IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Determining and the Vigilance of Road Accident Hotspots using Machine Learning Techniques

CH.Rajya Lakshmi[1], B.Neelaveni[2], K.P.Satya Kiran[3], CH.Deepak Chowdary[4]

[1]Graduation Student, Bachelor of Technology, Department of Information Technology, Dhanekula Institute of Engineering & Technology, Ganguru, Vijayawada, India

[2]Graduation Student, Bachelor of Technology, Information Technology, Dhanekula Institute of Engineering & Technology, Ganguru, Vijayawada, India

[3]Graduation Student, Bachelor of Technology, Information Technology, Dhanekula Institute of Engineering & Technology, Ganguru, Vijayawada, India

[4]Graduation Student, Bachelor of Technology, Information Technology, Dhanekula Institute of Engineering & Technology, Ganguru, Vijayawada, India

Abstract - The statistical analytic technique investigates the influence of the accident Severity Index (SI) on temporal-spatial patterns of accident hotspots related to specific time intervals.10 years of Andhra Pradesh (AP) Road Traffic Accident (RTA) data were used to analyze and test this approach. Finding the Accidents Hotspots according to the severity is the most crucial task. This study abridged to find the severity of the accident and map the locations. The model which we enact using supervised machine learning algorithms, such as Decision Tree (DT), Logistic Regression (LR), and Random Forests (RF) are implemented on traffic accident data. SMOTE algorithm is used for handling the data imbalance. An RF model can provide insights into the severity of traffic injuries, according to the result of this study. RF algorithm has shown better performance with 70% accuracy than LR with 72.5%, and RF with 63.6% accuracy.

Key Words: **Decision Tree (DT), Logistic Regression (LR), Random Forests (RF), Severity Index (SI).**

I. INTRODUCTION

Traffic accidents are a daily source of death, injury, and property damage on roadways resulting in huge losses at economic and social levels. According to the World Health Organization (WHO), in 2017 around 1.5 million different road users die every year from traffic accidents and half of them die due to traffic crashes. It is also expected that in the absence of sustainable traffic, traffic crashes will become the leading cause of death by 2030. As the demand for vehicles rises, the number of vehicles on the road and traffic jams increase, particularly during rush hours. Therefore, road traffic accidents are among the leading causes of death and injury worldwide. According to the report by the Michigan Traffic Crash Decade-At-A-Glance, there were over 314,921 traffic accidents in the US in 2017, which cost Americans over $230 billion each year. Over 1,028 people lost their lives while over 78,394 people were injured. Classification methods are among the most commonly used techniques in mining traffic accidents, where the goal is to build classifiers that can predict accidents. These classifiers are built using training sets of data in which accident factors are known.

Thus, investigative and predictive methods, such as machine learning algorithms are vital to making smart decisions that will eradicate avoidable accidents on freeways. Can machine learning algorithms assist in saving lives? This inspires the researchers of this study to use machine learning algorithms to predict and analyze freeway crashes based on the roadway, human, and environmental factors. The primary objective of this study is to achieve accuracy and identify the factors behind Traffic Accident Severity that could be helpful to reduce accident frequency and severity in near future, thus saving many lives and wealth, as well as many other things. Additionally, the study aimed to establish models to select a set of influential factors and to build up a model for classifying the severity of injuries that can be used by the Michigan Traffic Agencies (MTA). This model will help MTA and other responsible agencies in Michigan to be more proactive in combating high-risk areas on freeways.

The rapid growth of road traffic has caused a rise in road traffic accidents, particularly severe vehicle crashes. Indeed, in recent years much attention has been paid to determine factors that significantly affect the severity of traffic accidents and several approaches have been used to study this problem. Traffic accidents are caused by a variety of factors, including environmental factors (i.e. weather conditions and road signs), vehicle type & its safety, and the characteristics of traffic users. In addition, some of these factors play a greater role in determining the severity of an accident than others. Therefore, it is apparent that the analysis of the determinant factors of accident severity will help reveal more patterns and knowledge that can be used in the prevention of accidents.

II. METHODOLOGY

The purpose of the methodology that was used in this research study is to build the prediction classification rules of the best-performing model (LR, DT, and RF).

Logistic Regression (LR):

Logistic Regression is a classification model. The idea of the algorithm is to map the results of linear functions to sigmoid functions. The linear regression model is a simple mathematical model and easy to implement.

Random Forest Model (RF):

The Random Forest model is an ensemble learning method that constructs a series of decision trees at training time and outputs the class which is the mode of the classes Prediction of individual trees based on classification (classification) or regression (regression). The minimum number of samples required to split a node was set to two, and the minimum samples per leaf are set to one.

III. PROCEDURE

In the improvement of models first analyze the dataset of the Road traffic accidents data in the model building primarily first upload the data, read the data then train and test the data to predict the results. According to the results generated first analyze them and generate the graph. The figure-1 illustrates the flow of road accident severity model generation.

According to the atmosphere condition, collision, Road catalog, and surface of the earth characteristics considered to analyze the accident severity. As the plot demonstrates atmospheric conditions and the surface are the main causes of to increase in the severity.

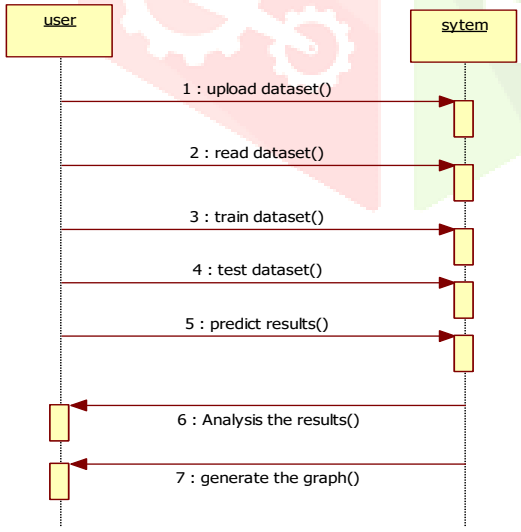


Figure-1: **Sequence Diagram**

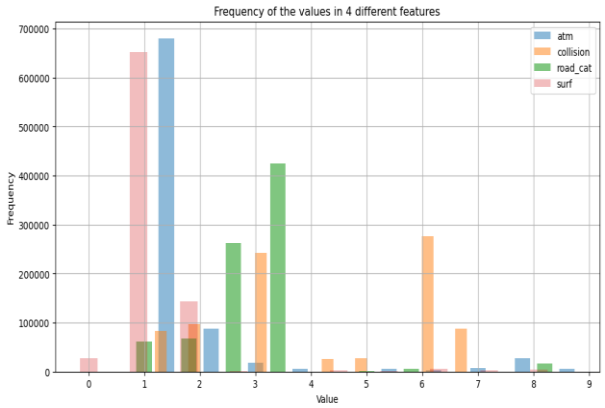
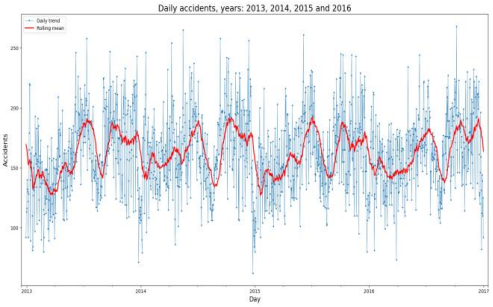
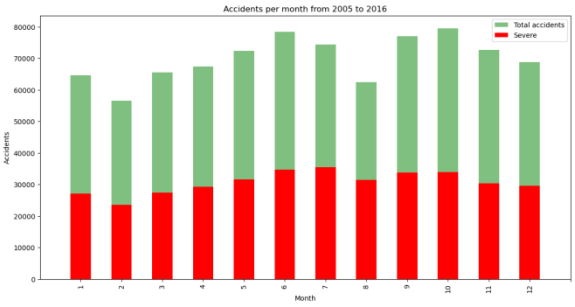


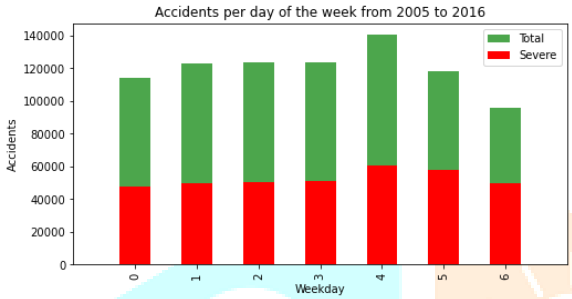
Figure-2: Frequency of the values in 4 different features



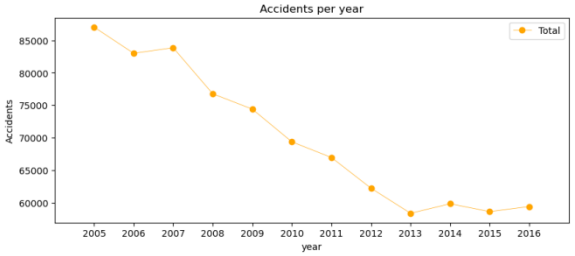
(a) Day wise Road accidents with the Severity mean



(c) Accidents per Month



(b) Accidents Per Day of the Week



(d) Accidents Per Year

Figure-3: Interrelationship between the Accident and time

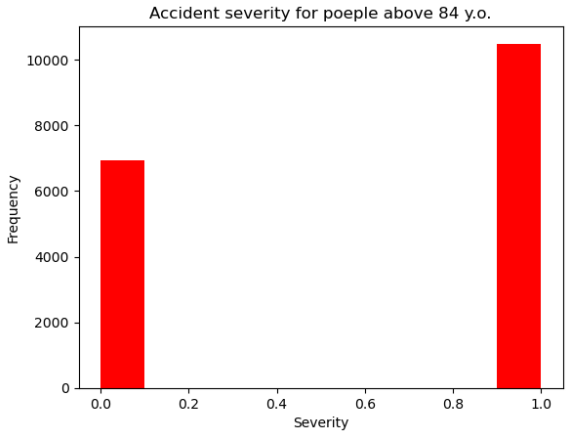
As for the model development and analysis of the Road accident hotspots, the data would be analyzed in different time intervals and age considerations. The time intervals chosen are day-wise, weekly, monthly, and yearly basis. The age considerations did it above 84 and below 84 are taken. In that the above 84 age ha the high fatality accident has occurred. It will be shown in figure-4.

Figure 3 (a) Illustrates the Day wise road accident severity with the mean value. That mean value would be considered as the severity of that day.

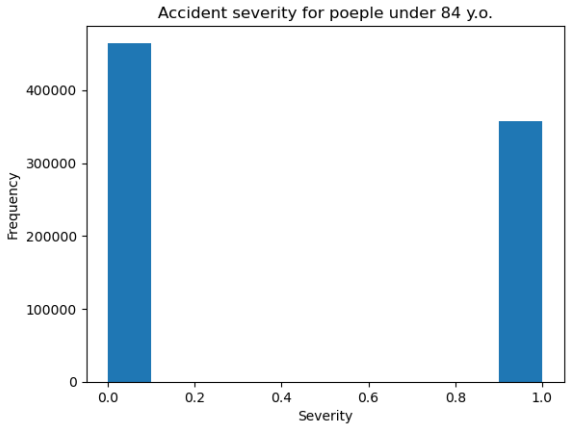
Figure 3 (b) analyzes the accidents weekly day-wise accident severity. In that two concerns are taken. They are total accidents and in that how many have severely occurred?

Figure 3 (c) would be analyzed the data according to the month-wise consideration. For that in the total of 12 months in which the month has the high events, it would be giving the high number of accidents with the severity consideration as high.

In Figure 3 (d) the number of accidents is calculated yearly. When compared to 2005 in 2016 the number of accidents would be a lot decreased. Due to the consideration of road traffic rules and the improvement of the road surfaces.



(a) Accident severity for people above 84 years of age



(b) Accident severity for people under 84 years of age

Figure-4: Accident Severity Under the age Consideration

IV. RESULT AND DISCUSSION

The proposed methods will give the best accuracy of road accident severities. Table 1 will show the percentages of the Accuracy that each model will give. And Figure 5 will be plotted the graph between each classification based on the labeled data and the predicted result accuracy. As the final illustration, the accident location would be indicated in cloud-based maps as shown in Figure 6. As the final scenario, it also shows the present weather conditions for safety purposes. The weather conditions are indicated and explained in Figure 7.

	Algorithm	Accuracy
0	Decision Tree	0.636431
1	Random Forest	0.720502
2	Logistic regression	0.672294

Table-1: Classification Models Accuracy

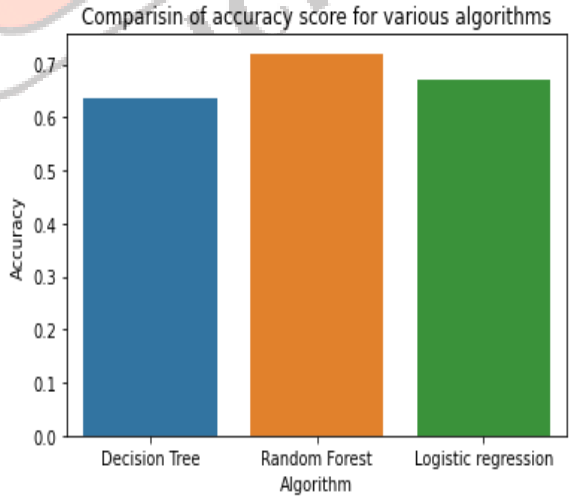


Figure-5: Comparison of accuracy score of various algorithms

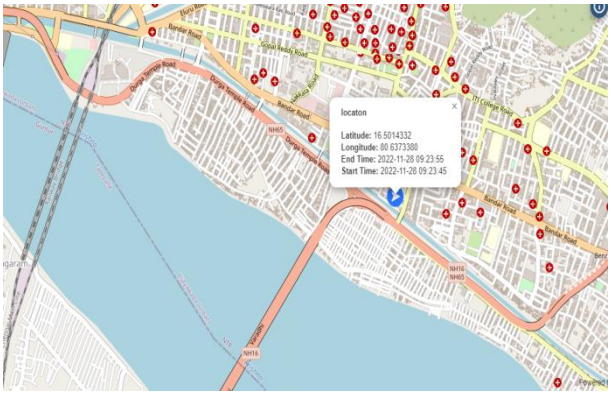


Figure-6: Accident Location indicator on the map

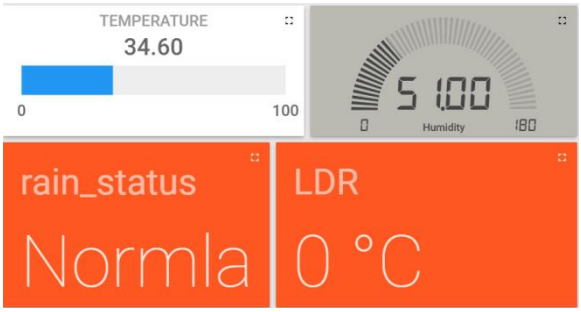


Figure-7: Present weather Condition on that certain accident hotspot

V. CONCLUSION

This study investigated the efficiency of the three machine learning algorithms to build classifiers that are precise and reliable. This includes the Random Forest (RF), Logistic Regression (LR), and Decision Tree algorithms. Based on the confusion (DT), the test results show that the Random Forest seemed to perform better than the other models. This research study shows that the algorithms can predict accidents with 75.50% accuracy. This study can help provide useful information for highway engineers and transportation designers to design safer roads. Further studies should be done to collect related information and investigate the impacts of these factors. It is recommended that Random forest (the best model for predicting freeway crashes) be applied in monitoring Fatal and serious injuries. The recommended predictive model can be used to rapidly and efficiently identify the Key factor causing traffic crashes. One limitation of the current study is that some of the factors (i.e. characteristics of the driver, passenger, and pedestrian, along with traffic conditions) may have possible effects on accident severity and duration, which are not considered because of the lack of suitable data.

VI. ACKNOWLEDGMENT

For allowing the authors access to the road traffic data, the authors are grateful to the national and regional road traffic departments in Andhra Pradesh. Dhanekula Institute of Engineering & Technology, Ganguru, Bachelor of Technology, Faculty of Information Technology Department, and our guide Assistant Professor P. Raghuveer through the provision of computational resources and a conducive working environment.

REFERENCES

[1] WHO | Road traffic injuries, 2017. WHO.
[2] Michigan State Police, Michigan Traffic Crash Decade-At-A-Glance, 2018.
[3] M. Chong, A. Abraham, M. Paprzycki, “Traffic accident data mining using machine learning paradigms”, Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary, 2004, pp. 415- 420.
[4] T. Dejene, A. Ajith, V. Snášel, and P. Krömer, “Knowledge discovery from road traffic accident data in Ethiopia: data quality, ensembling and trend analysis for improving road safety”, Neural Network World, vol. 22, no. 3, 2012, pp. 215–244.
[5] S. Krishnaveni and M. Hemalatha, “A perspective analysis of traffic accident using data mining `techniques,” International Journal of Computer Applications, vol. 23, no. 7, 2011, pp.40-48.
[6] T. Beshah and S. Hill, “Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia,” AAAI Spring Symposium, 2010.
[7] G. Chen, Z. Zhang, R. Qian, R. A. Tarefder, and Z. Tian, “Investigating Driver Injury Severity Patterns in Rollover Crashes Using Support Vector Machine Models,” Accident Analysis and Prevention, vol. 90, 2016, pp. 128–139.
[8] S. Nazneen, M. Rezapour, and K. Ksaibati, “Determining causal factors of severe crashes on the fort peck Indian reservation,” Journal of Advanced Transportation’ Montana, 2018, pp. 1-8.
[9] S. Dissanayake and U. Roy, “Crash Severity Analysis of Single Vehicle Run-off-Road Crashes,” Journal of Transportation Technologies, vol. 4, 2014, pp. 1-10.
[10] D. Shinstine, S. Wulff, and K. Ksaibati, “Factors associated with crash severity on rural roadways in Wyoming,” Journal of Traffic and Transportation Engineering, vol. 3, no. 4, August 2016, pp. 308-323.
[11] M. I. Ratnayake, “Effectiveness of seat belts in reducing Injuries: a different approach based on KABCO injury severity scale,” Midwest Transportation Consortium, Iowa, 2006.

[12] N. Chawla, K. Bowyer, L. Hall, and W. KegelMeyer, "SMOTE: synthetic minority oversampling technique," Journal of Artificial Intelligence Research, vol. 16, 2002, pp. 321-357.
[13] L. Lei and X.-D. Wang, "Improved ad boost ensemble approach based on loss function," Journal of Computer Applications, vol. 32, no. 10, 2013, pp. 2916–2919.
[14] G. James, D. Witten, T. Hastie, and R. Tibshirani, 2013. "An introduction to statistical learning with applications in R," Springer Texts in Statistics Series, New York, 2013.
[15] Dong. C, Shao. C, Li. J, and Xiong. Z, "An improved deep learning model for traffic crash prediction," Journal of Advanced Transportation, 2018, pp.1-13.