Predictive Analysis of Road Accidents Using Data Mining and Machine Learning

Bhuvanesh K[1], Sathya. M.R[2]

[1]*Bhuvanesh K, Jayalakshmi Institute of Technology*

[2]*Sathya M.R, Jayalakshmi Institute of Technology*

Abstract- **Prescient investigation utilizing information mining and machine learning is fundamental for making strides street security and decreasing accident-related fatalities. This considers applies different procedures to analyze chronicled street mischance information, centering on variables like climate, street sort, time, and activity thickness. Machine learning models such as Choice Trees, Arbitrary Woodlands, Bolster Vector Machines, and Neural Systems are utilized to foresee mishap seriousness, areas, and causes. Preprocessing strategies like cleaning and normalization improve show precision. The discoveries highlight the adequacy of prescient analytics in recognizing mishap designs, helping activity administration, and making difference policymakers actualize measures to anticipate mischances and spare lives.**

Keyword-: **Accident prediction, Machine Learning, Data mining, Data preprocessing, Classification**

I. INTRODUCTION

Street mischances are a basic issue in Tamil Nadu due to urbanization, rising vehicle numbers, and lacking foundation. The state reliably reports tall mishap and casualty rates, especially in thick urban regions like Chennai, Coimbatore, Madurai, and Salem. Key causes incorporate destitute street plan, need of signage, driver carelessness, and unfavorable climate conditions. From 2005 to 2025, a consistent rise in mischances is obvious over both urban and country districts. To address this, information mining and machine learning (ML) are progressively utilized for prescient investigation. These apparatuses prepare chronicled data such as time, area, and vehicle types to recognize mishap hotspots, foresee seriousness, and reveal patterns. Procedures like clustering, choice trees, and relapse offer assistance identify designs, whereas progressed models like neural systems figure accident-prone times and ranges. Geospatial examination assist improves exactness by mapping chance zones based on GPS information.ML calculations moreover anticipate activity infringement, such as overspeeding or path changes, making a difference activity specialists take preventive measures. With these prescient bits of knowledge, policymakers can execute focused on interventions like progressing street foundation, overseeing activity stream, and improving flag systems to diminish mischances and make strides street security in Tamil Nadu.

II. LITERATURE SURVEY

Later progressions in activity mischance expectation emphasize the utilize of machine learning and profound learning strategies. Lu Wenqi et al. proposed a Convolutional Neural Arrange (CNN) demonstrate to precisely anticipate mischance seriousness, leveraging its capability to analyze complex activity information. Fu Huilin and Zhou Yucai highlighted the restrictions of conventional straight models and illustrated that neural systems offer moved forward precision by capturing nonlinear designs. Thineswaran Gunasegaran and Yu-N Cheah presented an developmental cross-validation calculation to upgrade prescient execution by optimizing dataset dividing. Simon Bernard, Laurent Heutte, and Sebastien Adam centered on selecting the foremost compelling choice trees inside Irregular Woodland gatherings to progress demonstrate results. Rafael Mantovan, Ricardo Cerri, and Joaquin Vanschoren examined the affect of hyper-parameter tuning on choice tree calculations, emphasizing its significance in refining show accuracy. Collectively, these ponders uncover that shrewdly data-driven approaches are fundamental for understanding activity mischance patterns and creating compelling procedures for mishap anticipation and street security administration.

III. THE PROPOSED SYSTEM

1. There is currently no specific system available to help traffic police predict accident-prone areas at particular times.
2. Traditional Back Propagation (BP) neural networks have limitations, including lower prediction accuracy.
3. Studies indicate that the BP network performs with 17% less accuracy compared to the proposed approach.
4. The proposed system utilizes advanced machine learning techniques for more accurate and dynamic accident prediction.
5. Machine learning models can effectively handle complex, non-linear patterns in traffic data.
6. The system will analyze historical accident records, traffic flow, environmental factors, and time-based trends.
7. It aims to provide real-time, location-specific accident risk predictions to assist traffic authorities in preventive actions.

IV. THE EXISTING SYSTEM

The existing framework may be a machine learning-powered web application planned to anticipate the seriousness of street mischances based on current conditions. It is prepared on a expansive dataset comprising 1.6 million mishap records collected between 2005 and 2025, which altogether improves its prescient exactness. The essential objective of this framework is to recognize the conditions that are most likely to lead to extreme mishaps, empowering opportune preventive activities. By analyzing variables such as climate, activity thickness, area, and time, the framework gives real-time seriousness forecasts. Moreover, it points to pinpoint potential future mischance areas more accurately, permitting crisis administrations to reply quicker and more successfully. When a high-severity chance is anticipated, robotized alarms are sent to activity specialists, empowering them to require vital security measures in progress. This framework underpins activity administration operations by combining data-driven estimating with quick communication to diminish mishap affect and progress open street security.

V. SYSTEM ARCHITECTURES

Describes the System Architecture in a diagrammatic representation.
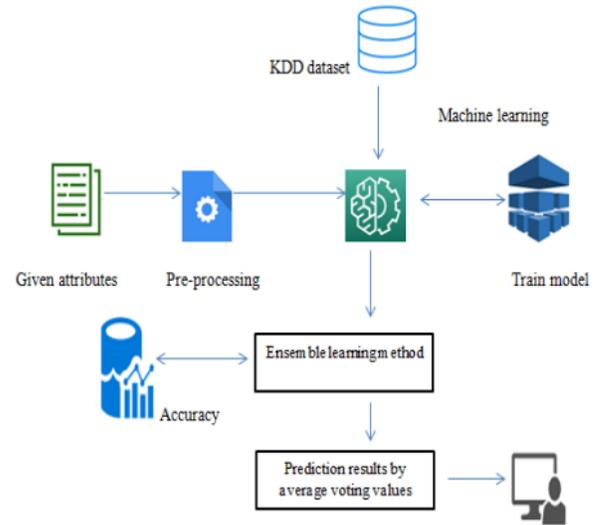
a) SSH Client

Secure Shell (SSH) may be a cryptographic organize convention that guarantees secure communication over an unsecured arrange. It is broadly utilized for inaccessible command-line get to, command execution, and securing any organize benefit in a client–server engineering. SSH makes a secure channel between an SSH client and an SSH server, supporting two adaptations: SSH-1 and SSH-2. Commonly working on TCP harbour 22, SSH is utilized on both Unix-like frameworks and Windows, with OpenSSH as the default client in Windows 10. Outlined to supplant Telnet and more seasoned unsecured conventions, SSH scrambles information to preserve privacy and keenness, anticipating unauthorized capture attempts or get to.

b) Diagrammatic Representation

A Data Flow Diagram (DFD) visually represents how data moves through a system or process. Using symbols like arrows, circles, and rectangles, it shows inputs, outputs, data stores, and process flow. DFDs help analyze existing systems or design new ones, making complex data processes easier to understand.
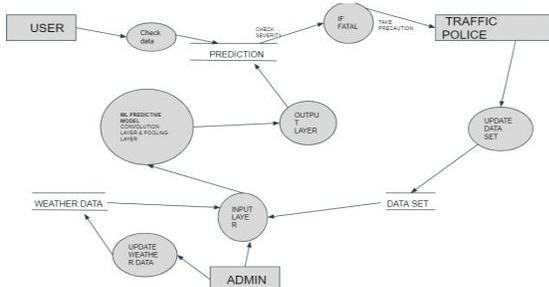
i) DFD level 0
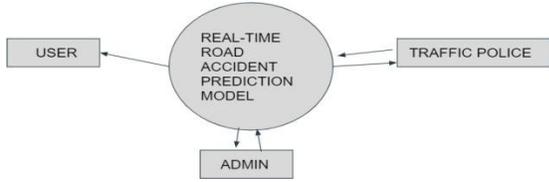
Admin: is responsible for building the ML model and maintaining it.
User: the person who views the output.
Traffic police: they take respective action according to the output predicted by the ML model.

ii) DFD level 1

The ML model is further divided into 3 layer
Input layer
Convolution layer or Pooling layer
Output layer

VI. RESEARCH METHODOLOGIES

The created web application coordinating machine learning for anticipating street mischance seriousness and comprises of four key components: front-end, back-end, machine learning demonstrate, and informing framework. The front-end collects client inputs such as area, age, sex, and other pertinent variables, which are at that point transmitted to the back-end server. The machine learning demonstrate is conveyed on the backend, where calculations like Choice Tree, Irregular Woodland, and Calculated Relapse are executed. To improve execution, hyperparameter tuning was connected, and the Arbitrary Timberland calculation, accomplishing the most elevated precision of 86.86%, was chosen for last sending. The show forms the input information and predicts mishap seriousness, categorized into three levels: 1 for Lethal, 2 for Genuine, and 3 for Slight. This expectation is sent back to the front-end and shown to the client. Furthermore, an SMS caution containing area arranges and seriousness is sent to the police, empowering incite preventive activity and making strides street security.

VII. CONCLUSION

The proposed framework presents an progressed and responsive approach to anticipating accident-prone locales, guaranteeing smoother activity stream and more secure street utilization. By analyzing numerous parameters such as area, natural conditions, and user-specific data the framework precisely decides whether an range is accident-prone or not. The user-friendly interface permits people to choose from different machine learning calculations, counting K-Nearest Neighbors, Arbitrary Woodland, and Calculated Relapse. The framework assesses these models and naturally receives the one with the most noteworthy precision, advertising clients exceedingly solid comes about. This flexibility upgrades believe within the system's predictions and permits clients to form educated choices, such as dodging high-risk zones. The application is appropriate for both open utilize in every day commuting & for specialists pointing to move forward street security. Moreover, the framework is adaptable and can be adjusted for individual utilize, making it effortlessly available. Generally, this arrangement holds awesome potential for legislative organizations to proactively avoid activity mischances and spare lives.

VIII. FUTURE ENHANCEMENTS

With access to more resources & comprehensive datasets, the system can be enhanced to provide **real time predictions and continuous alerts** to law enforcement agencies at regular intervals. This will enable authorities to take timely preventive measures and reduce accident risks significantly. Integration with **Google Maps** can allow for live tracking of accident-prone zones, enabling traffic police to monitor high-risk areas dynamically. A **fully developed web application** can be launched, supporting direct interaction between users & traffic authorities for improved responsiveness and safety management. Furthermore, the application can be scaled for use across **various Indian states and cities**, provided accurate and up-to-date accident data is made available by government agencies. This expansion would facilitate broader adoption of predictive traffic systems & support nationwide efforts to reduce road accidents and fatalities. Overall, the system holds strong

potential for development into a reliable, large-scale public safety tool in India.

IX. REFERENCE

[1] Ganesh Gopal Devarajan, U. Kumaran, C. Gopalakrishnan, Rajendra and Prasad Mahapatra, "Next-Generation Imaging Methodology: An Intelligent Transportation System for Consumer Industry", IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 3680-3687, Feb. 2024.

[2] M. Rithik and S. Chokkalingam, "Image Based Classification for Vehicle Accident Detection and Insurance Claim Using Artificial Intelligence", International Conference on Emerging Systems and Intelligent Computing, pp. 521-525, 2024.

[3] Shakil Ahmed et al., "A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance", Transportation Research Interdisciplinary Perspectives, vol. 19, pp. 100814, 2023, [online] Available: doi.org/10.1016/j.trip.2023.100814.

[4] Anand, J. V. "A Methodology of Atmospheric Deterioration Forecasting and Evaluation through Data Mining and Business Intelligence." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 02 (2020): 79- 87

[5] Zheng CT, Liu C, Wong HS. (2018). Corpus based topic diffusion for short text clustering. Neurocomputing 275: 2444-2458. http://dx.doi.org/10.1504/IJIT.2018.090859.

[6] ArunPrasath, N and Muthusamy Punithavalli. "A review on road accident detection using data mining techniques." International Journal of Advanced Research in Computer Science 9 (2018): 881-885.

[7] Sarkar S, Pateshwari V, Maiti J. (2017). Predictive model for incident occurrences in steel plant in India. In ICCCNT 2017, IEEE, pp.1-5.http://dx.doi.org/10.14299/ijser.2013.01

[8] Stewart M, Liu W, Cardell-Oliver R, Griffin M. (2017). An interactive web-based toolset for knowledge discovery from short text log data. In International Conference on Advanced Data Mining and Applications. Springer, pp. 853-858. http://dx.doi.org/10.1007/978-3- 319- 69179- 4_61.

[9] George Yannis, Anastasios Dragomanovits, Alexandra Laiou, Thomas Richter, Stephan Ruhl, Francesca La Torre, Lorenzo Domenichini, Daniel Graham, Niovi Karathodorou, Haojie Li (2016). "Use of accident prediction models in road safety management – an international inquiry". Transportation Research Procedia 14, pp. 4257 – 4266.