iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70424

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 13 Issue V May 2025- Available at www.ijraset.com

Development of Traffic Imaging for Accident Detection System Using AI

Prof. Rahul Bhandekar[1], Sanjana Fulzele[2], Chitaksha Bhoyar[3]
[1]Project Guide, [2, 3]Students, Department of Computer Science & Engineering, Wainganga College of Engineering and Management, Nagpur, Maharashtra.

Abstract: Traffic accidents are one of the leading causes of fatalities and severe injuries, endangering the lives and health of individuals. These accidents might have a variety of reasons, some internal to the driver while others are external. When there is low visibility because of unfavorable weather conditions like rain, clouds, and fog, driving can be challenging and even dangerous. Using algorithmic machine learning and approaches for clustering, this project aims to provide a summary of advanced methods for traffic accident predicting. The rising global vehicle accident rate has profound implications for all aspects of human life. Despite their importance, factors including causality assessment, traffic features, and the connections between different contributing components have typically been ignored. Moreover, the majority of the data on traffic accidents that is now accessible is used for data extraction and basic statistical analysis, which offers limited understanding of patterns and statistics. Through the identification of significant contributing factors and the development of preventative methods, this road accident information category seeks to lessen the severity of subsequent accidents. Machine learning algorithms are used to analyze data, identify hidden patterns, predict the impact of an event, and swiftly disseminate this information.
Index Terms: Road accident data, Machine learning, K-means Clustering, Analysis, Visualization, prediction etc.

I. INTRODUCTION

Recent data on fatalities from the World Health Organization indicates a startling number of collisions that occur on roadways throughout the globe each year. Anywhere on the road can experience an accident. The amount of traffic on today's roadways is increasing at an exponential rate, leading to an increase in accidents. Consequently, accident prediction is one of the most important research areas in transportation safety. Numerous factors, including road geometry, traffic patterns, individual driving styles, and ambient conditions, can have a substantial impact on the likelihood of traffic accidents. Countless studies have been conducted with the aim of determining hazardous areas or "hot spots," examining the features of traffic collisions, and forecasting the frequency of accidents.

Some inquiries focus on the mechanics of accidents. Other issues include the road's visibility and the weather. There's no special technique used by traffic cops to pinpoint an accident-prone area. The weather, people, automobiles, and roads are among the many unpredictable elements that contribute to traffic accidents. For this reason, forecasting traffic accidents is essential to the integrated scheduling and management of traffic. Machine learning algorithms have the capacity to handle categorization factors on a large scale, allowing for the identification of intriguing patterns. It can handle massive volumes of data processing and is scalable.

Furthermore, the clustering technique facilitates the study and display of data pertaining to auto accidents.

II. PROBLEM IDENTIFICATION

The World Health Organization recently revealed death figures, and they indicate an alarming amount of crashes on roads worldwide every year. Accidents on the road may happen anywhere. Today's traffic is expanding at an enormous pace, which causes many more accidents on the roads. Therefore, one of the most significant study fields in safety for transportation is road accident prediction. The probability of traffic accidents is significantly influenced by road shape, traffic flow, driver characteristics, and their surrounding environment. Many research, including those on identifying dangerous locations or "hot spots," analyzing the features of road accidents, and predicting accident frequency have been all carried out.

The workings of accidents is the subject of certain investigations. Weather and the visibility of the road are additional concerns. The traffic police lack a particular method for identifying a hazardous location. The prediction of traffic accidents is crucial to the coordinated organizing and overseeing of traffic since there is a lot of unpredictability in the factors that cause them, including people, cars, roads, weather, and other nonlinear variables. Computer learning.

In order to find meaningful patterns, algorithms could go through a vast number of categorization parameters. It is scalable as well as capable of processing vast volumes of data. Additionally, the clustering approach aids in the evaluation and visualization of data related to traffic accidents.

III. LITERATURE SURVEY

Currently, road traffic accidents require the attention of investigators, civic organizations, automobile companies, governments, and corporate societies around the world as both a development and public health issue.

Masashi Toyoda et. al. 2017, In the framework [1] To assess the frequency of traffic accidents at road crossings, massive amounts heterogeneity accident data is employed. The authors proposed two object identification methods: XGBoost to extract features from driving logs and route maps, and Fast R-CNN to extract image features. Overall experimental results indicate that the strategies employed in this investigation are capable of successfully identifying potentially dangerous crossings.

Jonardo R. Asor et. al. 2018, Nave Bayes, Decision Trees, and Rule Inference are used in [2] to classify and uncover concealed trends using a data set that comprises crash records received from the Philippine National Police (PNP). They examine the accident data using a Rapid miners data mining programme. The locations of incidents don't significantly affect whether or not victims die, according to the authors. The results emphasise significant components that lead to accidents and show that the most crucial variables when assessing the severity and demise of victims of traffic accidents are day and hour, and the algorithms operate with the predicted accuracy.

Sadiq Hussain et. al. 2018, This research article states that a large number of studies have been conducted to predict the main factors that influence accidents, including the causes of collisions, locations that are prone to collisions, the magnitude of the collision, as well as the sort of vehicle involved, in order to improve the effectiveness of the DM categorization [3]. The J48, Multi-layer Perceptron is, and Bayes Net classifiers in 150 instances in the dataset were assessed by the authors using Weka and Orange, two data mining tools. Evaluation criteria such as recall, accuracy, precision, or sensibility are used to analyse data mining methodologies and choose the best algorithm with the accident dataset prediction. The results of the experiment show that Multi-layer Perceptron is the most accurate in predicting the mishap database, with an accuracy of 85.33%.

Tibebe Beshah Tesemaet. al. 2012, The scientists employed a genetic algorithm to develop a symbolic fuzzy classification after gathering accident data collected by the Addis Abeba transportation office. the classifier that selects features from the erroneous dataset using symbols. The result shows that the developed classifier is capable of differentiating and classifying injuries, and that the characteristics that were utilised to identify the data are readily retrievable and explorable [4].

Tibebe Beshah Tesema et. al. 2011, This author also provided an experimental research on machine learning using data on traffic crashes collected in Ethiopian [5]. Using the CART, which stands Random Forest, Random Forest MARS, and Forest Net algorithms, they developed a model for predictions that investigated the issue of reliability of data and forecasted the impact of driving patterns on potential injury risks. The reasons of the disaster's severity that are connected to humans can be identified using simulations. The mix of methodologies used in this experiment enhanced the forecast precision.

Girija Narasimhan et. al. 2017, was developed with the goal of predicting the number of fatalities in Oman in the future by combining advanced machine learning algorithms with predictive analytics. The author employed a boosting tree regress model, that is based upon a decision tree using a multiplicative model, to increase the prediction accuracy [6]. This article indicates that nonhuman variables account for the remaining 9% of accidents, whereas people account for around ninety-one percent of all mishaps as the primary or significant contributing factor.

Li, L, Shrestha, et. al. 2018, Apriori, Naïve Bayes, K-means clustering algorithms are used in [7] to investigate the relationship between death rates and other parameters such driving while inebriated, light condition, accident style, weather condition, and road surface conditions. This study attempts to highlight the elements that are highly associated with fatal accidents. The result indicates that human factors—like drunk drivers, for example—contribute to a high fatality rate.

Ms. Nidhi. R,et. al. 2018, On the other hand, patterns in collisions were found using the Nave Bayes as well as Apriori approaches in the research in [8]. In order to anticipate the accident types that are likely to happen on new roads, the authors of this study developed a model for forecasting based on the law of association. The results of the analysis show that cars under five years old are involved in most accidents, and that the fatality rate is higher in rural areas.

Data mining techniques are important for assessing and projecting the value of traffic accident data in the future and for identifying patterns in the event elements that impact different metrics, according to multiple studies. Additionally, the enormous potential of information mining prediction approaches contributes significantly to avoiding and tracking the issues with road accident safety.

IV. METHODOLOGY

During the development and evaluation stages, the system makes advantage of machine learning. The proposed model appears to be,

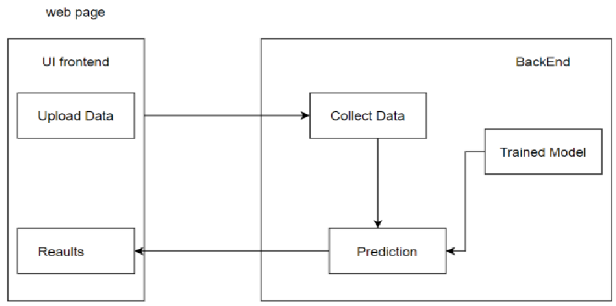


Fig.1. Block Diagram of system

The data are the most crucial component of any data analysis. It is crucial to get the correct sort of data. It is important to pay close attention to analyzing and comprehending the data's structure and substance. The data utilised in this research came from Kaggle or government sources.

The next step is to analyse the data once it has been collected. We require a tool that makes the job simpler in order to analyse the data. We were certain that Python would be employed for coding.

Pandas and Numpy were the two programmes that had the biggest impact on the analysis. Pandas is an analytical and editing tool for data. It has particular algorithms and data structures for handling time series and numerical tables. It provides quick, easy-to-use data analysis tools and frameworks.

"NumPy" is an acronym for Numerical Python. It is a rapid array or matrix processing software that is open source. NumPy is the primary NumPy Python package for scientific computing.

I'll now discuss the method we employed. Numerous algorithms are available to aid with data analysis. Techniques like artificial intelligence and data analytics are quite helpful in this area. We selected the Regression Research algorithm.

A group of statistical techniques known as logistic regression analysis are used to estimate the associations between variables. When the emphasis is on the connection between the variable that is dependent and any number of independent variables, it encompasses a variety of strategies for modelling and analyzing multiple variables.

A. Flow Diagram :

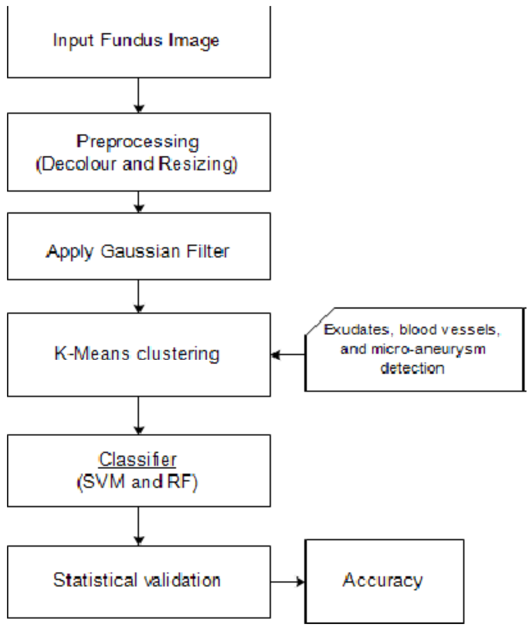


Fig. 2. Flow Chart

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

B. Architecture of system:

The most important part of any analysis of data is the data. It is essential to have the right kind of information. Understanding and interpreting the structure and content of the data requires careful attention to detail. The information used in this study was obtained from government or Kaggle sources.

After the data is gathered, the analysis process begins. To analyse the data, we need a tool which renders the work easier. We knew that the coding would be done in Python.

The two programmes which influenced the analysis the most were Numpy and Pandas. Pandas is a data analysis and editing tool. For managing time series & numerical tables, it contains specific algorithms and data structures. It offers quick and simple frameworks and tools for data analysis.

The term "NumPy" stands for Numerical Python. It's an open source quick array or matrix computing software. The main NumPy Python library for scientific computing is called NumPy.

I'll now go over the approach we took. There are many different algorithms available to help in data analysis. In this context, methods such as artificial intelligence and information analytics are quite beneficial. The Regression Analysis algorithm was our choice.

Logistic regression analysis is a statistical approach used to assess the connections between variables. It includes a range of approaches to multiple variable modelling and analysis where the focus is on the relationship between the dependent variable and any number of variables that are independent.

V. IMPLEMENTATION AND RESULTS

Accident data records are used to develop models that may be used to comprehend various factors, such as weather patterns and road conditions. This can assist users in calculating safety precautions that are helpful in preventing mishaps. By contrasting two situations based on projections that are outside of the sample, the statistical approach based on directional graphs may be demonstrated. In order to discover statistically significant elements that may be utilised to perform a threat and minimise it, the model is run to estimate the probability of collisions and injuries.

Here, a road incident study is conducted by data analysis using pertinent research questions. questions like when driving is the riskiest and what percentage of accidents happen in rural, urban, and other places. What is the annual trend in the amount of accidents, are there more casualties in crashes in locations with high speed limits, and so on. Microsoft Excel may be used to access these data and retrieve the necessary response. The goal of this research is to identify the information that is most crucial to a traffic accident so that predictions may be formed.

A. Data Importing:

In order to analyse the data, we have imported it here (fig. 3). A few characteristics of this data include area, alarm kind, ehourcat, visibility, weather, accident severity, and pothole severity. and information. Head(10) examines the data frame's top 10 rows.

```
In [36]: import pandas as pd
         import numpy as np

In [37]: data = pd.read_excel('DATASETS.xlsx')

In [38]: data.head(10)
Out[38]:
```

	Area	Alarm_Type	ehourCat	weather_condition	visibility	Accident_Severity	Pothole_Severity
0	Kadugodi	PCW	Early	fog	Low	Medium	Medium
1	Kadugodi	PCW	Early	fog	Low	Medium	Medium
2	Garudachar Playa	FCW	Early	fog	Low	Medium	Medium
3	Garudachar Playa	FCW	Early	fog	Low	Medium	Medium
4	Hudi	Overspeed	Early	fog	Low	Medium	Medium
5	Hudi	Overspeed	Early	fog	Low	Medium	Medium
6	Kadugodi	HMW	Early	fog	Low	Medium	Medium
7	Kadugodi	HMW	Early	fog	Low	Medium	Medium
8	Hudi	Overspeed	Early	fog	Low	Medium	Medium
9	Hudi	Overspeed	Early	Rainy	Low	Medium	Medium

Fig 3 : Datasets

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

B. Converting the data into numerical form:

In [43]:
```
data.head(10)
```
Out[43]:

	Area	Alarm_Type	ehourCat	weather_condition	visibility	Accident_Severity	Pothole_Severity
0	1	1	1	1	0	2	2
1	1	1	1	1	0	2	2
2	2	2	1	1	0	2	2
3	2	2	1	1	0	2	2
4	3	3	1	1	0	2	2
5	3	3	1	1	0	2	2
6	1	4	1	1	0	2	2
7	1	4	1	1	0	2	2
8	3	3	1	1	0	2	2
9	3	3	1	2	0	2	2

Fig 4 : Numerical Data

Here, we have numerically transformed the data (figure 4) as, as we all know, machine learning model training requires numerical data. and these particulars.Head (10) shows the top ten rows of the data frame where the numerical data is created.

C. Used k-means clustering algorithm:

We used the k-means cluster approach after the datasets were transformed to numerical form because our datasets did not contain labels or binary outcomes at the conclusion. The datasets are therefore unlabeled. For this reason, we are using an unsupervised learning strategy. Thus, we forecasted the road accident forecast zone outcomes at the conclusion of the final column that is in high/low utilising the k-means clustering technique after employing the unsupervised learning strategy.

D. Kmeans Algorithm :

1) Select K points as the intial centroids.
2) Repeat,
3) From K clusters by assigning all points to the closest centroid,
4) Re-compute the Centroid of each cluster,
5) Until the centroid don't change.

E. Clustering:

In [47]:
```
data.head(10)
```
Out[47]:

	Area	Alarm_Type	ehourCat	weather_condition	visibility	Accident_Severity	Pothole_Severity	labels
0	1	1	1	1	0	2	2	1
1	1	1	1	1	0	2	2	1
2	2	2	1	1	0	2	2	1
3	2	2	1	1	0	2	2	1
4	3	3	1	1	0	2	2	1
5	3	3	1	1	0	2	2	1
6	1	4	1	1	0	2	2	1
7	1	4	1	1	0	2	2	1
8	3	3	1	1	0	2	2	1
9	3	3	1	2	0	2	2	1

Fig 5 : Numerical Labeled Data

As you can see in Figure 5, the results of our application of k-means clustering indicate whether the accident forecast zone is high or low. The labels for the last column are found here. That is the numerical result. In this case, High is represented by (0) and Low by (1).

Additionally, you may view the outcome in String dataset form below the table (fig. 6).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Konena Agrahara	UFCW	PeakM	fog	High	High	Low	High
A Narayanapura	UFCW	PeakM	fog	Low	High	High	High
Hoysala Nagar	FCW	PeakM	Rainy	Low	Low	Low	Low
Hoysala Nagar	FCW	PeakM	Rainy	Low	Low	Low	Low

Fig 6: Labeled datasets

Following clustering, we used the Logistic Regression approach to make predictions, and the results showed 86% accuracy (fig 7).

```
Machine Learning Model Build

LR Alg Accuracy 86
output Prediction LR =  [0 0 0 0 0 1 0 0]
```

Fig 7 : Logistic Regression Accuracy

F. *Visualization:*

```
ehourCat  Early  PeakM  RegularM
labels
High         78     76       143
Low          21     43        38
```

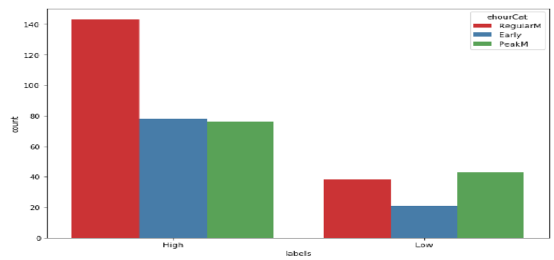


Fig 8: Graph of Labels vs Ehourcart

This graph, which is based upon the ehourCat (fig. 8), displays the counts rate outcome depending on the three ehourCat qualities, namely Early, PeakM, and RegularM. There are three hues: green, blue, and red. That is predicated on the outcome of high and low. In this case, red denotes early, blue indicates PeakM, and green indicates RegularM.

```
weather_condition  Cloudy  Rainy  fog
labels
High                  101    100   96
Low                    44     29   29
```

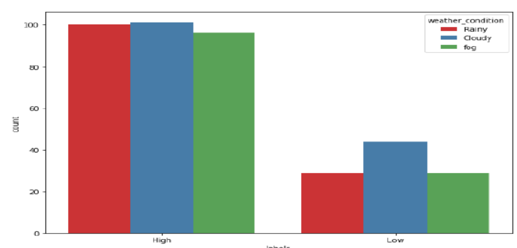


Fig 9: Graph of Labels vs weather condition

This graph (fig. 9) plots the counts' findings according on the meteorological conditions. In this instance, the colours red, blue, and green denote fog, rain, and clouds, respectively.

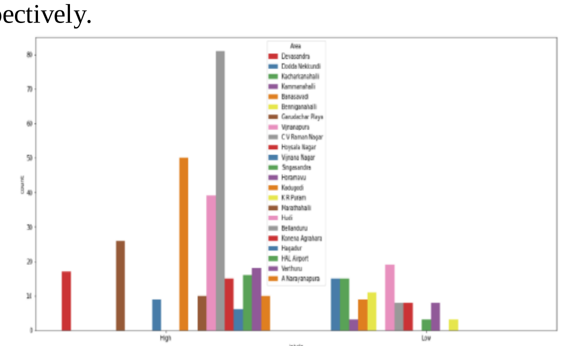


Fig 10: Graph of Labels vs Areas

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

This graph (fig. 10) displays the findings of counts depending on regions.

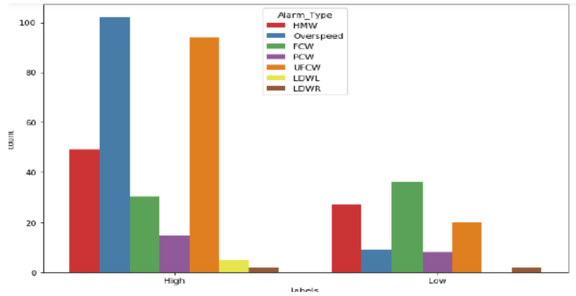


Fig 11: Graph of Labels vs Alarm_Type

Figure 11 plots this line according to the alarm kinds, which include FCW, HMW, LDWL, LDWR, overspeed, PCW, and UFCW.

visibility	High	Low	low
labels			
High	57	44	1
Low	100	194	3

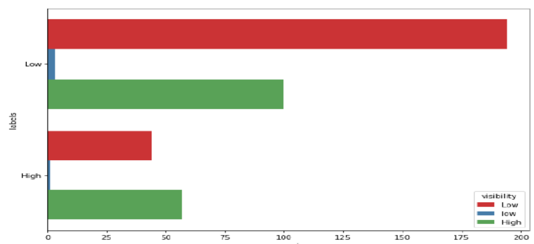


Fig 12: Graph of Labels vs Visibility

This is a visibility graph (fig. 12), which also displays the numbers based on low and high label results. In this case, the hue red denotes low results, whereas the colour blue denotes great results.

Alarm_Type			
FCW	13	23	30
HMW	32	13	31
LDWL	2	0	3
LDWR	0	2	2
Overspeed	38	0	73
PCW	8	2	13
UFCW	42	7	65

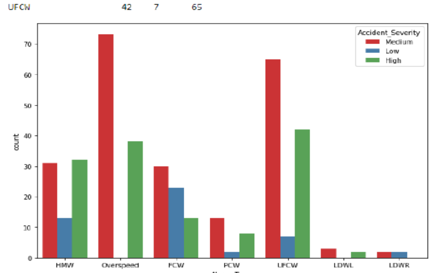


Fig 13: Graph of Alarm Type vs Accident Severity

Here, the accident severity is shown as medium, high, and low in this graph (fig. 13). as opposed to alarm types FCW, HMW, LDWL, LDWR, Overspeed, PCW and UFCW.

Pothole_Severity	High	Low	Medium
labels			
High	142	25	130
Low	7	62	33

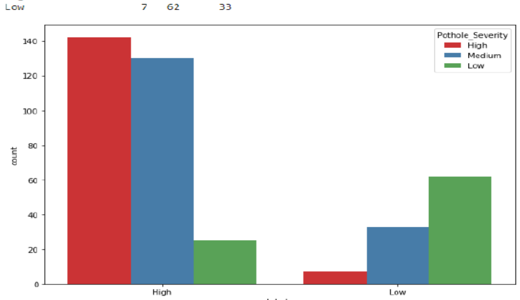
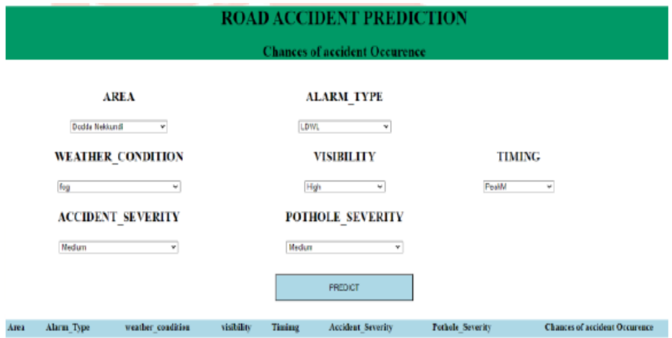


Fig 13: Graph of Labels vs Pothole Severity

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Here, the pothole severity is shown as medium, high, and low on this graph (fig. 13). Additionally, it displays the numbers depending on the low and high label findings. In this case, red denotes a medium outcome, blue denotes a high result, and green denotes a low result.

G. Web page:



Here, information must be entered and the predict button must be hit in order to determine the likelihood of an accident.

VI. CONCLUSION

Road Accidents are caused by various factors. By going through all the research papers it can be concluded that Road Accident cases are hugely affected by the factors such as types of vehicles, age of the driver, age of the vehicle, weather condition, road structure and so on. Thus we have build an application which gives efficient prediction of road accidents based on the above mentioned factors.

The classification algorithm of the entire dataset. In the Road Accident prediction final result is to find the percentage of accident in particular area. Having lower number of features helps the algorithm to converge faster and increases accuracy. In the Road Accident prediction final result is to find the percentage of accident in particular area. Then we apply logistic regression on these features and obtain the least error

Since k-means clustering is an unsupervised learning technique used for unlabeled data, the data in this article are not labelled into any particular cluster. Additionally, regression analysis with a sizable accident data set was employed in this work to determine the causes of traffic accidents. Plotting of the identified contributing elements to the accident is done through analysis and is shown as a graph. This provides a great deal of insight into accident situations and causes. And in the end, this aids the government in modifying road safety regulations to account for various accident and circumstance kinds.

REFERENCES

[1] Sachin Kumar, Durga Toshniwal, "A data mining approach to characterize road accident locations", J. Mod. Transport. 24(1):62–72.
[2] Tessa K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots", Accident Analysis and Prevention 41,359–364.
[3] Shristi Sonal and Saumya Suman "A Framework for Analysis of Road Accidents" Proceedings of International Conference on Emerging Trends and Innovations in Engineering and Technological Research (ICETIETR)
[4] Analysis of road accidents in India using data mining classification algorithms- E. Suganya,S.Vijayrani.
[5] Hao, W., Kamga, C., Yang, X., Ma, J., Thorson, E., Zhong, M., & Wu, C., (2016), Driver injury severity study for truck involved accidents at highway-rail grade crossings in the United States, Transportation research part F: traffic psychology and behavior, 43, 379-386.
[6] Li, L., Shrestha, S., & Hu, G., (2017), Analysis of road traffic fatal accidents using data mining techniques, In Software Engineering Research, Management and Applications (SERA), IEEE 15th International Conference on (pp. 363-370). IEEE.
[7] El Tayeb, A. A., Pareek, V., & Araar, A. (2015). Applying association rules mining algorithms for traffic accidents in Dubai. International Journal of Soft Computing and Engineering.
[8] Bahram Sadeghi Bigham ,(2014),ROAD ACCIDENT DATA ANALYSIS: A DATA MINING APPROACH, Indian Journal Of Scientific Research 3(3):437-443.
[9] Divya Bansal, Lekha Bhambhu, "Execution of Apriori algorithm of data mining directed towards tumultuouscrimesconcerningwomen", International Journal of AdvancedResearch in Computer Science and Software Engineering, vol. 3, no. 9, September 2013.
[10] S. Krishnaveni, M. Hemalatha, "A perspective analysis of traffic accident using data mining techniques", International Journal of Computer Applications, vol. 23, no. 7, pp. 40-48, June 2011.
[11] Masashi Toyoda, Daisaku Yokoyama, Junpei Komiyama, Masahiko "Itoh 2017 IEEE International Conference on Big Data," Boston, MA, 11-14. 2017
[12] Jonardo R. Asor, Gene Marck B. Catedrilla, "A study on the road accidents using data Investigation and visualization in Los Baños, Laguna, Philippines" 2018 International Conference on Information and Communications Technology (ICOIACT) Yogyakarta, Indonesia. Page s: 96 – 101. 6 March 2018

[13] Sadiq Hussain, L. J. Muhammad, F. S. Ishaq, Atomsa Yakubu and I. A. Mohammed "Performance Evaluation of Various Data MiningAlgorithms on Road Traffic Accident Dataset" Information and Communication Technology for Intelligent Systems pp 67-78, 30 December 2018.
[14] TibebeBeshahTesema, Ajith Abraham, Dejene Ejigu, Learning the Classification of Traffic Accident Types"2012 Fourth International Conference on Intelligent Networking and Collaborative Systems. Page s: 463 – 468, 19-21Sept. 2012
[15] Tibebe Beshah Tesema, Dejene Eijigu, Ajith Abraham, "Knowledge Discovery from Road Traffic Accident Data In Ethiopia" 2011 World Congress on Information and Communication Technologies Page s: 1241 – 1246.Year: 2011
[16] Girija Narasimhan, Ben George Ephrem, et al, "Predictive Analytics of Road Accidents in Oman using Machine Learning Approach"2017 International Conference on Intelligent Computing, Instrumentation, and Control technologies (ICICICT) Page s: 1058 - 1065 July 2017
[17] Li, L, Shrestha, S, and Hu, G, "Analysis of Road Traffic Fatal Accidents using Data Mining Techniques," 2017IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) Page s: 363 – 370, Year: 2017
[18] Ms. Nidhi. R, Ms. Kanchana V, "Analysis of Road Accidents Using Data Mining Techniques" International Journal of Engineering & Technology, Volume 7 (3.10) page 40-44, (2018)
[19] Sadiq Hussain "Survey on Current Trends and Techniques of Data Mining Research "London Journal of Research in Computer Science and Technology Volume 17 | Issue 1 | Compilation 1.0 (March 2017)
[20] Jesus Alcal ´ a-Fdez, Rafael Alcal ´ a, and Francisco Herrera "A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems With Genetic Rule Selection and Lateral Tuning" IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 19,NO. 5, OCTOBER 2011
[21] Niklas Donges Machine learning-blog.com Tutorials and explanations about applied Machine Learning "The Random Forest Algorithm," 06 February 2016.