Available Online at www.ijcsmc.com/journal

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 11, Issue. 3, March 2022, pg.83 – 94

TRAFFIC ACCIDENT PREDICTION USING MACHINE LEARNING

Ummadi Swathi; T.Lakshmi Prasanna; Shyam Mohan J S
Department of Computer Science and Engineering
Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University, Kanchipuram, Tamilnadu
ummadiswathireddy3@gmail.com, t.lakshmiprasanna2000@gmail.com;
jsshyammohan@kanchiuniv.ac.in
DOI: https://doi.org/10.47760/ijcsmc.2022.v11i03.010

Abstract: This report presents the results from the research study on applying large scale data mining methods into analysis of traffic accidents on the finnish roads. The data sets collected from traffic fatal accidents are huge, multidimensional, and heterogeneous. Moreover, they may contain incomplete and erroneous values, which make its exploration and understanding a very demanding task. The target data of this study was collected by the finnish road administration datasets. The intention is to investigate the usability of robust clustering, association and frequent itemsets, and visualization methods to the road traffic accident analysis. While the results show that the selected data mining methods are able to produce understandable patterns from the data, finding more fertilized information could be enhanced with more detailed and comprehensive data sets. Machine learning algorithm takes accident frequency count as a parameter to cluster the locations. Then we used association rule mining to characterize these surface condition. The rules revealed different factors associated with road accidents at different drunk and drive with varying accident frequencies. The association rules for high-frequency accident location disclosed that intersections on highways are more dangerous for every type of fatal accidents.
Keywords: Traffic accident, CNN, Safety Measures, Data mining, Fatal accidents.

INTRODUCTION:

There are a lot a vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accidents involves fatality, means people die in that accident. As a human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous decades. Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent itemset mining. A classical association rule mining method is the Apriori algorithm who main task is to find frequent itemsets, which is the method we use to analyze the roadway traffic data. Classification in data mining methodology aims at constructing a model from a training data set that can be used to classify records of unknown class labels. The naive Bayes technique is one of the very basic probability based methods for classification that is based on the Bayes hypothesis with the presumption of independent between each pair of variables. We used the FARS dataset for our study. The fatal accidents dataset contains all fatal accidents on public roads in 2007 reported to the national highway transportation safety administration. The dataset is downloaded from California polytechnic state university and all data originally came from FARS. The dataset contains 37,248 records and 55 attributes. The data description can be found in the document FARS.

Scope of the Project: Road traffic injury is a major global public problem. Rapid motorization in low and middle-income countries along with the poor safety quality of road traffic systems and the lack of institutional capacity to manage outcomes contribute to a growing crisis.

More than 1.24 million people die each year on the worlds roads. Many more suffer permanent disability, and between 20 and 50 million suffer non-fatal injuries. These are mainly in amongst vulnerable road users and involve the most social-economically active citizens.

LITERATURE SURVEY:

Title 1: Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women

Author: Divay Bansal and Lekha Bhambhu

Apriori algorithm is the most popular and useful algorithm of association rule mining of data mining. As association rule of data mining is used in all real life applications of business and industry. Objective of taking Apriori is to find frequent itemsets and to uncover the hidden information. This paper elaborates upon the use of association rule mining in extracting patterns that occur frequently within a dataset and showcases the implementation of the Apriori algorithm in mining association rules from a dataset containing crimes data concerning women. As for this WEKA tool is used for extracting results. For this one dataset is taken from UCI repository and other data is collected manually from the session court of sirsa to collect data on heart melting crimes against women. The main motive is use UCI is to first check the proper working of dataset and then apply Apriori on real dataset against crimes on women which extracts hidden information that what age group is responsible for this and to find where the real culprit is hiding. In last the comparison is done between Apriori and predictive apriori algorithm in which apriori is better and faster than predictive apriori algorithm.

Title 2: Applying association rules mining algorithm for traffic accidents in Dubai.

Author: Mira A EI Tayeb,Vikas Pareek,and Abdelaziz Araar

Association rule mining algorithms are widely used to find all rules in the database satisfying some minimum support and minimum confidence constrains. In order to decrease the number of generated rules, the adaptation of the association rule mining algorithm to mine only a particular subset of association rules where the classification class attribute is assigned to the right-hand-side was investigated in past research. In this research, a dataset about traffic accidents was collected from Dubai traffic department, UAE. After data prepossessing, Apriori

and predictive Apriori association rules algorithms were applied to the dataset in order to explore the link between recorded accidents factors to accident severity in Dubai.

Title 3: A Perspective analysis of traffic accident using data mining techniques.

Author: S.Krishnaveni and M.Hemalatha

Data mining is taking out of hidden patterns from huge database.it is commonly used in a marketing, surveillance, fraud detection and scientific discovery. In data mining, machine learning is mainly focused as research which is automatically learnt to recognize complex patterns and make intelligence decisions based on data. Nowadays traffic accidents are the major causes of death and injuries in this world. Roadway patterns are useful in the development of traffic safety control policy. This papers deals with the some of classification models to predict the severity of injury that occurred during traffic accidents.

Title 4: Analysis road accident data using association rule mining.

Author: Sachin Kumar and Durga Toshniwal

Road accident is one of the crucial areas of research in India. A variety of research has been done on data collected through police records covering a limited portion of highways. The analysis of such data can only reveal information regarding that portion only; but accidents are scattered not only on highways but also on local roads. A different source of road accident data in India is emergency management research institute(EMRI) which serves and keeps track of every accident record on every type of road and cover information of entire states road accidents. The results can be utilized to put some accident prevention efforts in the areas identified for different categories of accidents to overcome the number of accidents.

Title 5: Extracting Hidden patterns within road accident data using machine learning techniques.

Author: KMA Solaiman, Md Mustafizur Rahman and Nashid shahrair, Avra

Road accident its may no best oppida together, but can be reduced. Driver emotions such as sad, happy, and anger can be one reason for accidents. At the same time, environment conditions such as weather, traffic on the road, load in the vehicle, type of the road, health conditions of driver, and speed can also be the reasons of accidents. Hidden patterns in accidents can be extracted so as to find the common features between accidents. These paper presents the results of the framework from the research study on the road accident data of major national highways. These datasets collected from police stations are heterogeneous. Incomplete and erroneous values are corrected using data cleaning measures, and relevance attributes are identified using attributes selection measures.

EXISTING SYSTEM:

The traffic accident using data mining technique that could possibly reduce the fatality rate. Using a road safety database enables to reduce the fatality by implementation road safety programs at local and national levels. Classification models to predict the severity of injury that occurred during traffic accidents. Association rules mining algorithm on a dataset about traffic accidents which was gathered from government traffic office, Apriori and predictive apriori association rules algorithm were applied to dataset to investigate the connection between recorded accidents and factors to accident severity.

Disadvantages:

- Here we are using data mining the fault traffic accident injuries and deaths.
- In this the cost of maintain and repairing the roads.
- This will not useful for short distance level.

PROPOSED SYSTEM:

This paper presents our research to model the severity of injury resulting from traffic accidents using artificial neural networks and decision trees. We have applied them to an actual data set obtained from the National Automotive Sampling System(NASS) General Estimates System(GES).Experiment results reveal that in all the cases the decision tree outperforms the neural network. Our research analysis also shows that the three most important factors in fatal injury are: drivers seat belt usage, light condition on the roadway, and drivers alcohol usage. Our experiments also showed that the model for fatal and non-fatal injury performed better than other classes. The ability of predicting fatal and non-fatal injury is very important since drivers family has the highest cost to society economically and socially.

Advantages of Proposed System:

- Speed was available.
- Improve the performance of analysis in fatal and non-fatal accidents.

ALGORITHM: K-means clustering is a simple unsupervised learning algorithm that is used to solve clustering problems. It follows a simple procedure of classifying a given data set into a number of clusters, defined by the letter ‘k’, which is fixed beforehand. The clusters are then positioned as points and all observations or data points are associated with the nearest cluster, computed, adjusted and then the process starts over using the new adjustments until a desired result is reached.

K-means clustering has uses in search engines, market segmentation, statistics and even astronomy. It is used mainly in statistics and can be applied to almost any branch of study. For example, in marketing, it can be used to group different demographics of people into simple groups that make it easier for marketers to target. Astronomers use it to sift through huge amounts of astronomical data; since they cannot analyze each object one by one, they need a way to statistically find points of interest for observation and investigation.

SYSTEM ARCHITECTURE:

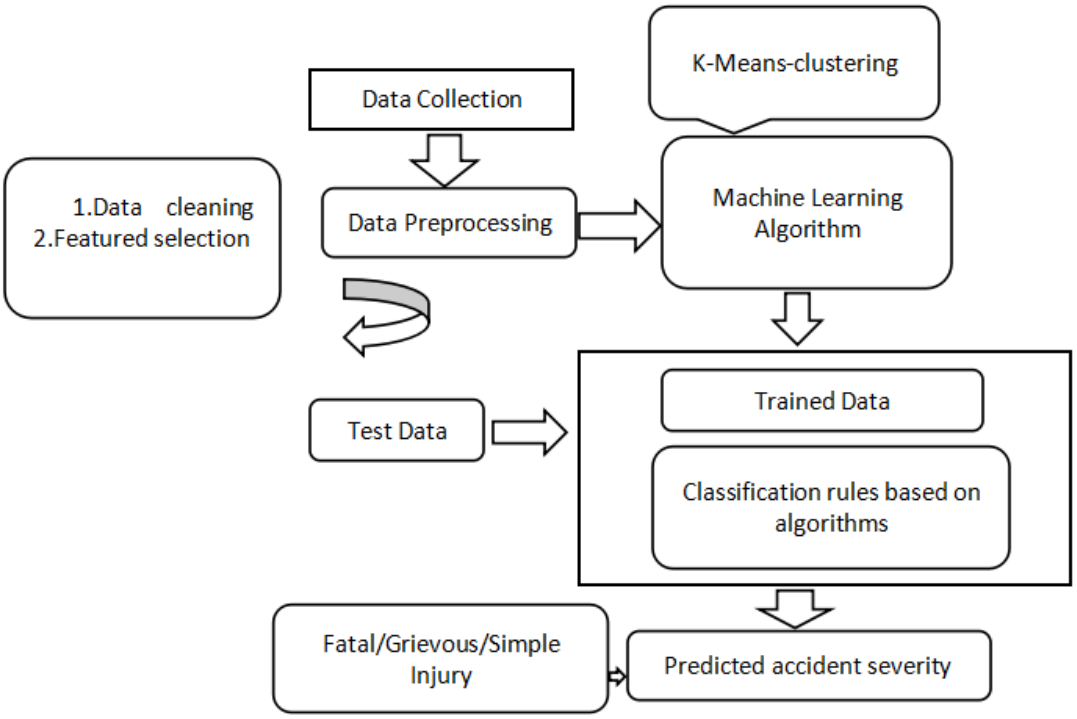

Fig 1. System Architecture

MODULE DESCRIOTION:

Modules:

- User Interface Module
- Data Collection Module
- Analysis Module
- Report Accident Module
- Prediction Module

User Interface Module: This Module provides user an interface to sign up or login into the system. Once the user is logged in he/she can select prediction or Analysis. In each of these modules the user can select

certain attributes and make a prediction or analysis. Once the user has completed the analysis or prediction he can logout.

Data Collection Module: In data collection module, accident data will be collected from kaggle. Such as year, month, state, time span.

Analysis Module: In this module the user specifies a particular attribute based on which an analysis with respect to its effect on accident can be determined. The analysis will be done using the training data set, support vector machine classifier and the attribute specified by the user. The output will in form of a graph.

Report Accident Module: The report page allows the user to interact with the site administrator by reporting the current state of victims and property involved in an accident as he/she travels along a particular route. This report is therefore sent to the incidence table of data base, where it is checked and validated by the administrator and then updated into the various fields of the data base for appropriate report generation. The roads users are allowed to also report the accident occurrences on a particular road at any point in time. The road user can also make reports even if he does not have an account.

Prediction Module: In prediction module, Traffic accident details will predict. It predict previous year and future year also it will predict based on year, month, state, and time span. Finally it will show the accuracy of traffic accident based on state.

RESULT:

Fig 2.Road accidents analysis and prediction

Fig 2 shows the home page of the road accidents analysis and prediction process. In this we have analysis and prediction and exit.

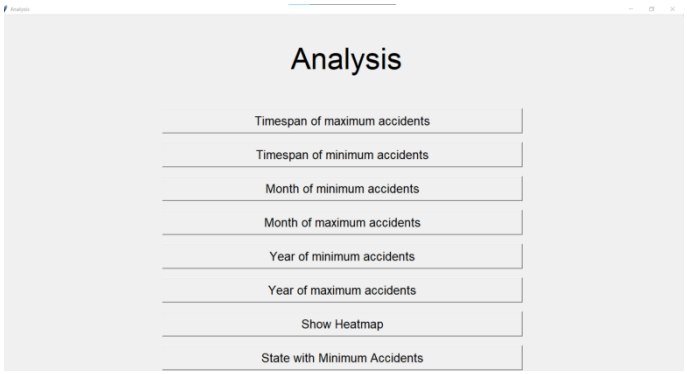

Fig.3 Accident Analysis

In fig 3 we can analysis the time span of max and min accidents, Month of max and min accidents, year of max and min accidents, heat map and state with max and min accidents.

Fig.4 month of minimum accidents

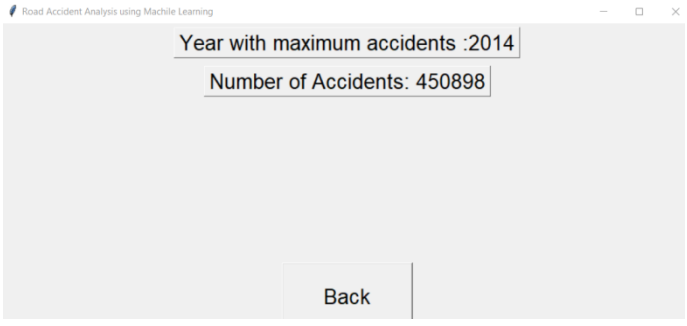

Fig.5 year of maximum accidents

In fig.4 displaying the minimum number of accidents occurred in a month.
In fig.5 displaying the maximum number of accidents occurred in a year.

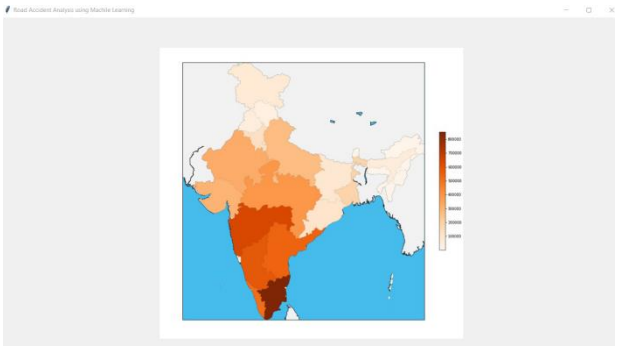

Fig.6 Heat map

In fig.6 It is displaying the heat map of the India about accidents.

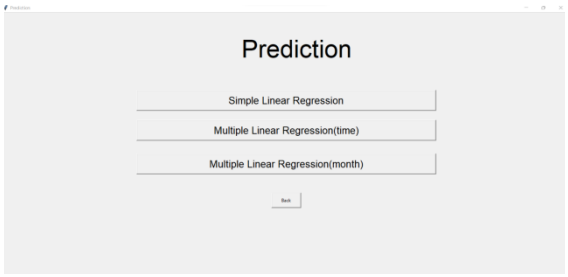

Fig.7 Accident Prediction

In fig.7 where we can predict simple liner regression, multiple linear regression based on time and months.

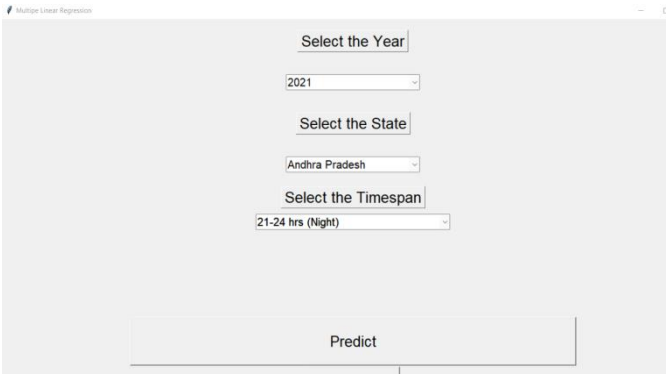

Fig.8 Multiple linear regression on time

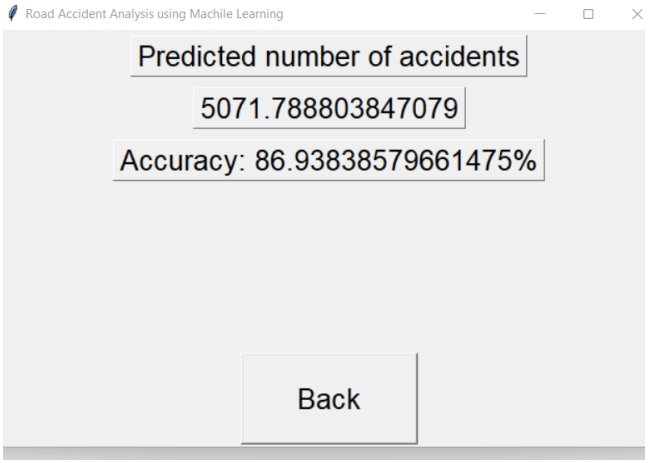

Fig.9 Predicted number of accidents

In fig.9 It will display the predicted number of accidents and accuracy.

CONCLUSION:

As seen in statistics, association rule mining, and the classification, the environmental factors like roadway surface, weather, and light condition do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have stronger affect on the fatal rate. From the clustering result we could see that some states/regions have higher fatal rate, while some others lower. We may pay more attention when driving within those risky states/regions. Through the task performed, we realized that data seems never to be enough to make a strong decision. If more data, like non-fatal accident data, weather data, mileage data, and so on are available, more test could be performed thus more suggestion could be made from the data.

REFERENCES

[1]. Divya Bansal and Lekha Bhambhu.Execution of apriori algorithm of data mining directed towards tumultuous crimes concerning women.International Journal of advanced research in computer science and software engineering,3(9),September 2013.

[2]. Amira A EI Tayeb,Vikas Pareek and Abdelaziz Araar.Applying association rules mining algorithms for traffic accidents in Dubai.International journal of soft computing and engineering,September 2015.

[3]. S.Krishnaveni and M.Hemalatha. A perspective analysis of traffic accident using data mining techniques.International journal of computer applications,23(70:40-48,June 2011.

[4]. Sachin Kumar and Durga Toshniwal. Analysing road accident data using association rule mining.In proceedings of international conference on computing,communication and security,pages1-6,2015.

[5]. KMA Solaiman,Ms Mustafizur Rahman,and Nashid Shahriar.Avra Bangladesh collection,analysis and visualization of road accident data in Bangladesh.In proceedings of international conference on informatics,Electronics and vision,pages1-6,IEEE,2013.
[6]. Trac Integrated SCM and Project management.Fatal accidents dataset. https://wiki.csc.calpoly.edu/datasets/wiki/ HighwayAccidents.
[7]. U.S. Census Bureau.Population Estimates. https://www.census.gov/popset/data/historical/2000s/vintage 2007/,2007.

