, there is no proper way to calculate the cost for a single passenger. In our proposed solution, we will be using multiple linear regression method to predict the fuel consumption, which results in estimating the cost for the ride. From the study of Aleksandar, Silvana and Valentina (2015), they have proved Mean Absolute Percentage Error (MAPE) of using Multiple Linear regression in their study is 3.0730601 using a trained set of data. From that we have identified error percentage of using regression model in critical analysis is comparatively lower than other regression models and we'll be using it for our cost calculation procedure to produce an accurate result [52].

1.3 Existing Products

UDIO

Codezync Ltd developed and design a carpooling mobile app for Sri Lanka. That helps the vehicle owners to earn some extra income and provides passengers with a comfortable, faster and safer ride to their destination. In their app, the driver is given the authority to set the cost for the trip. In the passenger point of view, it might be unfair. If the user is a passenger, he/she can select a trip and send a join request to the driver. If the driver accepts the join request within a specific period, then notification is received asking for the payment. If and only if payment is success driver will share the contact information [6].

Carpooling.lk

Carpooling.lk provides from the ShareColmbo service. This service can be used for daily travel or to get one-time travel experience with comfortable and economical rideshare. When a ride request passenger can see all offers, which vehicle owner posted. Once click the post, get details about the relevant trip and cost as well. The passenger can find a ride using enter the source, destination, and date. If the user is a driver, he/she should mention the vehicle, source, destination, route, ratio, departure time and no of available seats when adding a post [7].

Note: Currently they do not have a mobile app.

RideShare.lk

Once login into the system, the user can use either offer a ride option or request a ride option. They use the same form to offer a ride and request a ride. After adding the source and destination, google map will be popped up with the details (currently it's not working).In addition to that, we should enter the contact number, the name of the contact person, ride type, schedule Type (Daily or One Time), occurs (Daily or weekly), depart time, expected amount and select to the preferences such as Non-smoker, male-only [8].

Note: Currently they do not have a mobile app

1.4 Research Gap

There are several ride sharing applications in the market. However, according to the literature survey and other findings done, we identified a research gap that they do not address major factors, which need to be addressed using a ride-sharing platform to ensure the reliability and the experience of the user. The following table implies a comparison of features between existing products and our proposed solution.

Features	UDIO	Carpooling.lk	RideShare.lk	+Go
System focused mainly on office staff	✗	✗	✗	✓
Matching the passengers' profile with the suitable drivers	✗	✗	✗	✓
Allow the spouse/guardian to check the passenger's trip details	✗	✗	✗	✓
Suggestion of drivers per passenger interests	✗	✗	✗	✓
Consider gender preference when registering to provide high security	✓	✗	✗	✓
Validating the user by NIC and license by processing the images of them	✗	✗	✗	✓
Analyze the reviews given by users based on their severity and categorizing them	✗	✗	✗	✓
Allowing the passengers to rate the driver, vehicle and co-passengers separately at the end of trip.	✗	✗	✗	✓
Dynamic cost calculation procedure instead of static pricing	✓	✗	✗	✓
Crowdsourcing to improve the optimum path by analyzing more than one algorithm	✗	✗	✗	✓

Table 1.4 - Comparison of Existing Products

1.5 Research Problem

In this country today, traffic control has become the most prevalent factor for the police force. There will be a substantial reduction in deploying many police officers when traffic congestion is reduced. According to the articles, surveys, and police reports, there is a gradual increase of vehicles that enter to the Colombo in the rush hours [7]. Moreover, with the increase in traffic congestion, the number of accidents too increases. From the surveys conducted by the Central Bank, they have identified low occupancy vehicles like cars has been in high number and resulted in traffic congestion and wasting of financial resources, moreover polluting the environment badly.

By considering some statistics collected from articles and reports, we identified that traffic is a major concern in urban areas, especially during office hours. It stated there is an Rs.500 million loss due to the daily traffic congestion [7]. Due to the heavy traffic in the rush hours, there is a huge loss of time wasted on the road. The answer is obvious if we can find a way of reducing, the number of vehicles enter the city; we can reduce the traffic congestion up to a certain extent. So we thought of introducing a ride-sharing app, Which could possibly become a **solution to the traffic congestion**. The basic idea behind that was to combine professionals who are traveling to work by their private vehicles. Hence, when one vehicle carries several people together, it will reduce the number of vehicles enter the city. To verify our hypothesis, we conducted a survey with a sample of more than 150 office crowd and we were able to identify that majority considered ridesharing as a good solution to reduce traffic in urban areas.

2. OBJECTIVES

2.1 Main Objectives

The main objective of this research is to develop an effective solution to minimize the traffic congestion during office hours in urban areas. Because of that, we thought of introducing a ride-sharing app for the working people (office staff) which will help to minimize the traffic congestion. Ride sharing service has become a convenient and felicitous transportation system to everyone in everywhere at any time. Apart from reducing the traffic congestion, there are some other tremendous positive impact on our proposed solution. Some of them are building the network among professionals with similar social status, which will help to reduce the stress and improve the productivity while travelling as a passenger, minimize the environment pollution and fuel consumption cost and help to save the car ownership cost and ensure the security of passengers too.

2.2 Specific Objectives

2.2.1 User Profile Management

In this proposed system, there is a function to identify the most suitable list of drivers for the passengers, which will build an inter-connection between same peer groups. To achieve this, system use ride matching algorithm to filter drivers from a driver pool. For this, we use Rule Based Machine Learning and K-Means Clustering algorithm to get the most related driver list to passengers by considering some specified factors like interest of the passenger, trajectory details, driver ratings, social status of driver and vehicle type etc.

2.2.2 Document Validation

In this, main focus is to validate the driving license, NIC and identify the expirations using an image processing algorithm which will help to minimize the risk of fake profiles getting registered in the system .In this we identify and check the compatibility of driving license and National Identity Card (NIC) by considering the most significant components in them.

2.2.3 Profile Rating Maintenance

In this proposed system, it will identify the response of the drivers and passengers concerning other passengers who joined to the trip and rate the people correspondingly. To achieve this, system has to detect some keywords, which were selected by the users regarding their experience in the ride sharing .Further this system classifies the reviews given by the users by using a sentiment analysis algorithm (Naïve Bayes).This component will help to ensure the security of the passengers and drivers as well.

2.2.4 Optimum Path Recognition

In this proposed system, Dijkstra's Algorithm and Travelling Salesman Problem (TSP) Algorithm will be applied to perceive the closest path with least traffic, which connect source and destination while minimizing the traffic congestion. Dijkstra Algorithm find the shortest path by considering weight of the edges in a weighted graph, Depending upon the user's option ,weight of the edges will be vary as distance or time according to the applicable conditions in it. In suggested system, Genetic Algorithm will be used in traffic optimization.

2.2.5 Cost Calculation

Since this component is considered as to be the business logic of this system, we introduce a new algorithm to calculate the fare of the journey. In here, we use Multiple Linear Regression considering some factors, which will affect to predict the price calculation for a ride. Other than that, total fare will be distributed among all the passengers and drivers by a specific ratio. That ratio will depend on the distance of a segment travelled, cost of fuel consumed and the number of passengers entailed in that relevant segment.

3. METHODOLOGY

3.1 System Design

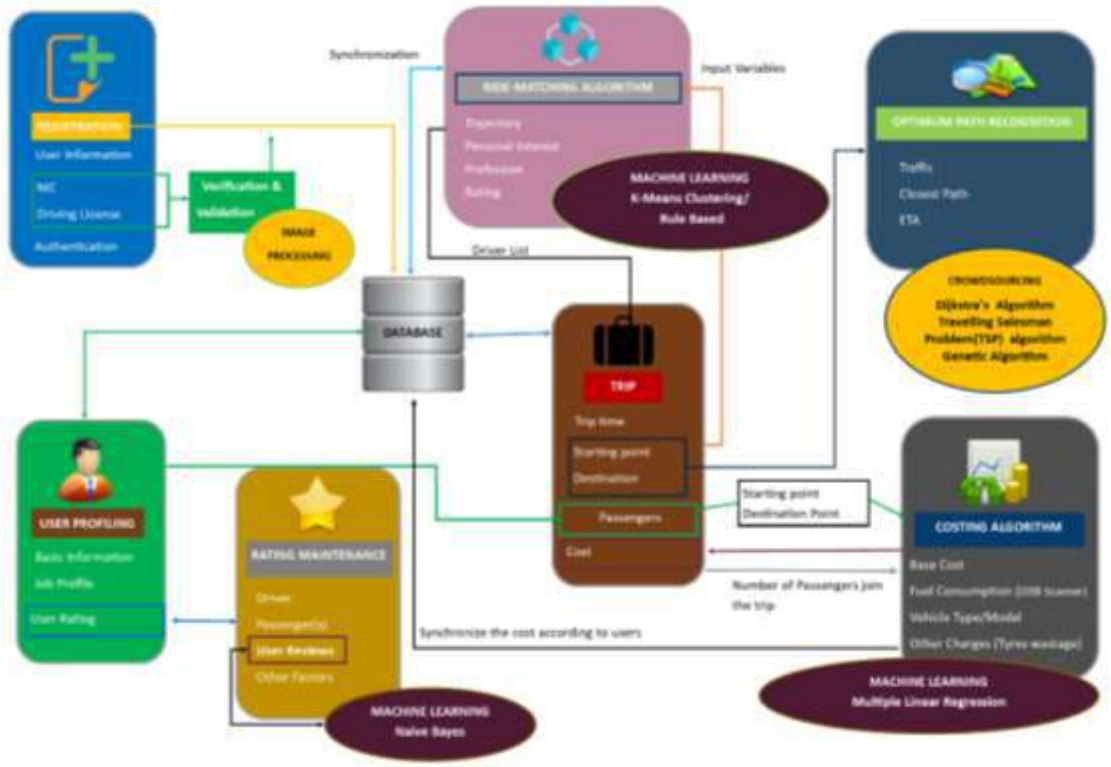

Figure 3.1 - Overall System Diagram

3.2 System Description

3.2.1 User Registration

In the user registration, both the passenger and the driver is considered as similar user roles. First, the user is verified by the image identification of the National Identity Card

(NIC) and then with the License. This is done to avoid any spammers getting register to the system. After that, all the basic information including name, address etc. is collected and personal interests and information of spouse/guardian is taken. All the users who are willing to register to the app should have a vehicle and vehicle details need to be added to create an account. To finalize the registration, he/she needs to verify the phone number and add a payment method to create a successful account. In our system, users can behave as a passenger or as a driver at different times.

3.2.2 User Profiling

User Profiling is the mechanism used to divide each user of the system based on basic information, job profile, interests, and ratings. It always communicates with the rating class to keep a track on the ratings and passenger details are taken from the trip class. All the results will be sent to the database and that information will be used in the Ride Matching algorithm.

3.2.3 Ride Matching Algorithm

We are introducing a new algorithm to recommend the most suitable driver list to the passengers upon searching for the destination. Ride Matching algorithm always synchronizes with the database to keep a track on the latest data to be processed. Basic inputs to the algorithm will be the personal details, trajectory, time, date, personal interests, ratings and more importantly, the profession. Initially, by using rule-based machine learning, we reduced the number of drivers eligible for the cluster and later uses that eligible list to get the most suitable cluster of drivers. From the result set, the most suitable driver list will be sent to the trip class for the final process.

3.2.4 Rating Maintenance

At the completion of the trip, the passenger will be asked to rate his/her experience. Then the user can simply rate with 5 is everything is good; the Driver, Vehicle and the

Co-Passengers will get the default rating 5. If the rating is below five, the passenger will be asked to specify which made them the journey uncomfortable. Further, they will be allowed to write their own review as well and the system will identify the user experience accordingly. Drivers too will get the chance to rate the passengers at the time they get down from the vehicle.

3.2.5 Optimum Path Recognition

In optimum path recognition, we use both Dijkstra's Algorithm and Travelling Salesman Problem (TSP) algorithm to identify the closest path with the least traffic. Because of that, it helps to minimize traffic congestion in urban areas. Dijkstra's Algorithm provides the optimal path to reach the destination by considering the weight of the edges. According to the user's concern, the weight of the graph gets vary. If the user's concern is the distance factor, the length of the edges will be identified as the weight. Similarly, if the user's concern is about traffic, accidents or road closures, weight becomes the calculated time duration to cover the specified distance. Travelling Salesman Problem (TSP) Algorithm provides the optimal path to travel through a city, which covers all the user-specified locations. When users set their locations, our supposed system will identify the order of the locations to visit. Then this algorithm finds the optimal path, which travels, through each location (node) in the listed user-specified locations. In here, registered users will be able to enter the live updates on the relevant path within a specified time range by uploading the pictures of the particular incident. We are introducing the Genetic algorithm to optimize the traffic.

3.2.6 Trip

The passenger is given the option to select the destination. After that, most data is sent to the database and most suitable driver list is displayed using the ride-matching algorithm and user profiling algorithms. Trip class mainly handle the data for each trip. The trip is

subdivided into several segments depending on the number of passengers traveling at that time. If a new passenger join the trip or a certain passenger finishes the trip, a new segment is created. Each Trip segment store, segment start time, start location and passengers details (Passenger's name, Current cost of the passenger). In addition to that, the trip class store overall trip started time (First segment start time), trip end time and details of the driver of the trip.

3.2.7 Costing Algorithm

We are introducing a new algorithm to calculate the cost for each passenger who is traveling in the vehicle. The estimated cost will be based upon fuel consumption of the vehicle, distance, the time is taken to get to the destination and other chargers (Tyres wastage, highway tolls etc.). Vehicle fuel consumption is predicted according to the data set which we gathered during the survey and we use multiple linear regression to formulate our cost formula. In the practical scenario, we decided to get fuel consumption using the odb2 scanner. After the cost is calculated, all the data will be synchronized with the database. To formulate the final cost, we consider a base cost, the cost for fuel consumption, other utility costs to bring out a more dynamic cost formula for the users.

3.3 Software Development Life Cycle

Agile Scrum Methodology is used as the software development methodology in our system. Since the Agile recommends user collaboration at the start of the project, this methodology will accept the late changes in the requirements within the development process. Therefore this methodology will facilitate the regular adoption to changing circumstances in a resilience manner. As this research is done by four members, this methodology will help to identify the research problem thoroughly and provide the pertinent solutions to the relevant problems with a common understanding about each an every research component by coordinating regular meetings.

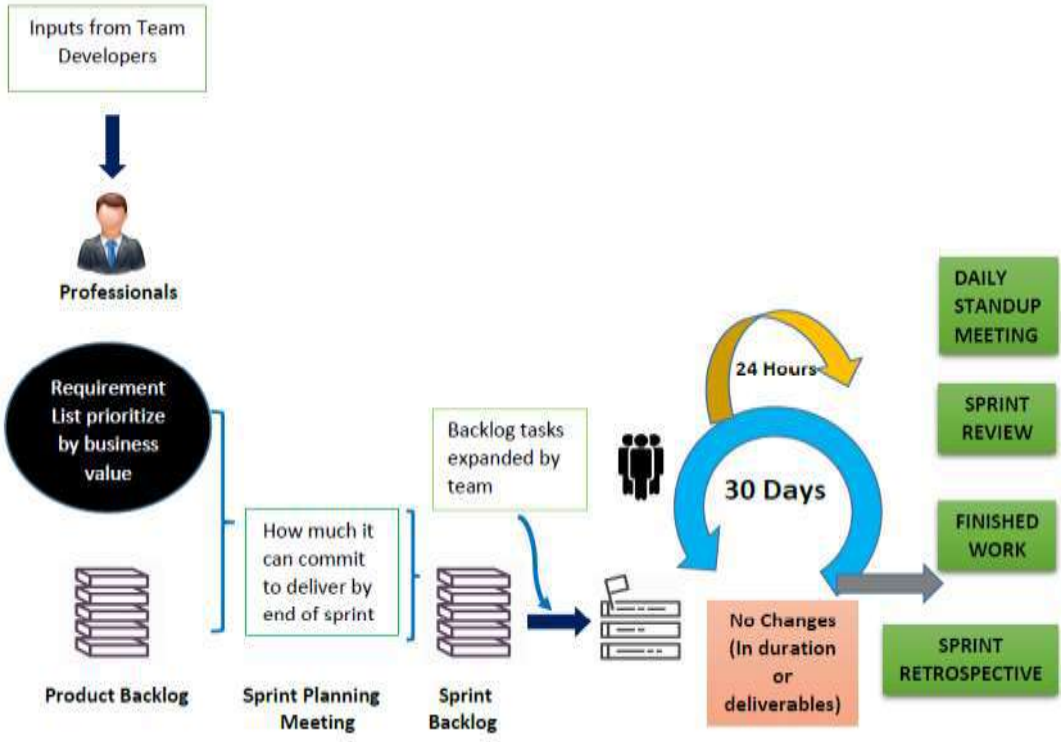

Figure 3.3 - Agile Methodology

3.4 Gantt Chart

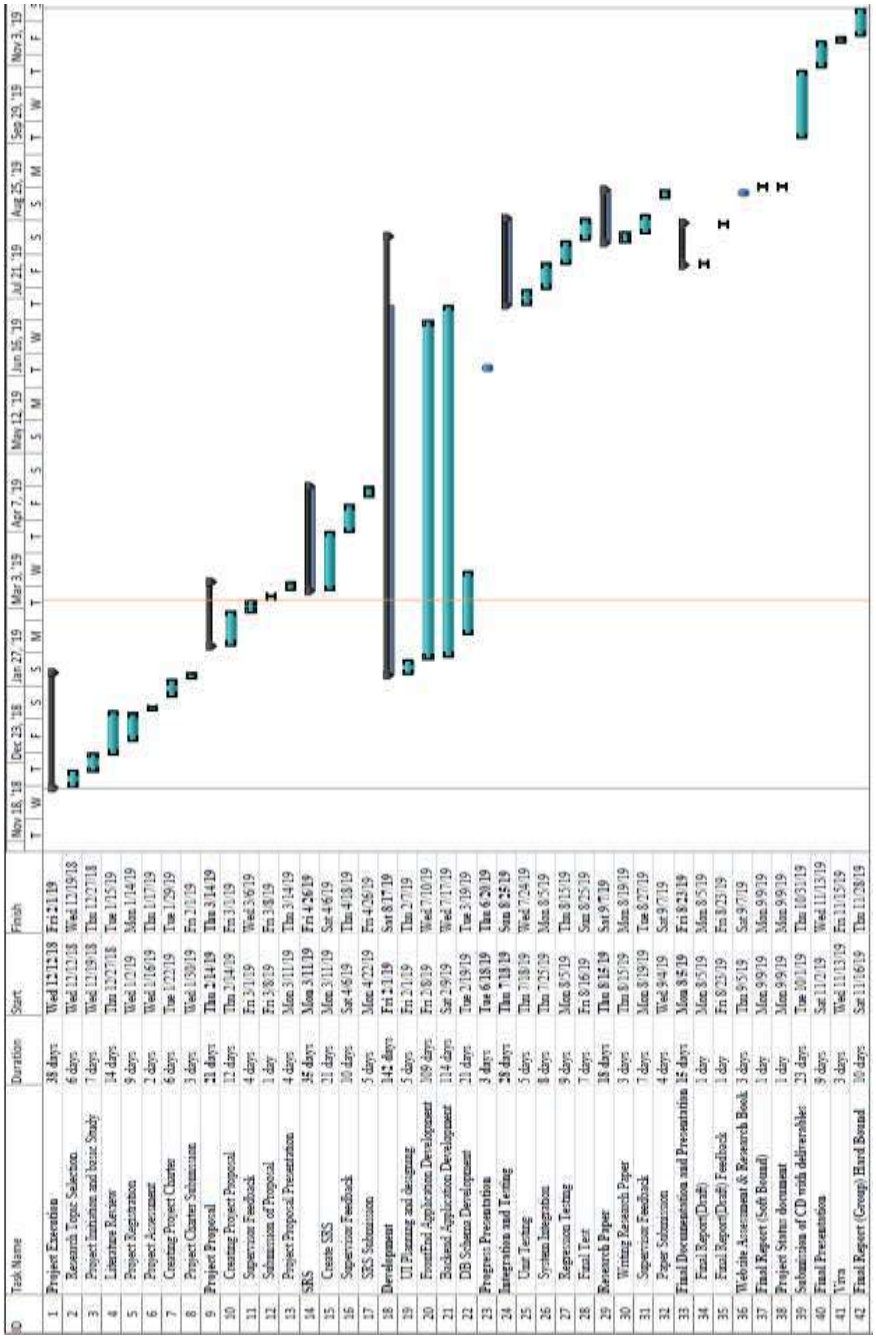

Figure 3.4 - Gantt chart

3.5 Work Breakdown Structure

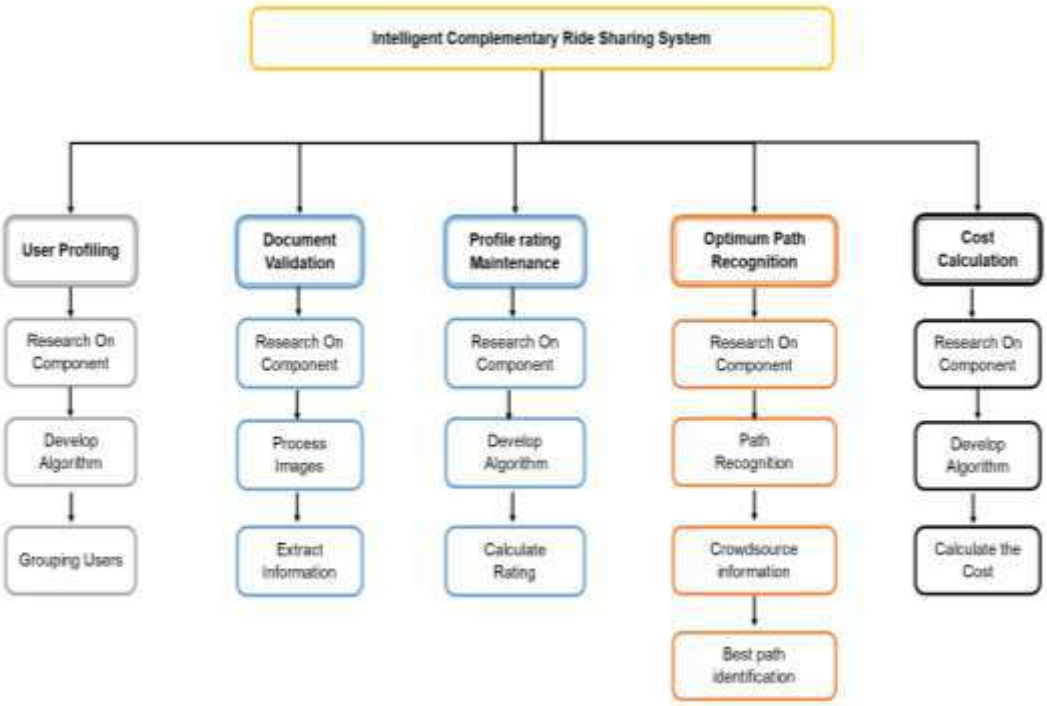

Figure 3.5 - Work Breakdown Structure

4. TOOLS AND TECHNOLOGIES

Following tools and technologies are expected to be used while developing the proposed system:

Tools and IDEs'	Technologies and Services
● Pycharm 2018.3.4 ● Android Studio 3.3.0 ● Jupyter Notebook	● Python 3.7.2 ● Android 4.4 upwards ● Google API
Libraries	**Database Engine**
● OpenCV ● PIL ● Tesseract ● Pandas ● Matplotlib ● Numpy ● Sklearn	● SQLite ● MySQL ● Firebase

Table 4.1 - Tools and Technologies

5. DESCRIPTION OF PERSONAL AND FACILITIES

Registration No	Name	Task Description
IT16030190	V.A. Wickramasinghe	• Develop a ride matching algorithm using rule based machine learning to filter drivers based on the interests, date and time, trajectory etc. of the passenger and K-means clustering to create a cluster of most suitable drivers • Implementation of user interface related to profile registration • Verifying user registration with the code sent to mobile phone • Develop the mobile application user interface for login, driver list and adding payment method • Documentation • Testing
IT16025936	A.E.Edirisinghe	• Develop an image processing algorithm to extract information from National Identity Card(NIC card) and license card • Validating the NIC and license by identification of human face in the cards • Develop the mobile application to capture images of NIC and license • Develop the mobile application user interface for user rating • Develop a sentiment analysis algorithm to analyse the reviews of users and generate the rating accordingly • Documentation • Testing

IT16011380	G.L.S.R. Gunawardena	• Development of an algorithm to predict the estimated cost of the journey • Calculating the actual cost of the journey using odb2 scanner • Implementation of user interface related to cost calculation • Development of interface of the system to be accessed by the spouse/guardian of the user • Documentation and Testing
IT16033474	R.M.A.N. Gunathilake	• Development of an algorithm to identify the optimum path suitable in the journey • Implementation of crowdsourcing to predict the most efficient route • Development of user interface to generate the optimum path visible in the system • Documentation • Testing

Table 5.1 - Description of Personal and Facilities

6. BUSINESS POTENTIAL

With the vast evolution of technology, ridesharing has a huge impact on these ride sharing platforms. This suggested system presents many prospective for society with several economic benefits. As a result of that, the supposed system will help to deliver the product with the highest business value.

Our ride-sharing app for the working people(office staff) which will help to minimize the traffic congestion. Since our ride-sharing app is mainly focused on working people, it will help to reduce traffic congestion during office hours in urban areas. Resultantly it will reduce environmental pollution too. From the business perspective, our supposed system will help to reduce the cost of traveling and in the meantime, it will provide the compensation fee on both passenger and driver for any delay other than specified waiting time. Besides, counter to the assertion this proposed ride-sharing application mainly focused on professionals, which, will lead to building a strong network among professionals. So, it will help to reduce stress and improve productivity while traveling as a passenger.

Turning an idea into a business is where the substantial challenge arises. From a business perspective, our supposed system will introduce dedicated parking slots for the carpoolers. Currently, the majority of the people in urban areas tend to use ridesharing method for the ease of use. Therefore, offer more embolden prizes to the people who use carpooling frequently. In order to keep the business more profitable, we hope to give more assets to top rating passengers such as promo codes etc. In other to achieve this, we can introduce loyalty cards for the passengers. As this is a business in the transportation sector, it is more important to centering the business more profitable.

7. BUDGET AND BUDGET JUSTIFICATION

Required Resources	Unit Cost in LKR	Unit Cost in US Dollars
ODB2 Scanner (2 units)	Rs.2321.82	$12.94
Web Server	Rs.5358.34	$29.97
Total	**Rs.7680.16**	**$42.91**

Table 7.1 - Budget and Budget Justification

8. REFERENCES

[1] [online] Available at: https://indi.ca/2015/10/colombo-vehicle-statistics-2015/ [Accessed 2 Jan. 2019].

[2] [online] Available at: http://www.worldometers.info/world-population/sri-lanka-population/
Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2017 Revision. (Medium-fertility variant). [Accessed 10 Jan. 2019].

[3] Urban Urban Transport Master Plan for Colombo Metropolitan Region and Suburbs. (2014). [ebook] Colombo ,Sri Lanka: Ministry of Transport, pp.25,26. Available at: http://www.transport.gov.lk/web/images/stories/CoMTrans%20Final%20Report_S.pdf [Accessed 13 Jan. 2019].

[4] Draft Urban Transport Master Plan - Ministry of Transport. (2013). [ebook] Colombo,Sri Lanka: Ministry of Transport, pp.66,74,75.
Available at: http://www.transport.gov.lk/web/images/stories/comtrans.pdf
[Accessed 21 Jan. 2019].

[5]Vehicleinfo.000webhostapp.com. (2019). *Vehicle Details*. [online] Available at: https://vehicleinfo.000webhostapp.com/

[6]Ft.lk. (2017). Innovative car pooling app UDIO goes live. [online] Available at: http://www.ft.lk/motor/Innovative-car-pooling-app-UDIO-goes-live/55-645554 [Accessed 18 Jan. 2019].

[7] Carpooling.lk. (2019). Carpooling. [online] Available at: http://www.carpooling.lk/ [Accessed 23 Jan. 2019].

[8] RideShare.lk. (2019). Carpool for better commute and quality of life. [online] Available at: http://www.rideshare.lk/ [Accessed 01 Feb. 2019].

[9] Ding, C.; Wang, D.; Liu, C.; Zhang, Y.; Yang, J. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. Transp. Res. Part A Policy Pract. 2017, 100, 65–80. [CrossRef]

[10] Schneider, R.J. Theory of routine mode choice decisions: An operational framework to increase sustainable transportation. Transp. Policy 2013, 25, 28–137. [CrossRef]

[11] Efthymiou, D.; Antoniou, C.; Waddell, P. Factors affecting the adoption of vehicle sharing systems by young drivers. Transp. Policy 2013, 29, 64–73

[12] Rayle, L.; Dai, D.; Chan, N.; Cervero, R.; Shaheen, S. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. Transp. Policy 2016, 45, 168–178

[13] Dewan, K.K. & Ahmad, Israr. (2007). Carpooling: A Step to Reduce Congestion (A Case Study of Delhi). Engineering Letters [Accessed 5 Jan. 2019].

[14] [online] Available at: https://www.newsfirst.lk/2017/03/16/rs-500m-loss-incurred-daily-traffic-congestion-transport-authorities-observe/ [Accessed 30 Dec. 2018].

[15] R. Manzini and A. Pareschi, "A Decision-Support System for the Car Pooling Problem," *journel on transportation technologies*, Vol.2 No.2 ,2012,pp. 85-101.

[16] Swati. R. Tare, Neha B. Khalate and Ajita A. Mahapadi,"International Journal of Advanced Research in Computer Science and Software Engineering 3(4)", ISSN: 2277 128X April - 2013, pp. 54-57

[17] S. Shang, R. Ding, B. Yuan, K. Xie, K. Zheng, and P. Kainis, "Useroriented trajectory search for trip recommendation," in Proc. EDBT, 2012, pp. 156–167.

[18] Fagin, R. and Williams, J. (1983). A Fair Carpool Scheduling Algorithm. IBM Journal of Research and Development, 27(2), pp.133-139.

[19] K. A. Abdul Nazeer, M. P. Sebastian,"Improving the Accuracy and Efficiency of the k-means Clustering Algorithm", Proceedings of the World Congress on Engineering 2009 Vol I, 2009 [Accessed 31 Dec. 2018].

[20] Chih-Ping Wei, Yen-Hsien Lee and Che-Ming Hsu, Empirical Comparison of Fast Clustering. Algorithms for Large Data Sets,2000, Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.

[21] Ketchen Jr DJ, Shook CL. The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. Strategic Management Journal. 1996; p. 441-458.

[22] [online] Available at: https://tradingeconomics.com/sri-lanka/co2-emissions-metric-tons-per-capita-wb-data.html
[Accessed 9 Jan. 2019]

[23] R. Valiente, M. T. Sadaike, J. C. Gutiérrez, D. F. Soriano, G. Bressan and W. V. Ruggiero, "A Process for Text Recognition of Generic Identification Documents Over Cloud Computing," 2016.

[24] A. Parwar, A. Goverdhan, A. Gajbhiye, P. Deshbhratar ,R. Zamare and P. Lohe, "Implementation to Extract Text from Different Images by Using Tesseract Algorithm," *International Journal Of Engineering And Computer Science ISSN: 2319-7242,* vol. 6 Issue 2, Page No. 20298-20300, Feb. 2017.

[25] V. Ohlsson, "Optical Character and Symbol Recognition using Tesseract", Dissertation, 2016.

[26] P. Chakraborty and A. Mallik, "An Open Source Tesseract based Tool for Extracting Text from Images with Application in Braille Translation for the Visually Impaired," *International Journal of Computer Applications (0975 – 8887),* vol. 68, no.16, April 2013.

[27] K. N. Natei, J. Viradiya and S. Sasikumar, "Extracting Text from Image Document and Displaying Its Related Information", *K.N. Natei Journal of Engineering Research and Application,* vol. 8, pp. 27-33, May 2018.

[28] A. Mordvintsev and K.Abid, OpenCV-Python Tutorials Documentation, Nov 05 2017.

[29] A. Clark, Pillow (PIL Fork) Documentation, Release 5.3.0, Oct 24 2018.

[30] S. Ramiah, T. Y. Liong and M. Jayabalan, "Detecting text based image with optical character recognition for English translation and speech using Android," 2015 IEEE Student Conference on Research and Development (SCOReD), Kuala Lumpur, 2015, pp. 272-277.

[31] R. Smith and Google Inc, "An overview of the Tesseract OCR Engine," Proc. 9th IEEE Intl. Conf. on Document Analysis and Recognition (ICDAR), 2007, pp. 629-633.

[32] C. Patel, A. Patel and D. Patel,"Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study," *International Journal of Computer Applications (0975 – 8887),* vol. 55, no. 10, October 2012.

[33] L. Dey, S. Chakraborty, A. Biswas, B. Bose and S. Tiwari,"Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier,"*International Journal of Information Engineering and Electronic Business,* vol. 6, Issue 4, pp. 54-62, 2016.

[34] P. Baid, A. Gupta and N.Chaplot, “Sentiment Analysis of Movie Reviews using Machine Learning Techniques”, 2017.

[35] Llombart, Òscar Romero, “Using Machine Learning Techniques for Sentiment Analysis,” 2017.

[36] V. U. Ramya and K. T. Rao,”Sentiment Analysis of Movie Review using Machine Learning Techniques,” *International Journal of Engineering & Technology*, vol. 7, pp. 676-681, 2018.

[37] Vidushi, G. S. Sodhi,”Sentiment Mining of Online Reviews Using Machine Learning Algorithms,” *International Journal of Engineering Development and Research (IJEDR)*, ISSN:2321-9939, vol.5, Issue 2, pp.1321-1334, May 2017.

[38] M. H. Hassan, S. P. Shakthi and R. Sasikala,”Sentimental analysis of Amazon reviews using naïve bayes on laptop products with MongoDB and R,” IOP Conference Series: Materials Science and Engineering, 2017, vol. 263, pp. 042090.

[39] Vaibhavi Patel and Prof. ChitraBaggar, “A Survey Paper of Bellman-Ford Algorithm and Dijkstra Algorithm For Finding Shortest Path in GIS Application,” *International Journal of P2P Network Trends Technology(IJPTT*), vol. 5, February 2014

[40] Ni Kai, Zhang Yaoting, Ma Yuepeng, "Shortest Path Analysis Based on Dijkstra's Algorithm in Emergency Response System," *Indonesian Journal of Electrical Engineering and COmputer Science*, vol. 12, no. 5,pp 133-139, 2015

[41] Dechuan Kong ,Yunjuan Liang, Xiaoqin Ma, Lijun Zhang “Improvement and Realization of Dijkstra Algorithm in GIS of Depot”

[42] Ojaswa S. , Darka M. , Francois A. and Girija D., 2005.TRAVELING SALESPERSON APPROXIMATION ALGORITHM FOR REAL ROAD NETWORKS Technical Report 388, Department of Geomatics Engineering, University of Calgary 2500 University Drive NW, Calgary, AB, Canada, T2N 1N4

[43] Adewole Philip, Akinwale Adio Taofiki and Otunbanowo Kehinde, “A Genetic Algorithm for Solving Travelling Salesman Problem,” *(IJACSA) International Journal of Advanced Computer Science and Applications*,vol. 2, no.1, January 2011

[44] Dweepna Garg and Saurabh Shah, “Ant Colony Optimization For Solving Traveling Salesman Problem,” *International Journal of Computer Science and System Analysis*,vol. 5, no. 1, January-June 2011, pp. 23-29, ISSN 0973-7448

[45] M.I.C., Tunon and M.R. Lopez. ”Design and Use of the CPAN Branch and Bound for The Solution of the Travelling Salesman Problem (TSP),” International Conference on Electronics, Communication and Computers, Conielecomp, 1 August 2005.

[46] Pragya, M.Dutta, and Pratyush. “TSP Solution Using Dimensional Ant Colony Optimization,” International Conference on Advanced Computing & Communication Technologies, 6 April 2015.

[47] M.M. Mena, R.H. Ucan, V.C. Vetina, and F.M. Ramirez, “Web Services Composition using the Bidirectional Dijkstra Algorithm,” IEEE Latin America Transaction, vol. 14, Issue 5, May 2016

[48] A. Ratnasari, F. Ardiani, F. Nurvita. “Penentuan Jarak Terpendek dan Jarak Terpendek Alternatif menggunakan Algoritma Dijkstra serta Estimasi Waktu Tempuh”. Universitas Islam Indonesia. ISBN: 979-26-0266-6. Yogyakarta 2013

[49] Zoepf, S., Chen, S., Adu, P. and Pozo, G. (2013). *The Economics of Ride-Hailing: Driver Revenue, Expenses and Taxes*. [ebook] MIT Center for Energy and Environmental Policy Research, p.1. Available at: https://orfe.princeton.edu/~alaink/SmartDrivingCars/PDFs/Zoepf_The%20Economics%20of%20RideHialing_OriginalPdfFeb2018.pdf

[50] O. Santos, D. and C. Xavier, E. (n.d.). *Dynamic Taxi and Ridesharing: A Framework and Heuristics for the Optimization Problem*. Eduardo C. Xavier, p.1.

[51] Riquelme, C., Banerjee, S. and Johari, R. (n.d.). *Pricing in Ride-share Platforms: A Queueing-Theoretic Approach*. [ebook] pp.6,7,15. Available at: http://www.columbia.edu/~ww2040/8100F16/Riquelme-Johari-Banerjee.pdf [Accessed 6 Jan. 2019].

[52] Aleksandar, P., Silvana, P. and Pancovska Valentina, Z. (2015). *Multiple Linear regression model for predicting bidding price*.

9. APPENDIX

Survey Questionnaire

This is a survey conducted to understand the transportation patterns of the office staff and to understand their interest on Carpooling.
Carpooling (also **car-sharing**, **ride-sharing** and **lift-sharing**) is the sharing of car journeys so that more than one person travels in a car, and prevents the need for others to have to drive to a location themselves.[wikipedia]

Personal Information

1) How do you travel between your workplace and home in each day ?
   - ❑ I drive
   - ❑ I use public transport
   - ❑ I walk
   - ❑ I get a lift
   - ❑ I don't need to commute

** If you are walking ,skip to question 20 .*

2) What is your position in the job?
   - ❑ Trainee
   - ❑ Driver
   - ❑ Engineer (Software /Civil /Mechanical)
   - ❑ Clerk
   - ❑ Accountant
   - ❑ Manager
   - ❑ CEO / CFO / CIO
   - ❑ Other _____

3) Please specify your gender
   - ❑ Male
   - ❑ Female

4) Please specify your age group
   - ❑ 17 - 21
   - ❑ 22 - 30
   - ❑ 31 - 40
   - ❑ 41 - 50
   - ❑ 51 - 60

- ❑ 60+

5) Which is the OS of the mobile phone you use?
- ❑ Android
- ❑ iOS
- ❑ Windows
- ❑ Other _____

6) Do you have a vehicle?
- a)Yes.
- b) No.

7) Vehicle type you own?
- ❑ Car
- ❑ Van
- ❑ Cab
- ❑ SUV
- ❑ No preference
- ❑ Other _____

8) Vehicle Brand
- ❑ Honda
- ❑ Toyota
- ❑ Nissan
- ❑ Suzuki
- ❑ Benz/BMW
- ❑ Other _____

9)Fuel Type
- ❑ Petrol
- ❑ Diesel
- ❑ Hybrid
- ❑ Electric
- ❑ Other _____

10)Fuel Consumption(km/l)
- ❑ less than 5
- ❑ 5-7
- ❑ 7-9
- ❑ 9-12

- ❑ 12-15
- ❑ greater than 15

11)What is the vehicle type you prefer to travel the most?

- ❑ Car
- ❑ Van
- ❑ Cab
- ❑ SUV
- ❑ No preference

12) Have you ever carpooled (either offered a ride in your vehicle or took a ride in others' vehicle)?

- ❑ Yes, on an informal basis
- ❑ Yes, on a formal paying basis
- ❑ No, I have never had a car sharing arrangement

13) Would you carpool in your car if you get paid by the people you are offering ride to?

- ❑ Yes
- ❑ I don't own a car but yes, I'd like to carpool
- ❑ I don't own a car & I wouldn't carpool
- ❑ Other _____

14) Would you be interested in a collaboration between Carpool platforms

- ❑ Yes
- ❑ No
- ❑ Maybe

** If you are not interested in carpooling ,skip to question 20.*

Carpooling Information

15)What would be your primary reason for opting for car sharing?

- ❑ To save car ownership costs
- ❑ To save fuel costs
- ❑ To meet interesting people
- ❑ For the benefit of the environment
- ❑ To avoid the hassle of looking for a parking space
- ❑ To make the journey to and from work quicker
- ❑ To minimize traffic congestion
- ❑ Other _____

16)What are the important factors to be considered before choosing car sharing service in the security and comfort perspective of the car owner?
(please tick the relevant)

	Very Important	Important	Neither important or unimportant	Unimportant	Not at all Important	Don't know
Type of vehicle						
Number of people car sharing in a given period of time						
Previous user reviews						
Luggage(s) allowed(size, etc)						
Insurance details						
Brief profile (Having access to their social profile,like Facebook etc)						

17)Please rate the following incentives that could be used to encourage the car sharing service.

- On a scale of 1 to 10,how comfortable would you be giving a stranger a lift?
  (1 - not comfortable , 10 - very comfortable)

  1( ) 2( ) 3( ) 4( ) 5( ) 6( ) 7( ) 8( ) 9( ) 10( )

- On a scale of 1 to 10,how comfortable would you be receiving a lift off a stranger?
  (1 - not comfortable , 10 - very comfortable)

  1( ) 2( ) 3( ) 4( ) 5( ) 6( ) 7( ) 8( ) 9( ) 10( )

18)What type of person would you consider having a formal car sharing arrangement with?

- ❑ A solo female driver
- ❑ A solo male driver
- ❑ An accompanied female driver
- ❑ An accompanied male driver
- ❑ I have no preference

19)What would you consider when calculating the pricing for a shared journey?

- ❑ Cost per Kilometer
- ❑ Depending on the distance and frequency ,agreeing for a fixed cost with the vehicle owner .
- ❑ Other(specify) _____

20)Suppose you travelling companion is working in a company where he use to travel more than your work place.What would you prefer (The total travelling cost gets divided among the number of passengers) ?

- ❑ Drop the companion at his office by travelling the extra distance
- ❑ Drop the companion at your working place and asking him to find another ride.
- ❑ Other _____

21)What are the main reasons for not interested in sharing a ride with others?

- ❑ I like the independence of having my car.
- ❑ Carpooling takes too much time.
- ❑ I like the privacy of driving alone.
- ❑ I need my car for business reasons.
- ❑ I need my car for other personal reasons before or after work.
- ❑ Inconvenient to wait for others.
- ❑ I can't have flexible work times.
- ❑ I have an irregular work schedule.
- ❑ Other (specify) _____

22) Do you think carpooling is a good option for Sri Lanka?

- ❑ Yes, I do

- ❑ No, I don't
- ❑ I don't have an idea

- Any Additional Comments

Survey to collect fuel consumption details