IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Road Accident Prediction Model Using Data Mining Techniques

Gone Ajay Kumar

School of computer science and engineering,

lovely professional university, Jalandhar, Punjab,

India

Pasala Kasi viswanath
School of computer science and engineering,
lovely professional university, Jalandhar, Punjab,
India

Mohammad Wamique
School of computer science and engineering,
lovely professional university, Jalandhar, Punjab,
India

Ishan Kumar

Assistant Professor

lovely professional university, Jalandhar, Punjab,

India

Abstract- Road accidents were always a calamity with a continually growing pattern. According to the Indian road safety campaign, a road accident happens every minute in India, more than 17 people die in road accidents every hour. There are several kinds of automotive incidents, including rear-end, head-on, and rollover collisions. State-recorded police reports, commonly termed as FIRs, are records that contain information regarding accidents. The people may report the occurrence themselves or the state police may record it. The frequency patterns of car crashes are estimated using Apriori and Nave Bayesian approaches in this paper. This pattern will aid the government or non-governmental organisations in improving road safety and taking prevention action in high density areas.

At this era, street mobility security is one of the most severe social issues in the world. For every death as a result, there are expected to be approximately four handicap injuries, like damage to the brain or neurological line, 10 genuine wounds, and 60 minor injuries. The rate of crashes occurring on a daily is also growing at an alarming rate owing to the exponentially growing number of automobiles on the road. With the rising traffic accidents and death these days, the transportation department's ability to predict the accident rate over a duration of time is critical

for making scientific decisions. In this case, it would be beneficial to investigate the causes of road accidents such that we can create methods to reduce them. Several stakeholders, but are not limited to government civil works departments, contractors, or other automobile industries, can gain from the results of this study in better designing highways and automobiles depending on the values predicted.

Keywords: Industries, Road traffic accidents, Government, Models that make accident predictions, Algorithm towards forecast, Patterns of data, Safety on the roads, Deaths.

Introduction:

Various research have looked into various elements of RTAs, with the majority of them focus on predicting or identifying the essential components that influence injury severity. Many data mining-related studies have been conducted to study RTA data locally and globally, with results vary widely based on the socioeconomic factors and technology of a particular region.

Various methods were employed to increase the accuracy of individual classifiers for two RTA intensity categories in order to investigate the association between RTA severity and operating surrounding parameters. Three alternative approaches have been used in neural and logistic regression individual classifiers: classifier fusion using the Participatory method, the Bayesian procedure, and the linear model; data ensemble fusion using sparking and dropping; and clustering using the k-means algorithm. However, it's among the world's biggest incidents, leading in death and physical damage. Identify key cause of road accidents will aid in the development of a suitable solution to reduce the negative impact of severity on people and damage to property. Severity on the road isn't accidental: it follows the pattern that can be foreseen and prevented. In a fraction of a second, human life and property were destroyed. It is one of the country's more frightening leading causes of mortality.

In the last couple of decades, one of the research areas in road safety has been the severity of RTAs. Just on road accident severity categorization based models, researchers used novel methodologies. The research looks at where to develop models using a standard statistical method. These methods aid in gaining insight into and identifying the underlying causes of automobile accidents and other issues that affect road safety. Machine learning now outperforms traditional statistical models in forecasting the model due to the large amount of available data.

There is, however, a scarcity of comparisons between state-of-the-art algorithms, Hybrid Machine Learning algorithms, and deep learning algorithms. Obtaining a suitable technique can make forecast accuracy more informative in some cases. As a result, selecting the best model aids in identifying important road accident elements. Furthermore, target-specific relevant aspects had not been discovered but was not a concern. To anticipate the seriousness of road accidents, the researchers used a combination of clustering and classification methods. Further, the suggested proposed method is compared to a deep learning network in order to compare it to other state-of-the-art classification techniques. Depending on categorization and performance metrics, the suggested classifier outperforms other classifications in the testing.

The purpose of this research is to determine the most relevant characteristics that affect the degree of injuries suffered by individuals involved in traffic incidents on these roadways, whereas by reducing or managing these factors, overall safety can be improved. They employed the CART method (Classification and Regression Tree).

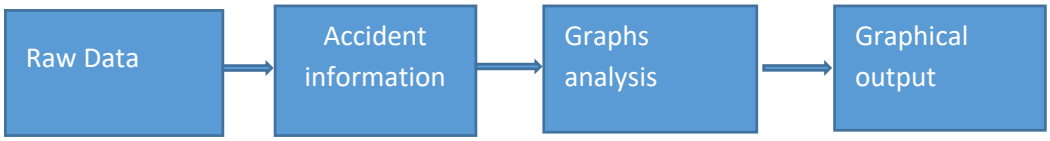

Fig 1. An Efficient Approach for the Recognition of data.

Literature review:

For the past few decades, traffic deaths have been the leading cause both injuries and deaths globally. When a road collides with some other vehicle, a person, an animal, or a geographical or physical obstruction, it is called a traffic collision. It also has the potential to cause damage, damage to property, and death. The place where essential data about society is gathered and preserved is the traffic control system. We can identify risk factors for car accidents, injuries, and fatalities using this data, and take precautions that can save lives. The intensity of injuries has societal consequences. Conventional statistical model-based strategies were utilised to forecast accident mortality and severity in the field of road safety. Classical statistical-based research include the mixed logit model based, ordered Probit model, and logit model. According to certain studies, the traditional statistical approach is more effective at detecting direct and indirect accident variables.

Data mining is a new and powerful tool that can help firms focus on most critical information in their database systems by extracting hidden prediction information from large databases. It's an useful tool for dealing with the requirement to move useful data from a database, such as hidden patterns.

Analysis of traffic injury severity: An example of how nonparametric classification tree approaches can be used: The goal of this study is to create the CART model that will be used to find connections among injuries and motorist characteristics, highway /environmental variables, and crash variables. They employed Logistic Regression and Back propagation Systems.

A Data Mining Approach to Identify Key Factory of Traffic Injury Severity: The purpose of this research is to determine the most relevant factors that influence overall degree of injuries sustained by drivers related to traffic incidents on these roads, so that by deleting or regulating these elements, overall safety can be improved. They employed using CART method (Classification and Regression Tree).

Mining Road Traffic Accident Data to Improve Safety in India: This research implemented data mining techniques to connect reported accident, driver, and roadway elements to accident severity in india, resulting in a total of guidelines that the indian Police might use to improve safety. They employ the Multilayer Perceptron and Bayesian Network methods.

Existing system:

Interface requirements- System permission is necessary for users at the start of the service. For all users, the login method is the same. They will provide a login and password that is legitimate or authorised. The user interface summary is detailed in general in the parts below.

Sign in- The user will be provided with a login screen whenever the Accident Analysis and Prediction System web address is opened. If an user has successfully registered in the system, he or she can log in using the username and password; if the user has not yet registered, the user should do so.

Upload data set- To evaluate or predict accidents based on parameters, the user should upload the data into the database server.

Data visualization- The system generates recommendations based on the uploaded data set, processes data, and predicts the outcomes. The proposal will just be displayed in the interface, as well as in the form of graphical visualisation.

Signout- When a people click the sign out button, the system's activities were stopped, and the user is routed to the login page.

Clustering Techniques:

A process of collecting data elements could be treated as a single entity. When undertaking clustering algorithm, we divide the data set into groups based on data similarity, and assign labels to the groups. Clustering has the advantage of being adaptive to changes & assisting in the identification of useful qualities that separate groups. Traffic accident data is now being collected in huge volumes. Multi - processor systems with a bunch of processing capacity. Various research have examined into various aspects of RTAs, with the majority of them focusing on anticipating or identifying the essential components that influence injury severity. Several database mining-related research have been conducted to evaluate RTA data locally and internationally, with results differing widely based on the economic status and technology of a given place.

Functional requirements:

A functional requirement describes a software program's or component's function. A function is made up of three parts: inputs, behaviour, and outputs. The application is loaded with numerous details and the heart illness that goes together with them. Users can use the app to share their cardio difficulties. It then examines the user's detailed info to see if there are any illnesses that could be linked to it. Here, we employ some advanced data mining techniques to determine a most accurate ailment that could be linked to the patient's information. They can contact a doctor for the further therapy based on the results. The system also allows users to view data on physicians. This system is used to provide free online cardiovascular disease consultation.

Hardware requirements:

Desktop/Laptop:	**OS: Windows 8 Minimum, Processor: Intel(R) Core (TM), i5-7400 CPU @ 3.00GHz, RAM: 4.00 GB, System type: 64-bit Operating System, X64-based processer, Full HD Display**
Mobile:	**Android Version: 4.4 KitKat higher, RAM: 3.00 GB, CPU: Octa-core Max 2.0GHz, Internal Storage: 6.00 GB, Mobile: Any Touch Mobile Accepted, Mobile Browser: MI Browser, Chrome, Firefox**
Tablet:	**RAM: 4 GB, Internal Storage: 128 GB, Processer: 2.30 GHz, Full HD Display, Wi-Fi+4G, Browser**

Software requirements:

Software Design / UI:	**Bootstrap based Web UI Kit**	**HTML5, CSS3, JS, jQuery, Node JS, Bootstrap, Grunt, Bower, SASS based solid framework.**
Software Development Kit	PhpMyAdmin version: 5.0.4	OS: Windows7/8.1/10, Database Server: MySql_V5.6.20, Web server Apache/2.4.10 (Win32), OpenSSL/1.0.1i PHP/5.5.15, Python
Browser:	Chrome, Firefox	Latest Versions of Mobile Bowsers with latest updates

It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification and the hardware requirements will help in the execution of the data into the database.

System architecture:

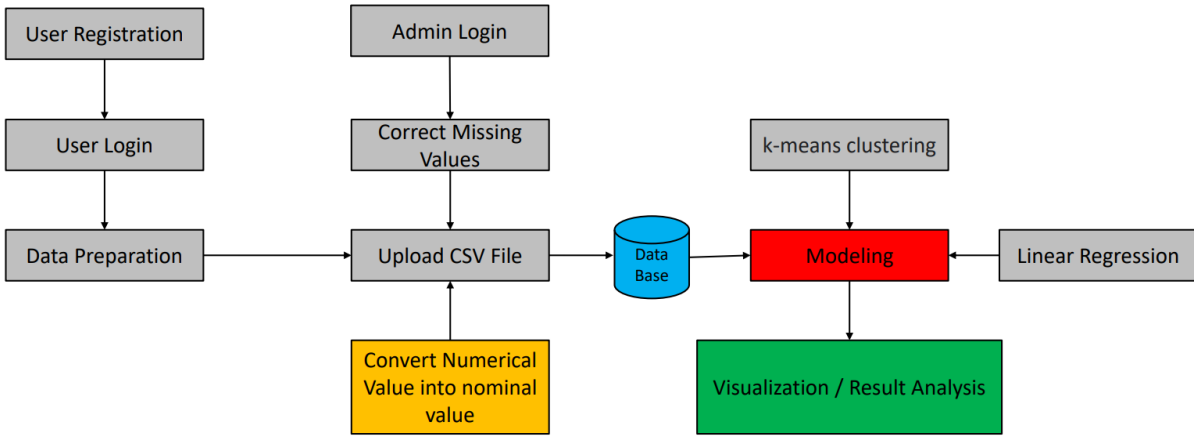

Fig. 2. The flow of the user to handle the data

A deployment diagram in the Unified Modeling Language shows how deploys are connected to construct larger deployment and/or software systems. They're used to show how the structure of extremely complex systems is represented. The user enters a primary enquiry, which is transformed into sub inquiries and sent to data collected

from various sources via data transmission. Data aggregators will display the data to the user. All of the boxes are components, and the arrows represent connections. Collecting sample data from the organisation is one of the more difficult aspects of this study. The initial dataset acquired from the authority was manually modified. The majority of the fields are blank. Most of the other fields are meaningless, and perhaps the most important fields are omitted from the manual document. It was impossible to read documents because they were invisible.

They're written poorly or in a panic.

Poison Progression

Although numerous regression analyses have been widely used, it has been shown that a Distribution can often better fit crash frequency. Another common blunder is to use an ordinary least regression to model crash data as time series. This strategy is ineffective since regression models can give non-integer projected values and also negative expected values, both of which are incompatible with data stream modelling. Moreover, many crash data distribution are positively skewed, with a large number of data points inside the tabular forms with a value of 0.0. A wide range of transport count data, includes crash frequency, has been subjected to Regression analysis. With two exceptions, a Poisson regression model is similar to a standard multiple linear regression. It starts by assuming that the mistakes have a Poisson (instead of a normal) distribution. Secondly, rather of modelling Y as a linear model of the linear regression, it models the natural log of Y, ln (Y), as a linear combination of the coefficients. The Poisson model can be written like this:

$$P(n_i) = \frac{\lambda i EXP(-\lambda i)}{n!}$$

Where,

P (n_i): the probability of n crashes occurring on a highway segment i.
n_i: the number of observations per time period (such as a year),
λ_i: the expected crash frequency on road segment i per time period

$$\lambda_i = EXP(\beta X_i)$$

Where,
X_i: a vector of the independent variables (i.e. risk factors),
β: a vector of the estimates (coefficients) of the independent variables X_i.

Factors affecting road traffic accidents:

Many factors, such as driver conduct, road layout, traffic volumes, vehicle, and surroundings, can all play a role in a traffic accident. Although the impact of such variables on crash occurrence may vary widely from case to case, both behavioral and non-behavioral factors such as road geometry, traffic flow conditions, vehicle, and environment are believed to have a substantial impact on traffic crashes. According to research, there are six key hazard factors that influence the likelihood of a road accident.

Driver behavior- Use of alcohol and drugs, careless driving, failure to correctly use occupant protection systems, mobile phone or texting use, and fatigue all are factors.

Vehicle Factors- vehicle type, as well as automotive engineering and safety design criteria. The design of windshield glass, for example, as well as the location and durability of gas tanks can all help improve safety. If employed, passenger protection devices in vehicles (such as air bags and seatbelts) can prevent or lessen injuries.

Roadways characteristics- well-designed curves and gradients, wide roads, enough line of sight, plainly visible striping, flaring guardrails, high quality shoulders, clear roadsides, well-located accident absorption devices, and well-planned use of traffic lights are all examples of road geometries and wayside conditions.

Traffic volumes- The annual average daily traffic (AADT), commonly referred as the number of miles travelled, is a measure of how much traffic travels through a city on a regular basis. The daily average traffic count is the number of cars that pass through a given location on a given day. The vehicle flow over a road section on an average day of the year is expressed by AADT. The distance travelled by automobiles on roadways is referred to as vehicle miles travelled. It's frequently used as a traffic indicator and also to assess mobility and transit patterns.

Environmental factor- weather conditions, and light conditions.

Time factors- the year's seasons, the month of the year, weekdays, and the hour of the crash

Preprocessing:

Raw datasets were terribly filthy, not in a format that computer machines could understand, and provided partial data to use as is. The effectiveness of the crash severity prediction model will be reduced if such datasets are used. As a result, irrelevant datasets should be eliminated in order to generate maximum data. Before designing the model, the researchers was using an expensive data preparation technique to obtain relevant and determinant potential risks, such as cleaning the data, missing value handling, outlier management, dealing with absolute value—encoding, and standardization.

Test Cases:

Sl. No	Test Scenario	User Action	Expected Result	Actual Result	Remarks
1.	Login	User greets hello.	The device greets back.	The device greets back successfully	Successful
2.	Query's	Users asks a query.	Query gets answered.	Query gets answered successfully.	Successful
3.	Play Video	User asks the device to play video.	video gets played.	video gets played successfully.	Successful
4.	Navigation	User asks to navigate to a page on the web.	The device opens the requested page on the web.	The device opens the requested page on the web successfully.	Successful
5.	Text Command	The device doesn't hear the user's voice properly.	The device takes the user's text command instead.	The device takes the user's text command instead successfully.	Successful

Table. 1. Test cases for the data flow.

Snapshots:

This project is implements like web application using PHP and the Server process is maintained using the socket & serversocket and the Design part is played by Cascading Style Sheet.

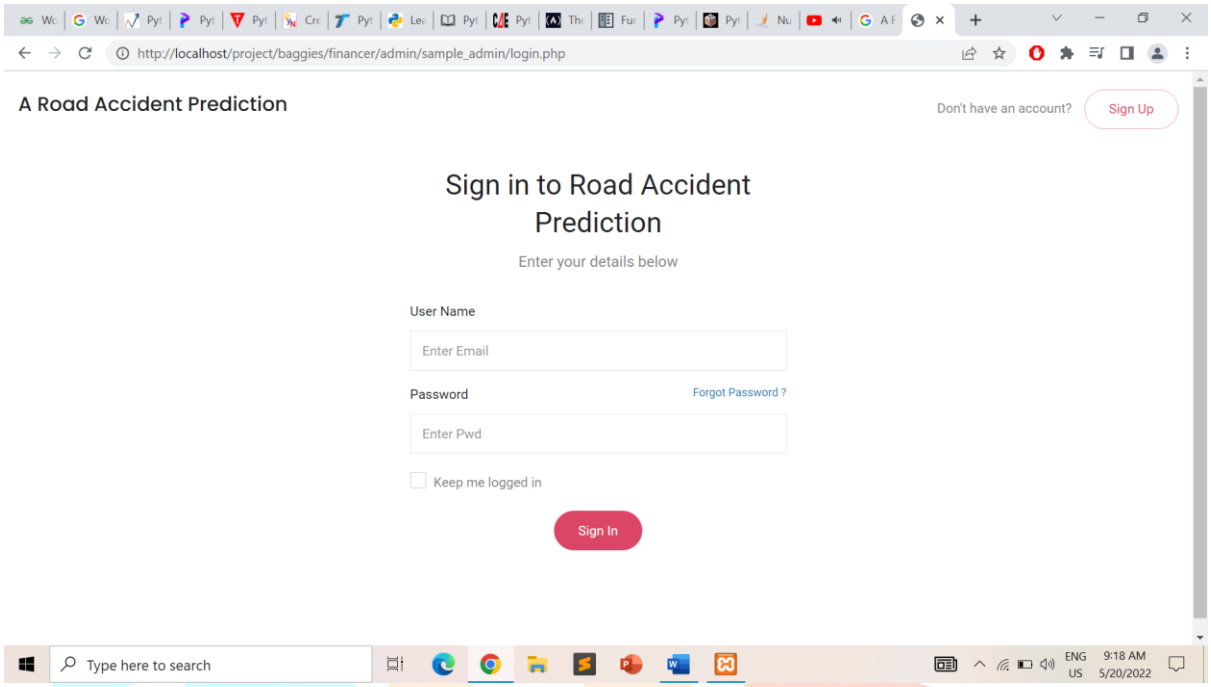

Fig. 3. Login information to road accident prediction

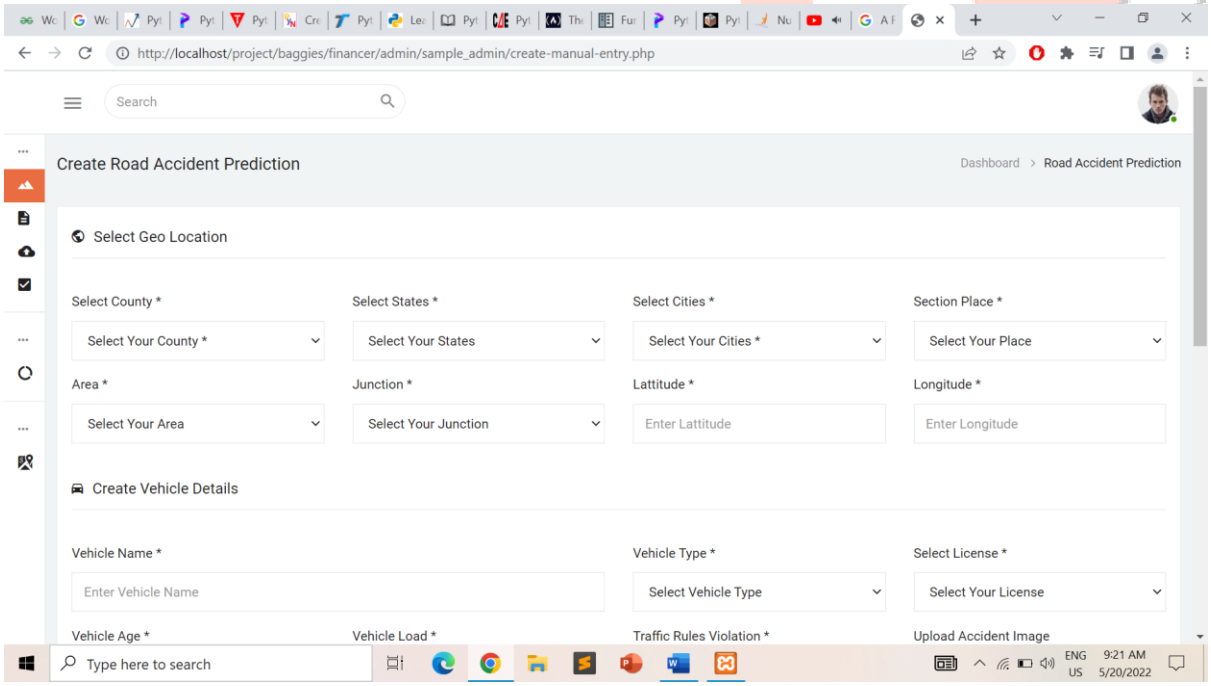

Fig. 4. Selecting the type of accident prediction

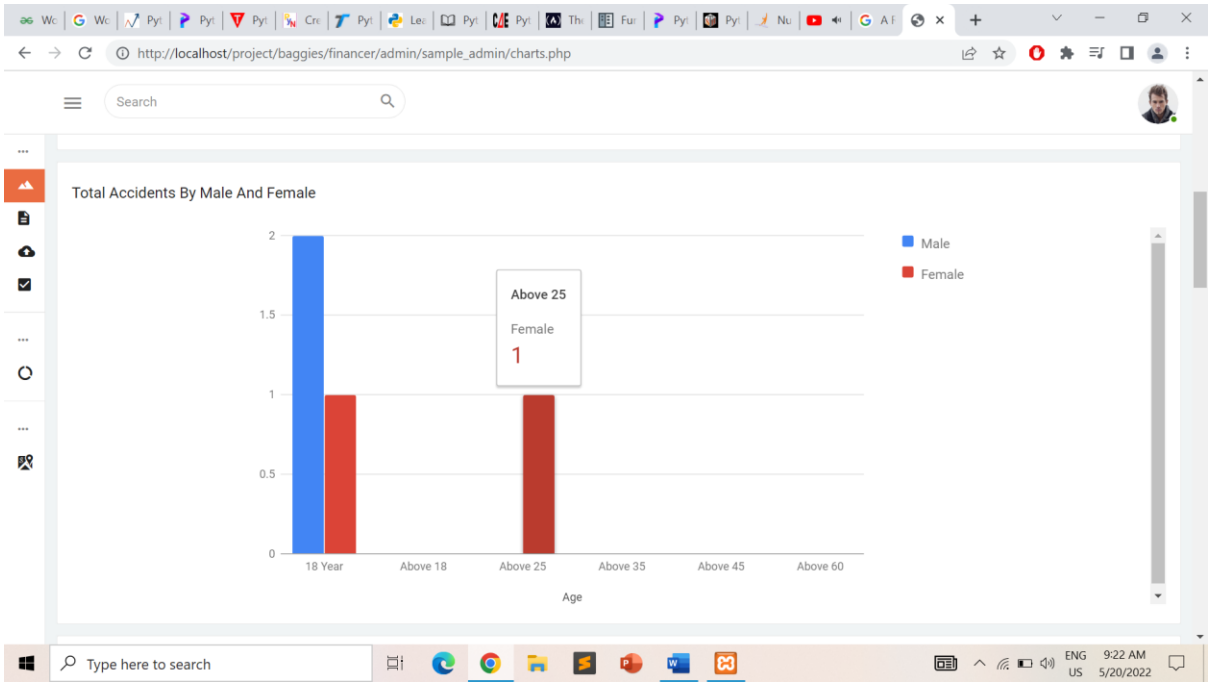

Fig. 5. Accidents done by Sex (M/F)

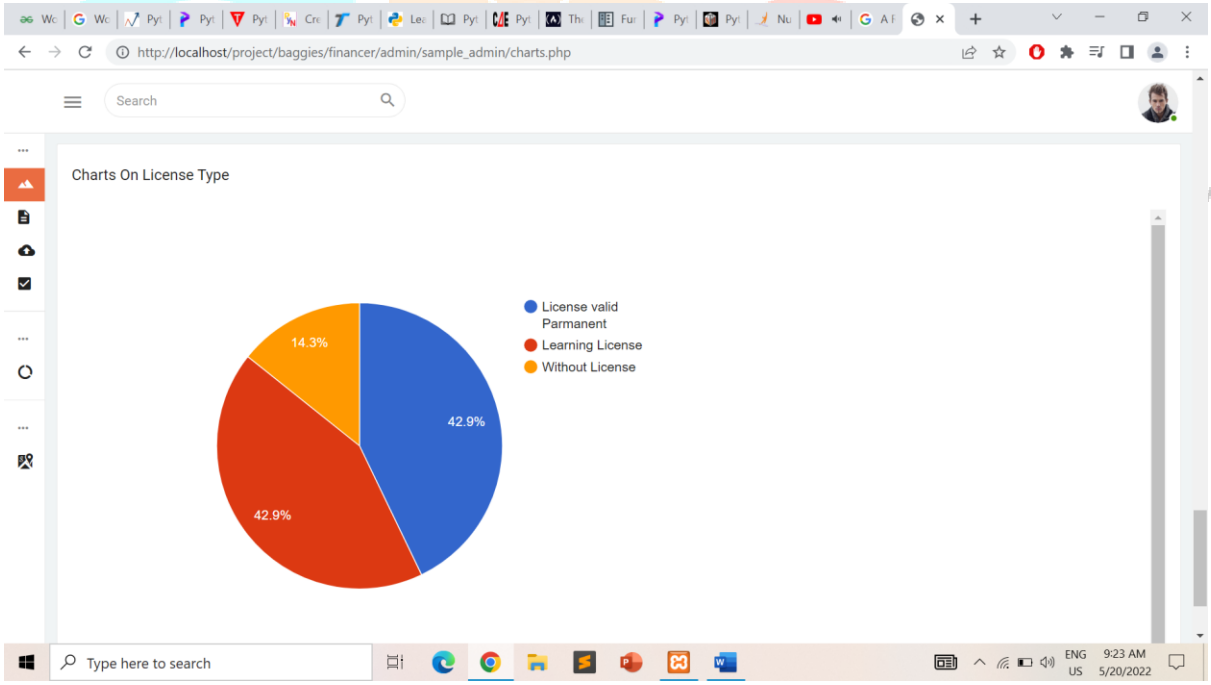

Fig. 6. People with proper license

Conclusion:

In the study, a hybrid- based approach developed to predict the severity of the TRA dataset. From the statistical results, it can be seen that the rural mortality rate is higher, while the city is lower. Statistical analysis also includes other limiting factors such as the age of the vehicle, the type of vehicle, the age group of the person, and the category of road users. The predicted data results are displayed in a graphical representation. Graphical representations help the public understand accident metrics that help reduce mortality. Furthermore, varying contribution was described in the study. Overall, the paper attempted to show the impacts of merging Clustering

and Classification to increase accuracy of the model, as well as identifying important contributing factors category from data obtained for traffic accident datasets. To get a better outcome, we'll use the other dataset to smooth out our model.

Rerfences:

1. F.M.O.I. Forensic Medicine Organization of Iran; Statistical Data, Accidents, online avail-able on: http://www.lmo.ir/?siteid=1&pageid=1347

2. A.T. Kashani et al., "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", PROMETTraffic& Transportation, 23(1), pp. 11-17, 2011.

3. L.Y. Chang, H.W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", Accident Analysis and Prevention, 38(5), pp. 1019-1027, 2006.

4. S. Yau-Ren et al. "The Application of Data Mining Technology to Build a Forecasting Model for Classification of Road Traffic Accidents", Mathematical Problems in Engineering, Volume 2015 (2015), pp. 1-8., 2015. F. Babi and K. Zuskáová • Descriptive and Predictive Mining on Road Accidents Data– 92

5. R. Nayak et al., "Road Crash Proneness Prediction using Data Mining". Ailamaki, Anastasia & Amer-Yahia , Sihem (Eds.) Proceedings of the 14th International Conference on Extending Database Technology, Association for Computing Machinery (ACM), Upp-sala, Sweden, pp. 521-526, 2011.

6. V. Shankar, J. Milton, F. Mannering, "Modeling accident frequencies as zero-altered probability processes: An empirical inquiry", Accident Analysis & Prevention, 29(6), pp. 829-837, 1997.

7. A Araar et al., "Mining Road traffic accident data to improve safety in Dubai", Journal of Theoretical and Applied Information Technology, 47(3), pp. 911-927, 2013.

8. S. Vigneswaran et al., "Efficient Analysis of Traffic Accident Using Mining Techniques", International Journal of Software and Hardware Research in Engineering, Vol. 2, No. 3, 2014, pp. 110- 118, 2014.

9. L. Martin et al. "Using data mining techniques to road safety improvement in Spanish roads", XI Congreso de Ingeniería del Transporte (CIT 2014), Procedia - Social and Behavioral Sciences 160 (2014), pp. 607–614, 2014.

10. P. Flach et al., "On the road to knowledge: Mining 21 years of UK traffic accident reports", Data Mining and Decision Support: Aspects of Integration and Collaboration, Springer, pp. 143-155, 2003.

11. H. Zhang et al., "In-Memory Big Data Management and Processing: A Survey", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 7, pp. 1920–1948, 2015.

12. J. Hipp, U. Güntzer, G. Nakhaeizadeh, "Algorithms for Association Rule Mining &Mdash; a General Survey and Comparison", SIGKDD Explor Newsl 2, pp. 58–64, 2000.

13. A.T. Kashani et al., "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", PROMETTraffic& Transportation, 23(1), pp. 11-17, 2011.

14. P.J. Ossenbruggen, J. Pendharkar et al., "Roadway safety in rural and small urbanized areas", Accidents Analysis & Prevention, 33(4), pp. 485-498, 2001.

15. R. Agrawal, T. Imieliski, A. Swami, "Mining Association Rules Between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, pp. 207–216, 1993.

16. R. Agrawal, R. Srikant, “Fast Algorithms for Mining Association Rules in Large Data-bases”, Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 487-499, 1994.

17. L. Breiman, “Random Forests”, Machine Learning, Vol. 45, pp. 5- 32, 2001.