IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Journal Europeen des Systemes Automatises
Vol. 52, No. 1, February, 2019, pp. 57-63
Journal homepage: http://iieta.org/Journals/jesa

Generating Road Accident Prediction Set with Road Accident Data Analysis Using Enhanced Expectation-Maximization Clustering Algorithm and Improved Association Rule Mining

Sakham Nagendra Babu[1*], Jebamalar Tamilselvi[2]

[1] R & D Center, Bharathiar University, Coimbatore 641046, India
[2] Jaya Engineering College, Thiruninravur 602024, Chennai, India

Corresponding Author Email: s.nagendrababu@gmail.com

https://doi.org/10.18280/jesa.520108

Received: 6 November 2018
Accepted: 29 January 2019

Keywords:
road accident, enhanced expectation-maximization, association rules, big data, clustering, accident prediction set

ABSTRACT

Prediction of Road Accidents has gained importance over the years however road accidents may not be stopped but rather can be controlled. Driver feelings, for example, tragic, sad, and anger can be one purpose behind accidents. In the meantime, weather conditions, for example, climate, traffic conditions, sort of road, health of driver, and speed can likewise be the purposes behind accidents. Big data is a term utilized for vast and complex informational collections for handling as the traditional data mining techniques are incomplete for preparing them. In this paper an Enhanced Expectation-Maximization (EEM) Algorithm is utilized which works dependent on the Gaussian dissemination. In the proposed work the entire dataset is divided into different clusters based on vehicle type and again these groups are separated into sub groups dependent on parameter on each vehicle type. Strong Association Rules using Improved Association Rule Mining (IARM) algorithm are designed for every vehicle class and for each parameter. The Congestion control using Machine Framework (CCMF) and Traffic Congestion Analyzer using Map Reduce TCAMP () algorithms are used for training the machine and to apply each and every association rule on the dataset and accurate prediction set is generated.

1. INTRODUCTION

Road Accidents keep on being a noteworthy issue on earth, both from the general security aspects and financial viewpoints. However every year, numerous vehicles are associated with accidents that cause large number of deaths and injuries. Comprehensively about 1.55 million individuals died in road accidents and around 60 million are injured [1]. Road Accidents are positioned as the ninth most purpose of death on the earth, and without new activities to upgrade road security, accidents will probably ascend to the third place constantly by 2020.

Roads get across more traffic than it was really intended to pass on. These outcomes result in expansion in the quantity of vehicles on Roads. This has thusly expanded traffic, vehicle accidents, and so on. In India, Road and Traffic Accidents represent a significant issue [2]. Enormous actions have been taken to enhance Road safety. Conventional strategies can't be utilized in such frameworks as the information produced is in huge volumes, which requests the utilization of machine learning calculations for calculation [3]. The framework of the proposed method is depicted in below Figure 1.

In this paper, an Enhanced EM calculation (EEM) is proposed. To start with, we utilize the dispersion rather than parameter as introductory condition so as to effortlessly control the EM emphasis. Second, the uniform appropriation of parameters is utilized to give the most irregular instatement [4]. That is, the proposed EEM calculation has been formulated dependent on the utilization of posteriori distribution as beginning condition that can accomplish worldwide optimality [5].

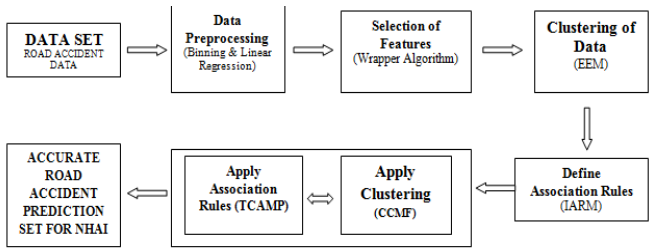

Figure 1. Proposed work architecture

2. LITERATURE SURVEY

As shown by the World Health Organization (WHO), 1.60 million people died consistently on the Roads, and upwards of 60 million are injured. Besides, the Centers for Disease Control and Prevention (CDCP) have announced that Road Accidents cost 100 billion every year.

Zheng et al. [1] used CART and MARS to dissect an epidemiological case-control examination of injuries coming about in view of motor vehicle accidents. They also perceived potential causes of danger, all things considered, brought about by the driver situation [10]. Sarkar et al. [2] used key backslide models to explore the components connected with accidents, and found that highways were more dangerous than normal streets. Sharma et al. [3] used the three data mining techniques for decision trees, neural frameworks and vital backslide methods to discover gigantic segments affecting the Road action.

Williams et al. [4] considered driver obligations using the ID3, J48, and multilayer perception (MLP) calculations to locate the related factors, and found that various segments specifically influence the cause of accidents, for instance, driver's age and experience. Beshah et al. [5] used CART and multinomial determined backslide method to research the parts played by the characteristics of drivers, and found that the CART procedure gave commonly correct results.

The methods include associated neural framework, decision tree classifiers, the Bayesian strategy, key model, and clustering using k-mean calculation. Their test comes about that batching framework was better than various methodologies.

Palamara et al. [6] associated unmistakable approaches and stood out them from discovering Accident seriousness factors. The authors at first used a course of action of theories to analyze data to check whether the data had complete information about the conditions related with the occasion of accidents, and a short time later contemplated these subordinate conditions. Their results showed that accidents came about in view of a mix of components. In addition, rules with high or low range exhibited particular features.

Chen et al. [7] concentrated the association between accident frequencies and the partition of the accidents from the zones of residence. As might have been anticipated, the Accident frequencies were higher closer to the zones of highways, possibly in light of higher presentation.

Verma et al. [8] apply clustering methods in the multi-organization speed relations. Clustering techniques are used to see the failure centers in a speed charts.

Srivastava et al. [9] concentrated on vehicle accidents that happened at signalized crossing points. The accident seriousness was isolated into three classes: no damage, conceivable damage and crippling damage. They thought about the execution of Multi-layered Perceptron (MLP) and Fuzzy ARTMAP, and found that MLP arrangement exactness is higher than Fuzzy ARTMAP.

Ghazizadeh et al. [10] utilized neural systems to investigate vehicle Accident that happened at crossing points. They picked feed-forward MLP utilizing BP learning. The model had 10 input hubs for eight factors (day or night, traffic streams circling in the crossing point, number of virtual accidents, number of real accidents, kind of junction, Accident type, Road surface condition, and climate conditions). The resultant hub was called Accident record, which was determined as the proportion between the quantity of accidents for a given convergence and the quantity of accidents at the most unsafe crossing point.

Stewart et al. [11] utilized genuine information for building up a multi-layered MLP neural system for accident predictions. They looked at the execution of the neural system demonstration and the occurrence recognition display in task on indian roads. Results demonstrated that neural system model could give quicker and increasingly discovering over the model that was in task on Indian roads. They additionally discovered that inability to give speed information at a station could altogether collapse execution inside that segment of the highway.

Abdat et al. [12] proposed a model for assessing Accident seriousness probability molded on the event of an Accident. They found that there is a more noteworthy likelihood of apparent damage or handicapping damage/casualty in respect to no obvious damage if no less than one driver did not exercise a self-control framework at the season of the Accident.

3. PROPOSED METHOD

3.1 Data preprocessing

The data set considered initially need to undergo pre processing for removal of noise values and to remove unwanted data. As the dataset considered contains raw information of all road accident types, the data set contains unwanted data and missing values also[6]. To get only useful data and to remove noise values and to fill missing values pre processing technique is applied on the data set.

In this proposed work, binning and linear regression methods are used for cleaning noisy data to improve the mining results [7].

3.2 Clustering

The proposed Enhanced Expectation-Maximization algorithm is applied on the dataset considered for road accidents and it divides the entire dataset into multiple clusters based on vehicle type and again each cluster is divided into sub groups based on parameters [8]. This paper improves Expectation Maximization (EM) calculation dependent on exceptional trademark properties like record count, parameters used and the vehicle types. EEM calculation is an unsupervised machine learning procedure which portrayed the structure from the covered information [9].

The posteriori distributions is calculated as w= q*h/N where q is the dataset and h is the record considered and N is count of records in dataset. Every component is estimated by a parameter (P_k) and it's probability with every other parameter that goes under different clusters (C_k). The probability of every parameter is calculated as

$$P(P_k \& C_k) = \frac{\sum_k K^2 * P_k}{(\lambda * W * \lambda')}$$

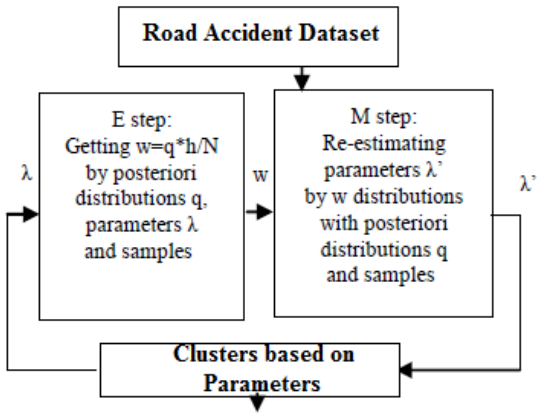

Figure 2. Proposed EEM workflow

The Enhanced Expectation Maximization algorithm is discussed below.

3.3 Algorithm

Enhanced_Expectation_Maximization
Input: Data sets Ds= {ds1,.., dsn}
(1). For Ds = 0 to MAX do
(2). Identify the Similarity measure Set S>= 90% for every vehicle class

(3). Sim (si) = Σvehicleclass ϵ Vtype
(4). Ds(i)=Sim(si)
(5). Repeat
(6). Split the dataset Ds to {Ds1, …., Dsn}, with vehicle class {VC1,…, VCn} ;
(7). Calculate the mean value of a record f(y) − f(z), z <=t <= y, where z is initial record, t is
threshold and y is the max count.
(8). Mean= f'(t)(y − z)
(9). For each cluster of vehicle type
For each parameter of vehicle class
For each record do
Calculate posteriori distributions
exp $\{yt,k,v\}$
Pd =$v$exp $\{yt,,\}$
until Ds(i)=NULL
end for
end for
end for
(10). For every parameter P_i create a cluster C_i
do
$C_i(P) = P_i$;
done
(11). Update cluster C_i.
$V_i(C) = C_i$;

Output: Generates clusters of each vehicle class and parameter Vi.

The above algorithm takes the original dataset as input and then each record similarity set is identified with the given parameters and if same vehicle class is identified, then they are grouped as a cluster [10]. The mean value of each record is identified based on the threshold value that z <=t <= y, where z is initial record, t is threshold and y is the max count. The dataset is divided into cluster based on vehicle class and again the cluster class is divided into sub clusters based on parameter. The posteriori distributions are calculated for every parameter which identifies the parameter to form a cluster.

The Improved Association Rule Mining (IARM) in information mining is a main stream approach that is utilized to break down the offered dataset to find fascinating examples or connections between the different things in the dataset. The idea of solid association rules was first utilized by recognizing the different associations selected between the things that are sold amid a substantial scale exchange database gathered from a grocery store utilizing a point framework. The connection between the things is distinguished in light of the buy design. The IARM strategy produces an arrangement of association rules winning between the different things of the given dataset in view of the quantity of events of these things blend in the dataset.

An association control is utilized to characterize the connection between any two things in the given dataset. Think about three things P, Q and R. The connection {P, Q} → R say that if a man purchases two things P and Q together, at that point he/she will in all likelihood purchase the thing R moreover [11]. That is, the relations between the things are produced by recognizing the different examples inside the dataset. The IARM procedure [3] comprises of two phases as takes after:

(1). Identify the item set that happen regularly in the dataset – The successive item set are those that have a help esteem (sup(item)) equivalent to or more prominent than the base help esteem that is pre-characterized. The help estimation of item set is figured as the quantity of exchanges that contains that thing. In the above illustration support of {P, Q} is figured as what number of exchanges have both P and Q [12].

(2). Association manage age utilizing regular itemset: In this stage the fascinating principles are produced by computing the certainty factor for all the regular itemset that are created in past stage [13]. The certainty esteem for the above case govern of {P, Q} → R will be sup ({P, Q})/sup(R).

In the proposed method strong association rules need to be built by the user in every aspect of vehicle class and road accident parameter [14].

The parameters/attributes considered for road accident prediction are given below.

Attribure Name	Values	Description
Attribure_ID	Integer	Identification of accident
Attribure_Type	Fatal, Injury, Property, damage	Accident type
Driver_Age	<20, [21-27], [28-60]>61	Driver age
Driver_Sex	M, F	Driver sex
Driver_Experience	<1，[2-4], >5	Driver experience
Vehicle_Age	[1-2], [3-4], [5-6] >7	Service year of the vehicle
Vehicle_Type	Car, Trucks, Motorcycles, other	Type of the vehicle
Light_Condition	Daylight, Twilight, Public lighting, Night	Light condition
Weather_Condition	Normal weather, Rain, Fog, Wind, Snow	Weather conditions
Road_Condition	Highway, Ice Road, Collapse Road, Unpaved Road	Road conditions
Road_Geometry	Horizontal, Alignment, Bridge, Tunnel	Road geometry
Road_Age	[1-2], [3-5], [6-10], [11-20] >20	The age of road
Time	[00-6], [6-12], [12-18], [18-00]	Accident time
City	Marrakesh, Casablanca, Rabat…	Name of city where accident occurred.
Particular_Area	School, Market, shops…	Where the accident occurred in school or Market areas
Season	Autumn, Spring, Summer, Winter	Seasons of year
Day	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday	Days of week
Accident_Causes	Alcohol effects, Fatigue, loss of Control, Speed, Pushed by another vehicle	causes of accident
Number_of_injuries	1, [2-5], [6-10] >10	Number of injuries
Number_of_death	1, [2-5], [6-10] >10	Number of deaths
Victim_Age	<1, [1-2], [3-5] >5	Victim Age

Based on the above parameters the association rules are designed for accurate prediction of road accidents. Some examples of designing association rules are

Accident_Type == {Injury}; criterion = 1, statistic = 42.195

Drive_Age == {[21–27], [28–60], <20, <21, >60, >61}; criterion = 0.801, statistic = 18.41

vehicle_Type == {Car}; criterion = 0.862, statistic = 14.079

* weights = 14

vehicle_Type == {Pedestrian, Truck}

* weights = 11

Drive_Age == {[21–25], <22, <23}

* weights = 7

Accident_Type == {Fatal, Property damage}

Drive_Age == {[28–60], <20, <23, <30}; criterion = 0.979, statistic = 23.916

* weights = 26

Drive_Age == {[21–27], <22, >60, >61}

Drive_Exp == {[2–6], <1, <3, >10, >6}; criterion = 0.963, statistic = 21.475

Season == {Summer}; criterion = 0.912, statistic = 9.75

* weights = 7

Season == {Autumn, Winter}

* weights = 7

Drive_Exp == {<2, <5, >8}

* weights = 17

Here for every parameter the data is analyzed and then for every parameter criteria and weights are also given based on analyzing the dataset. After designing the association rules, then these rules are to be applied on each and every parameter for all vehicle types which is a time-consuming process and also does not yield accurate outcomes. so to make the process much easier, big data analytics with machine learning methods are used [15].

Initially the data cluster of a vehicle type is given as trained data to the machine and then for every vehicle class each parameter value is trained to the machine [16]. The machine has to construct a tree based on the data and the vehicle type [17]. The clustering process will be internally done by the machine which accurately develops clusters of relevant parameters.

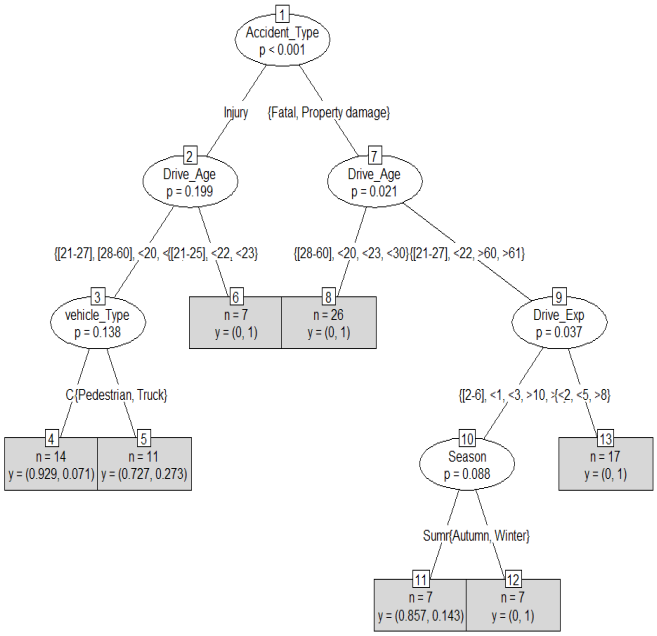

Figure 3. Storage architecture of data

The working of CCMF( ) algorithm is

Algorithm CCMF ()

{

Initially take the training data $t_1,t_2…..t_n$

Give each and every aspect of training data to the machine

After the completion of training go for testing

Give test data to the machine

If any failure

Go back to step 1 and include the failure case to train data

Or test all the test case

And derive the prediction's

}

The tree is constructed based on the CCMF( ) algorithm

Input: Training dataset T and attribures/parameters.P

Output: Multiple clusters based on parameters of vehicle type using decision Tree.

If (count(T) is NULL)

Stop

Else if (count(P) is NULL)

Stop

Else if (|T| OR |P|) is 1

Then only parameter is considered in the dataset and one node is formed a parent node.

Else

For p1 ε P and P ε T

If (p1 ε VTk)

split(T) = p1;

end if

end for

end.

Based on the above process the tree is constructed and then each association rule is applied on the data cluster for prediction of road accident.

The TCAMP( ) algorithm applies each association rule on the parameter set and generate a prediction set for road accident cases.

Algorithm TCAMP ()

{

Input Traffic data set

Output predicting the accidents

Step-1: Take the traffic data set

Step-2: Apply pre-processing

Step-3: After data clusters are formed apply Association rules.

For vehicleclass (V)

For Dc(i) to MAX do

Apply Association rule on parameter P(i).

End for

End for

Step-4: Display Prediction set PS.

}

The above algorithms explains the TCAMP( ) algorithm in which ɵ is the threshold value for every parameter. FS is the feature subset in which the prediction set is stored. Only the relevant feature data is compared and association rules are applied on them which results in prediction set.

4. RESULTS

Here for experimental setup we use a machine with 16GB Ram, 1TB HDD, with Ubuntu 16.04lts and it was with Hadoop and apache mahout installed in it [18]. The attributes considered in the dataset are listed in Table 1 below. The Data

set is considered from https://data.gov.in/dataset-group-name/road-accidents which provides a huge amount of data related to Road accidents of different states in India and on this data set pre-processing is applied for removing or filling of missed data [19]. In this proposed method 121868 records of data for year 2017 is gone through pre-processing [20]. In the proposed approach instead of manually performing the tasks the user after performing clustering operations the dataset is provided to machine which in turns compares with the dataset and prediction is accurate as all the association rules are applied by the machine to the dataset [21].

Table 1. Road accident analysis parameters

Attribure Name	Values	Description
Attribure ID	Integer	Identification of accident
Attribure Type	Fatal, Injury, Property, damage	Accident type
Driver Age	<20, [21-27], [28-60] >61	Driver age
Driver Sex	M, F	Driver sex
Driver Experience	<1, [2-4], >5	Driver experience
Vehicle_Age	[1-2], [3-4], [5-7] >10	Service year of the vehicle
Vehicle_Type	Car, Trucks, Motorcycles, other	Type of the vehicle
Light_Condition	Daylight, Twilight, Public lighting, Night	Light condition
Weather_Condition	Normal weather, Rain, Fog, Wind, Snow	Weather conditions
Road_Condition	Highway, Ice Road, Collapse Road, Unpaved Road	Road conditions
Road Age	[1-2], [3-5], [6-10], [11-20] >20	The age of road
Time	[006], [6-12], [12-18], [18.00]	Accident time
Particular_Area	School, Market, shops…	Where the accidents occurred in school or Market areas.
Season	Autumn, Spring, Summer, Winter	Seasons of year
Accident_Causes	Alcoholeffects, Fatigue, loss of control, Speed, pushed by another vehicle, Brake Failure	Causes of accident
Number_of_death	1, [2-5], [6-10] >10	Number of deaths

Table 2. Dataset considered

Manipur	99	128	100	111	137	173	165	136	139	161	135	167	136	117	330
Meghalay	91	101	105	109	79	79	122	82	81	86	80	91	117	102	315
Mizoram	32	13	9	26	36	31	18	29	25	16	40	51	34	27	240
Nagaland	25	35	77	54	58	19	18	14	13	19	41	94	11	0	213
Odisha	1324	1466	1466	2088	2198	1964	2386	2062	2129	2333	3433	3507	4074	3328	3541
Punjab	1295	1398	1147	1434	1047	1497	1431	1376	1962	2064	2122	1519	1965	2101	2314
Rajasthan	2430	2596	2380	2175	2870	2581	2913	3119	2625	2723	3029	3774	3638	3695	3908
Sikkim	39	66	101	34	26	36	159	49	170	32	83	71	85	70	283
Tamil Nad	13524	16417	17042	17013	17848	24912	18944	20722	20920	21810	20984	21441	23165	23405	23618
Telangana	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3623	3783	4090	4303
Tripura	124	234	73	211	306	438	464	526	424	432	382	374	268	217	430
Uttarakha	555	441	416	464	335	269	293	358	378	392	281	391	408	568	781
Uttar Prad	4214	5391	4975	5961	7396	8130	8783	8591	8861	9362	13196	11669	9320	11715	11928
West Ben	2610	4067	5040	3591	3170	3237	2600	4074	3832	3340	3832	3404	3802	3793	4006
Andaman	0	0	0	0	0	0	0	0	0	0	0	56	32	19	232
Chandigar	0	2	0	0	0	100	81	146	0	0	0	0	0	0	232
Dadra and	0	0	0	0	0	0	0	0	0	0	0	0	0	0	213
Daman an	0	0	0	0	0	0	0	0	0	0	0	0	0	0	213
Delhi	1062	1045	1300	1592	1133	919	952	916	0	0	0	0	0	0	213
Lakshadw	0	0	0	0	0	0	0	0	0	0	0	0	0	0	213
Puduchen	0	0	0	0	386	395	276	146	41	0	58	152	256	373	586
All india	90963	101191	103827	107632	116908	123972	115992	122406	122239	118835	124358	123408	120518	121655	121868

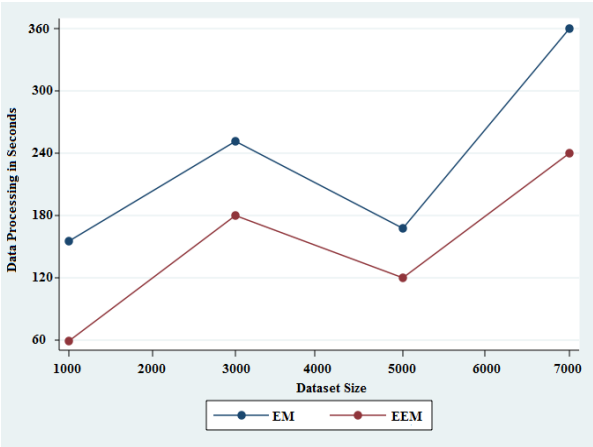

Figure 4. Time for clustering

The proposed EEM algorithm less time to divide the data into multiple clusters and this method is more accurate than the existing EM method. The Figure 4 below illustrates the time taken to perform clustering.

Some of the Association rules generated are given below.

Table 3. Association rules for road accident prediction

Rule	Best Rule
1	If Driv_gen=M 47 ==> Class=Death 47 conf:(1)
2	If Road_surface=Dry 47 ==> Class=Death 47 conf:(1)
3	If Weather_cond=Clear 46 ==> Class=Death 46 conf:(1)
4	If Driv_gen=M Road_surface=Dry 46 ==> Class=Death 46

5	If Weather_cond=Clear Road_surface=Dry 46 ==> Class=Death 46 conf:(1)
6	If Driv_gen=M Weather_cond=Clear 45 ==> Class=Death 45 conf:(1)
7	If Driv_gen=M Weather_cond=Clear Road_surface=Dry 45 ==> Class=Death 45 conf:(1)
8	If Driv_drink=Not_checked 42 ==> Class=Death 42 conf:(1)
9	If N_of_p_injured=1 Driv_gen=M 31 ==> Class=Death 31 conf:(1)
10	If Driv_gen=M and Driv_drink=Not_checked and Weather_cond=Clear and Road_surface=Dry 40 ==> Class=Death 40 conf:(1)

The performance analysis of the improved association rule mining algorithm is discussed below.

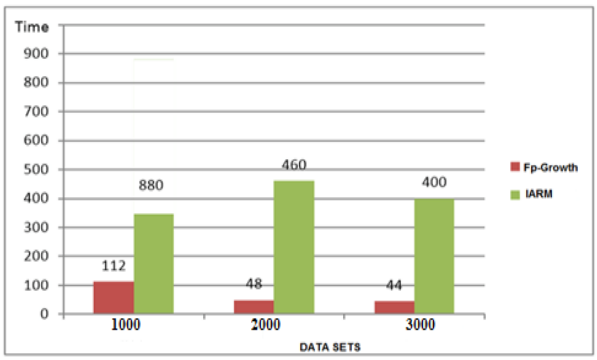

Figure 5. Performance level analysis

Examination like kind of vehicles (bicycle, auto, transport, lorry, jeep, truck, etc.) is done to foresee accidents on various criteria like speed purpose of repression, and harm importance [22]. Tantamount examination is done on different criterias, for instance, reason for accidents age of the driver, Accident area, speed measure, Accident time and season too.

Table 4. Top factors for road accidents

Contributing factor	Percentage of accidents (%)
Rash driving	62.57
Object hit	26.67
Lane change	8.1

The above Table 4 determines the level of accidents caused dependent on various parameters.

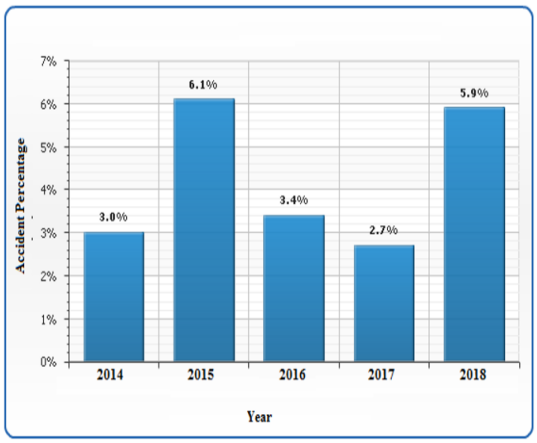

Figure 6. Accidents by speed limit

The above Figure 6 speaks to the accidents brought about by speed limit which shows the level of accidents occurred.

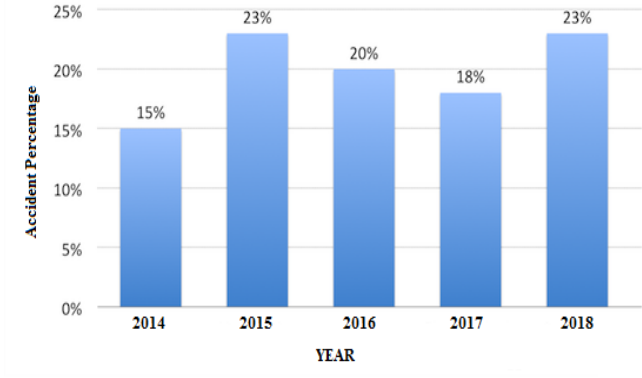

Figure 7. Accidents by injury severity

The above Figure 7 speaks to the harmed seriousness brought about by speed limit which demonstrates the level of accidents occurred.

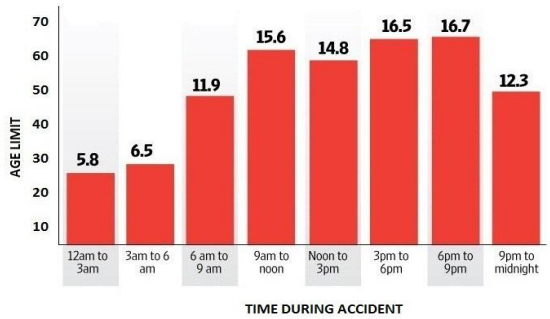

Figure 8. Final Road accident prediction based on time and age

In Figure 8, we can propose an answer for Road Accident forecast dependent on time on the day and the age of the driver who is driving a vehicle which meets with Accident [23]. In light of the above diagram clients are guided about the time in which larger part accidents happens and furthermore caution about age limit for driving.

5. CONCLUSION

The Proposed work discusses about the road accident investigation, which is clearly identified. In this proposed work, we proposed a system for predicting road accident causes for various kinds of accidents that influences utilization of big data analytics methods using machine learning techniques for accurate prediction. Initially machine learning based clustering technique is applied on the multi-dimensional dataset using the EEM algorithm and then association rules are developed for road accident prediction. Then the CCMF() and TCAMP() methods are used for automatic prediction of accidents by applying association rules to each and every parameter. We played out a few tests on road accident information to calculate reason for accidents and analyzing the data using map reduce methods. The outcomes exhibit that the proposed approach is far better in prediction of road accidents than the existing methods.

REFERENCES

[1] Zheng CT, Liu C, Wong HS. (2018). Corpusbased topic diffusion for short text clustering. Neurocomputing 275: 2444-2458. http://dx.doi.org/10.1504/IJIT.2018.090859

[2] Sarkar S, Pateshwari V, Maiti J. (2017). Predictive model for incident occurrences in steel plant in India. In ICCCNT 2017, IEEE, pp. 1-5. http://dx.doi.org/10.14299/ijser.2013.01

[3] Sharma S, Sabitha AS. (2016). Flight crash investigation using data mining techniques. In Information Processing (IICIP), 2016 1st India International Conference on. IEEE, pp. 1-7. http://dx.doi.org/10.14299/ijser.2013.01

[4] Williams T, Betak J, Findley B. (2016). Text mining analysis of railroad accident investigation reports. In 2016 Joint Rail Conference. American Society of Mechanical Engineers V001T06A009-V001T06A009. http://dx.doi.org/10.14299/ijser.2013.01.

[5] Sarkar S, Vinay S, Raj R, Maiti J, Mitra P. (2018). Application of optimized machine learning techniques for prediction of occupational accidents. Computers & Operations Research (Elsevier), pp. 343-348. http://dx.doi.org/10.1145/3075564.3078884

[6] Palamara F, Piglione F, Piccinini N. (2011). Selforganizing map and clustering algorithms for the analysis of occupational accident databases. Safety Science 49(8-9): 1215-1230. http://dx.doi.org/10.1109/HPCSim.2016.7568393

[7] Chen ZY, Chen CC. (2015). Identifying the stances of topic persons using a model-based expectationmaximization method. J. Inf. Sci. Eng 31(2): 573-595. http://dx.doi.org/10.1504/IJASM.2015.068609

[8] Verma A, Maiti J. (2018). Text-document clustering based cause and effect analysis methodology for steel plant incident data. International Journal of Injury Control and Safety Promotion, 1-11. http://dx.doi.org/10.1080/17457300.2018.1456468

[9] Srivastava AN, Zane-Ulman B. (2005). Discovering recurring anomalies in text reports regarding complex space systems. In Aerospace Conference, IEEE. IEEE 3853-3862.

[10] Ghazizadeh M, McDonald AD, Lee JD. (2014). Text mining to decipher free-response consumer complaints: Insights from the nhtsa vehicle owner's complaint database. Human Factors 56(6): 1189-1203. http://dx.doi.org/10.1504/IJFCM.2017.089439

[11] Stewart M, Liu W, Cardell-Oliver R, Griffin M. (2017). An interactive web-based toolset for knowledge discovery from short text log data. In International Conference on Advanced Data Mining and Applications. Springer, pp. 853-858. http://dx.doi.org/10.1007/978-3-319-69179-4_61

[12] Abdat F, Leclercq S, Cuny X, Tissot C. (2014). Extracting recurrent scenarios from narrative texts using a bayesian network: Application to serious occupational accidents with movement disturbance. Accident Analysis & Prevention 70: 155-166. http://dx.doi.org/10.1016/j.aap.2014.04.004

[13] Abbas OA. (2008). Comparisons between data clustering algorithms. Int. Arab J. Inf. Technol 5(3): 320-325. http://dx.doi.org/10.1504/IJIDS.2016.075787

[14] Sarkar S, Patel A, Madaan S, Maiti J. (2017). Prediction of occupational accidents using decision tree approach. In INDICON 2017 (IEEE). IEEE 1-6. http://dx.doi.org/10.1109/INDICON.2016.7838969

[15] Sarkar S, Vinay S, Pateshwari V, Maiti J. (2017). Study of optimized svm for incident prediction of a steel plant in India. In INDICON (IEEE). IEEE 1-6. http://dx.doi.org/10.1109/INDICON.2016.7838894

[16] Sarkar S, Baidya S, Maiti J. (2017). Application of rough set theory in accident analysis at work: A case study. In ICRCICN 2017, IEEE 245-250. http://dx.doi.org/10.1109/ICRCICN.2017.8234514

[17] Oztekin A, Al-Ebbini L, Sevkli Z, Delen D. (2018). A decision analytic approach to predicting quality of life for lung transplant recipients: A hybrid genetic algorithms-based methodology. European Journal of Operational Research 266(2): 639-651. https://doi.org/10.1016/j.ejor.2017.09.034

[18] Sarkar S, Lohani A, Maiti J. (2017). Genetic algorithm-based association rule mining approach towards rule generation of occupational accidents. In Communications in Computer and Information Science (Springer). Springer, Singapore 776: 517-530. https://doi.org/10.1007/978-981-10-6430-2_40

[19] Wang Y, Xu W. (2018). Leveraging deep learning with lda-based text analytics to detect automobile insurance fraud. Decision Support Systems 105: 87-95. https://doi.org/10.1016/j.dss.2017.11.001

[20] Wang Z, Ren J, Zhang D, Sun M, Jiang J. (2018). A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. Neuro Computing 287: 68-83. https://doi.org/10.1016/j.neucom.2018.01.076

[21] Zheng YJ, Chen SY, Xue Y, Xue JY. (2017). A Pythagorean-type fuzzy deep denoising autoencoder for industrial accident early warning. IEEE Transactions on Fuzzy Systems 25(6): 1561-1575. https://doi.org/10.1109/TFUZZ.2017.2738605

[22] Ding L, Fang W, Luo H, Love PE, Zhong B, Ouyang X. (2018). A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory. Automation in Construction 86: 118-124. http://dx.doi.org/10.1504/IJIT.2018.090878

[23] Fang W, Ding L, Luo H, Love PE. (2018). Falls from heights: A computer vision-based approach for safety harness detection. Automation in Construction 91: 53-61. https://doi.org/10.1016/j.autcon.2018.02.018