IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ROAD ACCIDENT ANALYSIS USING MACHINE LEARNING

Maneesha Kumari[1], Prof. Vivek Rai[2]

[1]M.Tech, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India

[2]Professors, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India

Abstract— In this research paper there are many inventories in in automobile industries to design and build safety measures for automobiles, but traffic accidents are unavoidable. There is a huge number of accidents prevailing in all urban and rural areas. Patterns involved with different circumstances can be detected by developing an accurate prediction models which will be capable of automatic separation of various accidental scenarios .These cluster will be useful to prevent accidents and develop safety measures. We believe to acquire maximum possibilities of accident reduction using low budget resources by using some scientific measures. There is a huge impact on the society due to traffic accidents where there is a great costs of fatalities and injuries. In recent years, there is a increase in the researches attention to determine the significantly affect the severity of the drivers injuries which is caused due to the road accidents. Accurate and comprehensive accident records are the basis of accident analysis .the effective use of accident records depends on some factors, like the accuracy of the data, record retention, and data analysis.

Keywords— automobile industries, traffic accidents, safety measures, urban and rural areas, drivers injuries.

I. INTRODUCTION

One of the most complicated and difficult daily needs is overland transportation. In India, more than 150,000 people are killed each year in traffic accidents. That's about 400 fatalities a day and far higher than developed auto markets like the US, which in 2016 logged about 40,000. Every year over 1 million vehicles are added to traffic averagely. 1.2 million People have died and over 50 million people have been injured in road accidents in the world every year. Studies on traffic have executed that road accidents and death- laceration ratio will increase. Design and control of traffic by advanced systems come in view as the important need. Assumption on the risks in traffic and the regulations and interventions in the end of these assumptions will reduce the road accidents.

An assumption system which will be prepared with available data and new risks will be advantageous. Data mining concept had been come up with by increasing and storage of data in the digital stage. Data mining involves the studies which will discover information from systematic and purposeful data structures obtained from disordered and meaningless data. Machine learning which is sub-branch of artificial intelligence supplies learning of computer taking advantage of data warehouses. Assumption abilities of computer systems have advanced in the event of machine learning. Utilization of machine learning is a widespread and functional method for taking authentic decisions by using experience. Machine learning is able to attain extract information from data and use statistical method. accidents have a great impact on the society. In recent years, researchers have paid increasing attention to determining factors that significantly affect severity of driver injuries caused by traffic accidents [29][30].

There are several approaches that researchers have employed to study this problem. These include neural network, nesting logic formulation, log-linear model, fuzzy ART maps and so on. Applying data mining techniques to model traffic accident data records can help to understand the characteristics of drivers' behaviour, roadway condition and weather condition that were causally connected with different injury severity. This can help decision makers to formulate better traffic safety control policies. Roh et al. [22] illustrated how statistical methods based on directed graphs, constructed over data for the recent period, may be useful in modelling traffic fatalities by comparing models specified using directed graphs to a model, based on out-of-sample forecasts, originally developed by Peltzman [23]. The directed graphs model outperformed Peltzman's model in root mean squared forecast error.

In this research paper section I contains the introduction, section II contains the literature review details, section III contains the details about existing system, section IV contains the proposed system details, section V shows architecture details, section VI provide data flow diagram details, section VII contains implementation details, section VIII describe the algorithm details, section IX provide result details and section X provide conclusion of this research paper.

II. LITERATURE REVIEW

Ossenbruggen et al. [24] used a logistic regression model to identify statistically significant factors that predict the probabilities of crashes and injury crashes aiming at using these models to perform a risk assessment of a given region. These models were functions of factors that describe a site by its land use activity, roadside design, use of traffic control devices and traffic exposure. Their study illustrated that village sites are less hazardous than residential and shopping sites.

Abdalla et al. [25] studied the relationship between casualty frequencies and the distance of the accidents from the zones of residence. As might have been anticipated, the casualty frequencies were higher nearer to the zones of residence, possibly due to higher exposure. The study revealed that the casualty rates amongst residents from areas classified as relatively deprived were significantly higher than those from relatively affluent areas.

Miaou et al. [26] studied the statistical properties of four regression models: two conventional linear regression models and two Poisson regression models in terms of their ability to model vehicle accidents and highway geometric design relationships. Roadway and truck accident data from the Highway Safety Information System (HSIS) have been employed to illustrate the use and the limitations of these models. It was demonstrated that the conventional linear regression models lack the distributional property to describe adequately random, discrete, nonnegative, and typically sporadic vehicle accident events on the road. The Poisson regression models, on the other hand, possess most of the desirable statistical properties in developing the relationships.

Abdelwahab et al. studied the 1997 accident data for the Central Florida area [2]. The analysis focused on vehicle accidents that occurred at signalized intersections. The injury severity was divided into three classes: no injury, possible injury and disabling injury. They compared the performance of Multi-layered Perceptron (MLP) and Fuzzy ARTMAP, and found that the MLP classification accuracy is higher than the Fuzzy ARTMAP. Levenberg-Marquardt algorithm was used for the MLP training and achieved 65.6 and 60.4 percent classification accuracy for the training and testing phases, respectively. The Fuzzy ARTMAP achieved a classification accuracy of 56.1 percent.

Yang et al. used neural network approach to detect safer driving patterns that have less chances of causing death and injury when a car crash occurs [17]. They performed the Cramer's V Coefficient test [18] to identify significant variables that cause injury to reduce the dimensions of the data. Then, they applied data transformation method with a frequency-based scheme to transform categorical codes into numerical values. They used the Critical Analysis Reporting Environment (CARE) system, which was developed at the University of Alabama, using a Backpropagation (BP) neural network. They used the 1997 Alabama interstate alcoholrelated data, and further studied the weights on the trained network to obtain a set of controllable cause variables that are likely causing the injury during a crash. The target variable in their study had two classes: injury and non-injury, in which injury class included fatalities. They found that by

controlling a single variable (such as the driving speed, or the light conditions) they potentially could reduce fatalities and injuries by up to 40%.

Sohn et al. applied data fusion, ensemble and clustering to improve the accuracy of individual classifiers for two categories of severity (bodily injury and property damage) of road traffic accidents [15]. The individual classifiers used were neural network and decision tree. They applied a clustering algorithm to the dataset to divide it into subsets, and then used each subset of data to train the classifiers. They found that classification based on clustering works better if the variation in observations is relatively large as in Korean road traffic accident data.

Mussone et al. used neural networks to analyze vehicle accident that occurred at intersections in Milan, Italy [12]. They chose feed-forward MLP using BP learning. The model had 10 input nodes for eight variables (day or night, traffic flows circulating in the intersection, number of virtual conflict points, number of real conflict points, type of intersection, accident type, road surface condition, and weather conditions). The output node was called an accident index and was calculated as the ratio between the number of accidents for a given intersection and the number of accidents at the most dangerous intersection. Results showed that the highest accident index for running over of pedestrian occurs at non-signalized intersections at nighttime.

III. EXISTING SYSTEM

The existing system provides little information on the number of accidents and the number of casualties. The casualty information at present is available for two injury levels, death and injuries. The police of each governorate are supposed to report accidents and casualties to the police headquarters in monthly reports. The police headquarters is responsible for reporting the data to the Central Statistics Organisation (CSO) in the Ministry of Planning. This organisation is responsible for producing the official statistics on road accidents.There is no specific form for collecting road accident data. The common way of reporting the accident is through narrative reports at all levels (i.e., from the policeman on the site of the accident to the police of the area or governorate, from hospitals to the police and from the police of the governorate to police headquarters). The police headquarters are responsible for extracting the information from the narrative reports and putting it in tabular form. It should be clear from the forgoing description that the existing Yemeni information system for road accident data is inadequate. The desired qualities of information can only partly be found in the existing system. The collected data suffer from deficiencies in both quantity and quality.

IV. PROPOSED SYSTEM

Models are created using accident data records which can help to understand the characteristics of many features like drivers behavior, roadway conditions, light condition, weather conditions and so on. This can help the users to compute the safety measures which is useful to avoid accidents. It can be illustrated how statistical method based on directed graphs, by comparing two scenarios based on out-of-sample forecasts. the model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injury that can be used to perform a risk factor and reduce it .Here the road accident study is done by analyzing some data by giving some queries which is relevant to the study. The queries like what is the most dangerous time to drive, what fractions of accidents occur in rural, urban and other areas What is the trend in the number of accidents that occur each year ,do accidents in high speed limit areas have more casualties and so on. These data can be accessed using Microsoft excel sheet and the required answer can be obtained. This analysis aims to highlight the data of the most importance in a road traffic accident and allow predictions to be made. The results from this methodology can be seen in the next section of the report.

V. ARCHITECTURE

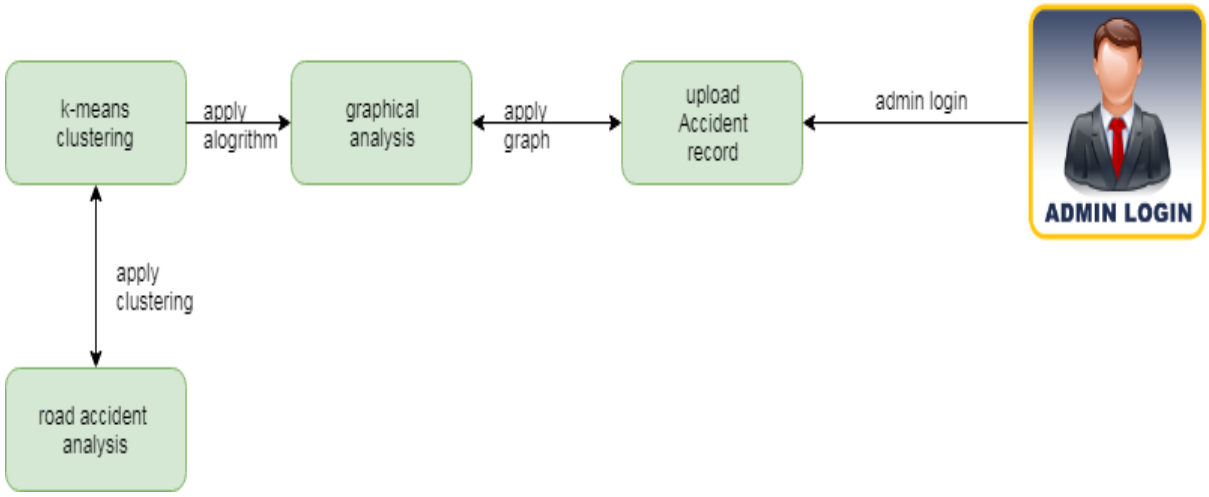

Figure 1: Architecture

VI. DATA FLOW DIAGRAM

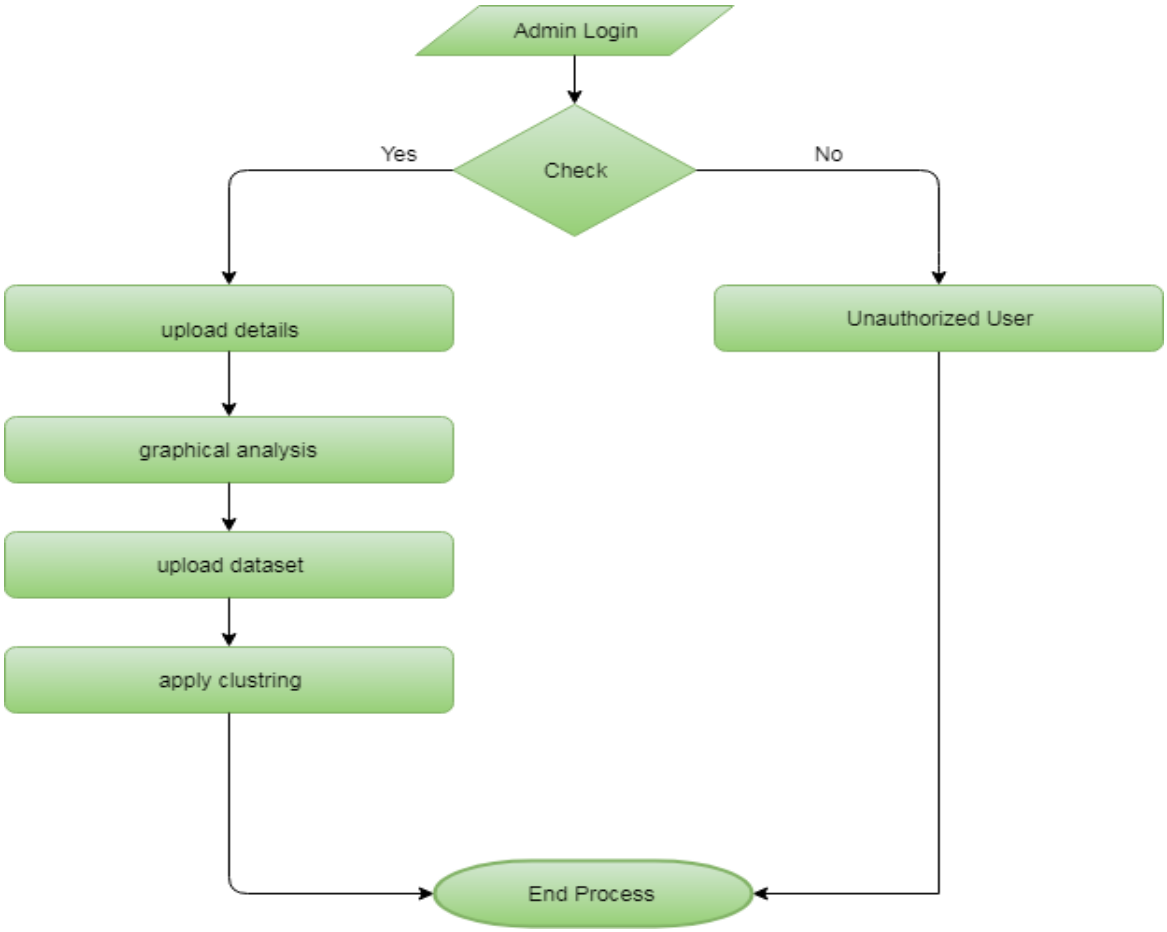

Figure 2: Data Flow Diagram

VII. IMPLEMENTATION

- **Admin Login**

Admin view, updates, delete customer and accident records .admin view update accident record. If any accident will constantly not good then admin can analysis accident.

- **Graph**

The analyses of proposed systems are calculated based on the approvals and disapprovals. This can be measured with the help of graphical notations such as pie chart, bar chart and line chart. The data can be given in a dynamical data.

VIII. ALGORITHM

- **k-means clustering algorithm**

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more.

IX. RESULT

The dataset used in the project to predict road accidents is based on values, and some of the data is written in plain English. Because of this, the data's numerical values are easy to predict and easy to calculate; however, the normal words are shown as they are or the data that cannot be predicted are dropped into the table.

Since there are a lot of columns and rows in this dataset, the forward fill method and the classification algorithm will be used to fill in all of the null values. The k-means clustering algorithm will be used in this classification algorithm.

1	1	1	30	2	1	10/8/2014	1	5 p.m.	1	1	3	2	27.215251	77.492786
4	2	2	30	2	3	8/8/2014	6	6:53 p.m.	1	1	3	2	11.933812	79.829792
3	1	2	30	1	1	9/8/2014	7	1:58 p.m.	1	1	3	2	29.691971	76.984483
2	1	2	30	2	1	9/8/2014	7	12:20 a.m.	1	1	3	2	8.177313	77.43437
3	1	1	60	2	1	10/8/2014	1	11 a.m.	1	1	3	1	10.785233	79.139093
4	1	1	70	2	1	10/8/2014	1	1:35 p.m.	1	1	3	2	25.775125	73.320611
4	1	1	30	1	1	10/8/2014	1	7 p.m.	1	1	3	1	23.836049	91.279386
4	1	2	20	2	1	11/8/2014	2	8:34 a.m.	1	1	3	1	15.503565	80.044541
4	1	2	30	1	1	8/8/2014	6	12:20 a.m.	1	1	3	1	19.798254	85.824938
4	1	1	30	2	1	12/8/2014	3	noon	1	1	3	2	10.362853	77.975827
4	1	2	30	1	1	8/8/2014	6	6:01 p.m.	1	1	3	1	22.025278	88.058333
4	2	2	30	2	2	6/8/2014	4	5:30 a.m.	1	1	2	1	28.403922	77.857731
4	2	2	30	2	2	2/9/2014	3	7:27 a.m.	1	1	3	2	25.776703	87.473655
4	1	2	30	1	1	3/9/2014	4	1:40 p.m.	1	1	3	2	14.7502	78.548129
4	1	2	30	2	1	3/9/2014	4	5:57 p.m.	1	1	3	2	28.460105	77.026352
3	1	2	30	2	1	5/9/2014	6	1:20 p.m.	1	1	3	2	21.273716	76.117376
2	1	1	30	2	1	5/9/2014	6	10:11 p.m.	2	1	3	2	16.187466	81.13888
2	1	2	30	1	1	6/9/2014	7	11:30 a.m.	2	1	3	2	28.793044	76.13968
2	1	2	30	1	1	6/9/2014	7	4:05 p.m.	2	2	3	2	15.477994	78.483605
2	1	2	40	1	1	6/9/2014	7	12:50 p.m.	2	1	2	1	21.043649	75.785058
2	1	2	30	1	1	5/9/2014	6	1:17 p.m.	2	2	3	2	27.598203	81.694709
2	1	1	30	3	1	8/9/2014	2	8:50 a.m.	2	2	3	1	26.168672	75.786111
4	1	2	30	2	1	9/9/2014	3	10:30 p.m.	2	1	3	2	29.534893	75.028981
2	1	2	30	2	1	9/9/2014	3	8:35 p.m.	2	2	3	2	18.11329	83.397743
2	1	2	30	2	1	10/9/2014	4	5:55 p.m.	1	1	3	2	12.905769	79.137104
2	1	1	40	2	1	10/9/2014	4	6:35 p.m.	1	1	3	2	9.494647	76.331108

Figure-3: Data set page

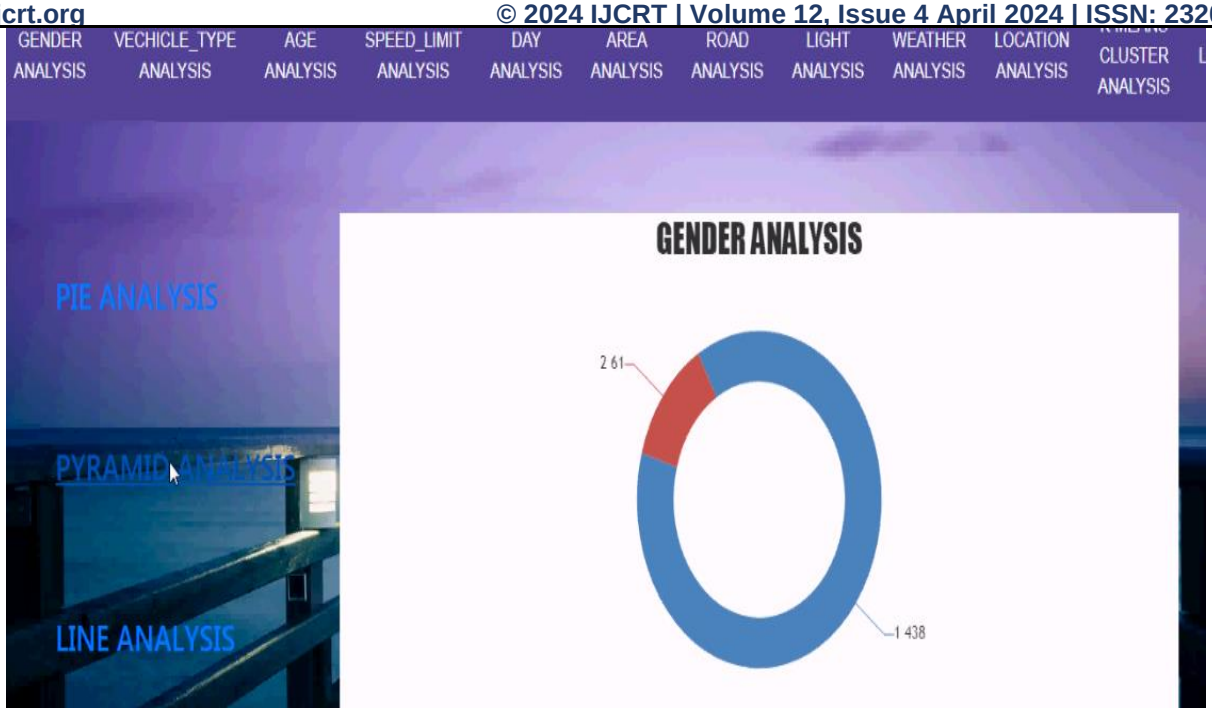

Figure-4: Graph for gender analysis

Figure-5: Graph for vehicle type analysis

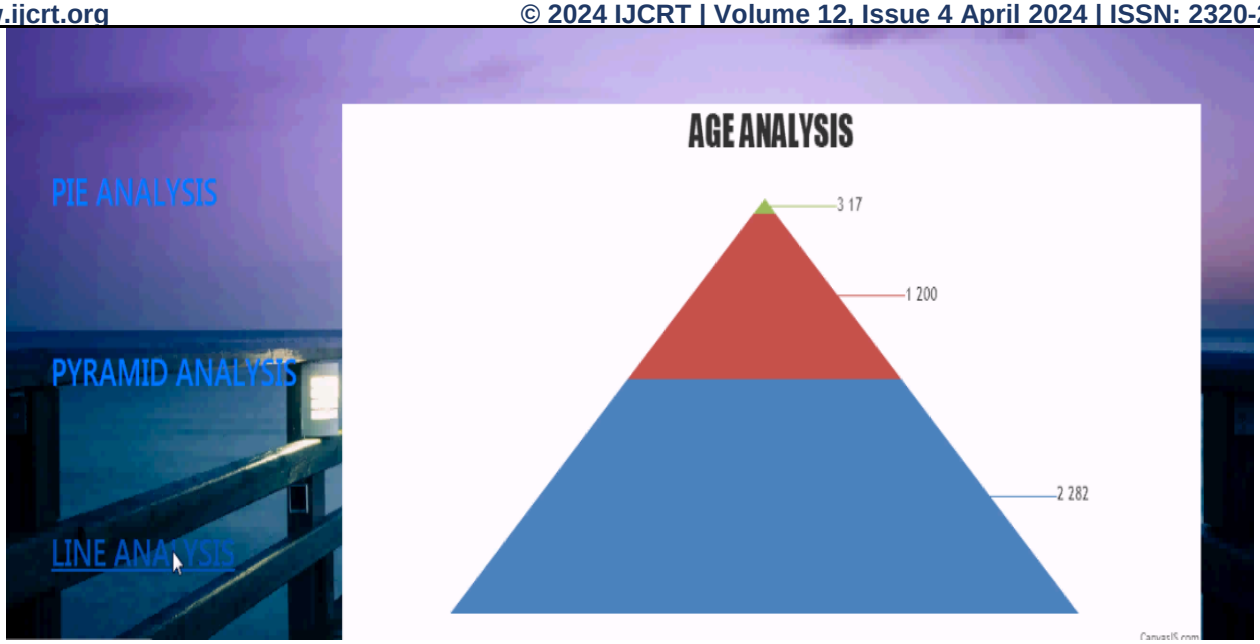

Figure-6: Graph for age analysis

X. CONCLUSION

In this research paper based approaches to predicting drivers' injury severity in headon front impact point collisions. The classification accuracy obtained in our experiments reveals that, for the non-incapacitating injury, the incapacitating injury, and the fatal injury classes, the hybrid approach performed better than neural network, decision trees and support vector machines. For the no injury and the possible injury classes, the hybrid approach performed better than neural network. The no injury and the possible injury classes could be best modeled directly by k means clustering. Past research focused mainly on distinguishing between no-injury and injury (including fatality) classes. It is well known that one of the very important factors causing different injury level is the actual speed that the vehicle was going when the accident happened.

REFERENCES

[1] Abdel-Aty, M., and Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles' Configuration and Compatibility. *Accident Analysis and Prevention*, 2003.

[2] Abdelwahab, H. T. and Abdel-Aty, M. A., Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. *Transportation Research Record 1746*, Paper No. 01-2234.

[3] Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. *Accident analysis and Prevention*, Vol. 34, pp. 717-727, 2002.

[4] Buzeman, D. G., Viano, D. C., & Lovsund, P., Car Occupant Safety in Frontal Crashes: A Parameter Study of Vehicle Mass, Impact Speed, and Inherent Vehicle Protection. *Accident Analysis and Prevention*, Vol. 30, No. 6, pp. 713-722, 1998.

[5] Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *Transportation Research C*, Vol. 5, No. 5, 1997, pp. 313-331.

[6] Evanco, W. M., The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities. Accident Analysis and Prevention, Vol. 31, 1999, pp. 455-462.

[7] Hand, D., Mannila, H., & Smyth, P., Principles of Data Mining. The MIT Press, 2001.

[8] Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention*,
Vol. 27, No. 4, 1995, pp. 469-481.

[9] Kweon, Y. J., & Kockelman, D. M., Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. Accident Analysis and Prevention, Vol. 35, 2003, pp. 441-450.

[10] Martin, P. G., Crandall, J. R., & Pilkey, W. D., Injury Trends of Passenger Car Drivers In the USA. Accident Analysis and Prevention, Vol. 32, 2000, pp. 541-557.
[11] Mayhew, D. R., Ferguson, S. A., Desmond, K. J., & Simpson, G. M., Trends In Fatal Crashes Involving Female Drivers, 1975-1998. Accident Analysis and Prevention, Vol. 35, 2003, pp. 407-415.
[12] Mussone, L., Ferrari, A., & Oneta, M., An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, Vol. 31,
1999, pp. 705-718.
[13] Ossiander, E. M., & Cummings, P., Freeway speed limits and Traffic Fatalities in Washington State. Accident Analysis and Prevention, Vol. 34, 2002,
pp. 13-18.
[14] Shankar, V., Mannering, F., & Barfield, W., Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention*, Vol.
28, No. 3, 1996, pp.391-401.
[15] Sohn, S. Y., & Lee, S. H., Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science*, Vol. 4, issue1, February 2003, pp. 1-14.
[16] Tavris, D. R., Kuhn, E. M, & Layde, P. M., Age and Gender Patterns In Motor Vehicle Crash injuries: Improtance of Type of Crash and Occupant Role. Accident Analysis and Prevention, Vol. 33, 2001, pp. 167-172.
[17] Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns By A Neural Network Approach. *ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining*, Vol. 9, pp 839-844, Nov. 1999.
[18] Zembowicz, R. and Zytkow, J. M., 1996. From Contingency Tables to Various Forms of Knowledge in Database. Advances in knowledge Discovery and Data Mining, editors, Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. AAAI Press/The MIT Press, pp.329-349.
[19] Abraham, A., Meta-Learning Evolutionary Artificial Neural Networks, Neurocomputing Journal, Elsevier Science, Netherlands, Vol. 56c, pp. 1-38, 2004.
[20] Moller, A.F., A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural
Networks, Volume (6), pp. 525-533, 1993.
[21] National Center for Statistics and Analysis http://www-nrd.nhtsa.dot.gov/departments/nrd-30/ncsa/NASS.html
[22] Roh J.W., Bessler D.A. and Gilbert R.F.**,** Traffic fatalities, Peltzman's model, and directed graphs**,** Accident Analysis & Prevention, Volume 31, Issues 1-2, pp. 55-61, 1998.
[23] Peltzman, S., The effects of automobile safety regulation. Journal of Political Economy 83, pp. 677–725, 1975.
[24] Ossenbruggen, P.J., Pendharkar, J. and Ivan, J., Roadway safety in rural and small urbanized areas. Accid. Anal. Prev. 33 4, pp. 485–498, 2001.
[25] Abdalla, I.M., Robert, R., Derek, B. and McGuicagan, D.R.D., An investigation into the
relationships between area social characteristics and road accident casualties. Accid. Anal. Prev. 29 5, pp. 583–593, 1997.
[26] Miaou, S.P. and Harry, L., Modeling vehicle accidents and highway geometric design
relationships. Accid. Anal. Prev. 25 6, pp. 689–709, 1993.
[27] SVMlight. http://www.cs.cornell.edu/People/tj/svm_light/. Access date: May, 2003.
[28] Vapnik, V. N., The Nature of Statistical Learning Theory. Springer, 1995.
[29] Chong M., Abraham A., Paprzycki M., Traffic Accident Data Mining Using Machine Learning Paradigms, Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary, ISBN 9637154302, pp. 415- 420, 2004.
[30] Chong M., Abraham A., Paprzycki M., Traffic Accident Analysis Using Decision Trees and Neural Networks, IADIS International Conference on Applied Computing, Portugal, IADIS Press, Nuno Guimarães and Pedro Isaías (Eds.), ISBN: 9729894736, Volume 2, pp. 39-42, 2004.
[31] Eui-Hong (Sam) Han, Shashi Shekhar, Vipin Kumar, M. Ganesh, Jaideep Srivastava, Search
Framework for Mining Classification Decision Trees, 1996. umn.edu/dept/users/kumar/dmclass.ps
[32] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.
[33] Abraham, Intelligent Systems: Architectures and Perspectives, Recent Advances in Intelligent Paradigms and Applications, Abraham A., Jain L. and Kacprzyk J. (Eds.), Studies in Fuzziness and Soft Computing, Springer Verlag Germany, Chapter 1, pp. 1-35, 2002.