Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

A Machine Learning Approach to Analyze and Predict the Severity of Road Accidents

1. KandasamySellamuthu
Assistant Professor (Sr.G), Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore.skandu23@gmail.com
2. Akshaya S
Student, Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore.akshayasandeep244@gmail.com
3. Anush J
Student, Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore.jpanush19@gmail.com
4. Arunkumar S
Student, Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore.arunkumarsubramaniam143@gmail.com

ABSTRACT:

Many developing countries are impacted by Globalization. India is one among the countries which benefited the foremost. The consumption level of the people is increased because of more economic activities. This made the people to travel more and to extend the use of vehicles. The increase in the count of vehicles is the reason for accidents. Two-wheeler accidents are the major reason for demise and affliction. It also ends up in the sudden demand for emergency medical care. Nearly above a million of people die every year due to traffic accidents. Also, 50 million are severely wounded or permanently impaired. A road accident prediction model would be helpful to take necessary precautions and save people from life threatening accidents.

INTRODUCTION:

The rapid urbanization caused a revolution within the increase of motorized vehicles. Considering the disease burden, the motor vehicles accidents are ranked ninth whereas at present due to massive increase of motor vehicles, the accidents rank third. A few researches have tried knowing the risk elements behind the road accidents in Indian cities. The accidents are acute in both highways and rural areas. In highway transportation, the major cause of accident is caused due to the specific way of traffic such as traffic along with people who walk along the road. In rural transportation, the accidents are caused because bad maintenance of roads, poor light quality, and lack of knowledge regarding road traffic and rules. We all know that road accidents are totally unpredictable and unavoidable, however, they can be reduced by using road accident prediction models that predict the important featured that cause accidents and a suitable traffic maintenance methodology. The analysis of accidents are made constantly at pre- eminent locations which paves way to conclude with effective decrease of accidents. The major purpose of the exploration are,

1. To study the reason behind road accidents in India.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

2.To plot and perform analysis of the accident-causing features and plot the clumsy areas which cause road accidents.

3.To know how the accidents occur with its type.

4. To know what kind of injuries happen when an accident occur.

RELATED WORKS:

Forecasting of Road Accident in Kerala: A Case Study

The important reason for mortality and injury is road accidents. Here, the prediction of road accidents was performed using the time-series analysis across all districts of Kerala. The data from January, 1999 to December, 2016 of Kerala was used in identifying patterns and developing a model wherein this idea is applied. The following are the models which we have used: “Holt-Winters (HW) exponential smoothing” and “Seasonal ARIMA (SARIMA)”.

A Driver-Centric Carpooling: Optimal Route-Finding Model Using Heuristic - Objective Search

A system to provide optimal route efficiently to user is introduced by introducing carpooling system. In order todo as such, we used NSGA. Particularly the multiple near-optimal results were made possible by using CDSA.

Computer Vision Based Road Traffic Accident and Anomaly Detection in the Context of Bangladesh

In Bangladesh, one important reason for mortality is road accidents. Here, the traffic pattern analysis is made sing the road side video data along with the vision based techniques to identify the kinetic features of vehicles. The result is the accident and its possibility. The accuracy level of this system in discernment of some special situations remained 85% approximately.

COVID-19 Cases in Iraq; Forecasting Incidents Using Box - Jenkins ARIMA Model

The pandemic outbreak of COVID-19 created panic all over the world. The mathematical principle in developing forecasting modes that aim to predict the number of future infections is considered crucial at this stage. The present investigation aims to analyze the time series using the Box-Jenkins method (Diagnostic, The Estimate, and selection, Forecasting) to find the best ARIMA model (Autoregressive Integrated Moving Average) for predicting the numbers of people infected with Covid-19 disease in Iraq. The results showed that the appropriate forecasting model is ARIMA (2,1,5). Depending on this model, they predict the numbers of those infected with COVID-19 daily and for thirty days. Predictive values are consistent with original series values, indicating the efficiency of the model.

EXISTING SYSTEM:

With the help of the machine learning algorithms such as Decision Tree, K- Nearest Neighbour Algorithm, Naïve Bayes, AdaBoost the road accidents were analyzed in existing systems. The

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

drawbacks were that Dataset contained a smaller number of features with less prediction. In the existing system, the accuracy was about 60 percentage.

PROPOSED SYSTEM:

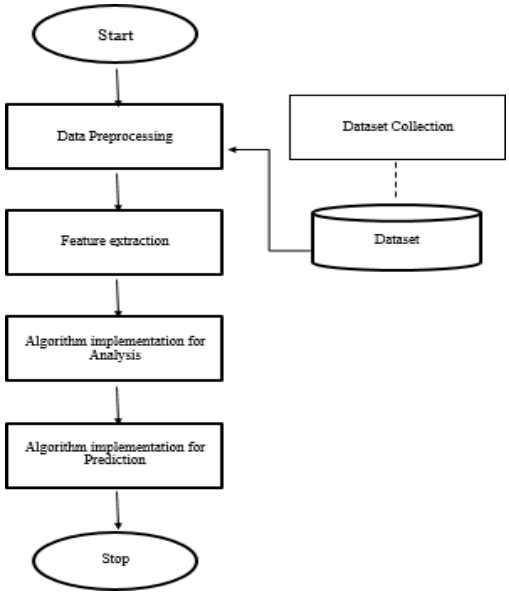
In this proposed approach, an effective machine learning model is built for analyzing road accidents along with the prediction of road accident severity with the use of Machine Learning algorithms. In already developed systems, there exists a prediction with 3 severity indicators with a related influences variable. The high affinity results of SVM model is greater than that of Random forest model in severity analysis. For the Government to take efficient measures in reducing theaccidents and improving the safety of people, the key influences of this model provides evocative results for through analysis of traffic data. The important dispute lies in the way of traffic in India which has a more congested situation where discernment of traffic and tracing becomes a strenuous task.

SYSTEM ARCHITECTURE:

The appropriate datasets for analyzing road accidents are collected from KAGGLE data source.The collected datasets are pre-processed using the algorithm like simple imputer and for feature extraction and filling the missing data. The dropna functions are used to drop the missing value row.

The pre-processed data is proceeded further for classification using classifier algorithms like SVM (Support Vector Machine Algorithm) and RFC (Random Forest Classifier Algorithm). The better result yielding algorithm is used for analyzing the severity based on different features in different categories. The analysis output is plotted as the bar chart for better understanding.

The final prediction is done using linear regression algorithm which produces good accuracy, and the output is displayed in the form of matplotlib graph.



Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

MODULES:

The following are the modules used to build our system:

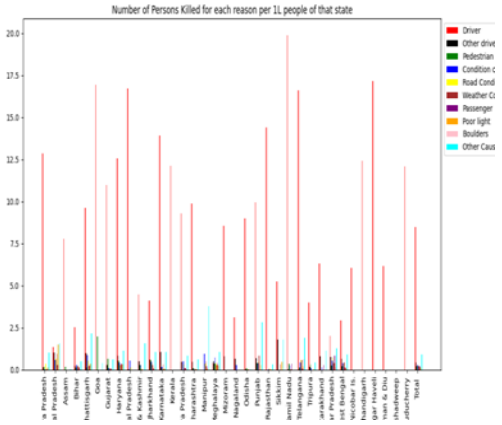
1. DATASET COLLECTION
2. PRE-PROCESSING
3. ALGORITHM IMPLEMENTATION/ ANALYSIS
4. PREDICTION

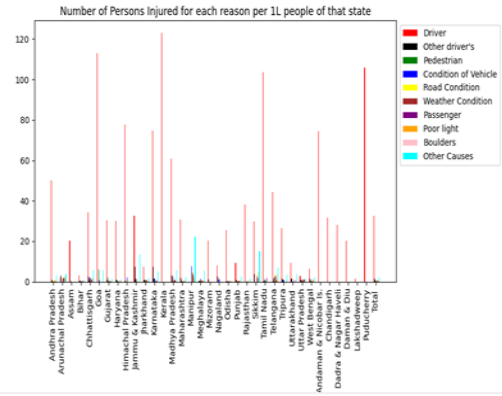
i. Model is built using the Python application. The dataset used in this study is the public dataset collected from the Kaggle website. We have collected three different Road datasets for the analysis and prediction.

ii. The data used in this study are categorical features, based on road accidents. We need to convert each text category to numbers for the machine to process those using mathematical equations. In this paper author apply a label encoder pre- processing technique to convert the categorical feature into numeric values. This proper dataset is split-up into training and testing. Training data used to train the machine. Training process contains both feature and label. The testing data should be used to predict the label class.

iii. In this study to analyze we have use supervised learning classification algorithm to compare the algorithm working process.

Reason for Accident:

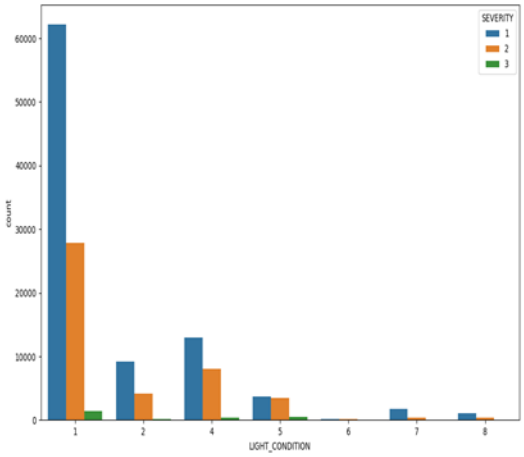




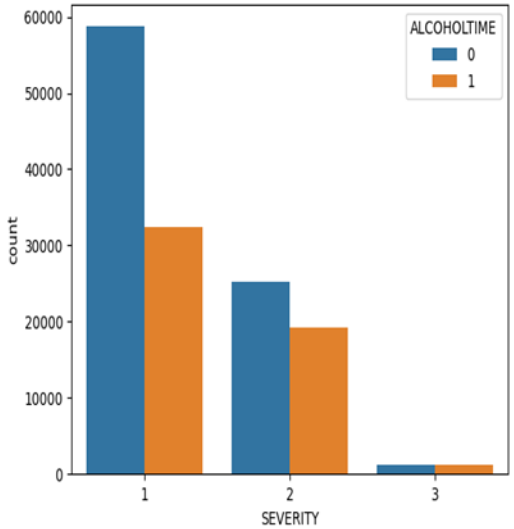
Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

Here, we have analyzed various features of various states of India. A few features that we have used for performing the analysis are specified here: driver, driver whether is a lady, Colliding area, Condition of road, Condition of vehicle, Condition of weather, Pedestrian, Passenger, Lack of Lighting. These features are denoted here using different colors. Here few of the features are shown for the rate of number of persons who get injured for different features per one lakh people in a particular state.

Severity Analysis:



Here, we have analyzed the severity of road accidents in three levels. The three levels are denoted using numeric values of one, two and three in different colors. Each feature is analyzed for the severity of accidents. The severity is analyzed based upon the count of accidents that occur with that particular feature. The severity analysis is plotted as a graph for better understanding and clear clarity on the occurrence of accidents in a particular situation owing to a particular feature.

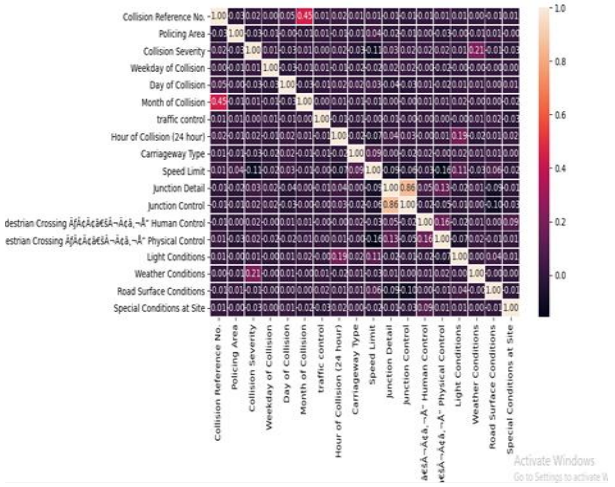


Few of the images of the severity analysis on road due to certain features are depicted here. The first image depicted here denotes the severity in terms of the feature of the Condition of Light in

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

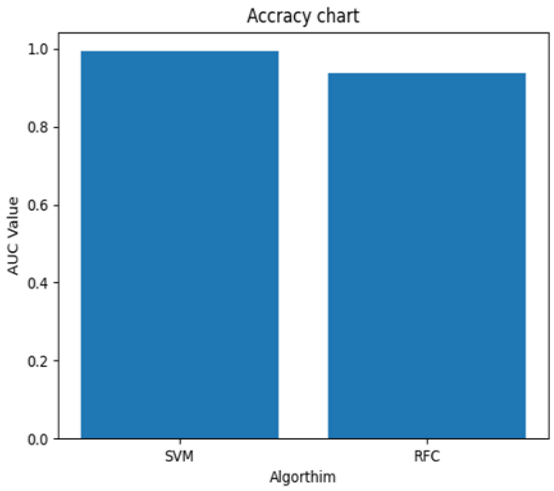
certain areas whereas the second image depicted here shows the severity of accident on road due to the driving of the vehicles by the drivers of various vehicles after the intake of alcohol.

iv. Models used for the prediction of Severity are SVM and RFC. Final output of the model is the prediction based on the analysis and severity.



The final prediction is made for providing the conclusion of severity of road accidents in a certain situation with a certain feature. The depiction shown above clearly states the features that cause high severity road accidents. Here the different variations of the color are plotted to differentiate the severity. The variations of colors are provided for the values ranging from 0.0 to 1.0. The low severity or no severity value always remains 0.0 whereas the moderate severity value remains 0.5 with the high severity value as 1.0. So, as per the prediction depiction shown the high severity features valued 1.0 should be treated with high importance. This can be used for taking any kind of preventive action to avoid or reduce the accidents that take place in that distinct scenario.

Accuracy Chart:

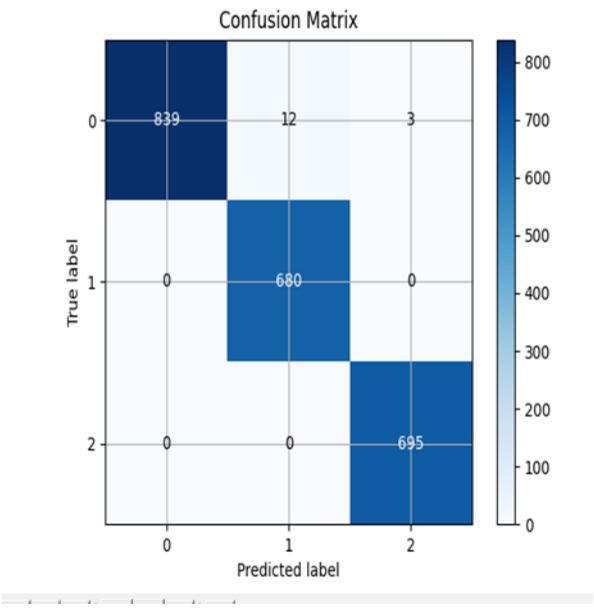


The comparison of the accuracy of the prediction/analysis of the road accident severity for the two algorithms namely Support Vector Machine (SVM) and the Random Forest Classifier (RFC) are depicted as a figure named Accracy chart is shown above. The SVM Algorithm obviously provides the more

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

accuracy in predicting/analyzing the severity of the road accidents. The SVM algorithm provides an accuracy of about 96% approximately in predicting the severity of the road accidents that occurred in different states of India based on the dataset that we have provided for analysis as well as the prediction. Final prediction of road accidents severity gives good accuracy.

Confusion Matrix:



Here, the confusion matrix is provided for further reference of the result. The dataset that we have used in this task of analysis and prediction of severity of accidents that take place on road are that of the data of accidents that have been occurred already in various states of India.

The performance of the system that we have built is clearly depicted here as an image that is the confusion matrix. Here, about 839 values matched the True label value 0 and about 680 values matched the True label value 1 and about 695 values matched the True label value 2. The performance of the model is at most good as most of the values of the true label that are available in the dataset that we have used for prediction matched exactly with the values of the predicted label.

CONCLUSION:

A system to predict the severity including the number of mortalities, number of bruises, and property damage along with the duration of accident was developed using an ML model. The comparison of SVM and RFC models enhances the accuracy of severity prediction estimation. Additionally, results of key influences identification of related elements on accident severity and duration helps the Government in taking steps to minimize accidental impacts and improvise safety of people. This serves as a technological way of alerting people about road accidents.

FUTURE ENHANCEMENT:

In the future, a recommender system developed as a mobile application with all the severity of accidents predicted accurately will be provided for the user to travel safely on roads which helps the user in a great way.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4241 - 4248
Received 05 March 2021; Accepted 01 April 2021.

REFERENCES:

1. T. Adhikary, A. K. Das, M. A. Razzaque, A. Almogren, M. Alrubaian, and M. M. Hassan. Quality of service aware reliable task scheduling in vehicular cloud computing. Mobile Networks and Applications, 21(3):482– 493, Jun 2016.
2. S. C. Huang, M. K. Jiau, and K. H. Chong. A heuristic multi-objective optimization for carpool services problem featuring high-occupancy vehicle itineraries. IEEE Transactions on Intelligent Transportation Systems, PP(99):1–12;, 2017.
3. I. H. Dridi, E. B. Alaia, and P. Borne. Heuristic approach for the optimization of delivery problem with time windows. In 2015 International Conference on Industrial Engineering and Systems Management (IESM), pages 733–739, Oct 2015.
4. C. Garrozi and A. F. R. Araujo. Multi- objective genetic algorithm for multicast routing. In 2006 IEEE International Conference on Evolutionary Computation, pages 2513–2520, 2006.
5. W. Herbawi and M. Weber. The ride matching problem with time windows in dynamic ridesharing: A model and a genetic algorithm. In 2012 IEEE Congress on Evolutionary Computation, pages 1–8, June 2012.
6. F. S. Hsieh. Carpooling based ontrajectories of drivers and requirements of passengers. In 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), March 2017.
7. M. K. Jiau, S. C. Huang, and C. H. Lin. Optimizing the carpool service problem with genetic algorithm in Service-based computing. In 2013 IEEE International Conference on Services Computing, pages 478–485, June 2013.