International Journal of Engineering & Technology, 7 (4.6) (2018) 63-67

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

R-Language Based Analytics System For Monitoring Railway Disasters

G S Prasada Reddy[1*], P Rama Rao[2], Vemula Harish[3] Nirmala Teegala[4]

[1]*Vardhaman College of Engineering, Hyderabad,India*
[2]*G Pullaiah College of Engineering and Technology*
[3]*Vardhaman College of Engineering, Hyderabad*
[4]*Department of Computer Science and Engineering MLR Institute of Technology, Hyderabad*
**Corresponding author E-mail: gsprasad.reddy56@gmail.com*

Abstract

The everyday workplace of railroads is backbreaking and frequently includes high-hazard activities. These operations has its impact on railways crew and public people who lives along the town and cities. To avoid the employees and public's exposure to risk, railway companies uses safety training programs[1] , advanced technology concepts like data analytics and machine learning to analyze about frequent accident occurring zones and safety precautions to be taken in advance. So in this paper we are proposing an visualization model for identifying the frequent occurrence of railway accident zones so as to take enough precautions to avoid them. Here we will be collecting the data on the frequently occurring railway accidents and the causes for it. After the data is been collected, we will be analyzing the data and presenting the information about most frequent and less frequent accident prone areas using visualization tools which provides the clear picture of incidents so as certain precautions or counter measures will be taken at proper times to minimize railway accidents. The results of the analysis done can be used to make and communicate decisions about accidents so as to minimize them.

Keywords*: railways, data analytics, visualization ,machine learning, R-language*

1. Introduction

The Indian Railways(IR)[2] with network size 121407 km and total track 67368km is fourth largest network. Indian railway is running more than 12500 passenger trains covering long distances over country. The Indian railways also in the job of carrying goods across the country.

The IR serving huge number of passengers as well as carrying goods. The railway network in India[15] is one of the biggest and busiest railway arranges on the planet, dealing with gigantic quantities of travelers and amounts of products every day. Railroads are the most well known methods for long-separate transportation in India; consequently the railway is frequently portrayed as the foundation of this present country's economy.

For example, factors, for example, the movement between real urban communities surpassing the arranged limit and over-used railroad tracks are bringing about trains traveling at lessened speeds and convey lesser measures of cargo, in this way expanding the cost and time of transportation. Likewise, Indian Railways have been developed way back by British, and have created during the time to its present state. Since the railroad mastermind outlines the establishment of system in India, the progression of the rail route can fill in as a marker of the money related improvement of the country also. Accordingly, the improvement of railroad has also been inspected in this proposition.

1.1 Recent railway accident history:

Despite of the many measures taken by Indian railways, more number of accidents taken place in the from 2017 onwards to till date. Here we collected railway accidents data from 2017 January onwards. There occurred a spate of railway accidents in the year 2017 resulting in loss of large number of human lives and damage to the nation property, disturbances to normal day to day life. There are nearly 9 major railway mishaps[2] and overall 35 above railway accidents in the year 2017 to till date resulting in loss of hundreds of human lives and injuring people. The below figure gives a glance on railway accidents taken place in 2017 year.

S.NO	Accident type	Accident date	Accident area	Killed people and Injured
1	Derailed	2017	Andhra Pradesh	41 and 68
2	Bomb explosion	2017	Madhya Pradesh	0 and 10
3	derailed	2017	Uttar Pradesh	0 and 52
4	derailed	2017	Rampur	0 and 24
5	derailed	2017	Uttar Pradesh	23 and 97
6	derailed	2017	Uttar Pradesh	0 and 100
7	derailed	2017	Uttar Pradesh	3 and 9
8	Hit	2017	Uttar Pradesh	5 and 2

Figure 1 : List of railway accidents in year 2017

The railways are facing some of the challenging problems. one of them is railway accidents which damages railway and public properties, huge loss of human life and employees of railways etc. There were many reasons for the railway disasters occurring in the daily operations of railways. Some of the most common railway accident types are :

- derailed
- bridge collapse

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- rail collision
- bomb explosion
- Fire
- bump
- signal
- Hitting vehicles

These kind of railway accidents causing heavy losses in terms of property and life. Hence to reduce these kind of losses at least to small extent, a railway accident analytics system is required to monitor the railway disaster incidents.

Present scenario:

- Data collection
- Precautions are implemented
- But no exact analysis

2. Proposed System:

Here we propose a analytics system which involves

- Collection of Data
- Analysis of Data using analytics package
- Frame Exact precautions depending on location and accident type
- Analysis on more and Less frequent accident prone areas

We can build the analytics system using analytics tool like R with the support of database.

3. Proposed Analytics System:

Work Flow of proposed analytics system

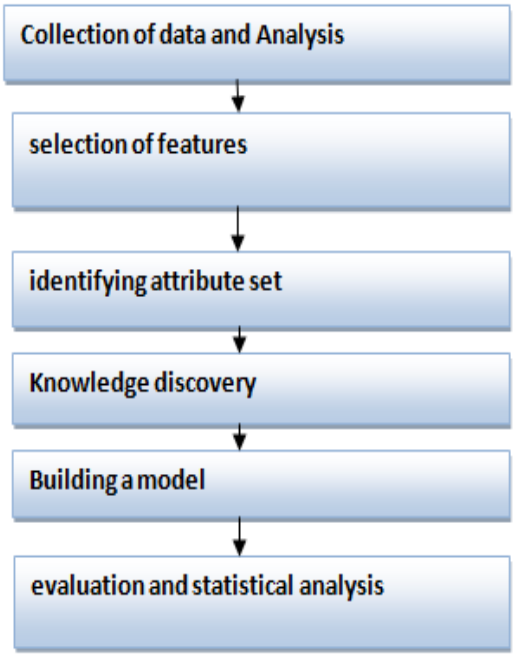


Figure 2: Analytics System Work Flow

The proposed analytics system involves the following steps:

Collection of Data: The data about different railway accidents occurred in the span of last 1-2 years is collected from various sources like railways reports etc. The data collected will have attributes like accident type, year of accident, location, number of people injured or died etc.

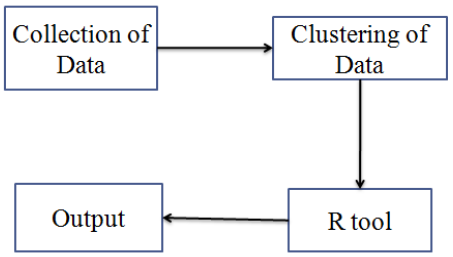


Figure 3: proposed system architecture

Clustering of Data: Then the clustering of data is done by dividing the data into similar groups based on similarity in any attribute or feature. This clustering of data is very helpful in data analysis. Here in this scenario, we can cluster all the collected data into different clusters based on location of accident or type of accident etc. With this clustering, we can identify most accident prone areas and least accident prone areas.

Using R-tool for building analytics system: We use R-tool for clustering data. We can use clustering algorithms like K-means algorithm for categorising data into cluster groups and to draw conclusions.

4. Modules:

For data collection, we use a web application based interface to import all necessary data into database.

The different modules of web interface includes:

- admin module
- server
- database
- client interaction interface

The essential thought fundamental secluded plan is to sort out a perplexing framework, (for example, a substantial program, an electronic circuit, or a mechanical gadget) as an arrangement of particular segments that can be created autonomously and afterward stopped together. In spite of the fact that this may show up a straightforward thought, encounter demonstrates that the adequacy of the method depends basically on the way in which frameworks are isolated into parts and the systems used to plug segments together.

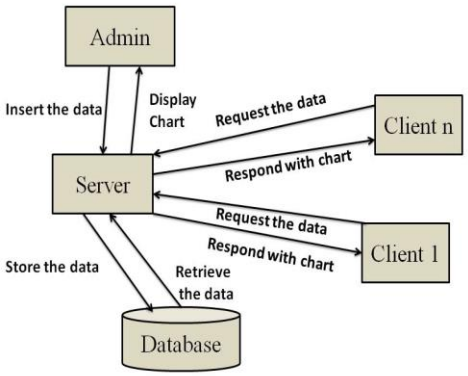


Figure 4: Modular Design approach

The details about different modules involved are specified below in detail :

- **Client/ User Module:** The user module allows users request for the data from the server. They get the response from the server in the form of Motion charts. They can even change the view of the chart from Motion

Chart to Histograms or else dot curves. The following are the user roles in our project,

- Request for the Data
- View the Charts

- **Admin Module:** The Admin modules are designed to help in managing the server. The admin module in the proposed system is used to do the following tasks,
  - Maintain Server to fulfill all the requirements
  - Maintain the updating of the data
- **Server Module:** The job of server is to take requests from client and respond accordingly. Security is the key concern for running servers on a dedicated computer. The server module in our project maintains the following data, such as
  - Database of the datasets inserted by the admin
  - Respond to the client on request

5. Design diagrams:

5.1 Visual Representation of functionality:

The business functionality can be represented by using use case diagrams. The identification of use case starts with listing the discrete business functions in software requirements document. The business functions identified can be declared as a potential use case.

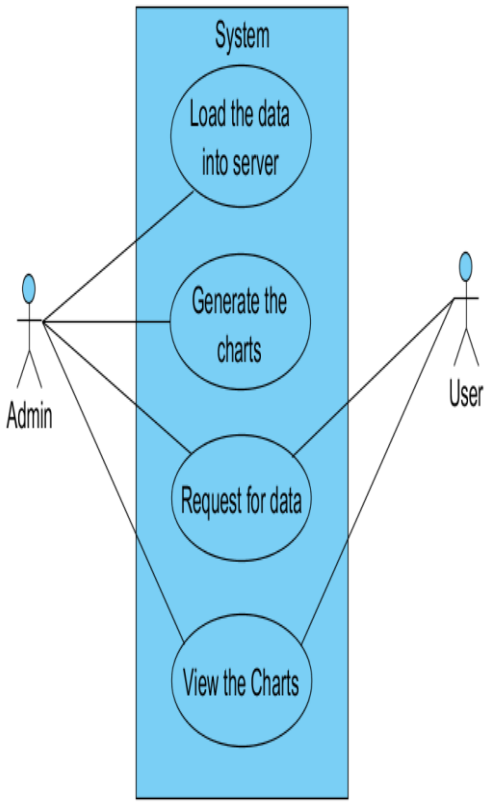


Figure 5: Use case diagram

5.2 Class Structure:

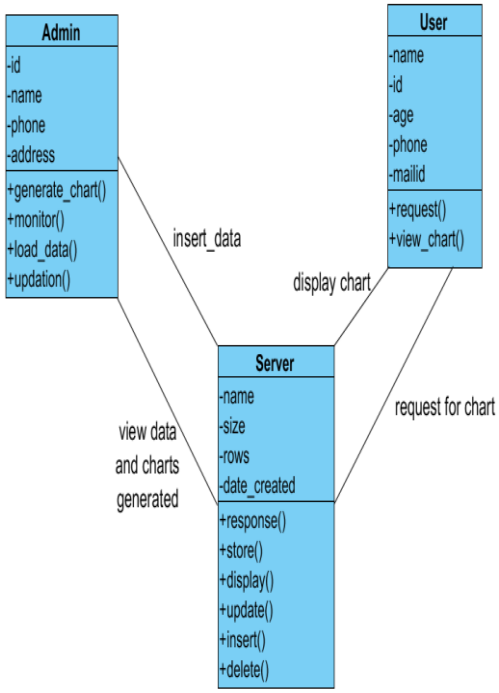


Figure 6: Class diagram

5.3 interactions between classes:

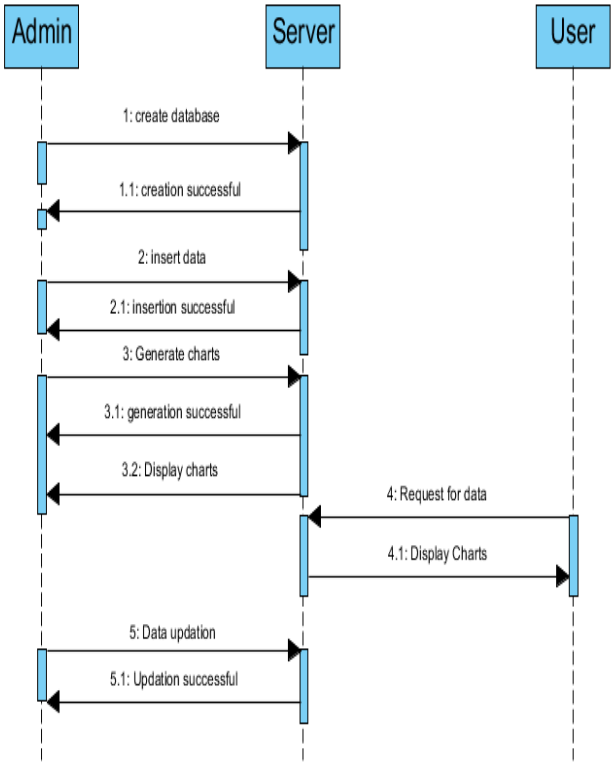


Figure 7: Sequence diagram

5.4 Workflow of System:

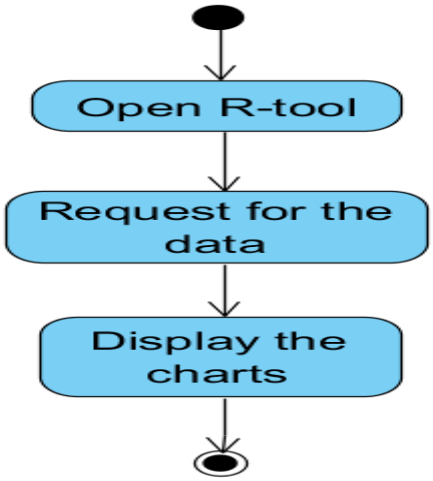


Figure 8: Work flow diagram

6. Introduction to R:

Statistical computing and visual graphics can be best implemented or developed by using R language which is open source. The different numeric statistical and visual graphics techniques supported by R-language are modelling(linear and non-linear), classification and clustering etc.By using R-language, clear and quality visual plots can be produced for any given data.

6.1 The R environment

R is an integrated package of software facilities for the data manipulation, calculation and graphical display.

It includes

- efficient handling of data and storage function,
- a set of operators for matrix calculations
- a huge set of intermediate tools for analysis of data
- visual graph representation facilitating data analysis and a robust programming language which includes Decision making constructs , loops, user-defined functions, recursion facility and I/O facilities.
- can be dynamically linked.

7. Sample data set of railway accidents in India:

Table 1: Sample data set of railway accidents

S.NO	Accident type	Accident date	Accident area	Killed people
1	Derailed	1950	Bihar	32
2	bridge	1950	Bihar	81
3	collision	1960	Bihar	11
4	Collision	1962	Bihar	25
5	Bomb	1966	Bihar	55
6	Bomb	1966	Bihar	40
7	Collision	1979	Bihar	50
8	river	1981	Bihar	800
9	Collision	1986	Bihar	50
10	Fire	2013	Bihar	35
11	Collision	1990	Bihar	60
12	Bump	2014	Bihar	35
13	Signal	2013	Bihar	20
14	fire	1990	Bihar	70
15	Derailed	1951	Delhi	10
16	Collision	1954	Delhi	10
17	Collision	1997	Delhi	12
18	Collision	1982	Delhi	50
20	Collision	1953	Andhra	5
21	Derailed	1954	Andhra	15
22	Derailed	2017	Andhra Pradesh	41
23	Bomb explo-sion	2017	Madhya Pradesh	10(injured)
24	derailed	2017	Uttar Pra-desh	52(injured)
25	derailed	2017	Rampur	24(injured)
26	derailed	2017	Uttar Pra-desh	23
27	derailed	2017	Uttar Pra-desh	100(injured)
28	derailed	2017	Uttar Pra-desh	3 and 9
29	Hit	2017	Uttar Pra-desh	5 and 2

The above data set is visualized as shown below by using tools like weka or R-tool.

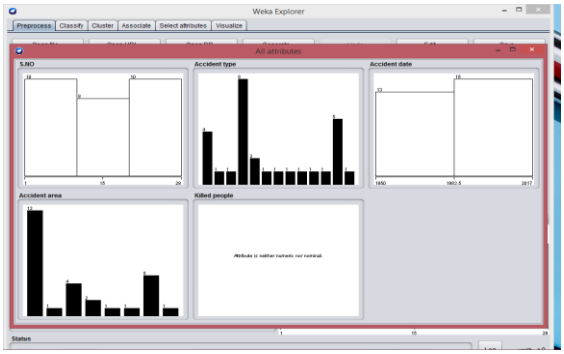


Figure 9: Analysis using WEKA

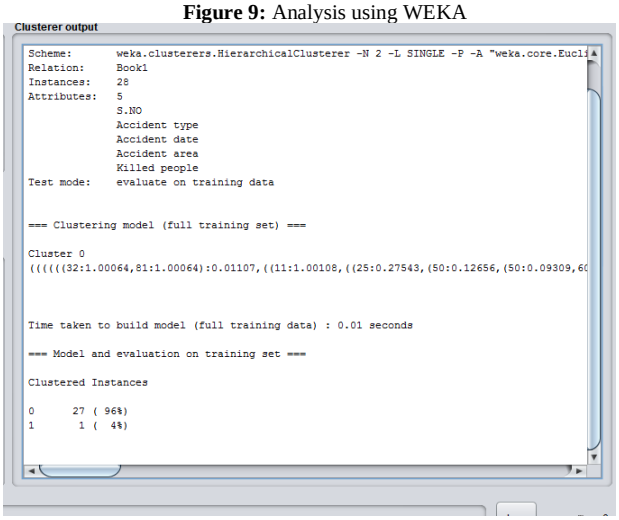


Figure 10: clustering Analysis using WEKA

Using Weka tool, The above given data set is analysed based on attribute accident type :

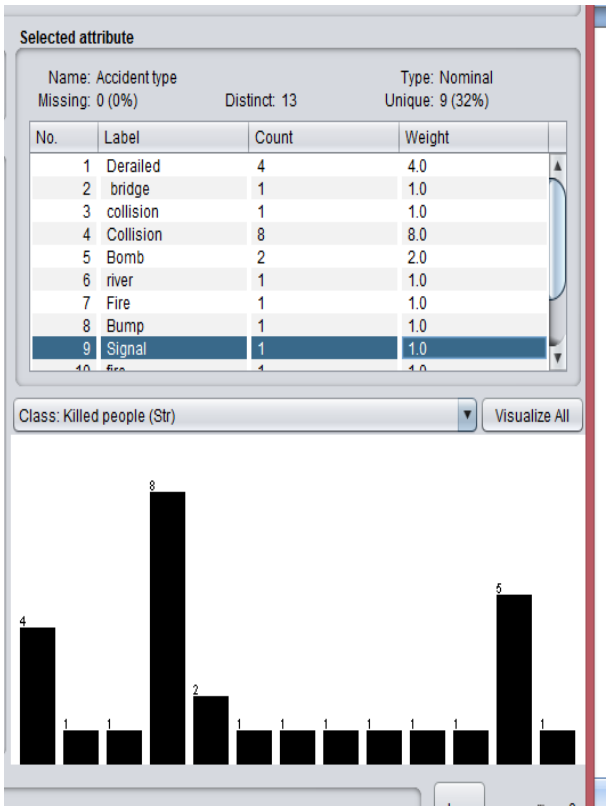


Figure 11: Attribute Analysis using WEKA

8. Motion Charts in R-Language:

Motion Chart which is possible in R-platform by using Google Vis brings your data to life. The basic requirements for generating motion charts are minimum four variables. One of the four variables represents the variable we want to plot, one related to time variable and remaining at least two numerical variables.

8.1 Commands in R to build analytics by using dataset given in Table 1:

- install.packages()
- library('googleVis')
- data1 <-read.csv(file.choose(),header=T)
- visualization1<- gvisMotion-Chart(data1,idvar='Train.per.million',timevar='Year')
- plot(visualization1)

8. 2 Motion Charts Generated using R-packages:

Motion charts are useful in visualization of data where data variables can be mapped on to two dimensional coordinate axis which helps us in understanding of large amounts of data

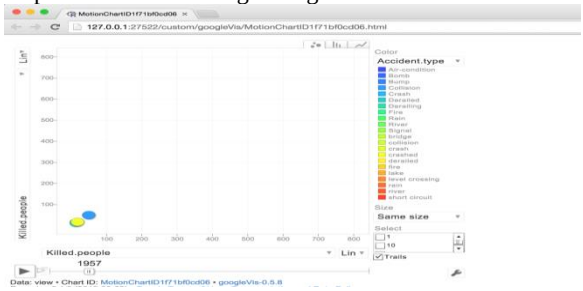


Figure 12: Motion Chart with parameter Accident type

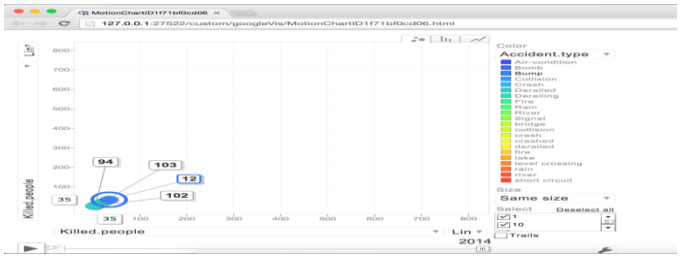


Figure 13: Motion Chart with parameter Accident type showing values inside chart with trails showing highlighting

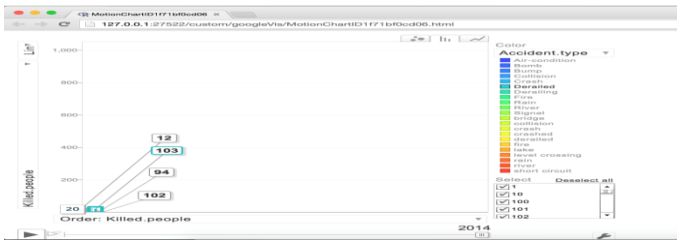


Figure 14: Histogram with parameter Accident type showing values inside chart with highlighting of values

9. Conclusion

With data set taken, we can draw some conclusions like which type of railway accident happened more times and at which locations, reason for accident etc. Hence we are proposing an analytical system based on R-language which can be used for monitoring railway disasters. The railway disasters can be analyzed region wise across country and can identify areas with less, more railway accidents prone areas based on motion charts generated by considering past history of accidents. This conclusions drawn from analytics system can be given to railway people so as they can take measures to avoid railway disasters saving human life and propety losses.

10. Future Scope

Extensive Investigation to be done to find other incident types which are not included in this project. The missing data in this project can also be introduced into the database in future so that the results will be more accurate and the precautions can be taken more precisely. Till now, with this analytics system, we are able to collect data and identify different attributes set so as to draw necessary conclusions.

References:

[1] Lira W.P. Alves, R. ; Costa, J.M.R. ; Pessin, G. ; Galvao, L. ; Cardoso, A.C. ; De Souza, C.R.B.,"A Visual-Analytics System for Railway Safety Management",Computer Graphics and Applications, IEEE (Volume:34, Issue: 5),26th July 2014,pp. 52-57.

[2] https://en.wikipedia.org/wiki/List_of_Indian_rail_accidents#2017

[3] https://ppiaf.org/sites/ppiaf.org/files/documents/toolkits/railways_toolkit/PDFs/RR 20Toolkit 20EN 20New 202017 2012 2027 20CASE6 20INDIA.pdf

[4] http://www.mospi.gov.in/statistical-year-book-india/2016/188

[5] https://analyticsindiamag.com/indian-railways-embrace-data-analytics-monetize-available-data/

[6] Sergio Saponara, Luca Fanucci, Fabio Bernardo, Alessandro Falciani, "Predictive Diagnosis of High-Power Transformer Faults by Networking Vibration Measuring Nodes With Integrated Signal Processing", *Instrumentation and Measurement IEEE Transactions on*, vol. 65, pp. 1749-1760, 2016, ISSN 0018-9456.