Railway Accident Factors Analysis using Modified Checkpoint Apriori Algorithm

Priyanka Joshi[1] Gori Shanker[2]
[1,2]St. Margaret Engineering College Neemrana, India

Abstract— Accidents in particular to talk about the railway accidents happen regularly in India but then can be reduced if we analyze the cause of accidents. In our study we are taking into consideration some factors like Road which are connected to the Junctions where the accidents took place, weather in which the accident took place, day time or night time when the accident occurs and more factors. And then we try to find out the combinations using the Modified Checkpoint based Apriori Algorithm the combinations of the factors which causes the most accidents and try to take them into consideration.

Key words: Railway Accident Factors, Apriori Algorithm

I. INTRODUCTION

Indian Railways is preparing for taking off fast traveler benefits even as it appears to be set to record its most exceedingly bad execution in three years on mishaps. Crisp information on wellbeing demonstrates complete number of mishaps per million train kilometers has found the middle value of 0.14 in the current money related year so far - higher than the numbers recorded in each of the last three years. [1]

The proportion of number of mischance's that jumped out at each million kilometer secured via trains has declined reliably from 0.29 in 2004-05 to 0.10 in 2013- 14. In any case, as indicated by railroads' most recent proficiency parameter report, the proportion remained at 0.14 in the middle of April and September 2014. On the off chance that kept up at the present level, the current financial would check the principal yearly ascent in the proportion in the past decade. [1]

Information mining is a promising zone for managing the expanded, put away information that has been produced in our times .It is the extraction of understood, already obscure and valuable information. In this paper we have examined a percentage of the information mining systems, devices, applications and web indexes for mishap examination and activity investigation. The vast majority of the mishap examination procedures depend on situations of the mischance event and re-enactment of mishap circumstance. The expenses of fatalities and wounds because of auto collision greatly affect society.

Architects and analysts in the car business have attempted to outline and assemble more secure vehicles, however auto collisions are unavoidable. As of late, scientists have been using genuine information in examining different parts of car crashes. So measures must be taken to diminish mischances. It is vital that the measures ought to be founded on investigative and target reviews of the reasons for mischance's and seriousness of wounds. Our study highlights different devices, strategies and utilizations of information mining in mischance examination will take out lack of different systems however covers their focal points. Our fundamental point is to beat the passing rate and the expanded rate of death toll by method for utilizing a few apparatuses, procedures or different calculations in the field of information mining utilizing the movement information bases [2].

II. IMPORTANCE AND RELEVANCE OF THE STUDY

Shen Bin 1, Liu Yuan 1, Wang Xiaoyi 1 propose four information digging models for the Internet of Things, which are multi-layer information mining model, circulated information mining model, Grid based information mining model and information mining model from multi-innovation joining point of view. Among them, multi-layer model incorporates four layers: 1) information gathering layer, 2) information administration layer, 3) occasion preparing layer, and 4) information mining administration layer. Conveyed information mining model can take care of issues from storing information at various destinations. Lattice based information mining model permits Grid structure to understand the elements of information mining. Information mining model from multi-innovation reconciliation point of view portrays the relating structure for the future Internet. [1]

Xindong Wu 2, Gong-Qing Wu 2, and Wei Ding 2 presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. They analyze the challenging issues in the data-driven model and also in the Big Data revolution. [2]

Feng Bao 3, Xu He 3, Fengzhi Zhao 3, addresses the features of the petro physical data, logging data, seismic data and geological data based on the concepts of the data mining. The mining ideas regarding the petro physical and logging data, seismic data and geological data are made based on their features. They uses different mining way to process the corresponding data, and describes the result from the perspective of the functions of data mining.

By information mining systems, the petro physical information are connected to discover the relations and conjecture repository the logging information will be utilized to assess the fluffy stores and perceive the powerful supplies in confused topographical conditions; the space mining consequence of the 3D seismic information; the graphs and content mining aftereffects of the geographical information. [3]

Ms Shweta 4, Dr. Kanwal Garg 4 considers information (bank information) and tries to acquire the outcome utilizing Weka an information mining instrument. Affiliation principle calculations are utilized to discover the best mix of various qualities in any information. In this paper creator utilizes Apriori to discover affiliation guideline. Here creator consider three affiliation principle calculations: Apriori Association Rule, Predictive Apriori Association Rule and Tertius Association Rule. Ms Shweta, Dr. Kanwal Garg analyzes the aftereffect of these three calculations and presents the outcome. By result acquired utilizing

information mining instrument creator find that Apriori Association calculation performs superior to the Predictive Apriori Association Rule and Tertius Association Rule calculations. [5]

A. "Study of Data Mining Tools in Knowledge Discovery Process" By Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam

Data mining, the extraction of hidden prognostic information from big databases, may well be a strong new technology with nice potential to facilitate companies focus on most important information in their info warehouses. It uses machine learning, applied mathematics and image techniques to discovery and gift information in a form that simple intelligible to humans. Varied well-liked data mining tools are get-able these days. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven choices. Data mining tools can answer business queries that traditionally were too time overwhelming to resolve.

In its simplest kind, data mining automates the detection of relevant patterns in a very information, use outline approaches and algorithms to look into current and historical data which can then be analyzed to predict future trends. As a results of data mining tools predict future trends and behaviors by reading through databases for hidden patterns, they allow organizations to make proactive, knowledge-driven picks and answer queries that were antecedent too long to resolve. [5]

B. Another Paper is "Mining Big Data in Real Time" by Albert Bife.

Streaming information analysis in real time is changing into the quickest and most ancient thanks to get helpful information from what is happening currently, permitting organizations to react quickly once issues seem or to notice new trends serving to to enhance their performance. Evolving knowledge streams square measure contributory to the growth of knowledge created over the last few years. We have a tendency to square measure making the same amount of information each 2 days, as we have a tendency to create from the dawn of time up till 2003.Evolving knowledge streams strategies are getting an inexpensive, inexperienced methodology for period of time on-line prediction and analysis. We have a tendency to discuss this and future trends of mining evolving knowledge streams, and the challenges that the eld can got to overcome throughout consecutive years.

Nowadays, the number of information that's created each 2 days is calculable to be five Exabyte's. This quantity of knowledge is similar to the quantity of information created from the dawn of your time up till 2003. Moreover, it absolutely was calculable that 2007 was the remainder year within which it absolutely was unattainable to store all the information that we have a tendency to be manufacturing. This large quantity of information opens new difficult discovery tasks. knowledge stream real time analytics square measure required to manage the knowledge presently generated, at AN ever increasing rate, from such applications as: device networks, measurements in network observance and track management, log records or click-streams in internet exploring, producing processes, decision detail records, email, blogging, twitter post sand others. In fact, all information generated are often thought-about as streaming knowledge or as a photo of streaming information, since it is obtained from AN interval of your time. In the knowledge stream model, information reach high speed, and algorithms that method them should do thus beneath terribly strict constraints of house and time. Consequently, knowledge streams create many challenges for data processing algorithmic program style. First, algorithms should create use of restricted resources (time and memory). Second, they have to handle with information whose nature or distribution changes overtime. [6]

III. PROPOSED CONCEPT

The goal of the proposed method is to reduce CPU time which is saved by reducing candidate set size. If candidate set size is less than time required to calculate the support of each candidate is less. We have proposed Method that reduces the number of candidate generated and time required to calculate the support of each candidate. In order to reduce CPU times, we have defined two type of checkpoint in dataset based on support value:

checkpoint1=Total transaction - support count+1; ∀ min_sup;
checkpoint2=support count+1; if min_sup>50.

Support property: If support count is n than any item to be frequent it must be appearing in at least n transactions in the dataset.

All new candidates after checkpoint1 cannot be frequent based on support property. At checkpoint1 if min_ sup<=50 or checkpoint2 if min_sup>50, scan the candidate set once and check the support value of all candidate. The estimated support value is used to remove infrequent itemsets at checkpoint.

In this work, the author has proposed a system in which we will first manage the data used in the Accident analysis into the various tables and using a GUI software system which summarize the causes of the accident using the Modified Checkpoint based Apriori Algorithm.

Proposed System has the following Database Schema

- Person Table
- Junction Table
- Roads Table
- Road_Check
- Accident Table

A. Person Table

This table is used for storing the information regarding the drivers driving the trains at the time of the accident.

B. Junction Table

This table is used for storing the details regarding the junctions.

C. Roads Table

This table is used for storing the information regarding the roads which are connecting roads or roads at the crossing to the nearest junctions.

D. Accident Table

This table is used for storing the details of accidents.

E. Road_check table

This table is used for categorization of the road.

In the figure 1, we have implemented the Aprori based implementation on the sample data.

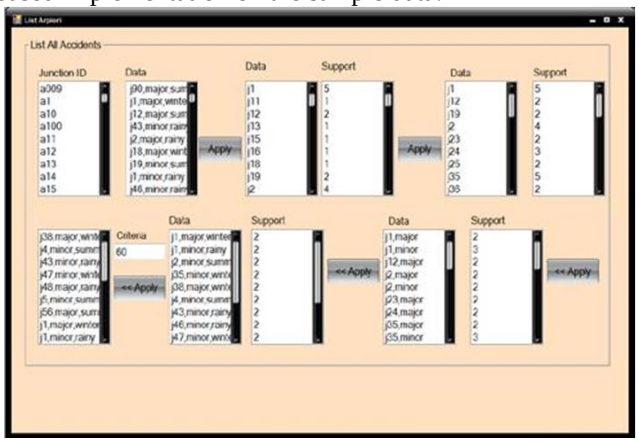

Fig. 1: Apriori Algorithm Implementation on the Sample Data.

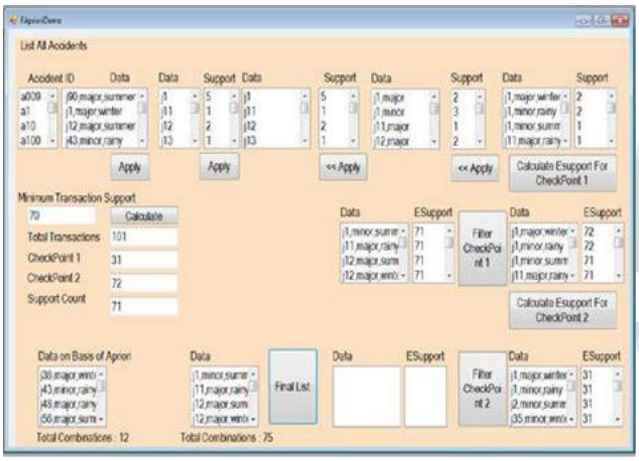

Fig. 2: Checkpoints based modified Aprori Algorithms

In the figure 1, we have shown the implementation of the checkpoint based apriori algorithm.

	A	B	C	D	E	F	G
1	accident_id	junction_id	road_id	accident_type	adate	time	weather
2	a1	j1	r4	major	22\1\1990	2:00am	winter
3	a2	j2	r5	minor	24\3\1989	3:00pm	summer
4	a3	j1	r7	major	1\2\2001	4:00pm	winter
5	a4	j6	r10	major	24\2\2004	8:00am	winter
6	a5	j3	r11	major	11\2\1994	6:00pm	summer
7	a6	j2	r21	minor	23\9\1998	7:00pm	summer
8	a7	j35	r34	minor	28\5\1996	8:00pm	rainy
9	a8	j47	r56	minor	10\08\1997	5:00pm	winter

Fig. 3 Example of the Sample Data Which We Have used in the Analysis Process.

Figure 3 shows the data which we have used in the proposed algorithm for the simulation of the algorithm.

IV. CONCLUSION & FUTURE SCOPE

This study is centered around how to take care of the effective issues of Apriori calculation and propose another changed check point based information mining apriori calculation. It concentrate on element finding for train mishaps examination and still information mining is an exceptionally inconceivable field so dependably a degree to enhance and grow new ideas for enhancing the aftereffects of the information mining operations.

REFERENCES

[1] Shen Bin, Liu Yuan, Wang Xiaoyi , "Research on Data Mining Models for the Internet of Things",Institute of Electrical and Electronics Engineers,2010
[2] Xindong Wu,Gong-Qing Wu, and Wei Ding, , "Data Mining with Big Data", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 1, January 2014
[3] Feng Bao, Xu He, Fengzhi Zhao, "Applying Data Mining to the Geosciences Data",International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE) ,2010
[4] Ms Shweta, Dr. Kanwal Garg "Mining Efficient Association Rules through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 6, June 2013.
[5] http://sfb649.wiwi.huberlin.de/fedc_homepage/xplore/ebooks/html/csa/node203.html
[6] http://users.jyu.fi/~samiayr/pdf/mining_road_traffic_accidents.pdf
[7] http://www.thearling.com/text/dmtechniques/dmtechniques.htm

