INVESTIGATION OF RAILWAY ACCIDENT USING OPTIMAL DATA MINING TECHNIQUE

[1]Manju Bala , [2]Anshu Bhasin
[1]Student in Dept of CSE at I.K.G Punjab Technical University Kapurthala, Punjab, India
[2] Assistant Professor in Dept of CSE, at I.K.G Punjab Technical University Kapurthala, Punjab, India

Abstract **-** Railways are a major network where millions of people travel throughout and act as main source of transportation. A rail accident occurs due specific reasons like crushes of two rails on track, strolling individual, creature, or some other normal deterrents. There are several major data mining techniques which have been developed and used in various data mining projects. In this work, an improved apriori algorithm is proposed in which the pruning step of apriori is optimized. In this railway traffic analysis will be done using improved apriori algorithm. Experimental results demonstrate that proposed technique outperforms existing technique.

Keywords – **Railway accident, Data mining, Apriori algorithm, Association rules, Railway Traffic safety.**

I. INTRODUCTION

Railways give the least expensive and most helpful method of traveller transport both for long separation and rural movement additionally it assumes a critical part in the improvement and development of enterprises. Railways help in providing crude materials and different offices to the plant locales and completed merchandise to the market. [1]. Train accidents make harm foundation and moving stock and also benefit interruptions, and may cause losses and mischief the earth. In like manner, enhancing train working wellbeing has for quite some time been a high need of the rail business and the administration. Prepare mishaps 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 additionally shifts broadly, contingent upon the specific mishap cause. Proficient allotment of assets to anticipate mishaps 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 moderate every mischance cause would then be able to be assessed and assets assigned with the goal that the best security change can be accomplished for the level of speculation accessible.

Train accidents are ordered into wrecking, impact, highway– rail level intersection mishap, and a few different less incessant writes. at the point when there is in excess of one kind of mishap, the sort of mischance that happened first would be assigned for all reports identified with it. for instance, a wrecking caused by a crash would be delegated an "impact." highway– rail level intersection accidents in the REA database incorporate just those that happen at the highway– rail interface and include no less than one parkway client [2]. Accidents occurred because of the carelessness of driving. A rail accident occurs due specific reasons like crushes of two rails on track, strolling individual, creature, or some other normal deterrents. It could bring about damage, property harm, and demise. Train crash investigation required investigation of the different factor influencing behind them. In overview its seen that rough 1.2 million passing and 50 million wounds evaluated worldwide consistently.

Rest of the paper is organized as follows. Section 2 discuss the related work, section 3 describes problem statement , section 4 describe proposed technique, 5 show the experimental result section 6 paper concluded with future work.

Generation of solid affiliation tenets of rail accidents. According to the system of affiliation rules learning for rail accidents, information arrangement and successive information cleaning are finished thusly, visit thing sets and solid affiliation rules are produced. In this analysis, the base help esteem is set to 0.5%, and the base certainty esteem is set to 0.50%.

Examination of solid affiliation tenets of rail accidents. Association rules investigation of rail accidents information can not just make quantitatively examination of single factor of accident causation and accident traits, however can likewise illuminate multi-factor affiliation connection mining which is a troublesome assignment for ordinary numerical examination strategy. From the investigation of the educated solid Association rules learning, we can realize that:

- From the measurements of accident outcome, we can see that mischances have minor result happened in the locale represented the principle part of the rail accidents.
- During the haze season Q territory of railway department have high-recurrence rail accidents.

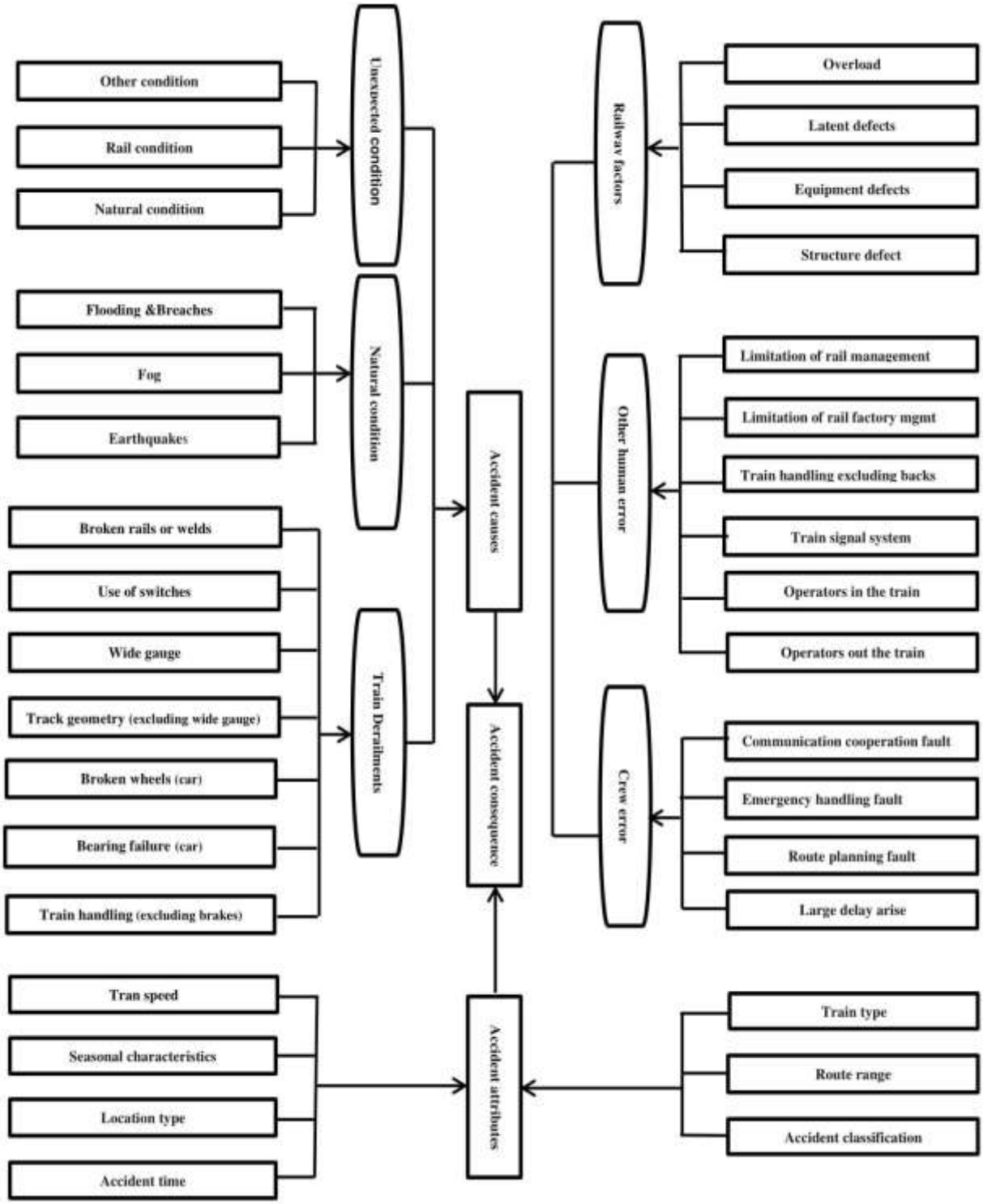

Figure 1: Model of Railway Accident

II. RELATED WORK

Arivazhagan et al. proposed an Automatic Railway Track Derailment Inspection System utilizing Machine Vision Algorithm to identify the splits in the railroad track. The proposed calculation is tried on an arrangement of constant examples gathered and the order rate got was tasteful [1]. Liu et al. presents Analysis of the reasons for prepare accidents is basic for judicious portion of assets to lessen mishap event in the most savvy way conceivable. Prepare wrecking information from the FRA rail gear mishap database for the interim 2001 to 2010 were investigated for each track write, with representing recurrence of event by cause and number of autos crashed [2].Wang et al. set forward the hypothesis of weakness and the hypothesis of transformation keeping in mind the end goal to defeat the inadequacies of substantial traveler movement caused by the security of urban rail travel, in view

of an enhanced strategy. The last delicate urban rail travel metro station is arranged. Urban rail travel metro activity control measures proposed and confirmed [3].Tschirner et al. show 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, talk about general and particular outline arrangements, as per human factors and clarify a method for presenting computerization that backings the movement controllers without meddling with their arranging [4]
.Kecman et al. presents a constant instrument for consistent online expectation of prepare activity utilizing a coordinated occasion diagram that catches every booked occasion and priority relations between them, for example, prepare runs and stops, associations, and least types of progress [5].Oneto et al. assemble an information driven prepare defer forecast framework that endeavors the latest investigation devices. The prepare defer expectation issue has been mapped into a multivariate relapse issue and the execution of portion techniques, troupe strategies and feed-forward neural systems have been looked at [6].
Viglioni et al. presents a model of the formal procedure of advancement of frameworks of disclosure of information in database for the forecast of railroad request, that incorporates a precise and thorough strategy, which coordinates the approachs: CRISP-DM, SEMMA, FAYYAD, and an intelligent situation for the usage of these frameworks [7]. Sridhar et al. joined robotized Train booking with the assistance of GPS framework and if there are simultaneous trains running in a similar way, planning is mechanized to maintain a strategic distance from halt situation and prepare holding up issues and consequently prepare voyaging time can be decreased [8]. Dinmohammadi et al. distinguished the potential dangers of startling disappointments happening in moving stock utilizing a disappointment mode, impacts and criticality investigation based approach [9].
Changhai et al. recognize potential causal connections among the numerous variables that assume a part in oceanic accidents. Correspondingly, affiliation govern learning is chosen as investigation approach, on account of its utility in getting affiliation administers through information mining on oceanic accidents information. In light of the investigation of affiliation administer taking in, this examination outlines the affiliation rules learning methodology of sea accidents and builds up the affiliation manage learning model of sea accidents [10].

III. PROBLEM STATEMENT

There are several major data mining techniques which have been developed and used in various data mining projects. In the existing technique maritime accidents had been analyzed using apriori algorithm. In Apriori algorithm minimum support is needed to generate the frequent itemset and the association rules by pruning the candidate sets using the user-defined minimum support threshold. Apriori algorithm have 2 problems: First, if minimum support is set too high, those rules that involve rare items will not be found. Secondly, to find rules that involve both frequent and rare items, minimum support has to be set very low. This may cause combinatorial explosion because those frequent items will be associated with one another in all possible ways. The Apriori is utilizing hit and trial method to find the required number of rules. Sometimes valuable association rules get pruned and sometimes not so required rules are generated, as the user is unaware of the required threshold for better rules. In order to solve the overcome, an improved apriori algorithm is proposed in which the pruning step of apriori is optimized. Also, lot of work has been done in road accidents; therefore, in this railway traffic accident analysis will be done using improved apriori algorithm.

IV. PROPOSED TECHNIQUE

1. Collection of raw data are collected from(Rail Equipment Accident/Incident Report (REAIR) form (FRA F 6180.54)) and then apply filtering techniques to make that raw data into structured format: Filtering techniques like Replace Missing Value filter

2. **Enhanced Apriori algorithm**
Enhanced-*Apriori* is the proposed algorithm which focuses in improving the pruning procedure of Apriori algorithm by using average support (sup_{avg}) instead of minimum support (sup_{min})It is used to generate probabilistic itemset instead of large itemset.

Here average support is not user defined value but it is calculated using the formula:

$$sup_{avg} = \frac{\sum_{k=1}^{m}(\sup(k))}{m}$$

where m is the number of items involved.

So, this formula would lead to better itemsets and this increases the number of better association rules that were left underestimated by Apriori algorithm.

Enhanced-Apriori contains following steps:

1. Calculate sup_{avg} from given database D

a. For itemset with sup ≥ sup_{avg} are inserted into probability itemset.
b. For itemsets with sup ≤ sup_{avg}, itemsets with sup ≥ S_p (where S_p = sup_{avg} / 1.2) and their probability of occurrence >80% (i.e. items with sup greater than 80% of the maximum support of itemset having sup>S_p) are inserted into probability itemset, while others are pruned.

2. Above step is repeated for all probability itemsets and from this probability itemset, association rules are produced.

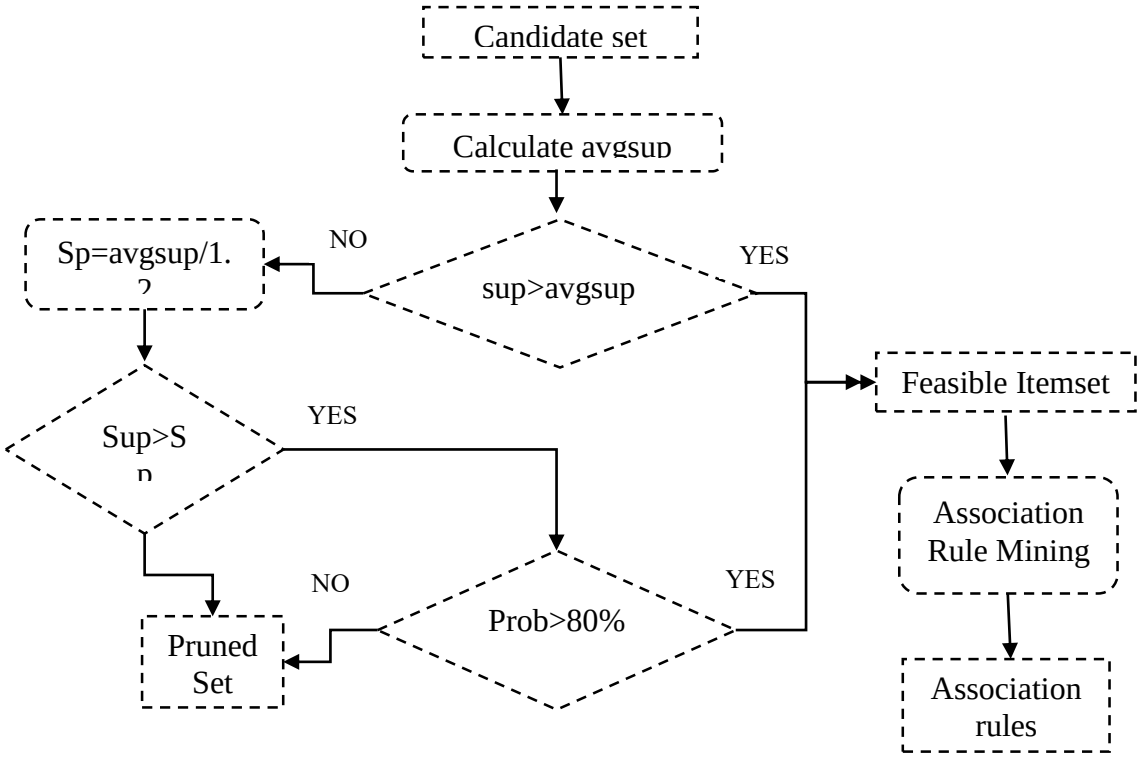

Figure 2: Flowchart of the proposed technique

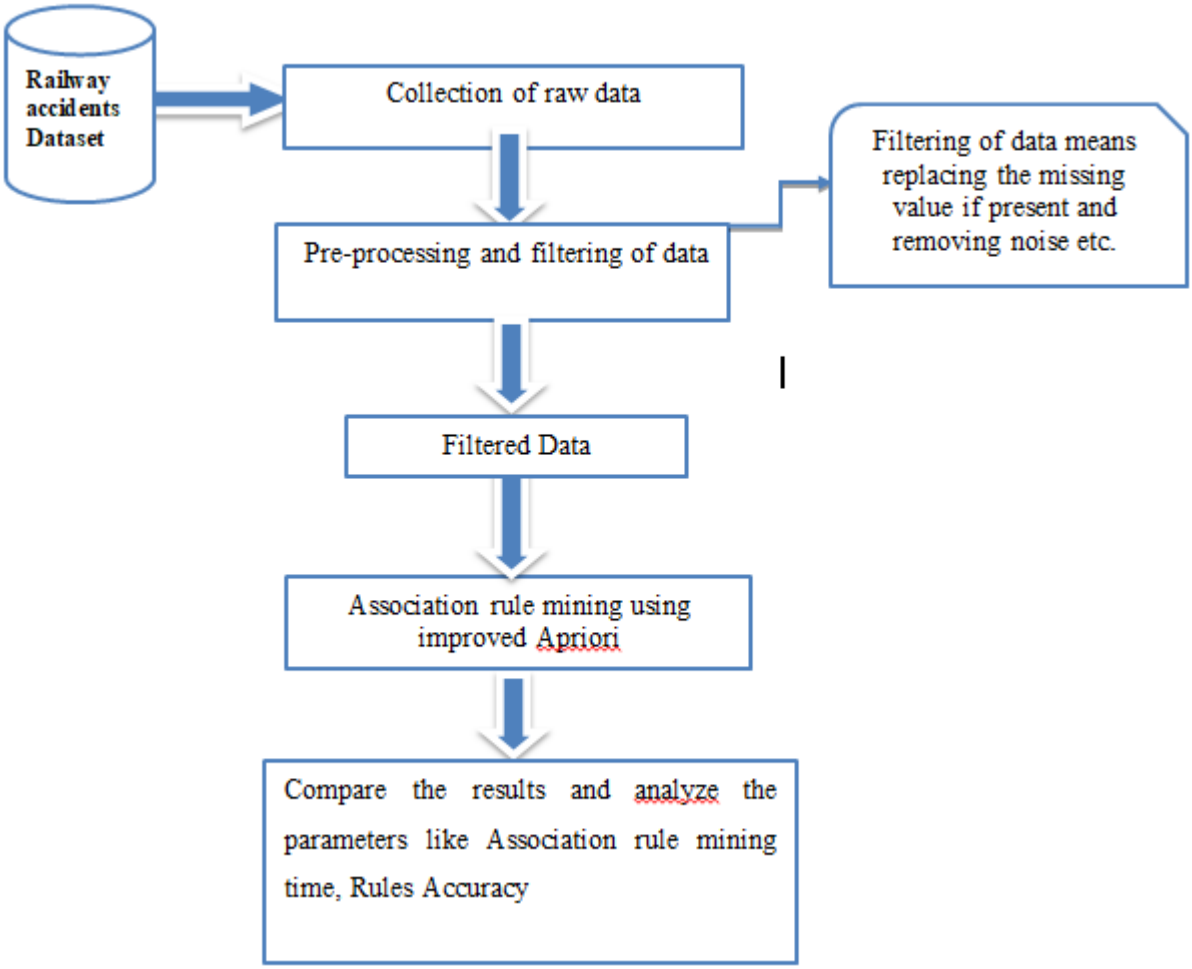

Figure 3: block diagram of proposed technique

v. EXPERIMENTAL RESULTS

This section presents experimental results of the proposed technique. Performance of proposed technique is evaluated on the basis of:

- Execution time
- Accuracy

Table 1: Comparison of execution time between Apriori and IP-Apriori

Min support	Min Confidence	Apriori	IP-Apriori
[0.5]	[0.5]	24648	23244
[0.6]	[0.6]	18574	18503
[0.7]	[0.7]	17658	16515
[0.8]	[0.8]	9750	9113
[0.9]	[0.9]	6740	6782

Above table shows results of proposed ip-apriori technique with apriori algorithm on the basis of execution time.

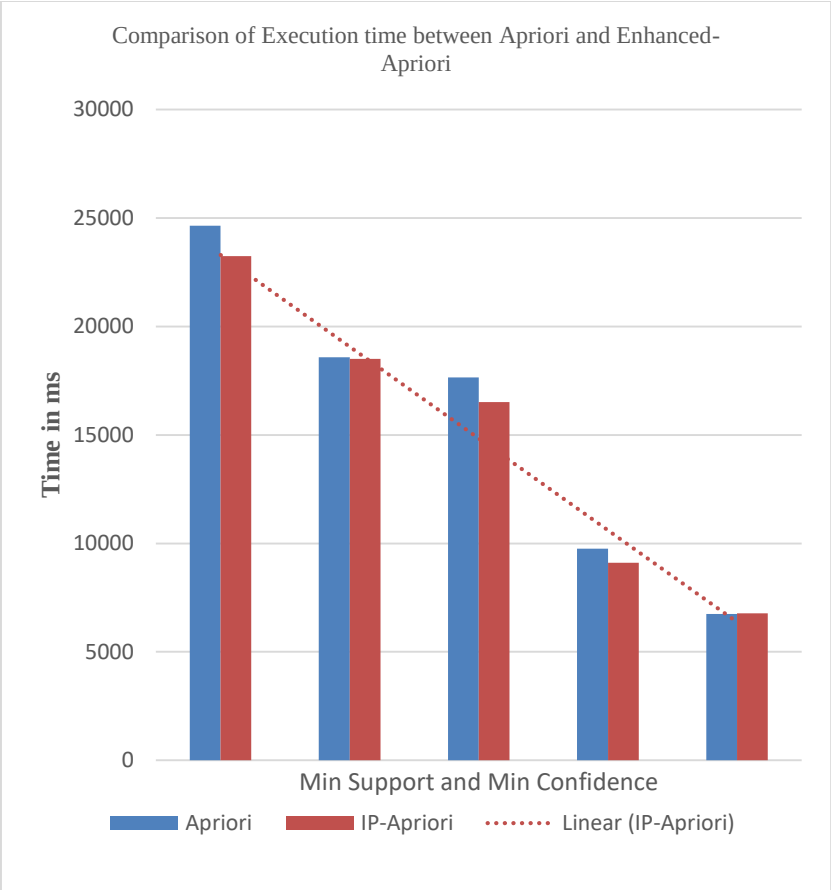

Figure 4: Showing comparison of execution time between Apriori and IP-Apriori

Above figure shows comparison of ip-apriori algorithm with apriori algorithm on the basis of execution time. It is shown that execution time of proposed technique is less than the existing technique.

Table 2: Comparison of accuracy between Apriori and IP-Apriori

Min support	Min Confidence	Apriori	IP-Apriori
[0.5]	[0.5]	89.3444	92.9077
[0.6]	[0.6]	89.69918	91.31137
[0.7]	[0.7]	91.24483	95.08196
[0.8]	[0.8]	94.44944	97.75756
[0.9]	[0.9]	97.73781	99.61452

Above table shows results of proposed ip-apriori technique with apriori algorithm on the basis of accuracy.

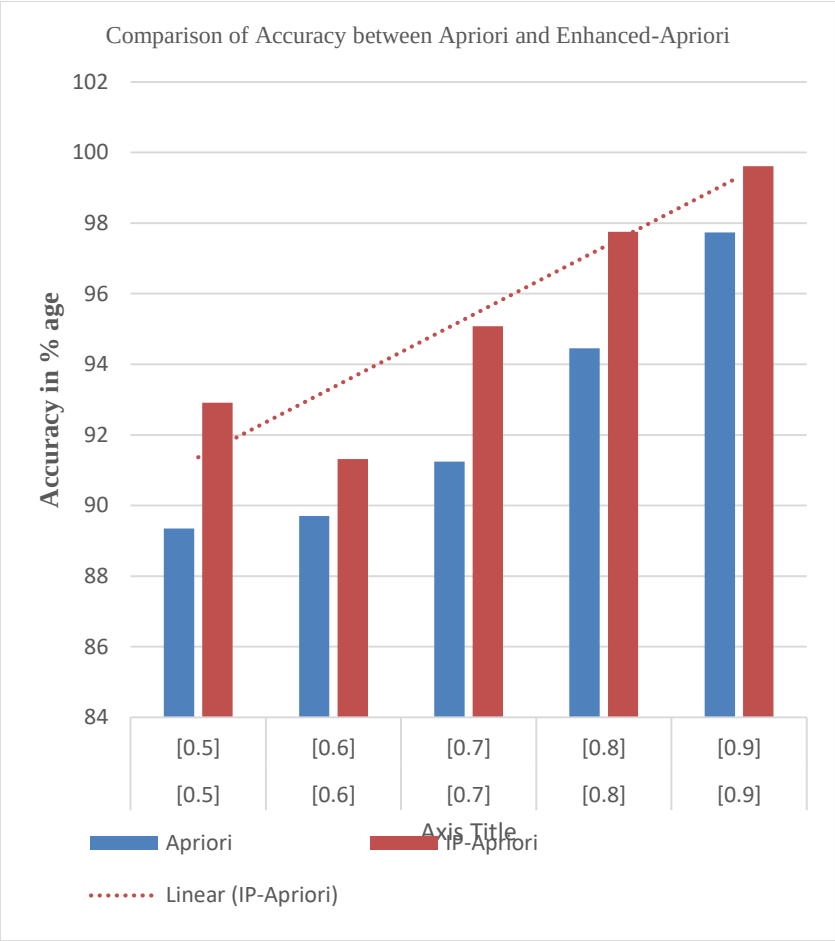

Figure 5: Showing comparison of accuracy between Apriori and IP-Apriori

Above figure shows comparison of ip-apriori algorithm with apriori algorithm on the basis of accuracy. It is shown that accuracy of proposed technique is more than the existing technique.

VI. CONCLUSION

This research work proposed improved apriori algorithm in which the pruning step of apriori is optimized. Also, lot of work has been done in street accidents; therefore, in this railway traffic accident analysis will be done using improved apriori algorithm. Performance of proposed technique with Apriori algorithm is evaluated on the basis of execution time and accuracy. Results of proposed apriori algorithm is compared with apriori algorithm. Experimental results show that execution time of proposed apriori algorithm is less than that of apriori algorithm and accuracy of proposed apriori algorithm is more than that of apriori algorithm.
In this work we improve apriori algorithm by optimizing pruning step. In future, we may use fpgrowth algorithm to further optimize the results. Also we may perform analysis of any other data set other than railway.

REFERENCES

[1] S. Arivazhagan, R. Newlin Shebiah, J. Salome Magdalene and G. Sushmitha, August 2015 , “Railway Track Derailment Inspection System Using Segmentation Based Fractal Texture Analysis”, Ictact Journal on Image and Video Processing, Vol. 6, Issue. 1, pp. 1060-1065.

[2] Xiang Liu, M. Rapik Saat, and Christopher P. L. Barkan, December 2012, “Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates”, Journal of the Transportation Research Board, pp. 154–163.

[3] Lei Wang, Yuntao Chen， Huaiyuan Zhai, Shouxin Song, July 2017 “Based on Large Passenger Flow Research on Safety Operation of Urban Rail Transit”, IEEE.

[4] Simon Tschirner, Bengt Sandblad, Arne W. Andersson , December 2014, "Solutions to the problem of inconsistent plans in railway traffic Operation", Journal of Rail Transport Planning & Management, Vol 4 issue 4,pp. 87–97.

[5] Pavle Kecman, Rob M.P. Goverde, May 2013 ,"An online railway traffic prediction model".

[6] Luca Oneto, Emanuele Fumeo, Giorgio Clerico, Renzo Canepa, Federico Papa, Carlo Dambra, Nadia Mazzino, and Davide Anguita, 2016, "Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather Data", IEEE, International Conference on Data Science and Advanced Analytics, pp. 458-467.

[7] Giovanni Melo Carvalho Viglioni, Instituto Militar de Engenharia, 2007, "Methodology for Railway Demand Forecasting Using Data Mining", Elsevier Publications, pp 161-170.

[8] Sridhar.T, V.Ulagamuthalvi, February 2017 "Safety Alert Signalling Measure in Train Transportation System and its Automated Behaviours Using GPS", IEEE, International Conference on Information, Communication & Embedded Systems.

[9] Fateme Dinmohammadi, Babakalli Alkali, Mahmood Shafiee, Christophe Be´renguer, Ashraf Labib , 2016, "Risk Evaluation of Railway Rolling Stock Failures Using FMECA Technique: A Case Study of Passenger Door System", Springer Publications , 2016, pp. 128–145.

[10] Huang Changhai, Hu Shenping , February 2018, "Factors correlation mining on maritime accidents database using association rule learning algorithm", Springer Publications.
.
[11] Jovita Vani Sequeira, Zahid Ansari, March 2015, "Analysis on Improved Pruning in Apriori Algorithm" International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, Issue 3, pp.894-902.