International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

TRAFFIC ACCIDENT ANALYSIS AND OPTIMAL PATH USING RANDOM FOREST

MD FAISHAL KHAN, K. AMUTHA, ABHISHEK KUMAR SINGH, MOHAMMAD SIRAJUDDIN, MERUGU SAI THARUN

ABSTRACT

This report presents the results from the research study on applying large scale data mining methods into analysis of traffic accidents and finding best path on the Finnish roads. The data sets collected from traffic fatal accidents are huge, multidimensional, and heterogeneous. Moreover, they may contain incomplete and erroneous values, which make its exploration and understanding a very demanding task. The target data of this study was collected by the Finnish Road Administration Datasets. The intention is to investigate the usability of robust clustering, association and frequent itemsets, and visualization methods to the road traffic accident analysis. While the results show that the selected data mining methods are able to produce understandable patterns from the data, finding more fertilized information could be enhanced with more detailed and comprehensive data sets. Machine Learning algorithm takes accident frequency count as a parameter to cluster the locations. Then we used association rule mining to characterize these Surface Condition. The rules revealed different factors associated with road accidents at different drunk and drive with varying accident frequencies. The association rules for high-frequency accident location disclosed that intersections on highways are more dangerous for every type of fatal accidents. The goal of this project is to create a Pathfinding Visualizer amd road accident analysis, which can be used to create a best possible path to connect all the landmark locations so that the travellers can feel comfortable to travel.

OBJECTIVE

- This report presents the results from the research study on applying large scale data mining methods into analysis of traffic accidents on the Finnish roads. The data sets collected from traffic fatal accidents are huge, multidimensional, and heterogeneous. Moreover, they may contain incomplete and erroneous values, which make its exploration and understanding a very demanding task.
- The goal of this project is to create a Pathfinding Visualizer, which can be used to create a best possible path to connect all the landmark locations so that the travellers can feel comfortable to travel.

BASE PAPER

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9675287

INTRODUCTION

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous decades. Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent itemset mining. A classical association rule mining method is the Apriori algorithm who main task is to find frequent itemsets, which is the method we use to analyze the roadway traffic data. Classification in data mining methodology aims at constructing a model (classifier) from a training data set that can be used to classify records of unknown class labels. The Naive Bayes technique is one of the very basic probability-based methods for classification that is based on the Bayes' hypothesis with the presumption of independence between each pair of variables. We used the FARS dataset for our study. The Fatal Accidents Dataset contains all fatal accidents on public roads in 2017 reported to the National Highway Transportation Safety Administration. The dataset is downloaded from California

Polytechnic State University and all data originally came from FARS. The dataset contains 37,248 records and 55 attributes. The data description can be found in the document FARS.

SCOPE OF THE PROJECT

Road traffic injury is a major global public health problem. Rapid motorization in low and middle-income countries along with the poor safety quality of road traffic systems and the lack of institutional capacity to manage outcomes contribute to a growing crisis.

More than 1.24 million people die each year on the world's roads. Many more suffer permanent disability, and between 20 and 50 million suffer non-fatal injuries. These are mainly in amongst vulnerable road users and involve the most socio-economically active citizens.

CHAPTER 2

LITERATURE SURVEY

1 INTRODUCTION

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating

system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

Title 1: Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women

Author: **divay bansal and lekha bhambhu**

Apriori Algorithm is the most popular and useful algorithm of Association Rule Mining of Data Mining. As Association rule of data mining is used in all real life applications of business and industry. Objective of taking Apriori is to find frequent itemsets and to uncover the hidden information. This paper elaborates upon the use of association rule mining in extracting patterns that occur frequently within a dataset and showcases the implementation of the Apriori algorithm in mining association rules from a dataset containing crimes data concerning women. As for this WEKA tool is used for extracting results .For this one dataset is taken from UCI repository and other data is collected manually from the session court of sirsa to collect data on heart melting crimes against women. The main motive to use UCI is to first check the proper working of dataset and then apply Apriori on real dataset against crimes on women which extracts hidden information that what age group is responsible for this and to find where the real culprit is hiding. In last the comparison is done between Apriori & PredictiveApriori Algorithm in which Apriori is better and faster than PredictiveApriori Algorithm

Title 2: Applying association rules mining algorithms for traffic accidents in dubai.

Author: mira A El Tayeb, Vikas Pareek, and Abdelaziz Araar

Association rule mining algorithms are widely used to find all rules in the database satisfying some minimum support and minimum confidence constraints. In order to decrease the number of generated rules, the adaptation of the association rule mining algorithm to mine only a particular subset of association rules where the classification class attribute is assigned to the right-hand-side was investigated in past research. In this research, a dataset about traffic accidents was collected from Dubai Traffic Department, UAE. After data preprocessing, Apriori and Predictive Apriori association rules algorithms were applied to the dataset in order to explore the link between recorded accidents' factors to accident severity in Dubai. Two sets of class association rules were generated using the two algorithms and summarized to get the most interesting rules

using technical measures. Empirical results showed that the class association rules generated by Apriori algorithm were more effective than those generated by Predictive Apriori algorithm. More associations between accident factors and accident severity level were explored when applying Apriori algorithm.

Title 3: A perspective analysis of traffic accident using data mining techniques.
Author: S. Krishnaveni and M. Hemalatha

Data Mining is taking out of hidden patterns from huge database. It is commonly used in a marketing, surveillance, fraud detection and scientific discovery. In data mining, machine learning is mainly focused as research which is automatically learnt to recognize complex patterns and make intelligent decisions based on data. Nowadays traffic accidents are the major causes of death and injuries in this world. Roadway patterns are useful in the development of traffic safety control policy. This paper deals with the some of classification models to predict the severity of injury that occurred during traffic accidents. I have compared Naive Bayes Bayesian classifier, AdaBoostM1 Meta classifier, PART Rule classifier, J48 Decision Tree classifier and Random Forest Tree classifier for classifying the type of injury severity of various traffic accidents. The final result shows that the Random Forest outperforms than other four algorithms.

Title 4: Analysing road accident data using association rule mining
Author: Sachin Kumar and Durga Toshniwal

Road accident is one of the crucial areas of research in India. A variety of research has been done on data collected through police records covering a limited portion of highways. The analysis of such data can only reveal information regarding that portion only; but accidents are scattered not only on highways but also on local roads. A different source of road accident data in India is Emergency Management research Institute (EMRI) which serves and keeps track of every accident record on every type of road and cover information of entire State's road accidents. In this paper, we have used data mining techniques to analyze the data provided by EMRI in which we first cluster the accident data and further association rule mining technique is applied to identify circumstances in which an accident may occur for each cluster. The results can be utilized to put some accident prevention efforts in the areas identified for different categories of accidents to overcome the number of accidents.

Title 5: Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques
Author: KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar. Avra

Road accident its may no best oppida together, but can be reduced. Driver emotions such as sad, happy, and anger can be one reason for accidents. At the same time, environment conditions such as sweather, traffic on the road, load in the vehicle, type of road, health condition of driver, and speed can also be the reasons for accidents. Hidden patterns in accidents can be extracted so as to find the common featuresbetweenaccidents.Thispaperpresentstheresultsoftheframeworkfromthe research study on road accident data of major national highways that pass through Krishna district for the year 2013 by applying machine learning techniques into analysis. These datasets collected from police stations are heterogeneous. Incomplete and erroneous values are corrected using data cleaning measures, and relevanceattributesareidentifiedusingattributeselectionmeasures.Clustersthatare formed using K-melodies, and expectation maximization algorithms are then analysed to discover hidden patterns using a priori algorithm. Results showed that the selected machine learning techniques are able to extract hidden patterns from the data. Density histograms are used for accident data visualization.

Chapter 3

Introduction

Design is a multi- step that focuses on data structure software architecture, procedural details, algorithm etc… and interface between modules. The design

Process also translate the requirements into presentation of software that can be

Accessed for quality before coding begins. Computer software design change

Continuously as new methods; better analysis and border understanding evolved. Software design is at relatively early stage in its revolution. Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are normally associated with more classical engineering disciplines. However techniques for software designs do exit, criteria for design qualities are available and design notation can be applied.

EXISTING SYSTEM

The traffic accident using data mining technique that could possibly reduce the fatality rate. Using a road safety database enables to reduce the fatality by implementing road safety programs at local and national levels. Classification models to predict the severity of injury that occurred during traffic accidents. Association rules mining algorithm on a dataset about traffic accidents which was gathered from Government Traffic Office, Apriori and Predictive Apriori association rules algorithms were applied to the dataset to investigate the connection between recorded accidents and factors to accident severity.

Disadvantages

- Here we are using data mining the fault traffic accident injuries and deaths
- In this the cost of maintain and reparing the roads
- This will not useful for short distance level

PROPOSED SYSTEM

- This paper presents our research to model the severity of injury resulting from traffic accidents using artificial neural networks and decision trees. We have applied them to an actual data set obtained from the National Automotive Sampling System (NASS) General Estimates System (GES). Experiment results reveal that in all the cases the decision tree outperforms the neural network.
- Our research analysis also shows that the three most important factors in fatal injury are: driver's seat belt usage, light condition of the roadway, and driver's alcohol usage. Our experiments also showed that the model for fatal and non-fatal injury performed better than other classes.
- The ability of predicting fatal and non-fatal injury is very important since drivers' fatality has the highest cost to society economically and socially.
- The goal of this project is to create a Pathfinding Visualizer amd road accident analysis, which can be used to create a best possible path to connect all the landmark locations so that the travellers can feel comfortable to travel.

ADVANTAGES OF PROPOSED SYSTEM

- speed was available
- improve the performance of analysis in fatal and non-fatal accidents

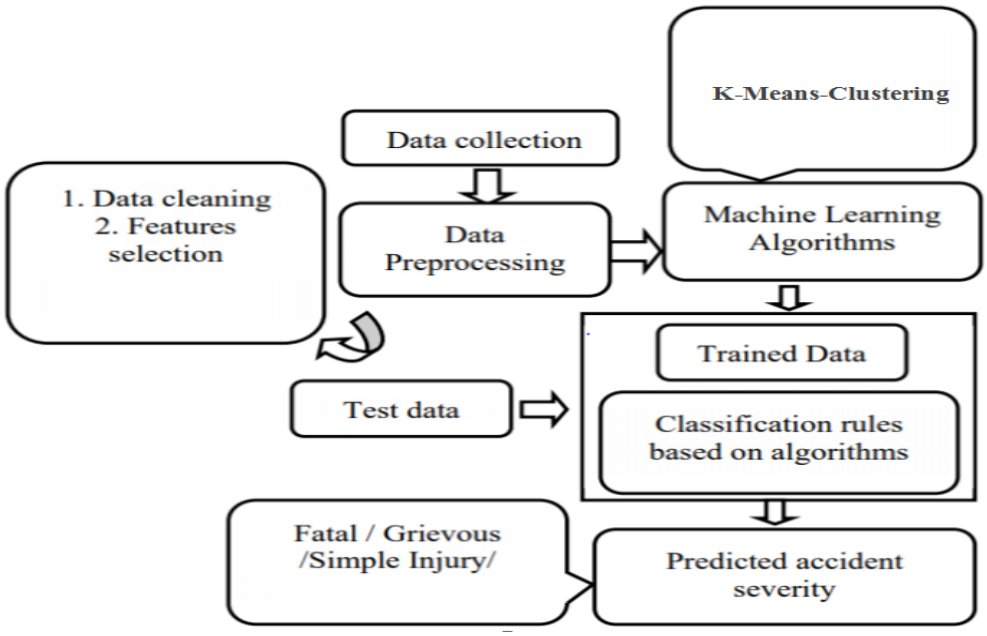
ALGORITHM

Random forest algorithm:

Random forest is a **Supervised Machine Learning Algorithm** that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

As the name suggests, "Random Forest is **a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset**."

SYSTEM ARCHITECTURE



SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- System - Pentium-IV
- Speed - 2.4GHZ
- Hard disk - 40GB
- Monitor - 15VGA color
- RAM - 512MB

SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding language - Python

SYSTEM DESIGN AND TESTING PLAN

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?

- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

DATA FLOW DIAGRAM

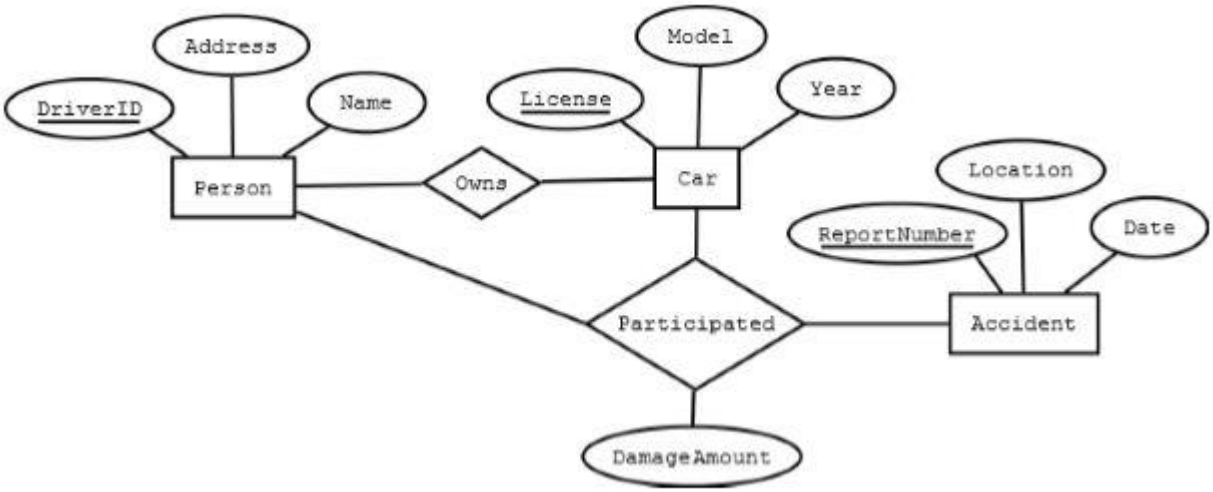
Data Flow Diagram (DFD) is a two-dimensional diagram that describes how data is processed and transmitted in a system. The graphical depiction recognizes each source of data and how it interacts with other data sources to reach a mutual output. In order to draft a data flow diagram one must

- Identify external inputs and outputs
- Determine how the inputs and outputs relate to each other
- Explain with graphics how these connections relate and what they result in.

Role of DFD:

- It is a documentation support which is understood by both programmers and nonprogrammers. As DFD postulates only what processes are accomplished not how they are performed.
- A physical DFD postulates where the data flows and who processes the data.
- It permits analyst to isolate areas of interest in the organization and study them by examining the data that enter the process and viewing how they are altered when they leave.

ER diagrams



UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

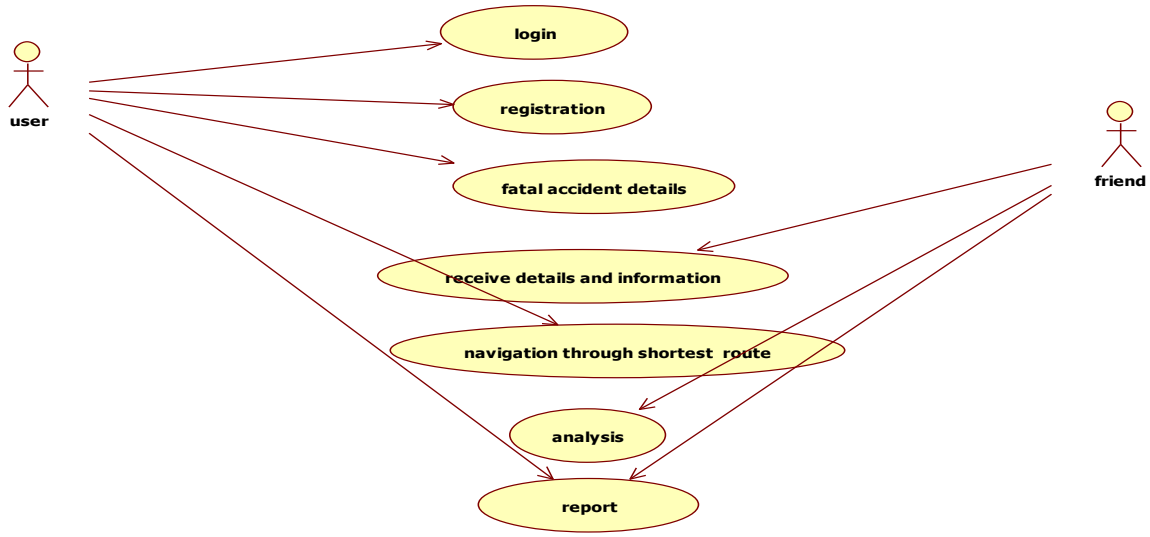
GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

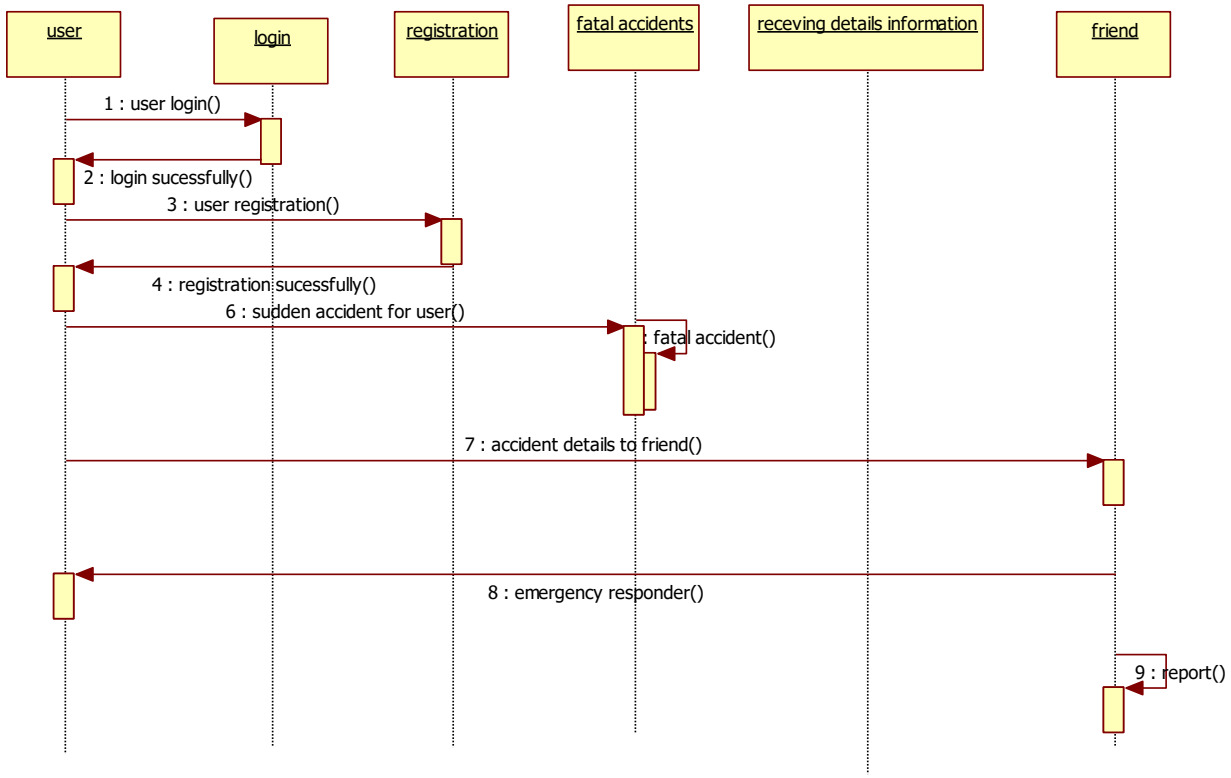
USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

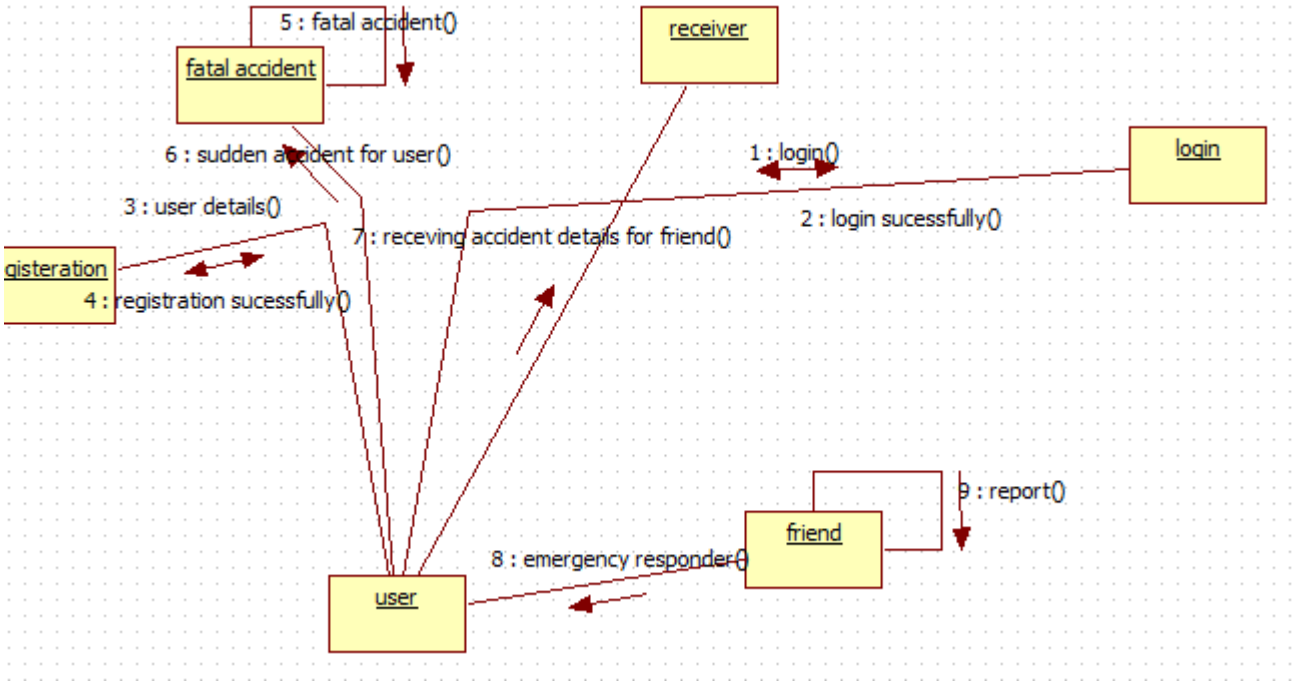


SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



COLLABRATION DIAGRAM



REFERENCE

[1] Divya Bansal and Lekha Bhambhu. Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women. International Journal of Advanced Research in Computer Science and Software Engineering, 3(9), September 2013.

Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar. Applying association rules mining algorithms for traffic accidents in dubai. International Journal of Soft Computing and Engineering, September 2015.

[3] William M Evanco. The potential impact of rural mayday systems on vehicular crash fatalities. Accident Analysis & Prevention, 31(5):455–462, September 1999.

[4] K Jayasudha and C Chandrasekar. An overview of data mining in road traffic and accident analysis. Journal of Computer Applications, 2(4):32–37, 2009.

[5] S. Krishnaveni and M. Hemalatha. A perspective analysis of traffic accident using data mining techniques. International Journal of Computer Applications, 23(7):40–48, June 2011.

[6] Sachin Kumar and Durga Toshniwal. Analysing road accidentdata using association rule mining. In Proceedings of International Conference on Computing, Communication and Security, pages 1–6, 2015.

[7] Eric M Ossiander and Peter Cummings. Freeway speed limits and traffic fatalities in washington state. Accident Analysis &Prevention, 34(1):13–18, 2002.

[8] KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar. Avra Bangladesh collection, analysis & visualization of road accident data in Bangladesh. In Proceedings of International Conference on Informatics, Electronics & Vision, pages 1–6. IEEE, 2013.

[9] Trac Integrated SCM & Project Management. Fatal Accidents Dataset. https://wiki.csc.calpoly.edu/datasets/wiki/ HighwayAccidents

[10] U.S. Census Bureau. Population Estimates. http://www.census. gov/popest/data/historical/2000s/vintage 2007/, 2007.