

MACHINE LEARNING APPLICATION FOR ROAD ACCIDENT SEVERITY PREDICTION

ABSTRACT

Road accidents are one of the most regrettable hazards in this hectic world. Road accidents lead to numerous casualties, injuries, and fatalities each year, as well as significant economic losses. There are many factors that contribute to road accidents, especially those related to the environment, vehicles and the travelers. Despite many precautionary measures and safety measures to reduce road accidents it remains one of the uncontrollable. By analyzing the severity of the road accidents that happened in the past, and the factors that caused it, it is possible to take precautionary measures to reduce road accidents rate significantly in the future. Machine learning plays a vital role in analyzing the data and makes future predictions based on the existing data. Data mining using classification algorithms can be utilized to develop a predictive model for predicting the severity of the road traffic accidents.

This project implements developing a machine learning model which can classify the severity of the accident, based on the accident influential factors. Various machine learning classifiers such as Decision Tree (DT), K Nearest Neighbor (KNN), Random Forest (RFF) and Gradient Boosting Classifiers (GBC) used to develop a predictive model. However, the results show that a Gradient boosting algorithm is capable of 90% accurately detecting the road accident severity. The results of the predictive model suggest the Gradient boosting classifier could be a useful tool for forecasting accident hotspots. Further, the model used in the web portal to develop an intelligent system for the accident severity prediction.

Keyword: Accident Severity, Classification, Ensemble, Data Analysis, Machine Learning.

Viva voce questions and answers

1. Explain the objective of the project?

The main objective of the project is to classify the road accidents into slight/ serious based

on the casualty and environment relevant factors. Further we have implemented the trained model into the web portal so that any technical/ non technical end users can predict the accident severity.

2. Why did you chosen this domain?

As a young engineers we whould like to solve a real world problems by using the technology.

We found that road accidents are one of the most regrettable hazards in this hectic world. Road accidents lead to numerous casualties, injuries, and fatalities each year, as well as significant economic losses. There are many factors that contribute to road accidents, especially those related to the environment, vehicles and the travelers. Despite many precautionary measures and safety measures to reduce road accidents it remains one of the uncontrollable. Hence, we have chosen the accident domain to give some solution using the technology

3. Why did you considered Machine Learning for this project ?

Machine learning is of the rapidly growing technology over the IT domain. Machine can perform the task more accurate than manual predictions. If a man predict something, definitely they will perorm the task based on the data in their mind. But while the machine learning the past numerous data, it will get more experience than human. As well as there is a lot of usefull data assests are storing on every single day. We would like to turn all the stored into the usefull assets to provide some solutions. Than we found that the machine learning is the best domain for what we actually thought. Thats why we have chosen ML for this project

4. Explain Machine Learning

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed

5. Difference between Programming problem solving & machine learning

Program : $\text{program} + \text{input} = \text{output}$

ML : $\text{input} + \text{output} = \text{program (Learning)}$

6. Types of ML Algorithms

a) Supervised Learning

Learning by the input as well as target (output) together

b) Un supervised Learning

Learning with inputs only, output will not present

c) Re-inforcement learning

Learn from the experience

7. Your project under which type

Supervised machine learning, because we input both the features as well the class(output)

8. Which algorithms used

KNN

Decision Tree

Random Forest

Gradient boosting classifier (ensemble)

9. What is ensemble

Ensemble learning is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem.

10. Dataset and features

(Please Refer Descriptive analysis file)

11. Pre – process performed

Categorizing or grouping the data

Label Encoding

12. Is there any null values present in your data

No

13. If any null values available in the dataset how will you fill the NA values

By using the Mean and median of the column values

14. Explain DT (search : Decision tree analytics vidhya)

15. Explain KNN (search : KNN analytics vidhya)

16. Explain GBC (search : GBC analytics vidhya)

17. Why GBC given best accuracy

Because GBC is a ensembling technique. If a machine leaning algorithm can given better accuracy alone, than the group of algorithms can given very best accuracy

18. Tech stacks used in this project

Front end

HTML

CSS

Javascript

Backend

Python – Flask

Machine Learning

Python - Pandas

Python - Numpy

Python – Scikit Learn

Database

MySQL

19. Future works on this project

This work can be extended by using various machine learning algorithms and with different input features inorder to predict the accident severity

20. Realtime Application

- 1) Reporters
- 2) Journals
- 3) Highway authorities can be use this idea for

Make a preventions, countermeasures for the road accidents.

21. Your contribution on this project

- 1) Most of the research paper suggested KNN, and DT for the road accident data. We have implemented using GBC and we proved GBS can give best accuracy for road accident data.
- 2) We have implemented the model in a web portal hence any end users can utilize
- 3) Strong user authentication provided for the security purpose
- 4) Every new predictions are stored into a database hence these new data can be utilized in the future
- 5) User can download the reports as well as they can share it with a mail
- 6) User can view the analytics of the road accident in the different web pages

--