

Development of Traffic Imaging for Accident Detection System Using AI: A Review

Prof. Rahul Bhandekar¹, Sanjana Fulzele², Chitaksha Bhojar³

¹Project Guide, Department of Computer Science & Engineering, Wainganga College of Engineering and Management, Nagpur, Maharashtra.

^{2,3}Students, Department of Computer Science & Engineering, Wainganga College of Engineering and Management, Nagpur, Maharashtra.

Abstract-

Road accidents are one of the most relevant causes of injuries and death. This is also one of the serious issues, which can possibly cause disabilities, injuries and even fatalities. There are many of reasons that contribute to accidents. Some of them are internal to the driver but many are external. For example, adverse weather conditions like rain, cloudy, and fog cause partial visibility and it may become difficult as well as risky to drive on such roads. This project aims to provide an Overview of the area of the art in the prediction of road accidents through clustering techniques and machine learning algorithms. The dramatic increase in road traffic accidents in the world is causing serious problems in every aspect of human lives. The most important and meaningful nature of traffic characteristics, causation analysis, and associations between different causal factors have been ignored. Moreover, the traffic accident data is only used to conduct a rudimentary statistical analysis and data mining efforts which results only in patterns and statistics. The main targets of this road accident data classification are to identify the major and key factors that cause the road traffic accident and form policies and preventive actions that would reduce the accident severity level. Machine learning algorithms are used to analyze the data, extract hidden patterns, predict the severity level of the accidents and summarize the information in a useful format.

Keywords – Road Accident Prediction, Machine Learning, K-Means Clustering, Traffic Safety, Accident Severity etc.

1. Introduction

According to the death statistics released by the World Health Organization, the number of traffic accidents occurring annually in the world is alarming. And Road and accidents are uncertain incidents. In now a day's traffic is increasing at a huge rate which leads to a large numbers of road accidents. So, Road accident prediction is one of the most important research areas in traffic safety. The occurrence of road traffic accidents is mainly affected by geometric characteristics of road, traffic flow, characteristics of drivers and environment of road. Many studies have been conducted to predict accident frequencies and analyze the characteristics of traffic accidents, including studies on hazardous location/hot spot identification, accident severities.

Some studies focus on mechanism of accidents. Other factors include weather and light conditions of the road. No specific approach available for the traffic police to predict which area is accident prone. The road accident prediction play an important role in the integrated planning and management of traffic, the reason which with much randomness about the traffic accident include some nonlinear elements, such as people, car, road, climate and so on. Machine Learning algorithms can process large number of classification parameters and are able to obtain useful patterns. It can process huge amounts of data efficiently and can be scalable. And also the clustering technique is also helps in analyzing and visualizing the road accident data.

According to a system concept based on manenvironment changes and instabilities of Muhlard and Lassare, the main reason of road traffic accident is usually by faults in three main constituents of road

safety such as the environment, the human, and the vehicle [4]. The environmental factors incorporate the natural and constructed environments as well as the transportation networks. The most common environmental factors affecting road traffic accidents are weather condition, dusty air, heavy wind, light conditions. The infrastructural factors under environment are road type, road condition, road lane type, poor or defective Road surface, inadequate road markings, defective traffic signs, and road layout. The human constitutes includes sex, age, educational level, human behaviors, driving skill, driving manner, risk recompense, and hazardous driving (use of drugs and alcohols), speeding, breaching traffic lights, driving behavior. The vehicle constituent includes the configuration, age, volume, and quality of vehicles like technical conditions and safety equipment, design defects.

2. Problem Statements

- Road accidents remain one of the leading causes of injuries and fatalities worldwide, posing significant challenges to public safety. The primary issue lies in the inability to predict and prevent accidents effectively due to the complex interplay of various factors such as driver behavior, road geometry, traffic flow, weather conditions, and environmental aspects.
- Traditional accident analysis systems depend on static data and statistical models, which fail to uncover intricate patterns and relationships among these factors. The lack of real-time data processing and predictive capabilities limits the effectiveness of these systems in reducing accident occurrences.
- Additionally, there is no standardized approach available for traffic authorities to identify accident-prone zones proactively. Manual data collection methods often result in incomplete or inaccurate datasets, further hampering the reliability of the analysis.
- A robust and intelligent system is required to analyze accident data comprehensively, identify high-risk zones, and provide actionable insights to enhance traffic safety measures.

3. Literature Review

Currently, road traffic accidents have been both developmental and public health concern and demanded the concern of researchers, civil societies, vehicle companies, governments, and business societies in the whole world.

Masashi Toyoda et. al. 2017, In the framework [1] a large-scale heterogeneous accident data is used to estimate the occurrence of traffic accidents on the road crossings. The authors proposed an object detection method called Fast R-CNN for photo features extraction and XGBoost for extracting road map and driving record features. The overall experimental outcome shows that the algorithms used in this work can extract the possibly dangerous crossings with a good performance.

Jonardo R. Asor et. al. 2018, To classify and discover hidden patterns using a dataset that contains the road accident records obtained from the Philippines National Police (PNP), Naïve Bayes, Decision Tree, and Rule induction are used in [2]. They use a Rapid miner data mining tool to analyze the accident data. The authors revealed that the places where accidents occur don't have a significant relationship on the fatality of the victims. The results show that key factors affecting the accident and found that time and day are the most critical causes for the severity and fatality of road accident victims and the algorithms attain the expected accuracy.

Sadiq Hussain et. al. 2018, This research article states that to predict the main influencing factors of accidents like reasons of the accident, accident-prone sites, the severity of the accident, and type of vehicle involved and so on many research studies conducted in order to increase the performance of DM classification [3]. The authors used two data mining tools, Weka and Orange to evaluate J48, Multi-layer Perceptron, Bayes Net classifiers of 150 instances of the dataset. Evaluation metrics for measuring data mining techniques such as accuracy, precision, recall or sensitivity and so on are applied to identify the best algorithm for the prediction of the accident dataset. The experimental result shows that Multi-layer Perceptron is the best for the prediction of the accident database with an accuracy of 85.33%.

Tibebe Beshah Tesema et. al. 2012, the authors used a genetic algorithm to develop a symbolic fuzzy classifier on traffic accident dataset obtained from Addis Ababa traffic office. The symbolic classifier used to select features from accident dataset. The result shows that the developed classifier is able to separate and classify the classes of injuries and the attributes used for data labeling are easily extracted and explored [4].

Tibebe Beshah Tesema et. al. 2011, This author also proposed a machine learning experimental research which uses a traffic accident data gathered in Ethiopia [5]. They used CART, Random Forest, MARS and Tree Net algorithms to develop a predictive model focused on exploring the issue of data quality and predicting the impact of road behaviors on potential injury risks. The models can identify the human-related causal factors for the accident severity. The combined techniques used in this work proved to be effective in terms of predictive accuracy.

Girija Narasimhan et. al. 2017, was developed to employ predictive analytics through innovative machine learning models to predict the future result of the accidents' number in Oman. The author utilized Boosted Tree Regression model which is based on the decision tree and Multiplicative model to increase the prediction results [6]. This work states that the sole or contributory factor for accident cause is related to human factors which counted about 91% of the total accidents and the rest 9% is non-human factors.

Li, L, Shrestha, et. al. 2018, Apriori, Naïve Bayes, and k-means clustering algorithms are used in [7] to analyze the FARS dataset with the aim of examining the relationship between fatal rate and other features such as a drunk driver, light condition, collision mode, weather condition, and road surface conditions. In this research work, the variables associated highly with fatal accidents are articulated. The result shows that the factors related to human factors like a drunk driver cause a high fatality rate.

Ms. Nidhi. R, et. al. 2018, While the work in [8] utilized Naïve Bayes and Apriori algorithms to predict patterns in the road accident. In this work, the authors developed a prediction model to predict the frequently occurring accident types on new roads based on the

association rule. From the analysis, results show that most of the accidents are occurred by vehicles age less than five years and there is a high mortality rate in the rural areas.

As stated by several researchers, data mining techniques have a vast role in analyzing and predicting the future value of road accidents records and in identifying the patterns of the components of accidents determining different factors. In addition, the great potential of data mining prediction techniques plays a major role in avoiding and monitoring the problems of road accident safety.

4. Research Methodology

Block Diagram:

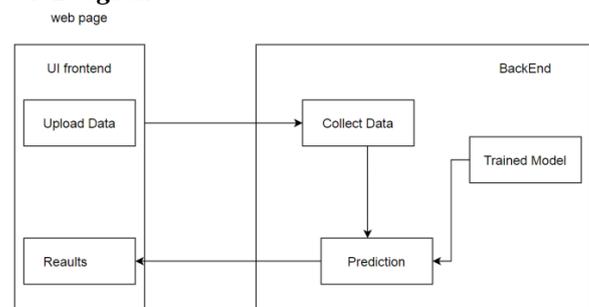


Figure 1: Block diagram of system

Working:

For any data analysis, the most important aspect is the data. Collecting the right kind of data is very important. Analyzing and understanding the content and structure of the data needs special attention. The data used here for the analysis is been taken from Kaggle and government websites.

Once the data is been collected, the task of analyzing it comes further. To analyze the data we need some tool which simplifies the work. We had a clear idea of using Python for coding.

The packages which played a major role in the analysis are pandas and numpy. Pandas is used for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It provides high- performance, easy to use structures and data analysis tools.

NumPy stands for Numerical Python or Numeric python. It is open source module which provides fast

computation on arrays and matrices. NumPy is the fundamental package for scientific computing with Python.

Now talking about the algorithm we used. There are a lot of algorithms present which help us in analyzing data. Machine learning and data analytics techniques are a boon in this field. The algorithm we opted for is Regression Analysis.

Logistic Regression Analysis is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between dependent variable and one or more independent variables.

Tools / Platform to be used :

As its definition states, learning is the activity of obtaining knowledge by studying something. In the particular case of Machine Learning, this learning activity is carried out by a computer, thus enabling the construction of computer program that automatically improve with experience. There are three niches for Machine Learning.

- **Data mining:** this kind of systems aim at using huge quantities of data which humans for themselves cannot handle in order to improve decisions. This offers, for example, a very useful application in the medical field, since it allows for the creation of medical knowledge based on medical records.
- **Software applications:** unbelievable as it may seem, not everything in this world can be programmed by humans. However, with the aid of Machine Learning techniques, these horizons can be broadened. For instance, these kinds of techniques are currently being successfully applied in fields like autonomous driving, speech recognition, image recognition or, as in the case of this project, Automatic Number.
- **Self-customizing programmers:** although most people might not realize, almost everybody has contact with this last niche every day. In effect, it is

this kind of technology the one that is behind the news feeds that people usually receive according to their personal interests when surfing the Internet.

The way all this is achieved, that is, the way Machine Learning systems actually work is by means of using a series of labeled training examples based on which the algorithms are able to create general target functions that applied to a new, unseen dataset are able to correctly predict the expected result. These training labeled examples are called training dataset and the new, unseen dataset is called testing dataset. In particular, a sufficiently good and complete training dataset is crucial in this kind of applications, since a poor training always offers poor results.

Final output is represented in the html page linked to the main system is shown in the GUI window. Output page consist of button "Load Data". After liking on this button the script is executed which extract the values from the dataset. Apart from the fundamentals that have just been explained, Machine Learning algorithms are so many and so varied that they have very little in common. In effect, even one single Machine Learning algorithm can be designed in infinite ways. Hence, for selecting the most appropriate design, a good evaluation is key, which is usually carried out by means of several statistics.

5. Advantages

1. **Accurate Prediction:** Machine learning algorithms enable precise identification of accident-prone zones.
2. **Data-Driven Insights:** Provides actionable insights by analyzing large datasets and uncovering hidden patterns.
3. **Proactive Safety Measures:** Helps in implementing safety measures to prevent accidents before they occur.
4. **Cost-Effective:** Reduces economic losses by minimizing accidents and associated costs.
5. **Improved Traffic Management:** Supports better planning and management of traffic flow.

6. Applications

- **Traffic Safety Planning:** Identifying hazardous zones and implementing safety measures.

- Urban Development: Assisting city planners in designing safer road infrastructure.
- Law Enforcement: Helping traffic police prioritize monitoring of high-risk areas.
- Insurance Analysis: Supporting insurers in assessing accident risks and claims.
- Real-Time Monitoring: Integration with IoT systems for dynamic traffic control and accident prevention.

7. Conclusion

Road accidents are a major cause of fatalities and injuries worldwide, often resulting from a combination of driver behavior, road conditions, and environmental factors. This project highlights the importance of using machine learning techniques, such as k-means clustering, to predict and analyze accident-prone zones effectively. By leveraging historical accident data, machine learning algorithms can uncover hidden patterns, identify risk factors, and provide actionable insights to reduce accidents.

The k-means clustering it's an unsupervised learning which is used for the unlabelled data therefore data are not labelled into any group of cluster. And also in this study, the techniques of regression with a large set of accident's data to identify the reasons of road accidents were used. Analysis is done for the identification of factors involved in the accident that occur together which is then plotted in a graph form. This shares a lot in understanding the circumstances and causes of accident. And this ultimately helps the Government to adapt the traffic safety policies with different types of accidents and situations.

The study emphasizes the significance of factors such as weather conditions, traffic flow, and road geometry in predicting accident severity and frequency. The use of unsupervised learning methods allows for better categorization of accident zones, supporting traffic authorities in implementing targeted safety measures. Ultimately, this approach aids in enhancing traffic management, reducing fatalities, and optimizing infrastructure planning. Machine learning offers a proactive solution, empowering stakeholders to prioritize safety and make data-driven decisions to mitigate road accidents effectively.

8. Future Scope

With more resources, continuous prediction and alerts can be sent to the police for every location at regular intervals of time to take preventive measures. The web app can be incorporated with Google Maps which can be live tracked by the police. A fully-fledged web app for user and police interaction can be published for use in real-time. It can be used for Indian states or cities, if proper data of accidents is provided by the Indian Government.

Acknowledgment

I would like to use this opportunity to show our sincere appreciation and respect to our project guide at the Wainganga Collage of Engineering and Management, Nagpur, who gave us direction and space to complete this assignment.

References

- [1] Masashi Toyoda, Daisaku Yokoyama, Junpei Komiyama, Masahiko "Itoh 2017 IEEE International Conference on Big Data," Boston, MA, 11-14. 2017
- [2] Jonardo R. Asor, Gene Marck B. Catedrilla, " A study on the road accidents using data Investigation and visualization in Los Baños, Laguna, Philippines" 2018 International Conference on Information and Communications Technology (ICOIACT) Yogyakarta, Indonesia. Page s: 96 – 101. 6 March 2018
- [3] Sadiq Hussain, L. J. Muhammad, F. S. Ishaq, Atomsa Yakubu and I. A. Mohammed "Performance Evaluation of Various Data Mining Algorithms on Road Traffic Accident Dataset" Information and Communication Technology for Intelligent Systems pp 67-78, 30 December 2018.
- [4] Tibebe BeshahTesema, Ajith Abraham, Dejene Ejigu, Learning the Classification of Traffic Accident Types"2012 Fourth International Conference on Intelligent Networking and Collaborative Systems. Page s: 463 – 468, 19-21Sept. 2012

- [5] Tibebe Beshah Tesema, Dejene Eijigu, Ajith Abraham, “Knowledge Discovery from Road Traffic Accident Data In Ethiopia” 2011 World Congress on Information and Communication Technologies Page s: 1241 – 1246. Year: 2011
- [6] Girija Narasimhan, Ben George Ephrem, et al, “Predictive Analytics of Road Accidents in Oman using Machine Learning Approach” 2017 International Conference on Intelligent Computing, Instrumentation, and Control technologies (ICICT) Page s: 1058 - 1065 July 2017
- [7] Li, L, Shrestha, S, and Hu, G, “Analysis of Road Traffic Fatal Accidents using Data Mining Techniques,” 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) Page s: 363 – 370, Year: 2017
- [8] Ms. Nidhi. R, Ms. Kanchana V, “Analysis of Road Accidents Using Data Mining Techniques” International Journal of Engineering & Technology, Volume 7 (3.10) page 40-44, (2018)
- [9] Sadiq Hussain “Survey on Current Trends and Techniques of Data Mining Research “London Journal of Research in Computer Science and Technology Volume 17 | Issue 1 | Compilation 1.0 (March 2017)
- [10] Jesus Alcal ´ a-Fdez, Rafael Alcal ´ a, and Francisco Herrera “A Fuzzy Association Rule-Based Classification Model for High- Dimensional Problems With Genetic Rule Selection and Lateral Tuning” IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 19, NO. 5, OCTOBER 2011
- [11] Niklas Donges Machine learning-blog.com Tutorials and explanations about applied Machine Learning “The Random Forest Algorithm,” 06 February 2016.