iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** I **Month of publication:** January 2021

DOI: https://doi.org/10.22214/ijraset.2021.32778

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: 7.429
Volume 9 Issue I Jan 2021- Available at www.ijraset.com

A Novel Approach for Traffic Accidents Analysis using Hierarchical Clustering Techniques

Kinnari Parikh[1], Dr. Gayatri S Pandi(Jain) [2]
[1]*Department of Computer Engineering, L.J Institute of Engineering and Technology, Ahmedabad, India*
[2]*Head of Department of Post-Graduation, L.J Institute of Engineering and Technology, Ahmedabad, India*

Abstract: Consistently, a few thousand individuals are killed on Indian road and many thousands get harmed. The most well-known, for all time happening reasons for road accident incorporate inability to consent to the standards of the road traffic, low driving aptitudes joined with a helpless evaluation of the circumstance out and about, absence of sufficient road framework just as helpless specialized states of vehicles. It despite of different measures that are being taken to improve security on roads, the number of the dead and injured because of road accident is still extremely high and the misfortunes acquired by the society are high also. In rate terms, in 2018, the quantity of accidents has expanded by 0.46 percent, people died has expanded by 2.4% and injuries have decreased by 0.33 percent over that of the earlier year i.e 2017. Data mining is the extraction of certain, previously unknown and possibly valuable data from information. We use hierarchical agglomerative cluster analysis characterize similarities in a bunch of crash information factors data variables, these clusters can then be used as the basis in test scenario development. Hierarchical clustering analysis framework is developed to identify benchmark and basic regions for successful road safety strategies.

Keywords: Road accident, traffic collision, road safety strategies, clustering analysis, crash data, historical driver risk

I. INTRODUCTION

While planning the road organizations, information identified with dangerous and safe stretches are gathered. This helps in arranging road improvement plans. Accident records can be deceiving as the recurrence of accident differs considerably. The issue of getting dependable assessments of the long-term road accident frequencies at singular road area is a difficult issue. The information mining of recently gathered information of road organizations will help in recognizing high danger destinations move of fluctuating recurrence of accident. The information mining activity could be completed utilizing the circumstances and logical results measure information identified with accident frequencies, environmental and different factors.

Traffic accident have gotten one of the main sources of passing and injuries around the world, with more than 1.25 million of deaths every year, somewhere in the range of 20 and 50 million harmed and a worldwide extent of 18 deaths for each 100,000 occupants (World Safety Organization, 2015). The majority of these deaths happen in lacking nations (16%) and in developing nations (74%), which changes over mortality from car crashes, not just into a general medical condition, yet in addition into a socioeconomic advancement issue. Hierarchical cluster analysis over an large arrangement of data to give efficient clusters. It is a well-known clustering method that seeks to group data by creating a hierarchy of clusters. Hierarchical clustering method is applied in driver classification to explore the contributing factors related to driver risk. It is an algorithm that performs the clustering of distance or dissimilarity of objects with a tree structure. It was used to classify drivers according to their driving risk level. In this publication, we examine the contributing elements related with singular driving danger considering authentic driving elements utilizing a dataset assembled utilizing naturalistic driving analyses and a self-report technique. Our emphasis is on characterizing driver conduct activities, for example, time progress, slowing down weight, and quickening to produce close crash occasions. At that point, close crashes are utilized as safety related occasions for driver arrangement.

II. RESEARCH MOTIVATION

Road accident are one of the greatest cause for death on Indian roads. According to media measurements, in India one individual dies in a road accident at regular intervals. The foundations for India's especially high number of on road setbacks incorporate - damage road, careless driver behaviour, defective road plan and designing, helpless authorization of traffic rules. India holds 85% of all road fatalities happening in the developing nations. It is needed to recognize the causal impacts of such significant level of accidents in each state or area. The Motor Vehicles Act of 1988 (MVA), the main demonstration that governs the road safety in India, has demonstrated ineffectual in tending to any of the previously mentioned issues unequivocally. The road safety is as yet not seen as a need issue in the nation.

III. CAUSES OF ROAD ACCIDENTS

Road accident are multi-causal and are the consequence of a transaction of different elements. These can comprehensively be classified into those identifying with (1) human error, (2) road environment and (3) vehicular condition.

A. *Human Error*

Accidents brought about by human mistake incorporate (1.1) cases of accidents caused by traffic rule violations, (1.2) driving without valid driver license and (1.3) non-use of safety devices.

1) *Traffic Rules Violations:* The over speeding is the primary violation related with accidents, accident related deaths and injuries in 2018 with over speeding representing 66.5 percent of the road accidents, 64.4 percent of all out deaths and 67.4 percent of absolute injuries. Driving on wrong side / lane indiscipline is the second most significant reason representing 5.3 percent of the road accidents, 5.8 percent of absolute deaths and 5.1 percent of all out injuries.
2) *Driving Without Valid Driver License:* Vehicles driven by undeveloped and untrained drivers are a genuine traffic risk and can cause accidents death and injuries. Though the fact that the issue is essentially an authorization issue, it should likewise be addressed to with better facilities and opportunities for preparing/skilling and assessment/testing
3) *Non-use of Safety Devices – Helmets and Seatbelts:* **N**on usage of safety gadgets, for example, helmets and safety belts don't cause accidents however are basic for deflecting fatal and unfortunate injuries in an event road accident happens.

B. *Road Environment*

The causes identified with the classification of road environment include (2.1) accidents happening in a particular geographical area (residential, commercial institutional etc), (2.2) those related to the type of road features including straight, curved, steep etc, (2.3) type of junction & type of traffic control, (2.4) weather condition etc.

1) *Road Accidents Classified By Type Of Neighbourhood:* Local location, institutional territory and market/business zone will in general have higher measures of traffic congestion and are more inclined to road accidents. Nonetheless, the information got demonstrated bigger portion of accidents and people killed in open areas in both 2017 and 2018, maybe as open regions have lower implementation presence and might be inclined to driving and traffic rule violations.
2) *Road Accidents Classified By Road Features:* Vehicle speed will in general be high on straight roads in open areas which proves the high level of road accident, people killed and injured (differing in the scope of 62-66%) on these roads in both 2017 and 2018.
3) *Accidents By Road Junction Type & Traffic Control:* Road junctions are traffic combining focuses and consequently are inclined to accidents. Any way the information outfitted by States/UTs show that just about 33% of the accidents in 2018 occurred at different types of junctions, with practically 67% of the accidents occurring in the "others" category.
4) *Accidents By Weather Condition:* Accidents under adverse weather climate conditions, for example, rainy, foggy and hail/sleet represented just 16.3 percent of road accidents during 2018. Further, both road accidents and number of accidents related deaths under various weather conditions have increased. in the year 2018 as compared to 2017.

C. *Vehicular Condition*

The third classification of vehicular condition includes cases of accidents related with overloading and the age of the vehicles.

1) *Accidents in Over-aged Vehicles:* It will be noted from the over that vehicles 10-15 years of age, represented 13.1% of accidents, vehicles more than 15 years represented 9.6 percent of total accidents and vehicles with age not known represented another 13.9% of the total accidents in 2018. Together these categories represented 36.6% of accidents.
2) *Overloading:* It will be seen that overloaded vehicles represented a portion of 10.1% of all out accidents,11.9% of total killed and 10% of the injured in 2018.

IV. LITERATURE SURVEY

A. *Definition of run-off-road crash clusters—For safety benefit estimation and driver assistance development, ScienceDirect (2018)*

Data determination gives Methodological steps from questioning the GIDAS information base to crash investigation. Factors for cluster analysis Provides variable name from the GIDAS information base. Crash examination having Frequency grouped into bigger classifications examination of the two factors proposed that few classifications could be assembled into bigger classes. In the Cluster investigation, the clustering process consists of three stages

1) A proportion of (dis-)closeness (or separation) between two components or clusters was picked.
2) An agglomeration criterion figures out which clusters to converge at each phase of the clustering.
3) A measure to decide the number of clusters should be utilized.

B. *Investigating the Significant Individual Historical Factors of Driving Risk Using Hierarchical Clustering Analysis and Quasi-Poisson Regression Model, MDPI (2020)*

In this paper Two fundamental points are thought of: grouping drivers dependent on their driving risk and researching risk factors identified with individual driving. hierarchical clustering investigation was adopted for gathering and arranging high-risk drivers. In the regression model, the objective of the examination was to assess the effect of individual regression risk factors on driving risk by looking at the critical elements of close crashes by thinking about the impact of the result variable value (close accident recurrence of drivers).

C. *Continuous Risk Profile and Clustering-based Method for Investigating the Effect of the Automated Enforcement System on Urban Traffic Collisions, Springer (2019)*

Continuous risk profile (CRP) is a procedure that uses the collision density of every road unit. It calculates each weighted average number of crashes in each sliding moving window as opposed to partitioning the road into areas. K-means clustering includes the assignment of a starting value for each cluster (k), framing clusters dependent on the task of each article (information) to the nearest starting value, forming clusters dependent on the task of each item (information) to the nearest starting value, and the restoration of each starting value by figuring the average of each cluster. Hierarchical clustering performs the consecutive clustering of distance or uniqueness of objects with a tree model. Cross connection coefficient measure of similarity of two series as a function of the displacement(τ) of one comparative with the other.

D. *Augmenting Classifiers Performance through Clustering: A Comparative Study on Road Accident Data, igiglobal, ACM (2018).*

Cluster Analysis utilized three cluster analysis strategies k-modes clustering, LCC clustering and clustering utilizing BIRCH. Classification Techniques Classify the accident data dependent on seriousness of accidents, it utilized Naïve Bayes (NB), random forest (RF) and super vector machine (SVM) classification techniques. In the cluster selection criteria, the best cluster model is the one expanding the Gap value. In the data set description, the traffic accident data set was acquired from police records of Muzzafarnagar area, Uttarpradesh. A sum of 2300 accident records were chosen for investigation.

E. *Hierarchical Clustering Analysis Framework of Mutually Exclusive Crash Causation Parameters for Regional road safety Strategies, Taylor and Francis (2017).*

Utilizing hierarchical clustering investigation, group states/districts independently for each crash causation limit. Estimate group average worth. These correlations are valuable in choosing the real benchmark regions or states and introduced as three unmistakable standards given in the following:

1) *Criteria 1:* For each crash causation boundary, regions or states having a place with groups that meet the predefined threshold value chose dependent on the achievable national average target.
2) *Criteria 2:* The average seriousness index growth rate of regions or states meeting the predefined threshold value chose dependent on the feasible national average seriousness index growth rate target.
3) *Criteria 3:* The average seriousness record development pace of locales or states meeting the predefined threshold value chose dependent on the national average severity index growth rate target.

F. *Accident Data of Ahmedabad City*

The year 2019 claimed 423 lives in road accidents over the city, the most elevated casualty figure since 2013. As per the city traffic police, 405 fatal accidents were accounted for in 2019 in which 341 men and 82 women lost their lives. The data uncovered that the number of women killed in accidents has everything except multiplied more than seven years; 42 women had died in 2013. Fatalities rose in 2019 however the number of accidents decreased. Traffic police had enrolled 1,371 accidents a year ago; 1,610 of every 2018; and 1,888 out of 2013.

Proposed Model

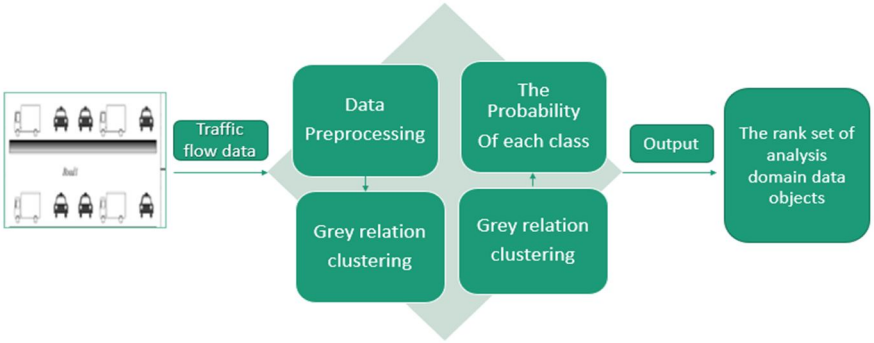


Figure 1: Proposed Model

Here, the proposed model contains various steps as follow,

1) *Traffic Accident Data:* Consistently the lives of roughly 1.35 million individuals are stopped because of a road traffic accident. Somewhere in the range of 20 and 50 million additional individuals people suffer non-fatal injuries, with many acquiring a disability because of their injury. To comprehend the real traffic conduct, it requires evaluation of a portion of the basic traffic flow qualities, for example, Speed, Flow, Density and Occupancy through which the limit can be inferred.
2) *Data Preprocessing:* Data pre-processing is a data mining procedure that includes changing raw data into a justifiable arrangement. True data is often incomplete, conflicting, or lacking in specific behaviors or trends, and is probably going to contain many errors. Data pre-processing is a demonstrated technique for resolving such issues.
3) *The Probability of Each Class:* In this step, pre-processed data classes are separated by them probability of occurrence. This classes are use the parameter like human error, road environment and vehicular condition.
4) *Gray relation Clustering:* The grey relational analysis use for analysing the closeness between given patterns. The methodology is utilized to take care of a few data clustering issues as specific illustrations. In every model, the presentation of the proposed algorithm is contrasted and other notable calculations, for example, the fuzzy c-means technique and the hard c- means method. Reproduction results exhibit the effectiveness and feasibility of the proposed method.
5) *The Rank Set of Analysis Domain Data Objects:* Ranking data set is valuable when statements on the request for perceptions are a higher priority than the magnitude of their differences and little is thought about the hidden distribution of the data.

Proposed system solution

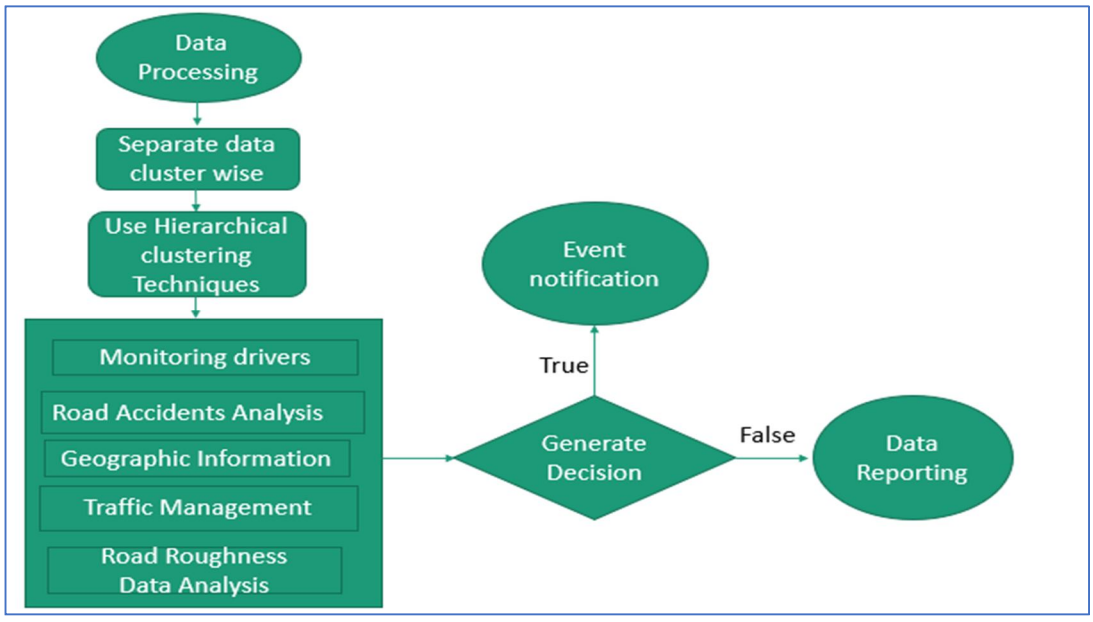


Figure 2: Proposed system solution

Methodology using in proposed solution system is given as follow,

a) *Data Processing:* Data handling is the collection and control of things of information to deliver significant data. In this sense it very well may be viewed as a subset of data preparing, "the difference in data in any way perceivable by an observer."

b) *Cluster Analysis:* Factors for the cluster analysis were picked identified with the kinematics of the accident (beginning velocity, departure point, and loss of control), the plan of the roadway (distance to road edge, type of the road, and sidelong acceleration at posted speed), or to the natural conditions (lighting conditions and road surface conditions).

c) *Hierarchical Clustering:* Hierarchical cluster investigation, is a calculation that similar group items into groups called clusters. This investigation gave data about pre-crash factors present in each cluster, and will function as a guide for future virtual testing.

d) *Event Detection:* Proposed a model can distinguishing a accident investigation utilizing various rules of accident accurse i.e. Road Roughness Data Analysis, Road Accidents Analysis, Geographic Information, Monitoring drivers, Traffic Management

V. CONCLUSION

Review of exploration papers give a knowledge of methods and calculations utilized in the analysis of road accident. Logical factors were discovered valuable to comprehend the idea of each cluster and give important insights. This examination explored the contributing individual historical components of driving risk utilizing hierarchical clustering analysis. hierarchical clustering analysis was utilized to aggregate drivers into groups dependent on their close accident recurrence. hierarchical agglomerative cluster analysis to characterize similitudes in a group of crash information variables, these clusters would then be able to be utilized as the premise in test situation development. It's evaluated traffic safety by considering drivers' conduct and execution under psychological workload in complex environment areas. The investigation induced that speed maintenance and path deviations alongside driving experience and gender orientation gatherings may assume a significant part in analysing and improving traffic security in road environments. By utilizing referenced methods in the proposed system will help in better analysis of road accident.

REFERENCES

[1] Sachin Kumar, Prayag Tiwari, Kalitin Vladimirovich Denis “Augmenting Classifiers Performance through Clustering:A Comparative Study on Road Accident Data” IGIGIglobal, International Journal of Information Retrieval Research(2018), Cited by [7].

[2] Daniel Nilsson, Magdalena Lindman, Trent Victor, Marco Dozza “Definition of run-off-road crash clusters—For safety benefit estimation anddriver assistance development” Accident Analysis and Prevention, ScienceDirect(2018), Cited by [9].

[3] Hasan A.H. Naji, Qingji Xue, Ke Zheng and Nengchao Lyu “Investigating the Significant Individual Historical Factors of Driving Risk Using Hierarchical Clustering Analysis and Quasi-Poisson Regression Model” MDPI(2020), Cited by [36].

[4] Shin Hyung Park, Shin Hyoung Park, Oh Hoon Kwon, Yunsick Sung “Continuous risk profile and clustering-based method for investigating the effect of the automated enforcement system on urban traffic collisions” The Journal of Supercomputing, springer(2019), Cited by [2].

[5] Avijit Maji, Nagendra R. Velaga & Yohan Urie “Hierarchical clustering analysis framework of mutually exclusive crash causation parameters for regional road safety strategies” International Journal of Injury Control and Safety Promotion, taylor and francis(2017), Cited by [17].

[6] Sachin Kumar1* and Durga Toshniwal2 “Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC)” Journal of big data, springer(2016), Cited by [42].

[7] Yulong Wang , Kun Qin, Yixiang Chen and Pengxiang Zhao” Detecting Anomalous Trajectories and Behaviour Patterns Using Hierarchical Clustering from Taxi GPS Data” MDPI(2018), Cited by [36].

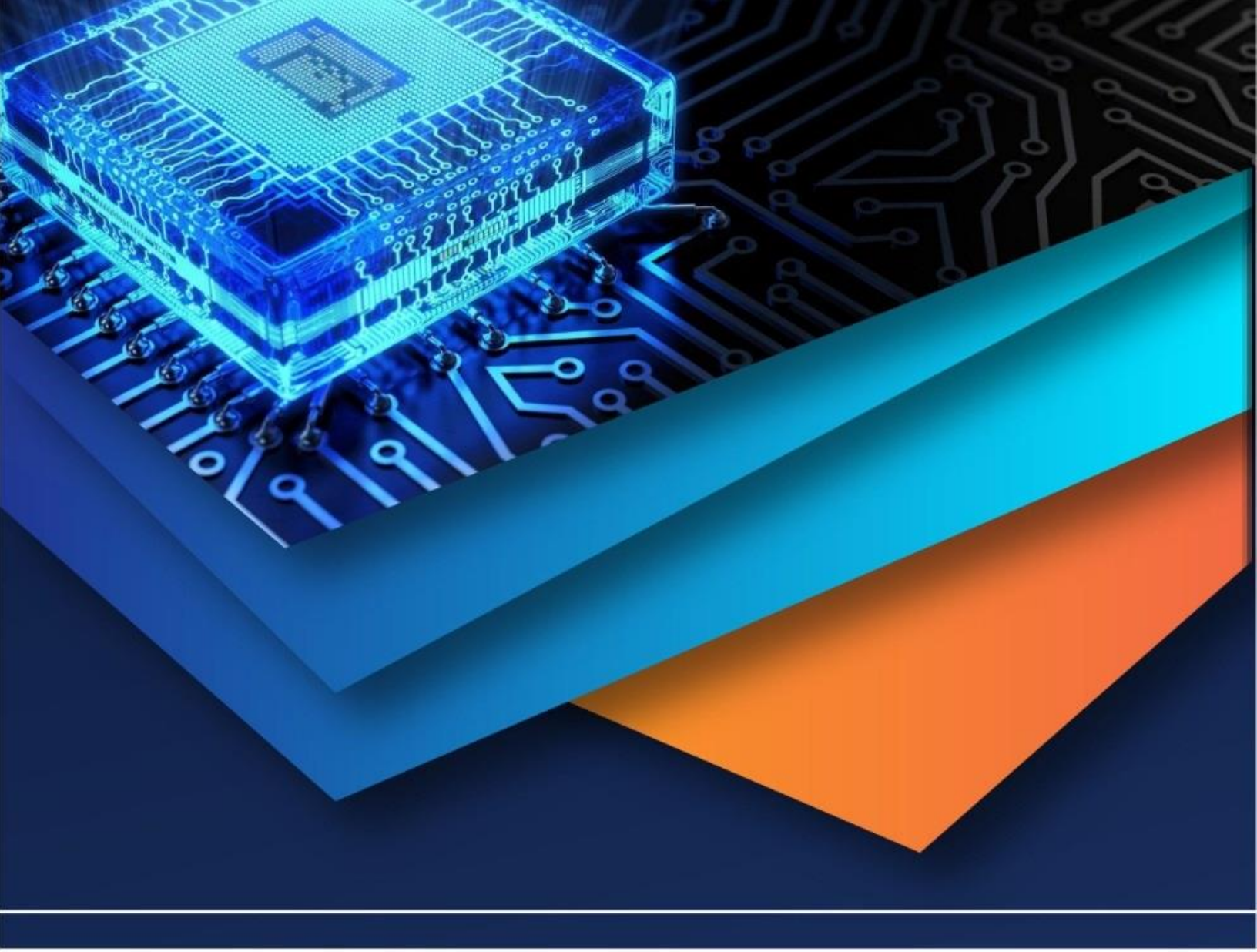
[8] Sudhir Kumar Barai “Data mining applications in transportation engineering” taylor and francis(2011), Cited by [60].

[9] Madhar Taamneh, Salah Taamneh & Sharaf Alkheder ”Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks” taylor and francis(2016), Cited by [15].

[10] Feifeng Jiang , Kwok Kit Richard Yuen , Eric Wai Ming Lee and Jun Ma “Analysis of Run-Off-Road Accidents by Association Rule Mining and Geographic Information System Techniques on Imbalanced Datasets”MDPI(2020), Cited by [7].

[11] Causes of road accidents “Road accidents in india – 2018”, Available on https://morth.nic.in/sites/default/files/Road_Accidednt.pdf.

[12] Accident Data of Ahmedabad city Available on https://timesofindia.indiatimes.com/city/ahmedabad/ahmedabad-at-423-2019-accident-toll-was-highest-in-seven-years/articleshow/73069071.cms.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY