ROAD SAFETY- Driver drowsiness, Road accidents using machine learning

by Ananya Adiga

Submission date: 15-Nov-2022 03:35PM (UTC+0530)
Submission ID: 1954653328
File name: TY-_Driver_drowsiness,_Road_accidents_using_machine_learning.pdf (301.16K)
Word count: 1259
Character count: 6859

ROAD SAFETY- Driver drowsiness, Road accidents using machine learning

15 November 2022 08:48

(I) *Authors and affiliations :*

Ananya Adiga *Btech, CSE student PES University* Bangalore, India

Amisha Mathew, *Btech, CSE , Student PES University* Bangalore, India

Amrita Patra, *Btech, CSE, Student* PES University Bangalore, India

Arindham Chaudhari, Professor *PES University* Bangalore, India

(II) ABSTRACT:

This document includes a detailed analysis of how to predict the conditions during which maximum road accidents occurs. The project consists of two parts- Driver drowsiness prediction and road accidents. The first part uses image classification to detect driver's drowsiness using a dataset consisting of images. The second part predicts the severity of an accident based on certain conditions such as weather or light conditions.

KEYWORDS:

Machine Learning , data leaning, label binalizer, NumPy, pandas, seaborn , matplotlib, logistic regression, TensorFlow, keras, label encoder, accident severity, pandasql, driver drowsiness.

(III) INTRODUCTION

According to research approximately 1.3 million lives are cut in road accidents. In India, traffic accidents are a major cause of annual fatalities, injuries, and property damage. According to the National Crime Records Bureau (NCRB) 2021 report, there were 155,622 fatalities, which is the highest number since 2014, and 69,240 of the deaths were related to two-wheelers. The reason being drink and drive, driver feeling drowsy, darkness during night, speeding and slippery roads during rain. The illegal overloading of vehicles and excessive driving hours provide a severe risk to other road users in the unorganized trucking sector. Commercial Vehicles (CV) or trucks transport the majority of the freight traffic along these roadways, which raises the risk of their being involved in a traffic accident. The economy of the nation is expected to prosper in the years to come, so the need for CVs is in line with international classification in the supply chain and will be essential. 13,532 traffic fatalities in 2019 were linked to CV occupants.

To overcome this, alerting the driver when the sensor detects the driver to be falling drowsy might prevent a serious accident from occurring. There is lot of data about road accidents recorded but kept unused therefore analyzing the conditions at which the the severity of accidents is high, will help traffic policemen or drivers to be aware and enforce measures to prevent accidents during those peak hours when accidents are prone to occur.

(IV) LITERATURE SURVEY

The literature survey for this project was done to provide users a better opportunity of this cross -platform application and native sort of application.

The papers links provided in our references contains information such as: -

According to the results cited in [8], Logistic regression is one of the best algorithms for predicting severity with an accuracy of 87%.

In paper [3] the researchers achieved an accuracy of 96% using neural networks

(V) METHODOLOGY

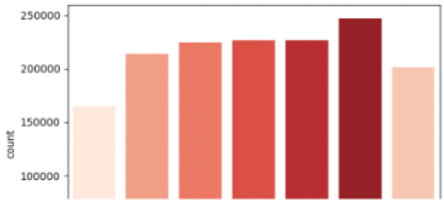
There are two datasets both obtained from Kaggle. The dataset for driver drowsiness consists of 4 training sets each for closed eyes, open eyes, yawn and not yawn. The dataset for accident severity consists of 1504150 instances and 33 attributes. In the driver drowsiness there are four target variables- closed eyes, open, yawn and not yawn. In the road accidents dataset, the target variable is accident severity. The attributes of road accidents include- accident index, locating easting, location northing, longitude, latitude, police force, accident severity, number of vehicles, number of causalities, pedestrian crossing, light conditions, weather conditions, road surface conditions, special conditions at site, carriageway hazards, urban or rural area, presence of police, LSOA, year.

We use multiclass classification with logistic regression to predict accident severity based on various conditions. We use the cross-entropy loss as the loss function and cross validation to check overfitting or underfitting of the model. Logistic regression is a binary classifier that is used for classification and predictive analysis. It estimates the probability of an event occurring or not. As part of the pre-processing we drop the columns which do not contribute much to the training process due to a lot of null values. We construct a correlation matrix to find the linear relationship between each attribute. Since speeding is also another reason for maximum road accidents, we consider speed limit also as an attribute by using pandasql to find the speed limit of the vehicles.

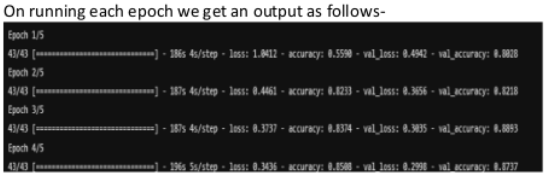
We read the images for driver drowsiness and convert them to an array to make pre-processing easy. We use data augmentation on image claasification to improve the robustness. Label binarizer is used to label each target variable. 0 for yawn, 1- for no yawn, 2- for closed eye and 3- for open eye. A sequential model of keras is used to train the model with activation function set to 'relu', which will output the input to the matrix if positive else puts zero. The evaluation metrics are calculated and the model is saved

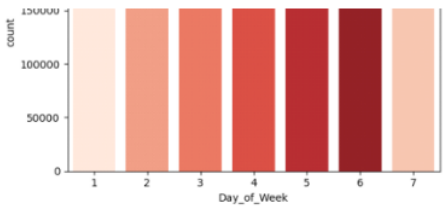
(VI) IMPLEMENTATION-

In road safety, as a part of pre-processing we first drop the columns which are logically unnecesary and then fill the columns which have a null values with an average value. Next we count the number of casualities on each day of a week and plot them.



In driver drowsiness, the images are read using the cv2 library and are converted into an array to make pre-processing easy. Each target variable are assigned a category index, with the help of sequential model, if the image exists then the index of the category it belongs to is returned to the array else an empty matrix is returned.
On running each epoch we get an output as follows-





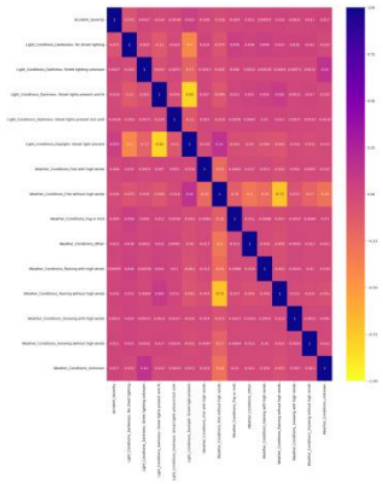
We calculate the maximum and minimum speed limits based on the days of week by generating a query using pandasql. Pandasql allows us to query on a pandas dataframe using SQL syntax.

	Day_of_Week	Casualties_Count	Max_Speed	Min_Speed
0	1	239532	70	10
1	2	284043	70	10
2	3	294476	70	10
3	4	297756	70	20
4	5	299044	70	10
5	6	331934	70	10
6	7	285261	70	10

We one-hot encode the categorial variables as models can only work only on numerical data.

	Accident_Severity	Light_Conditions_Darkness: No street lighting	Light_Conditions_Darkness: Street lighting unknown	Light_Conditions_Darkness: Street lights present and lit	Light_Conditions_Darkness: Street lights present but unlit
1422658	3	1	0	0	0
33042	3	1	0	0	0

A correlation matrix is plotted between 15 variables to check their linear relationship with each other.



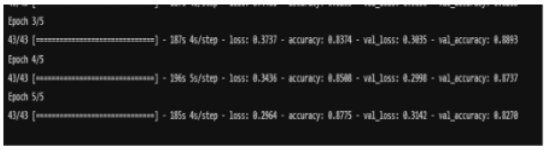
On implementing multi-class classification using logistic regression and evaluating the metrics we get an accuracy of 85.1%

(VII) CONCLUSION

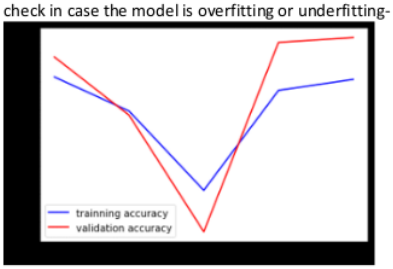
This paperwork aims at helping drivers prevent accidents, increase awareness and not drive at inappropriate speed and rash drive especially during risky hours. With the analysis done in this paperwork , it is safe to conclude that Neural networks is the best technique which can be used compared to other techniques. As part of future work this can be extended to implementation on hardware using sensors.

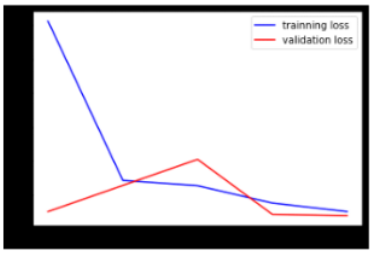
(VII) REFERENCES

List of papers we referred for our project are:-



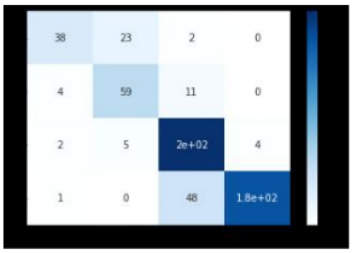
We evaluate the training and validation accuracy as well as the loss to check in case the model is overfitting or underfitting-





After implementing the keras module of tensorflow to train the model we get the metrics, confusion matrix and accuracy as follows-

```
[[ 38  23   2   0]
 [  4  59  11   0]
 [  2   5 204   4]
 [  1   0  48 177]]
0.8269896193771626
```



The above image is a graphical representation of the confusion matrix which can be read as , for example- 38 is the number of images which are yawn and classified correctly as yawn.

The final classification report is evaluated which gives the value of every evaluation metrics-

	precision	recall	f1-score	support
yawn	0.84	0.60	0.70	63
no_yawn	0.68	0.80	0.73	74
Closed	0.77	0.95	0.85	215
Open	0.98	0.78	0.87	226
accuracy			0.83	578
macro avg	0.82	0.78	0.79	578
weighted avg	0.85	0.83	0.83	578

[1] https://www.researchgate.net/publication/347769816_Machine_learning_applied_to_road_safety_modeling_A_systematic_literature_review

[2] https://www.frontiersin.org/articles/10.3389/fbuil.2022.860805/full

[3] https://link.springer.com/chapter/10.1007/978-981-16-2008-9_7

[4] https://link.springer.com/chapter/10.1007/978-3-030-20521-8_24

[5] https://ieeexplore.ieee.org/abstract/document/9039851

[6] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3884784

[7] https://www.sciencedirect.com/science/article/pii/S0022437512000266?casa_token=TjTut00y5wgAAAAA:CUBbPK_ygYKbJxdBCy3_g6xEKu6J_ngvW7uhaY36uoqqn-rOgsOBQdQ1asfxOsDsvCmN-Vm4Sg

[8] https://www.hindawi.com/journals/mpe/2021/9974219/

ROAD SAFETY- Driver drowsiness, Road accidents using machine learning

ORIGINALITY REPORT

4%	3%	0%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Coventry University Student Paper	2%
2	github.com Internet Source	1%
3	data.wu.ac.at Internet Source	1%

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		