e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

PREDICTIVE MODEL FOR PREVENTING ROAD ACCIDENT CONSIDERING VARIOUS INFUENCING FACTORS

Karthik NS*1, Dr. Nuthan AC*2

*1Students, GMIT Mandya, India.

*2HOD of DE & CS and Professor, GMIT Mandya, India.

DOI : https://www.doi.org/10.56726/IRJMETS29604

ABSTRACT

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion.

I. INTRODUCTION

Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes, while business intelligence covers data analysis that relies heavily on aggregation, focusing on business information. Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous decades. Regression algorithm is method we use to analyze the road accident data and helps in predicting the future result based on the previous data sets. Similarly the K-means clustering algorithm is use to find out which states are similar to each other considering fatal rate, and which states are safer or more risky to drive, clustering algorithm was performed on the fatal accident dataset. We used the road accident datasets for our study. The datasets are downloaded from the website data.gov.in . we collected data sets from 2010-2020. The analysis part involves the collecting the data and performing the statistics operation on the previous datasets and predicting the result of 2022

II. METHODOLOGY

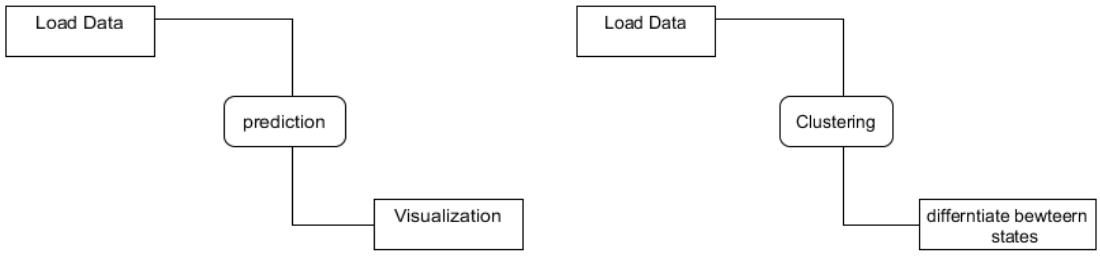


Fig 1: Data Flow Diagram

A data flow diagram (DFD) is a way of representing a flow of data of a process or a system (usually an information system). The DFD also provides information about the output and inputs of each entity and the process itself. A data flow diagram has no control flow, there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart. There are several notations for displaying data flow diagrams.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

III. MODELING AND ANALYSIS

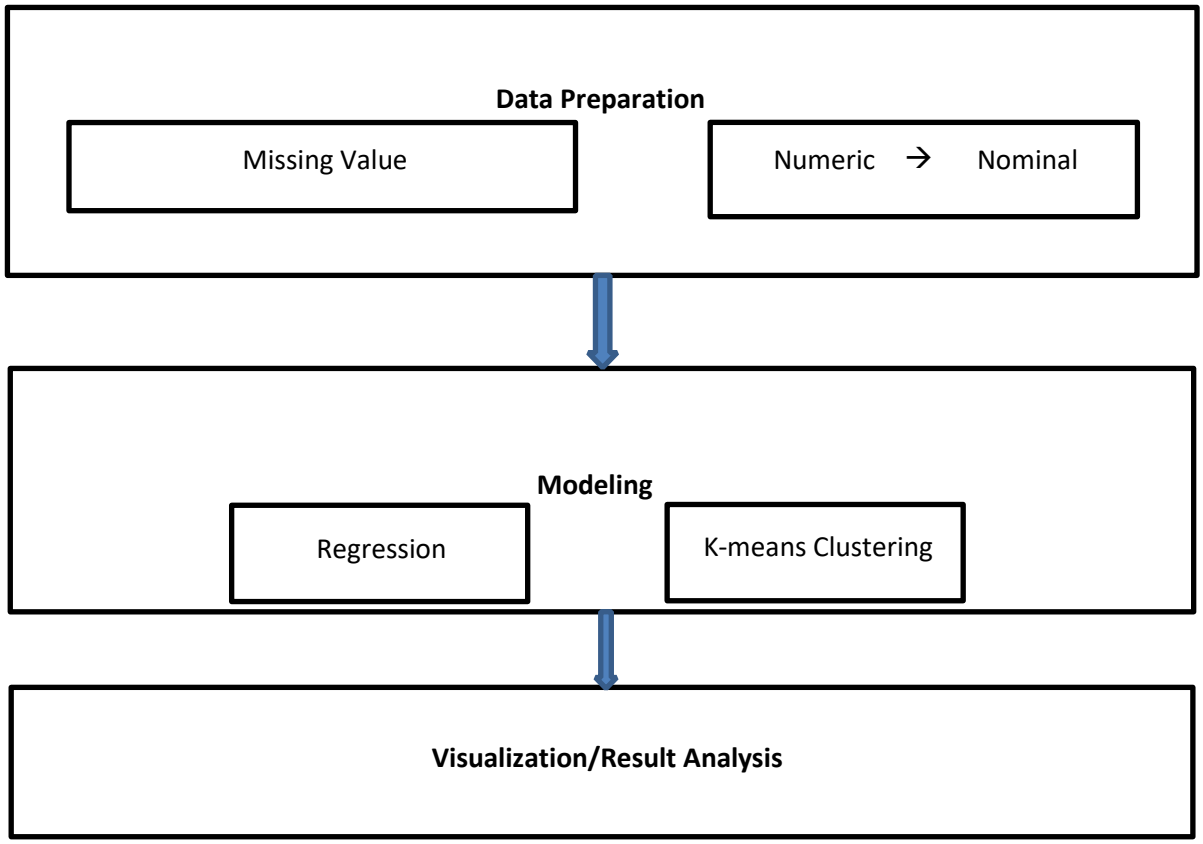


Fig 2: System Architecture

DATA PREPARATION

Data preparation was performed before each model construction. All records with missing value (usually represented by 0 in the dataset) in the chosen attributes were removed. All numerical values were converted to nominal value according to the data dictionary.

MissingValues

Occurs when the no data value is stored for the observation

Modeling

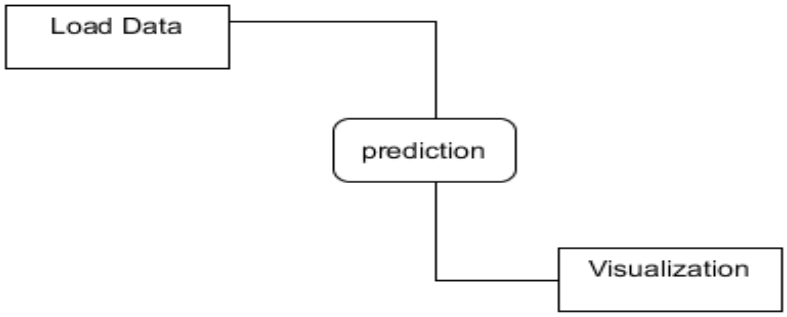
We first calculate several statistics from the dataset to show the basic characteristics of the education, then applied Regression and clustering relationships among the attributes and the patterns. The results of our analysis include Prediction rules among the variables, clustering of states in the INDIA on their populations and number of educational rate. We used a data analytic tool Highcharts to perform these analysis.

DATA FLOW DIAGRAM

Data flow diagram (DFD) is a way of representing a flow of data of a process or a system (usually an information system). The DFD also provides information about the output and inputs of each entity and the process itself. A data flow diagram has no control flow, there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart. There are several notations for displaying data flow diagrams.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

DFD: 0 Level



DFD 0; Clustering

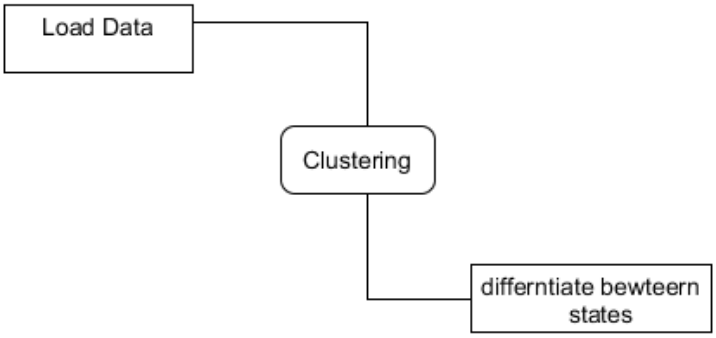


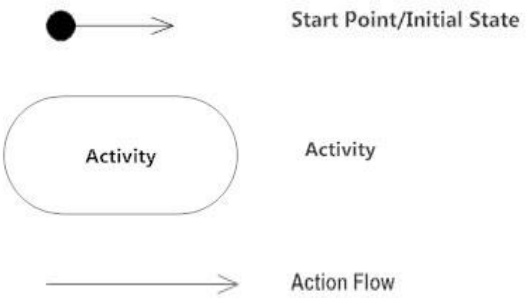
Figure 3

ACTIVITY DIAGRAM

An activity diagram visually presents a series of actions or flow of control in a system similar to a flowchart or a data flow diagram. Activity diagrams are often used in business process modelling. They can also describe the steps in a use case diagram. Activities modelled can be sequential and concurrent. In both cases an activity diagram will have a beginning (an initial state) and an end (a final state).

VISUALIZATION

- Prediction of Dashboard for all the above.
- The year wise datasets from 2010-2022 are included in the prediction dashboard as bar graphs, pie charts, etc.
- The dashboard is more use full for the respective authorities that help them to know about the accident that goings to happen in the current years, so they can take some preventive measures.
- The purpose of hierarchical assessment component to analyses drivers speed adjustment under different zone.
- The Application should produce the yearly report about the road Accident.
- Year wise comparison of Road Accident.



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

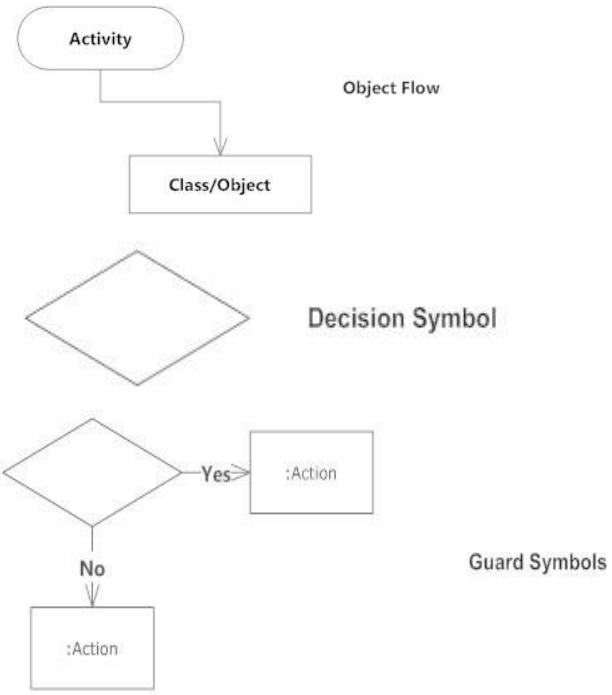
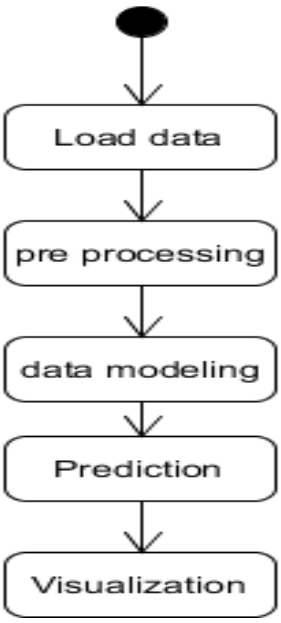


Fig 4: Activity Diagram

ACTIVITY DIAGRAM:



CLUSTERING:

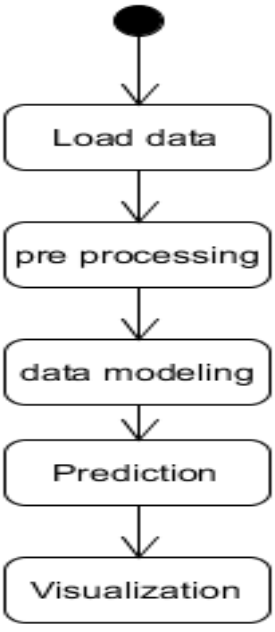


Fig 5: Activity Diagram

SEQUENCE DIAGRAM

Sequence diagrams describe interactions among classes in terms of an exchange of messages over time. They're also called event diagrams. A sequence diagram is a good way to visualize and validate various runtime scenarios. These can help to predict how a system will behave and to discover responsibilities a class may need to have in the process of modelling a new system.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

Sequence Diagram:

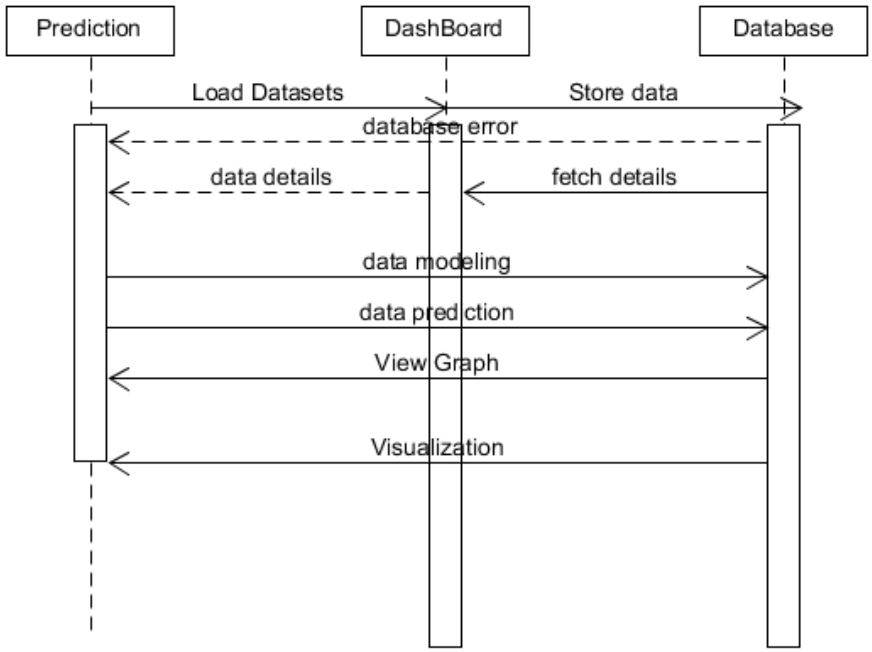


Fig 6: Sequence Diagram

IV. RESULTS AND DISCUSSION

Experimental Setup

To calculate the performance of the proposed algorithms, experiments are conducted against five variants accident data sets(climate data, Location data, vehicle defect, Alcohol based, and junction based). Table shows the characteristics of data set. Experiments were performed on a computer with 2.20 GHz Intel Core processor and 4gb of memory, running windows 8.1.All the Algorithms implemented in java.

Characteristics of Datasets:

Weather Condition:

Table 1: Weather condition

Data sets	States	#Trans	Date
Fog	29Statse+7 Territories	2224	2010-2020
Snow	29Statse+7 Territories	2224	2010-2020
Hot	29Statse+7 Territories	2224	2010-2020
Cold	29Statse+7 Territories	2224	2010-2020
Heavy Rain	29Statse+7 Territories	2224	2010-2020
Light Rain	29Statse+7 Territories	2244	2010-2020

Location Type:

Table 2: Location Type

Data sets	States	#Trans	Date
Near By school	29Statse+7 Territories	2224	2010-2020
Near By Temple	29Statse+7 Territories	2224	2010-2020
Near By Factory	29Statse+7 Territories	2224	2010-2020
Near By Bazar	29Statse+7 Territories	2224	2010-2020
Near By Hospital	29Statse+7 Territories	2224	2010-2020
Near By Busstop	29Statse+7 Territories	2244	2010-2020

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

Vehicle Defect:

Table 3: vehicle defect

Data sets	States	#Trans	Date
Defective Breaks	29Statse+7 Territories	2224	2010-2020
Bald Tyres	29Statse+7 Territories	2224	2010-2020
Punctured	29Statse+7 Territories	2224	2010-2020
Defective Steering	29Statse+7 Territories	2224	2010-2020
Others	29Statse+7 Territories	2224	2010-2020

Junction Based Accident:

Table 4: Junction

Data sets	States	#Trans	Date
T Junction	29Statse+7 Territories	2224	2010-2020
Y junction	29Statse+7 Territories	2224	2010-2020
Four Arm	29Statse+7 Territories	2224	2010-2020
Rail Cross	29Statse+7 Territories	2224	2010-2020
Road cross	29Statse+7 Territories	2224	2010-2020

Alcohol Based Accident

Table 5: Alcohol Based

Datasets	States	#Trans	Date
Alcohol	29 states + 7 territories	2340	2010-2020

Test Cases:

A test case is a set of conditions or variables under which a tester will determine whether a system under test satisfies requirements or works correctly. The process of developing text causes can also help find problems in the requirements or design of an application.

Table 6: Test Cases

C No	scenario	Required Input	Expected output	Actual output	Test Result
1	Enter Prediction values	Enter a valid values	Should predicted successfully	predicted successfully	Pass
2	Enter clustering values	State,year, type	Should cluster successfully	cluster successfully	Pass
3	Enter clustering values	Accident Categories, Subcategories, year	Clustering of states done successfully	cluster successfully	Pass
4	Enter Analysis values	Accident Categories, Subcategories, year	Analysis done Successfully for the year 2019	Analysis successfully	Pass
5	Enter Prediction values	Enter a valid values	Should predicted successfully	Database error	fail
6	Enter clustering	State,year,	Should cluster	Database error	Fail

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

	values	type	successfully		
7	Enter clustering values	Accident Categories, Subcategories, year	Clustering of states done successfully	Data is not fetched properly	Fail
8	Enter Analysis values	Accident Categories, Subcategories, year	Analysis done Successfully for the year 2019	Data is not fetched properly	Fail

The following tables show the various test causes scenarios that are generated along with the required inputs o the given scenarios, expected outputs, actual output and the result whether the test passes or fail

Result Analysis

The Result Analysis mainly don on the accuracy measures. Accuracy measures involves comparing each algorithm's result with the actual result. The result for each algorithm are represented in tabular format ordered by their rank. The actual result is obtained by Road accident per fatality and representing them in the format similar to algorithmic result.

To calculate the Accuarcy of Algorithm we predicted the result of 2017 by taking the datasets from the year 2009-2016. The predicted Result and The Actual result are compared.

Result Analysis for Alcohol

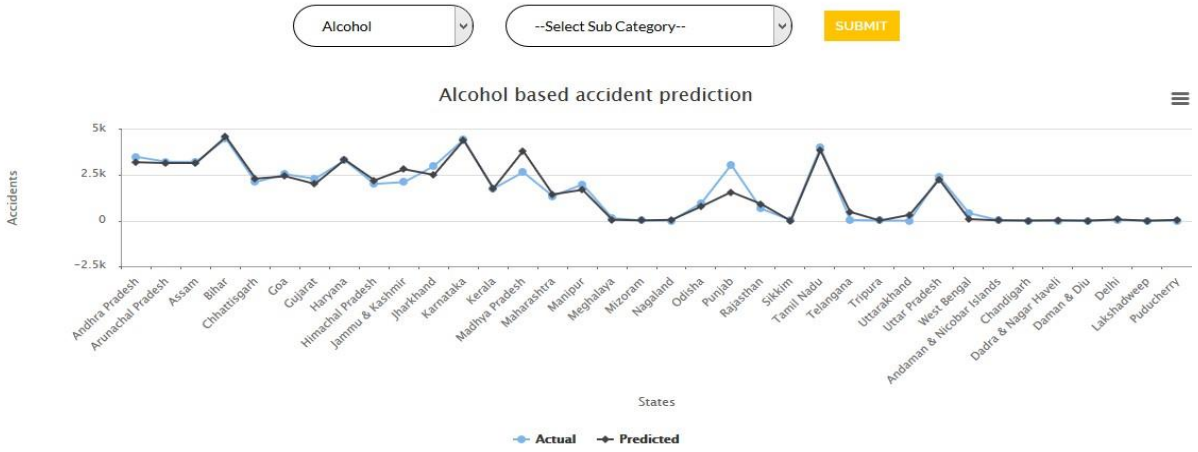


Fig 7: Analysis of Accident due to Alcohol

Result Analysis for Climate

Fog

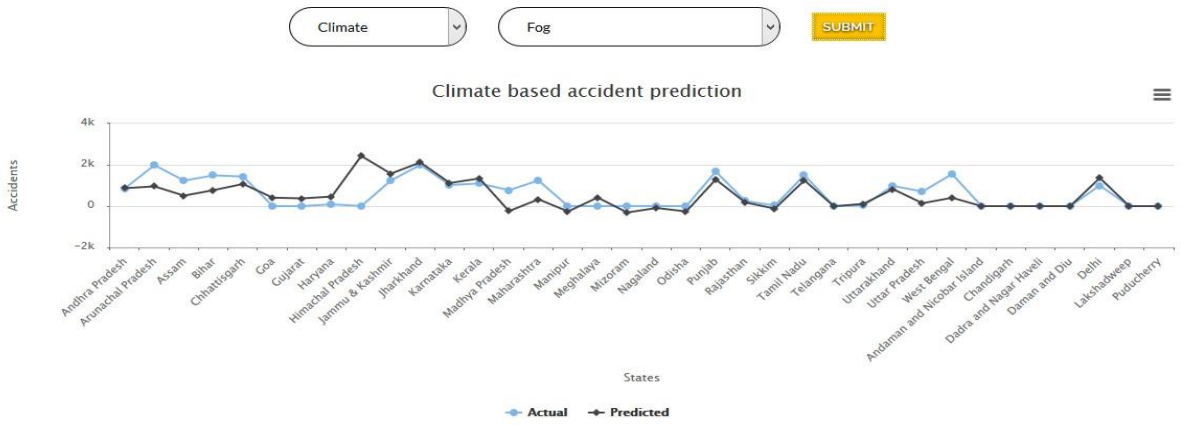


Fig 8: Analysis of Accident due to fog

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

Cloud

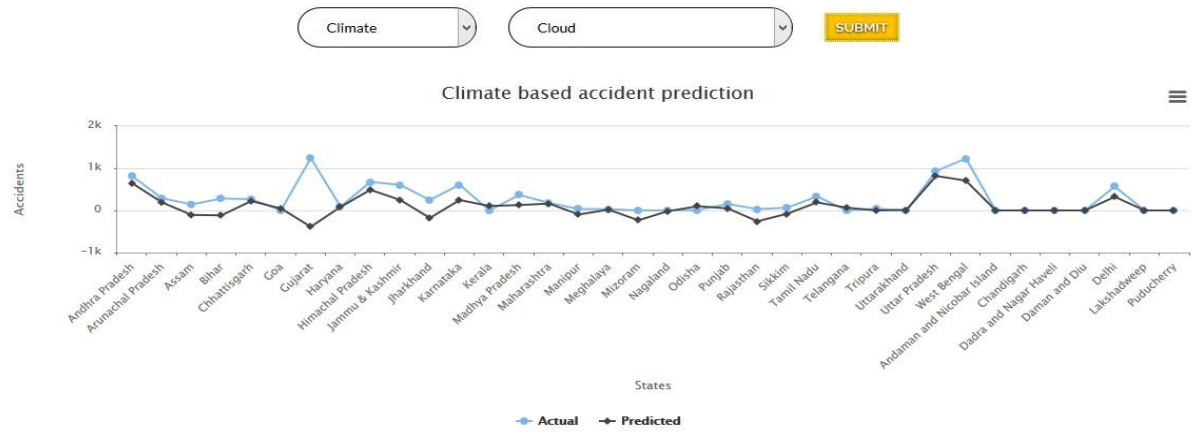


Fig 9: Analysis of Accident due to cloud

Light Rain

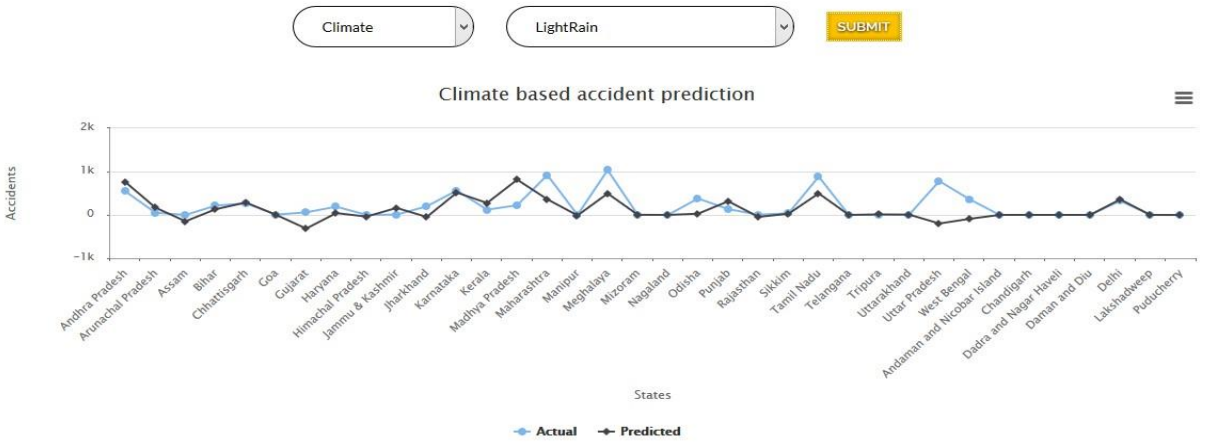


Fig 10: Analysis of Accident due to Light Rain

Heavy Rain

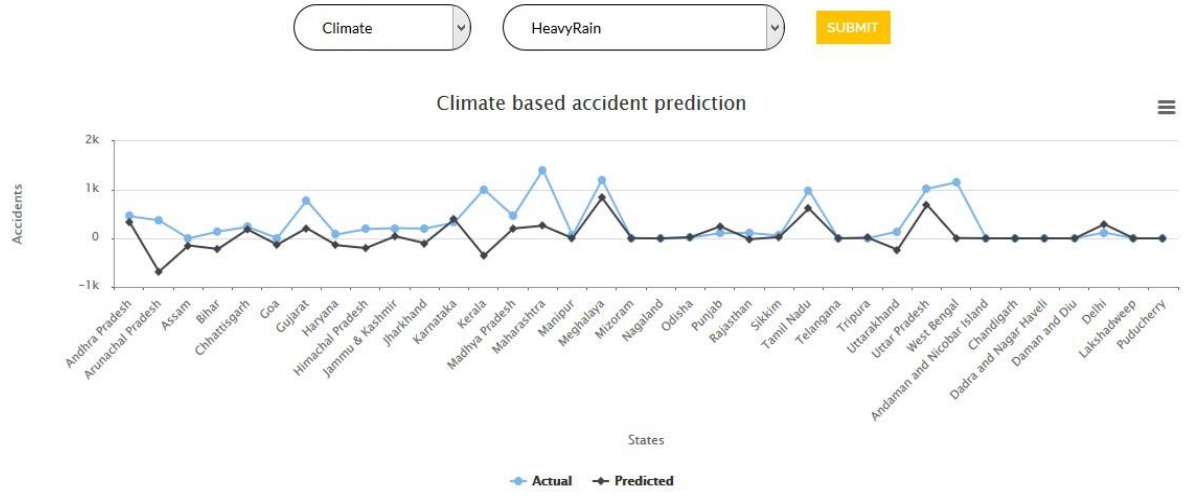


Fig 11: Analysis of Accident due to Heavy RAin

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

Snow

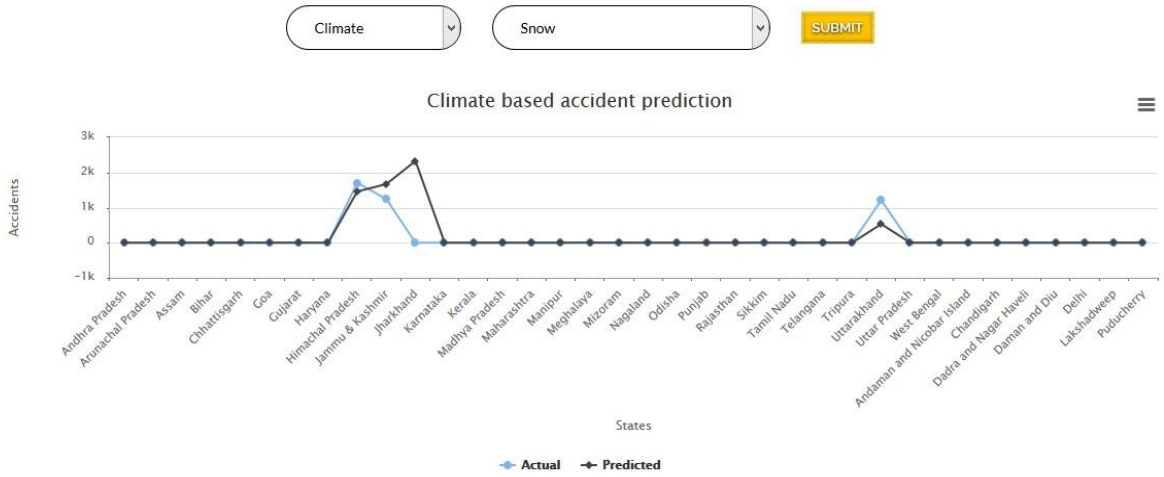


Fig 12: Analysis Accident due to snow

Cold

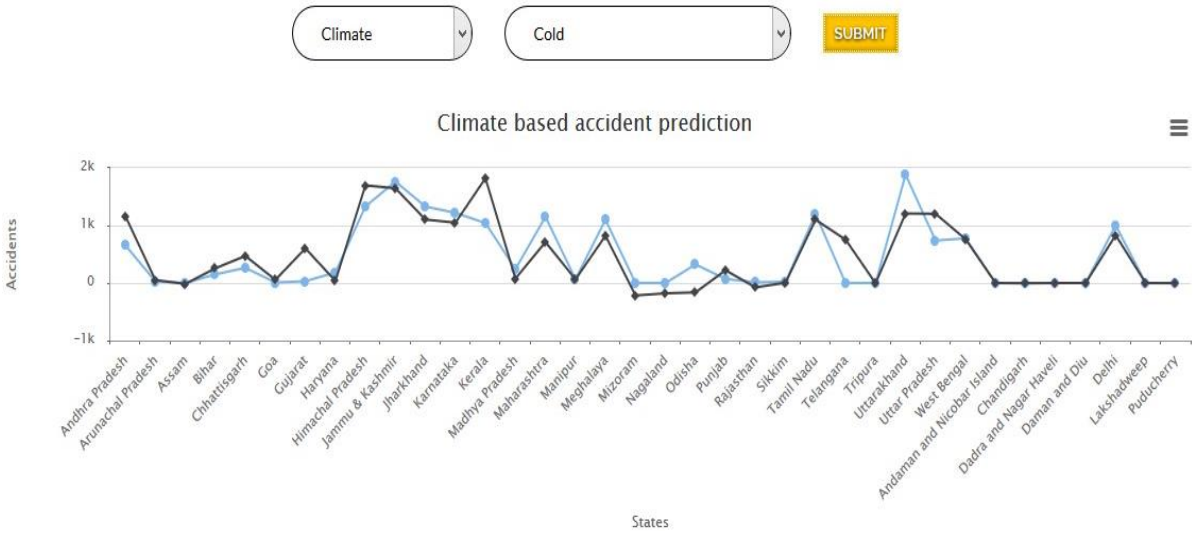


Fig 13: Analysis of Accident due to cold

Result Analysis for location

School

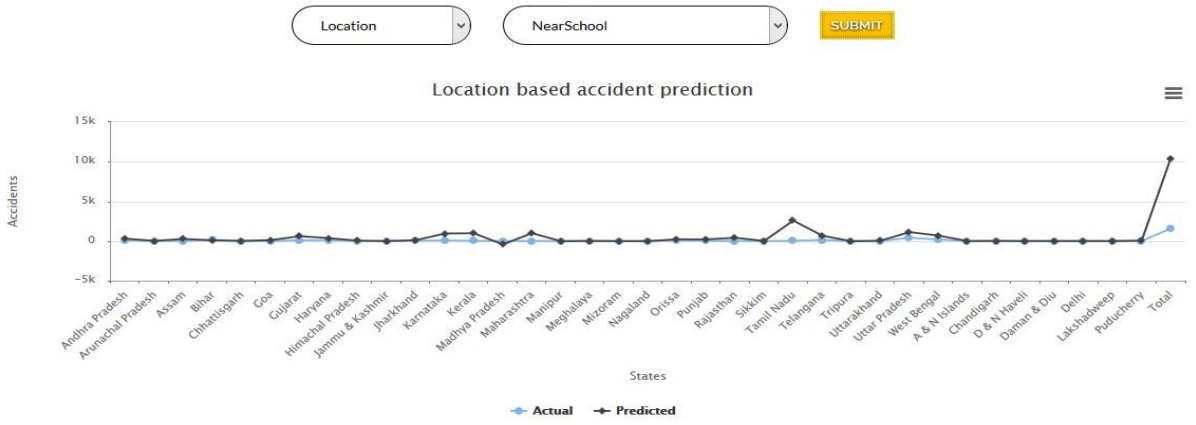


Fig 14: Analysis of Accident nearby school

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

Hospital

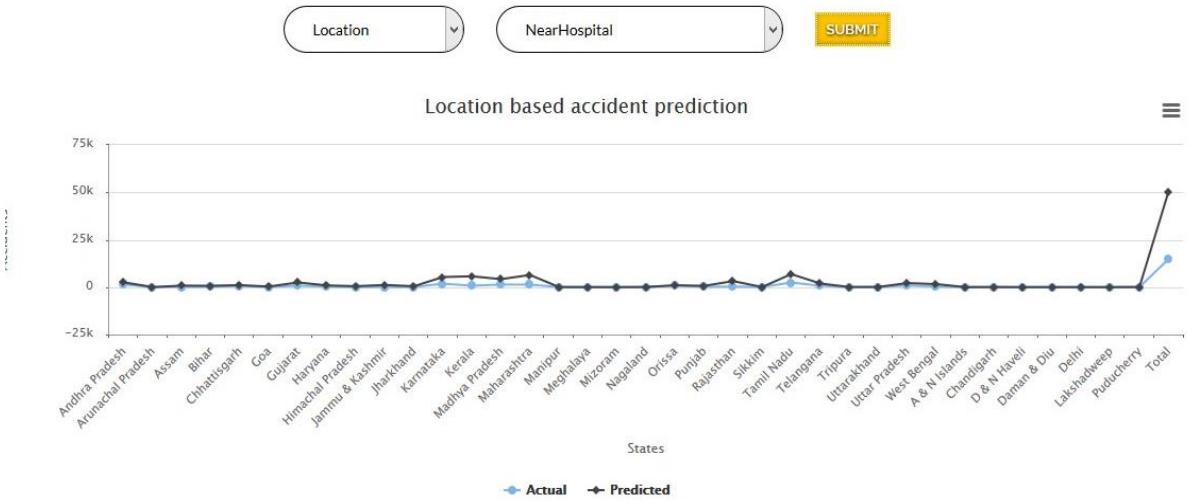


Fig 15: Analysis of Accident nearby Hospital

Factory

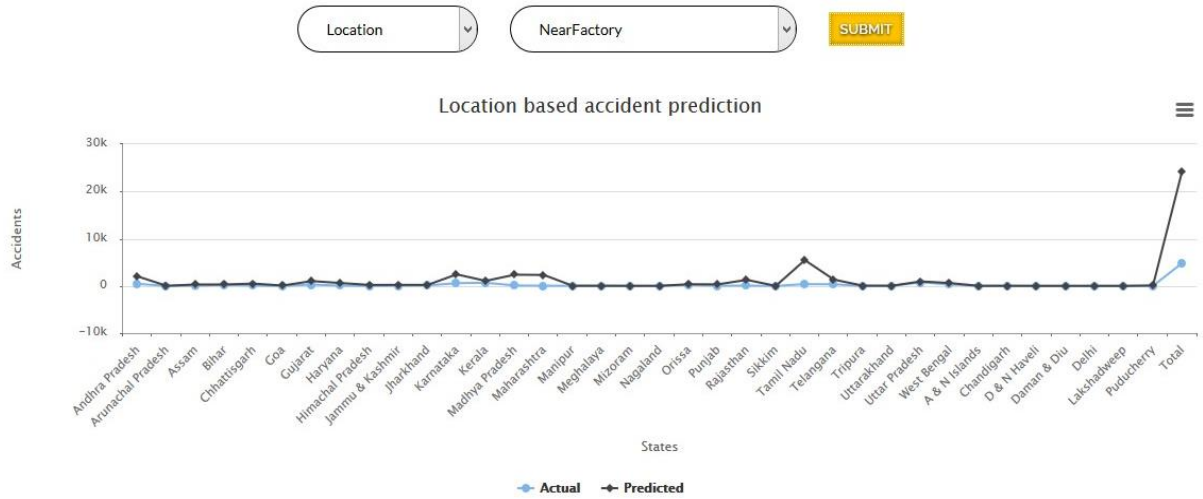


Fig 16: Analysis of Accident nearby Factory

Bazaar

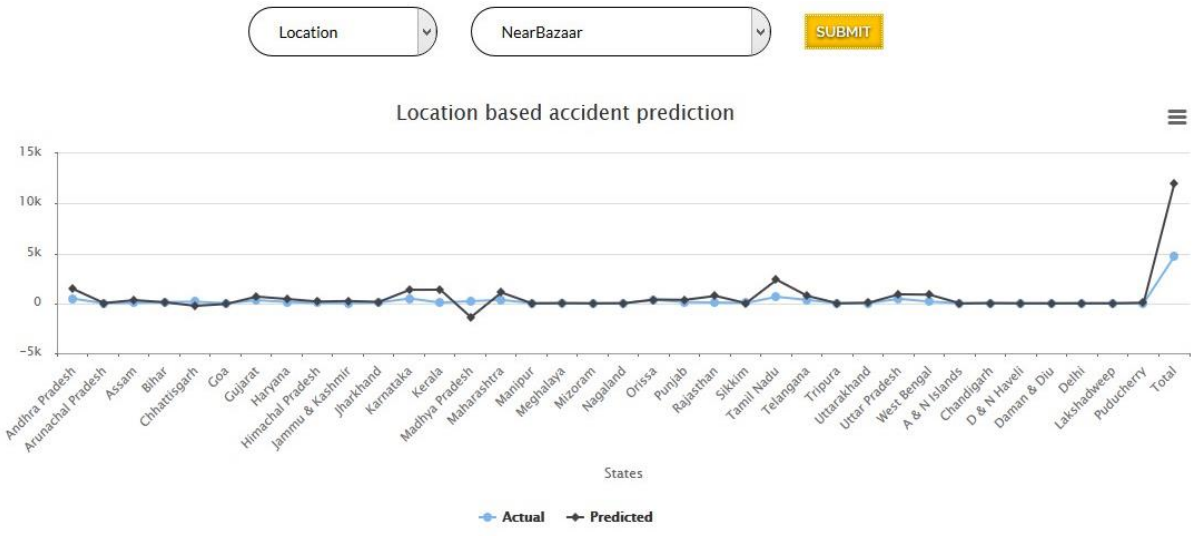


Fig 17: Analysis of Accident nearby Bazaar

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

Result Analysis for junction based

T-junction

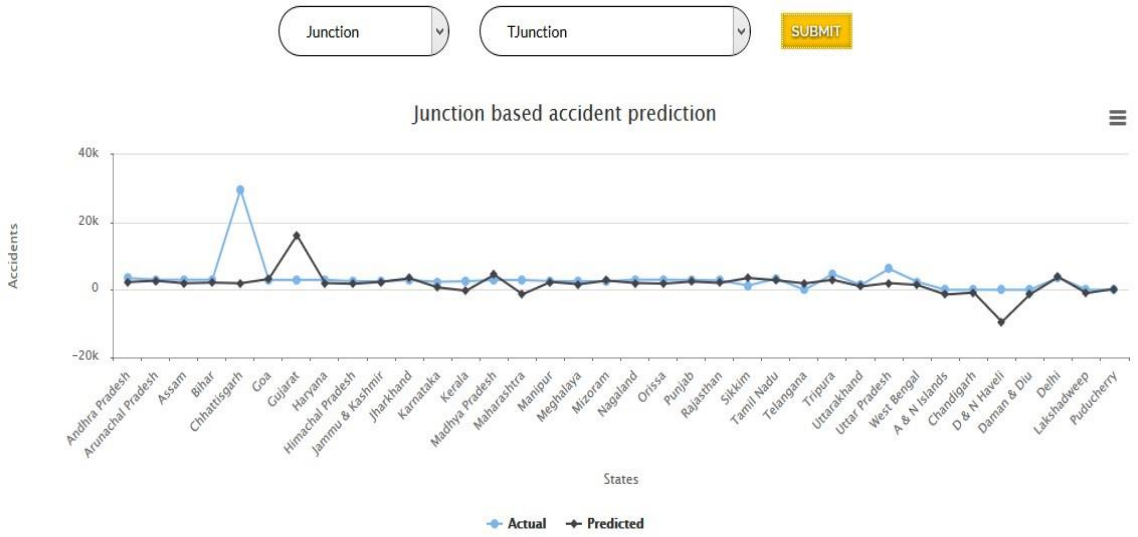


Fig 18: Analysis of Accident due to junction Y-junction

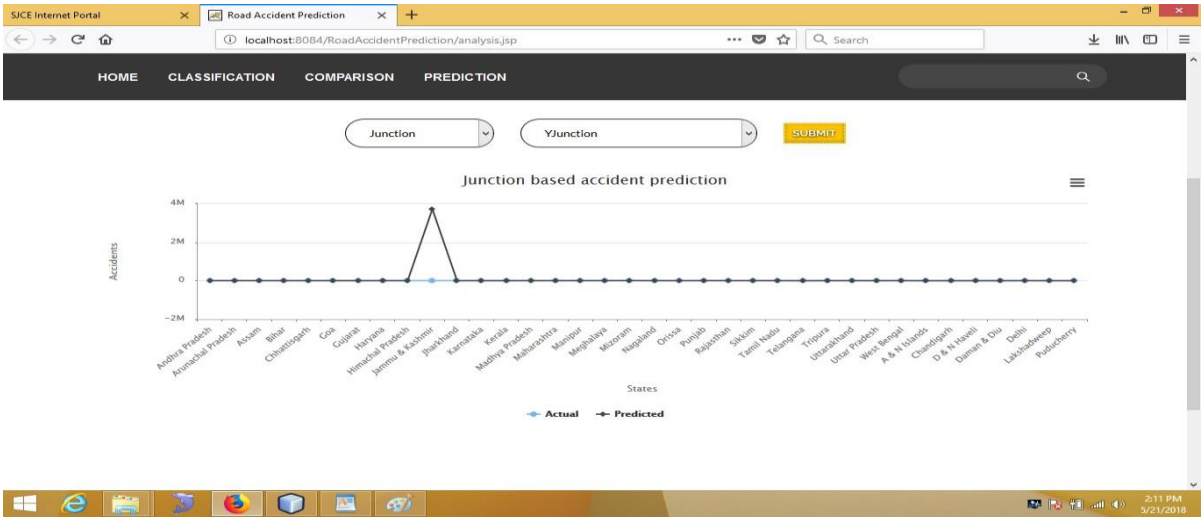


Fig 19: Analysis of Accident due Y junction

Four arm:

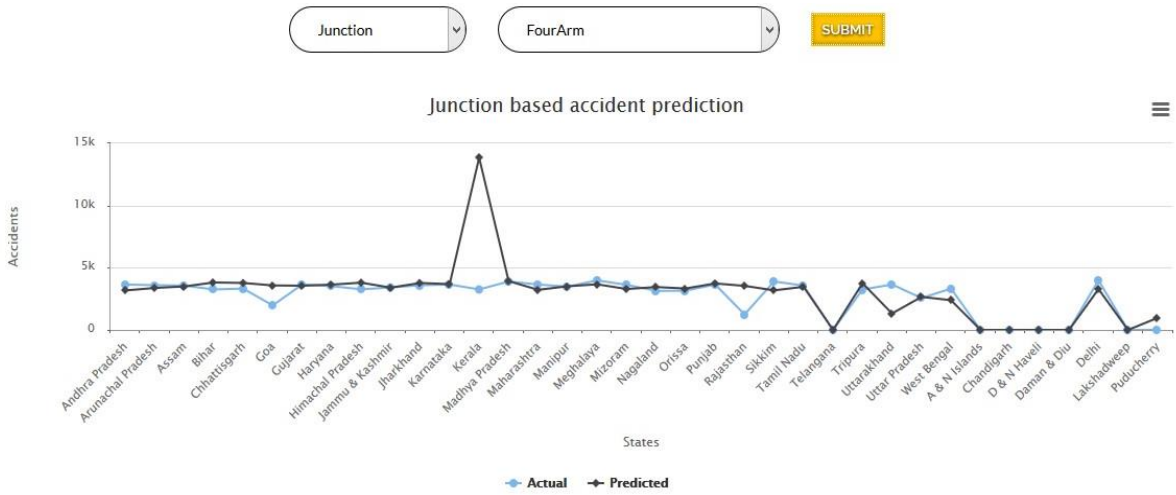


Fig 20: Analysis of Accident due to Four Arm Railcross

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

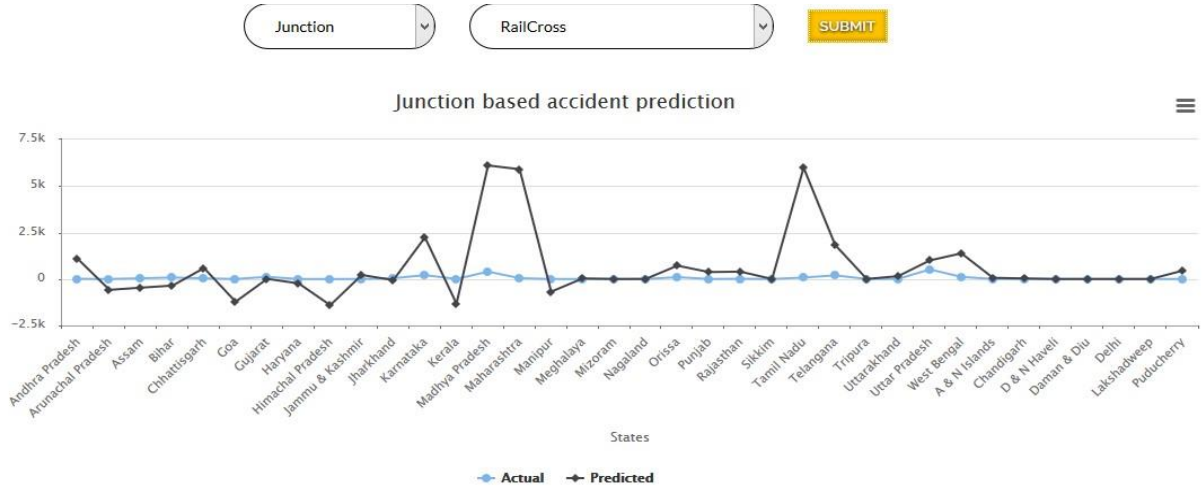


Fig 21: Analysis of Accident due to Rail Cross

V. CONCLUSION

As seen in statistics, linear regression, and the classification, the environmental factors like roadway surface, weather, and light condition do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have stronger affection the fatal rate. From the clustering result we could see that some states/regions have higher fatal rate, while some others lower. We may pay more attention when driving within those risky states/regions. Through the task performed, we realized that data seems never to be enough to makes strong decision. If more data, like non-fatal accident data, weather data, mileage data, and so on, are available, more test could be performed thus more suggestion could be made from the data.

VI. REFERENCES

[1] Yannis George "Investigation of road accident severity per vehicle type" World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016.

[2] Sanjay Kumar Singh " Road Traffic Accidents in India: Issues and Challenges" World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016.

[3] Liling Li, Sharad Shrestha, Gongzhu Hu "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques" Department of Computer Science Central Michigan University, USA 2017 may.

[4] Poul Greibe " Accident prediction models for urban roads "Danish Transport Research Institute, Knuth Winterfeldts Allé, DK-2800 Kgs. Lyngby, Denmark Received 23 May 2001; received in revised form 16 November 2001; accepted 4 December 2001.

[5] Francesca La Torre "Development of a transnational accident prediction model" 6th Transport Research Arena April 18-21, 2016.

[6] Poojitha Shetty "Analysis of road accidents using data mining techniques" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 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 2012 Vol I WCECS 2012, October 24-26, 2015, San Francisco, USA

[8] Ms.P.Shivaranjani1, Dr.K. Karthikeyan "A Review of Weather Forecasting Using Data Mining Techniques" International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 5 Issue 12 Dec. 2016.

[9] Sachin Kumar and Durga Toshniwal. "Analysing road accident data using association rule mining". In Proceedings ofInternational Conference on Computing, Communication and Security, pages 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 Received 01 Sept 2017, Accepted 01 Nov 2017, Available online 10 Nov 2017, Vol.7, No.6 (Nov/Dec 2017).