

# A Review on Analysis of Railway Traffic Accident with Data Mining Techniques

Manju Bala<sup>1\*</sup>, Anshu Bhasin<sup>2</sup>

<sup>1</sup>Department of CSE, I.K.G Punjab Technical University Kapurthala, Punjab, India

<sup>2</sup>Department of CSE, I.K.G Punjab Technical University Kapurthala, Punjab, India

\*Corresponding Author: manjuagarwal741@gmail.com

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 19/Jun/2018, Published: 30/Jun/2018

**Abstract** - Accident examination assumes an imperative part in transportation framework. Examination of mishap is critical on the grounds that it can uncover the connection between the distinctive kinds of ascribes that adds to a accident. Qualities that influence the accident can be characteristic, condition properties, movement traits and so on. Breaking down accident can give the data about the commitment of these characteristics which can be used to defeat the accident rate. These days, Data mining is a famous system for inspecting the railway accident dataset. This paper presents various research work done in past in the field of rail accident analysis using data mining as a review and also discussed about the cause of accidents and role of data mining in the analysis of accidents.

**Keywords** – Accident, Analysis, Data Mining, Rail Accident, Traffic Management

## I. INTRODUCTION

Railways are a noteworthy system where a great many individuals travel all through and go about as fundamental wellspring of transportation. With headway of present day advances like GPS being consolidated into the transportation frameworks clears route for better computerization and area-based administration changes. Prepare incidents happen because of impacts, crashes, fire and various situations in which crisis therapeutic help is a key concern. Maintaining center around the crisis medicinal help for any incidents that may happen. Prepare mischances make harm foundation and moving stock and also benefit disturbances, and may cause setbacks and damage nature. In like manner, enhancing train working security has for some time been a high need of the rail business and the legislature [1].

Train accidents happen because of a wide range of causes; be that as it may, some are significantly more common than others. Besides, the recurrence and seriousness of accidents likewise fluctuates broadly, contingent upon the specific mishap cause. Productive designation of assets to avoid mischances in the most savvy way conceivable requires understanding which factors represent the most serious hazard, and under which conditions. Appraisal of the advantages and expenses of procedures to relieve every mishap cause would then be able to be assessed and assets distributed with the goal that the best wellbeing change can be accomplished for the level of investment accessible [2].

Activity administration is a conceivable method to diminish rail accidents. Movement Management is turning into an all the more usually utilized term with regards to rail tasks, to depict the administration of rail activities as a proportionate to air or street Traffic Management. Rail Traffic Management Systems will be frameworks intended to help dealing with the

innate intricacy of rail administrations and rail arranges by giving a coordinated and all-encompassing perspective of operational execution, empowering rail operational staff to better adjust the occasionally contending requests of viewpoints, for example, track access and prepare team assets when taking choices. Rail Traffic Management Systems bolster rail activities procedures and strategies and are frameworks in view of the idea of a solitary, coordinated and steady arrangement of operational information, empowering abnormal amounts of rail tasks productivity [3].

Decreasing the seriousness of wounds from rail crashes is a standout amongst the best intends to enhance thruway security. To decrease the quantity of individuals slaughtered and additionally harmed in auto collisions, numerous exploration thinks about have been directed to recognize the hazard factors that can altogether impact the damage results of train crashes[4].

Rest of the paper is organized as follows. Section 2 describes cause of accident, section 3 describes data mining and accident, section 4 presents related work, finally section 5 conclude the paper.

## II. CAUSE OF ACCIDENTS

Accidents happened on account of the heedlessness of driving. FRA prepare accident cause codes are progressively sorted out and arranged into significant reason gatherings—track, hardware, human elements, flag, and random. Inside every one of these significant reason gatherings, FRA arranges singular reason codes into subgroups of related causes, for example, roadbed and track geometry, inside the track gathering and comparable subgroups inside the other real reason gatherings. A minor departure from the FRA subgroups created by Arthur D. Little, in which comparable reason codes were joined into bunches based on master feeling, was utilized (9, 10). The Arthur D. Little groupings are like FRA's subgroups however are all the more fine-grained, in this manner permitting more noteworthy determination for specific causes. For instance, FRA consolidates broken rails, joint bars, and rail grapples in a similar subgroup, while the Arthur D. Small gathering recognizes broken rail and joint bar surrenders. These gatherings were utilized to break down reason particular crash recurrence and seriousness. The reason bunches are positioned in diving request by number of crashes and aggregate number of autos wrecked, separately. The previous metric relates to crash recurrence, though the last records for wrecking recurrence and seriousness. Diverse positioning strategies may prompt distinctive security change prioritization choices [2].

## III. DATA MINING AND ACCIDENTS

Data Mining is a computational strategy to manage huge and complex informational index and these informational indexes can be of typical, ostensible and blended. It is very simple to use in assortment of area have a place with science and administration; additionally, it could be utilized as a part of extortion distinguishing proof and numerous more logical cases and also in mishap seriousness issue. Parcel of items in a gathering of groups or in a homogeneous set is a crucial activity of information mining. Information mining method is perceived as solid strategy for investigation of train crash seriousness issue and discovering factors behind them. Information mining is broadly utilized as a part of a few investigations identified with transportation frameworks, for example, train accident information examination, track asphalt information examination, track harshness information investigation and so on. Harm like property, individuals because of rail accident are bothersome. Ordinarily, it happened that train mishap episodes are more typical at specific places that can help in distinguishing factors behind them [4].

## IV. BACKGROUND

Wasnik et al. [4] conveyed two year review investigation of railroad related deadly cases in the Department of Forensic Medicine and Toxicology, Indira Gandhi Government Medical College, Nagpur. Railroad casualty cases were remain for 5.99% (Total 173) of all post-mortems (n=2888)

directed amid the time of January 2001 to December 2002. The fatalities were overwhelmingly found in mails

Nireesha et al. [5] depicts the utilization of mining with a mix of strategies to consequently find accident qualities that can admit a superior comprehension of the supporters of the mishaps. Rail mishaps portray a huge wellbeing worry for the transportation business in different nations. To better comprehend the supporters of these outrageous accidents, the Federal Railroad Administration has built up the railways associated with mishaps to submit reports that contain both settled capacity passages and accounts that clarify the attributes of the mishap. While various investigations have seen the settled fields, nothing has completed a broad examination of the accounts. The outcomes demonstrate that prescient exactness for mishap costs fundamentally enhances the utilization of highlights found by mining and prescient precision additionally enhances using present day troupe strategies. Significantly, this investigation besides appears through case illustrations how the discoveries from mining of the accounts can enhance comprehension of the supporters of rail mishaps in ways unrealistic on just settled field examination of the mischance reports.

Letia et al. [6] takes care of the railway traffic control issue utilizing the asset allotment. The railway activity is portrayed as a vast and dynamic framework with unverifiable properties identified with asset stacking, prepare entries and disappointments. Regardless of these vulnerabilities, the control framework is required to ensure that every one of the trains carry on as per their timetables. The trains are considered errands with determined transient practices that need to satisfy their due dates. The arrangements in light of open circle, shut circle with free, planned and hierarchical controllers are characterized and thought about. The control signals are actualized and confirmed utilizing time Petri nets. A few calculations for control framework execution are given. The strategy assessments are performed utilizing the meter capacities: utility, usage, reservation and effectiveness. The outcomes got through recreations demonstrate that the proposed conveyed controllers take care of enough of the control issues and can be utilized for expansive scale usage.

Ghomi et al. [7] distinguish vehicle driver damage seriousness variables of expressway railroad level intersection (HRGC) accidents with a specific end goal to recognize communications and also dissimilarities among mishap factors. At this point, information mining methods were utilized to break down the communication of numerous variables in huge databases. This paper applies Classification-Regression Tree (CART) and Association Rules calculations on the U.S. Government Railroad Administration (FRA) HRGC accident database for the time of 2006 - 2013 to recognize vehicle driver damage seriousness factors at HRGCs. Both the arrangement trees and the principles disclosure were viable in giving significant experiences about

accident factors and their connection. The consequences of the two calculations were never opposing. Moreover, a large portion of the discoveries of this investigation were reliable with the consequences of past examinations which utilized diverse systematic procedures, for example, probabilistic models of mishap damage seriousness. The outcomes demonstrate that prepare speed, sort of street vehicle, driver age and sex, position of street vehicle before mishap, kind of mischance and thruway asphalt write are the key elements affecting the driver damage seriousness.

Mirabadi et al. [8] break down the information from past mishaps of the Iranian Railway (RAI) by applying affiliation rules information mining systems with a specific end goal to find and uncover obscure connections and examples among the information. The request to movement by rail is regularly expanding in light of the fact that it benefits the two travelers and cargo; hence it is of most extreme significance for railroad managers to convey travelers and cargo securely to their goals. Experiencing wellbeing techniques and creating security frameworks require consciousness of what is causing hazardous conditions. This can be expert by gaining from the past. By the use of CRISP-DM as the information mining approach and using Clementine 12.0 as the product device, the said targets of this paper were satisfied. For this examination approximately 6500 mishap records were chosen from the accidents database from 1996 to 2005. A definitive connections and examples removed can be used to create controls and guidelines. This exploration considers mishap conditions and connections found among the most well-known mischance factors (human mistake, wagon and track) with different fields of the database keeping in mind the end goal to keep them from happening later on.

Wang et al. [9] set forward hypothesis of defenselessness and the hypothesis of change with a specific end goal to defeat the inadequacies of expansive traveler activity caused by the security of urban rail travel, in view of an enhanced strategy. To begin with, the essential information of urban rail transport presents movement issues and escape clauses. At that point, another defenselessness hypothesis depends on the transformation hypothesis. Furthermore, it is utilized for activity issues in the field of urban rail travel tasks. 3 long stretches of information on the sheltered transportation of urban rail travel applications. In view of the above techniques, a case of urban rail travel security is broke down. The last delicate urban rail travel tram station is arranged. Urban rail travel tram movement control measures proposed and confirmed. An illustration demonstrates that the calculation strategy can take care of this issue.

Corman et al. [10] introduce a study of the current methodologies on online railroad movement rescheduling issues, which show dynamic and stochastic (or, at any rate, not totally deterministic) viewpoints. Railroad timetables are produced to make tasks hearty and strong to little

postponements. Nonetheless, unsettling influences annoy the everyday design, and dispatchers change the arrangement to keep tasks plausible and to confine postpones proliferation. Rescheduling approaches go for refreshing the disconnected timetable, best case scenario, within the sight of postponements. Indeed, while online static rescheduling has achieved a wide level of spread, much is still to be finished with respect to the probabilistic idea of the railroad movement rescheduling issues, and furthermore how to best consider for future states. Open difficulties for the future research are at last illustrated.

Kecman et al. [11] presents an ongoing instrument for consistent online forecast of prepare activity utilizing a coordinated occasion chart that catches every single planned occasion and priority relations between them, for example, prepare runs and stops, associations, and least degrees of progress. Expectation of prepare positions in time and space is required for activity control and traveler data. Nonetheless, by and by just the last estimated prepare delays are known and dispatchers must foresee the entry times of trains without sufficient computer support. Vigorous evaluations for the base procedure times are determined online by registering little percentiles (contingent on current defer where pertinent) for struggle free running circumstances that are acquired by pre-preparing verifiable prepare describer information. The chart is refreshed consistently when new data ends up accessible on prepare positions or movement control choices. The acknowledgment times of all occasions in the diagram are anticipated considering the utilization of running time supplements and cradle times, and in addition time misfortune because of course clashes in light of a contention identification plot inside the forecast calculation.

Arivazhagan et al. [12] proposed an Automatic Railway Track Derailment Inspection System utilizing Machine Vision Algorithm to distinguish the breaks in the railroad track. Crashes occur when a prepare keeps running off its rails and are truly dangerous to human security. The vast majority of the Railway Track surrenders which prompt crash are recognized physically via prepared human administrators strolling along the track. The info picture is disintegrated by Gabor channel and surface highlights were removed utilizing Segmentation based Fractal Texture Analysis (SFTA) and the highlights are named deformity and imperfection free classes utilizing AdaBoost Classifier. The proposed calculation is tried on an arrangement of continuous examples gathered and the characterization rate acquired was agreeable.

Oneto et al. [13] construct an information driven prepare defer expectation framework that adventures the latest investigation instruments. Best in class prepare postpone forecast frameworks neither adventure authentic information about prepare developments, nor exogenous information about wonders that can influence railroad activities. They depend, rather, on static standards worked by specialists of the railroad

foundation in light of established univariate measurements. The prepare postpone expectation issue has been mapped into a multivariate relapse issue and the execution of piece strategies, gathering techniques and feed-forward neural systems have been looked at. Right off the bat, it is demonstrated that it is conceivable to construct a dependable and hearty information driven model construct just with respect to the recorded information about the prepare developments. Moreover, the model can be additionally enhanced by including information originating from exogenous sources, specifically the climate data gave by national climate administrations. Results on certifiable information originating from the Italian railroad organize demonstrate that the proposition of this paper can surprisingly enhance the present cutting edge prepare defer expectation frameworks. Besides, the performed recreations demonstrate that the incorporation of climate information into the model has a noteworthy positive effect on its execution.

Tschirner et al. [14] investigates the idea of a constant activity design (RTTP) to facilitate coordinated effort between the diverse performing artists, and shows how it can be executed in frameworks for prepare movement control and driver data. The requests on current railroad movement frameworks are high. Higher productivity is required, which means better usage of framework limit and lessened vitality utilization. Auspiciousness has a high need and wellbeing must be unequivocal. The task of railroad activity incorporates numerous performers in various parts and separate associations. These investigations of prepare activity control have demonstrated that enhanced cooperation between the performing artists and propelled control frameworks are expected to meet the levels of popularity. Rather, numerous performers are following their own plans in view of their own objectives and inadequate data. This paper introduces the activity control framework STEG and the driver warning framework CATO. The two frameworks are being used, permitting re-arranging and sharing of such a RTTP. In light of these frameworks, we talk about general and particular

**A brief review of research and development in the subject: -**

outline arrangements, as per human factors and clarify a method for

presenting computerization that backings the movement controllers without meddling with their arranging. With these frameworks, it can demonstrate that a more all-encompassing way to deal with prepare movement control, in light of a RTTP, is actually practical and that imparting this arrangement to the prepare drivers generously enhances characteristics in prepare activity control.

Sridhar et al. [15] progression of remote correspondence and GPS frameworks prepared for better observing in Railway Transportation frameworks. The GPS based observing of trains will track trains, better booking and maintains a strategic distance from any setbacks. In the current framework, the planning of trains, movement railroad flagging is manual and less exact in foreseeing train landing goal time and no/or less crisis ready communicate sent to from and to station if there should arise an occurrence of any setbacks. Robotized Train booking will be consolidated with the assistance of GPS

framework and if there are simultaneous trains running in a similar way, planning is computerized to stay away from halt situation and prepare holding up issues and thus prepare voyaging time can be diminished. Activity Signaling is better taken care of with extra proactive measures by sending communicate message to the closest from or to stations with current scope area subtle elements of prepare. The closest station where prepare has been ended can be found by utilizing the most brief separation utilizing proposed Hybrid most brief separation calculation utilizing scope transformation to weight and applying dijkstra to locate the most limited separation. Exact Train plan timings are classified and expected prepare landing in the goal will be sufficiently exact as classifications are dynamic. Thus fundamental advances can be completed for false activity and prompt medicinal administrations can be accommodated setbacks and decreases the rate of accidents.

S.no.	Paper	Author	Work Performed	Data set	Tools	Methods
1.	Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates	Liu et al.(2012)	Identify the causes of train accidents and their effect on accidents rates	Transportation railroads FRA U.S gov. data , the period 2001 to 2010 dataset.	Weka tool	Chi-square analysis
2.	Rail traffic management systems (TMS)	Davey et al.(2012)	Presents Railway operation management in Austria - and Rail Traffic management for GB rail.		Simulation tool	Remote Control, Generic system functions

3.	Text Mining the Contributors to Rail Accidents	Brown et al.(2016)	Discussed severity of road accident survey.	11 years from 2001 to 2012, dataset (FRA) U.S.	Weka tool	text mining techniques, random forests , partial least squares, latent Dirichlet allocation,
4.	Analysis of Railway Fatalities in Central India	Wasnik et al.(2010)	Presents two years study relevant to railway accident cases.	Two years Jan 2001 to Dec 2002 railway accident dataset. Nagpur , India	Weka tool	
5.	Analyzing the Train Accident Injuries using Mining Techniques”	Nireesa et al. (2017)	Describe techniques to identify characteristics of accident.	11 years 2001 to 2012 rail accidents U.S. Data set	Weka tool	Association rules, Apriori algorithm
6.	Train Traffic Control Based on Distributed Resource Allocation	T. S. Letia, et al.(2010)	Describe railway traffic control problem using the resource allocation.		Time Petri nets tool	Resource allocation, Scheduling approaches:- Open loop control, Closed loop control
7.	Identifying vehicle driver injury severity factors at highway-railway grade crossings using data mining algorithms	Ghomi et al. (2017)	Identify injury of drivers in factors of highway-railway grade crossing.	U.S (FRA) accident dataset the period of 2006 – 2013	STATA13 software package tool	Classification Regression Tree(CART), Association Rules
8.	Application of association rules in Iranian Railways (RAI) accident data analysis	Mirabadi et al. (2010)	Identify previous data of Iranian railway accidents.	RAI accidents data set , Two different parts: (1) 1996–2005 (2) 2006	Clementine 12.0 tool	Association rules, CRISP reference model
9.	Based on Large Passenger Flow Research on Safety Operation of Urban Rail Transit	Wang et al.(2017)	Presents vulnerability and mutation theory.	3 years of data on the safe transportation of urban rail transit		Vulnerability, mutation theory.
10.	A Review of Online Dynamic Models and Algorithms for Railway Traffic Management	Corman et al.(2015)	Presents a survey on current techniques on online railway traffic rescheduling issues.		Meta tool	Rescheduling approaches: - Open-Loop Approaches Dynamic approaches:- Closed- Loop Approaches
11.	An online railway traffic prediction model	Kecman et al. (2013)	Presents tool to predict continuous online traffic of trains.	Dutch train system TROTS Data	Online prediction tool	Microscopic graph model TEG, Prediction model
12.	Railway Track Derailment Inspection System Using Segmentation Based Fractal Texture Analysis	Arivazhagan et al.(2015)	Presents automatic system to inspect railway track.	256MB RAM Camera Primary 5MP, 2592 * 1944 pixels having the resolution of 320 * 240 pixels. Some of the samples in our dataset	Ada Boost tool - MATLAB	Gabor Filter, Data Fusion, Feature Extraction:-(SFTA)
13.	Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather Data	Oneto et al.(2016)	Presents prediction system to detect delay of trains.	RFI 6 months of data set related to two main areas in Italy,	Simulation tool - MATLAB	Multivariate regression problem, Kernel methods, Neural networks, data-driven model
14.	Safety Alert Signaling Measure in Train Transportation System and its Automated Behaviors Using GPS	Sridhar et al. (2017)	Incorporated automatic scheduling of the train using GPS system.		Simulation tool - MATLAB	GPS system, Hybrid shortest distance algorithm, Dijkstra algorithm

15.	Solutions to the problem of inconsistent plans in railway traffic Operation	Simon Tschirner et al (2014)	Presents real-time traffic plan (RTTP) to manage traffic of trains.		STEG tool	CATO method, Open control loop, Closed control loop
-----	---	------------------------------	---	--	-----------	---

## V. CONCLUSION

This paper presents various research work done in past in the field of rail accident analysis using data mining as a review. Today, accidents are one of the main reasons of destruction of life and property. Various techniques of different authors are discussed to prevent rail accidents. Further this paper also discussed about the cause of accidents and role of data mining in the analysis of accidents.

## REFERENCES

- [1] Sridhar.T, "Safety Alert Signalling Measure In Train Transportation System And Its Automated Behaviours Using Gps", IEEE, International Conference On Information,Communication & Embedded Systems, 2017.
- [2] Xiang Liu, M. Rapik Saat, Christopher P. L. Barkan, "Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates", Journal of the Transportation Research Board, 2012, pp. 154-163.
- [3] E. Davey, " Rail Traffic Management System (Tms)", 2012, Pp. 126-137
- [4] Donald E. Brown, "Text Mining the Contributors to Rail Accidents", IEEE, Transactions on Intelligent Transportation Systems, Vol.17,No.2,2016,pp. 346 - 355.
- [5] Ramesh NanajiWasnik, "Analysis of Railway Fatalities in Central India", pp. 311-314.
- [6] CH. Nireesha, N. Vijay Kumar, V. Babu, "Analysing the Train Accident Injuries using Mining Techniques", International Journal for Research in Applied Science & Engineering Technology, Vol. 5, Iss. 4, 2017, pp. 1159-1162.
- [7] T. S. Letia, A. Astilean, R. Miron, M.M. Santa, "Train Traffic Control Based on Distributed Resource Allocation", Symposium on Telematics Applications, 2010, pp. 1-6
- [8] Haniyeh Ghomi, Morteza Bagheri, "Identifying vehicle driver injury severity factors at highway-railway grade crossings using data mining algorithms", IEEE, International Conference on Transportation Information and Safety, 2017, pp. 1054-1059.
- [9] Ahmad Mirabadi, Shabnam Sharifian, "Application of association rules in Iranian Railways (RAI) accident data analysis", 2010, pp. 1427-1435.
- [10] Lei Wang, Yuntao Chen, Huaiyuan Zhai, Shouxin Song, "Based on Large Passenger Flow Research on Safety Operation of Urban Rail Transit", IEEE, 2017.
- [11] Francesco Corman, Lingyun Meng, "A Review of Online Dynamic Models and Algorithms for Railway Traffic Management", IEEE, Transactions on Intelligent Transportation Systems, Vol. 16, No. 3, 2015, pp. 1274-1284
- [12] Pavle Kecman, Rob M.P. Goverde, "An online railway traffic prediction model", (2013) pp. 1-19.
- [13] S. Arivazhagan, R. Newlin Shebiah, J. Salome Magdalene, G. Sushmitha, "Railway Track Derailment Inspection System Using Segmentation Based Fractal Texture Analysis", ICTACT Journal

on Image and Video Processing, Vol. 6, Iss. 1, August 2015, pp. 1060-1065.

- [14] Luca Oneto, Emanuele Fumeo, Giorgio Clerico, Renzo Canepa, Federico Papa, Carlo Dambra, Nadia Mazzino, and Davide Anguita, "Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather Data", IEEE, International Conference on Data Science and Advanced Analytics, 2016, pp. 459-467.
- [15] Simon Tschirner, Bengt Sandblad, Arne W. Andersson, "Solutions to the problem of inconsistent plans in railway traffic Operation", Journal of Rail Transport Planning & Management, 2014.

## Authors Profile

Manju Bala pursued Bachelor of technology from Punjab Technical University, India in Year 2015. She is Currently pursuing M.Tech in Department of computer science and Engineering from Punjab Technical University main campus, kapurthala, India. Her main research work focuses on Data Mining, Big Data Analytics.



Anshu Bhasin pursued Master degree from Punjab Technical University, India in year 2006, and She has done PhD in CSE in Optical Communications She is currently working as Assistant Professor in Department of Computer Science and engineering, IKGPTU Main Campus, Kapurthala since August, 2017). She has certifications in Java programming (Sun Microsystems), Penetration Testing and Virtualization, Ethical Hacking and Cyber Security. She has several research publications in SCI and Scopus Journal and is there viewer of various with international journals. She is the sole author of a book on "Computer Networks" and is the co-author of a book on "Web Technologies on ASP.Net". Her main research work Computer Networks, Fiber Optics based Networks, Optimization Techniques using heuristics. She has more than 18 years of teaching experience.

