TRAFFIC ACCIDENT SEVERITY PREDICTION BASED ON DECISION LEVEL FUSION OF MACHINE AND DEEP LEARNING

Motupally Yashaswi[1], Pandavula Thanuja[2], Renukuntla Sandhya[3]
Dr.G.Thirupathi Associate Professor[4]
SVS GROUP OF INSTITUTIONS, BHEEMARAM(V), Hanamkonda T.S. India -506015

ABSTRACT

Due to the exponentially increasing number of vehicles on the road, the number of accidents occurring on a daily basis is also increasing at an alarming rate. With the high number of traffic incidents and deaths these days, the ability to forecast the number of traffic accidents over a given time is important for the transportation department to make scientific decisions. In this scenario, it will be good to analyze the occurrence of accidents so that this can be further used to help us in coming up with techniques to reduce them. Even though uncertainty is a characteristic trait of majority of the accidents, over a period of time, there is a level of regularity that is perceived on observing the accidents occurring in a particular area. This regularity can be made use of in making well informed predictions on accident occurrences in an area and developing accident prediction models. In this paper, we have studied the inter relationships between road accidents, condition of a road and the role of environmental factors in the occurrence of an accident. We have made use of data mining techniques in developing an accident prediction model using Apriori algorithm and Support Vector Machines. Bangalore road accident datasets for the years 2014 to 2017 available in the internet have been made use for this study. The results from this study can be advantageously used by several stakeholders including and not limited to the government public work departments, contractors and other automobile industries in better designing roads and vehicles based on the estimates obtained.

Keywords : traffic, survey, accident, prediction, apriori algorithm

1 INTRODUCTION

The alarming rate of increase of accidents in India is now a cause for serious concern. According to some recent statistics, India accounts for roughly six percent of global road accidents while owning only one percent of the global vehicle population. There are a lot of accident cases reported due to the negligence of twowheelers, whereas over-speeding is also another contributing factor. Accidents caused while under the influence of alcohol or during general traffic violations are also common. In spite of having set regulations and the highway codes, the negligence of people towards the speed of the vehicle, the vehicle condition and their own negligence of not wearing helmets has caused a lot of accidents. While the major cause of road accidents is attributed to the increasing number of vehicles, the role played by the condition of the roads and other environmental factors cannot be overlooked.

The number of deaths due to road accidents in India is indeed a cause for worry. The scenario is very dismal with more than 137,000 people succumbing to injuries from road accidents. This figure is more than four times the annual death toll from terrorism. Accidents involving heavy goods vehicles like trucks and even those involving commercial vehicles used for public transportation like buses are some of the most fatal kind of accidents that occur, claiming the lives of innocent people. Weather conditions like rain, fog, etc., also play a role in catalysing the risk of accidents. Thus, having a proper estimation of accidents and knowledge of accident hotspots and ca sing factors will help in taking steps to reduce them. This requires a keen study on accidents and development of accident prediction models.

To implement a well designed road framework management system for looking into road security aspects, it is often desired to have an optimized accident prediction model which can analyze potential issues arising due to infrastructure fallbacks and to estimate the effect of existing models in reducing the occurrence of accidents. The main challenge involved in the creation of such a model include the evaluation of the weight that can be attributed to the impact of each variable in contributing to the accident and assessing how the model can be best designed to incorporate the effect of all such variables.

Data mining techniques and models have in the past been found useful for the purpose of data interpretation in a variety of domains including but not limited to credit risk management, fraud detection, healthcare informatics, recommendation systems and so on. Approaches involving artificial intelligence and machine learning have further helped to augment these studies. For this paper, we have investigated the inter-relationship between the occurrences of ro d accidents and the roles played by the underlying road conditions and environmental factors in contributing to the same. Since such a study requires us to cover several aspects affecting accidents, we can make use of data mining techniques to analyze this data to extract relevant details from them, as these huge volumes of data would 1618otherwise be meaningless without the right interpretation applied to them

In this paper, we are discussing the effects of such an accident prediction model in identifying the risks involved in road accident scenarios. The next section discusses the prior works done with respect to analyzing the different accidents that have taken place over the years. This is followed by a summarized description of the methodology used in this work. Further, the different components of implementation including the system architecture, software and languages used, simulation, user interface and screenshots of the developed application are discussed. Finally, the discussion and conclusions derived from the present study and the future scopes are outlined in the last two sections. The results from this study have been used to propose a model that can be used as a tool to estimate the possibility of road accidents in a particular area chosen by the user

Due to more vehicles on the road, accident rates are steadily increasing. So being a citizen it's our responsibility to avoid accident rate .now a days as machine learning techniques became more popular therefore we can use that technique to overcome from this problem. As we have already accidental data by using this we can build a new application to overcome accidents rate . As we have noticed that accidental data has been used in existing system to avoid accident's as per our research previous data provide basic idea that where the accident going to be happen .in early days where d roads were not properly built but nowadays roads are built in proper manner supposed if we are using earlier data to predict the place where accident going to happen it does make sense.so there is huge need to take different scenario to build an application.

The data set that we need to consider are wind weather, road conditions , population, traffic, time , speed limit by knowing pattern of all this data set we can draw the algorithm so that it will help us to predict the area where accident will happen. As it is difficulty to enter our data all the time it's better to have mapping system to predict the area where accidents will happen.

2. LITERATURE SURVEY

Vision-based traffic accident detection using sparse spatio-temporal features and weighted extreme learning machine

Author: Yuan long Yu1, Miaoxing Xu1,Jason Gu

Publication: The Institution of Engineering and Technology 2019 (www.ietdl.org) **Abstract:** Vision-based traffic accident detection is one of the challenging tasks in intelligent transportation systems due to the multi-modalities of traffic accidents. The first challenging issue is about how to learn robust and discriminative spatio- temporal feature representations. Since few training samples of traffic accidents can be collected, sparse coding techniques can be used for small data case. However, most sparse coding algorithms which use norm regularisation may not achieve enough sparsity. The second challenging issue is about the sample imbalance between traffic accidents and normal traffic such that detector would like to favour normal traffic. This study proposes a traffic accident detection method, including a self-tuning iterative hard thresholding (ST-IHT) algorithm for learning sparse spatio-temporal features and a weighted extreme learning machine (W-ELM) for detection. The ST- IHT algorithm can improve the sparsity of encoded features by solving an norm regularisation. The W-ELM can put more focus on traffic accidentsamples. Meanwhile, a two- point search strategy is proposed to adaptively find a candidate value of Lipschitz coefficients to improve the tuning precision. Experimental results in our collected dataset have shown that this proposed traffic accident detection algorithm outperforms other state-of-the-art methods in terms of the feature's sparsity and detection performance.

Design and implementation of an eye blinking detector system for automobile accident prevention

Author: Tariq Jamil ,Iftikhar Uddin Mohammed , Medhat H. Awadalla.

Publication: Southeast Con 2016

Abstract: Ever increasing number of fatal traffic accidents around the world can be significantly reduced if modern technology is incorporated within the automobile toassess the physical condition of the driver at regular intervals during the movement of the vehicle and preventive measures are automatically taken for the safety of all concerned entities, both within the vehicle and outside the vehicle. In this paper, design of an eye blinking detector system is presented which can monitor the physical state of the driver at regular intervals during his/her driving and, if needed, can raise an audible alarm within the vehicle to alert the.

Principal Component Analysis of Fatal Traffic Accidents Based on VehicleCondition Factors

Author: Tang Youming, Zhong Deliang , Zha Xinyu

Publication: 2018 11th International Conference on Intelligent Computation Technology Automation (ICICTA) and

Abstract: In order to find out several pivotal influencing factors of fatal traffic accidents, the number of fatal injuries recorded in the FARS database of the NationalHighway Traffic Safety Administration of the United States from 2010 to 2016 was calculated. The principal component analysis (PCA) method of multivariate statistical analysis is used to analyse the traffic conditions, and several pivotal influencing factors of fatal traffic accidents are obtained. The results show that tire wear, rim damage, exhaust system failure and coupling failure are the most important factors.

A Study on Road Accidents in Abu Dhabi Implementing a Vehicle Telematics System to Reduce Cost, Risk and Improve Safety

Author: Omar Kassem Khalil

Publication: 2017 10th International Conference on Devel- opments in e Systems Engineering (DeSE)

Abstract: Road accident study in the Emirate of Abu Dhabi is imperative as it will allow the government to improve its transport system and realize the vision of Abu Dhabi Department of Transport to contribute to quality of life, economic growth and environmental sustainability of the Emirate of Abu Dhabi. Despite sustainable and continuous reduction annually, road accidents still remain a serious phenomenon inthe Emirate of Abu Dhabi. Road accidents killed 5,564 people in the UAE over thepast 6 years, an average of more than two each day, with Abu Dhabi being the main victim. A further examination of the severity index, Police records have shown thatthe deaths during the years 2006-2011 were a result of nearly 56,700 accidents, which also injured 63,406 people. Research has shown that over 80% of road accidents are related to human factors. Bad driving habits are dangerous and can lead to loss of life. This study will propose possible method of adopting Telematics System for Abu Dhabi Department of Transport to enhance its road safety programs. Telematics System can help reduce fatalities and injuries, thus improving road safetyand efficiency of driving performance

The steady increase in the rate of accidents in India have prompted many researchers to look into the factors affecting road accidents and study about it. Since data mining techniques do not require certain assumptions between dependent and independent variables which are required in traditional statistical techniques, variouscategories of data mining techniques have been made use of in creating prediction models for road accidents in the past. Researchers have focussed on different sets of attributes in developing such models.

Srivastava et al. and Ghazizadeh et al. have mainly concentrated on studying the accidents occurring at intersection points. While the former has lookedinto categorizing the accidents based on their levels of seriousness using a Multi- layered perceptron (MLP) technique that was found to be more effective, the latter has made use of a feed forward MLP that utilizes back propagation learning to analyse the effect of several factors such as day or night, traffic conditions etc. on accidents.

Chen et al. in their studies have found highways to be the common area where a majority of accidents had been reported to occur. Williams et al. have found through their studies that the age and experience of a driver also play a major role inthe occurrence of accidents.

3. PROBLEM STATEMENT

Williams et al. have found through their studies that the age and experience of a driver also play a major role in the occurrence of accidents. Suganya, E. and S. Vijayarani in their paper have analysed the road accidents in India and compared the performance of different classification algorithms such as linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, Random Forest and gradient boosting algorithm using accuracy, error rate and execution time as a measure of performance. They have found the performance of KNN to be better thanthat of the others.

Sarkar et al. have done a comparative study on the type of roads that are prominent in accidents. While exploring the other components associated with accidents, they have found that the occurrence of accidents in highways is more common than in a normal road similar to. Stewart et al. have utilized original data in building a neural network model to predict accidents. They found that this modelwas able to give quicker results than those being used in the models built on Indianroads.

Zheng et al. have studied the range of injuries that come forth in a motor vehicleaccident and have also analysed the emotions of the drivers involved in the accidents that could have been a causal factor. Arun Prasath N and Muthusamy.

Punithavalli have conducted an extensive survey on the different techniques usedin road accident detection over the years, the approaches implemented in them and discusses their merits and de-merits.

George Yannis et al., in their paper, have discussed about the current practices used in the development of accident prediction models on an international level.Detailed information on various models have been collected with the help of questionnaires and they have made use of this data to identify which could be the most useful model that can be applied for accident prediction.

Anand, J. V has developed a method to determine the effect of different variablesin the detection and prediction of atmospheric deterioration all over the world. Fuzzy C means clustering, R-studio, and the ARIMA frame work have been made use of in creating this method.

A similar approach can also be tried in evaluating the impact of various factorson road accidents. Analysing the original cause of accidents is important because this will tell us the impact factor and contribution of each attribute towards road accidents. Tiwari et al. have made use of self-organizing maps, K-mode clustering techniques, Support Vector Machines, Naïve Bayes and Decision tree to classify thedata from road accidents based on the type of road users.

3.1 LIMITATION OF SYSTEM

The system doesn't have facility to train and test on large number of numbers. The system doesn't measure an accurate road accident due to poor classificationmodel.

4. PROPOSED SYSTEM

In the proposed system, the system has built an application that is capable ofpredicting the possibility of occurrence of accidents based on available road accidentdata. Data pre-processing is done on this road accident data to obtain a dataset. Thedata preprocessing step includes cleaning to remove the null and garbage values, and normalization of the data, followed by feature selection, where only relevant features from the original dataset are selected to be included in the final dataset. The dataset is then subjected to different data mining techniques. Clustering is performed on this dataset. The clusters are then subjected to other algorithms like Support Vector Machines (SVM) and Apriori. Since the data being used for the study has an unknown distribution and we need to sort out the frequent and infrequent items in the dataset, the former (SVM) is used to predict the probable risk of accidents while the latter (Apriori) is applied to perform rule mining, that is, to generate a frequent item set based on given support and confidence values.

Rules have been set considering different combinations of factors which have caused accidents of varying nature and severity in different road types and weather conditions. For the frequently occurring item sets, the chosen support and confidence values imply the higher probability of the particular combination of attributes in leading to an accident. For example, based on the rule mining done, the probable risk of an accident occurring even during fine weather in a junction on account of over-speeding is high and could prove to be fatal based on the training dataset.

SVM classification has been used to characterize each accident event into a high or a low risk category. Various data mining techniques and exploratory visualization techniques are applied on the accident dataset to get the interpreted results..

4.1 Advantages of system

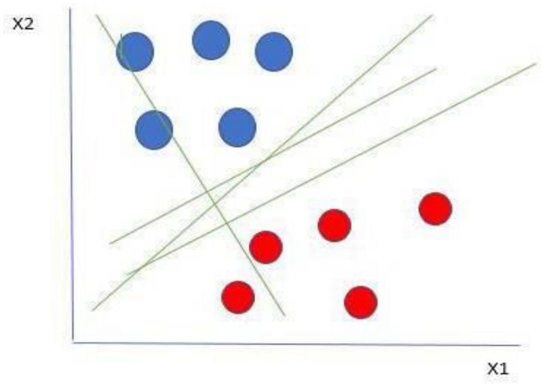
These optimized models can be efficiently utilized by the government to reduce road accidents and to implement policies for road safety. The overall model has helped to give an understanding of the combinations offactors that have proven fatal in accident scenarios.

5. ALGORITHMS USED

5.1 SVM

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

Let's consider two independent variables x1, x2 and one dependent variable which is either a blue circle or a red circle.



Linearly Separable Data points

rom the figure above its very clear that there are multiple lines (our hyperplane here is a line because we are considering only two input features x1, x2) that segregates our data points or does a classification between red and blue circles. So how do we choose the best line or in general the best hyperplane that segregates our data points.

5.2 K- Nearest Neighbours Classifier:

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K-NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostlyit is used for the Classification problems. K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data. It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

6. MODULE DESCRIPTION

6.1 Dataset Collection

Dataset is simply a collection of data pieces that can be analysed and predicted by a computer. Data collection should be uniform, and able to be understood by machines that don't see things the same way that humans do. Hence, it is crucial that after collecting data, it is cleaned, completed, and then annotated with meaningful tags readable by computers. The data collection module is responsible for gathering relevant data related to traffic accidents. This can include various types of data such as:

- **Accident data**: Information about the accidents themselves, including location, date and time, weather conditions, road conditions, and severity ofthe accidents (e.g., minor, serious, fatal).
- **Vehicle data**: Details about the vehicles involved in the accidents, such as type of vehicle, make and model, speed, and any mechanical issues.
- **Driver data**: Information about the drivers involved, including age, gender, driving experience, and any prior traffic violations.
- **Environmental data**: Data related to the environment surrounding the accident, such as nearby landmarks, traffic signals, signage, and visibility conditions.
- **Geographic data**: Geographic information system (GIS) data that provides spatial context to the accidents, including road networks, intersections, and traffic flow patterns.

The data collection module may utilize various sources to gather this data, such as police reports, insurance claims, traffic cameras, sensors installed on roads, and publicly available datasets. Additionally, data cleaning and preprocessing techniques are often applied to ensure the data is accurate, consistent, and ready foranalysis and modeling.

6.2 Data Pre Processing

Data pre processing is an information mining method that is used to convert the raw facts into a useful and efficient format. It is the essential step while developing a system learning model. While developing a system mastering mission,It isn't always constantly a case that we come across easy and formatted facts. And at the same time as doing any operation with statistics, it's far obligatory to ease it and put it in a formatted way. So for this, we use statistics pre processing task. The cause is that after the dataset is break up into train and take a look at sets, there will not be sufficient statistics inside the schooling dataset for the model to analyze an powerful mapping of inputs to outputs. There will also not be enough information inside the check set to correctly compare the version overall performance. The splitisfinished via 80% as education set and 20% as test set.

6.3 Train set

The educate set would incorporate the facts in an effort to be fed into the version. In easy terms, our version would analyze from these facts. As an instance, a regression model would use the examples in this record to discover gradients to be able to reduce the value characteristic. Then these gradients will be used to lessen the value and are expecting information efficaciously.

6.4 Test set

The test set contains the facts on which we check the educated and verified model. It tells us how green our ordinary version is and the way likely is it going to predict something which does now not make us feel. There are a plethora of assessment metrics (like precision, do not forget, accuracy, and so forth.) that canbe used to measure the performance of our model. The data preprocessing module is responsible for preparing the raw data collected from various sources for analysis and model training. Here's a breakdownof what it typically involves:

Data Cleaning: This step involves handling missing values, outliers, and inconsistencies in the data. Missing values may be imputed using techniques like mean, median, or mode imputation. Outliers may be treated or removedbased on domain knowledge.

Feature Selection/Extraction: Identifying the most relevant features (variables) for predicting accident severity is crucial. Feature selection techniques such as correlation analysis, feature importance ranking, or domain knowledge may be used. Additionally, new features may be derived or extracted from existing ones to enhance predictive power.

Feature Scaling/Normalization: Features may need to be scaled or normalized to ensure that they are on a similar scale. This prevents certain features from dominating others during model training. Common techniquesinclude Min-Max scaling or standardization.

Handling Categorical Variables: Categorical variables, such as weather conditions or road type, need to be encoded into numerical format for the model to interpret. Techniques like one-hot encoding or label encoding maybe used for this purpose.

Data Splitting: The dataset is typically split into training, validation, and testsets. The training set is used to train the model, the validation set is used to tune hyperparameters and evaluate model performance during training, and the test set is used to assess the final model's performance.

Data Balancing (if needed): Imbalanced datasets, where one class (e.g., severity level) is significantly more prevalent than others, may lead to biased models. Techniques like oversampling, under sampling, or generating synthetic samples can be used to address this imbalance.

Data Augmentation (if applicable): In some cases, especially when dealingwith image data or time series data, data augmentation techniques may be applied to increase the diversity of the training dataset and improve modelgeneralization.

6.5 Data Transformation (if needed): Depending on the nature of the data andthe chosen machine learning algorithms, additional transformations such as log transformation, polynomial transformation, or Box-Cox transformation may be applied to meet the assumptions of the models.

Overall, the data preprocessing module plays a critical role in ensuring that the input data is clean, relevant, and properly formatted for the machine learning models to effectively learn from it and make accurate predictions of traffic accident severity.

6.6 Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Datasetsand Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Road Accident Prediction, Find Road Accident Prediction Type Ratio, View Road Accident Ratio Results, Download Predicted Data Sets, View All Remote Users.

A service provider module typically refers to a component responsible for managing and delivering services related to the prediction model. This module mightinclude functionalities such as:

Model Deployment: Deploying the trained machine learning model to a server or cloud platform to make predictions accessible via an API.

API Development: Creating an interface that allows users or other systems to interact with the prediction model, usually through RESTful APIs.

Data Preprocessing: Handling incoming data requests by preprocessing thedata to meet the model's input requirements, such as encoding categorical variables or scaling numerical features.

Prediction Handling: Utilizing the deployed model to make predictions on incoming data and returning the predicted severity of the traffic accident.

Error Handling: Managing errors that may occur during the prediction process and providing appropriate responses or logging for debugging purposes.

Scalability: Ensuring that the service can handle a high volume of requests efficiently, either through horizontal scaling (adding more server instances) or vertical scaling (upgrading server resources).

Security: Implementing security measures to protect the API from potential attacks or unauthorized access, such as authentication and authorization mechanisms.

Overall, the service provider module acts as a bridge between the trained predictionmodel and the end-users or client applications, facilitating the seamless integration and utilization of the predictive capabilities within the project's ecosystem.

6.7 View and Authorized Users

In this module, the admin can view the list of users who all registered. In this,the admin can view the user’s details such as, user name, email, address and adminauthorizes the users.

The "View Users" module typically allows authorized users to access information about registered users within the system. This module might include functionalities such as searching for users, viewing their profiles, and accessing relevant data associated with them, such as accident history or personal information.

On the other hand, the "Authorize Users" module is responsible for managing user permissions and access rights within the system. This module enables administrators or authorized personnel to grant or revoke access privileges for different users based on their roles or responsibilities. It involves functionalities like creating user accounts, assigning roles, setting permissions, and managing authentication mechanisms to ensure secure access to the system.

6.8 Remote Users

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user nameand password. Once Login is successful user will do some operations like PREDICTROAD ACCIDENT STATUS,VIEW YOUR PROFILE.

The "Remote Users" likely refers to the functionality that allows users to accessthe system from remote locations or over a network. This module would enable usersto interact with the application from different geographical locations, possibly usingvarious devices such as computers, tablets, or smartphones.

7. METHODOLOGY

7.1 Artificial neural network

An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain. Here, each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another. Artificial neural networks (ANNs), or connectionist systems, are computing systems vaguely inspired by the biological neural networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task specific rules. An ANN is a model based on a collection of connected units or nodes called "artificial neurons", which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit information, a "signal", from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. In common ANN implementations, the signal at a connection between artificial neurons is a real number, and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs. The connections between artificial neurons are called "edges". Artificial neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection.

7.2 Decision trees

Decision tree learning uses a decision tree as a predictive model to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining, and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data, but the resulting classification tree can be an input for decision making.

7.3 Support vector machines

Support vector machines (SVMs), also known as support vector networks, are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. An SVM training algorithm is a non-probabilistic, binary, linear classifier, although methods such as Platt scaling exist to use SVM in a probabilistic classification setting. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

7.4 Regression analysis

Regression analysis encompasses a large variety of statistical methods to estimate the relationship between input variables and their associated features. Its most common form is linear regression, where a single line is drawn to best fit the given data according to a mathematical criterion such as ordinary least squares. The latter is often extended by regularization (mathematics) methods to mitigate overfitting and bias, as in ridge regression. When dealing with non-linear problems, go-to models include polynomial regression (for example, used for trendline fitting in Microsoft Excel, logistic regression (often used in statistical classification) or even kernel regression, which introduces non-linearity by taking advantage of the kernel trick to implicitly map input variables to higher-dimensional space.

7.5 Bayesian networks

A simple Bayesian network. Rain influences whether the sprinkler is activated, and both rain and the sprinkler influence whether the grass is wet. A Bayesian network, belief network, or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independence with a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. Efficient algorithms exist that perform inference and learning. Bayesian networks that model sequences of variables, like speech signals or protein sequences, are called dynamic Bayesian networks. Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams.

7.6 Genetic algorithms

A genetic algorithm (GA) is a search algorithm and heuristic technique that mimics the process of natural selection, using methods such as mutation and crossover to generate new genotypes in the hope of finding good solutions to a given problem. In machine learning, genetic algorithms were used in the 1980s and 1990s. Conversely, machine learning techniques have been used to improve the performance of genetic and evolutionary algorithms.

7.7 Training models

Usually, machine learning models require a lot of data in order for them to perform well. Usually, when training a machine learning model, one needs to collect a large, representative sample of data from a training set. Data from the training set can be as varied as a corpus of text, a collection of images, and data collected from individual users of a service. Overfitting is something to watch out for when training a machine learning model. Trained models derived from biased data can result.

8. OUTPUT SCREENSHOTS

Figure:8.1 Sign_in screen

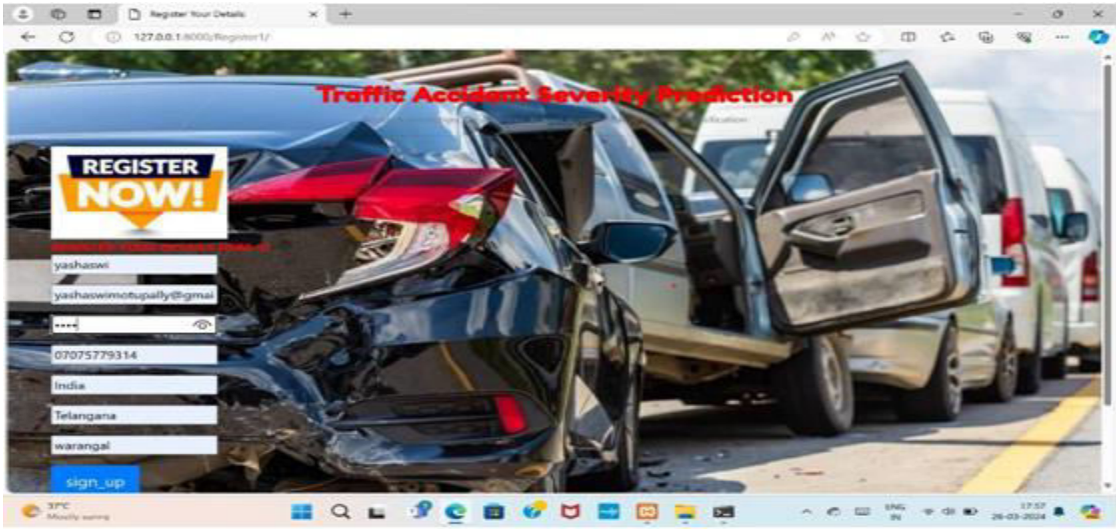


Figure:8.2 Register Your Details

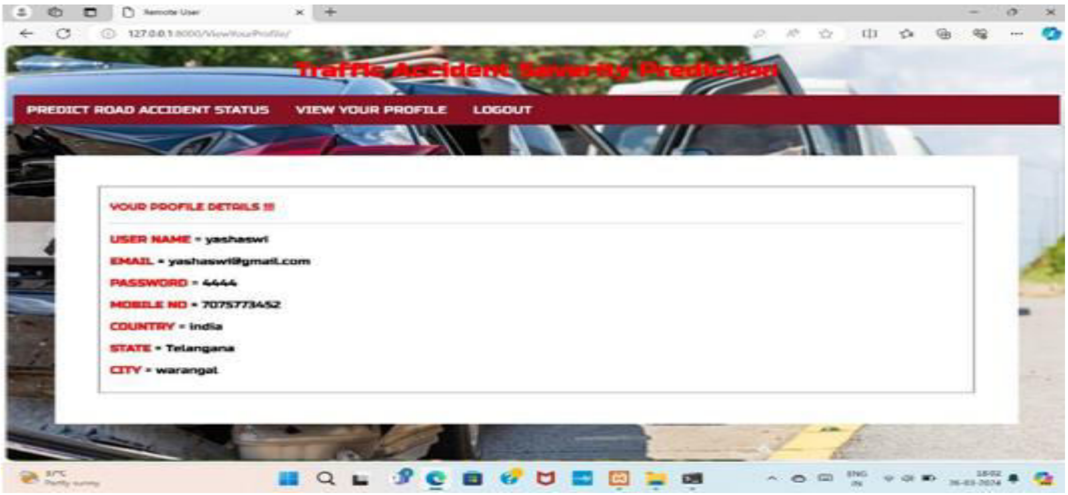


Figure:8.3 View Your Profile Details

Figure:8.4 Predict Road Accident Status with 'ACCIDENT'

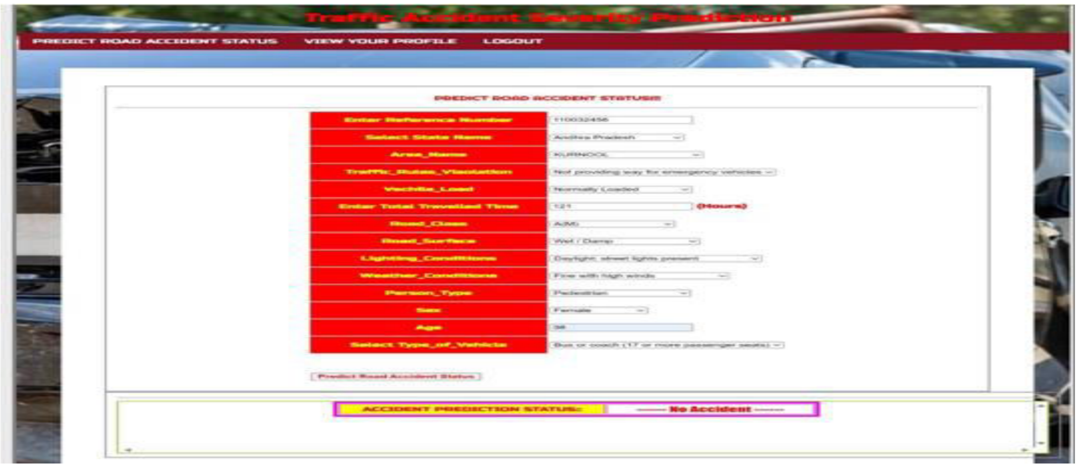


Figure:8.5 Predict Road Accident Status with ‘NO ACCIDENT’

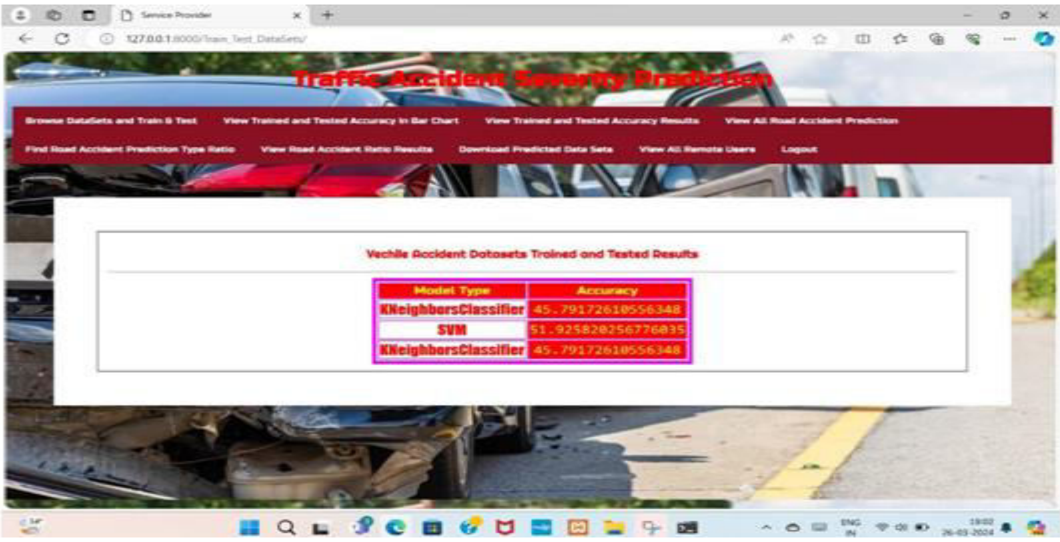


Figure:8.6 Vechile Accident Datasets Trained and Tested Results

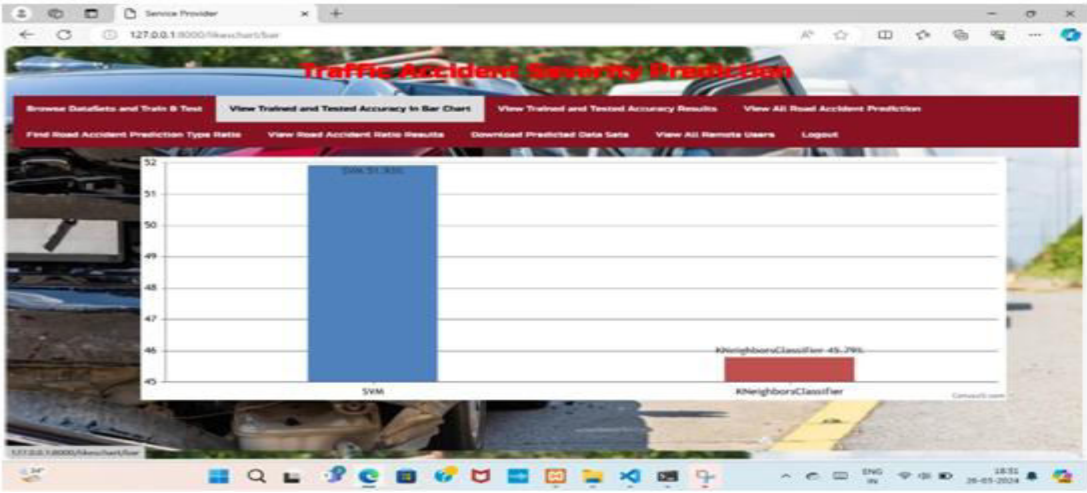


Figure:8.7 View Trained and Tested Accuracy In Bar Chart

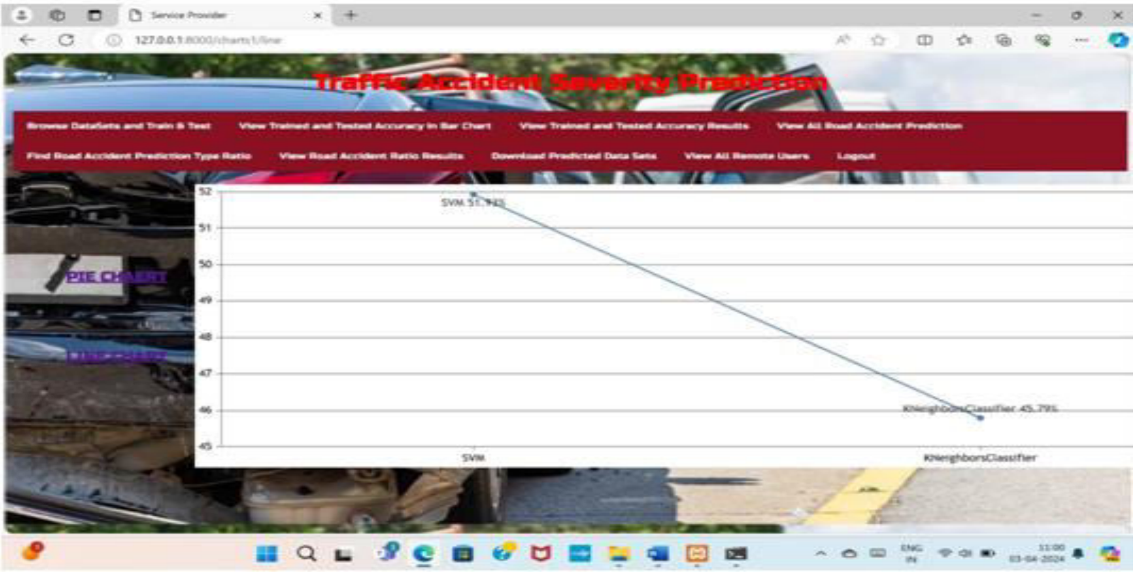


Figure:8.8 View Trained and Tested Accuracy Result in Line chart

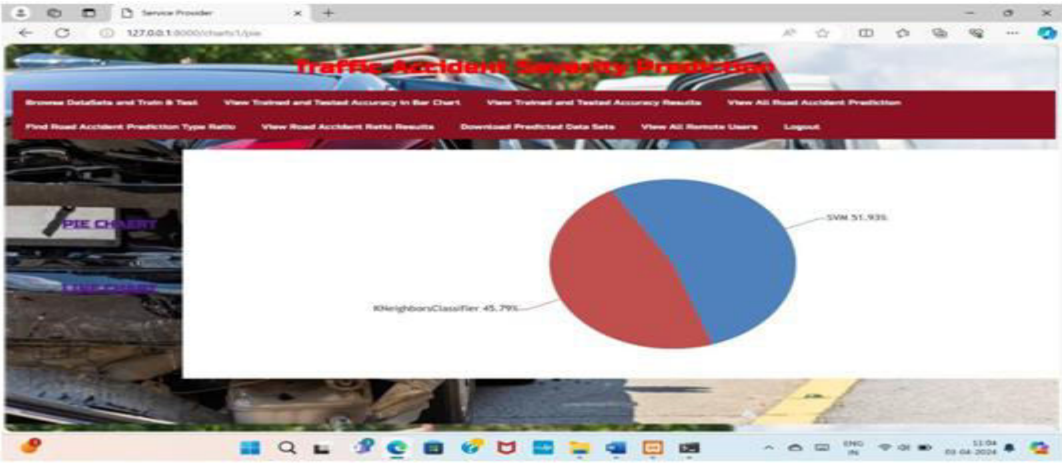


Figure :8.9 View Trained and Tested Accuracy Result in Pie chart

Figure :8.10 View All Road Accident Prediction Status

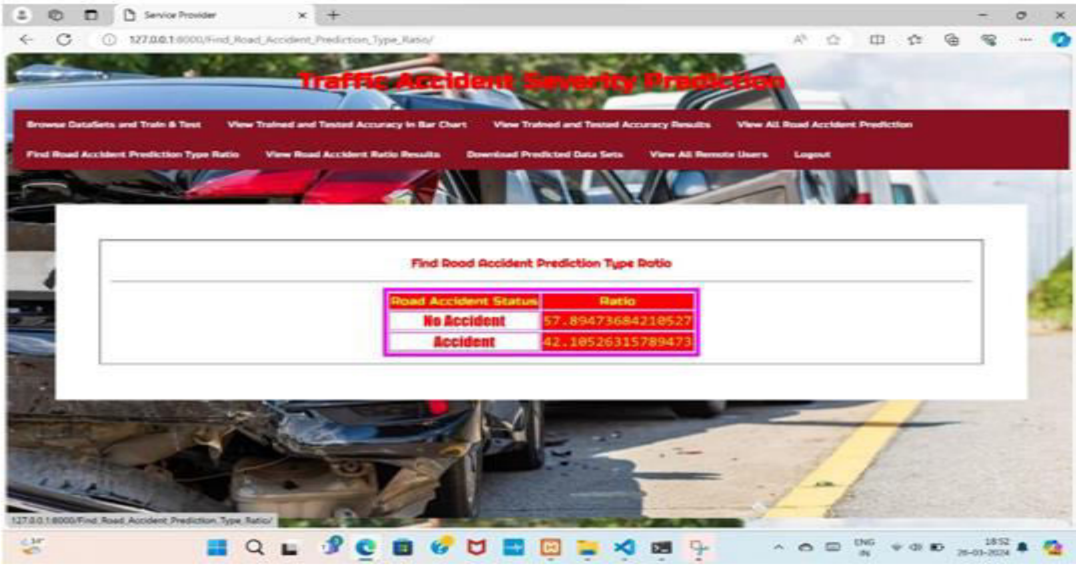


Figure :8.11 Find Road Accident Prediction Type Ratio

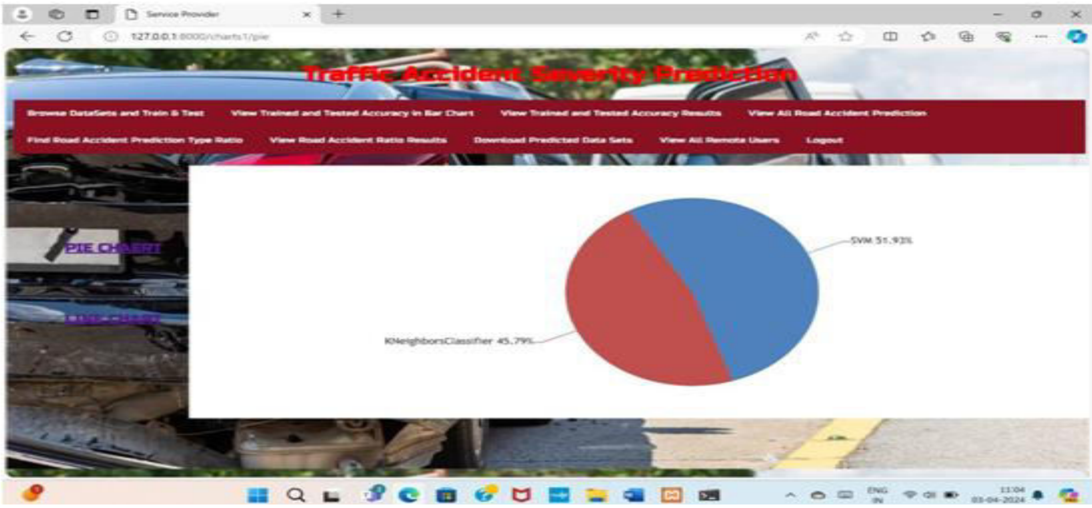


Figure :8.12 View Road Accident Ratio Results with Pie Chart

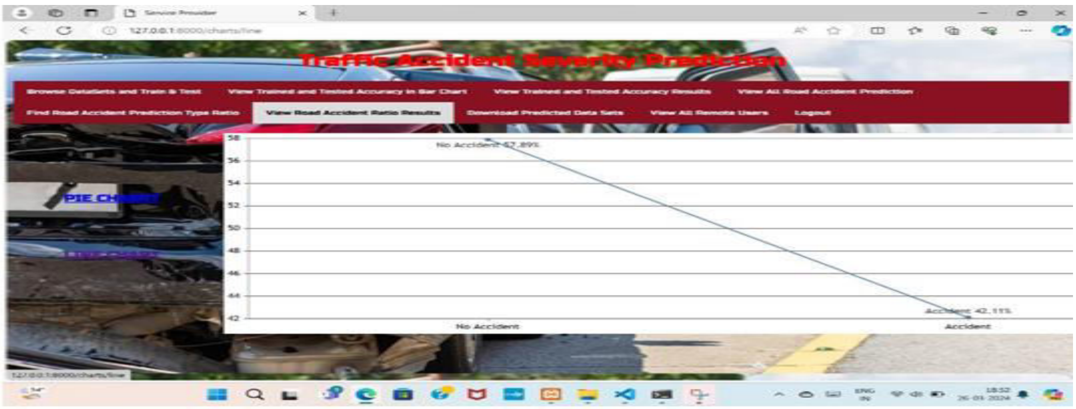


Figure : 8.13 View Road Accident Ratio Results with Line Chart

9. CONCLUSION

An accident can change the lives of many people. It is up to each of us to bring down this increasing number. This can be made possible by adopting safe driving measures to an extent. Since all instances of accidents cannot be attributed to the same cause, proper precautionary measures will also need to be exercised by the road development authorities in designing the structure of roads as well as by the automobile industries in creating better fatality reducing vehicle models. One thing within our capability is to predict the possibility of an accident based on previous data and observations that can aid such authorities and industries. This project was successful in creating such an application that can help in efficient prediction of road accidents based on factors such as types of vehicles, age of the driver, age of the vehicle, weather condition and road structure, This model was implemented by making use of several data mining and machine learning algorithms applied over a dataset for Bangalore and has been successfully used to predict the ris probability of accidents over different areas with high accuracy.

10. FUTURE ENHANCEMENT

The model can be further optimized in future to include several constraints that have been left out in the current study. These optimized models can be efficiently utilized by the government to reduce road accidents and to implementpolicies for road safety. Another scope of this work would be to develop a mobile app that will help the drivers in choosing a route for a ride. A call out to the driver through the maps service can also be implemented that would also announce the risk probability in a chosen route along with the directions. This can then be implemented by service provider companies such as Uber, Ola and so on in future. This will alsobe useful in having a better surveillance of accident prone areas and providing emergency services in the event of an accident. Better road safety instructions can also be installed along the highways taking into account the risks obtained from thismodel.

11. REFERENCES

[1] https://www.statista.com/topics/5982/road-accidents-in-india/

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

[3]Ghazizadeh M, McDonald AD, Lee JD. (2014). Text mining to decipher free-response consumer complaints: Insights from the nhtsa vehicle owner’s complaint database. Human Factors 56(6): 1189-

1203. http://dx.doi.org/10.1504/IJFCM.2017.089439.

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

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

[6]Suganya, E. and S. Vijayarani. "Analysis of road accidents in India using data mining classification algorithms." 2017 International Conference on Inventive Computing and Informatics (ICICI) (2017): 1122-1126.

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

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

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

[10]ArunPrasath, N and Muthusamy Punithavalli. "A review on road accident detection using data mining techniques." International Journal of Advanced Research in Computer Science 9 (2018): 881-885.

[11]George Yannis, Anastasios Dragomanovits, Alexandra Laiou, Thomas Richter, Stephan Ruhl, Francesca La Torre, Lorenzo Domenichini, Daniel Graham, Niovi Karathodorou, Haojie Li (2016). "Use of accident prediction models in road safety management – an international inquiry". Transportation Research Procedia 14, pp. 4257 – 4266.

[12]Anand, J. V. "A Methodology of Atmospheric Deterioration Forecasting and Evaluation through Data Mining and Business Intelligence." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 02 (2020): 79-87.

[13]Prayag Tiwari, Sachin Kumar, Denis Kalitin (2017). "Road-User Specific Analysis of Traffic Accident Using Data Mining Techniques". International Conference on Computational Intelligence, Communications, and Business Analytics. 10.1007/978-981-10-6430- 2_31.

[14]Kaur, G. and Er. Harpreet Kaur. "Prediction of the cause of accident and accident prone location on roads using data mining techniques." 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (2017): 1-7.

[15]Irina Makarova, Ksenia Shubenkova, Eduard Mukhametdinov, and Anton Pashkevich, “Modeling as a Method to Improve Road Safety During Mass Events”, Transportation Research Procedia 20 (2017)