iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** VII **Month of publication:** July 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.45977

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue VII July 2022- Available at www.ijraset.com*

# Road Accident Analysis and Prediction Model using a Data Mining Hybrid Technique

Dr. Shivashankara S[1], Archana B J[2], Vijayalakshmi D[3], Meghana Gowda D[4], Anusha[5]
[1]*Dept. of Computer Science and Engg., Government Engineering College, Krishnarajapet, Karnataka, India*
[2, 3, 4, 5]*UG Scholar, Dept. of Computer Science and Engg., Government Engineering College, Krishnarajapet, Karnataka, India*

***Abstract:*** ***In India, Roads plays major role of Accidents. There are many reasons for occurring accident in India, like consuming alcohol, atmospheric condition, connection points, vehicle fault, speed, road defect etc.. The major goal of this paper is to make up analytical mock up for road accident taking into consideration of earlier years datasets from 2011-2021 and predicting the future result of 2022. The required datasets are collected from data.gov.in website for further processing. An hybrid technique is developed using the data mining techniques such Linear Regression, K-Means Clustering, Association Rule and Naive Bayes algorithms. There are 3 elements of disaster extremity has been researched by forecasting the outcome of 2022 considering all places, and on the basis high and low-level frequency of disasters Clustering of places (states) are done. Comparatively, the proposed model yields the moderately better accuracy rate.***
***Keywords: Association Rule Mining, Classification, Data Mining, Naive Bayes, Prediction of accident, Random Forest.***

## I. INTRODUCTION

Every day lot of vehicles moving on the road, and road accidents may happen anywhere at any time. Due to some accident humans can also loss their lives. As human being, we all want to develop a new techniques to stay away from road accident and save life. Data mining technique is used to find solution to control road accident and drive safer. Accidents are caused by the various influencing factor, like influence of alcohol, vehicle condition, road condition, road negligence in control high speed, not wearing helmets and during normal traffic violations. Increasing the number of vehicles is one of main reason for road accidents. Weather condition such as fog, rain, etc also caused the road accidents. Trucks, buses and heavy vehicles are cause a death kind of accidents. To reduce the accident rates, knowledge about accident spots and causing factor will help. With the advancement of recent technology like data mining and big data analytics, solution for reducing the road accidents can be achieved and saves the human and other animal life too.

Remaining part of this paper is organized as follows: Section II covers the state of art of the domain; Datasets, Description, Preprocessing of data and Parameters considered are discussed in section III; section IV gives the approaches of prediction; The proposed work and outcome is discussed in section V; at the end, conclusive part and references are placed at section VI and VII respectively.

## II. STATE OF ART

Some of the researchers have contributed their research work on prediction of accidents in and around the India are discussed in this section. Sanjay Kumar Singh [2], have presented prediction of accidents. According to authors the distribution of traffic accidental deaths and injuries in India differ according to age, gender, month and year. Age group 35- 60 years is the most susceptible population group, though male category have more death and injuries compared to female category. Furthermore, traffic accidents are relatively higher in poor weather and during working hours.

In [3], authors applied the statistics method and data mining method on the FARS Accident death dataset as an attempt to deal with this problem. Association rules were discovered by Apriori algorithm, classification algorithm was built by Naive Bayes classifier, and K-means clustering algorithm. A new techniques to finding a road accidents in cities based on the junction like 'T' junction and 'Y' junction and road links is proposed [4]. The prediction method helps to avoid the possibility of road accidents in urban junctions. By using this methods recognize the spot, where the accidents more occur and take a safety driving in that particular areas. A descriptive mining technique is applied on earlier accidents data in combination with some other main information as weather condition, speed limit or road defect creates an interesting another potentially useful and helpful outcome for all involved stakeholders is proposed by Poojitha Shetty, Scahin PC [6]. In the proposed system, authors used the Apriori method to predict the patterns of road accidents by analyzing earlier accidents data.

©IJRASET: All Rights are Reserved 

Shanthi, R. Geetha Ramani [7] proposed a research work on highlighting the importance of Data Mining classification method in predicting the factors that influence the road accidents precise to injury severity. Additionally, applied feature selection methods to select the related road accident factors and Meta classifier Arc-X4 to develop the accurateness of the classifiers.

## III.DATASET AND PARAMETRES CONSIDERED

A dataset is a group of data, usually accessible in table.Every vertical portrayal a specified fluctuating. Every horizontal corresponds to provided number in the information set in questions. The record parameters for every values consist like weight and height of an object.The improvement of the prognostic mock up the information set was obtained interior in subordinate form. Subordinate information involve numerical matter or data are not initiated or available from the examiner, but also acquire from others values or given sources like agencies of central government .There are many reason for happening road accident like, due to vehicle defect, weather, location, educational qualification, junction, speed, alcohol etc… For developing the predictive model for road accident we consider mainly main Five variables for example, environment state, drink and drive, spot, vehicles fault with type of different junction point.These five variables again have sub-type like

*1)* Type of Weather state (fog ,hot,Rain, cloud, , snow).
*2)* Type of Junction Point ( Junction, Four arm, Round etc..).
*3)* Type of Location (near bazaar, Industries, College, temple, others).
*4)* Vehicle Faults (Bald tyres, Tyre punctured, Brakes fault, Steering fails, others).

In this proposed work, with the help of data.gov.in website, the data sets are collected. The earlier year's datasets from 2011-2021 were downloaded from website. There are total 10000 of datasets upto 100MB of datasets are collected.

### *A. Dataset Description*

This dataset record consist of 2500 instances and 5 attributes. The Table 1 allows the report of the dataset by considering 5 distinct factors such as Alcohol, Weather, Junction, Location and Vechicles defects.

TABLE I
Dataset Used In The Proposed Work

| Sl. # | Variable Name | Variable Type | Distinct Count |
|---|---|---|---|
| 1. | Alcohol | Integer | 1900 |
| 2. | Weather | Integer | 1750 |
| 3. | Junction | Integer | 1740 |
| 4. | Location | Word | 8000 |
| 5. | Vehicles defects | Integer | 8000 |

### *B. Data Preprocessing*

In the data mining process, the main step is data pre-processing. Data accumulation. Data accumulation methodology are frequently restricted and its outcomes is not possible to information combination, unknown parameter etc... It is essential to change some information into representation to make sure correct, well-organized, or significant study.

All records with missing value (represented by 0 in the datasets) in the chosen attributes were removed. All numerical values were converted to nominal values according to the data dictionary. Nominal data is a type of data that is used to label variables without providing any quantitative value. It is the simplest form of a scale of measure. All examine approved in Weka.

## IV. APPROCHES CONSIDERED FOR PREDICTION

### *A. Linear Regression*

With considering the Linear regression algorithm to identify the past correlation in between an independent and dependent parameter to prognostic the potential data of the dependent parameters. To predict the future behaviour of regression uses model between the parameters of the past correlation. The linear regression model advantages in prognostic the upcoming behaviour of road accident with help of the statistical methods. To examine the algorithm of the variance sets of the depend parameters and the formula to prognostic is applied like $Y=b_0+b_1*X$ for the upcoming behavior.

*B. K-Means Clustering*

This algorithm is used to explore the low and high-frequency accident spot. The algorithm go after uncomplicated along with uncomplicated method to categorize a provided information put along definite numeral of clump firm to deduce. Major goal is to describe 'K' essential 1 for every clump. 'M' essential need to be put down them greatly practicable distant along with one and all. Further measure is to get hold of every spot be in belonging to a provided information set and connect to close essential. While aimless is remaining, the 1st stride is accomplished and an untimely category during age is done. With respect to that point re-evaluate 'K' new essential barry center of centriod from the before steps. Those 'K' essential, a latest wrapping has to be completed in middle or between the similar information set prong or points & the nearby new essential centriod. Curve has been implemented. As the outcome of this curve we can observe the 'K' essential replace their place steps up until no other interchanges done.

*C. Association Rule Mining*

Ahead implementing Algorithms, from chosen attribute the multiple variable with unknown values were removed. In record with information vocabulary in user manual the nominal values are extracted from statistical values.

Good information was kept in .com setup and set to be examine at the information study device. The good information order extract and codification hold 42000 group 4 form accredit, and conclusion accredit. Full desirability changed to statistical value. Following implement inferred design with 0.4 rock bottom hold up and 0.6 bottom assured in wood hen, alliance rule death outlay of dextral conclusion are created. All can watch the death accident connecting alcohol consumption have the more death figure, that way alcohol consumers are most threatening compare to many alternatives. As well fog state along morning will have large death figure, these display not accident figure is larger, also appear in primary information moreover death figure are larger

*D. Naive Bayes*

Navie bayes classifier were construct in good information. The overall 35,675 datasets, 23,995 was appropriately categorized providing 68% exact figure. Navie bayes Categories present the death figure weakly turn on provided accredit. Even though all measured attribute with association to different features in information sets.

## V. PROPOSED WORK AND OUTCOME

Figure 1 demonstrates the proposed system architecture of the accident prediction model. This architecture consists of three main phases such as pre-processing, modelling and result analysis. For predicting controlling the road accidents, different techniques are used and performance estimation of classifiers are discussed in the following parts.

Before construction of each model data preparation was performed. All missing values are removed and all the numerical values are converted to nominal values according to data dictionary. Also the removing of unnecessary attributes are performed.

In Modeling, to show the basic characteristics of the accidental deaths the statistical are calculated. Then classifying the different accident rates in different areas. Based on the dataset collected, future accident rates are predicted. Later comparison of the accident rates in different states with distinct datasets.

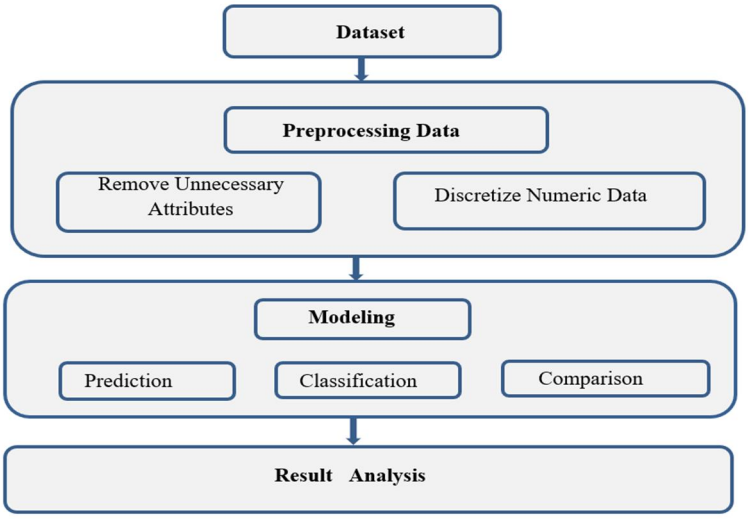



Figure 1. Proposed System Architecture

TABLE III
COMPARATIVE ANALYSIS OF EXISTING WORKS WITH RESPECT TO TECHNOLOGIES USED AND REMARKS

| Ref.& Year | Technologies Used | Results/Remarks |
|---|---|---|
| [11], 2018 | ARIMA & ARIMAX | Finding vehicle & environment condition in time series analysis of crash dataset. |
| [12], 2019 | BPNN LSSVM | Accuracy in prediction was not much good in BPNN & LSSVM to provide adequate distance prediction. |
| [1], 2016 | Log Normal Regression | Accident severity consistently affected by crash types. |
| [3], 2017 | Aprior Associatio, Naïve Bayes | Naïve Bayes Classifier result shows that which states have more death rate. |
| [10], 2017 | KNN, K-Means Clustering | The study help us to derive the statistical model using various techniques. |
| [9], 2015 | K-Modes Clustering and Association Rules | Using rule association predict the accurate fatal rate. |
| [13], 2021 | SVM, Apriori | Predict the risk possibility of road accidents over special areas with more accuracy. |
| Proposed Work | Hybrid Technique | Comparatively better prediction rate for the factors considered such as Alcohol, Weather, Junction, Location and Vehicle defects. |

In the result analysis, the accuracy of different techniques are analyzed. In the proposed work, An hybrid technique is developed using the data mining techniques such Linear Regression, K-Means Clustering, Association Rule and Naive Bayes algorithms.
The table 2 shows, summarizes the technologies used and their respective remarks of the existing works. Also, Figure 2 depicts the comparative analysis considering accuracy the of existing works and proposed work. From the figure 2, it is noticed that, the proposed work yields the summarized and average accuracy of 82% which is moderately better compared to existing methods.

## VI. CONCLUSION

In this paper, an effort has been placed to address about some of the existing works in the related field. It is also presented about the datasets used, data descriptions, and approaches to be used in this proposed work. The system architectures and its three modules are also discussed. At the end, comparative analysis of various techniques used with respect to their remarks in shown in table 2 and overall average accuracy of the existing works and proposed work comparison is depicted in figure 2. From the figure 2, it is observed that the proposed work yields the better accuracy of 82%, which is moderately better compared to existing works.

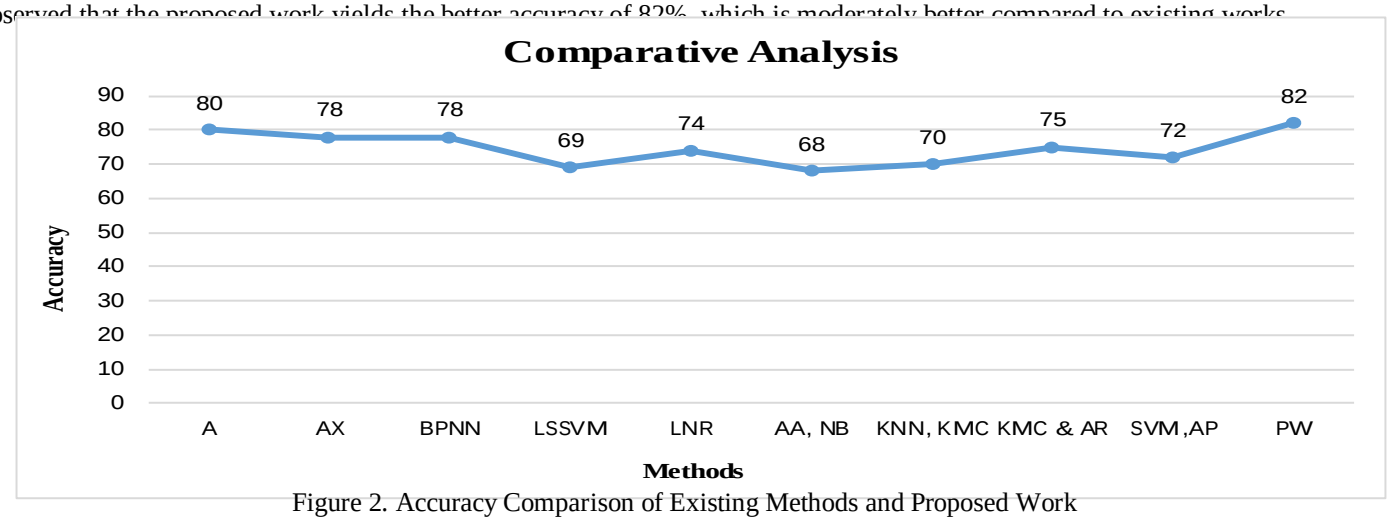
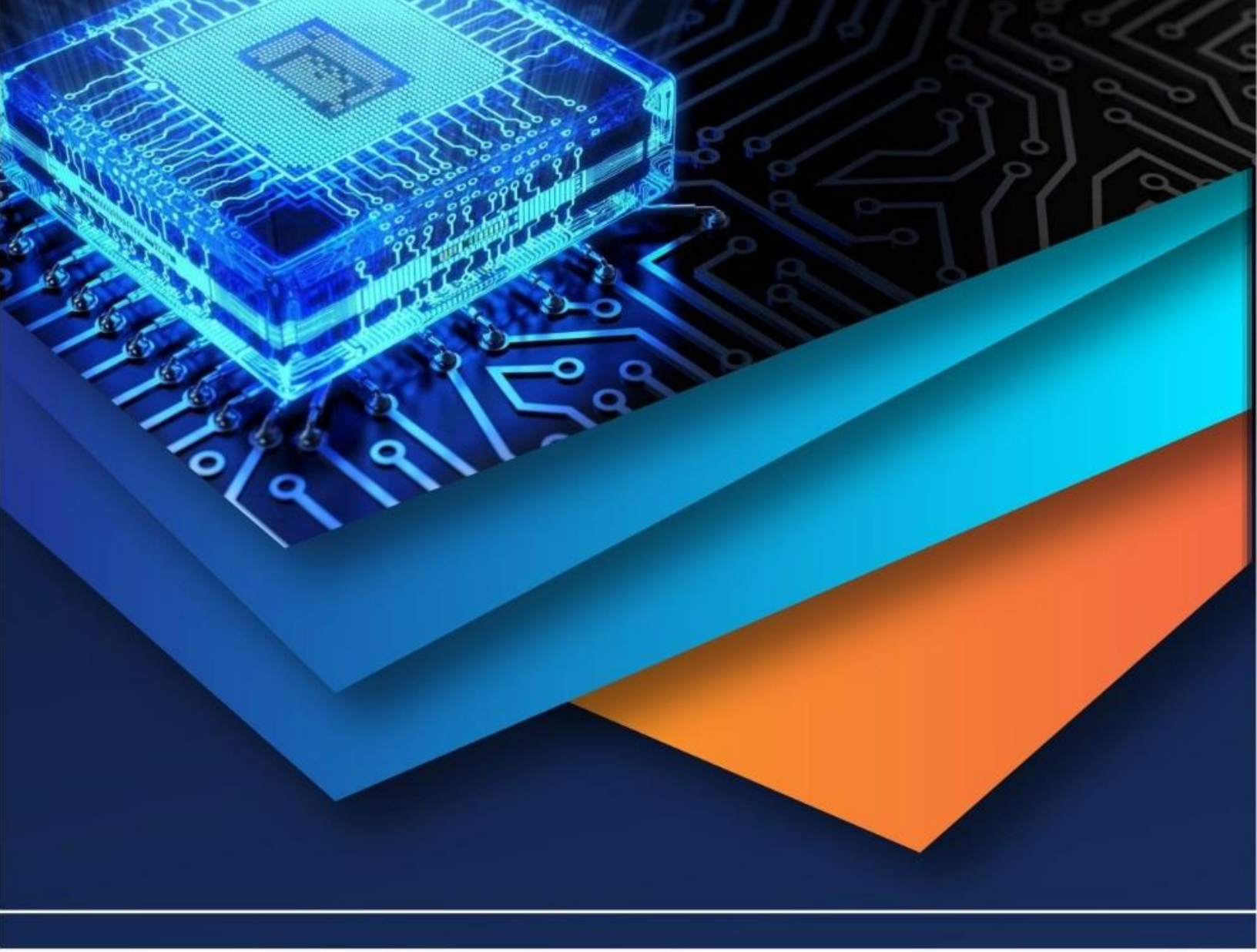


Figure 2. Accuracy Comparison of Existing Methods and Proposed Work

**Note:** A: Arima, AX: ArimaX, BPNN: Back Propagation Neural Networks, LS: Long Short Support Vector Machine; LNR: Log Normal Regression; AA, NB: Aprior Association, Naïve Bayes; KNN: K Nearest Neighbour; K Means Clustering; KMC: K Mode Clustering & Association Rule; Support Vector Machine, Apriori; PW: Proposed Work.

## REFERENCES

[1] Yannis George, "Investigation of road accident severity per vehicle type", World Conference on Transport Research - WCTR 2016 Shanghai, Jul 2016.
[2] Sanjay Kumar Singh, "Road Traffic Accidents in India: Issues and Challenges", World Conference on Transport Research - WCTR 2016 Shanghai, Jul 2016.
[3] Liling Li, Sharad Shrestha, Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques",IEEE Computer Society, pp.363-370, Jun 2017.
[4] Poul Greibe,"Accident prediction models for urban roads", Danish Transport Research Institute, Knuth Winter feldts Allé, DK-2800 Kgs. Lyngby, Denmark, Dec 2001.
[5] Francesca La Torre, "Development of a transnational accident prediction model", 6th Transport Research Arena, Apr 2016.
[6] Poojitha Shetty, "Analysis of road accidents using data mining techniques", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, Vol: 04, Issue: 04, Apr 2017.
[7] S. Shanthi, R. Geetha Ramani, "Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques",Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA. Vol1, Oct 2015.
[8] Dr.K.P.Shivaranjani Karthikeyan, "A Review of Weather Forecasting Using Data Mining Techniques", International Journal of Engineering and Computer Science, ISSN: 2319-7242, Vol 5, Issue 12, Dec 2016.
[9] Sachin Kumar, and Durga Toshniwal, "Analysing road accident data using association rule mining", In Proceedings of International Conference on Computing, Communication and Security, pp:1–6, 2015.
[10] Baye Atnafu, and Gagandeep Kaur", Survey on Analysis and Prediction of Road Traffic Accident Severity Levels using Data Mining Techniques in Maharashtra, India", Dept of CS/IT, Symbiosis Institute of Technology, Pune, India, Vol7, No.6, Nov 2017.
[11] Chukwutoo C. Ihueze, Uchendu O. Onwurah, "Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria", Accident analysis-and-prevention, Vol 112, pp.21-29, 2018.
[12] Junhua Wang, Boya Liu, Ting Fu, Shuo Liu, Joshua Stipancic, "Accident Analysis & Prevention" Vol 112, Sep 2019
[13] Preethi K, Nandini R, "A Road Accident Prediction Model Using Data Mining Techniques", May 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)