Intel Unnati Grand Challenge – Data Analysis

Team TriShakti – Akash Rai, Ankur, Harsh Ranjan

- 10th September 2023

Repo Link: <https://github.com/akashrai2003/IntelUnnatiGrandChallenge>

Abstract:

India has been one of the major hubs for road accidents worldwide with the highest number of fatal accidents in the year 2021. Road accidents claim around 1.5 million lives in India every year, contributing 11 per cent to the total number of such fatalities globally. The worst affected age group in Road accidents is 18-45 years, which accounts for about 67 per cent of total accidental deaths. Tamil Nadu is the state with the highest number of accidents taking place and so we are here to analyze and obtain insights from the data gathered by a device from AI-based ADAS devices. ADAS devices are being connected to the vehicles and thus warn the drivers when there is a chance of collision and similar events with the help of a buzzer and displaying the warning too.

Introduction:

The data gathered by ADAS devices which has information of over 35,000 km travelled by 12 vehicles that were being monitored for a period of 2 months is provided as a CSV file and we are aiming to derive insights from the data available. The dataset can be described with the features as follows:

* *Alert:* 4 different types of alerts i.e. Forward Collision Warning(FCW), Pedestrian Collision Warning(PCW), Lane Departure Warning(LDW) and Headway Monitoring Warning(HMW).
* FCW: Any rear-end collision about to happen within 2 secs is warned.
* PCW: If there are any object/person in front of the vehicle.
* LDW: Changing lanes without an indicator warning.
* HMW: To maintain a safe distance between vehicles moving ahead.
* *Date:* The date in the format DD-MM-YYYY .
* *Time:* The time at which the warnings popped up in the format HH:MM:SS.
* *Lat:* Contains the geographical Lattitude of the location where the warning appeared.
* *Long:* Longitude of the location where the warning appeared.
* *Vehicle:* The 12 different vehicles which were numbered.
* *Speed:* Speed at which the vehicle was travelling at the time of alert.

|  |
| --- |
| Related Work:  A lot of work has been done in this domain majorly focusing on Tamil Nadu as it is the state with the highest number of accidents in India. The government has also started using GIS-based techniques to monitor road safety and databases such as RADMS have been maintained. Previously a lot of analysis was done based on the blackspots observed and also trying to predict accidents by using Time-Series forecasting methods. The majority of accident types are due to overspeeding and overtaking issues and thus there should be proper traffic controls at all heavily barged places.  Data Sources:  Road safety datasets made available to the public by the state governments were available and most of the data was analyzed properly and then displayed the causes and the visualisations based on road accidents. Kaggle datasets were available but the latest dataset was from the year 2020 so we just decided to try to take insights from the data given to us and then go for additional datasets such as the nearest hospitals from the highways and roads, this could easily be done by using applications like QGIS, ArcGIS etc.  Also for plotting the geographical coordinates, we used Python libraries too which include folium, Geopandas, plotly etc.  Beyond the Data:  The features were described earlier and the dataset was already clean without any NaN values.  As we know “Data Analysis is not only about the visualizations but about the inference made from the data, but visualization can make it easier for everyone to understand the meaning.”   1. Getting the geographical locations plotted onto the Map.   We can see that the data collected here is based on the Northern region of Tamil Nadu, thus we can focus majorly on districts like Chennai, Tambaram, Maduranthakam etc.     1. Was overspeeding one of the important factors for these Alerts?   Here we can see that the highest no. of alerts were due to Vehicle speed of -  59 kmph. But all the remainig alerts were at lower speeds and thus we can’t clearly determine if overspeeding was an issue here.  So we will now look more into the Speed vs. Alerts plottings which would make us clear on this topic.       * For FCW’s and LDW’s overspeeding was the root cause of these Alerts, and overtaking was most probably the reason of increase of alerts for LDW * For PCW’s we can say that the vehicle’s weren’t too fast and in the case of HMW’s vehicles were at an average speed as expected on highways…  1. Do peak hours cause more no. of alerts? And specifying the peak hours:   In major metropolitan cities like Chennai, there’s a lot of rush during peak hours and thus these times require higher traffic management. We have defined the peak hours using the following plot -    From the circular barplot, we can infer that the peak hours are between 0700 – 0900 hrs in the morning whereas from 1600 - 1800 hrs. This is the usual departure and arrival time for people working or studying.  So the traffic management should be at its maximum manpower as this rush can cause many unprecedented accidents.     * From the individual plots below we see that all the plots are similar except the PCW plot.   There are more PCW’s due to the fact that early mornings include children going to school, people leaving their homes and also people going for walks/workouts.   1. Insights from the Distribution of Alerts:     HMW’s are present in the maximum numbers which is to be expected as not most of the time there’s a chance for an accident. But the higher no. of LDW indicates that people changing lanes without an indicator is very high and a factor which leads to accidents.  Pie Chart  Also, the high PCW’s are due to highly populated regions in Northern Tamil Nadu due to cities like Chennai & also due to the fact that an International Airport is nearby in the region.  Further, we have analyzed based on the geolocation which gives the most valuable inferences/insights.     1. Based on Geolocation for better Visualization & understanding location of Blackspots :   These GIFs show the heatmap plotted onto the region and thus we can see that each frame represents 1 hour passed and thus covers 24 hours.  HMW  FCW    PCW  LDW  Results:  After the analysis of the dataset given, we can come up with a few solutions in order to minimize the chances of an accident occurring:  Tamil Nadu, an area in the Northern region is a hotspot for most accident causes including overspeeding, overtaking, lane jumping, wrong-side accidents, signal avoidance and other causes.  Road networks in Southern Region areas like Tamil Nadu plagued with overspeeding accidents should have more speed bumps and speed limit signs added  Proper road marking and overtaking signage placement in networks with a high number of overtaking accidents  Imposing stiff penalties on road users violating traffic rules  Installation of cameras along the roads  ClusterMap – To completely view all the locations with pinpoint accuracy against each Alert respectively.  (Double click to Open)  A lot more inferences could’ve been made here too.  References:  <https://towardsdatascience.com/analysis-of-car-accidents-in-barcelona-using-pandas-matplotlib-and-folium-73384240106b>  <https://www.hindawi.com/journals/mpe/2019/2151284/>  <https://kanoki.org/python-plotting-latitude-and-longitude-from-csv-on-map-using-basemap-folium-geopandas-and-plotly> |