VEGUETA

Vol 26 Issue 01, April 2026

ISSN NO: 1133-598X

AI Based Road Safety Hazard Forecasting and Risk Estimation Model

Morla Venkata Sai Krishna[1], Madugu Shiva Kumar[2], Mora Upendhra Chary[3], P V Rama Gopala Rao[4]

Student, Department of CSE , Teegala Krishna Reddy Engineering College Hyderabad, India.

E-mail: saikrishnamorla144@gmail.com

Student, Department of CSE , Teegala Krishna Reddy Engineering College Hyderabad, India.

E-mail: madugushivakumar565@gmail.com

Student, Department of CSE , Teegala Krishna Reddy Engineering College Hyderabad, India.

E-mail: uppyuppy020@gmail.com

Assistant professor, Teegala Krishna Reddy Engineering College

rao.ramagopal1975@gmail.com

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 project, 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 crash prediction, ML models, Feature engineering, Class imbalance, Oversampling, Feature selection*

INTRODUCTION

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 causing 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 challenges 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 effects 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 project, we have investigated the inter-relationship between the occurrences of road 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 otherwise be meaningless without the right interpretation applied to them.

In this project, 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.

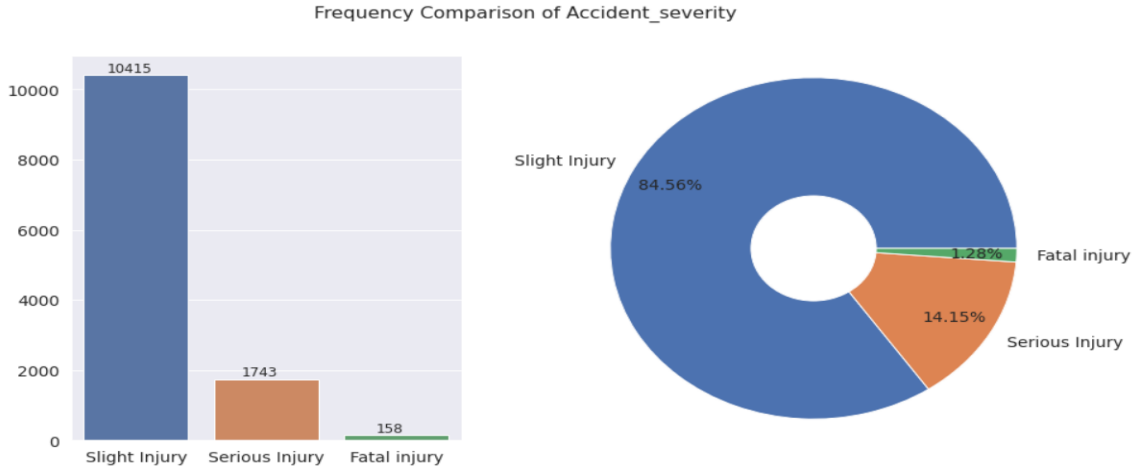
SYSTEM ANALYSIS

Dataset Details:

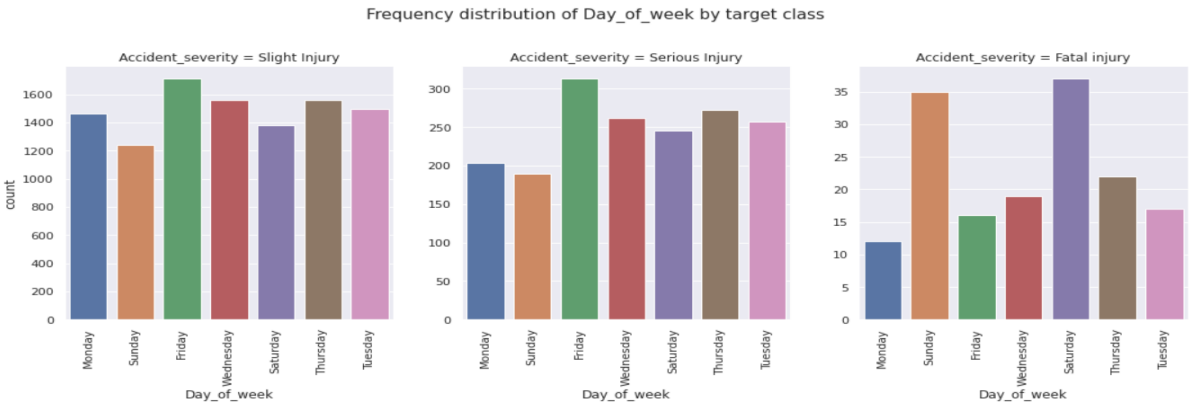
We have used the RTA dataset which is an open source dataset from Kaggle. The data set has been prepared from manual records of road traffic accident of the year 2017-20. All the sensitive information have been excluded during data encoding and finally it has 32 features and 12316 instances of the accident. Then it is preprocessed and for identification of major causes of the accident by analyzing it using different machine learning classification algorithms algorithms. RTA Dataset.csv is the dataset before preprocessing and cleaned.csv is the preprocessed dataset.

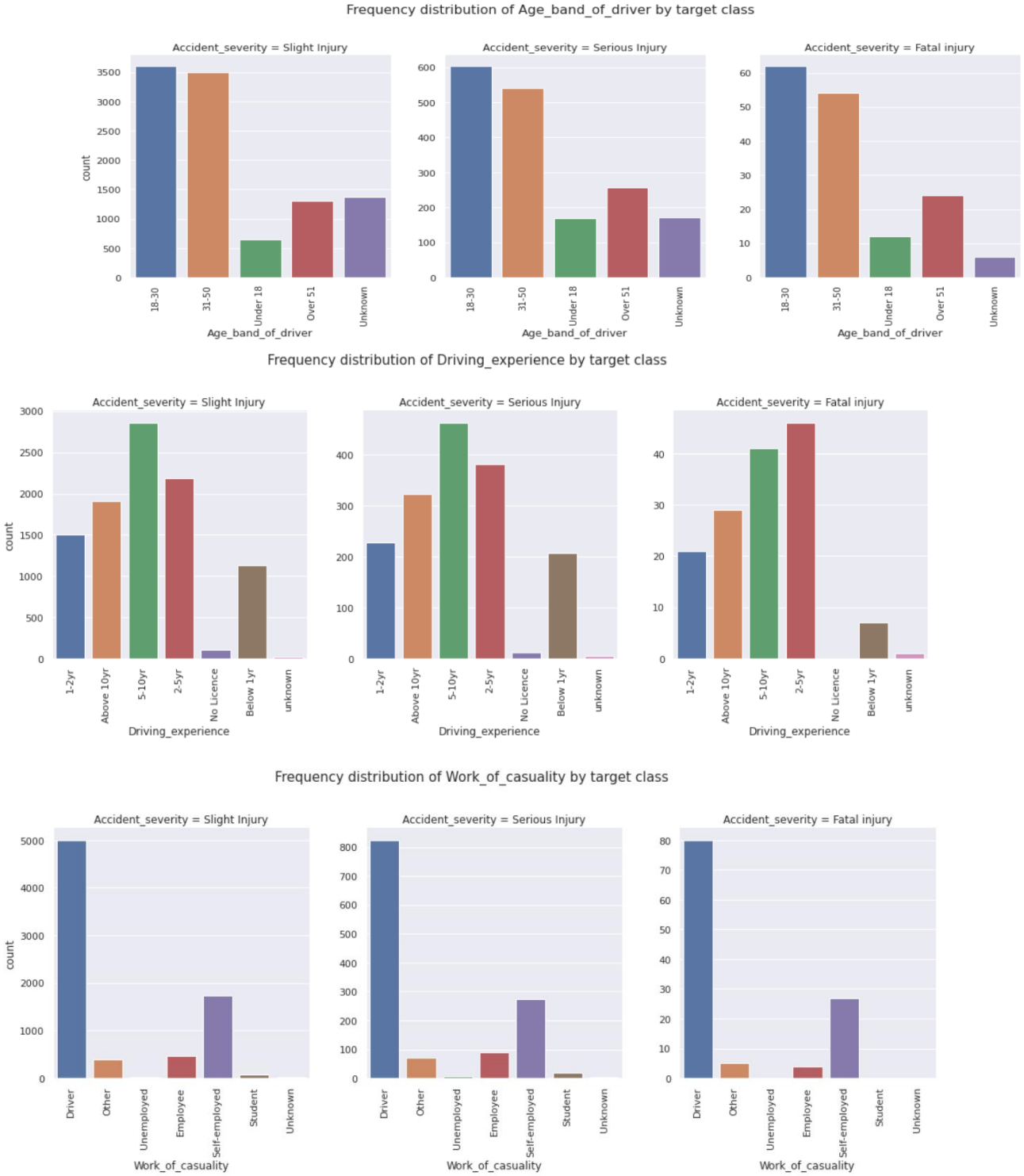
Data Analysis:

1. Read the dataset
2. Remove Duplicates
3. Fill missing values
4. Create a donut plot for frequency comparision of Accident severity



5. We observe that the dataset is imbalanced
6. Create a Catplot to compare frequency distributions of features (except 'Time') across target classes.





ALGORITHMS AND TECHNIQUES

Data Pre-processing:

1. Detect the outliers
2. Impute missing details using simple imputer
3. Encode categorical data
4. Split train data and test data
5. Perform feature selection using Chi-Squared method
6. Feed the dataset to multiple machine learning algorithms.
7. Below are the accuracy results:

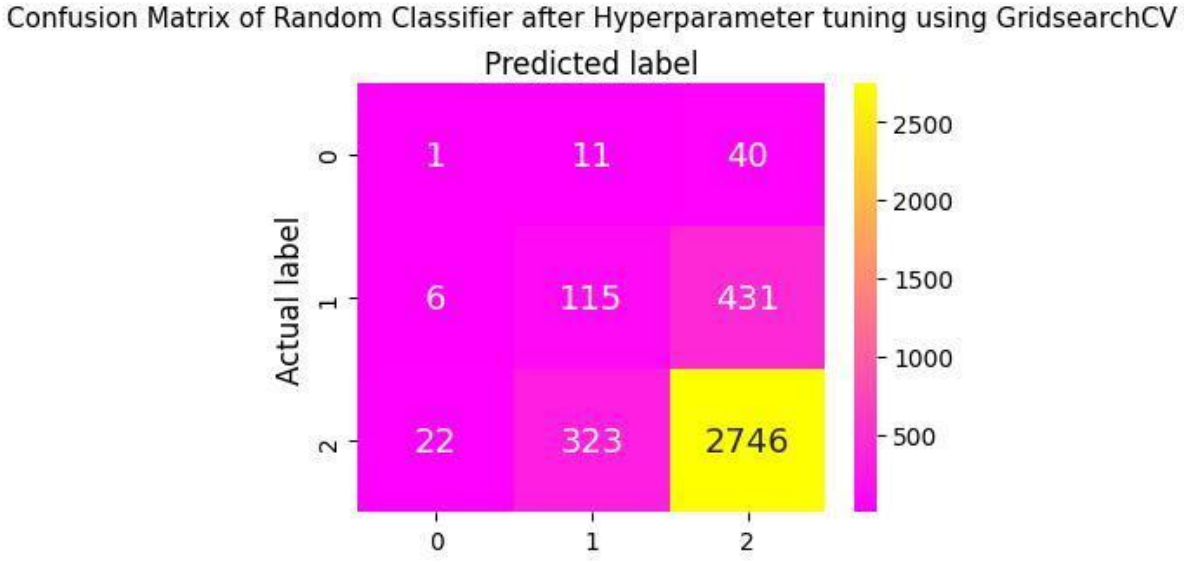
Model Creation :

Create a random forest classifier after hyper parameter tuning.

We achieve accuracy of 91.68% for the hyper tuned random forest classifier.

Final Classifier using Random Forest and GridsearchCV

Best Accuracy is 0.9175912568061043



Classification Report - Tuned Random Forest:

	precision	recall	f1-score	support
0	0.03	0.02	0.02	52
1	0.26	0.21	0.23	552
2	0.85	0.89	0.87	3091

accuracy			0.92	3695
macro avg	0.38	0.37	0.38	3695
weighted avg	0.75	0.77	0.76	3695

PROPOSED MODULAR IMPLEMENTATION

Below is the proposed modular implementation of the project. It consists of two modules:

1. Admin

Admin Module:

1. Login
2. Upload Road Traffic Accidents dataset that was downloaded from Kaggle
3. Exploratory Data Analysis
4. Data Preprocessing
    a. Check for duplicates in the dataset.
    b. Fill missing values
    c. Transform Categorical features using label encoding.
    d. Drop unnecessary features
    e. Balance the dataset using SMOTE
    f. Split the data into Training and Testing Datasets.
5. Feeding the dataset to multiple classification algorithms
    a. Random Forest
    b. Decision Trees
    c. Support Vector Machine
    d. Logistic regression
    e. K-Nearest Neighbour
    f. Gaussian Naïve Bayes
    g. Adaboost
    h. Gradient Boosting
6. Creation of model using Gradient Boosting Classifier

PROJECT EXECUTION AND TESTING

Compare Algorithm Summary: On this page, the admin can feed the dataset to various Algorithms to train them, get the test accuracy for each algorithm and their accuracies are summarized here.

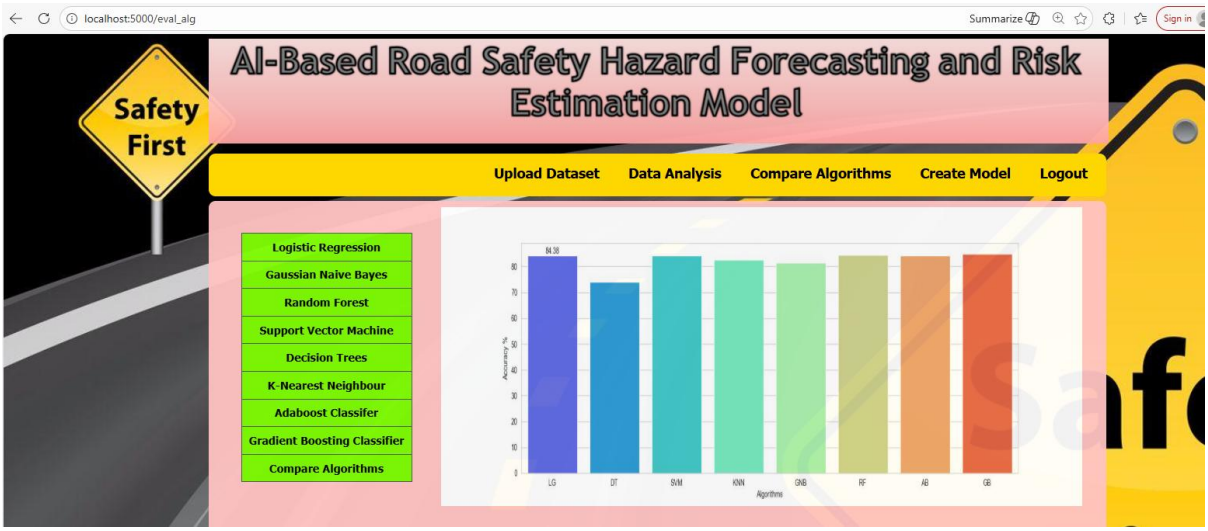


Fig: Compare Algorithm Summary

Create Model:

This screen shows the Accuracy of the Final Classifier Model is 91.68%.

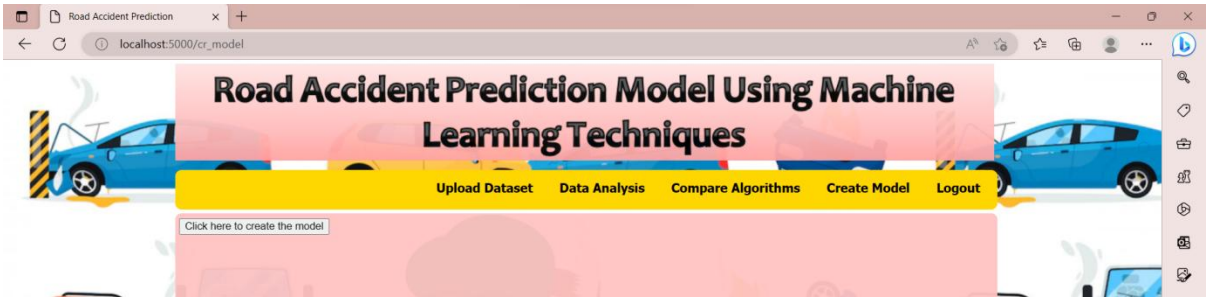
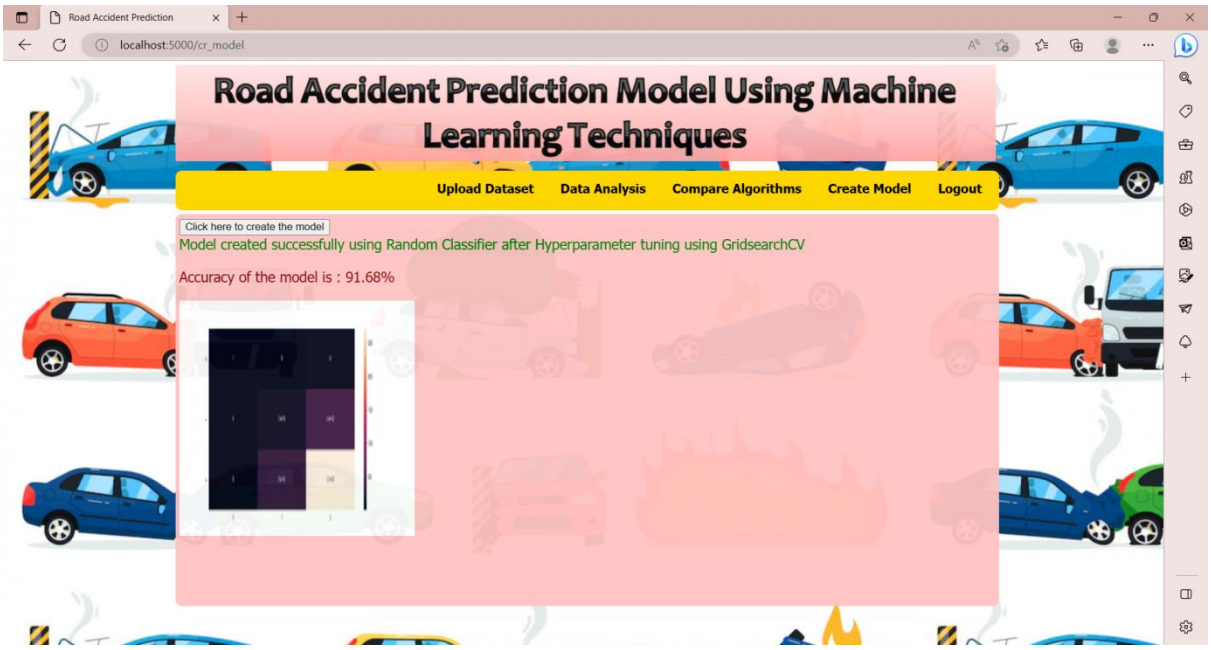


Fig: Create Model



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, so on. 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 risk probability of accidents over different areas with high accuracy.

Future Scope: 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 implement policies 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 also be 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 this model.

REFERENCES

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

[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. Text mining to decipher free-response consumer complaints: Insights from the nhtsa vehicle owner's complaint database. Human Factors 56(6): 1189-1203.

[4] Chen ZY, Chen CC.. Identifying the stances of topic persons using a model-based expectationmaximization method. J. Inf. Sci. Eng 31(2): 573-595.

[5] Williams T, Betak J, Findley B. Text mining analysis of railroad accident investigation reports. In Joint Rail Conference. American Society of Mechanical Engineers V001T06A009- V001T06A009.

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

[7] Sarkar S, Pateshwari V, Maiti J. Predictive model for incident occurrences in steel plant in India. In ICCCNT, IEEE, pp. 1-5.

[8] Stewart M, Liu W, Cardell-Oliver R, Griffin M. 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.

[9] Zheng CT, Liu C, Wong HS. Corpus based topic diffusion for short text clustering. Neurocomputing 275: 2444-2458.

[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: 881-885.