I J C T A, 10(9), 2017, pp. 783-793 ISSN: 0974-5572
© International Science Press

Mapping of Food and Driving Behavior Flavored with Interestingness: An Incremental Learning Approach

Preeti Mulay*, Rahul Raghvendra Joshi* and Selam Mulatu*

ABSTRACT

Safety comes first while driving on roads. Entities like food intake, gender, driving time and interestingness of driver play major role in relation to road safety. Hybrid Expert System for Knowledge Management (HES-KM) and associative rule mining are discussed in this paper. Hierarchical regression model is implemented to predict results and generated knowledge about road safety, ownership of vehicles and food offerings by chain of restaurants on high ways is augmented. This model is further implemented in WEKA to enhance knowledge of various users in open forums relating to effectually modified association mining algorithm. Obtained results quality is supported by Mutual Information (MI) matrix, Predictive Accuracy matrix and use of Cobweb Incremental Clustering algorithm are clearly visible in results section of this research paper.

Keywords: Hybrid expert system; Machine learning; Knowledge management; Intelligent computational model; Traffic accident prediction model.

1. INTRODUCTION

Now-a-days, the rate of traffic accidents is increasing like a whirlwind. Knowledge augmentation (KA) techniques need to be applied to find out causes behind traffic accidents [1, 3]. Using efficient and effective knowledge management paradigm an "*optimal intelligent traffic accident prediction computational model*" can be build which in turn can play a crucial role in tackling the said rate of traffic accidents. Best plausible machine learning paradigms applied on a huge volume of traffic accidents data will help in increasing knowledge sharing about accident prediction. Figure 1 shows graphical representation of proposed research. In this paper optimal intelligent computational model for traffic accident prediction is proposed. Also, Food Driving Interestingness (FDI) algorithm is implemented by considering driver's food habit, gender, road type and timing of driving [2].

Predictive association rule mining or Apriori algorithm with interestingness (IPARMI) [4] given in equation 1 is based on conditional entropy and acts as a metric for measuring interestingness [5] of a rule (a semi-supervised incremental learning approach) which can be used for building patterns which helps to predict traffic accident based on driver's food habit, gender, road type and time of driving.FDI algorithm along with Cobweb and modified Cobweb algorithm make use of modified category utility function which is given in equation 2. Training data set includes US crash [6] traffic data (FARS) and Nutritional survey data (NHANES) along with WEKA tool.

$$MI(X,Y) = H(Y) - H\left(\frac{Y}{X}\right) \tag{1}$$

* CS/IT Dept. of Symbiosis Institute of Technology (SIT), Pune affiliated to Symbiosis International University (SIU), Pune, India, *Emails: preeti.mulay@sitpune.edu.in, rahulj@sitpune.edu.in, selam.mulatu@sitpune.edu.in*

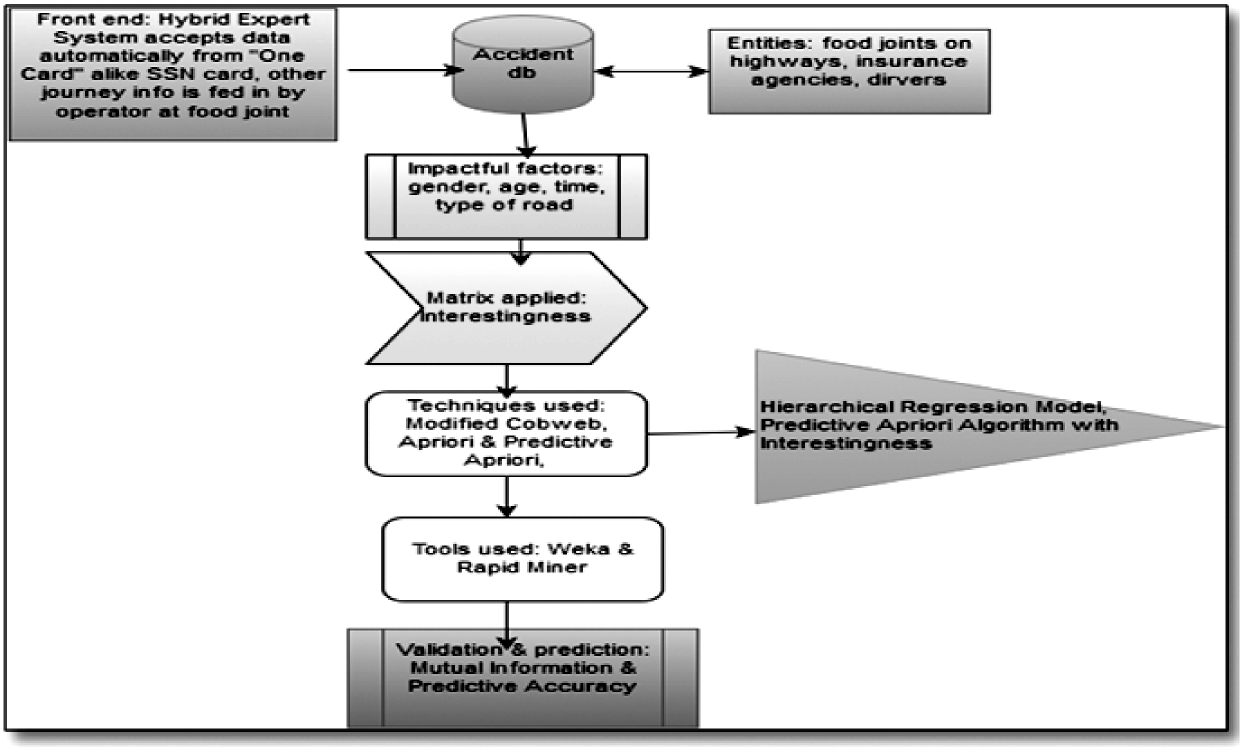

Figure 1: Graphical representation of undertaken research

Here, the rule is X => Y, H(X) is entropy value of consequent variable Y. $H\left(\frac{Y}{X}\right)$ is the conditional entropy between the antecedent variable X and consequent variable Y.

$$Cu\left(C_1, C_2, \ldots C_k\right) = \frac{1}{|C|} + \sum_{i=1}^{|C|} \Pr\left[C_l\right] \frac{1}{2\sqrt{\pi}} \sum_{i=1}^{|A|} \left(\frac{1}{\sigma_{il}} - \frac{1}{\sigma_i}\right) \tag{2}$$

Here $|C|$ - number of clusters in a training data, $|A|$ - number of attributes in each instance, σ_{il}-Standard deviation of attribute a_i and $\Pr|C_l|$ is probability of cluster C_l.

Factors related to accident related may vary depending on accident location and time of driving. By taking into account variability of each attribute within and in between, a 3-level (level-1 - accident location – equation 3, level-2 - time of driving – equation 4 and level-3 - accident factor - equation 5) equation is proposed. Hierarchical Regression Model is developed to predict about accident, its related causes and time of driving is given in equation 7.

1. Level -1 - Accident Location

$$V_{ijk} = \pi_{0jk} + \sum_{p=1}^{n} \pi_{pjk} a_{pjk} + e_{pjk} \tag{3}$$

Here, for variability of level - 1 units σ2 = var (e_{pjk}). Y_{ijk} is an function for chance of an accident to occur. π_{0jk} is the intercept for chance of accident to occur. p is level -1 predictor variable for accident location. π_{pjk} is regression coefficient associated with a_{pjk} for the j[th] level. a_{pjk} is value of number of accidents and e_{pjk} is random error associated with the i[th] level nested within the j[th] level.

2. Level - 2 - Time of driving /Accident Time

$$\pi_{pjk} = \beta_{p0k} + \sum_{p=1}^{n} \beta_{pqk} x_{qjk} + r_{pjk} \tag{4}$$

Here, variability of level - 2 units r_{pjk}, Tπ = var (r_{pjk}). β_{p0k} is intercept for the j[th] level-2 unit i.e. intercept for time of driving. x_{qjk} is value of the level-2 predictor for evening time. β_{pqk} is regression coefficient associated with accident location relative to level-2 intercept. r_{pjk} is random error of the j[th] level-2 unit adjusted for x_{qjk} on the intercept.

3. Level -3 - Accident Factors

$$\beta_{pqk} = y_{pq0} + \sum_{p=1}^{n} y_{pqs} w_{sk} + u_{pqk} \tag{5}$$

Here, variability of level - 3 units u_{pqk}, Tβ = var (u_{pqk}). y_{pq0} is intercept for the j[th] level-3unit i.e. intercept for accident factor. w_{sk} is value of the level-3 predictor for accident factor. y_{pqs} is regression coefficient associated with w_{sk} accident factor relative to level- 2 intercept. u_{pqk} is random error of the j[th] level-3 unit adjusted for w_{sk} accident time/time of driving on the intercept.

2. LITERATURE REVIEW

This research includes various algorithms, technologies and tools to predict accident details accurately. Online interactive incremental data mining tool (OIIDM) [7-11] trains data and form clusters in an incremental mining fashion by using Cobweb algorithms modified category utility (CU) function [1-2]. Results of multiple sleep latency test [12-13] indicate that consumption of high fat food increases sleepiness where as carbohydrates are related to alertness. After studying sleep related crash cases of North Carolina for finding out major factors related to sleep related crashes and it is observed that peoples who are shift workers have sleep hours below the average one are more prone to crashes. The knowledge security model [14] helps managers to minimize loop holes of securing useful information of an organization by doing identification of information [15] risk factors. In [16] Bayesian network and clustering techniques were used to build a traffic accident prediction model to find accidents caused by a particular factor and by integrating it with traffic accident data and real time weather conditions data. In [17], association rule mining is used to investigate the association between crash factors related to geometric design found in Italy for better design of roads. In [18], study is focused on eating habits of a driver to road traffic accidents using association rule mining. In [19], accident prediction model based on Support Vector Machines and Gaussian kernel was presented, in [20] road accidents count was analyzed by using hierarchical clustering and cophenetic correlation coefficients and [21] characterizes accident locations using data mining. These studies indicate eating habits of driver is one of the causalities of road traffic accident which in turn may change driving behavior and attitude.

3. INTELLIGENT TRAFFIC ACCIDENT PREDICTION COMPUTATIONAL MODEL

Studies shows that drivers' whose age group is in between 18 to 35 and >= 65 are more likely to be engaged in sleep related crashes than those whose age group is in between 36 to 55. Also, driving in between 4:00 AM - 6AM is very risky as during this period human body is at lowest level of alertness. Proposed intelligent traffic accident prediction computational model (Fig. 2) can act as a catalyst for coming out of such scenarios. Food and driving interestingness algorithm (FDI) is as given below:

Input: Accident and food data

Output: Set of rules to predict road type, driver's age, gender, profession and eating habit related to accident

1. Apply Cobweb/ modified Cobweb incremental (CI) clustering algorithm to generate clusters
2. Apply association rule mining, with interestingness as major feature by calculating mutual information based on conditional entropy value using (equation 1).
3. Note outcomes of association rule mining algorithm, CI and other numbers, save generated rules in first iteration in db

4. On arrival of new data, run incremental clustering algorithm in next iteration
5. Either update existing clusters or form new clusters based on Category Utility(CU)Value
6. Continue to calculate interestingness, rules etc and observe the changes in them
7. Arrive at conclusion about accidents,
8. Repeat these steps with every iteration
9. Incrementally learn about driver's profession, age, and interestingness, eating habits and their mapping,
10. Predict about accidents using these values / outcomes

Figure 2: FDI Algorithm

Intelligent traffic accident prediction computational model uses incremental learning approach to increase its performance, as and when new data is made available or when predicted data is feed into the same system. This research also recommends menus containing low fat food by considering customer's age and service time, in particular for late night travelers. Fast food joints can collect customer's basic information through "One Card" similar to SSN card in USA and remaining details can be captured through questionnaire, matching them with the facts derived from the knowledge base system.

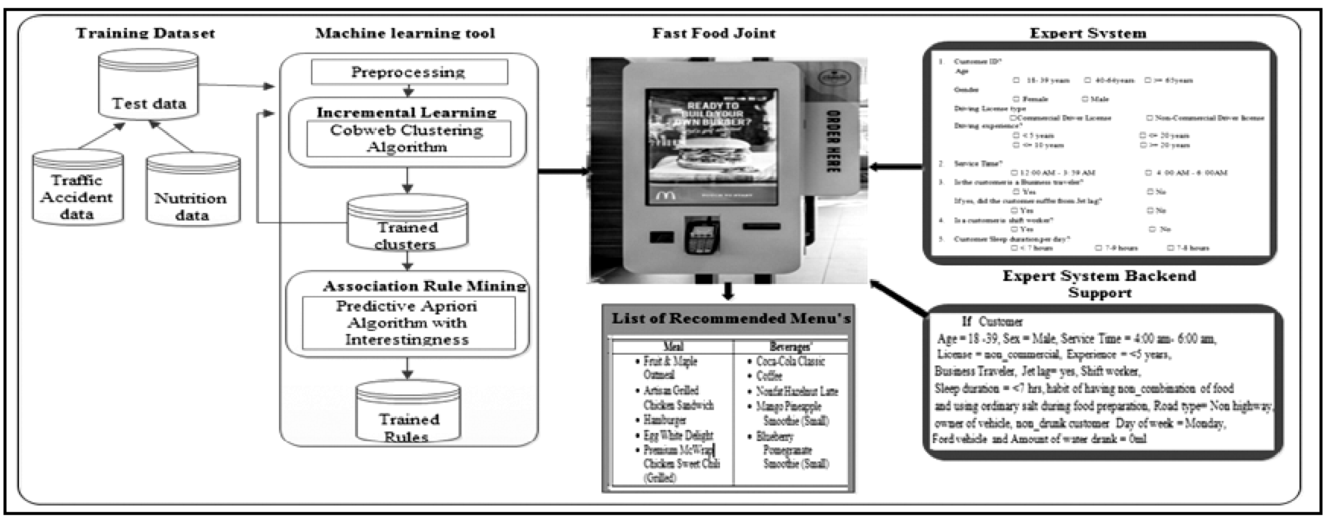

Figure 3: Computational model for intelligent traffic accident prediction

4. HES-KM QUESTIONNAIRE

In intelligent traffic accident prediction computational model (Figure 2) details are partly fetched through computerized system using a unique ID card and through a customer or a counter waiter at the time of purchase of food item to recommend a menu by mapping driving time /age / gender of a driver.

Remark: Questions written in green, red and black represent questions answered/filled by counter waiter, cash counter machine by reading from the system and customer of the fast joint respectively.

1. Customer ID? Age ☐ 18-39 years ☐ 40-64 years ☐ > = 65 years
   Gender ☐ Female ☐ Male Driving License type ☐ Commercial Driver License ☐ Non-Commercial Driver license Driving experience ☐ < 5 years ☐ <= 20 years
   ☐ <= 10 years ☐ >= 20 years
2. Service Time and Day of week?

3. Is the customer is a Business traveler? ☐ Yes ☐ No
   If yes, did the customer suffer from Jet lag? ☐ Yes ☐ No
4. Is a customer is shift worker? ☐ Yes ☐ No
5. Customer Sleep duration per day? ☐ < 7 hours ☐ 7-9 hours ☐ 7-8 hours
6. Does the customer have having habit of non-combination of food? ☐ Yes ☐ No
7. Type of salt used by the customer during food preparation at home?
   ☐ Non- sodium adjusted salt £ Sodium adjusted salt
8. Location of the fast food joint shop on road?
9. Does the customer have alcohol? ☐ Yes ☐ No
10. Is the customer is owner of the vehicle? ☐ Yes ☐ No
11. Amount of water the customer drink per day? ☐ Very less ☐ <= 1 liter ☐ >1 liter

Figure 4: Questionnaire for the intelligent traffic accident prediction model

Facts related to high risk, moderate risk sleep related crash group scenarios are considered and a list of recommended menu's is prepared for 10 such scenarios out which only two samples for high and low risk sleep related crash are shown in tables 1 to 4 by taking into account information provided by the customer and risk levels based on age group and driving time based on nutritional facts [22-24].

5. HIGH RISK SLEEP RELATED CRASH GROUP

a. If (Customer Age = 18 - 39, Gender = Male, Service Time = 4:00 am- 6:00 am, License = non-commercial, Experience< =5 years, Business Traveler, Jet lag = yes, Shift worker, Sleep duration <=7 hrs, habit of having non-combination of food and use ordinary salt during food preparation, Road type = Non highway, owner of vehicle, non-drunk customer Day of week = Monday, and Amount of water drank = 0 liter)

Table 1
Recommended menu for scenario 'a'

Meal	*Beverages*	*Remark*
• Fruit & Maple Oatmeal • Artisan Grilled Chicken Sandwich • Hamburger • Egg White Delight • Premium McWrap Chicken Sweet Chili (Grilled)	• Coca-Cola Classic, Coffee • Nonfat Hazelnut Latte • Mango Pineapple Smoothie (Small) • Blueberry Pomegranate Smoothie (Small)	Foods and beverages contain very low amount of fat and high carbohydrate contents

b. If (Customer Gender = Female, habit of having non-combination of food and non salty food and all other conditions are similar to a except Gender)

Table 2
Recommended menu for scenario 'b'

Meal	*Beverages*	*Remark*
• All Meal types mentioned in Table 1. • Premium Southwest Salad with Grilled Chicken • Grilled Onion Cheddar	• Beverages mentioned in Table 2 • Sprite® (Small pack) • Fat Free Chocolate Milk Jug • Fruit n Yogurt Parfait	Foods and beverages containing very low amount of fat and high carbohydrate contents(little bit difference than male drivers)

6. MODERATE RISK SLEEP RELATED CRASH GROUP

c. If (Customer Age = 40 - 64, Gender = Male, Service Time = 4:00 am- 6:00 am, License = non-commercial, Experience < = 10 years, Business Traveler, Jet lag = Yes, Shift worker, Sleep duration < =7 hrs, habit of having non-combination of food and non _ sodium adjusted salt used in food preparation, Road type = Non highway, non drunk driver, owner of vehicle, and Day of week = Monday and Amount of water drank = 0 liter)

Table 3
Recommended menu for scenario 'c'

Meal	*Beverages*	*Remark*
· Fruit & Maple Oatmeal · Artisan Grilled Chicken · Sandwich · Premium Asian Salad (without chicken) · Premium Southwest Salad with Grilled Chicken	• Coca-Cola Classic, Coffee • Nonfat Hazelnut Latte • Mango Pineapple Smoothie (Small Pack) • Blueberry Pomegranate Smoothie (Small Pack) • Iced Strawberry Lemonade (Small) • Vanilla Reduced Fat Ice Cream Cone	Foods and beverages with very high calorie, small fat and high carbohydrate contents

d. If (Customer Gender = Female and all conditions mentioned in 'c' except Gender = Male)

Table 4
Recommended menu for scenario'd'

Meal	*Beverages*	*Remark*
• Meals listed in Table 4 • Premium Southwest Salad (without chicken) • Premium Asian Salad with Grilled Chicken	• Listed beverages in Table 4	Foods and beverages with less calories than that of 'c', with small fat and high carbohydrate contents

Traffic accident data stores are dynamic in nature so effective communication link between traffic accident databases and machine learning can be created by using following optimized network-model.

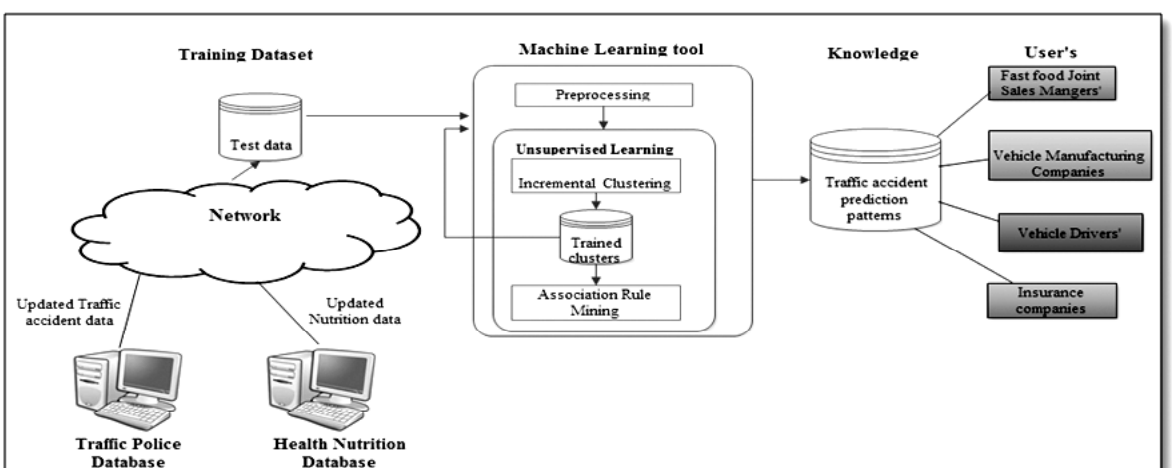

Figure 5: Communication network-model across traffic accident databases and traffic accident perdition model

7. EXPERIMENTAL SET UP AND RELATED OUTCOMES

An experiment was conducted by using WEKA tool in incremental fashion by considering 400 instances for 5 iterations having 39 attributes related to causes of accident and nutrition information taking into account FDI algorithm.

7.1. Results of 1st iteration of Cobweb clustering algorithm

The first iteration of Cobweb (Fisher & Douglas, 1987) builds 241 clusters from 400 total instances with acuity = 0.1 and Cutoff = 0.6.

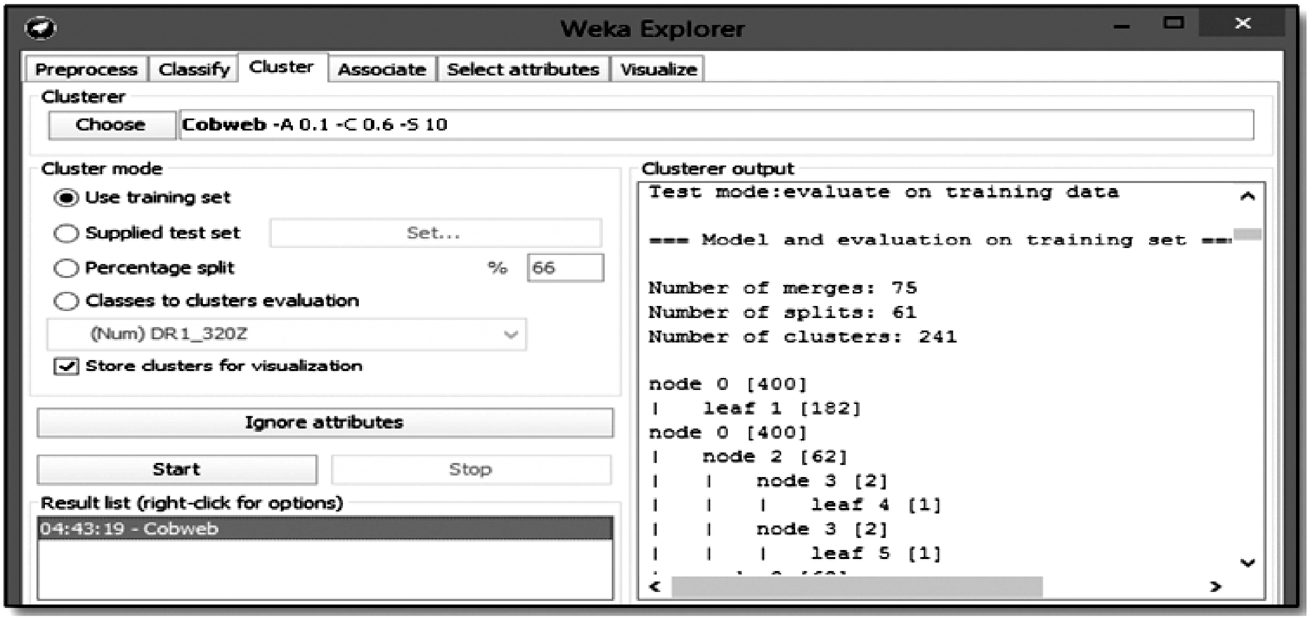

Figure 6: Results of Cobweb clustering algorithm's 1st iteration results

7.2. Result of Predictive Apriori algorithm's 1st iteration with interestingness

The 10 rules are trained by using predictive Apriori algorithm with Interestingness indicates that most of the accidents happened on Non highway road during clear weather condition and the factors related to accidents are male drivers having habit of consuming non-combination, non home prepared food, non drunked driver and absence of sodium adjusted salt used in food preparation.

7.3. Results of 5th iteration of Cobweb clustering algorithm

In 5th iteration number of merges increased from 87 to 224 and number of splits from 76 to 194 as compared to 4th iteration and number of clusters from 3 to 7.

7.4. Results after 5th iteration of Predictive Apriori algorithm with interestingness

The result of predictive Apriori algorithm with mutual information (IPARMI) shows for 2000 accidents the main were drivers whose dietary status meets minimum criteria and most of the accidents were took place in major collector roads and on the first day of week.

7.5. Number of splits, merges and cluster formed during each iteration and time taken by them

In 5 iterations - the number splits increased and number of clusters formed by Cobweb algorithm is decreased and vice versa as shown in Table 5. The comparative information in terms time taken for undertaken scenario for both Cobweb clustering algorithm is shown in Table 6.

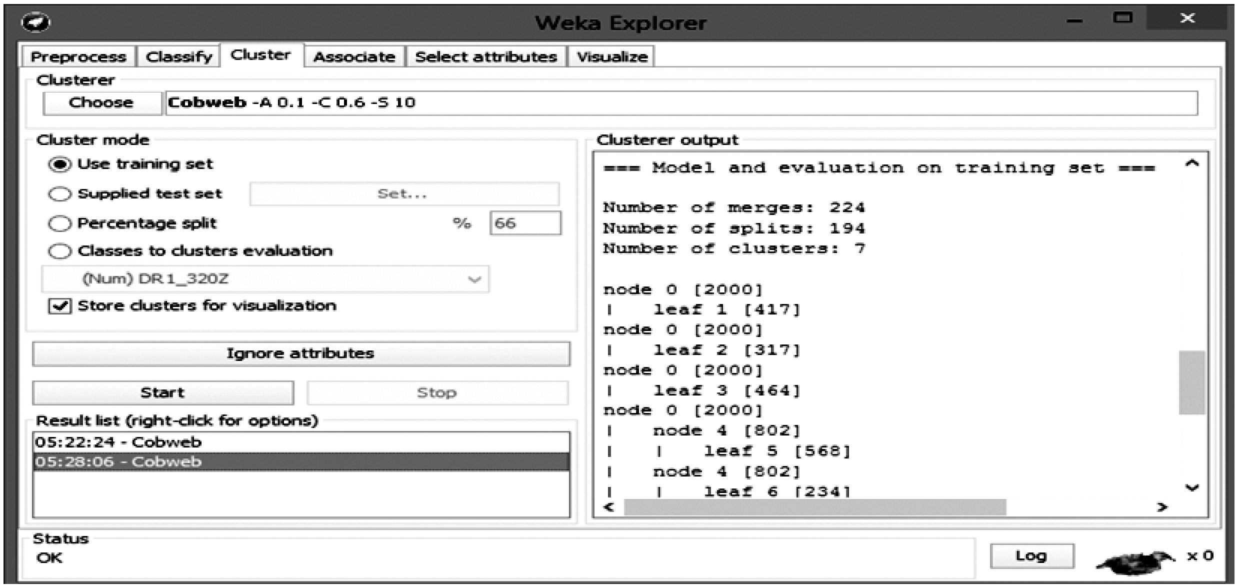

Figure 7: Screenshot for Cobweb algorithm's 5th iteration

Table 5
Number of instances, merges, splits and clusters formed in 5 iterations

No. of iterations	*No. of instances*	*No. of merges*	*No. of splits*	*No. of clusters formed*
1	400	75	61	241
2	800	187	162	555
3	1200	143	128	5
4	1600	87	76	3
5	2000	224	194	7

Table 6
Time taken by Cobweb and Modified Cobweb clustering algorithm to form clusters

No. of iterations	*No. of clustersformed*	*Time taken to form clusters (in seconds)*	
		Cobweb algorithm	*Modified Cobweb algorithm*
1	241	4.49	3.91
2	555	10.3	9.07
3	5	10.28	8.78
4	3	6.61	5.68
5	7	14.68	12.91

7.6. Accuracy and mutual information measurement

Table 7
Accuracy of the rules in each of iteration

No. of iteration	*No. of Instances*	*Total No. of Rules*	*Predictive Accuracy*	*Mutual Information*
1	400		99%	50%
2	800		60%	80%
3	1200	10	90%	40%
4	1600		70%	80%
5	2000		60%	70%

7.7. Prediction results of intelligent accident prediction computational model

The rules trained by IPARMI indicate male drivers are mostly prone to accident compared to females and probability for chances of accident on non-highway roads is greater than that of highway roads. The following figures 8 to 10 indicate accident causes for different types of scenarios.

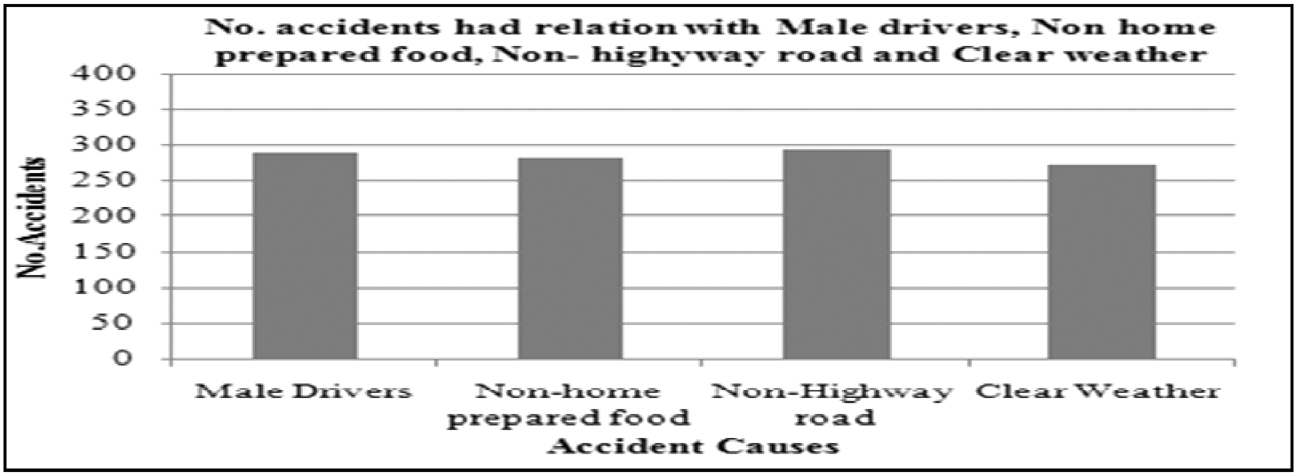

Figure 8: Accidents because of male drivers; consumption of non-home prepared food, on non-highway road and due to clear weather condition

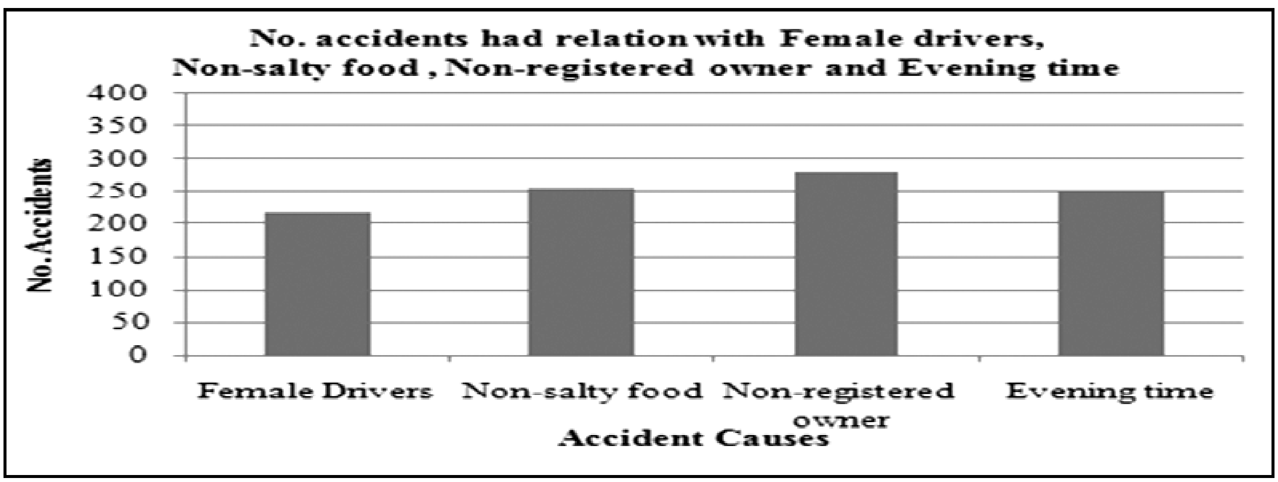

Figure 9: Accidents because of females, non-salty food, non-registered owner and evening time

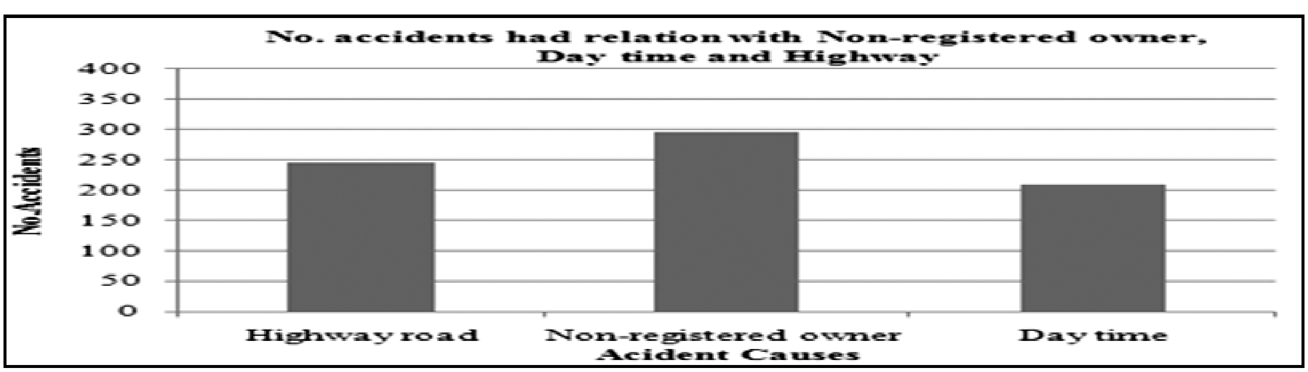

Figure 10: Accidents because of non-registered owners, day time and highway road

7.8. Applicability of traffic accident prediction expert system

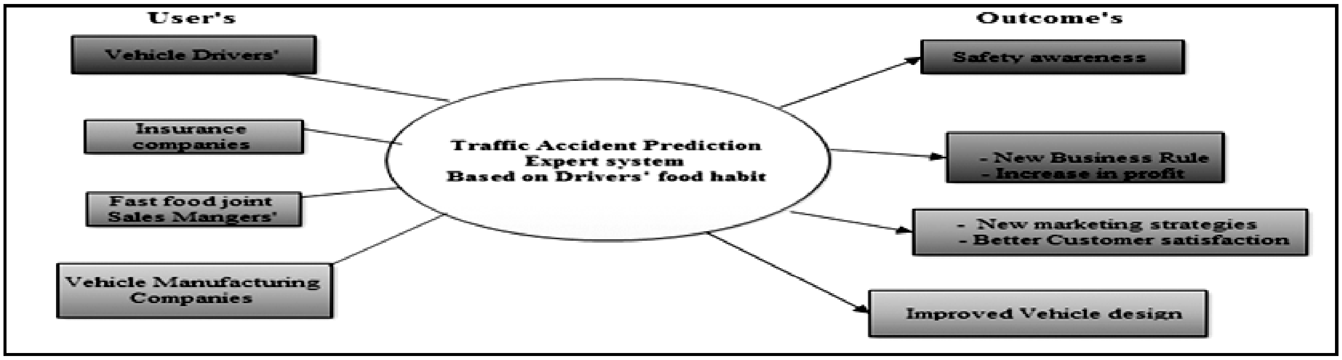

Figure 11: Applicability of Traffic Accident Prediction Expert System and its outcomes

8. CONCLUSION

Incremental semi-supervised learning along with an effective knowledge management approach is applied to build an optimal intelligent traffic accident prediction computational model. This model facilitate various entities such as fast food joints, vehicle manufacturers, vehicle drivers and insurance companies to collaborate together to tackle the risk of road traffic accident. This in turn increases profit margins of various vendors by implementing knowledge sharing; customization techniques can be developed based on the results obtained by analyzing influence of driver's food habit and road traffic. Also the outcomes of the prediction model may help to design better quality vehicle design and fast joints to increase their profit by planning new marketing strategies especially during late night service hours.

9. FUTURE DIRECTIONS

In this paper accident prediction model is developed based UK and Australian accident datasets by considering features like food, interest, time of driving, type of road, location and history or background of a driver. This study needs to be elaborated further by considering Indian accident datasets along with additional features viz., profession of driver, roads, type of car (luxury or ordinary), occasion like 31st December, Zodiac Sign [25] etc.

REFERENCES

[1] Mulay, P., & Kulkarni, P. A. (2013). Knowledge augmentation via incremental clustering: new technology for effective knowledge management. International Journal of Business Information Systems, 12(1), 68-87.

[2] Mulay, P., & Mulatu, S. (2016). What You Eat Matters Road Safety: A Data Mining Approach. *Indian Journal of Science and Technology*, *9*(15).

[3] Mulay, M. P. Incremental Learning.

[4] Mutter, S., Hall, M., & Frank, E. (2004, December). Using classification to evaluate the output of confidence-based association rule mining. In *Australasian Joint Conference on Artificial Intelligence* (pp. 538-549). Springer Berlin Heidelberg.

[5] Stefan, R. (2001). Incremental learning with support vector machines. In *Data Mining, IEEE International Conference on* (p. 641).

[6] Montella, A. (2011). Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. *Accident Analysis & Prevention*, *43*(4), 1451-1463.

[7] Borhade, M., & Mulay, P. (2015). Online interactive data mining tool. *Procedia Computer Science*, *50*, 335-340.

[8] Data mining software in Java. http://www.cs.waikato.ac.nz/ml/weka/ (Accessed on 1st September 2016).

[9] Margahny, M. H., & Mitwaly, A. A. (2005, December). Fast algorithm for mining association rules. In *the conference proceedings of AIML, CICC, pp (36-40) Cairo, Egypt* (pp. 19-21).

[10] Mutter, S., Hall, M., & Frank, E. (2004, December). Using classification to evaluate the output of confidence-based association rule mining. In *Australasian Joint Conference on Artificial Intelligence* (pp. 538-549). Springer Berlin Heidelberg.

[11] Dr. Preeti Mulay, Rahul Raghvendra Joshi, Aditya Kumar Anguria, Alisha Gonsalves, Dakshayaa Deepankar, D. G. (2017). Threshold Based Clustering Algorithm analyzes Diabetic Mellitus. In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications (Vol. 2, pp. 1-7). Springer Nature Singapore Pte Ltd. 2017. DOI: 10.1007/978-981-10-3156-4_3.

[12] Stutts, J. C., Wilkins, J. W., Osberg, J. S., & Vaughn, B. V. (2003). Driver risk factors for sleep-related crashes. *Accident Analysis & Prevention*, *35*(3), 321-331.

[13] Drowsy Driving and Automobile crashes http://www.nhtsa.gov/people /injury/drowsy_driving1/drowsy.html #II.BIOLOGY OF HUMAN SLEEP (Accessed on 27th September 2016).

[14] Ilvonen, I., Jussila, J. J., & Kärkkäinen, H. (2015). Towards a Business-Driven Process Model for Knowledge Security Risk Management: Making Sense of Knowledge Risks. *International Journal of Knowledge Management (IJKM)*, *11*(4), 1-18.

[15] Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

[16] Paikari, E., Moshirpour, M., Alhajj, R., & Far, B. H. (2014, August). Data integration and clustering for real time crash prediction. In *Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on* (pp. 537-544). IEEE.

[17] Kritikou I., Pejovic S., Vgontzas A., Pernandez M. J., Basta M., Bixler E.. High fat intake is Associated with Physiological sleepiness in healthy non- obese adults, *Journal of sleep and sleep research disorder*, 2013, Vol. 36.

[18] Sharma, B., Katiyar, V. K., & Kumar, K. (2016). Traffic Accident Prediction Model Using Support Vector Machines with Gaussian Kernel. In *Proceedings of Fifth International Conference on Soft Computing for Problem Solving* (pp. 1-10). Springer Singapore.

[19] Kumar, S., & Toshniwal, D. (2016). Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC). *Journal of Big Data*, *3*(1), 1-11.

[20] Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations. *Journal of Modern Transportation*, *24*(1), 62-72.

[21] Facts and Stats http://drowsydriving.org/about/facts-and-stats/ (Accessed on 16th October 2016).

[22] The Royal society for the prevention of accidents. Driver Fatigue and Road accidents, Registered Charity, 2011, No: 207823.

[23] Pawar. Use of decision support system in management functioning and decision making, Global Journal of Computational Intelligence Research (GJCIR), 2016, 5(1).

[24] McDonald's USA Nutrition Facts for Popular Menu Items http://nutrition.mcdonalds.com/getnutrition/nutritionfacts.pdf. (Accessed on 18th October 2016).

[25] AHIRE, E. P. R., & MULAY, P. (2016). DISCOVER COMPATIBILITY: MACHINE LEARNING WAY. *Journal of Theoretical and Applied Information Technology*, *86*(3).

This document was created with Win2PDF available at http://www.win2pdf.com.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.