PROPOSING K-MODE BASED METHODOLOGY FOR ROAD ACCIDENTS WITH IMPROVED APRIORI

Inderpreet kaur
CTIEMT, Jalandhar

Ashish Kumar Luhach
Associate Professor, CTIEMT,Jalandhar

Pooja
Assistant Professor, CTIEMT, Jalandhar

Abstract- Road traffic accidents area unit a serious public health concern, leading to Associate in Nursing calculable 1.2 million deaths and fifty million injuries worldwide annually. within the developing world, road traffic accidents area unit among the leading reason behind death and injury. The target of this study is to judge a group of variable that contribute to the degree of accident severity in traffic crashes. the problem of traffic safety has raised nice issues across the property development of contemporary traffic and transportation. The study on road traffic accident cause can establish the key factors quickly, expeditiously and supply tutorial ways to the traffic accident interference and road traffic accident reduction,that might greatly cut back personal casualty by road traffic accidents. Mistreatment the ways of traffic knowledge analysis, can improve the road traffic safety management level effectively.

Keywords: Data Mining,Association Rule Mining,Road Accident Data.

I. INTRODUCTION

Deaths and injuries happening due to road accident have always been an issue of concern. However with the increase in vehicles on road and casual behavior of people, the rate of road accidents has increased many folds over past many years. As per the statistics available on ndtv.com [9], 1214 road crashes occur daily in India. More of such details are represented below under the heading "Stats on road accidents". In this paper we tend to analyse the data available on such road accidents so as to insight the major causes leading to accidents, believing that such an analysis will result in finding preventive measures towards minimizing road accidents.

1.1 Stats on Road Accident

- One serious road accident within the country happens each minute and sixteen die on Indian roads each hour.
- 1214 road crashes occur each day in Bharat.
- Two wheelers account for twenty fifth of total road crash deaths.
- 20 kids below the age of fourteen die each day as a result of road crashes in within the country.
- 377 folks die each day, like a gargantuan jet bloody each day.
- Two folks die each hour in province – State with most variety of road crash deaths.
- Tamil Nadu is that the state with the most variety of road crash injuries.

Analysis of data is done using techniques nad methododlogies of data mining domain. Same has been explored in this paper for the analysis of road accident data.

1.2 Data Mining

Data Mining is the term which means the overuse of data to obtain valid inferences.It is the discovery of useful summaries of data. Data Mining is the process to discover the knowledge or hidden pattern from the large databases and is a useful method required for the intersection of machine learning, dbase systems, AI and statistics. It works on the principle of retrieving relevant information from the data.Mainly in use of financial analysts, banks, insurance

companies, retail stores, business intelligence organizations, and hospitals, to extract the useful information that they require from the large databases. The goal of process is to create and find accurate patterns that are not previously known by us.The overall goal of the data mining is to extracting and obtain information from the databases and transform it into an understandable format for use in future. In actual data mining task can be automatic to extract the unknown patterns.

1.2.1. Data Mining Techniques

Many data processing techniques and systems are designed.These techniques will be classified supported the the information to be discovered, techniques to be used, and information.
Based on the info:There are several database systems that are employed in organizations, such as, group action info, abstraction info, multimedia system database, computer database, internet info, and legacy database. A Data Mining system can be classified based on the type of database for which it is designed.
Based on the techniques: Data mining systems can be analysed according to Data Mining techniques to be used as a Data Mining system can be categorized by driven method,that i data driven mining, autonomous knowledge mining, interactive Data Mining, and query-driven mining techniques.But, it can also be classified by its mining approach, such as statistical- or mathematical-based mining, integrated approaches, generalization primarily based mining, and pattern-based mining.
Based on the knowledge: As mentioned earlier in section one.1, data processing systems discovers varied kinds of information, together with classification, association, prediction, call tree bunch, and consecutive patterns. The information may also be classified into construction information, primitive-level information and public knowledge.

1.2.2. ASSOCIATION RULE MINING

Association Rule Mining may be a data processing operate that discovers the chance of co-occurrence of things in transactional information. Association rule mining may be a most significant and one in every of the well researched techniques among data processing, that was introduced in. It aims to extract attention-grabbing associations, casual structures, correlations or frequent patterns among sets of things in information repositories or group action databases.Let I is item-set of m distinct attributes, I = and group action information, D = , wherever T C I and there square measure 2 item-sets X and Y, such

X C T and Y C T. Then association rule, X=>Y holds wherever X C I and Y C I and X ∩ Y = Ø. Here, X is thought as antecedent whereas Y as resultant. There square measure 2 necessary basic measures for association rules, support(s) and confidence(c). Thresholds for support and confidence square measure predefined by users to eliminate the uninterested rules.

Support(s) of Associate in Nursing association rule is outlined because the share or fraction of transactions in D that contain X ∪ Y. Support(s) will be calculated by the subsequent formula:

$$\sup(XUY) = \frac{count(XUY)}{Count(D)}$$

Eq no.(1)

Confidence may be a live of strength of the association rules. Confidence is outlined by the share or fraction of the quantity of transactions in D that contain X conjointly contains Y. 'IF' part is Antecedent and 'THEN' part is resultant. It will be calculated by dividing the chance of things occurring along to the chance of prevalence of antecedent. Confidence(c) is calculated by the subsequent formula:

$$conf(X => Y) = \frac{\sup(X \cup Y)}{\sup(X)}$$

Eq no.(2)

Association rule mining is to seek out out association rules that satisfy the predefined supmin and confmin from a given information Objective of ARM is to seek out the universal set S of all valid association rules

Association rule for cluster one shows that 2 wheeler accidents square measure chiefly happens on specific road segments like intersections at community areas, i.e., colony, markets and hospitals. Intersections in colonies close to highways square measure a lot of liable to 2 wheeler accidents than colonies on non-highways. conjointly market square measureas are a lot of probably to own 2 wheeler accidents with 2 or a lot of injuries at evening around 4:00 p.m. to 8:00 p.m. Rules disclosed that hospitals space also are related to 2 wheeler accidents however most of the accidents at this place have happened in the dead of night time when 8:00 p.m.

Association rules for cluster two indicates that forest space and agriculture that are aside of bound highways are dangerous for 2 wheeler accidents as explosive bend, slope in the dead of night time will cause imbalance of driver and will cause accidents. Rules show that curves on craggy highways involves 2 injuries and principally tykes square measure concerned in such accidents. Also, no light-weight areas like forest also are liable to accidents in getting dark. Highways with agriculture aside are found to be accident prone areas.

Association rule for cluster three shows that almost all of the vehicle-fall from height accidents concerned quite two gashed. it's found that vehicles falling from height on craggy highways square measure severe accident wherever quite 2 gashed persons square measure there. the rationale can be the vehicle kind is hackney or similar class that transports quite two persons at a time. Also, it shows that principally vehicles fall from height from hill location square measure thanks to a curve on road that's the most characteristics of the hills.

Association rules for cluster four indicate that multi-vehicular and glued object/divider hit accidents square measure principally occurred in the dead of night time on road roads. Intersections on highways square measure another road feature for such style of accidents. principally the areas with no light-weight condition.are a lot of accident prone in getting dark and leads to crucial accidents. Rules show that curve at agriculture land and forest space and intersection at highways square measure a lot of dangerous in the dead of night time because it is troublesome for a speedy vehicle to guage the vehicles from opposite facet and glued object to avoid collision.

II. WORK DONE TILL NOW

Jain, G. Ahuja, Anuranjana3 D. Mehrotra, "Data Mining Approach to Analyse the Road Accidents in India", 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Sep. 7-9, 2016[1] To create a model that not solely smooths out the nonuniformity of the information by grouping similar objects along to search out the accident prone areas within the country with relation to totally different accident-factors however additionally helps confirm the association between these factors and casualties.Cluster analysis helps U.S.A. to work out the accident prone states and territories of Asian country. These clusters square measure labeled to be classified with the assistance of call tree to conclude the dominant issue, backing the accidents

S. Kumar, D. Toshniwal, "A Data Mining framework to analyze road accident data", Journal of Big Data, Springer, vol.2-no.1, pp 1-18, 2015.[2] To identify the most factors related to a road and traffic accident.However, heterogeneous nature of road accident information makes the analysis task troublesome.Data segmentation has been used wide to beat this nonuniformity of the accident information.K-modes clump technique as a preliminary task for segmentation ANd association rule mining square measure accustomed establish the varied circumstances that square measure related to the incidence of an accident for each the whole information set (EDS) and therefore the clusters known by K-modes clump formula.

D. Khera, W. Singh, "A Review on Injury Severity in traffic System using various Data Mining Techniques", International Journal of Computer Applications, vol.100-no.3, pp 0975- 8887, 2014.[4] To evaluate a collection of variables that contribute to the degree of injury severity in traffic crashes. the problem of traffic safety has raised nice considerations across the world and it's become one in every of the key problems difficult the property development of recent traffic and transportation. victimization the strategy of traffic information analysis, will improve the road traffic safety management level effectively.

A. T. Kashani, A. Shariat-Mohaymany, A. Ranjbari, "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", Promet Traffic & Transportation, Vol. 23, No. 1, pp11-17, 2011. [6] To spot the factors that have an effect on injury severity of drivers concerned in traffic crashes on these roads, so by eliminating or dominant such factors AN overall safety improvement may be accomplished. victimization the Classification and Regression Tree (CART), one in every of the powerful data processing tools, the crash information concerning the last 3 years (2006-2008) were analyzed. The variable choice procedure was administered on the idea of Variable Importance live (VIM) that is one in every of the CART methodology outputs.

B. Depaire, G.Wets and K. Vanhoof, "Traffic accident segmentation by means of latent class clustering, accident analysis and prevention", vol. 40. Elsevier; 2008.[7] The effectiveness of a clump technique, i.e. latent category clump, for distinguishing undiversified traffic accident varieties.Firstly, a heterogeneous traffic accident

information set is segmental into seven clusters, that square measure translated into seven traffic accident varieties. Secondly, injury analysis is performed for every cluster.Latent category clump analysis will reveal hidden relationships and might facilitate the domain skilled or traffic safety scientist to phase traffic accidents.

DipoT.Akomolafe, Akinbola Olutayo," Using Data Mining Technique to Predict Cause ofAccident and Accident Prone Locations on Highways", American Journal of Database Theory and Application 2012, 1(3): 26-38. The cost of deaths and injuries thanks to traffic accidents encompasses a nice impact on society. In recent years, researchers have paid a good attention at determinant the issue that considerably affects accident severity in traffic system. The author in presents a random forest & rough pure mathematics to spot the factors considerably influencing single vehicle crash severity.

Liping et al.,"Traffic Incident Duration Prediction Based on Artificial Neural Network" 2010 International Conference on Intelligent Computation Technology and Automation. They

predict traffic accident period of incident and driver system. The author in used numerous data processing techniques and tells the random forest outperforms than different classification algorithms. In paper , author talks concerning the importance of knowledge mining classification algorithmic rule in predicting the issue that influences road traffic accident.

Tanagra–a Free Data Mining Software for Teaching and Research, Available at: http://eric.univlyon2.fr/~ricco/tanagra/en/tanagra. html, (Accessed 20 June 2014). The author in accustomed explore the attainable application of knowledge mining technology for developing a classification model and therefore the result shows that developed model might classify accidents inside an affordable accuracy. it's necessary to analyse these datasets to extracts helpful data. data processing is a good tool for analysing information to extract helpful data The severity of injuries measured for crash records has each continuous and categorical characteristics. Hence several previous studies have used models with ordered structure to analyse risk issue and their impact on severity of injuries sustained in traffic crashes.

Table 1. Comparison of different types of techniques for road accidents

Author	Objective	Data Mining Techniques	Influential Factors
Chaozhong et.al (2009)	To Identify The Factors Significantly Influencing Single Vehicle Crash Severity.	Random Forest Rough Set Theory	Weather, ordinance, Lightingconditions, Collision Factors, Gender, Age, Experience, life belt, Vehicle sort, Severity Of Svc
Ali et.al (2010)	To Identify Most Important Factors Which Affect Injury Severity	Classification & Regression Tree	Injury Severity, Gender, Age, Seat Belt, explanation for Crash, Collision sort, Vehicle sort, Location sort, Lighting Conditions, climatic conditions, paved surface Condition, Occurrence, Shoulder sort, Shoulder dimension
Liping et.al (2010)	To Predict Traffic Accident Duration Of Incident And Driver Information System	Artificial Neural Networks	No. Of Trucks concerned, change Vehicle, Facility injury, Degree Of traffic congestion, No. Of Fatalities, No. Of Severe Personal Injuries, Road Pollution Involvement, Hazard Material, hearth Involvement, Police Involvement, Patrol Vehicle Involvement, fire truck Involvement
S.krishnaveni, Dr. M.hemalatha (2011)	To Predict Severity Of Injury Using Data Mining Techniques &	Naïve Bayes Adaboost M1 Meta Classifier Part J48,	Casualty, human death, Slight Accident, Killed relation, Serious Injury, Slight Injury, Road Users,

	Compare Algorithm Performance.	Random Forest	Vehicles concerned
Dipot.akomolafe, Akinbola olutayo (2012)	To Predict Causes Of Accidents And Accident Prone Locations	Decision Tree: Id3, Functional Tree	Vehicle sort, Time Of The Day, Season, Causes
S.Shanthi, R.Geetha Ramani (2012)	Significance Of Data Mining Classification Algorithms In Predicting The Factors Which Influence Road Traffic Accident.	Classfication Techniques: C4.5,Id3, Cs-Crt,Cr-T,Cs Mc4, Naive Bayes, Random Forest	Key Value, State, County, Month, Date, Time, Day, Harmful Event, Manner Of Collision, Person sort, Seating Position, Age, Gender, Injury Severity, Air Bag, Protection System, Ejection, Ejection Path, Year_ oftentimes Death, Month_ oftentimes Death, Alcohol check, Drug Test, Drug Involvement, Accident Location, connected Factors
Tibebe et.al (2013)	To Explore The Possible Application Of Data Mining Technology For Developing A Classification Model	Classification & Regression Tree	Accident_Id, Driver_Age, Driv_Exp, Vehic_Age, Vehic_Type, Road_Surf_Type, Road_Cond,
Farzaneh Moradkhani et. al (2014)	To Explore The Accidents Data Recorded By The Police Information System, And Discover Patterns And Predict Future Behaviors And Effective Decisions To Be Taken To Reduce Accidents.	Data Mining, Market Basket Analysis, Association Rules, Pattern Mining	Day Of Week,Weather Conditions, Junction Location At Impact,Vehicle sort
Mehdi mansouri1 et. el (2014)	Extracted Along With Extracting A Local Data Model	Clustering, Data Mining, Decision Tree	System Of The Faulty Vehicle , Airbag , ABS Brake,Gender Male/Female,Age
Sachin kumar et.al (2015)	To Identify The Main Factors Associated With A Road And Traffic Accident.	Clustering And Data Preprocessing	Number Of Injury,Age,Gender, Time Of Day,Month, Road Type, Accident Severity, sort of Accident, Accident Severity
Ayushi Jain et. al (2016)	To Find The Accident Prone Areas In The Country With Respect To Different Accident-Factors	Classification,Clustering And Data Preprocessing	Overloading,Overcrowding, Accident Rate
Ramya V et. al (2016)	Provides Details About How Road Accidents And Traffic Data Can Be Analysed And Used To Predict The Probability Of An Accident To Occur	Data Preprocessing	Accident Severity, Road Type,Date, Junction management, lightweight Conditions, Pedestrian, Hit And Run

III.   PROPOSED METHODOLOGY

3.2 Objectives of proposed methodology

1. To apply pre-processing and filtering on the road accidents raw data and convert into formatted dataset.
2. To apply K Mode clustering and Rule mining algorithm on road accidents dataset,

- Using K mode algorithm for clustering on the formatted road accidents data.
- Improved Apriori algorithm on clustered data for mining the association rules from the clusters of road accidents data.

3. To compare and analyze the results of proposed technique with K mode clustering and Apriori algorithm on the basis of following parameters:

- Clustering time
- Accuracy Rate(AR)
- Association rule mining time

3.2 Phases of proposed methodology

The main improvement of our algorithmic program is to optimize the frequent single things and people things co-occurrence with them.The data structure Bitable is additionally used horizontally and vertically to calculate the token array and count supports, severally. token array and also the corresponding computing technique area unit planned. By computing the token, those item sets that co-occurrence with representative item is known quickly. The frequent item sets, as well as representative item and having constant support as representative item, is known directly by connecting the representative item with all the combos of things in its subsume token therefore, the value for process this sort of item sets is lowered , and also the potency is improved.
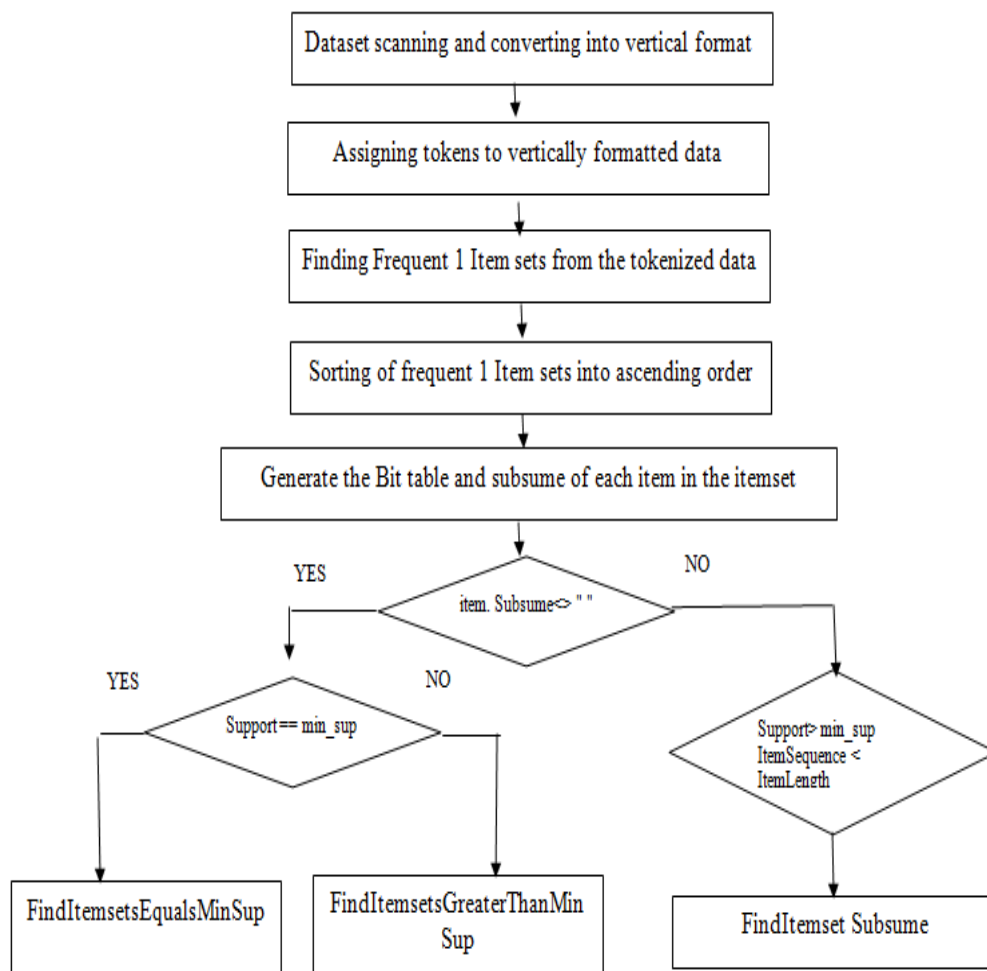
Fig.2 Flow chart of research methodology

Flowchart of Proposed Technique-The flowchart description are as follows:

A. **Read Road data Accident**- In this we take the image of objects from database.The input image is used to acquire an image from the databases.

B. **Pre-processing**- To analyze the info, we have a tendency to develop a framework as shown in Fig. 1. The elaborate description of the framework is as follows:
Data preprocessing
- Data preprocessing is one amongst the necessary tasks in data processing. information preprocessing principally deals with removing noise, handle missing values, removing unsuitable attributes so as to create the info prepared for the analysis. during this step, our aim is to preprocess the accident information so as to create it applicable for the analysis.
- Clustering algorithm There area unit many clump algorithms exist within the literature. the target of clustering algorithmic program is to divide the info into totally clusters or teams specified the objects inside a bunch area unit like whereas objects in other clusters area unit different from one another

C. **Read Road data Accident**- In this we take the image of objects from database.The input image is used to acquire an image from the databases.
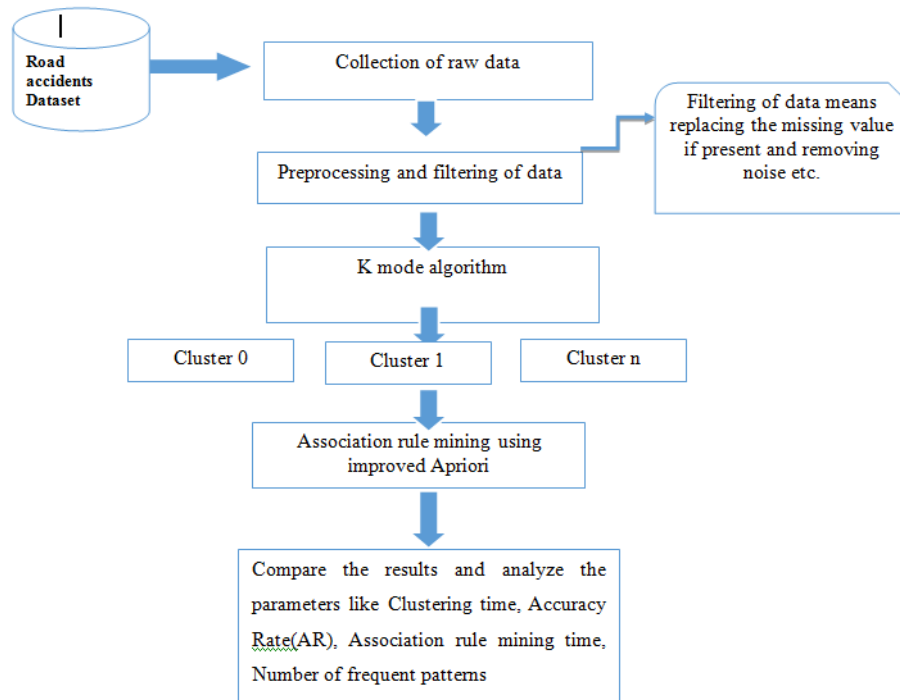

Figue-3 Block Diagram of Proposed System

D. **Read Road data Accident**- In this we take the image of objects from database.The input image is used to acquire an image from the databases.

E. **Pre-processing**- To analyze the data, we develop a framework as shown in Fig. 1. The detailed description of the framework is as follows:
Data preprocessing
- Data preprocessing [14] is one of the important tasks in data mining. Data preprocessing mainly deals with removing noise, handle missing values, removing irrelevant attributes in order to make the data ready for the analysis. In this step, our aim is to preprocess the accident data in order to make it appropriate for the analysis.
- Clustering algorithm There are several clustering algorithms [14, 18] exist in the literature. The objective of
clustering algorithm is to divide the data into different clusters or groups such that the objects

within a group are similar to each other whereas objects in other clusters are different from each other [19].

F. **K-Mode Clustering-**The K-modes clustering technique is an enhanced version of traditional k means. algorithm.The major extensions to the k means algorithm to k modes algorithm is the distance measure and the clustering process which are explained below:

- **Distance measure** Given a data set D, the distance between two objects X and Y, where X and Y are described by N categorical variables, can be computed as follows:

$$d(X, Y) = \sum_{i=1}^{N} \delta(Xi, Yi)$$

where,

$$\delta(Xi, Yi) = \begin{cases} 0, & Xi = Yi \\ 1, & Xi \neq Yi \end{cases}$$

In the on top of equations, Xi and Yi area unit the attribute i values in object X and Y. This distance live is commonly referred as easy matching unsimilarity live. The a lot of the quantity of variations in categorical values of X and Y a lot of the various 2 objects area unit.

K-mode clump procedure:
In order to cluster the info set D into k cluster, K-modes clump algorithmic program perform the subsequent steps:
1. AB initio choose k random objects as cluster centers or modes.
2. Notice the gap between each object and also the cluster centre exploitation distance live outlined in combining weight. 1.
3. Assign every object to the cluster whose distance with the thing is minimum.
4. Choose a brand new center or mode for each cluster and compare it with the previous worth of centre or mode; if the values area unit totally different, continue with step a pair of.In K-mode, clustering can be done on the bases of:-
Casualty: A person killed or injured in an accident, there may be more than one
Slight accident; In traffic accident all persons involved either not detained in hospitals
Serious accident; In traffic accident, one or more persons injured and detained in hospital
Serious injury; An injury for which a person is detained in hospital as an 'in-patient' for more than twelve hours and the injuries causing death 30 or more days after the accident are also included in this category

Vehicles involved; Vehicles whose drivers or passengers are injured, which hit a pedestrian.

G. **Association Rule Mining(Improved Apriori)-** Association rule mining could be a very fashionable data processing technique that extracts attention-grabbing and hidden relations between varied attributes in an exceedingly massive information set. Association rule mining produces a collection of rules that outline the underlying patterns within the information set.The associativity of 2 characteristics of accident is decided by the frequency of their incidence along within the information set. A rule A → B indicates that if A happens then B also will occur.Use of bit wise data structure for storing the support count.And assigning the Id's or token to each data item in dataset. Processing data through Id's and bitset will reduce lot of time.

IV. CONCLUSION

In this paper, we tend to projected a framework for analyzing accident patterns for various types of accidents on the road that makes use of K modes agglomeration and association rule mining algorithmic program to boost the apriori algorithmic program.

V. REFERENCES

1. D. Khera, W. Singh, "A Review on Injury Severity in traffic System using various Data Mining Techniques", International Journal of Computer Applications, vol.100-no.3, pp 0975-8887, 2014.
2. P. Verma, D. Kumar, "Association Rule Mining Algorithm's Variant Analysis", International Journal of Computer Applications, vol.78-no.14, pp 0975- 8887, 2013.
3. S. Kumar, D. Toshniwal, "A Data Mining framework to analyze road accident data", Journal of Big Data, Springer, vol.2-no.1, pp 1-18, 2015.
4. A. Jain, G. Ahuja, Anuranjana, D. Mehrotra, "Data Mining Approach to Analyse the Road Accidents in India", 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Sep. 7-9, 2016.
5. A. T. Kashani, A. Shariat-Mohaymany, A. Ranjbari, "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", Promet Traffic & Transportation, Vol. 23, No. 1, pp11-17, 2011.

6. B. Depaire, G. Wets and K. Vanhoof, "Traffic accident segmentation by means of latent class clustering, accident analysis and prevention", vol. 40,Elsevier,2008.

7. https://www.datascience.com/blog/introduction-to-k-means-clustering-algorithm-learn-data-science-tutorials, accessed on 24/1/2017

8. V. Vijayalakshmi, A. Pethalakshmi, "Mining of Frequent Itemsets with an Enhanced Apriori Algorithm", International Journal of Computer Applications , Vol. 81 – No.4, pp 0975 – 8887, November 2013.

9. http://sites.ndtv.com/roadsafety/important-feature-to-you-in-your-car-5/ accessed on 25/1/2017.

10. Ali et al.," A Data Mining Approach to identify key factors of traffic injury severity" Traffic & Transportation, Vol. 23, 2011, No. 1, 11-17.

11. Bouckaert Remco, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, and AlexSeewald, 2008.WEKA Manual for Version 3-6-0. University of Waikato, New Zealand.

12. Brijesh Kumar Baradwaj, Saurabh Pal," Mining Educational Data to Analyze Students Performance" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.

13. Chaozhong et.al.,"Severity Analyses of Single-Vehicle Crashes Based on Rough Set theory" 2009 International Conference on Computational Intelligence and Natural Computing.

14. DipoT.Akomolafe, Akinbola Olutayo," Using Data Mining Technique to Predict Cause ofAccident and Accident Prone Locations on Highways", American Journal of Database Theory and Application 2012, 1(3): 26-38.

15. Han, Jiawei and Kamber, Micheline. (2006)," Data Mining: concepts and Techniques. San Fransisco", Morgan kufman Publishers.

16. Hand, D.J., Mannila, H., and Smyth, P. (2001) ,"Principles of Data Mining" ,MIT Press.

17. Liping et al.," Traffic Incident Duration Prediction Based on Artificial Neural Network" 2010 International Conference on Intelligent Computation Technology and Automation.

18. Mehmed Kantardzic(2003),"Data Mining: Concepts, Models, Methods, and Algorithms" ISBN13: 9780471228523, John Wiley & Sons Publisher.

19. Pasko Konjevoda and Nikola Stambuk, "Open-Source Tools for Data Mining in Social Science," Theoretical and Methodological Approaches to Social Sciences and Knowledge Management, pp.163-176.

20. S.Krishnaveni, Dr. M.Hemalatha," A Perspective Analysis of Traffic Accident using Data Mining Techniques", International Journal of Computer Applications (0975 – 8887) Volume 23– No.7, June 2011.

21. S.Shanthi, R.Geetha Ramani " Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques" Proceedings of the World Congress on Engineering and Computer Science 2012 Vol. I WCECS 2012, October 24-26, 2012, San Francisco, USA.

22. Tanagra – a Free Data Mining Software for Teaching and Research, Available at: http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html, (Accessed 20 June 2014).

23. Tibebe Shah, Shawndra Hill (2013)" Mining Road Traffic Accident Data to Improve Safety Safety: Role of Road- related Factors on Accident Severity in Ethiopia".

24. Poch M, Mannering F. Negative binomial analysis of intersection-accident frequencies. J Transp Eng. 1996;122.

25. Abdel-Aty MA, Radwan AE. Modeling traffic accident occurrence and involvement. Accid Anal Prev Elsevier. 2000;32.

26. Joshua SC, Garber NJ. Estimating truck accident rate and involvements using linear and poisson regression models. Transp Plan Technol. 1990;15.;33:387–410.