iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59991

Road Accident Severity Prediction Using Random Forest Algorithm

Asst. Prof. Tejashwini A G[1], Chaitanya Kumar Singh[2], Chitransh Sinha[3], Dheeraj Karnani[4], Ashutosh Pandey[5]
Computer Science and Engineering Dayananda Sagar College of Engineering Bangalore, India

Abstract: The surge in accident rates, intensifying human casualties, has propelled the integration of cameras and fixed speed cameras into daily activities. This study focuses on predicting road accident severity, a crucial advancement in road accident management, especially for urban emergency logistics. Employing Machine learning methodologies such as Random Forest which is commonly used for predictive analysis, Naive Bayes, and logistic regression, we rigorously evaluate their efficacy in densely populated areas. The research implements and compares these algorithms, utilizing a confusion matrix to illustrate interclass impacts on pedestrians, vehicle or pillion passengers, and drivers or riders. Notably, the severity prediction for road accidents achieves an impressive 86.8% accuracy with Random Forest, surpassing SVM's 82%. This exemplifies the effectiveness of machine learning in enhancing accuracy and reliability, providing valuable insights for proactive road safety measures.
Keywords: Machine learning, Road accident, severity prediction, Random Forest Model, Google Maps API

I. INTRODUCTION

The prevalence of road accidents, ranging from minor incidents to life-threatening collisions, has surged, prompting a global imperative to enhance traffic safety. The escalation inaccidents has led to a rise in casualties, compelling the adoption of measures such as the widespread deployment of cameras and fixed speed cameras. Addressing this challengerequires proactive strategies, and the utilization of machine learning, stands out as a promising approach. Machine learning employs algorithms to analyze extensive datasets, including information from online sources, to discern patternsand relationships crucial for driving safety. Multiple factors contribute to the occurrence of accidents, such as weather conditions (fog, rain), road conditions, and driver age. By comprehensively studying these variables, computers canprovide valuable insights and advice to drivers, promoting safer practices on the road. Traditionally, managing road accidents relied heavily on post-hoc analysis and surveys conducted at problematic intersections. This reactive approach hindered creating enduring solutions, resulting in temporary fixes that proved ineffective in ensuring sustained road safety. Recognizing these limitations, a shift towards predictive models implementing machine learning algorithms has gained traction. such as Random Forest, Logistic Regression, Decision Tree and Naive Bayes, achieving an impressive accuracy of 86.8% in forecasting the severity of traffic accidents based on coordinates using Google Maps API. Additionally, it offers recommendations based on the received predictions and conditions. For instance, if the severity level is rated as 1, indicating a high likelihood of an accident, it will propose alternative actions to mitigate the severity and reduce the risk. This model transcends theoretical applications and can be employed in real-world scenarios to predict accident risks. Furthermore, transportation companies can utilize the Machine Learning (ML) model to assess the suitability of drivers based on provided coordinates, enhancing overall road safety and reducing the likelihood of accidents. In essence, the integration of ML into road safety strategies represents a proactive and effective paradigm shift, ensuring the creation of safer road environments.

II. MATERIALS AND METHODS

Over the last century, ensuring transportation security has emerged as a vital element for global economic and social advancement. Recognizing its profound impact on development, nations worldwide have prioritized addressing this issue to safeguard the well-being of their citizens and foster ongoing progress. Presently, countries employ variousmeasures, including the installation of speed cameras and enforcing speed limits at intersections and accident-prone areas, to enhance road safety. Fig. 1 illustrates the landing page of the website that was developed.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

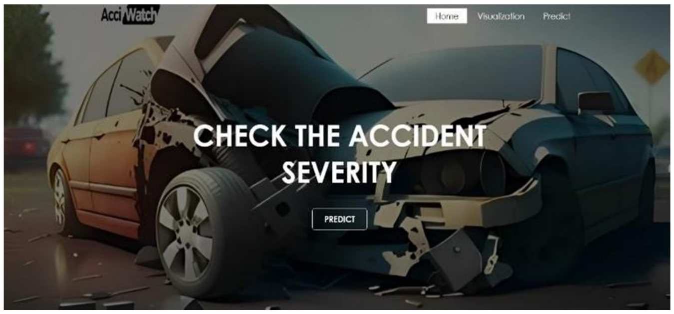


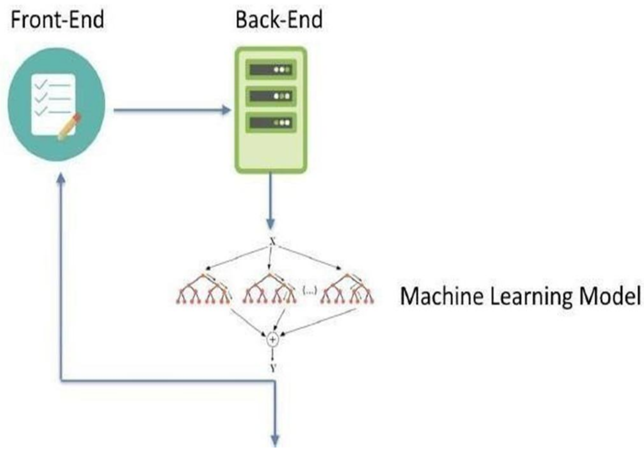
Fig. 1.

We've launched a web application featuring three integral components:

1) Front-End: The user interface allows individuals to input prediction factors, which are then transmitted to the backend server.
2) Back-End: This serves as the operational hub where the ML model is deployed. Input data is channeled into the model for analysis.
3) Machine Learning Model: Utilizing decision tree, logistic regression and random forest algorithms, we incorporated hyperparameter tuning to enhance efficiency. Notably, random forest algorithm exhibited the highest accuracy at 86.86%, leading to its selection for our model. The algorithm processes the input data, Estimate the severity of the incident on the scale: 1 for Most likely, 2 for likely, and 3 for less likely for accident to happen.

The model executes its predictions, and the results are relayedback to the front-end, through which they are displayed to theuser. This dynamic interaction ensures a seamless and informative experience for users seeking insights into the potential severity of traffic incidents. By amalgamating advanced ML techniques with user-friendly web interfaces, we aim to empower individuals with important information that can contribute to safer road practices.

Our web app signifies a significant leap towards proactive accident prevention. It not only provides a platform for usersto input relevant data but also integrates sophisticated machine learning algorithms, ultimately delivering accurate predictions on accident severity. The emphasis on the randomforest algorithm showcases our commitment to precision andreliability, ensuring that users receive meaningful and trustworthy insights through our predictive model. The flow of data across the enclosed application is shown below in Fig. 2



Severity Prediction and Suggestion.

Fig. 2.

There are four important steps:

A. *Preprocessing:*

1) Data Cleaning is a critical step in the data analysis process, encompassing the identification and handling of noisy or irrelevant data. Visualization aids in understanding the importance of various factors. Additionally, identifying and addressing missing values are crucial for robust analysis. In the provided dataset, 2 types of missing values, '-1' and 'NaN', are observed. A thorough investigationof columns with missing values is undertaken. Instead of imputing mean or median values, the dataset's size allows for analysis without such imputations.

2) To enhance data integration, a join method is employed to combine the accidents and vehicles files, leveraging the common primary key Accident Index. This consolidationfacilitates a more comprehensive analysis of the dataset.
3) Data Visualization becomes pivotal for extracting meaningful insights. Initial explorations involve understanding accident timing and the age of drivers involved. Analysis of accidents based on day of the week reveals that Thursday has the highest occurrence of accidents from 2005 to 2015. Although, it is essential to consider that accident numbers may be affected by varying traffic volumes on specific days.
4) Examining the time of accidents tells us that a concentration of incidents in the afternoon, suggesting heightened traffic during post-lunchhours, potentially associated with people leaving work.
5) Age Band of Casualties is examined bygrouping age bands into 11 codes and creating labels for representation. The analysis gives anidea about the distribution of age groups involved in accidents.
6) We examine the correlation among variables toidentify relationships within the numeric dataset. It's worth noting that strong correlations are generally limited, except for a positive correlation observed between the speed limit and the classification of the area as Urban or Rural.
7) The Speed of Cars is investigated, revealing that a majority of accidents occur in areas with a speed limit of 100. This finding contrasts with expectations, as one might anticipate more accidents on highways or major roadways. Possible contributing factors to accidents on lower speed limit roads include stop signs, lanechanges, or turns into parking lots.

B. Training
The Percentage of data that has beenused to train the model is 80%.

C. Testing
The Percentage of data that has beenused to test the trained model is 20%.

D. Web app integration
After this, the integration of the front-end and back-end components has been successfully completed.
The suggestion model functions by pinpointing the particular factor accountable for the alteration resulting in the anticipated result. It suggests adjusting the condition influencing the result. In instances where the severity level is assessed as low, it signals that everything is satisfactory. This approach streamlinesdecision-making processes by targeting key variables and proposing actionable adjustments, ensuring optimal outcomes are achieved while also highlighting areas where no immediate action is required due to minimal impact.
The severity rate is then presented to the Numerous variables influence outcomes, such as the age of individuals involved, day-of-week incidents, and time of day/night. Please see the image below to understand the events surrounding the accident.

III. RESULTS AND DISCUSSION
The front-end interface works as an entry point, prompting important forecasts such as the user's age, gender, vehicle type, vehicle age, engine capacity, and weather conditions. The user must enter these parameters explicitly. The prediction model is sent in the backend process. User input from the front end is then fed into the machine learning model.
The depiction illustrates user-input parameters, encompassing vehicle type, age, gender, speed limit, vehicleage, engine capacity, and additional variables. The diverse range of factors considered reflects the comprehensive natureof data collection for analysis.
For our prediction, we use the Random Forest algorithm, chosen for its high accuracy, resulting in an incredible accuracy of 86.86% in our model. The algorithm works by processing input data and creating an estimate of the severityof the situation. The severity category is defined as follows: 1means more likely, 2 means likely, and 3 means less likely. Figure 3 depicts the architectural diagram of the Model presented above. Once the prediction process is completed, the output is sent back to the front end.

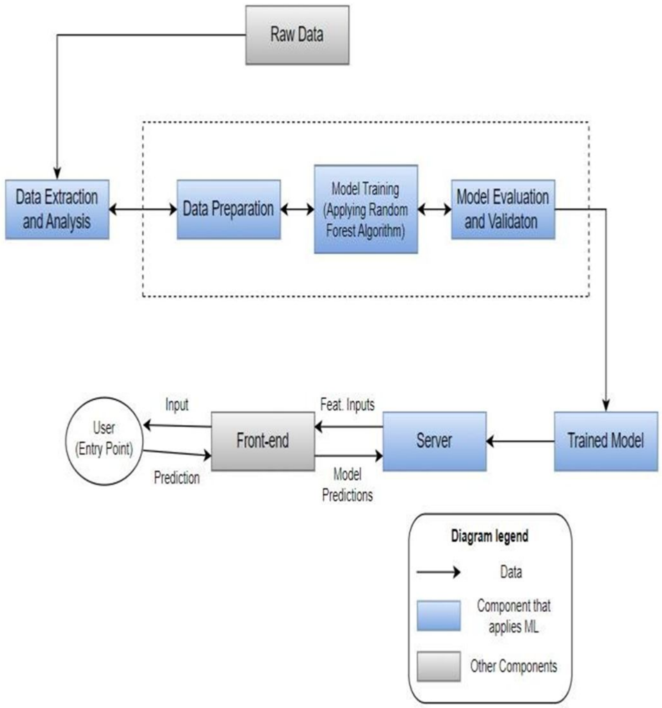


Fig. 3.

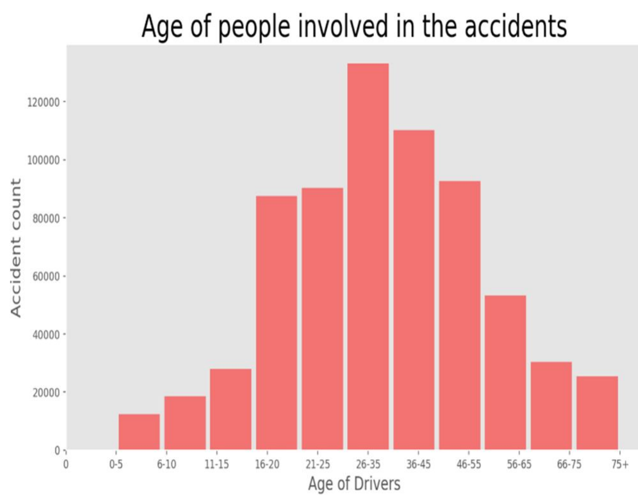


Fig. 4.

Examination of Fig 4 (mainly for driver age) reveals an important fact: people aged 26-35 are most likely to be involved in an accident. Our data analysis relies on the consensus of data analysis to guide the selection of relevant products. These findings provide a deeper understanding of crash patterns and help shape our dataset.

Likewise, the analysis of Figure 5 highlights a significant observation: a predominant number of accidents transpire onroads with a 30-speed limit. Surprisingly, drivers on thisroad often exceed the prescribed speed limit, a major contributor to the high incidence of accidents. Contrary to expectations, major highways did not witness as many accidents. Contributing factors include violations such as ignoring stop signs, lane changes, or turning into parking lots.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

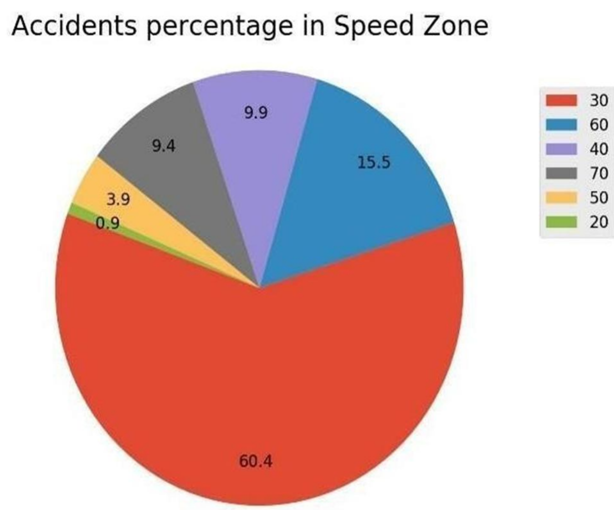


Fig. 5.

Illustrated in Figure 6 is the comprehensive user dataset. Upon the user's prediction request, the data is transmitted to the backend. Subsequently, it undergoes processing through our designated machine learning algorithm, specifically Random Forest. The resultant prediction is categorized based on severity: 1 for More likely, 2 for likely, and 3 for less likely, providing users with valuable insights into potential outcomes.

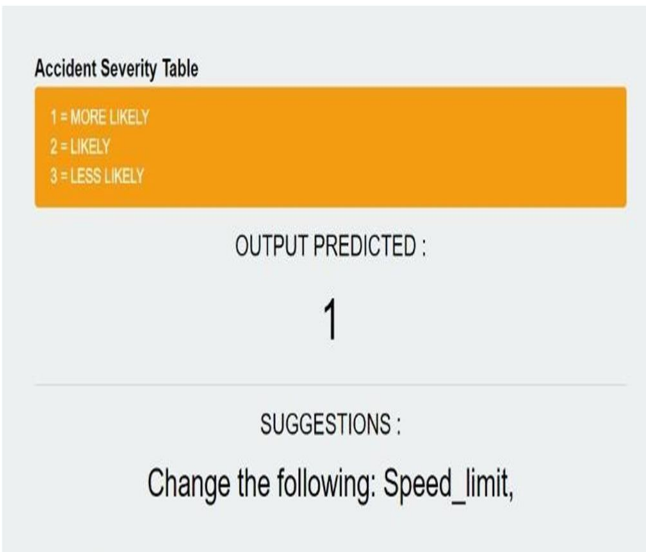


Fig. 6.

IV. CONCLUSION AND FUTURE WORK

This project focuses on employing Machine Learning classification techniques to anticipate the severity of accidents at specific locations. Harnessing the strength of ML allows us to analyze data more effectively, surpassing human capabilities. Our model, boasting an accuracy surpassing 17%compared to conventional systems, has been encapsulated intoa web-based application, utilizing the most precise algorithm. This innovation holds significant potential for governmental use, offering a proactive tool for accident prevention. By leveraging advanced technologies, this project contributes to enhanced accuracy in predicting accident severity, ultimatelypromoting safer road conditions and aiding in the strategic planning of preventative measures by government entities.

By allocating additional resources, we can establish a system for ongoing prediction and alerts, notifying law enforcement periodically about potential accident-prone locations. Integrating the web app with Google Maps allows real-time tracking by the police. The development of a comprehensive web application for direct interaction between users and law enforcement can be deployed for immediate use.

The web apphas great potential for use in Indian states or cities, dependingon whether the Indian government can provide accurate accident data. This strategic integration not only facilitates continuous monitoring and alerting for preventive measures but it also enhances police responsiveness through live tracking on familiar platforms, fostering a more efficient and timely approach to accident prevention and law enforcementinterventions.

REFERENCES

[1] Lu Wenqi, Luo Dongyu & Yan Menghua, “A Model of Traffic Accident Prediction” INSPEC Accession Number: 17239218 DOI: 10.1109/ICITE.2017.8056908

[2] Thineswaran Gunasegaran Yu-N Cheah, “Evolutionary Cross validation” INSPEC Accession Number: 17285520 DOI: 10.1109/ICITECH.2017.8079960

[3] Simon Bernard, Laurent Heutte and Sebastien Adam, “On the Selection of Decision Trees in Random Forests” INSPEC Accession

[4] Number: 10802866 DOI:10.1109/IJCNN.2009.5178693

[5] Rafael G.Mantovan,, Ricardo Cerri, Joaquin Vanschoren, “Hyper- parameter Tuning of a Decision Tree Induction Algorithm” INSPEC Accession Number: 16651860 DOI: 10.1109/bracis.2016.018

[6] Fu Huilin, Zhou Yucai, “The Traffic Accident Prediction Based on Neural Network”, 2011

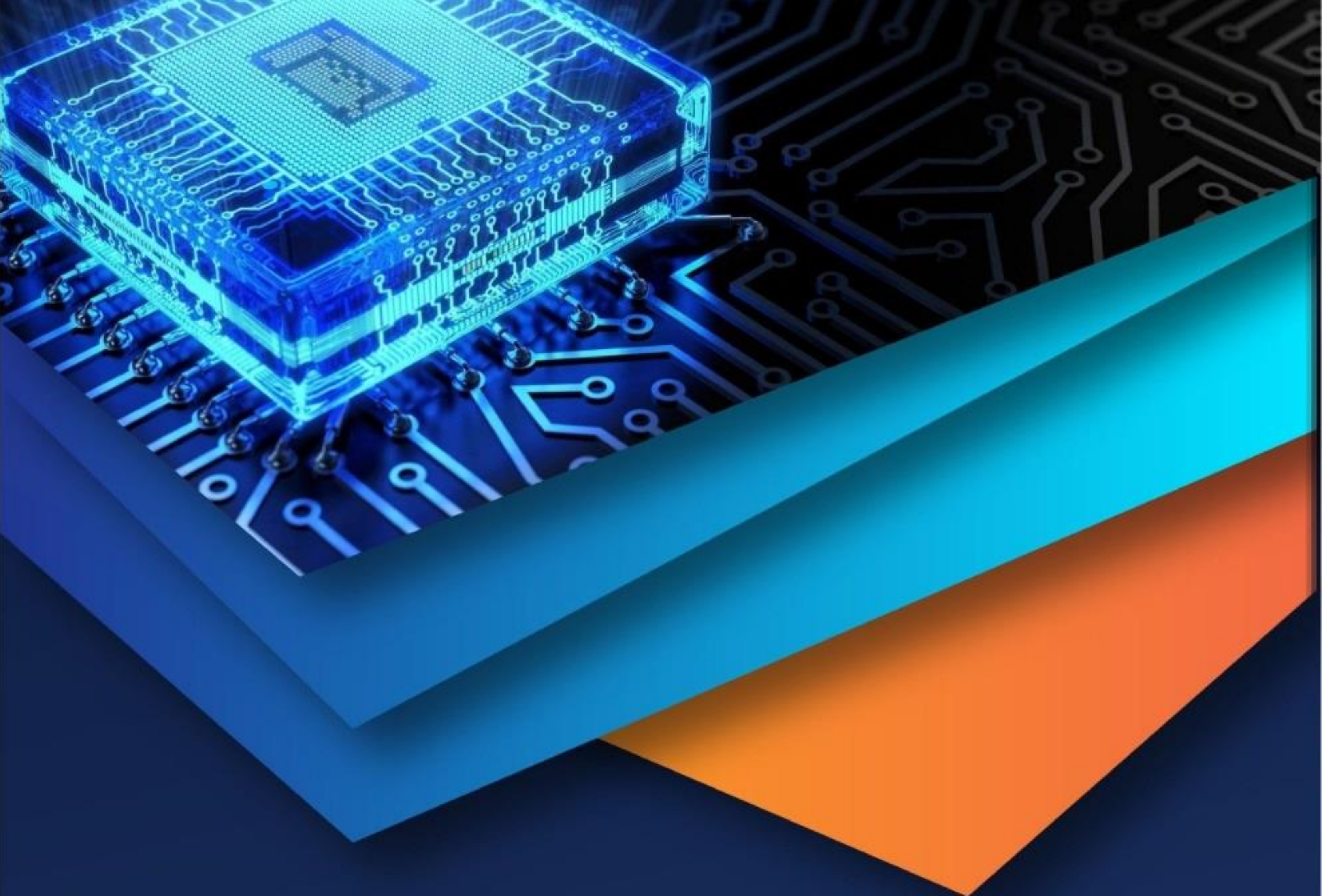
[7] Lin, L, Wang, Q, Sadek, A.W, 2014. Data mining and complex networks algorithms for traffic accident analysis. In: Transportation Research Board 93 Annual Meeting (No. 14- 4172).

[8] Gunasegaran, T., & Cheah, Y.-N. (2017). Evolutionary cross validation. 2017 8th International Conference on Information Technology (ICIT). doi:10.1109/icitech.2017.8079960

[9] Bernard, S., Heutte, L., & Adam, S. (2009). On the selection ofdecision trees in Random Forests. 2009 International Joint Conference on Neural Networks. doi:10.1109/ijcnn.2009.5178693

[10] Matthew Bihis ; Sohini Roychowdhury, “A generalized flow for multi- class and binary classification tasks: An Azure ML approach”

[11] Abdel-Aty, M., N. Uddin, and A. Pande. Split Models for Predicting Multivehicle Collisions during High-Speed and Low- Speed Operating Conditions on Freeways. In Transportation Research Record: Journal of the Transportation Research Board, No. 1908, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 51–58.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429