Accident Severity Prediction Using Machine Learning Techniques

S.V.V.Sai Kalyan, Shaik Mohammed Saif, Dr. Shanthi Thangam Manukumar

Sathyabama Institute of Science and Technology, mshanthithangam@gmail.com

Abstract.

Motor car accidents are the main purpose of death, incapacity, and hospitalization within the United States. Reducing street accidents and human casualties is vital. With this goal in thoughts, several techniques had been proposed, starting from vintage statistical models to new strategies stimulated with the aid of the advent of device getting to know. This research paper affords a comparative examination of numerous such models, aiming to analyse and derive a green method to predicting avenue twist of fate dangers. The aim of this observe is to expect the danger of injuries on the street by analysing elements which include vehicle type, age, gender, time of day, and climate earlier. Drivers display avenue conditions. In addition, reading geolocation facts the use of a good clustering algorithm allows us to examine wherein accidents occur. By reading the elements already diagnosed the use of these techniques, drivers attain estimates of the possibility of injuries, which enables them to make informed picks to lessen them. The purpose of this challenge is to increase the predictions of site visitors on the website and analyse road accidents.

Keywords. Machine Learning, Decision Tree, Random Forest, Logistic Regression, Road Accidents, Vehicle Safety, Climate.

1. INTRODUCTION

Energy With so many motors on the streets each day, street accidents can show up every time, anywhere. Some accidents are deadly, which means human beings die in accidents. As a human, everyone wants to stay accurately without getting harm. By applying statistics mining methods to the whole avenue harm dataset, we can advantage valuable insights into a way to drive accurately and offer tips to drivers. Data mining uses a wide variety of strategies and algorithms to discover relationships in large-scale facts. It has been taken into consideration one of the maximum vital gears within the IT business enterprise for decades. Association rule mining algorithm [2] is a famous set of policies to pick out meaningful relationships between information stored in a huge database, and it regularly plays a completely crucial role in content material mining. A common place approach for locating affiliation guidelines is a priori rule set, whose principal goal is to find common item sets that is the technique used for traffic data mining. One of the main probabilistic techniques of this type is Naive Bayes, which is specifically based on Bayesian hypothesis. For our study, we used the FARS dataset. This deadly crash dataset contains all deadly site visitors' crashes suggested via the National Highway Traffic Safety Administration in 2007. This dataset becomes downloaded from California Polytechnic State University and all the facts changed into acquired from FARS. This dataset has 37,248 statistics factors and fifty-five features. A description of the statistics may be found within the FARS report.

2. RELATED WORK

The Apriori set of rules is the maximum popular and efficient algorithm for association rule mining. This is because affiliation rule is utilized in all the modern-day global business and industrial tasks in data processing. The aim of the use of Apriori is often to stumble on devices and discover hidden records. The very last distinction is among a priori and a priori predictive algorithms, wherein a priori prediction is greater accurate and quicker than a priori [1]. Association rule mining algorithms are broadly used to locate all guidelines in a database that satisfy minimal nice assist and self-assurance constraints. To lessen the huge variation of generated regulations, preceding research have used a membership rule mining set of rules to extract an easy finite subset of affiliation policies, relationships among random components and twist of fate severity tiers have been explored the use of a priori strategies [2]. Data mining extracts hidden styles from massive databases. It is usually used in marketing and advertising, surveillance, fraud detection, and medical discovery. To classify the injury severity elegance of various road incidents, I compare the naive Bayes classifier, the AdaboostM1 met classifier, the component rule classifier, the J48 choice tree classifier, and the random wooded area classifier

[3]. Road accidents are one of the most important areas of research in India. There is many research on statistics gathered from police facts regarding precise sections of two-lane roads. Statistics mining strategies to explore the statistics provided via EMRI, wherein we first divide the matching statistics into clusters to the results are used to take coincidence prevention measures in areas diagnosed for specific types of injuries to reduce the wide variety of accidents [4]. The clusters generated via K-tunes and estimate maximization algorithms are analyzed to extract latent styles the use of apriori algorithms. The implications of this demonstration display that the chosen device mastering strategies can extract hidden styles from the information. Density histograms are used to visualize the twist of fate facts [5].

3. RESEARCH METHODOLOGY

Classification models to are expecting the severity of damage in avenue person accidents. Based at the statement rule inversion dataset accrued from the State Transportation affiliation rule mining algorithm, a priori and predictive club rule algorithms have been carried out at the dataset to track the connection between the desired crashes and the components that affect the crash severity. The information accrued on street fatalities is very significant, multifaceted and heterogeneous. The purpose of this work as shown in figure 3.1 is to expand the evaluation competencies of observers, avenue twist of fate analysis and navigation visualization. It can be used to create a first-class itinerary that connects all key locations so that passengers feel safe at any factor in their journey. This paper modeling the severity of harm in avenue traffic accidents the use of synthetic neural networks and decision bushes. Based on real global facts acquired from the General Assessment System (GES) of the National Automobile Sampling System (NASS). Experimental consequences display that the choice tree performs better than the neural network in all instances. In our take a look at, it was also determined that the motive force's seat belt use, avenue situations and driver alcohol consumption are the three most crucial elements for deadly injuries. Our tests showed that this version performs a far extra critical role in dangerous and deadly accidents than in different instructions. Since driving force deaths cause large economic and social losses to society, it's miles critical to evaluate the danger of fatal and extreme accidents.

Figure 3.18. Proposed System Architecture of Accident Severity Prediction

The facts about series module, twist of fate data is collected from the Cargill. For instance, the year, month, United States of America, time of the programming language. In this module, someone specifies a selected characteristic, from which an estimate of its accident impact is decided. The education dataset is analyzed the use of the assist vector kind and a user-described characteristic. The report web page allows you to talk with the website administrator via reporting the recent recognition of fatalities and houses even as traveling on the selected path. This report is sent to the database occasions desk, in which its miles are demonstrated and displayed by means of the administrator, and several fields in the database are up to date to create a valid document. Finally, it affords accurate records approximately traffic accidents with the aid of nation.

4. RESULTS AND DISCUSSIONS

The experimental consequences display that choice trees continually outperform neural networks in all cases, highlighting their effectiveness in processing categorized facts and identifying key choice thresholds. The evaluation identified three important elements contributing to fatal accidents, driver seat belt use, avenue lighting fixtures conditions, and motive force alcohol consumption. Furthermore, the model accomplished excellently in distinguishing among deadly and non- deadly injuries, that is especially critical given the vast financial and social fees related to motive force fatalities.

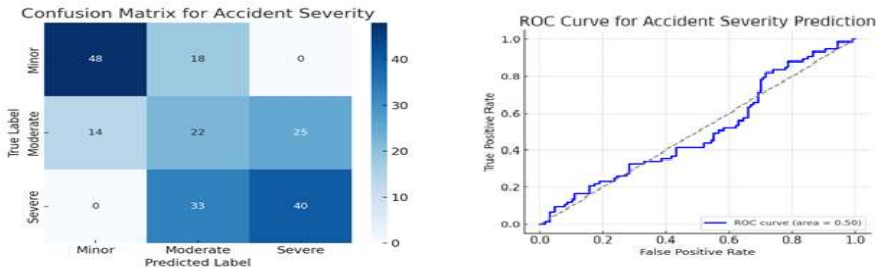

Figure 4.19. Confusion Matrix and Roc Curve

Confusion matrix shows how well the model predicts accident severity. The rows represent the actual accident severity (Minor, Moderate, Severe), while the columns represent the predicted severity. The ROC (Receiver Operating Characteristic) curve evaluates the model's ability to distinguish between severe and non-severe accidents. The X-axis (False Positive Rate) measures how often the model incorrectly predicts a severe accident when it isn't. The Y-axis (True Positive Rate) measures how well the model correctly identifies severe accidents. The AUC (Area Under Curve) value indicates performance—closer to 1 means better prediction ability. The Above ROC (Receiver Operating Characteristic) curve evaluates the model's ability to distinguish between severe and non-severe accidents. The results of Confusion Matrix and Roc Curve are as shown in Figure 4.1.

5. CONCLUSION

The affiliation's analysis of guidelines and publications indicates, environmental factors together with road floor, climate, and lights do now not have a first-rate impact on fatality quotes, not like human elements such as alcohol consumption or kind of collision. High effect on fatality costs. From the aggregation results, we see that a few states/territories have excessive fatality rates, while others have low fatality prices. The extra information about fatality statistics, climate records, mileage information, etc., can do more checks, because of this may draw greater conclusions from the records. The proposed methodology gives improved results than the existing system.

6. REFERENCES

[1] Fu, X.; Meng, H.; Wang, X.; Yang, H.; Wang, J. A hybrid neural network for driving behavior risk prediction based on distracted driving behavior data. PLoS ONE 2022, 17.

[2] Mukherjee, D.; Mitra, S. Pedestrian safety analysis of urban intersections in Kolkata, India using a combined proactive and reactive approach. J. Transp. Saf. Secur. 2022, 14, 754–795.

[3] Jeong, H.; Kim, I.; Han, K.; Kim, J. Comprehensive Analysis of Traffic Accidents in Seoul: Major Factors and Types Affecting Injury Severity. Appl. Sci. 2022, 12, 1790.

[4] Yang, Z.; Zhang, W.; Feng, J. Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework. Saf. Sci. 2022.

[5] Wang, J.; Lu, H.; Sun, Z.; Wang, T.; Wang, K. Investigating the Impact of Various Risk Factors on Victims of Traffic Accidents.Sustainability 2020, 12, 3934.

[6] Elassad, Z.E.A.; Mousannif, H.; Al Moatassime, H. Class-imbalanced crash prediction based on real-time traffic and weather data:A driving simulator study. Traffic Inj. Prev. 2020, 21, 201–208.

[7] Banerjee, K.; Bali, V.; Sharma, A.; Aggarwal, D.; Yadav, A.; Shukla, A.; Srivastav, P. Traffic Accident Risk Prediction Using MachineLearning. In Proceedings of the 2022 International Mobile and Embedded Technology Conference (MECON), Noida, India,10–11 March 2022; pp. 76–82.

[8] Moghadam Gilani, V.N.; Hosseinian, S.M.; Ghasedi, M.; Nikookar, M. Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models. Math. Probl. Eng. 2021, 2021, 1–11.

Biographies

Mr. S.V.V.Sai Kalyan, a student of B.E. degree in the specialization of CSE with AIML, in Sathyabama Institute of Science and Technology, Chennai. . His research interests are in the area of Artificial Intelligence and Machine Learning and Optimization Algorithms

Mr. Shaik Mohammed Saif, is a student of B.E. degree in the specialization of CSE with AIML, in Sathyabama Institute of Science and Technology, Chennai. His research interests are in the area of Artificial Intelligence and Machine Learning and Optimization Algorithms.

Dr. Shanthi Thangam Manukumar is currently working as Assistant Professor in Sathyabama Institute of Science and Technology. She completed Ph.D. degree in Network Engg., at College of Engineering Guindy, Anna University, Chennai. She received her M.E. degree in Network Engineering from Anna University Coimbatore and B.E. degree in CSE, Anna University. Her research interests are in the area of Cloud Computing, Computer Networks, and AI. She has been serving as a reviewer for many highly-respected journals.