



A ROAD ACCIDENT PREDICTION MODEL USING DATA MINING TECHNIQUES

Dhanush G¹, Vignesh D²

Mr. Senthil K.P, Dept of CSE, Paavai Engineering College, Pachal, Namakkal ¹.

Associate Professor, Dept of Computer Science, Paavai Engineering College, 637018²

ABSTRACT - In the project, the machine learning based road accident based prediction system has been implemented. The analysis of the road accident is done by using various kind of visualization tools or packages called plotly, seaborn and matplotlib. K-means clustering is an unsupervised machine learning techniques which is used to cluster the road accident points the high number of traffic incidents and deaths these days, the ability to forecast the number of traffic accidents over a given time is important for the transportation department to make scientific decisions. In this scenario, it will be good to analyze the occurrence of accidents so that this can be further used to help us in coming up with techniques to reduce them. Even though uncertainty is a characteristic trait of majority of the accidents, over a period of time, there is a level of regularity that is perceived on observing the accidents occurring in a particular area. This regularity can be made use of in making well informed predictions on accident occurrences in an area and developing accident prediction models. In this paper, we have studied the inter relationships between road accidents, condition of a road and the role of environmental factors in the occurrence of an accident. We have made use of data mining techniques in developing an accident prediction model using Apriori algorithm and Support Vector Machines.

Keywords -- Data Mining, Machine Learning, Exploratory Data Analysis, Apriori Algorithm, Support Vector Machines

1 INTRODUCTION

According to some recent statistics, India accounts for roughly six percent of global road accidents while owning only one percent of the global vehicle population. There are a lot of accident cases reported due to the negligence of two-wheelers, whereas over-speeding is also another contributing factor. Accidents caused while under the influence of alcohol or during general traffic violations are also common. In spite of having set regulations and the highway

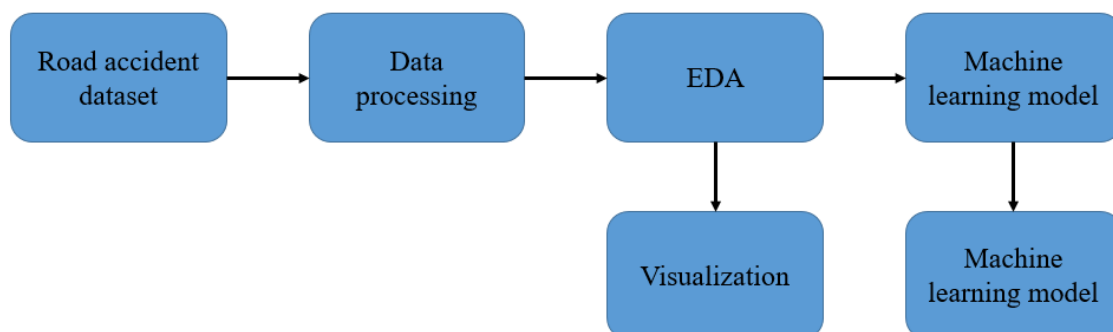


codes, the negligence of people towards the speed of the vehicle, the vehicle condition and their own negligence of not wearing helmets has caused a lot of accidents.

The objective of this paper is to present an extensive literature review and a questionnaire survey among researchers worldwide, both conducted to define the current state of the art, the current practices for road safety assessment within National Road Authorities (NRAs) and in research organizations as well as to identify the data availability for developing and calibrating reliable accident prediction models. The paper continues with an analysis and discussion of relevant findings and concludes by providing useful remarks on current situation in accident prediction and highlighting next steps of the PRACT project and of pertinent research in general.

1) OVERVIEW OF EXISTING SYSTEM

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.



Road Accident Dataset : The purpose is to find the possible causes of accident related to driver vehicle, and roadway. Accident analyses are made to develop information such as:

- 1) Driver and Pedestrian - Accident occurrence by age groups and relationships of accidents to physical capacities and to psychological test results.
- 2) Vehicle - Accident occurrence related to characteristic of vehicle, severity, location and extent of damage related to vehicles.



Data Processing : Data processing refers to the entire process of collecting, transforming (i.e. cleaning, or putting the data into a usable state), and classifying data. Raw data is the data collected from various sources, in its original state. It is usually not in the most proper format for data analysis or modeling.

EDA: Exploratory data analysis popularly known as EDA is a process of performing some initial investigations on the dataset to discover the structure and the content of the given dataset. It is often known as *Data Profiling*. It is an unavoidable step in the entire journey of data analysis right from the business understanding part to the deployment of the models created. EDA is where we get the basic understanding of the data in hand which then helps us in the further process of Data Cleaning & Data Preparation.

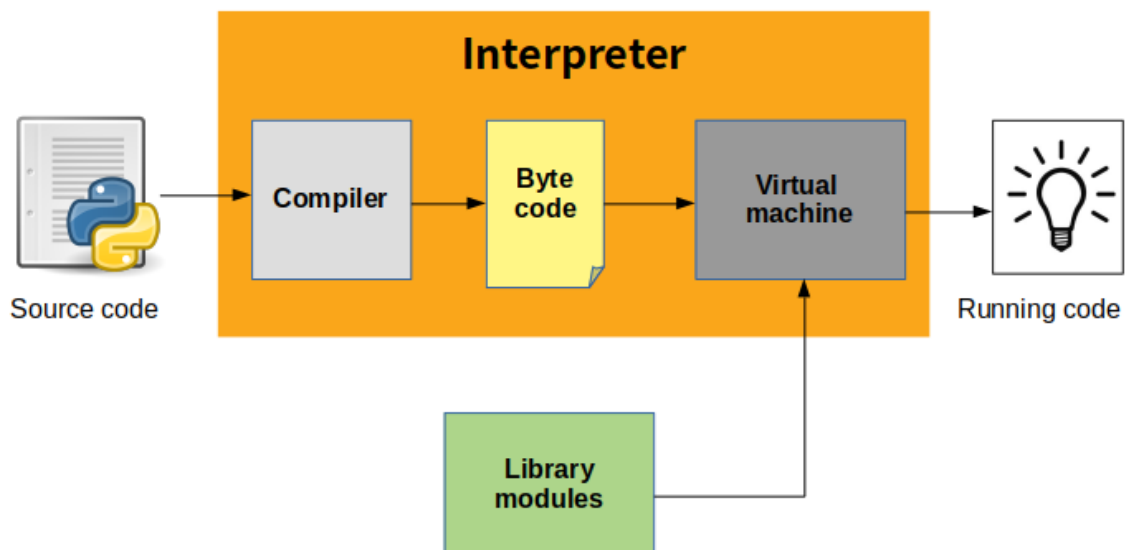
Machine Learning Algorithm : Machine learning (ML) algorithms are broadly categorized as either supervised or unsupervised. Supervised learning algorithms have both input data and desired output data provided for them through labeling, while unsupervised algorithms work with data that is neither classified nor labeled.

Visualization : Algorithm visualization illustrates how algorithms work in a graphical way. It mainly aims to simplify and deepen the understanding of algorithms operation. Within the paper we discuss the possibility of enriching the standard methods of teaching algorithms, with the algorithm visualizations.

2.1 PROPOSED WORKFLOW

Python's developers strive to avoid premature optimization, and reject patches to non-critical parts of the Python reference implementation that would offer marginal increases in speed at the cost of clarity. When speed is important, a Python programmer can move time-critical functions to extension modules written in languages such as C, or use Python, a just-in-time compiler. Python is also available, which translates a Python script into C and makes direct C-level API calls into the Python interpreter.

An important goal of Python's developers is keeping it fun to use. This is reflected in the language's name a tribute to the British comedy group Monty Python and in occasionally playful approaches to tutorials and reference materials, such as examples that refer to spam and eggs (from a famous Monty Python sketch) instead of the standard foo and bar.



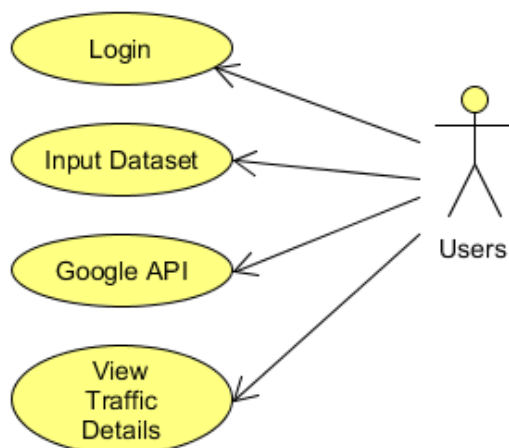
Python uses duck typing and has typed objects but untyped variable names. Type constraints are not checked at compile time; rather, operations on an object may fail, signifying that the given object is not of a suitable type. Despite being dynamically typed, Python is strongly typed, forbidding operations that are not well-defined (for example, adding a number to a string) rather than silently attempting to make sense of them.

2.2 USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



CLASS DIAGRAM



3 SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

The dataset for the project is collected from the Kaggle Bangalore road accident dataset. To implement a well-designed road framework management system for looking into road security aspects, it is often desired to have an optimized accident prediction model which can analyze potential issues arising due to infrastructure fallbacks and to estimate the effect of existing models in reducing the occurrence of accidents.

4 PRODUCTIVITY AND SPEED

It is a widespread theory within development circles that developing Python applications is approximately up to 10 times faster than developing the same application in Java or C/C++. The impressive benefit in terms of time saving can be explained by the clean object-oriented design, enhanced process control capabilities, and strong integration and text processing capacities. Moreover, its own unit testing framework contributes substantially to its speed and productivity.



PYTHON IS POPULAR FOR WEB APPS

Web development shows no signs of slowing down, so technologies for rapid and productive web development still prevail within the market. Along with JavaScript and Ruby, Python, with its most popular web framework Django, has great support for building web apps and is rather popular within the web development community.

OPEN-SOURCE AND FRIENDLY COMMUNITY

As stated on the official website, it is developed under an OSI-approved open source license, making it freely usable and distributable. Additionally, the development is driven by the community, actively participating and organizing conference, meet-ups, hackathons, etc. fostering friendliness and knowledge-sharing.

PYTHON IS QUICK TO LEARN

It is said that the language is relatively simple so you can get pretty quick results without actually wasting too much time on constant improvements and digging into the complex engineering insights of the technology. Even though Python programmers are really in high demand these days, its friendliness and attractiveness only help to increase number of those eager to master this programming language.

5 IMPLEMENTATIONS

5.1 REFERENCE IMPLEMENTATION

Python is the reference implementation of Python. It is written in C, meeting the C89 standard with several select C99 features. It compiles Python programs into an intermediate bytecode which is then executed by its virtual machine. CPython is distributed with a large standard library written in a mixture of C and native Python. It is available for many platforms, including Windows and most modern Unix-like systems. Platform portability was one of its earliest priorities.

5.2 OTHER IMPLEMENTATION

Python is a fast, compliant interpreter of Python 2.7 and 3.10. Its just-in-time compiler brings a significant speed improvement over CPython but several libraries written in C cannot be used with it.



Stackless Python is a significant fork of CPython that implements microthreads; it does not use the C memory stack, thus allowing massively concurrent programs. PyPy also has a stackless version. MicroPython and CircuitPython are Python 3 variants optimized for microcontrollers. This includes Lego Mindstorms EV3.

6 RESULT

6.1 ENTER THE DETAILS ABOUT MODEL

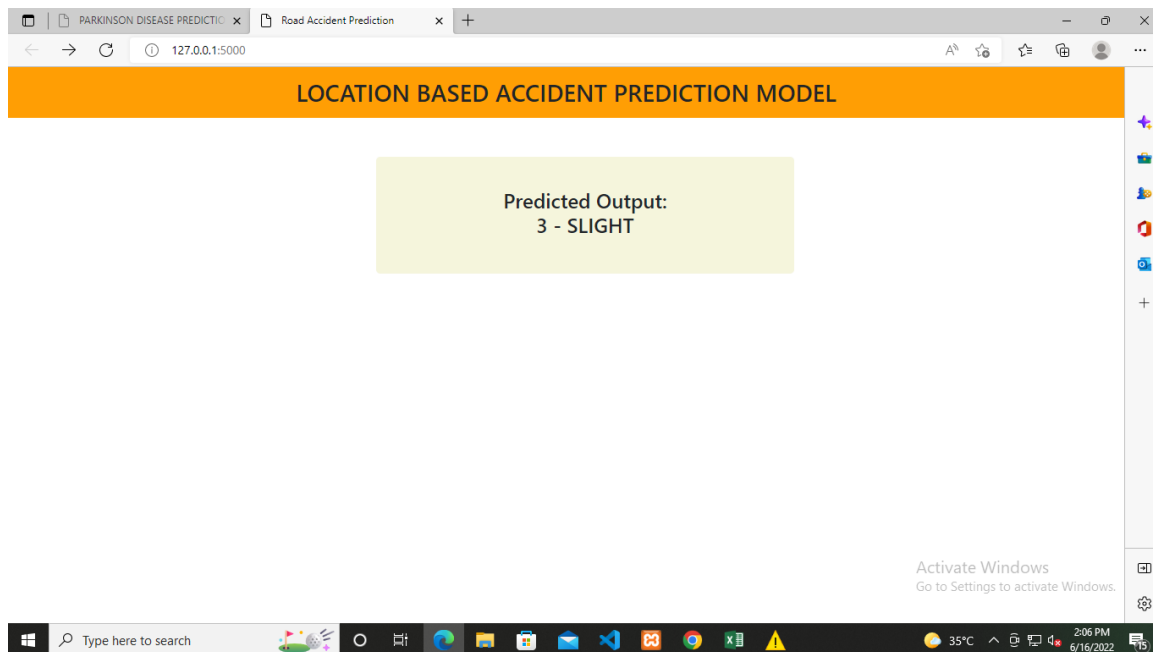
The screenshot shows a web browser window with the title "Road Accident Prediction". The address bar shows "127.0.0.1:5000". The page has an orange header with the text "LOCATION BASED ACCIDENT PREDICTION MODEL". Below the header is a light orange form with the following fields:

- Age of Driver :
- Vehicle Type :
- Age of Vehicle :
- Sex of Driver :
- Speed limit :

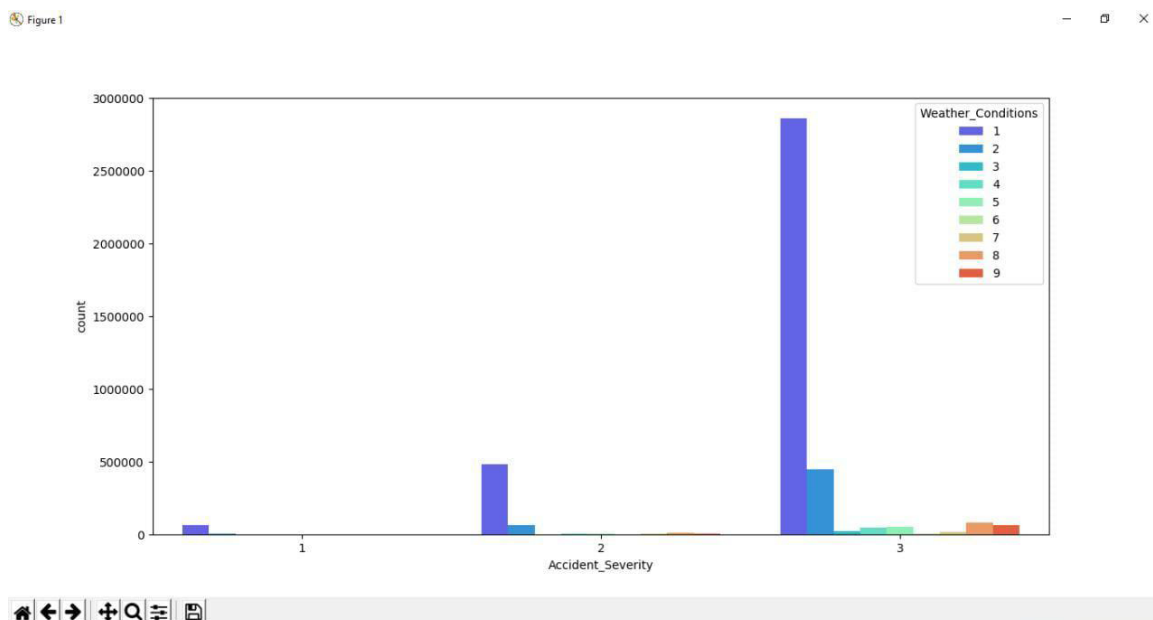
A yellow "Submit" button is located at the bottom right of the form. The Windows taskbar is visible at the bottom of the screen.



6.2 NO OF ACCIDENT IN LOCATION



6.3 ACCIDENT SEVERTY





7 CONCLUSION

An accident can change the lives of many people. It is up to each of us to bring down this increasing number. This can be made possible by adopting safe driving measures to an extent. Since all instances of accidents cannot be attributed to the same cause, proper precautionary measures will also need to be exercised by the road development authorities in designing the structure of roads as well as by the automobile industries in creating better fatality reducing vehicle models.

8 FUTURE WORK

The model can be further optimized in future to include several constraints that have been left out in the current study. These optimized models can be efficiently utilized by the government to reduce road accidents and to implement policies for road safety. Another scope of this work would be to develop a mobile app that will help the drivers in choosing a route for a ride. A call out to the driver through the maps service can also be implemented that would also announce the risk probability in a chosen route along with the directions. This can then be implemented by service provider companies such as Uber, Ola and so on in future. This will also be useful in having a better surveillance of accident prone areas and providing emergency services in the event of an accident. Better road safety instructions can also be installed along the highways taking into account the risks obtained from this model.

9 REFERENCE

- [1] Suganya, E. and S. Vijayarani. "Analysis of road accidents in India using data mining classification algorithms." 2017 International Conference on Inventive Computing and Informatics (ICICI) (2017): 1122-1126.
- [2] 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.
- [3] 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.



- [4] 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>.
- [5] 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.