Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Density-based clustering for road accident data analysis

Abdullah S. Alotaibi *

Computer Science Department, Shaqra University, Shaqra, Saudi Arabia

ARTICLE INFO

Article history:
Received 9 April 2018
Received in revised form
11 June 2018
Accepted 13 June 2018

Keywords:
Accident analysis
DBSCAN
Road accident dataset
FP growth
Weka

ABSTRACT

Now days, Road accidents due to traffic are increasingly being recognized as key issue for transportation agencies as well as common people. A considerable unexpected output of transportation systems is road accidents with injuries and loss of lives. In order to suggest safe driving, precise study of road traffic data is serious to discover elements that are related to mortal accidents. In this research paper, we discover factors behind road traffic accidents problem solving by data mining algorithms together with DBSCAN and Parallel Frequent mining algorithm. We initially divide the accident places into k clusters depends on their accident frequency with DBSCAN algorithm. Next, parallel frequent mining algorithm is apply on these clusters to disclose the association between dissimilar attributes in the traffic accident data for realize the features of these places and analyzing in advance them to spot different factors that affect the road accidents in different locations. The main objective of accident data is to recognize the key issues in the area of road safety. The efficiency of prevention accidents based on consistency of the composed and predictable road accident data using with appropriate methods. Road accident dataset is used and implementation is carried by using Weka tool. The outcomes expose that the combination of DBSCAN and parallel frequent mining explores the accidents data with patterns and expect future attitude and efficient accord to be taken to decrease accidents.

1. Introduction

The Road traffic accidents are a key issue concern for transportation leading traders as well as common people. Road accidents are damage the public life with multi-level of injuries (Kumar and Toshniwal, 2015). The number of factors that influence these incidents like Environmental conditions, motorway design, and type of accident, driver characteristics, and vehicle attributes (Karlaftis and Tarko, 1998). The key objective of accident data analysis recognizes the major parameters associated to road traffic accidents (Savolainen et al., 2011). However, various natures of accident data generate the task analysis is tough context. The key problem in this accident data analysis disturbs the human life. Thus heterogeneity have to be measured during data analysis (Depaire et al., 2008), a few correlation between the data may remain out of sight. Although, researchers used partition of the data to decrease this heterogeneity using few measures such as professional knowledge, but there is no security that will guide to a best possible partition which consists of similar type of clusters of road accidents. Data partition has been used broadly to overcome this dissimilarity of the accident data (Kumar and Toshniwal, 2015).

In order to provide safe driving instructions, cautious road traffic of statistics is critical to discover variables that are related to mortal accidents. Data analysis has the ability to recognize the various logics behind road accidents (Ma and Kockelman, 2006). In this paper, we are building data mining methods to make out high-frequency accident places and additional data to identify the different factors that influence road accidents at different locations. We initially split the accident places into m clusters with the support of accident frequency via DBSCAN clustering algorithm (Jones et al., 1991). The frequent pattern mining algorithm is imposed on these for expose the connection between dissimilar attributes of accident data with dissimilar places. Hence, our major accent will be the understanding of the results. The key idea of this research inspect the responsibility of human, vehicle, and infrastructure-with correlated factors in accident sternness by applying data mining learning techniques on road accident information (Miaou and

* Corresponding Author.
Email Address: a.shawan@su.edu.sa
https://doi.org/10.21833/ijaas.2018.08.014

Lum, 1993). Fig. 1 shows proposed system architecture.

Road accidents are major issue of fatality and disability across the world. Road accident can be considered as an event in which a vehicle bumps with other vehicle, person or other objects. A road accident not only provides property damage but it may lead to partial or full disability and sometimes can be fatal for human being. Increasing of road accidents is not a fine sign for the safety transportation. The only solution requires for accident data analysis to know diverse reasons of road accidents corresponding preventive measures taken. Different research studies used various techniques to examine road accident data by data mining methods and offer fruitful outcomes. Different other research use data mining methods to simplify road accident data and data mining methods are novel and superior to classical statistical techniques. Although, both the techniques offer good results that surely helpful for traffic accident prediction, expose that different road accident information exists and should be detached prior to the analysis of data. They also suggested that use of suitable clustering techniques to the analysis of accident data reduces the accidents and reveal the hidden data.

TRW (2014) presented in its account every year; there are 0.4 million accidents in India, this is huge accident rate. This statement shows that there is a negative tendency of accidents from 2012 to 2013; however, as accidents are erratic and can take place in any type of conditions, there is no security that this leaning will assist in future also. Kononov and Janson (2002) declared that efficient connection between accident frequency and additional variables such as geometry of road, road side situations, traffic in sequence and vehicle position can assist to increase effective accident avoidance.

Lee et al. (2002) presented that statistical design were a fine choice for analyze road accidents with geometric factors. Chen and Jovanis (2000) stated that analyzing huge dimensional datasets using classic statistical methods may outcome in certain issues such as inadequate data in great contingency tables. Statistical models have own design specific assumptions these can lead to few erroneous outcomes. Due to these drawbacks of statistical methods, data mining methods are being used to examine road accidents. Data mining is a combination of methods to extract new and unseen information from huge datasets. Barai (2003) discussed different ways of data mining in transportation like road accident data analysis. Several data mining methods like clustering algorithms, classification and association rule mining are broadly used for road accidents data analysis (Tan, 2006).

The rest of paper is prepared as follows: Section 2 states brief description road accident problems are analyzed in facts. In section 3, propose road accident data framework. Section 4 presents a comparison on road accident analysis techniques. Conclusion of the study is presented in section 5.

2. Analysis of road accidents

2.1. Reasons for road accidents

Different reasons for road accidents are: 1. Road Users - lack of care, High speed rash driving, abuse of traffic rules, sleep, fatigue and alcohol etc. 2. Vehicle - Defects such as brakes failure, steering system of vehicle, tire burst, and lighting system. 3. Road Condition - Skidding surface of roads, pot holes on roads. 4. Road design - imperfect geometric design of roads, insufficient breadth of roads, awkward curve design, improper traffic maintenance and poor lighting. 5. Environmental factors -critical weather conditions like smoke, snow, mist and heavy rainfall which bound the normal visibility and makes driving is not safe (https://www.coursehero.com).

Figs. 2 and 3 display road accidents in India, 2017. It clearly shows that Road user's mistakes are the most important factor accountable for accidents. Drivers fault for 79% of total accidents in 2017. Within in the type of drivers fault, accidents are exceeding lawful speed accounted for a large share of 55.6%. The environmental conditions and road design problems appears to be trivial; they account only 1.9% and 2.6% of total accidents. The reason of accidents due to defects in road condition and motor vehicle condition is negligible comparison with drivers fault. They accounted only 1.8% and 1.4% of total road accidents.

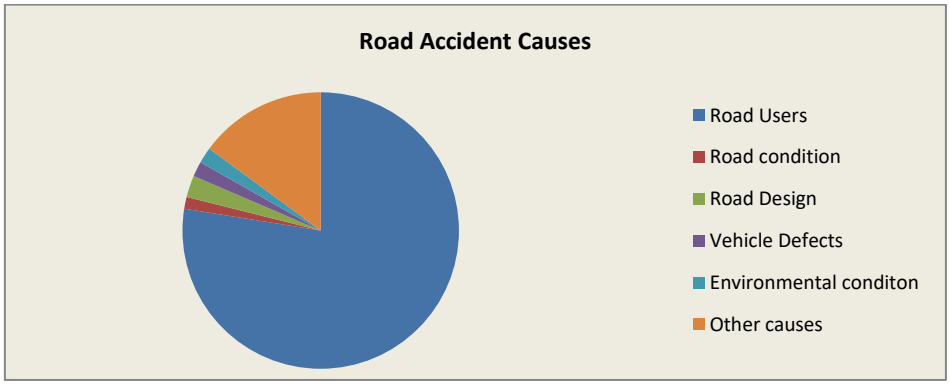

Fig. 1: 2017 road accidents causes

2.2. Road accidental injuries and deaths

There has been advance in road accidental deaths in India over the last few years. Road accidental deaths have increased 9 times, from 14,500 in 1970 to 138,400 in 2017. In comparison to 2005, injuries in 2017 are superior by 53,000 and 87,000, respectively. Table 1 represents all about information. From 2005 to 2013, fatalities have increased of 5% rate per year while the population of the country has larger than before at rate of 1.4% per year. Consequently, road accidental deaths per one lack people, has greater than before 8.9 in 2005 and 11.2 in 2017. In India fatality risk is very high level compared with developed countries. This type of risk in India is high than in the United Kingdom, Sweden. Road accidental deaths occurred due to one lack vehicles, as of 87.5 in 1970 to 8.6 in 2017, it is still quite high compare with developed countries.

Table 1: 1970-2017 road accident statistical data

Year	Road accidents	Road accidental deaths	Accident risk	Road accidental injuries	Fatality risk	Fatality rate	Accident severity index
1970	114.1	14.5	21.6	70.1	2.7	87.5	12.7
1980	153.2	24.6	23.1	109.1	3.7	54.4	16.1
1990	282.6	54.1	34.4	244.1	6.6	28.2	19.1
2000	308.3	80.1	30.8	340.2	8.0	16.6	26.0
2005	336.4	84.4	31.5	382.9	7.9	12.6	25.1
2010	430.6	133.9	36.3	470.6	11.3	10.5	31.1
2013	443.0	137.4	36.1	469.9	11.2	8.6	31.0
2015	446.1	138.1	36.0	472.1	11.4	9.1	32.1
2017	451.6	138.4	36.3	471.4	11.1	8.7	32.5

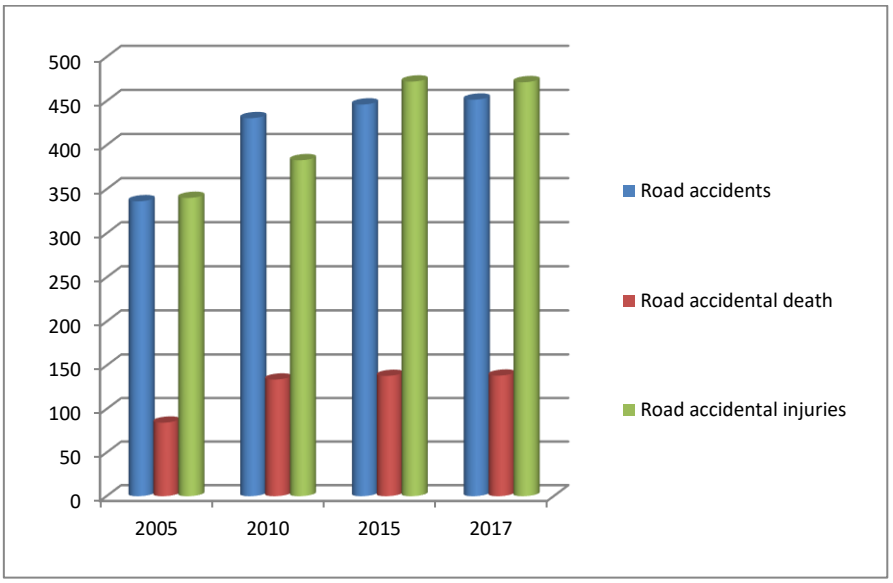

Fig. 2: Road accidents in India, 2017

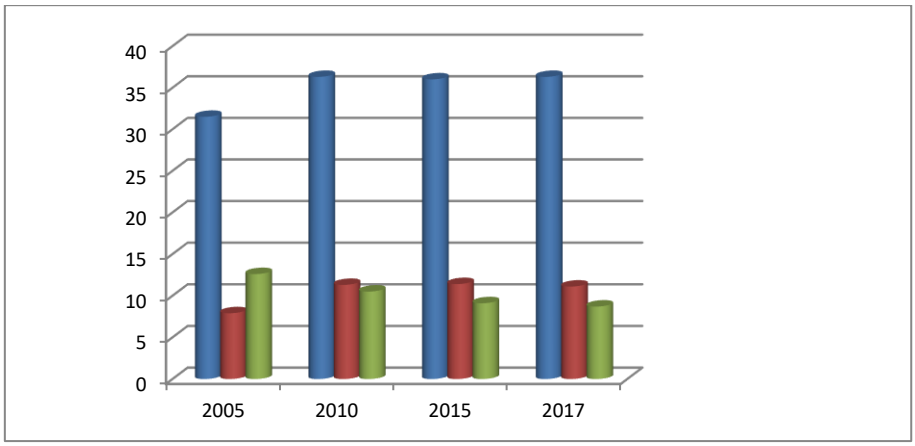

Fig. 3: Graphical representation of 1970-2017 road accident statistical data

2.3. Road accidental distribution based on age, time and sex

The road accident distribution based on age (statistics are not mentioned) is clearly shows that the most creative age group, 25-40 years, is the flat to road accident fatality in India. Age group of 20-42 years comprises 23% of Indian population, faces roughly 37% of total road accidents. During the previous 10 years from 2005 to 2017, number of

fatalities faced by this age cluster has also improved significantly. The middle age (30-40) group 12% of the total population, but fatality faces 21%. So age group 30-59 years, the inexpensively energetic age group, is the most susceptible population cluster in India. Half of the road accidents are faced by this group of people which counts for less than 1/3 of the entire population. Sex wise allocation of injuries and road accidental deaths in India for the year 2005 and 2017 presents that the males for 85.2% of all fatalities 81.1% of injuries in 2017. Past 10 years, total number of fatalities by males has improved by 68.3%.

3. Materials and methods

Data pre- processing is the primary step for remove noise from given input. In second phase attribute selection done by DBSCAN algorithm. parallel frequent mining algorithm is apply on these clusters to disclose the association between dissimilar attributes in traffic accident data for realize the features of these places and analyzing in advance those to spot different factors that affect the road accidents. Finally visualize the patterns of performance evaluation as shown in the Fig. 4.

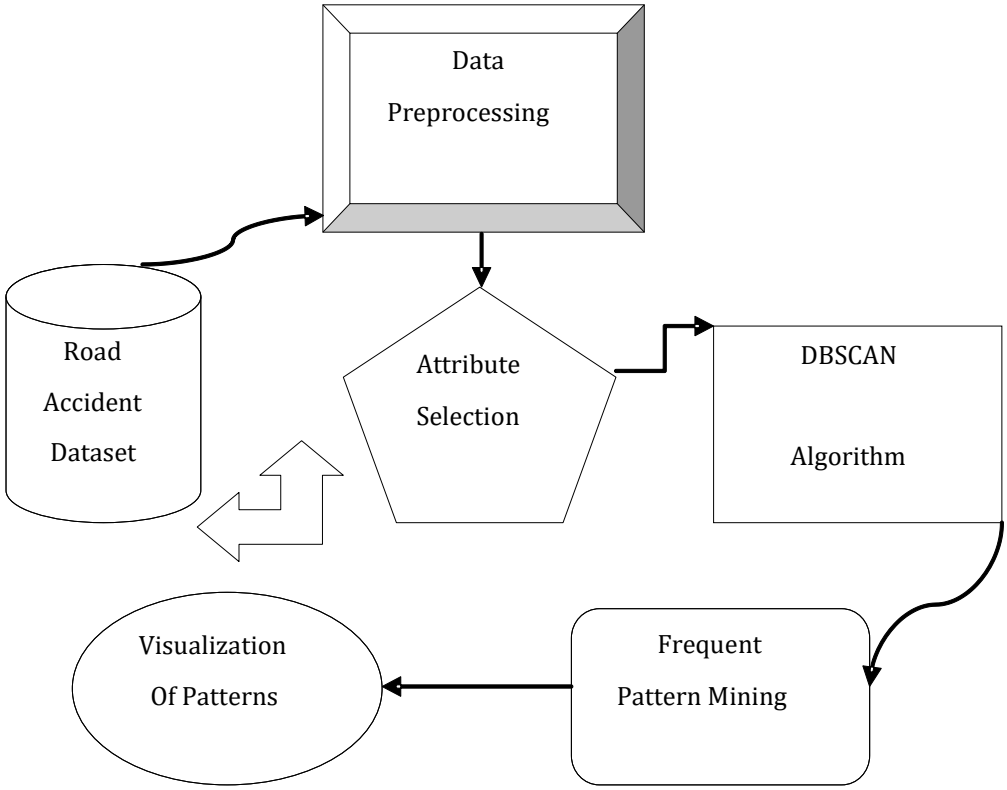

Fig. 4: Proposed system architecture

Cluster analysis split the data components into different groups in a way that maximizes the homogeneity of components within the different clusters. This technique is known as an unsupervised learning algorithm as the accurate number of clusters and their shapes are unknown. Generally, cluster analysis is a procedure of repetitively maximizing the intra cluster components. These similarity-based clustering methods calculate similarity using a specific distance function and measures for components with qualitative. The well-known among similarity-based method is density-based approach (Madhulatha, 2012).

3.1. Density-based road accident analysis

DBSCAN is a density-based clustering algorithm and it is designed to overcome large data sets with noise and is capable of determining different sizes and shapes. Density-based means that cluster are connected points where the density of points is equal to or more than a threshold. If the density is less than the threshold, the data are considered as noise. When a data set is given, DBSCAN divides it into segments of clusters and a set of noise points (https://algorithmicthoughts.wordpress.com). The density threshold condition is that there should be at least MinPts number of points in ε-neighborhood. Clusters contain core points and boundary points. A core point is a point that meets the density condition, and a boundary point is a point that does not meet the density condition but is close enough to one or more core point's ε-neighborhood. Points that are not core points or boundary points are considered as noise. Below is the pseudo code, prepared as functions for road accident data analysis. The function of regionQuery ( ) proceeds the points within the n-dimensional sphere. The function expandCluster ( ) returns for every points in the sphere, the DBSCAN algorithm is presented below in Fig. 5.

Density-Based Road Accident analysis

Aim: *Road accidents analysis using data mining algorithm called DBSCAN*
Input: *D, data set of road accidents; K, no of clusters; M, mean for each cluster*
Algorithm *DBSCAN (D, epsilon, min_points):*

```
C = 0
for each unvisited point P in dataset
        mark P as visited
        sphere_points = regionQuery(P, epsilon)
 if sizeof(sphere_points) < min_points
        ignore P
 else
       C = next cluster
 expandCluster(P, sphere_points, C, epsilon, min_points)
 expandCluster(P, sphere_points, C, epsilon, min_points):
       add P to cluster C
 for each point P' in sphere_points
    if P' is not visited
        mark P' as visited
        sphere_points' = regionQuery(P', epsilon)
    if sizeof(sphere_points') >= min_points
        sphere_points = sphere_points joined with sphere_points'
 if P' is not yet member of any cluster
        add P' to cluster C
 regionQuery (P, epsilon):
return all points within the n-dimensional sphere centered at P with radius epsilon (including P).
```

Output: k cluster groups

Fig. 5: Density based road accident analysis algorithm

The idea behind DBSCAN and its developments is the notion that points are assigned to the similar group if they are density-reachable from every other cluster. To know this model, we will go through the definitions used in DBSCAN and associated algorithms. Clustering starts with dataset E containing a set of point's p ∈ E. DBSCAN estimates the density around a point using the concept of ϵ-neighborhood (http://technodocbox.com).

1. ϵ -Neighborhood. The ϵ -neighborhood, N ϵ(a), of a data point p is the set of points within a specified radius ϵ around p.

$$M\epsilon(a) = \{ b \mid d(a,b) < \epsilon \}$$

where d is some distance measure and $\epsilon \in R^{+\cdot}$ Note that the point p is always in its own ϵ -neighborhood, i.e., a ∈ M ϵ(a) always holds. Following this definition, the size of the neighborhood |M ϵ (a)| can be seen as not normalized kernel density estimate around p using a uniform kernel and a bandwidth of ϵ. DBSCAN uses minPts, detect dense areas for classify the points in a dataset into core, border, or noise points of the cluster.

2. Point classes, A point a ∈ E is classified as a center point if M ϵ(a) has large density, i.e., |M ϵ(a)| minPts where minPts ∈ Z^+ is a user defined density threshold,
3. Directly density-reachable, A point b ∈ E is density-reachable from a point a ∈ E with respect to ϵ and minPts if, and only if,

i. |M ϵ(a)| ≥ minPts,

and

ii. b ∈ M ϵ(a).

That is, p is a core point and q is in its ϵ -neighborhood.

4. Density-reachable, A point p is reach to density from q if their exist in E in sequence of points (a_1, a_2, ..., a_n) with b = a_1 and a = a_n such that a_i+1 directly density reach from a_i ∀ i 2 {1, 2, ..., n − 1}.
5. Connected density, A point a ∈ E is connected density to a point b ∈ E if there is a point o ∈ E, both a and b is density-reachable from o.

In the DBSCAN algorithm, core points of the same cluster, self-governing of the sequence in which the points in the dataset are computed. It is dissimilar for all border points in a cluster. Border points might be density-reachable from core points in clusters and the algorithm assigns them to the primary of these clusters computed which depends on assemble of the data points and the execution of the algorithm (http://technodocbox.com).

3.2. Parallel frequent association mining

Association rule mining is an extremely popular data mining method that extracts attractive and hidden relations between dissimilar attributes in a huge dataset. Association rule mining generates different rules that illustrate the underlying patterns in the dataset. The FP-growth algorithm using for the issue of discovery frequent patterns recursively add the suffix. This algorithm uses minimum frequent items as a suffix; it is well selection for the process reduce the search cost and extracts the frequent

patterns (Pandya and Rustom, 2017). The FP growth algorithm is shown in Fig. 6.

Frequent Pattern Mining Algorithm

Algorithm *FP- growth (FPT, S, P)*
// FPT – Tree on Frequent Items
// S-Minimum Support and P-Current Item set Suffix.
Begin
1. *If FPT is a single path do*
2. *For every C of nodes in path do*
   a. *Inform all patterns C ∪ P;*
   Else
   b. *For every item i in FPT do*
   Begin
      i. *Produce pattern Pi = set i ∪ P;*
      ii. *Inform pattern Pi as frequent;*
   End
3. *Use pointer to extract condition prefix paths for item one;*
4. *Construct conditional Frequent Pattern Tree FPTi from condition*
5. *From prefix paths after eliminating infrequent items;*
6. *If ($FPT_i \neq \emptyset$) FP- growth (FPT_i, P_i, S)*
   End
7. *End*

Fig. 6: Frequent pattern mining road accident data analysis

Using Bayes rule, we can find the probability of label given the observation of a frequent pattern FP_i as:

$$P(K/FP_i) = \frac{P\left(\frac{FP_i}{K}\right) * P(K)}{P(FP_i)} \quad (1)$$

P(K) is the probability of the label which is assumed constant, given by NK/T where NK is the number of images of the class K, and T is the total number of images across all classes (www.cse.iitm.ac.in).

We guess equal number of training data items for all classes i.e., $NK_i = NK_j$. the above assumption P (FP_i | K) can be rewritten as N_{FP}^{K} / N_K. The probability of observing a frequent pattern P (FP_i) is NFP_i / T i.e., the number of data items on which FP_i fired regardless of the label, separated by the total number of images (www.cse.iitm.ac.in).

Substituting all of these in the above equation we have,

$$P(K/FP_i) = \frac{P\left(\frac{FP_i}{K}\right) * P(K)}{P(FP_i)} = N_{FP_i}^{K} X \frac{N^K}{T} X \frac{T}{N_{FP_i}}$$

therefore

$$P(K/FP_i) = \frac{N_{FP_i}^{K}}{N_{FP_i}} \quad (2)$$

The above outcome displays that for testing a label, the operator sets should be ordered according to the ascending order of $\frac{N_{FP_i}^{K}}{N_{FP_i}}$. This is for an operator set to get a better score in this phase, either the frequency of observing the operator set FP_i for the particular label is high or that the probability of the operator set FP_i firing for other classes is less (www.cse.iitm.ac.in).

There are different clustering algorithms exist in the literature. The objective of clustering algorithm is to partition the data into different clusters such that the objects within a group are similar to every other object in other clusters are diverse from each other. DBSCAN clustering method, after that we can use FP growth algorithm of association rule mining for computing the clusters in Fig. 7.

Data pre- processing is the primary step for remove noise from given dataset. Next level attributes selection done by DBSCAN algorithm. It can be constructing as a groups based on attributes. parallel frequent mining algorithm is apply on these clusters to disclose the association between dissimilar attributes in traffic accident data for realize the features of these places and analyzing in advance those to spot different factors that affect the road accidents. Finally visualize the patterns of performance evaluation.

3.3. Dataset description

The road accident dataset consists of 11,574 road accidents from 2012 to 2016 for last 6 years period. After preprocessing of the data, 11 variables were recognized for the research with satisfactory. The dataset comprised of accident features time, type of accident, and number of injured victims age, gender, road type, and area around accident location (Kumar and Toshniwal, 2015). Data add number of people, vehicles involved, road surface, location and weather conditions. The Easting's and Nothings are generated at the roadside where the accident occurred. Attributes of dataset include reference number Northing and Easting number of vehicles, Accident Date, Time (24× 7) 1st road class, Road Surface, Lighting Conditions, Weather Conditions with reference number Grid Ref: Easting Grid Ref: Northing Number of vehicles Accident Date Time 21G0539, 427798, 426248, 5, 16/01/2015, 1205; 21G0539, 427798, 426248, 5, 16/01/2015, 1205; 21G1108, 431142, 430087, 1, 16/01/2015, 1732; 21H0565, 434602, 436699>, 1, 17/01/2015, 930; 21H0638, 434254, 434318, 2, 17/01/2015, 1315; 21H0638, 434254, 434318, 2, 17/01/2015, 1315.

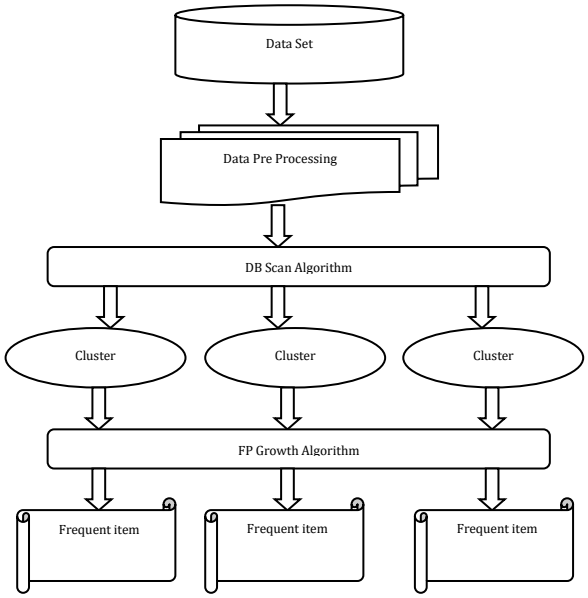

Fig. 7: A Framework of proposed method

Table 2 shows the sample road accident dataset with different parameters of road surface, Lightening conditions, weather conditions, casualty class, sex of casualty, age of casualty and type of vehicle. These parameters most helpful for finding the reasons behind the accidents on roads. In order to suggest safe driving, precise study of road traffic data is serious to discover elements that are related to mortal accidents.

4. Results and discussion

A diversity of data mining methods, algorithms and tools are proposed for road traffic accident data analysis accident location tracking, prediction and identification of different contributory factors that affect the accident cruelty levels. Garib et al. (1997) they have been construct statistical design using stepwise regression analysis method for guessing incident duration.

The result analysis displays that over 85% of differences in occurrence duration can be predicted by the eight factors implicated in the regression model. DBSCAN and Frequent pattern mining algorithms are used for clustering, and the following clusters are constructed.

Cluster 1 represents the traffic clusters in such a way accidents occur because of high traffic. Cluster 2 represents the time of accident cluster in which accidents happen during day and night time. Cluster 3 represents the age of the drivers cluster. Cluster 4 presents the accident occurred every month. Cluster 5 states the weather condition at the time of accident. Cluster 6 is the lightening condition issue on the roads. Cluster 7 describes about type of accident the road condition. Cluster 8 describes the speed limit of vehicles at the time of accident. F-measure is used for cluster analysis because it throughput node-based analysis using the following equations.

Table 2: Sample road accident dataset

Ref. Number	Grid Ref: Easting	Grid Ref: Northing	Number of Vehicles	Expr1	Accident Date	Time (24hr)	Road Surface	Lighting Conditions	Weather Conditions	Casualty Class	Sex of Casualty	Age of Casualty	Type of Vehicle
2181280	418241	442351	2	Leeds 2016	8/1/2016	1905	Dry	Darkness: street lights present and lit	Fine without high winds	Driver or rider	Male	38	Motorcycle > 500cc
2191037	424993	432898	2	Leeds 2016	9/1/2016	1615	Dry	Darkness: street lights present and lit	Fine without high winds	Driver or rider	Female	50	Car
2CQ0870	431159	436397	2	Leeds 2016	15/01/2016	1645	Dry	Daylight: street lights present	Fine without high winds	Driver or rider	Male	26	Car
2CQ0870	431159	436397	2	Leeds 2016	15/01/2016	1645	Dry	Daylight: street lights present	Fine without high winds	Vehicle or pillion passenger	Female	22	Car
3111091	439313	432376	2	Leeds 2016	1/1/2016	956	Wet / Damp	Daylight: street lights present	Fine without high winds	Driver or rider	Male	57	Pedal cycle
3111178	426994	439957	2	Leeds 2016	1/1/2016	1115	Dry	Daylight: street lights present	Fine without high winds	Driver or rider	Male	59	Pedal cycle
3111395	427813	431257	1	Leeds 2016	1/1/2016	1352	Wet / Damp	Daylight: street lights present	Fine without high winds	Driver or rider	Female	53	Car
3111981	431496	432727	2	Leeds 2016	1/1/2016	2015	Wet / Damp	Darkness: street lights present and lit	Raining without high winds	Vehicle or pillion passenger	Female	22	Car
3120560	431880	430498	2	Leeds 2016	2/1/2016	1110	Wet / Damp	Daylight: street lights present	Fine without high winds	Vehicle or pillion passenger	Female	20	Car
3120872	429425	433999	1	Leeds 2016	2/1/2016	1638	Wet / Damp	Darkness: street lights present and lit	Raining without high winds	Driver or rider	Male	22	Car
3120872	429425	433999	1	Leeds 2016	2/1/2016	1638	Wet / Damp	Darkness: street lights present and lit	Raining without high winds	Vehicle or pillion passenger	Female	26	Car
3130442	435325	433760	5	Leeds 2016	3/1/2016	1150	Wet / Damp	Daylight: street lights present	Raining without high winds	Driver or rider	Male	35	Car
3130442	435325	433760	5	Leeds 2016	3/1/2016	1150	Wet / Damp	Daylight: street lights present	Raining without high winds	Vehicle or pillion passenger	Female	21	Car
3130442	435325	433760	5	Leeds 2016	3/1/2016	1150	Wet / Damp	Daylight: street lights present	Raining without high winds	Vehicle or pillion passenger	Male	28	Car

$$Precision = TP/(TP + FP) \quad (3)$$
$$Recall = TP/(TP + FN) \quad (4)$$
$$F - Measure = (1+\propto)/((1 + precision) + (\propto /Recall)\ Where\ \propto = 1 \quad (5)$$

Cluster based analysis findings and road accident dataset analysis are compared. The outcome reveal that the mixture of DBSCAN clustering and frequent pattern mining is extremely inspirational as it generates important data that would remain hidden, if no partition has been performed prior to produce frequent item sets. Weka is data mining software that uses a collection of machine learning algorithms. These algorithms can be applied directly to the data. Table 3 shows data mining algorithms, Comparison for road accident analysis of different methodologies, classifiers and their result.

Fig. 8 shows the graphical representation of Table 3 values. Table 3 statistical results prove the DBSCAN with combination of FP growth generates better results compare to other methods. In this combination of methodology datasets with altering densities are tricky. So they can be working aggressively up to datasets are not alter.

5. Conclusion

Data mining has been verified as a reliable method in analyzing road accident data. So many authors used data mining method for analyzing road accident data of different countries. The data mining methods like association rule mining, clustering and classification are broadly used recognized multiple reasons that affect the serious of road accidents. In this scenario present safe driving suggestions and careful road traffic data analysis is dangerous to discover factors that are strongly related to destructive accidents. In this research, we locate so many factors behind road accidents, these accidents are analysis by using data mining algorithms like DBSCAN and Parallel Frequent mining algorithm. We initially split the accident places into k clusters based on their frequency of accident results by means of DBSCAN algorithm. Next, parallel frequent mining algorithm is exposing the association between dissimilar attributes in accident data, when it is applied on clusters. Understand the features of these places and additionally analyzing them to recognize different factors affect the road accidents at different locations. The major objectives of road accident data are scrutiny to recognize the key issues in the area of road safety. The efficiency of accident avoidance depends considerably on the reliability of composed and estimated data and the appropriateness of the methods. Road accident dataset is used and execution is carried by using Weka tool. The outcomes reveal that dataset for road accident and its analysis using DBSCAN and FP mining algorithm demonstrate that this procedure can be reused on new accident data with extra attributes to recognize different factors connected with road accidents.

Table 3: Data mining algorithm for comparison of road accident analysis

Application	Methodology	Classifiers	Result	Efficiency
Traffic accidents analysis based on road users	K-modes Clustering	Support Vector Machine	75.99 %	Low
A prospective traffic accident analysis	PART algorithm	Random Forest Tree	84.66%	High Low
Classification of vehicle crash structure in road accidents	CS-MC4 algorithm	Naive Bayes Classifier	80.59%	Medium
Traffic accidents in Dubai	Apriori Algorithm	Association rules	78.63%	Low
Detection of key factors for traffic injury harshness	CART Algorithm	Rule Induction	72.49 %	High
predicting cause of accident places on highways	ID3 Algorithm	Decision Tree	77.70%	Low
Traffic incident duration calculation based on ANN	ANN Algorithm	Artificial Neural Network	85.35%	Low
Road accidents in Korea	DT Algorithm	Artificial Neural Network	85.2%	High
A data mining framework for road accident data analysis	K-Modes Algorithm	Association Rule Mining	79.5%	High
Gender-specific classification of road accident patterns	C4.5 Algorithm	Random Tree	83.02%	Low
Imbalanced traffic accidents datasets	Naive Bayesian Algorithm	Bayesian networks classifiers	.78.2%	Low
Identifying accident-prone locations	fuzzy K-NN Algorithm	Bayesian networks classifiers	78.6%	Medium
Accident Data analysis	DBSCAN Algorithm	FP-Growth	97.6%	High

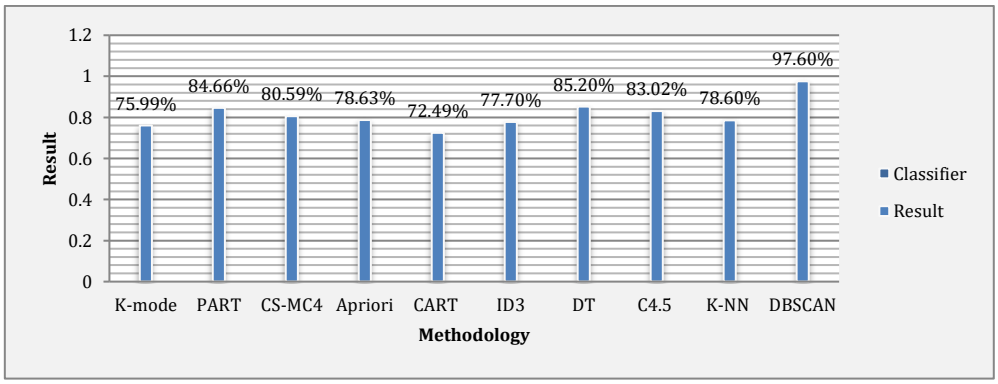

Fig. 8: Assessment of data mining algorithms

References

Barai SK (2003). Data mining applications in transportation engineering. Transport, 18(5): 216-223.

Chen WH and Jovanis P (2000). Method for identifying factors contributing to driver-injury severity in traffic crashes. Transportation Research Record: Journal of the Transportation Research Board, 1717: 1-9.

Depaire B, Wets G, and Vanhoof K (2008). Traffic accident segmentation by means of latent class clustering. Accident Analysis and Prevention, 40(4): 1257-1266.

Garib A, Radwan AE, and Al-Deek H (1997). Estimating magnitude and duration of incident delays. Journal of Transportation Engineering, 123(6): 459-466.

Jones B, Janssen L, and Mannering F (1991). Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention, 23(4): 239-255.

Karlaftis MG and Tarko AP (1998). Heterogeneity considerations in accident modeling. Accident Analysis and Prevention, 30(4): 425-433.

Kononov J and Janson B (2002). Diagnostic methodology for the detection of safety problems at intersections. Transportation Research Record: Journal of the Transportation Research Board, 1784: 51-56.

Kumar S and Toshniwal D (2015). A data mining framework to analyze road accident data. Journal of Big Data, 2(26): 1-18.

Lee C, Saccomanno F, and Hellinga B (2002). Analysis of crash precursors on instrumented freeways. Transportation Research Record: Journal of the Transportation Research Board, 1784: 1-8.

Ma J and Kockelman K (2006). Crash frequency and severity modeling using clustered data from Washington State. In the IEEE Intelligent Transportation Systems Conference, IEEE, Toronto, Canada: 1621-1626.

Madhulatha TS (2012). An overview on clustering methods. IOSR Journal of Engineering, 2(4): 719-725.

Miaou SP and Lum H (1993). Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention, 25(6): 689-709.

Pandya JP and Rustom MD (2017). A survey on association rule mining algorithms used in different application areas. International Journal of Advanced Research in Computer Science, 8(5): 1430-1436.

Savolainen PT, Mannering FL, Lord D, and Quddus MA (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis and Prevention, 43(5): 1666-1676.

Tan PN (2006). Introduction to data mining. Pearson Education India, Bengaluru, India.

TRW (2014). Road accidents in India 2013. Ministry of Road Transport and Highways Transport Research Wing, Government of India, New Delhi, India.