OPEN ACCESS

Volume: 11

Special Issue: 1

Month: July

Year: 2023

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Impact Factor: 3.025

Received: 08.05.2023

Accepted:13.06.2023

Published: 01.07.2023

Citation:
Deepa, KR, and S. Abhishek. "In Order to Forecast the Likelihood of a Traffic Collision, Machine Learning Is Used." *Shanlax International Journal of Arts, Science and Humanities*, vol. 11, no. S1, 2023, pp. 1–7.

DOI:
https://doi.org/10.34293/sijash.v11iS1-July.6306

In Order to Forecast the Likelihood of a Traffic Collision, Machine Learning is Used

Deepa K.R
Department of Master of Computer Applications
RajaRajeswari College of Engineering, Bengaluru

Abhishek S
Department of Master of Computer Applications
RajaRajeswari College of Engineering, Bengaluru

Abstract

Road accidents continue to be among the leading causes of death, disability, and hospitalisation inside the nation. This makes traffic accident risk prediction critical to be capable reduce them and save lives. A number of models had been developed to accomplish the same goal, ranging from ancient statistical models to modern models influenced by advent of machine learning. This work compares many of these models in an effort to examine and derive a helpful method to traffic accident risk prediction. Because drivers are in charge of the road, the study's goal is to give traffic accident risk prediction to drivers by examining information they would be aware of beforehand, as an instance vehicle type age, gender, time of day, weather, and so forth. additionally the usage of Random Forest and Logistic Regression, Optimal Classification Trees is a model capable of producing such outcomes that make intuitive sense to the driver. Furthermore, geo-location data analysis utilising the K-means clustering technique can offer information about high-accident locations.The road has gotten more difficult in the design and management sectors as the quantity of cars on the road has increased. Traffic accidents are a major source of worry on a worldwide scale because they have a considerable impact on people's safety, health, and well-being. The World Health Organisation (WHO) estimates that 1.35 million people die in automobile accidents each year. The result isthey represent a substantial field of research for utilisation of cutting-edge methodologies and algorithms for analysis and prediction. While many traffic accidents occur by external factors, some are caused by the driver. Unfavourable weather conditions, such as rain, clouds, and fog, for example, impair visibility and make driving on such roads difficult and often deadly. The present system prediction model evaluated only many probable causal factors.

Keywords: Road Accident, Traffic Accident, Machine Learning, K-means, Geo-Location.

Introduction

The Traffic Accident Risk Prediction project wants to develop a system that predicts the possibility of road accidents using (ML) Machine Learning algorithms. Road accidents are a major public safety problem, with millions of incidents happening each year throughout the world, resulting in thousands of fatalities and injuries. Existing traffic accident prediction systems frequently depend on statistical models and heuristic-based techniques that may be incapable of capturing the underlying data's complexity. Machine Learning algorithms, on the opposite hand, a more data-driven strategy that can evaluate huge datasets and find patterns and linkages that are not immediately apparent visible.The suggested system would analyse historical traffic accident data using a Random Forest Classifier algorithm, this will consist of items such as location, time, weather conditions, road conditions, and other important aspects. The algorithm will be optimised to deliver reliable forecasts of traffic accident risk, and it will feature a user-friendly interface that allows people to enter pertinent information and obtain a risk score indicating the chance of a traffic accident occurring. The technology has the ability to improve road safety and minimise the appearance of traffic accidents by giving useful insights on the risk factors related to traffic accidents to drivers, transportation authorities, and other stakeholders. Furthermore, the system can aid in the creation of targeted initiatives and policies aimed at alleviating traffic congestion.

Literature Survey

Vehicle collisions[1] are among the leading causes of injury and global death, and as such, they represent an important area of investigation into utilisation ofmodern algorithms and techniques to assess and anticipate traffic accidents, in addition identify the most critical elements that lead to accidents on the road. The goal of road accident prediction research is to answer to the problem of creating a more safe transportation environment and, eventually, saving lives.This paper's goal is to give an overview of the state-of-the-art in road accident prediction using deep learning architectures, advanced information-analysis methods, and machine learning algorithms, including convolutional neural networks and long short-term memory networks. This article also aggregates and examines the most often used data sources for predicting traffic accidents.A classification based on its origin and characteristics, such as open data, measuring technology, onboard equipment, and social media data, is additionally proposed..For information analysis, the many methods used to forecast Road mishaps are discussed and contrasted, also their relevancebased on the kind of data being examined, the conclusions obtained, and is simple to comprehend and analyze.In this study, a systematic technique for identifying abrupt braking occasion incidence association[2] as time and place is provided. The suggested method, which is built regarding real-time and batch clustering approaches, uses data from the past and data from the present to anticipate the moment of extreme braking and its location occurrences. To construct groups representing the original correlation patterns, clustering in a batch accomplished using a mix fuzzy c-means clustering and subtractive clustering. Then, real-time correlation patterns are generated and maintained based on the evolving Gustafson Kessel Like (eGKL) approach. batch clustering. current driving information from operating automobiles equipped with the suggested technique is validated using data gathering and a wireless communication platform. Drivers maymade aware of the circumstance. Road and traffic accidents[3] are a major worry all over the world. Road accidents not only endanger public health by causing varying degrees of harm, but they also cause property loss. Data analysis can determine the many causes of road accidents, such as traffic factors, weather characteristics, road features, and so on. Several research on road accident data analysis have been conducted. have previously demonstrated its significance. Some research concentrated on identifying factors related with accident severity, whereas others concentrated on identifying factors associated with accident occurrence. Traditional statistical approaches as well as data mining methods were utilised in these study investigations. In current study, data mining is a popular tool for analysing road accident data. Another major study field in the world of road accidents is trend analysis. Data mining[4] has been demonstrated to be a viable approach for analysing road accidents and producing useful findings. The vast majority of road accident data analysis use data mining techniques, with the goal of discovering characteristics that influence the severity of an accident. Any harm caused by a car accident, however, is always undesirable in relations of health, property loss, and other economic issues. It is well known that traffic accidents occur more often in some areas. The examination of these places can aid in finding key road accident issues that contribute to the frequency of road accidents in these areas. Association rule mining is a prominent data mining technique for knowing correlations in various aspects of a traffic accident. We originally used in this paper. Road traffic accidents[5] are among the top sources of death and injury across the world. In AbuDhabi in 2014, 971 road incidents occurred, resulting in 121 fatalities and 135 serious injuries. Several variables, including driver-related factors, road-related factors, and accident-related factors, all contribute to the cruelty of an injury. Based on 5,973 traffic accident records in AbuDhabi from 2008 to 2013, data-mining techniques were to conduct this study to create models (classifiers) to forecast the injury severity of every new accident with fair accuracy. Furthermore, the study attempted to develop a set of regulations that the United Arab Emirates (UAE) Traffic Agencies might utilise to regulate the primary elements that lead to accident severity. WEKA (Waikato Environment for Knowledge Analysis)

Existing Model

The current system study analysed and investigated several techniques to application it and found that traffic accident risk projections generated from previously known to drivers elements such as personal descriptors, vehicle descriptors, and location made a lot of intuitive sense. Once individuals are aware of such high risk variables, they have a certain degree of power to lessen the danger. Because drivers are the ones in charge on the road, having this info can assist individuals in making better travel decisions, minimizing the chance of traffic accidents and saving lives.The current system in use The K-means clustering technique is used. The present method analyses accident geolocation data employing algorithms for clustering, like K-means to categorise them into high risk hotspots in a specific region. Once gathered, clusters can be subjected to a classification algorithm to discover which are the characteristics responsible for raising the risk. These factors might include time of day , weather conditions, road conditions, and so on.

Sensitivity to initialization: K-means clustering is extremely sensitive to centroids' initial location. If the starting centroids are not correctly positioned, the method may converge to a inadequate answer that does not suffice represent the underlying data distribution.K-means clustering has a limited application since it only applies to datasets with a spherical or circular form. It might not be appropriate for datasets with uneven forms or clusters with changing densities.Scalability: K-means clustering is not scalable to huge datasets since processing massive amounts of data demands a substantial amount of memory and computer resources.Cluster size and form bias: K-means clustering assumes isotropic clusters with equal variances. This may not be the case for all datasets, resulting in a skewed cluster size.Outlier sensitivity: K-means clustering is sensitive to outliers and noisy data points. Outliers might cause misleading clusters to develop or the algorithm to converge to poor solutions.Difficulty in establishing the appropriate number of clusters: The best amount of clusters is not always known a priori and may need trial and error. This might take time and may need specialist knowledge or subject competence.

Proposed Methodology

The suggested system is a Traffic Collision Risk Prediction system that predicts the possibility of traffic accidents using a Random Forest Classifier algorithm. The system tries to increase the accuracy of traffic collision risk prediction compared to previous systems, and it also includes a geo-location component, which is significant.

The suggested system would make using a variety of historical traffic accident data, including geo-location, time, location, weather conditions, road conditions, and other pertinent aspects, sourced from the Kaggle repository. The data will be cleaned and transformed into features that capture the data's underlying patterns and trends.The Random Forest Classifier algorithm will be trained on the preprocessed data and will generate predictions using an ensemble of decision trees. The algorithm will be tuned to maximise accuracy while minimising mistake rate. The Random Forest Classifier algorithm's performance will be tested using several metrics such as accuracy, precision, recall, and F1-score. The system will also have a user interface that will let users to enter important information such as location, time, weather conditions, and road conditions to obtain a risk score indicating the potential for a traffic collision occurring. Improved accuracy: The Random Forest Classifier algorithm is very accurate and can accurately forecast the potential for traffic collisions. This can help minimise the amount of traffic collision and save lives.

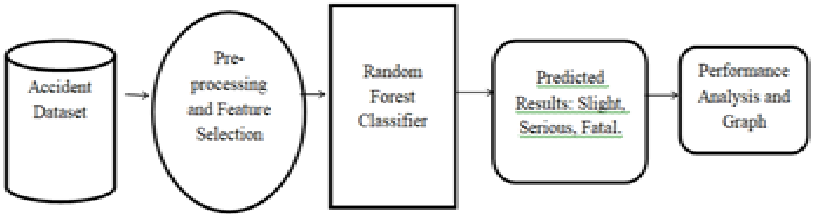

Figure 1 Proposed Architecture

Robustness: Because the Random Forest Classifier method is resistant to noise and outliers in the data, it is less prone to mistakes and can handle a larger variety of data. Non-linear connections: The method can handle non-linear relationships that may exist in the underlying data between characteristics and the output variable. As a result, the system can capture more complicated patterns and enhance its accuracy.

Interpretability: The algorithm gives insights into the most significant contributing factors to traffic accident risk. This information is available. Scalability: Because the Random Forest Classifier method is very scalable, it can handle big datasets with millions of data points. This enables more precise and efficient prediction of traffic accident risk. The system has a user-a user-friendly interface that enables enter important information and obtain a risk score indicating the potential for a traffic accident occurring.

Implementation

Data Collection

Collecting data is the initial real step towards the actual construction of a machine learning model. This is a vital phase that will possess knock-on effect on how successful the model is; the more and better data we have, the better our model will perform.

There are several methods for gathering data, including online scraping and manual interventions. The dataset might be found in the model folder. The dataset is from the well-known dataset repository Kaggle.

Preprocessing of Data

Gather and arrange data for training. Clean up everything that needs it (remove duplicates, rectify mistakes, deal with missing numbers, normalisation, data type conversions, and so on). Randomise data to remove the impacts of the sequence in which we acquired and/or otherwise prepared our data.Visualise data to helping to identify meaningful correlations between variables or class imbalances (bias alert!), or do other exploratory analysis.Sets are classified as follows: training and assessment sets.

Model Selection

- The Random Forest Classifier (ML)Machine learning algorithm was utilized. We applied this method after achieving an accuracy of 99.0% on the training set.
- The Algorithm of Random Forests
- Let's go out the algorithm in layman's words. Assume you want to go on a trip and you want to go somewhere you would love.
- So, how do you go about finding a place you'll like? You may conduct an internet search, read reviews on travel blogs and websites, or ask your friends.
- Assume you chose to question your pals and inquired about their prior trips to various areas. Every buddy will give you some recommendations. You must now construct a list of the recommended locations.

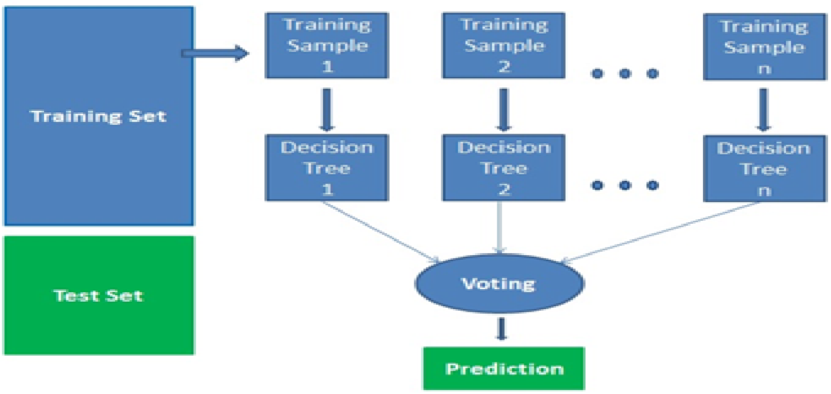

Figure 2 Random Forest Classifier

Feature

A random forest is also an effective feature selection indication. With the model, Scikit-learn includes an additional variable that reflects the relative value or contribution of each attribute to the prediction. During the training phase, it automatically computes the relevance score of each feature. The relevance is then scaled down so that the total of all ratings is 1.This result will help you in selecting the most vital characteristics and eliminating the least important ones for model development. Random forest calculates the relevance utilising the Gini significance for each characteristic or mean reduction in impurity (MDI). Gini significance is frequently referred to as complete decrease in node impurity.

Results

Figure 3 Home Page

The introductory page of a website, typically serving as a table of contents for the site.

Figure 3 Login Page

The login page allows a user to gain access to an application by entering their username and password or by authenticating using a social media login.

Conclusions

Finally, the Traffic Accident Research on Risk Prediction Using Machine Learning has shown a lot of promise for enhancing road safety and lowering the frequency of traffic accidents. The suggested method can reliably forecast the likelihood of traffic collisions, then select the most pertinent risk variables by analysing historical traffic accident data and applying powerful approaches for machine learning such as Random Forest Classifier.

Future work might concentrate on increasing system performance by combining other data sources, such as traffic flow data and driver behaviour data, as well as integrating real-time data to offer up-to-date risk rankings. Furthermore, the system might be improved to deliver more personalised risk rankings based on individual driver profiles, perhaps encouraging safer driving behaviour.

Overall, the Traffic Accident Risk Prediction project has shown the capacity of Machine Learning in enhancing road safety and reducing traffic accidents, ain addition to the relevance of adopting data-driven techniques to address complex transportation concerns.

References

1. Camilo Gutierrez-Osorio and César Pedraza, "A review of contemporary data sources and methods for the analysis and forecasting of traffic accidents." The Journal of Traffic & Transportation Engineering (English Edition), volume 7.4 (2020), pages 432-446.
2. G. Cao, J. Michelini, K. Grigoriadis, B. Ebrahimi, and M. A. Franchek, "Cluster-based severe braking event connection with time andlocation," 2015, pp. 187-192, doi: 10.1109/SYSOSE.2015.7151986.
3. Kumar, S., and D. Toshniwal (2016). Hierarchical the clustering and cophenetic correlation coefficient (CPCC) were used to analyse hourly traffic accident numbers. 1-11 in Journal of Big Data, 3(1).
4. Kumar, S., and D. Toshniwal (2016). A data mining strategy to characterising the sites of traffic accidents. 62-72 in Journal of Modern Transportation.
5. Taamneh, M., S. Alkheder, and S. Taamneh (2017). In the United Arab Emirates, strategies for data mining used to model and forecast traffic accidents. Transportation Safety & Security, 9(2), pp. 146-166.
6. Tiwari, P., Dao, H., and G. N. Nguyen (2017). On traffic accident analysis, the performance of slow, decision tree classifier, and multilayer perceptron is evaluated. 41(1), Informatica.
7. Ait-Mlouk, A., and T. Agouti (2019). A case study on a road accident using DM-MCDA, a web-based tool for data mining and multiple criteria decision analysis. SoftwareX, vol. 10, no. 100323.