

Road Accident Analysis

Dr. Anitha Patil^a, Prithvish Kumble^b, Naresh K^c, Srihari^d

^aHOD, Department of Computer Science and engineering, Nagarjuna College of Engineering and Technology, Bangalore, India

^{b,c,d}B.E. Students, Department of Computer Science and Engineering, Nagarjuna College of Engineering and Technology, Bangalore, India

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: India being a highly populated country, the number of people using vehicles for commuting everyday is also high. With this there are a lot of accidents that takes place everyday. These accidents tend to effect a family very adversely with risk of lives or to endure the expense after it happens. This paper is to try and figure out ways to find accident prone regions and warn daily commuters regarding the accidents happening in that particular area. Accidents are something that happen without any prior intimation, but if we being a user of this interface we can be more careful when it comes to areas where accidents happen majorly. The user interface will notify a user regarding the accident prone regions being high medium and low. We will be using different algorithms in Machine learning to process the data to train and test a model. We are using data from a couple of years regarding accidents as datasets to divide them into these specific areas. This paper of ours follows the saying "It's better safe than sorry", and we see to help the users to avoid any accident that they might face

Keywords: Accidents, Machine learning, data set, training and testing.

1. Introduction

Road accidents are unquestionably the most frequent cause of damage. It's one of the most significant causes of the fatalities. The reasons for this are the extremely dense road traffic and the relatively great freedom of movement given to drivers. Accidents that involve heavy

goods vehicles (like Lorries, trucks) and even the commercial vehicles with the public transportation like buses are one of the most fatal kinds of accidents that occur, claiming the lives of innocent people. Highways are always a soft spot for these accidents with injuries and deaths. Various weather conditions like rain, fog etc play a role in catalyzing the risk of accidents. Having a proper estimate of accidents and knowing the hotspots of accidents and its factors will help to reduce them. Providing timely emergency support even when the casualties have occurred is needed, and to do that a keen study on accidents is required.

Models are created using accident data records which can help to understand the characteristics of many features like drivers behavior, roadway conditions, light condition, weather conditions and so on. This can help the users to compute the safety measures which is useful to avoid accidents. It can be illustrated how statistical method based on directed graphs, by comparing two scenarios based on out-of-sample forecasts. the model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injury that can be used to perform a risk factor and reduce it. Here the road accident study is done by analyzing some data by giving some queries which is relevant to the study. The queries like what fractions of accidents occur in rural, urban and other areas, What is the trend in the number of accidents that occur each year, do accidents in high speed limit areas have more casualties and so on ... These data can be accessed using Microsoft excel sheet and the required answer can be obtained. This analysis aims to highlight the data of the most importance in a road traffic accident and allow predictions to be made.

2. Literature Survey

Tessa K. Anderson et al. [1] proposed a method of identifying high-density accident hotspots, which creates a clustering technique that determines that stochastic indices are more likely to exist in some clusters, and can therefore be compared in time and space. Sachin Kumar et al. [2], used data mining techniques to identify the locations where high frequency accidents are occurred and then analyze them to identify the factors that have an effect on road accidents at that locations. The first task is to divide the accident location into k groups using the k-means clustering algorithm based on road accident frequency counts. S. Shanthi et al.

[3] proposed data mining classification technology based on gender classification, in which RndTree and C4.S use AdaBoost Meta classifier to provide high-precision results. From the Critical Analysis Reporting Environment (CARE) system provided by the Fatal Analysis Reporting System (FARS) used by the training data set.

Principal Component Analysis of Fatal Traffic Accidents Based on Vehicle Condition Factors

In order to find out several pivotal influencing factors of fatal traffic accidents, the number of fatal injuries recorded in the FARS database of the National Highway Traffic Safety Administration of the United States from 2010 to 2016 was calculated. The principal component analysis (PCA) method of multivariate statistical analysis is used to analyze the traffic conditions, and several pivotal influencing factors of fatal traffic accidents are obtained. The results show that tire wear, rim damage, exhaust system failure and coupling failure are the most important factors.

Vehicle Accident Prevent cum Location Monitoring System

Rate of road accidents is increasing day by day. Fatal road accidents can be easily avoided by understanding the psychological state of a driver. Majority of road accidents occur during night driving due to the drowsiness state of a vehicle driver. The paper provides mechanism to reduce accidents to a large extent by monitoring eye blinking of the driver which indicates the drowsiness, obstacles located in the road and the drunken state of the drivers. Automatic pre- cautionary system is activated based on the above alarming condition. Accident and its probable location are also generated at the nearby police station that helps initiating medical help. In normal cases no medical help is received due to the non- availability of accident information. This happens mainly at night and in roads where the traffic is low.

3. System Architecture

Models are created using accident data records which can help to understand the characteristics of many features like drivers behavior, roadway conditions, light condition, weather conditions and so on. This can help the users to compute the safety measures which is useful to avoid accidents.

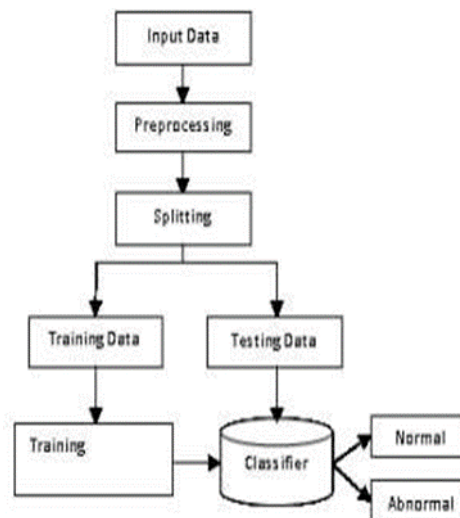


Fig 1: System Block Diagram

It can be illustrated how statistical method based on directed graphs, by comparing two scenarios based on out-of-sample forecasts. the model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injury that can be used to perform a risk factor and reduce it.

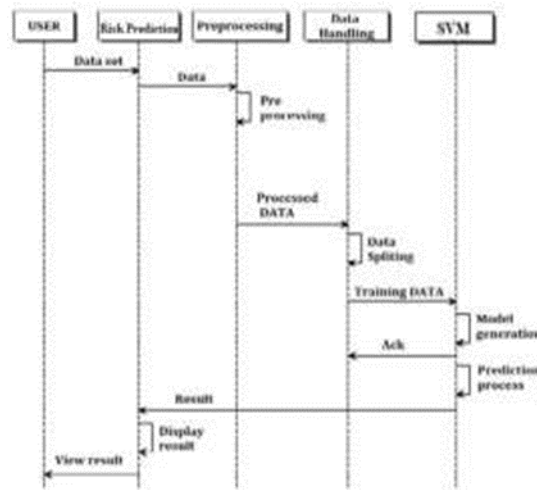


Fig 2: Sequence Diagram of System Here the road accident study is done

by analysing some data by giving some queries which is relevant to the study. The queries like what is the most dangerous time to drive , what fractions of accidents occur in rural, urban and other areas. What is the trend in the number of accidents that occur each year, do accidents in high speed limit areas have more casualties and so on. These data can be accessed using Microsoft excel sheet and the required answer can be obtained. This analysis aims to highlight the data of the most importance in a road traffic accident and allow predictions to be made.

The working of the project is divided in to parts.

Data Set Selection

Data is the most import part when you work on prediction systems. It plays a very vital role your whole project i.e., you

system depends on that data. So selection of data is the first and the critical step which should be performed properly, For our project we got the data from the government website. These datasets were available for all. There are other tons of websites who provide such data. The dataset we choose wad selected based on the various factors and constraints we were going to take under the consideration for our prediction system.

Data Cleaning and Data Transformation

After we have selected the dataset. The next step is to clean the data and transform it into the desired format as it is possible the dataset we use may be of different format. It is also possible that we may use multiple datasets from different sources which may be in different file formats. So to use them we need to convert them into the format we want to or the type that type prediction system supports. The reason behind this step is that it is possible that the data set contains the constraints which are not needed by the prediction system and including them makes the system complicated and may extend the processing time. Another reason behind data cleaning is the dataset may contain null value and garbage values too. So the solution to this issue is when the data is transformed the garbage values are replaced. There are many methods to perform that.

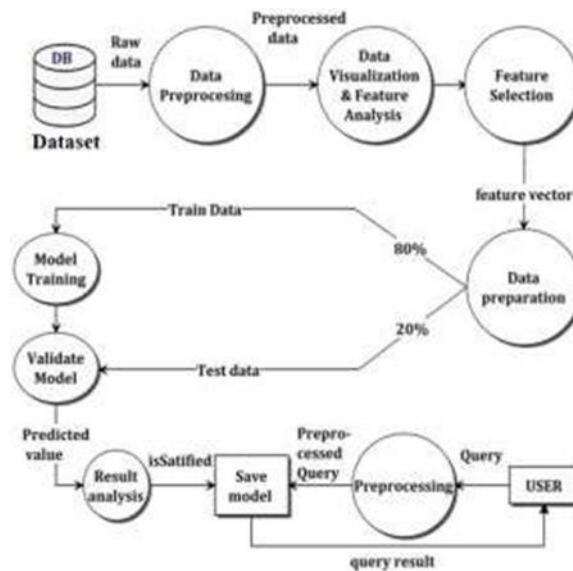


Fig 3: Data Flow of System

Data Processing and Algorithm Implementation

After the data is been cleaned and transformed it's ready to process further. After the data has been cleaned and we have taken the required constraints. We divide the whole dataset into the two parts that can be either 70-30 or 80-20. The larger portion of the data is for the processing. The algorithm is applied on that part of data. Which helps the algorithm to learn on its own and make prediction for the future data or the unknown data. The algorithm is executed in which we take only the required constraints from the cleaned data. The output of the algorithm is in 'yes' and 'no'. It gives the error rate and the success rate.

Output and User Side Experience

After the prediction system is ready to use. The Website is developed for the user. The user just has to fill a form which consists of different options they need to select. They are like the type of climate, the type of vehicle and so on. Once the user submits the form the algorithm is triggered and the input given by the user is passed to the prediction system. The user is given how accident prone the road can be in percentage.

4. Software And Languages Used

Jupyter

Jupyter exists to develop open- source software. It is used for open-standards, and services for interactive computing across dozens of programming languages. It is an opensource web application that allows you to create and share documents and code live. Which is a very big advantage of Jupyter. It can be used for data cleaning and transformation, numerical simulation, statistical modelling, machine learning and

much more. We used Jupyter to run the algorithm. The Jupyter Notebook is an interactive computing environment that enables users to author notebook documents that include: Live code, Interactive widgets, Plots, Narrative text, Equations, Images, Video. These documents provide a complete and self-contained record of a computation that can be converted to various formats and shared with others using email, Dropbox,

The Jupyter Notebook combines three components: The notebook web application: An interactive web application for writing and running code interactively and authoring notebook documents. Kernels: Separate processes started by the notebook web application that runs users' code in a given language and returns output back to the notebook web application. The kernel also handles things like computations for interactive widgets, tab completion and introspection. Notebook documents: Self-contained documents that contain a representation of all content visible in the notebook web application, including inputs and outputs of the computations, narrative text, equations, images, and rich media representations of objects. Each notebook document has its own kernel.

Python

Python is an interpreter, high-level and a general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It is the most used programming language presently. It provides constructs that

enable clear programming on both small and large scales. The logistic regression in the system is implemented in jupyter and the algorithm is written in python language.

SVM: A New Generation of Learning Algorithm

SVM is a supervised machine learning algorithm which can be used for classification or

regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs. Simply put, it does some extremely complex data transformations, then figures out how to separate your data based on the labels or outputs you've defined.

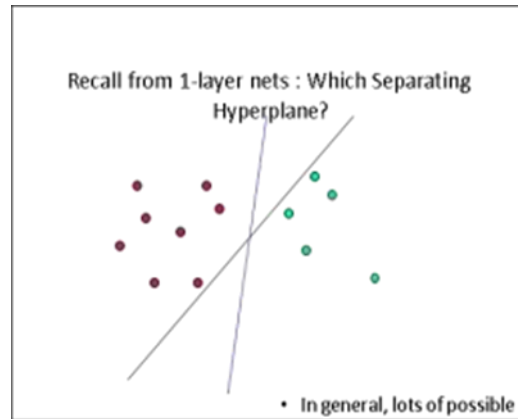


Fig 4: Support Vector Machine (SVM)

In particular I'll be focusing on non-linear SVM, or SVM using a non-linear kernel. Non-linear SVM means that the boundary that the algorithm calculates doesn't have to be a straight line. The benefit is that you can capture much more complex relationships between your datapoints without having to perform difficult transformations on your own. The downside is that the training time is much longer as it's much more computationally intensive.

5. Conclusion and Future Scope

Road Accidents are caused by various factors. By going through all the research papers it can be concluded that Road Accident cases are hugely affected by the factors such as types of vehicles, age of the driver, age of the vehicle, weather condition, road structure and so on. Thus we have build an application which gives efficient

prediction of road accidents based on the above mentioned factors. There are many reasons for choosing Artificial neural network including: The ability to extract information from the data which are incomplete and noisy. Acquire experience and knowledge through self-training and organization of the knowledge. The potential for very fast optimization. Their suitability for problems in which algorithmic solutions are difficult to develop or do not exist.

References

1. Akgungor AP, Dogan E. "Estimating road accidents of Turkey based on regression analysis and artificial neural network approach". *Advances in Transportation Studies International Journal*.
2. Chang L. Analysis of freeway accident frequencies: "Negative binomial regression versus artificial neural". *Safety Science*. 43(8): 541-557.
3. Kalyoncuoglu SF, Tigdemir M. "An alternative approach for modelling and simulation of traffic data: artificial neural networks". *Simul Model Pract Theory*. 12(5):351-362.
4. Khair S. Jadaan, Muaath Al-Fayyad, and Hala
5. F. Gammoh, (2014) "Prediction of Road Traffic Accidents in Jordan using Artificial Neural Network (ANN)", *Journal of Traffic and Logistics Engineering* Vol. 2, No. 2.
6. Galal A. Ali and Awadalla Tayfour ,(2012) , "Characteristics and Prediction of Traffic Accident Casualties In Sudan Using Statistical Modeling and Artificial Neural Networks", *International Journal of Transportation Science and Technology* • vol. 1 • no. 4 • pages 305 – 317.