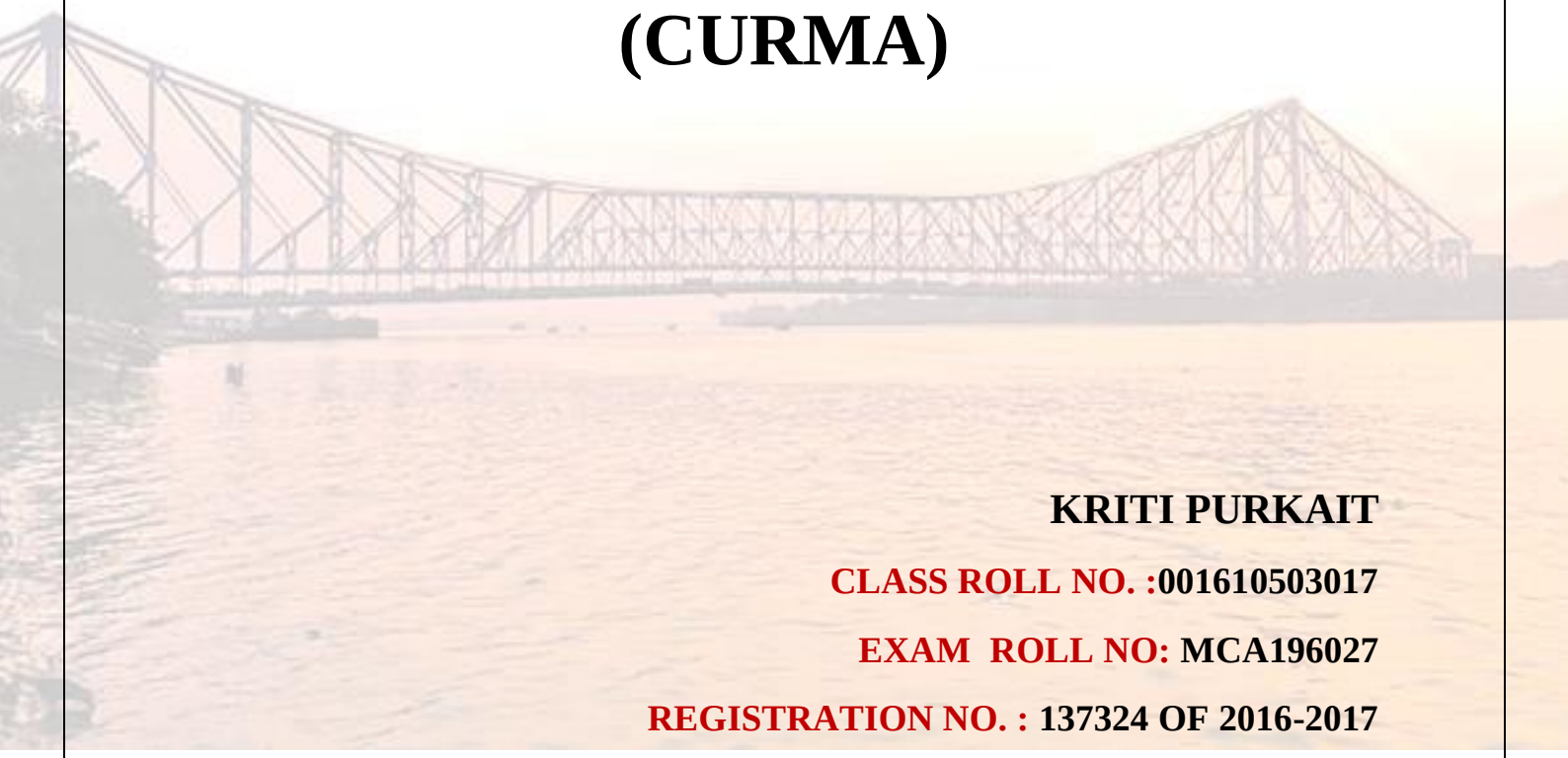
ROUTE OPTIMIZATION FOR CROWDSOURCED URBAN AREA ROAD MONITORING APPLICATION

(CURMA)



KRITI PURKAIT

CLASS ROLL NO. :001610503017

EXAM ROLL NO: MCA196027

REGISTRATION NO. : 137324 OF 2016-2017

MCA THIRD YEAR (SECOND SEMESTER)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

JADAVPUR UNIVERSITY

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF RECOMMENDATION

This is to certify that the project entitled "ROUTE OPTIMIZATION FOR CROWDSOURCED URBAN AREA ROAD MONITORING APPLICATION (CURMA)" has been satisfactorily completed under my guidance and supervision by **Kriti Purkait** (University Registration No.:137324 of 2016-17, Examination Roll No.: MCA196027,Class Roll No.: 001610503017).I hereby recommend that the project be accepted in partial fulfilment of the requirement for the Degree of Master of Computer Application, Department of Computer Science and Engineering in Faculty of Engineering and Technology, Jadavpur University, Kolkata for the academic year 2018-2019.

_____

Dr. Chandreyee Chowdhury (Project Supervisor)
Assistant Professor
Department of Computer Science and Engineering
Jadavpur University, Kolkata-700032

Countersigned

_____

Prof. Mahan Tapas Kundu
Head, Department of Computer Science and Engineering,
Jadavpur University, Kolkata-700032.

_____

Prof. Chiranjib Bhattacharjee
Dean, Faculty of Engineering and Technology,
Jadavpur University, Kolkata-700032.

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL

This is to certify that the project entitled "ROUTE OPTIMIZATION FOR CROWDSOURCED URBAN AREA ROAD MONITORING APPLICATION (CURMA)" is a bona fide record of work carried out by Kriti Purkait in partial fulfilment of the requirements for the award of the degree of Master of Computer Application in the Department of Computer Science and Engineering, Jadavpur University during the period of February 2019 to May 2019. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn there in but approve the thesis only for the purpose for which it has been submitted.

_____
Signature of Examiner
Date:

_____
Signature of Supervisor
Date:

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this project entitled "ROUTE OPTIMIZATION FOR CROWDSOURCED URBAN AREA ROAD MONITORING APPLICATION (CURMA)" contains literature survey and original research work by the undersigned candidate, as part of her Degree of Master of Computer Application. All information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: KRITI PURKAIT
University Registration No. : 137324 of 2016-17
Examination Roll No. : MCA196027

Project Title: ROUTE OPTIMIZATION FOR CROWDSOURCED URBAN AREA ROAD MONITORING APPLICATION (CURMA)

_____
Signature
Date:

ACKNOWLEDGEMENT

I would like to thank my respected guide, Prof. Dr. CHANDREYEE CHOWDHURY for guiding me and helping me with my project. I am grateful to her, for her constant cooperation and encouragement which has helped me complete my project/thesis.

I would like to convey my gratitude to our respected HOD sir, Prof. MAHAN TAPAS KUNDU and our coordinator Prof. DIGANTA SAHA. I would also like to thank my teammate TANMOY KR DAS and PHD scholar Mrs MOHANA BAKSHI for their kind cooperation and help.

I am highly indebted to Jadavpur University for providing me the opportunity and the required infrastructure to carry on my project. Last but not the least; I would like to thank my parents, for giving me constant encouragement and mental support throughout my work.

KRITI PURKAIT
University Registration No. 137324 of 2016-17
Examination Roll No. : MCA196027
Master of Computer Application
Department of Computer Science and Engineering
Jadavpur University

CONTENTS

CHAPTER 1

INTRODUCTION

- CROWDSOURCING A GROWING TREND
- APPLICATIONS OF CROWDSOURCING
- MOTIVATION
- CONTRIBUTION
- ORGANIZATION OF THE REPORT

CHAPTER 1: INTRODUCTION

1.1. CROWDSOURCING A GROWING TREND

In recent years crowdsourcing has evolved significantly as a problem solving and business model. Many top global brands like Nestle, Coca-Cola, Toyota etc are reaching out to the crowd for their innovation and to augment their marketing efforts. Now the question arises, where from this term 'crowdsourcing' come from, and what is it? The term crowdsourcing is actually a combination of crowd and outsourcing which was coined in the year 2006 by Jeff Howe in an article "The Rise of Crowdsourcing" written by him in Wired Magazine. Crowdsourcing is an emerging way of involving human intelligence in performing information seeking and computational tasks where large crowds may take part in social computations for a variety of motivations, which include non-monetary ones, such as public recognition, fun, or the genuine wish of contributing their knowledge to a social process [1]. It consists of building data sets with the help of a large group of people. The people i.e. the crowd act as a source and as data suppliers who are willing to enrich the data with relevant, missing, or new information.

The spread of the internet, decreasing cost of production and increasing accessibility of information makes it easier for the crowd to participate and is regarded as a factor leading to the success of crowd sourcing. Crowd sourcing allows people to do their tasks anytime and anywhere. Crowdsourcing help reduce costs, speed up project timelines, tap in to crowd intelligence and creativity, and engage citizens at all levels of corporate and government processes. It's not just for new startups; iconic organizations including NASA and GE have specialist teams managing mass engagement with external stakeholders. There are several types of crowdsourcing such as crowd voting, crowd funding etc [2].Crowd sourcing is advantageous if human intelligence is used in the right direction and if valid information is contributed by the users i.e. the crowd, otherwise it may lead to poor quality entries and misleading information. In crowd sourcing applications there are two groups of people namely service providers and data providers. In some cases both are distinct group of people but in other cases they are the same group of people. There are four primary types of crowdsourcing: – 1) Crowd Wisdom; 2) Crowd Creation; 3) Crowd Voting and 4) Crowd Funding described as follows:

- **Crowd Wisdom**: The "Wisdom of Crowds" principle attempts to utilize the knowledge of crowd in order to solve problems or predict future outcomes or for help in direct corporate strategy. Examples of crowd wisdom comprise idea jams and prediction markets such as the Iowa Elections Market, Hollywood Stock Exchange and SIM Exchange.

- **Crowd Creation**: The pre-eminent known forms of crowdsourcing are "creation" activities such asking individuals to film TV commercials, perform language translation or solve challenging scientific problems. Examples of Crowd Creation include NASA's ClickWorkers, Threadless.com etc.

- **Crowd Voting**: Crowd voting utilizes the community's judgment to organize, filter and stack-rank content like newspaper articles, music and movies. It is the most popular form of crowdsourcing, which generates the highest levels of participation. Google's search engine is built upon the principle of Crowd Voting. Reality TV shows add to another example of Crowd Voting. Threadless.com uses crowd voting to decide which T-shirts it will manufacture and sell on its web site.

- **Crowd Funding:** Crowd-Funding evades the traditional corporate establishment to offer financing to individuals or groups that might otherwise be denied credit or opportunity. Kiva a micro lending portal is an example of crowdsourcing. It provides a marketplace for aspiring entrepreneurs in developing nations to seek out financing for projects that are not readily available in their home markets. Sellaband provides a similar value proposition to Garage bands that have been turned down by the major record labels in Hollywood.

1.2 APPLICATIONS OF CROWDSOURCING

One of the first ever instance of crowdsourcing is the Oxford English Dictionary. The project aimed to list all the words that have any recognized lifespan in the Standard English language with their definition and explanation of usage. That task was a huge one. So the dictionary creators invited the crowd to help them on a voluntary basis.

By now, crowdsourcing has firmly ingrained in both our societies and our economies. This paved the way for the next-level crowd sourcing sites such as Amazon Mechanical Turk, Fiverr, Upwork, TaskRabbit, 99designs[3].Some applications of crowdsourcing are described as follows:

- Amazon Mechanical Turk (MTurk) is an Internet crowdsourcing marketplace enabling individuals and businesses (known as Requesters) to coordinate human labor to perform tasks that computer is currently unable to do. It is operated under Amazon Web Services and is owned by Amazon. Employers post jobs known as Human Intelligence Tasks (HITs), such as identifying specific content in an image or video, writing product descriptions, or answering questions, among others. Workers, colloquially known as Turkers or crowdworkers, browse among existing jobs and complete them in exchange for a rate set by the employer. To place jobs, the requesting programs use an open application programming interface (API), or the more limited MTurk Requester site.

- Wikipedia, a free, web-based, multilingual and collaborative encyclopedia built on a non-for-profit business model. The platform has more than 100,000 active volunteer contributors who add new knowledge to the system daily.

- In 2008 Starbucks launched a crowdsourcing initiative to raise customer satisfaction levels and innovation within its outlets, called My Starbuck Ideas.A team of Idea Partners has so far read over 312,000 customer submissions. Taking in to account public votes and their level of innovation, a selection of ideas are regularly presented to Starbucks' key decision makers to figure out how to implement them.

- The US Air Force is currently short more than 1,500 pilots, largely due to pilots choosing to go to private airlines rather than re-enlist. A major part of on-going efforts to improve re-enlistment of all personnel is through involvement with the Airmen Powered by Innovation initiative. It was launched in November 2014 to crowdsource ideas and contributions from serving airmen to improve procedures remove pointless activities and reduce costs.18 months later API had received 6,791 ideas, of which 192 were approved by Air Force leadership and they represent $121.3 million in projected savings. A year on to the end of May 2017 and the Air Force has now received about 12,000 ideas, and also currently receives about 5,000 visits per day from airmen interested in commenting, sharing and voting on ideas that will improve squadron units.

1.3 MOTIVATION

Road conditions in Indian cities have improved a lot in last few years. But still due to poor drainage system in certain places, water logging takes place during monsoon season which leads to pandemonium during busy hours with increasing traffic. Due to increasing population, the traffic conditions of many cities have worsened. There are frequent traffic jams and some places are prone to accidents. Moreover though the country has taken a lot of measures for the empowerment of women but still there are roads in its city and other places that are unsafe for women at night. Another problem for people here is, there are not sufficient amount of public toilets in roads and with the growing era of digitalization and internet there are very few roads in the city that have a Wi-Fi zone.

Now, often it will be beneficial for people if they come to know about the road conditions of the city i.e. whether water logging occurs or not in a particular route between places during monsoon, whether the route is safe for women at night, whether the route is prone to traffic jams or accidents or not. Keeping this present scenario in mind we have thought that crowd sourcing i.e. the common people of the city who have travelled through different routes can help others with all these information, through their feedback on different routes. Now Google maps provide us with the shortest route between any two places and recently it has added a feature which shows the current traffic jam of an area and Google Maps users can now report accidents, speed traps [4]. But still it does not provide the places where water logging takes place or whether it is safe for women, Wi-Fi zone area and moreover what common people have to say about a particular route.

All of the above reasons led us in developing a Crowdsourced Urban Area Road Monitoring Application(CURMA).The application we have designed based helps users to get information about these particular parameters with the scope of what percentage of people have a positive or negative feedback. It is always better to take information from the common people or pedestrians who are actually using the road. And nowadays since internet is accessible to almost all so giving a feedback is not that difficult. Now there can be several other parameters like public toilets, sufficient number of dustbins etc on which feedback can be taken for the routes. Later this crowdsourced data is used to give information about the routes between areas and what is the best (optimized) route depending on the criteria that the user gives more priority on. But we have currently worked with only five of the parameters on which user feedback is taken and two of them which is sensitive is used for optimization which are further described in Chapter 3 : ‘Description of CURMA’ in the following pages.

Now the question arises why have we done Route Optimization for CURMA. Now suppose the users of CURMA receive information generated based on the crowdsourced data in the form of feedback about say more than three alternative routes between any two places

depending on the order of priorities given by them on a large number of parameters(more than 5) like traffic condition, women safety, accident vulnerability etc. If we leave it for them to decide which is the best route for them depending on their prioritized criteria it will take a lot of time for them to decide which is the optimized route .i.e. the best route for them to travel to. Thus to find the best possible route depending on certain criterion and constraints and to save their time, and increased convenience for the users route optimization is done by the application.

1.4 CONTRIBUTION

Crowdsourced Urban Area Road Monitoring Application (CURMA) is a team project. It has mainly two modules in it. First being 'Validation framework for CURMA' and second being the 'Route optimization for CURMA'.I have done the second module. Route optimization for CURMA' is done using Genetic algorithm .After validation of the crowdsourced data which is contributed by the crowd on various routes between any two places in Kolkata in the form of feedback ,the statistics of various alternative routes between two selected areas are generated depending on the list of priorities given by the users. After the generation of the route statistics for the alternative routes i.e. the candidate solutions Genetic algorithm is used to determine the best possible (optimized) route for the users based on their priorities such as traffic condition, women safety. This helps the users to decide which route is best for them to travel quickly.

I have also developed the backend of the application along with the database connectivity.

1.5 ORGANIZATION OF THE REPORT

- **RELATED WORK:**

  It consists of the previous works done on route optimization based on crowdsourced data, their description and references.

- **DESCRIPTION OF CURMA:**

  This portion of the report consists of an overview of what CURMA is and what are the modules. CURMA mainly consists of two modules namely the Validation Module and the Route Optimization module. Each of the modules again consists of two sub modules namely the Input Module and the Output Module.

- **ROUTE OPTIMIZATION:**

  This portion consists of displaying all the candidate solutions i.e. all the alternative routes with route statistics between any two areas chosen by users based on their priorities and the valid data which after validation was stored in the database. It consists of what Genetic algorithm is and how Genetic algorithm is applied on all the candidate solutions to find the best possible (optimized) route between any two selected areas.

- **IMPLEMENTATION DETAILS:**

  It consists of how the CURMA and Route Optimization of CURMA was implemented and what tools were used to implement it .It consists of the screenshots of the real database ,user interface, the result i.e. how the displayed alternative route statistics look like when displayed and finally the optimized route according to the user priorities.

- **CONCLUSION AND FUTURE WORK**

It comprises of how CURMA can be used not only for monitoring roads but the same data validation and optimization framework can be used for other crowd sourcing applications which use user rating and feedback. The report generated by the 'Route optimization of CURMA' which displays the route statistics and the optimized route can be even used by some government authority to survey road condition and improves roads.

CHAPTER 2

RELATED WORK

CHAPTER 2: RELATED WORK

With growing population and technical advancement researchers are trying to utilize the crowd to solve various complex problems. Till present day many Crowdsourcing platforms have come up like MTurk, Upwork etc. Researchers have also proposed certain algorithms for optimization of certain tasks in crowdsourcing system. One such is an efficient algorithm for database query optimization in crowdsourcing system. Since Query optimization provides declarative query interface in relational database management system it is important to crowdsourcing system. In their proposed approach, a technique called crowd optimization is proposed in which the user has to submit an SQL query and the system compiles the query, generating the execution plan and evaluating that query, giving an optimized plan to the user. In relational database systems, query optimization provides query interfaces, which are important for crowdsourcing. The system considers cost and latency in query optimization and generates query plans that give a good balance between the cost and latency. Efficient algorithms are used for optimizing four types of queries are used i.e. selection queries, join queries, complex queries and order-by queries. [5]Researches have been done on Route optimization for fully electric vehicles via crowdsourcing. Route planning for fully electric vehicles (FEVs) must take energy efficiency into account due to limited battery capacity and time-consuming recharging. In addition, the planning algorithm should allow for negative energy costs in the road network due to regenerative braking, which is a unique feature of FEVs. A framework has been proposed for energy-driven and context-aware route planning for FEVs. It has two novel aspects: 1) It is context aware, i.e., the framework has access to real-time traffic data for routing cost estimation; and it is energy driven, i.e., both time and energy efficiency are accounted for; which implies a bi objective nature of the optimization. In addition, in the case of insufficient energy on board, an optimal detour via recharge points is computed. [6]

CHAPTER 2: RELATED WORK

CHAPTER 3

DESCRIPTION OF CURMA

CHAPTER 3: DESCRIPTION OF CURMA

In Crowdsourced Urban Area Road Monitoring Application (CURMA) crowdsourced means the source of data for the application is the crowd, i.e. the common people. The term Urban Area Road Monitoring is given since the application can be used to collect data from the crowd on different routes between various places in urban area.
In CURMA we have considered the city of Kolkata as the urban area and the crowd is the people of Kolkata. Through CURMA information is collected on several road parameters like traffic condition, accident vulnerability, water log, women safety, wifi facility, number of public toilets in the road, whether there is potholes on the road etc from the crowd as feedback and later this data collected is used to generate a report or statistics describing road conditions depending on these parameters and an optimized roué based on user given priority on these parameters.

Now the question comes why did we think of developing such an application and why have we depended on the crowd to contribute data to our application. The reason behind the thought it is that, maps like Google maps and others do not provide information about the road parameters that I have mentioned in the earlier Para and do not allow users to give priority on their choice of parameters on which they want to see an optimized route between two selected areas. We have considered the crowd i.e. the people of Kolkata as our data source because it is the people who actually live in Kolkata can give the actual information about the route conditions and in today's 21st century with the increasing availability of internet it will be easier and convenient for people to contribute data i.e. give feedback as well as easier for them to get route information.

Generally in every crowdsourcing application after data collection part there is data validation part. After data validation it is stored in a database for further analysis. For any crowd sourcing application, building data sets with the help of large group of people is the most elementary level work. The big question is that how data can be collected in an organized manner so that it can fit into specified applications. We can use web based survey tool or shared spreadsheets [7] to collect data from the crowd. Online tolls like Google forms, Survey Monkey allow us to build survey quickly and cheaply.

CURMA mainly has two modules namely:

- **Validation module** where the validation of crowdsourced data is done before storing it in the database for further analysis of the data.
- **Optimization module** where the optimization is done on the validated data to find out the optimal i.e. the best solution depending on certain constraints.

CHAPTER 3: DESCRIPTION OF CURMA

The proposed workflow of the validation and optimization module is shown in fig 1 and the descriptions of the modules are given as follows:

- **Validation module**: This module again consists of the input module and the output module. The input module comprises of data collection from the crowd as feedback on different parameters like traffic condition, accident vulnerability, women safety, water log etc. on alternative routes between two user selected areas. The output module consists of the data validation. Generally collected data are in form of raw data and needed to make sense of it very quickly. Raw data can be stored into the database in tabular forms after passing through some validation process. Validation can be carried out through Automatic validation, Authoritative data comparison and Model based validation [8]. In CURMA data is validated using different three phases of validation. The first being Geo location based validation where it is validated whether the user who is giving the feedback is actually present on the route or not. Second is the screen time based validation where it is checked whether the user takes the required time to properly read the feedback page and give their feedback or not. The third and final phase of validation consists of the machine learning based validation where a classifier is to classify data as valid and finally store in the database.

- **Optimization module:** This module again consists of the input module and the output module. The input module comprises of data collection in the form of user priority on the route parameters like traffic condition, women safety etc. The output module consists of the generation and display of all the candidate solutions i.e. the route statistics of each of the alternative routes containing the percentage of users saying about the different conditions of the routes and the generation of the optimized solution (route) based on user priority. Now for users convenience optimization of the routes is done based on user priority using an optimization algorithm. Optimization algorithms are used to narrow down a set of candidate solutions to an optimum solution. Among the optimization algorithms we have chosen (GA) Genetic Algorithm to find out the optimal i.e. the best route according to user priority. The route optimization using GA is discussed in detail in Chapter 4: Route Optimization section 4.4.

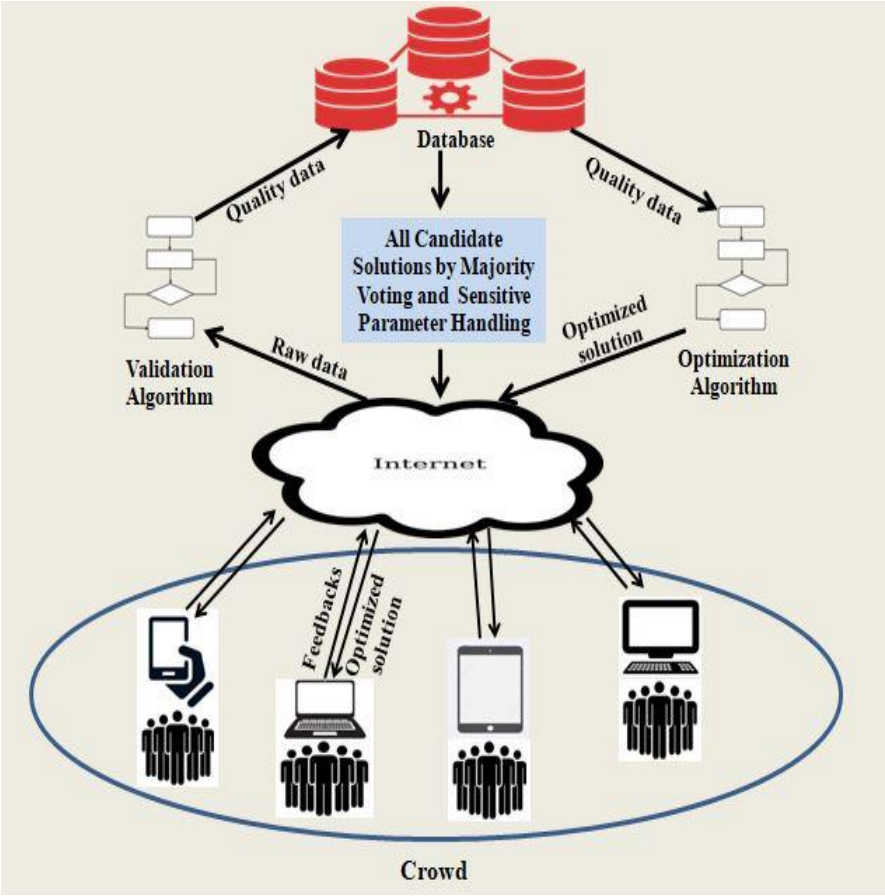


Figure 1: Proposed Workflow for Crowd Sourced Data Analysis using Pre Validation and Post Optimization

CHAPTER 4

ROUTE OPTIMIZATION

- NEED FOR ROUTE OPTIMIZATION
- GENETIC ALGORITHM
- USE OF GENETIC ALGORITHM IN ROUTE OPTIMIZATION
- GRAPHS AND CHARTS

CHAPTER 4: ROUTE OPTIMIZATION

Optimization means finding an alternative with the most cost effective or highest achievable performance under the given constraints, by maximizing desired factors and minimizing undesired ones. In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function.

The purpose of optimization is to achieve the "best" design relative to a set of prioritized criteria or constraints. These include maximizing factors such as productivity, strength, reliability, longevity, efficiency, and utilization.

4.1 Need For Route Optimization

Route Optimization means obtaining the best possible root between any two areas based on certain priorities or constraints. It is beneficial as getting the optimized route between two areas depending on parameters like traffic condition will make it easier for people to decide, travelling through which roads will save their time; reduce their vehicle's fuel costs. Less time spent on the road will reduce the need for maintenance on fleet vehicles. Compared to manually designed routes and other less-advanced route mapping solutions, route optimization software or application can reduce the amount of time fleet managers spend on trip planning. It reduces transportation delays.

Route optimization based on women safety can help women of the cities to avoid using routes which are marked unsafe according to the feedback given by most of the users. There can be several other such parameters like accident vulnerability, water log etc. that can be used in route optimization which will help people choose routes according to their priority. Route Optimization for CURMA can help users to obtain the best route depending on the priorities they want according to their need i.e. whether they want it based on traffic condition or women safety or it can be with any other parameter like accident vulnerability, water log, wifi facility, number of public toilets etc. by obtaining opinion directly from the general people who are actually using the road. This is unlike Google Maps or other maps that provide users with the shortest route based on distance, and only allow users to mark places that are prone to accidents(recent feature added in Google Map) that too with their own algorithm and do not allow users to give priority on which parameters they want the best route. I have done with only two parameters i.e. Traffic condition and Women safety depending on which optimized rote will be generated. The procedure is described below in 'Use of Genetic Algorithm in Route Optimization '

CHAPTER 4: ROUTE OPTIMIZATION

Optimization algorithms are used to narrow down a set of candidate solutions to an optimum solution. Among the optimization algorithms I have chosen (GA) Genetic Algorithm for route optimization of CURMA to find out an optimum solution i.e. the best (optimum) route because in Route optimization of CURMA is multi-objective optimization is done by calculating the fitness function using the classical approach Weighted Sum Approaches and being a population based approach, GA is well suited to solve multi-objective optimization problems.

4.2 Genetic Algorithm:

Genetic algorithms (GA) are search algorithms based on the principles of natural selection and genetics. It was introduced by J Holland in the 1970's and inspired by the biological evolution of living beings. In the case of GA, the candidate solutions are the individuals in the population which are evolved by the algorithm. The evolution usually starts from a population of randomly generated individuals. It is an iterative process, with the population in each of the iteration called a generation. In each generation, the fitness of every individual in the population is evaluated using a fitness function. The more fit individuals are selected from the current population, and each individual's genome is modified to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm. The algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population [9].

Genetic Algorithm can be summarized as follows:

1) Randomly initialize population P

2) Determine fitness of population

3) Until convergence repeat:

   a) Select parents from population

   b) Crossover and generate new population

   c) Perform mutation on new population

   d) Calculate fitness for new population

Genetic algorithm works on a set of initial population i.e. a set of candidate solutions.

4.3 Generation And Display Of Candidate Solutions

After validation and storage of data in a crowdsourced application a set of candidate solution is generated using some concept on which optimization is done. In our application after data validation, it is checked whether a sufficient amount of feedback i.e. data (say >n) is available for a particular route say R1 between two areas X and Y or not. If yes, it is further seen what majority of the users say about a particular user prioritized parameter for e.g. traffic condition etc. Then percentage is calculated on it and candidate solutions i.e. the route statistics for route each of the routes $R1, R2, \ldots\ldots Rm$ between the two areas X and Y is generated and displayed which contains what majority of the crowd says about the particular parameter for e.g. traffic condition , whether it is high , moderate or low. If sufficient amount of feedback is not available i.e. data (<n) then it is displayed insufficient amount of data is available. Now route optimization can be done on the set of these routes (candidate solutions) using Genetic Algorithm (GA) which is described below:

4.4 Use of Genetic Algorithm in Route Optimization

Basic steps of GA according to the problem formulation in our application are given as:

Begin

1. Initialize population with the set of candidate solutions (alternative routes) generated between two user selected areas and keep the reference in a suitable data structure.
2. Compute fitness value using fitness function

   $$Fitness\ value = \sum(prioritized\ parameter\ value \times 100 \times weight)$$

   Where, weight $= 1$ if parameter is first priority

   $= 0.5\ if\ parameter\ is\ second\ priority$
3. If a distinct minimum fitness value is achieved go to Step 8.
4. Select the two chromosomes (alternative routes between the selected areas) with identical minimum fitness value from the population as parents.
5. Crossover the parents i.e. swap the second priority (tail) values of the chromosomes to generate new population.
6. Perform mutation i.e. swap the head and tail values of each chromosome itself (using swap mutation) to generate the new population.
7. Go to step 2
8. Output the best solution (one with minimum fitness value) along with its reference and stop.

End

The basic steps of GA are now described in detail as follows:

4.4.1 Population Initialization

Initial population consists of all candidate solutions i.e. say n routes (chromosomes) between two places X and Y. Each chromosome is encoded or represented either by binary representation or integer representation. We have used integer representation where the range lies from 1 to 3. Now care must be taken so that after crossover and mutation the number does not go outside the range. Each route is evaluated on the basis of k specific parameters (genes) chosen based on user priority. The parameter values (alleles) of k parameters lie in a specific range as already mentioned in the Domain details. We have used an integer representation for encoding the chromosomes. The population initialization is depicted in figure 2 where $P1, P2, \ldots\ldots\ldots\ldots\ldots Pk$ are the k parameters (genes) chosen based on user priority and $R1, R2, \ldots\ldots\ldots Rn$ are the n routes (chromosomes) between two places X and Y. $Value11, Value12 \ldots\ldots\ldots Value1k$ are the k priority based parameter values for each of n routes.

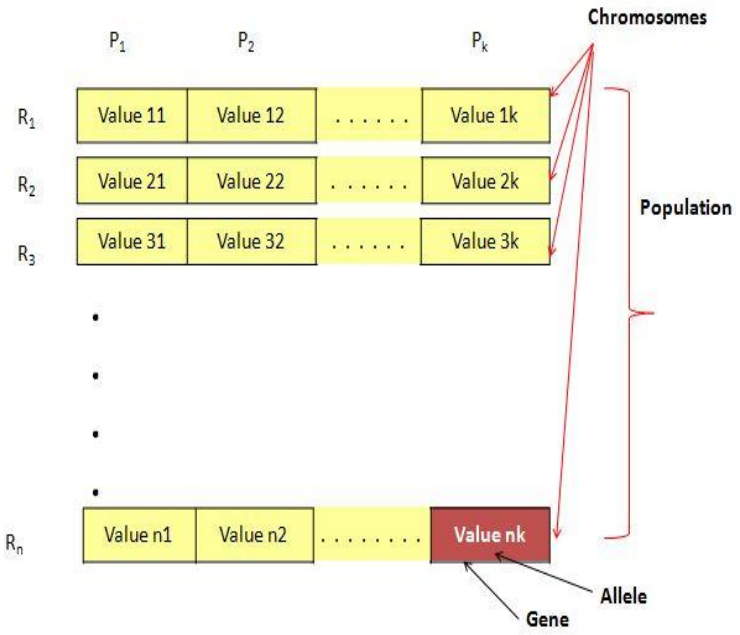


Figure 2: Population initialization

For example: Let our initial population consists of 3 roads between two areas namely Jadavpur and Howrah as given in Table 1:

Route	Traffic condition	Women safety
Jadavpur Howrah 1	3	1
Jadavpur Howrah 2	1	2
Jadavpur Howrah 3	2	2

Table1: All possible candidate solutions according to our problem.

Let us assume "women safety" is chosen as first priority and "traffic condition" as second priority.

Parameters:

1. Traffic Condition (1st priority)

2. Women Safety (2nd priority)

4.4.2 Fitness function computation

Fitness Function (also known as the Evaluation Function) evaluates how close a given solution is to the optimum solution of the desired problem. It generates a fitness score or value which determines how fit a solution is. There is no hard and fast rule that a particular function should be used in a particular problem [11].

$$Fitness\ value = \sum(prioritized\ parameter\ value \times 100 \times weight) \qquad (1)$$

Where, weight = 1 if parameter is first priority
= 0.5 *if parameter is second priority*

From the initial population of the example mentioned above using the fitness function (1) we get:

Fitness value of route Jadavpur Howrah 1=$3 \times 0.5 \times 100 + 1 \times 1 \times 100 = 250$

Fitness value of route Jadavpur Howrah 2= $1 \times 0.5 \times 100 + 2 \times 1 \times 100 = 250$

Fitness value of route Jadavpur Howrah $3 = 2 \times 0.5 \times 100 \times 2 \times 1 \times 100 = 300$

4.4.3 Selection

The idea of selection phase is to select the fittest individuals and let them pass their genes to the next generation. Two pairs of individuals (parents) are selected based on their fitness values or scores. Individuals with high fitness have more chance to be selected for reproduction. In Genetic algorithms either the maximum or minimum fitness value is taken depending upon the fact that whether it is a maximization problem or a minimization problem. In our application we have taken the minimum fitness value since for us (1=best, 2=moderate, 3=worst). If only one value is achieved then we terminate the process and return the corresponding route as the optimized route. Else if more than one fitness value has same minimum value then we proceed to the crossover step. Now for our example as mentioned previously we have achieved more than one minimum value, so we proceed to the crossover step.

4.4.4 Crossover

The crossover operator is analogous to reproduction and biological crossover. In this, more than one parent is selected and one or more off-springs are produced by exchanging the genes of parents among themselves until the crossover point is reached. There are different types of crossovers namely one point crossover, multipoint crossover, uniform crossover etc. In our application we have used one point crossover. Usually a chromosome is depicted in the form of head and tail. Generally the head is the portion with better quality gene than the tail. In one point crossover, a random point is chosen from the tails of both parents. So the genes of the tail are swapped to generate new offspring [12]. This is depicted in Fig 3 and Fig 4 where value 1, value 2 and value 3, value 4 are the genes of chromosome 1 and chromosome 2 respectively.

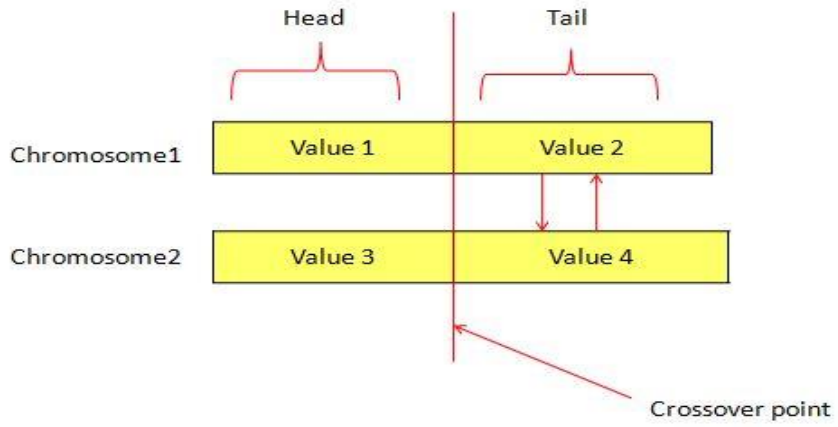


Figure 3: Before one-point crossover between chromosomes.

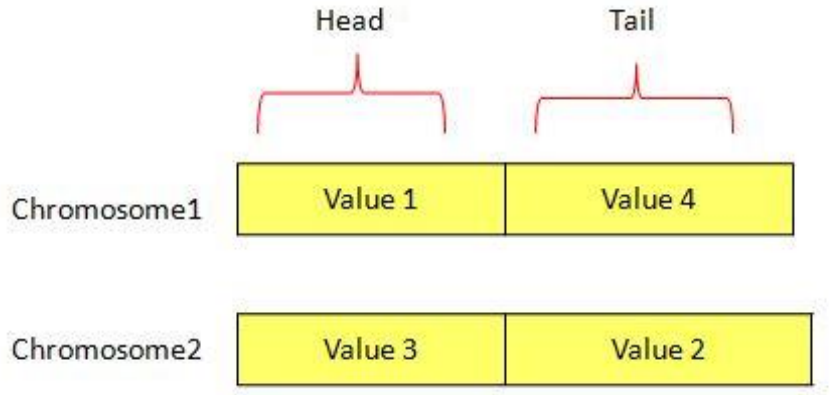
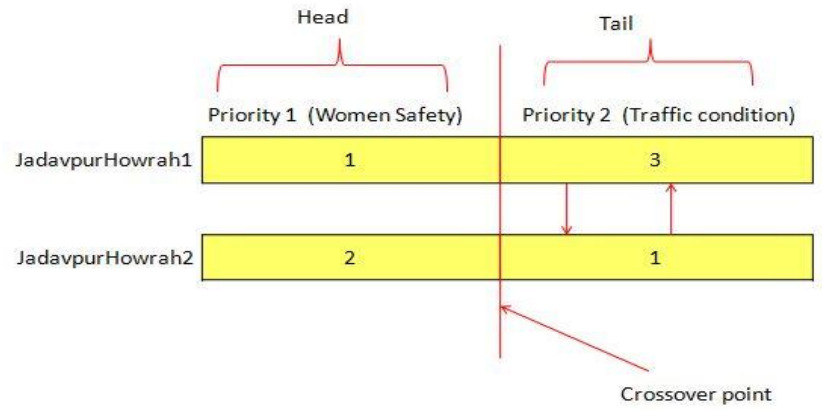


Figure 4: After one-point crossover between chromosomes.

Now for our problem the gene (parameter) with higher user priority is the head and the one with lower user priority is the tail. So in our application the parameter values with low priority are swapped to generate a new offspring. From our previous example “Traffic condition” was the

parameter with low user priority as compared to "Women safety "which has higher priority. So the values of parameter "traffic condition" i.e. 1 and 3 are swapped (crossover).This is depicted in Figure 5 and Figure 6.



.

Figure 5: Before one- point crossover between two routes as per the application.

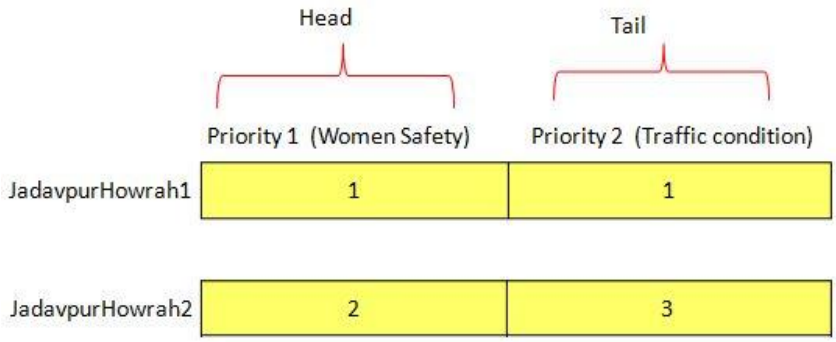


Figure 6: After one-point crossover between two routes as per the application.

4.4.5 Mutation

In mutation for each offspring, some genes are selected and its value is changed. Mutation varies based on the chromosome representation but it is up to us to decide how to apply mutation. If the encoding is binary (i.e. the value space of each gene have just two values 0 and 1), then flip the bit value of one or more genes.

But if the gene value comes from a space of more than two values such as 1, 2, 3 as in case our application then the binary mutation will not be applicable and we should find another way. One way is by swap mutation which we have used in our application. In swap mutation we select two positions on the chromosome at random, and interchange the values. The set of all newly generated individuals will be the new population that replaces the previously used old population. Each population created is called a generation. The process of replacing the old population by the new one is called replacement. From the above example mutation can be done as shown in the Figure 6.The head and tail values of each of the chromosomes are swapped and in route JadavpurHowrah1 even after mutation the values remain same whereas in JadavpurHowrah2 the value of head becomes 3 instead of 2 and tail becomes 2 instead of 3 as shown in Figure 7.

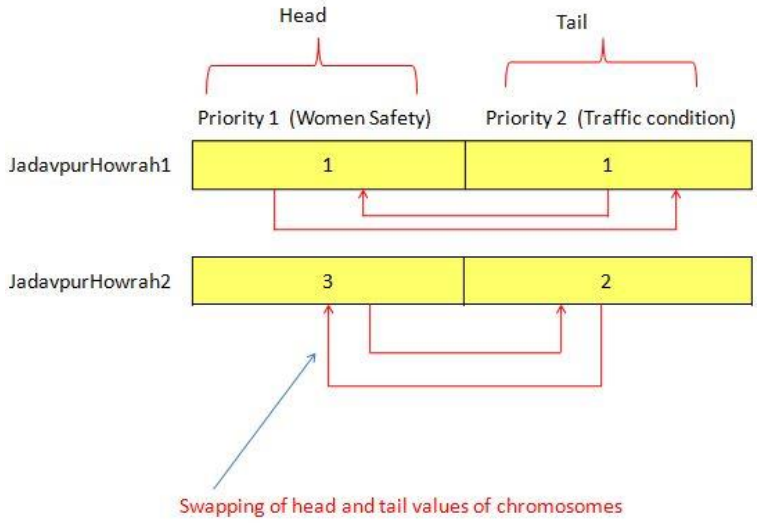


Figure 7: Performing swap mutation on the two routes as per our application

4.4.6 Calculation of fitness function for new population

After mutation new offspring is placed in a new population. Then newly generated population is used for further run of the algorithm i.e. again fitness value is calculated, till the termination condition is reached and the best solution of the current population is returned .In our case the termination is a distinct minimum fitness value. Now though the current population contains newly generated offspring still the first priority field in the chromosomes remains unchanged in the new offspring. Thus the offspring contains characteristics of its parents and hence by back tracking the reference to the offspring's parent we consider the new offspring as JadavpurHowrah1 and JadapurHowrah2 respectively i.e.as they came from the respective parents having these names and calculate their fitness functions to decide the optimized route.

The fitness value evaluation is as follows:

Fitness value of route Jadavpur Howrah 1= $1 \times 0.5 \times 100 + 1 \times 1 \times 100 = 150$

Fitness value of route Jadavpur Howrah 2= $2 \times 0.5 \times 100 + 3 \times 1 \times 100 = 400$

Now we will choose the minimum fitness value which is 150 in this case and return the corresponding route as the optimized route.

Jadavpur Howrah 1 is the optimized route in this case.

4.5 Graphs and Charts

Considering n routes between any two places graphs are plotted for each of the n routes, taking the parameters as X axis and user rating as Y axis. This is depicted by Figure 8.

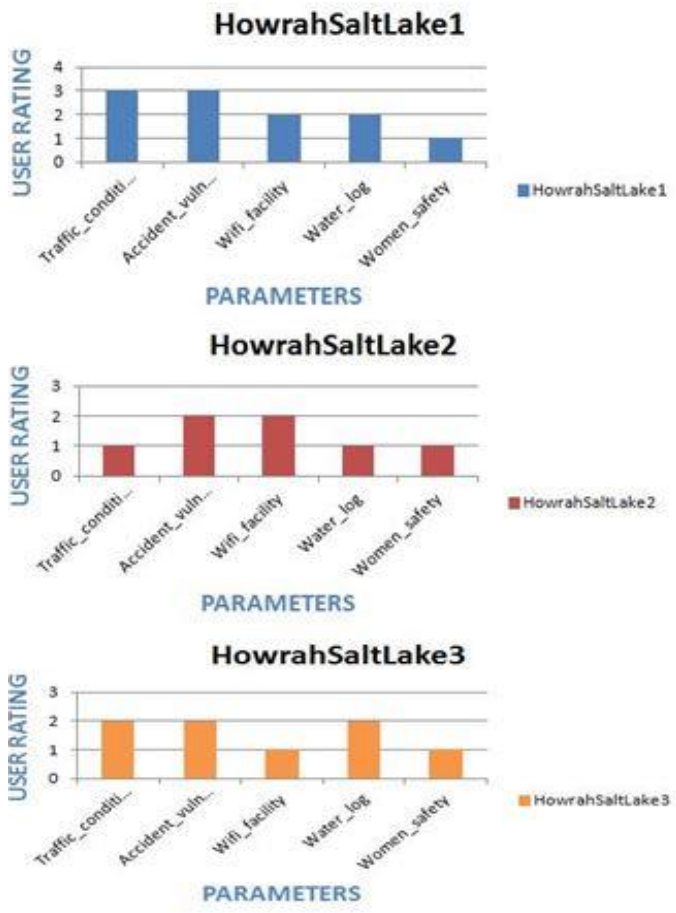


Figure 8: User rating for different parameters between three routes between Howrah and SaltLake.

The statistics of the percentage of feedback given by users saying 'yes' or' No' for each of n routes present between two places i.e. whether wifi facility is available for that particular route or not is depicted in Fig 9.

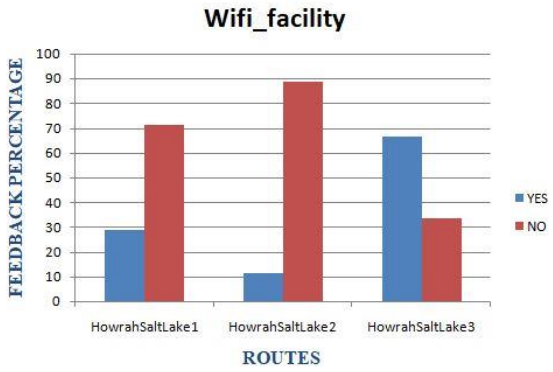


Figure 9: Percentage of feedback given by user for Wifi Facility of three routes between Howrah and SaltLake.

Fig 10 depicts the statistics of the percentage of feedback given by users saying yes or No for women Safety for each of n routes present between two places i.e. whether that particular route is safe for women or not.

Figure 10: Percentage of feedback given by user for Women Safety of three routes between Howrah and SaltLake.

CHAPTER 5

IMPLEMENTATION DETAILS

- SOFTWARE AND TOOLS REQUIREMENT SPECIFICATION
- USER INTERFACES IN ROUTE OPTIMIZATION MODULE

CHAPTER 5: IMPLEMENTATION DETAILS

5.1 Software and Tools Requirement Specifications

- **Operating System Used**

  CURMA is implemented on Windows 10 platform. But it can also be implemented on Linux, Mac OS platform with some changes in specifications.

- **System Configuration**

  - **RAM:** 4 GB or more for better performance.
  - **HDD:** Minimum 20 GB or more.
  - **Processor:** Intel Pentium 4 or any upgraded version.

- **Browser Compatibility**

  Google Chrome, Mozilla Firefox, Internet Explorer, Microsoft Edge, Opera Mini etc.

- **Local Host Server**

  XAMPP Server

- **Back End Tool**

  PHP

- **Front End Tool**

  HTML, CSS

- **Scripting Tool**

  JAVASCRIPT

5.2 User Interfaces in Route Optimization Module

CURMA has two modules as mentioned earlier in 'Description of CURMA', first the Validation Module and second the Optimization module. The user interface and the implementation details of the Optimization module is described as follows:

The name our application is 'Tilottomar Oli Goli'. There are two buttons on the home page of the application as shown below in Fig 11. One is the 'Feedback Form' which leads to the validation module of crowdsourced data where users are allowed to select two areas and submit feedback on any of the routes between them which is stored in the database after three phases of validation. The second button 'Know your roads' leads to the Route optimization module.

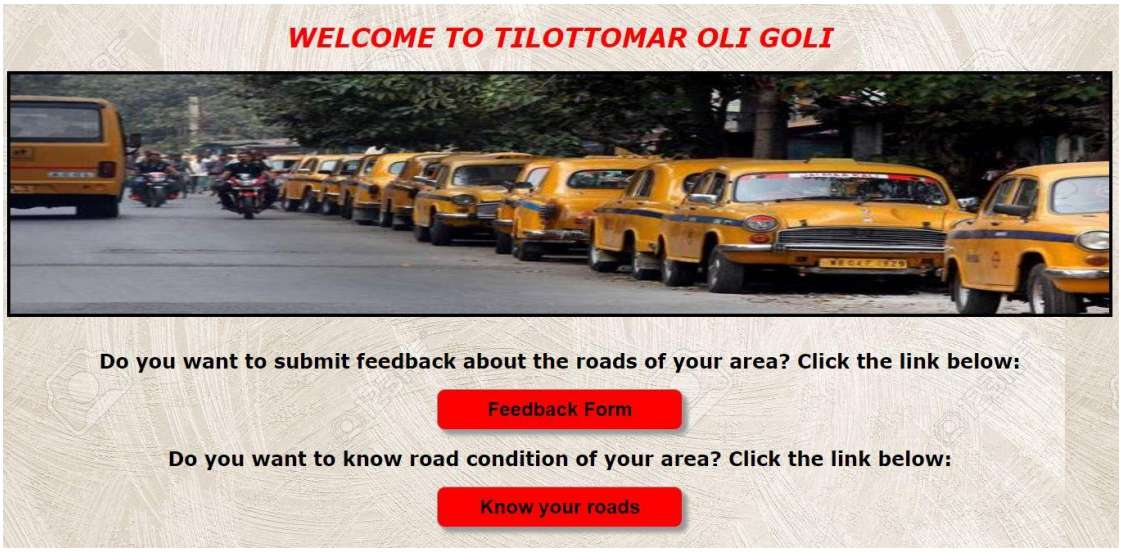


Figure 11: Home page of CURMA

After clicking on 'Know your roads' button we can see the next page as shown in Fig: 12 where the user is allowed to select two areas from the drop down list of areas .The user is also allowed to set priorities between two parameters as first and second priority. One is 'Traffic Condition' another is 'Women Safety'.

CHAPTER 5: IMPLEMENTATION DETAILS

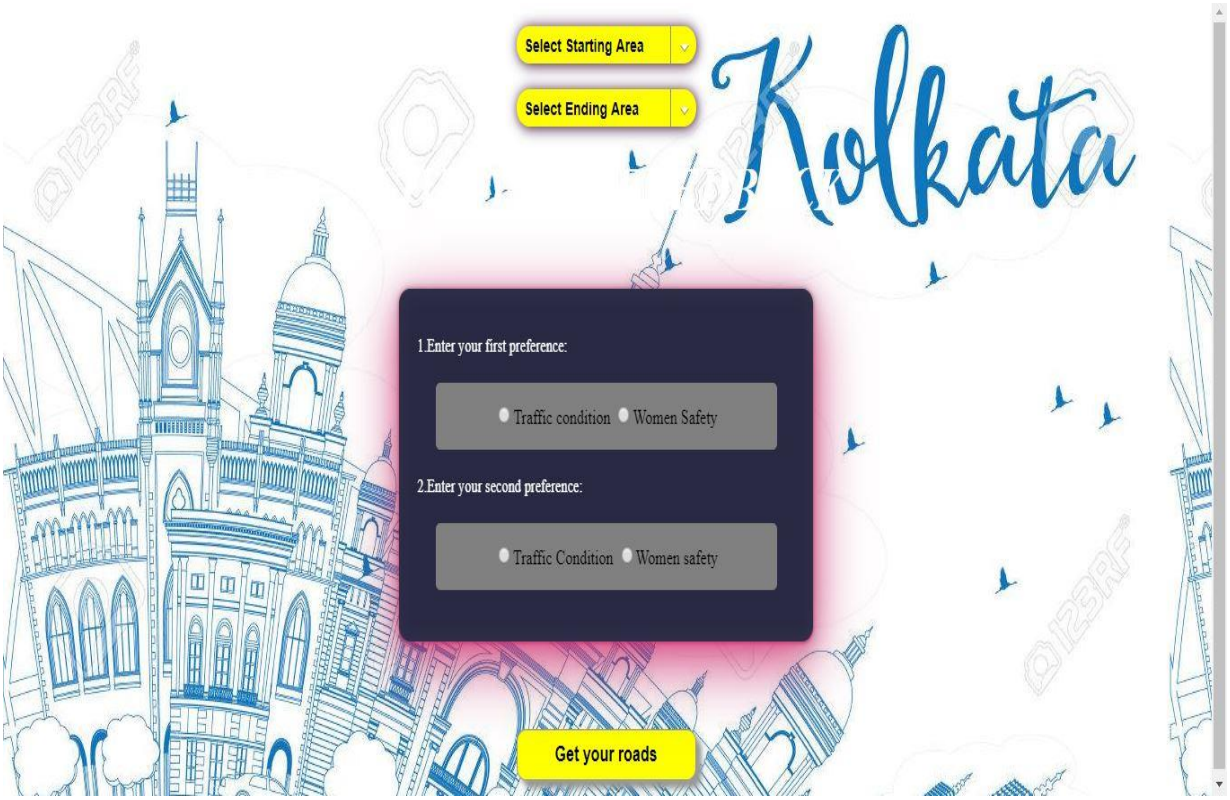
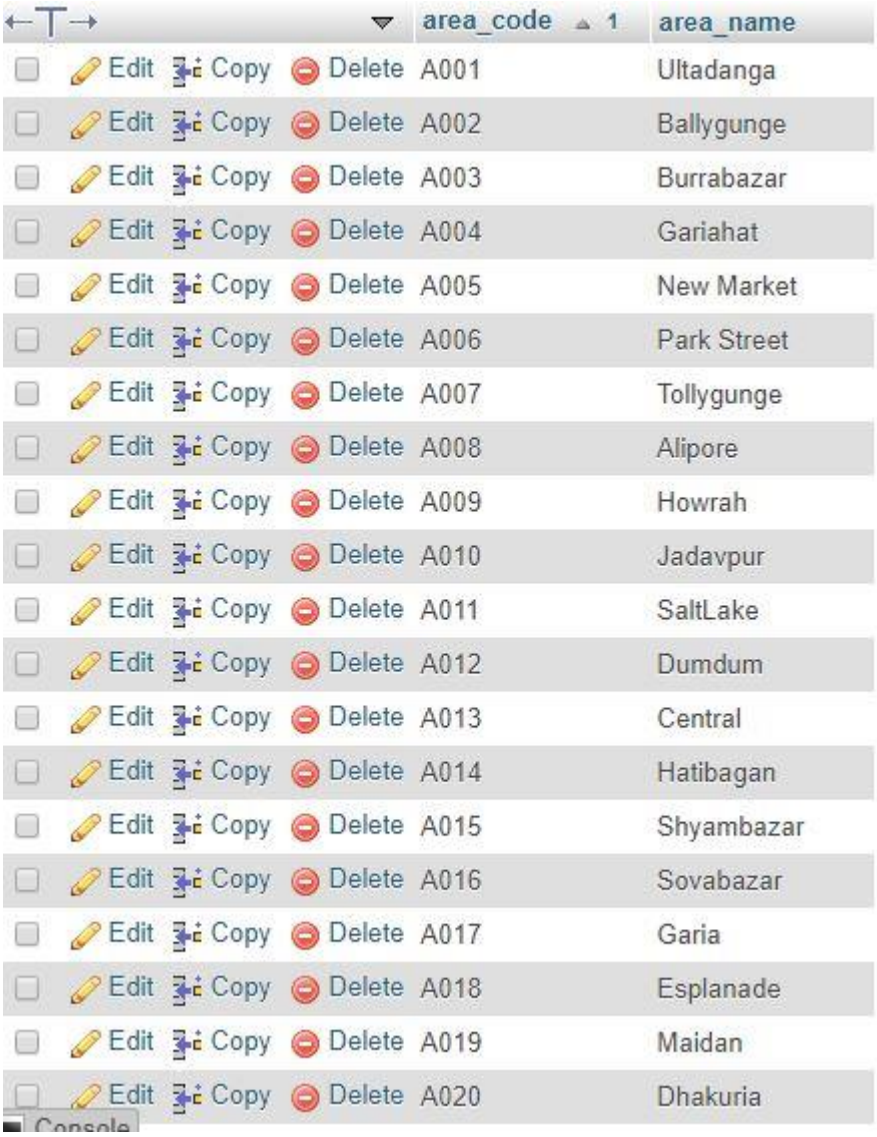


Figure 12: Second page of Route Optimization module of CURMA after clicking on 'Know your roads' button

Now in CURMA we have taken a list of 25 areas of Kolkata. The list of areas is stored along with a unique area code in the database, in a table named ‘area’. This is shown in Fig 13. So among the list of areas we are fetching it in a dropdown box and allowing users to select from it. For e.g. let us select Howrah from the first dropdown box and SaltLake from the second dialogue box. Let the first priority given is Traffic condition and second priority given be Women Safety as shown in Fig 14 below.



	area_code 1	area_name
Edit Copy Delete	A001	Ultadanga
Edit Copy Delete	A002	Ballygunge
Edit Copy Delete	A003	Burrabazar
Edit Copy Delete	A004	Gariahat
Edit Copy Delete	A005	New Market
Edit Copy Delete	A006	Park Street
Edit Copy Delete	A007	Tollygunge
Edit Copy Delete	A008	Alipore
Edit Copy Delete	A009	Howrah
Edit Copy Delete	A010	Jadavpur
Edit Copy Delete	A011	SaltLake
Edit Copy Delete	A012	Dumdum
Edit Copy Delete	A013	Central
Edit Copy Delete	A014	Hatibagan
Edit Copy Delete	A015	Shyambazar
Edit Copy Delete	A016	Sovabazar
Edit Copy Delete	A017	Garia
Edit Copy Delete	A018	Esplanade
Edit Copy Delete	A019	Maidan
Edit Copy Delete	A020	Dhakuria

Console

Figure 13: Snapshot of the area table in the database showing list of areas in Kolkata used in the application

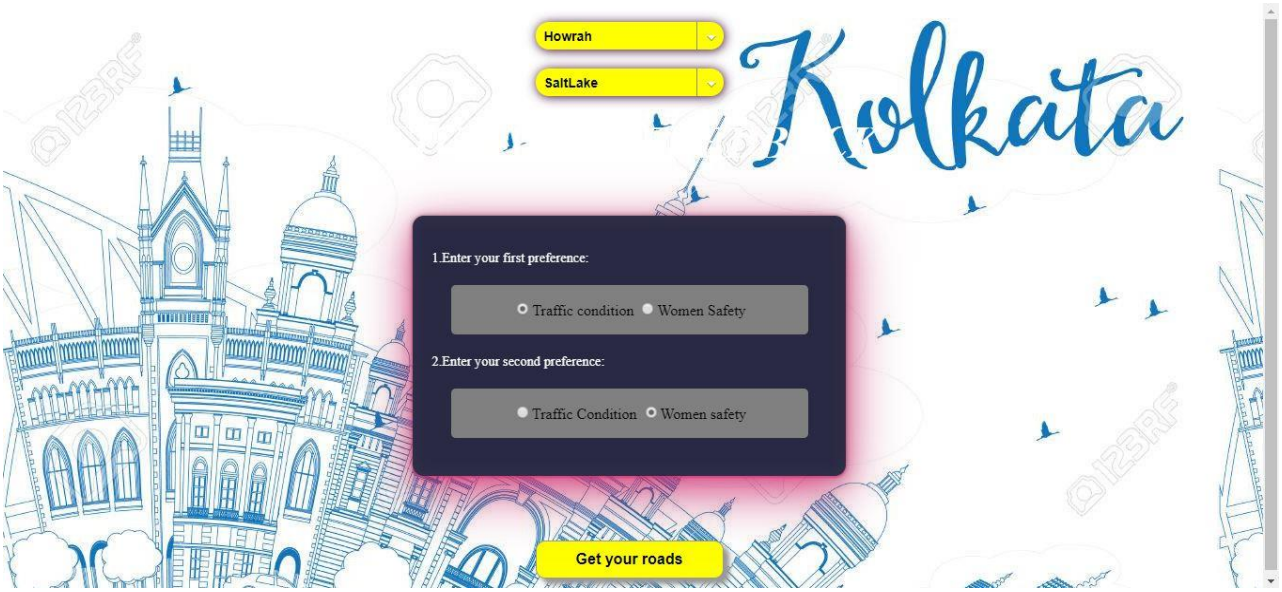


Figure 14: Second page of Route Optimization module of CURMA after clicking 'Know your roads' and selecting the areas and priorities.

Now after clicking on 'Get your roads' button we are directed to the next page where the map of Kolkata is shown using Mapquest API with the two areas Howrah and SaltLake mapped on it and the three alternative routes shown between Howrah and SaltLake. The right hand side dialogue box shows the directions corresponding to the road selected on the map among the three alternative routes given. This is depicted in Fig 15 and Fig 16 below. As we scroll down the page we see the route descriptions of each of the routes are shown in Fig 17.

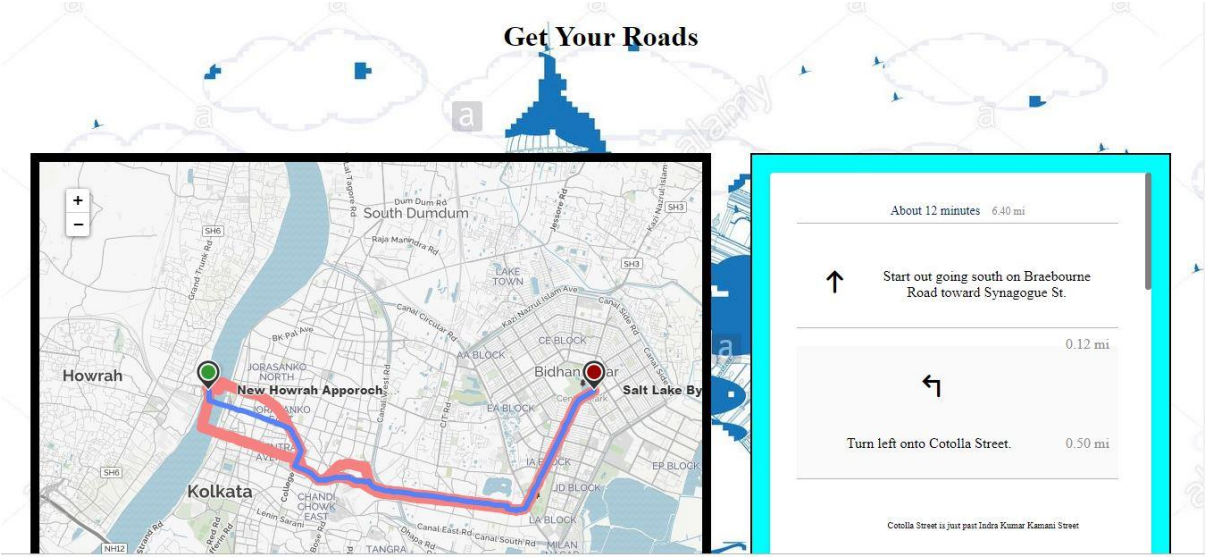


Figure 15: Next page after clicking 'Get your roads button' showing the map and directions

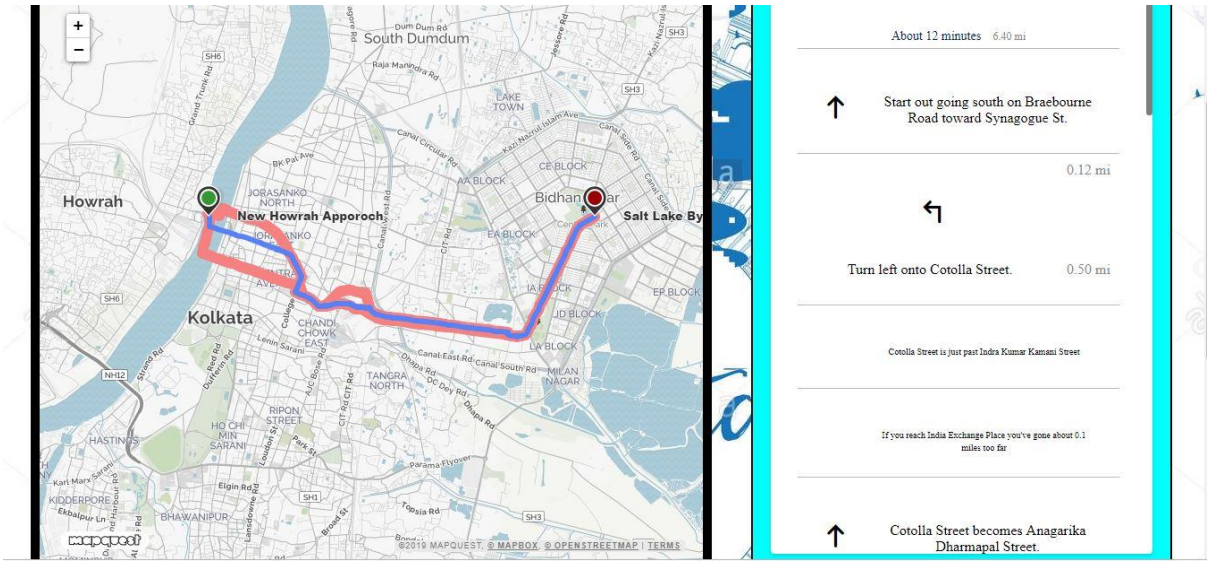


Figure 16: Next page after clicking 'Get your roads button' showing the map and directions as we scroll down

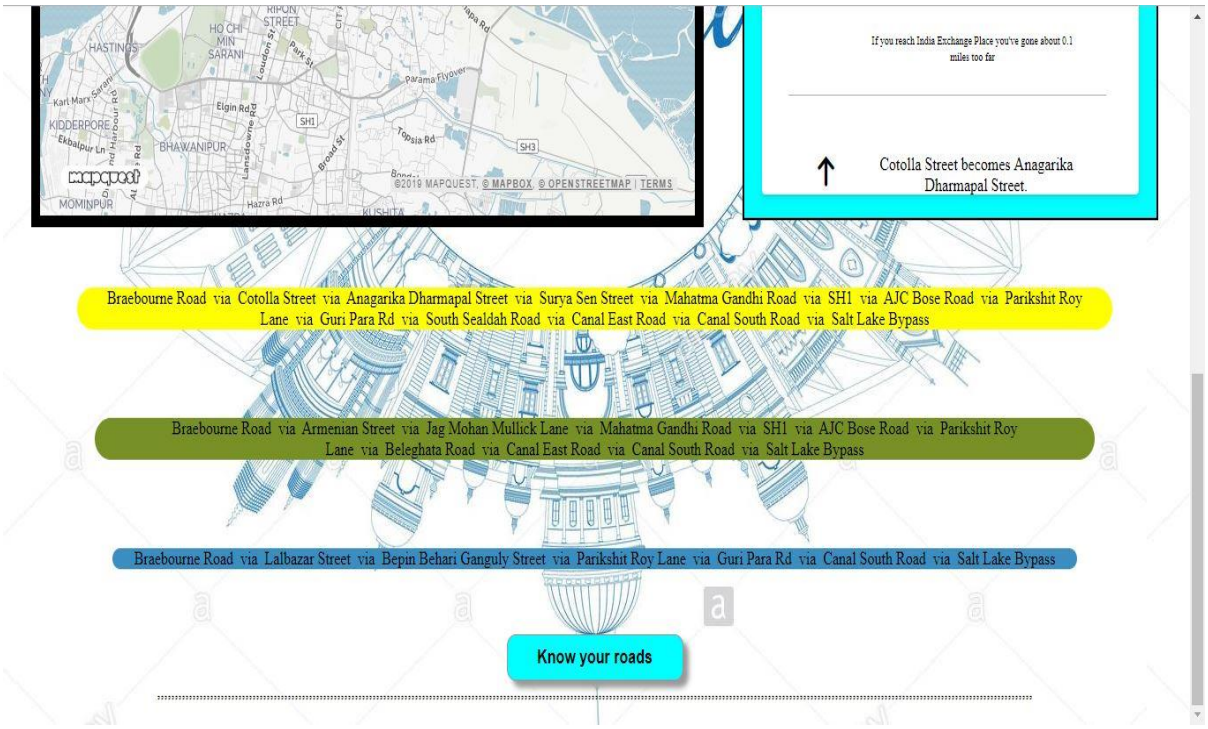


Figure 17: Next page after clicking 'Get your roads button' showing the map and directions and the route descriptions of alternative routes as we scroll down further

Now, after clicking on Know your roads the candidate solutions are displayed showing the route statistics of each of the routes between Howrah and Saltlake by calculating the percentage of what majority of the users say about the five parameters taken i.e. traffic condition, accident vulnerability, wifi facility, water log, women safety .This percentage is calculated based on the feedback stored in the feedback table of the database after validation is performed in the validation module as described earlier in 'Description of CURMA. .

For any crowd sourcing application data storage is considered as an administrative and a server side implementation which provides the backbone for the entire application. According to the flow of our work, now we have stored values (i.e. user ratings) of 5 specific parameters of a route between two places A and B say in tabular form using MYSQL RDBMS in XAMPP. Here each row of the table represents a complete feedback accepted from a user on a particular route in between A and B. Each route in between A and B is uniquely identified by a key. The key value satisfying the criteria of RDBMS is used for further mapping with the output module and displaying all candidate solutions.

CHAPTER 5: IMPLEMENTATION DETAILS

In our application we are receiving the values of the parameters in form of user ratings like “High”, “Moderate”, “Low” or “Yes/ No” or “Not Sure” etc. Now before storing into the database, these ratings will be encoded to some integer values as follows:

Ratings	Corresponding Integer value
High	3
Moderate	2
Low	1
Yes	1
No	2
Not Sure	3

Table 2: Rating Conversion

Encoding the values of parameters in above mentioned manner (Table 2) requires specific domain range information [13]. So it will be advantageous if the domain range of each parameter is made limited and small in size. In our application, we have used the domain range from “Low” to “High” for parameters like “Traffic Condition”, and “Accident Vulnerability”. But parameters like “Water Log” and “Wi-Fi Facility” we have different type of range as “Yes “or “No”. For “Women Safety” we have different type of range as “Yes”, “No” or “Not Sure”. Assignment of domain range can be application specific or parameter specific or some other specifications based. In our application, “Traffic Condition” and “Accident Vulnerability” are ordinal attributes which is a classification of qualitative attribute [14]. In accordance with the nature of both parameters, we have defined their possible values as “Low”, “Moderate” and “High” with corresponding integer encoding as 1, 2, and 3 respectively. In the other hand “Wi-Fi Facility “and “Water Log” are asymmetric binary attributes which is also a classification of qualitative attribute [14]. So we have set limitations on their values to be either “Yes” or “No” with corresponding integer encoding as 1 and 2 respectively. For parameter “Women Safety” we have defined possible values as “Yes”, “No” and “Not Sure” with corresponding integer encoding as 1, 2, and 3 respectively. Let us consider, a user has chosen a route between two places A and B and provided some feedbacks for some parameters like “Traffic Condition” as “High”, “Accident Vulnerability” as “Low”, “Wi-Fi Facility” as “No”, “Water Log” as “Yes” “Women Safety” as “No” etc. This mapping is depicted in Fig 18.

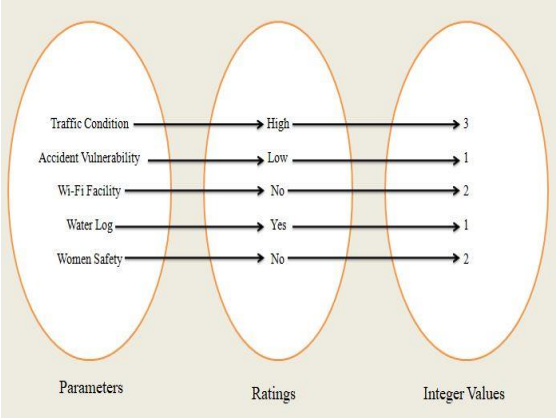


Figure 18: Mapping of Sample User Ratings into Integer Values

By applying the above mentioned methodology, some encoded ratings (i.e. the integer values) have been generated to be stored into the database. The snapshot of the feedback table in the database is shown in Fig 19.

The candidate solutions i.e. the route statistics for the three routes are depicted in Fig 20.Now for all the three routes we see that for accident vulnerability and women safety the percentage is not displayed. The idea behind this is if 70% or more of the people say accident chances are low in the route then only we say that accident chances are low in the road else we say accident chances are high in the road. Again if 70% or more of the users say that the route is safe for women then we consider it safe for women else we consider it unsafe for women at night. Now on these candidate solutions i.e. the route statistics of each route the optimized route is found out using Genetic algorithm and the optimized route is displayed. For this case the optimized route displayed is the third route depending on first priority as traffic condition and second priority women safety as depicted in Fig 21.

	area_code	traffic_condition	accident_vulnerability	wifi_facility	water_log	women_safety	sr_no	validity_status
Edit Copy Delete	JadavpurSaltLake1	3	1	2	2	1	1	
Edit Copy Delete	HowrahJadavpur3	2	2	1	2	2	2	
Edit Copy Delete	UltadangaGaria1	3	2	2	2	1	3	
Edit Copy Delete	New MarketDhakuria2	3	2	2	1	1	4	
Edit Copy Delete	HowrahEsplanade1	3	2	2	2	1	5	
Edit Copy Delete	HowrahShyambazar1	3	1	2	1	1	6	
Edit Copy Delete	HowrahGariahat1	2	1	1	2	1	7	
Edit Copy Delete	HowrahNew Market1	2	1	2	1	1	8	
Edit Copy Delete	HowrahPark Street 1	2	2	1	2	1	9	
Edit Copy Delete	HowrahTollygunge2	1	2	2	1	2	10	
Edit Copy Delete	HowrahAlipore1	1	2	2	2	1	11	
Edit Copy Delete	HowrahHatibagan2	3	1	2	1	1	12	
Edit Copy Delete	HowrahSovabazar1	2	1	2	1	2	13	
Edit Copy Delete	HowrahMaidan2	1	2	1	2	1	14	
Edit Copy Delete	HowrahDhakuria1	3	1	2	1	2	15	
Edit Copy Delete	BurrabazarNew Market1	1	1	2	1	1	16	
Edit Copy Delete	BurrabazarNew Market2	3	1	2	1	1	17	
Edit Copy Delete	GariahatNew Market1	2	2	2	1	1	18	
Edit Copy Delete	Park Street New Market1	2	1	1	2	1	19	
Edit Copy Delete	JadavpurNew Market1	3	2	2	1	2	20	

Figure 19: Snapshot of feedback table showing various feedbacks stored on 5 different parameters for each of the routes

Figure 20: Next page after clicking 'Get your Roads' button showing the route statistics of the alternative routes.

Figure 21: Next page after clicking 'Get your Roads' button and scrolling down, showing the optimized route,

CHAPTER 6

CONCLUSION AND FUTURE WORK

CHAPTER 6: CONCLUSION AND FUTURE WORK

Route Optimization of CURMA has paved a better way for the users of the application to know the routes from the crowd i.e. the common people who actually use it which will help them to choose the best routes depending on the criteria they want to. Even information like whether there are pot holes in the road, sufficient dustbins are present in the route or not, or public toilets are available or not etc many other factors can be obtained.

Through this application in future a report can be generated and submitted to the local government authorities like Kolkata Municipal Corporation or state government authorities like State Road Safety Council for the State of West Bengal. By taking into account the different road conditions between various areas as collected from the common people (crowd) they can improve the road conditions, take measures to minimize accidents in accident prone areas, build sufficient amount of public toilets in the routes where there are none, keep dustbins in various routes to keep the roads clean, improve drainage system in roads where water logging occurs, Thus this application can help the authorities to keep a safer and cleaner city

This application can not only help in optimizing routes through crowdsourced data but in general can be used for optimizing any other product in future which can be evaluated based on parameters through user feedback. For e.g.: cars, any household product, mobile phones etc.

BIBLIOGRAPHY

References:

[1] Crowdsourcing Definitions and Its Features: An Academic Technical Report
URL
http://eprintd.boumemouth.ac.uk/21773/1/List%20of%20Crowdsorcing%20Definitions%20Features.pdf

[2] Types of Crowdsourcing
URL
http://www.thinksimpletoday.com/Simple%20Thoughts/files/2ba9fbe4298cc7d746afaa4bd9685699-10. html

[3] THE EVOLUTION OF CROWDSOURCING: AN OLD IDEA WHOSE TIME HAS COME

URL

https://99designs.com/blog/crowdsourcing/evolution-crowdsourcing/

[4] Google Maps users can now report accidents, speed traps through latest updates. March 16, 2019-04-09
URL
https://www.themobileindian.com/news/google-maps-users-can-now-report-accidents-speed-

[5] Miss. Pariyarath Jesnaraj, Dr. K. V. Metre

An Efficient Algorithm for Database Query Optimization in Crowdsourcing System

Vol-3 Issue-4 2017

URL - https://pdfs.semanticscholar.org/71c9/bce46963f538efb1ede98ca28bb68c5f661f.pdf

[6] Yan Wang, Jianmin Jiang , Tingting Mu

Context-Aware and Energy-Driven Route Optimization for Fully Electric Vehicles via Crowdsourcing
Volume: 14, Issue: 3, Sept. 2013

URL - https://ieeexplore.ieee.org/abstract/document/6516898

[7] JENNY RICHARDS: Two methods for crowd sourcing data. April 11, 2017
URL
https://www.tableau.com/about/blog/2017/4/two-methods-crowdsourcing-data-68134

[8] Sam Meek, Mike J Jackson, Didier G Leibovici: A flexible framework for assessing the quality of crowd sourced data. AGILE 2014 – Castellón, June 3-6, 2014, ISBN: 978-90-816960-4-3

[9] Genetic algorithm
URL
https://en.wikipedia.org/wiki/Genetic_algorithm

[10] Ming-DerYang, Yeh-FenYang, Tung-Ching Suand Kai-Siang Huang: An Efficient Fitness Function in Genetic Algorithm Classifier for Landuse Recognition on Satellite Images, 2014 Feb 18.
URL
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3948504/

[11] How to define a Fitness Function in a Genetic Algorithm? Jul 21, 2017

URL

https://towardsdatascience.com/how-to-define-a-fitness-function-in-a-genetic-algorithm-be572b9ea3b4

[12] Nitasha Soni, Dr .Tapas Kumar: Study of Various Crossover Operators in Genetic Algorithms
URL
http://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit2014050673.pdf