Safarlabs

Feasibility Report

August 2023

Harnessing the power of data and AI/ML to enhance road safety:

A feasibility report of
Safar Labs

Solution Summary and Broad Overview:

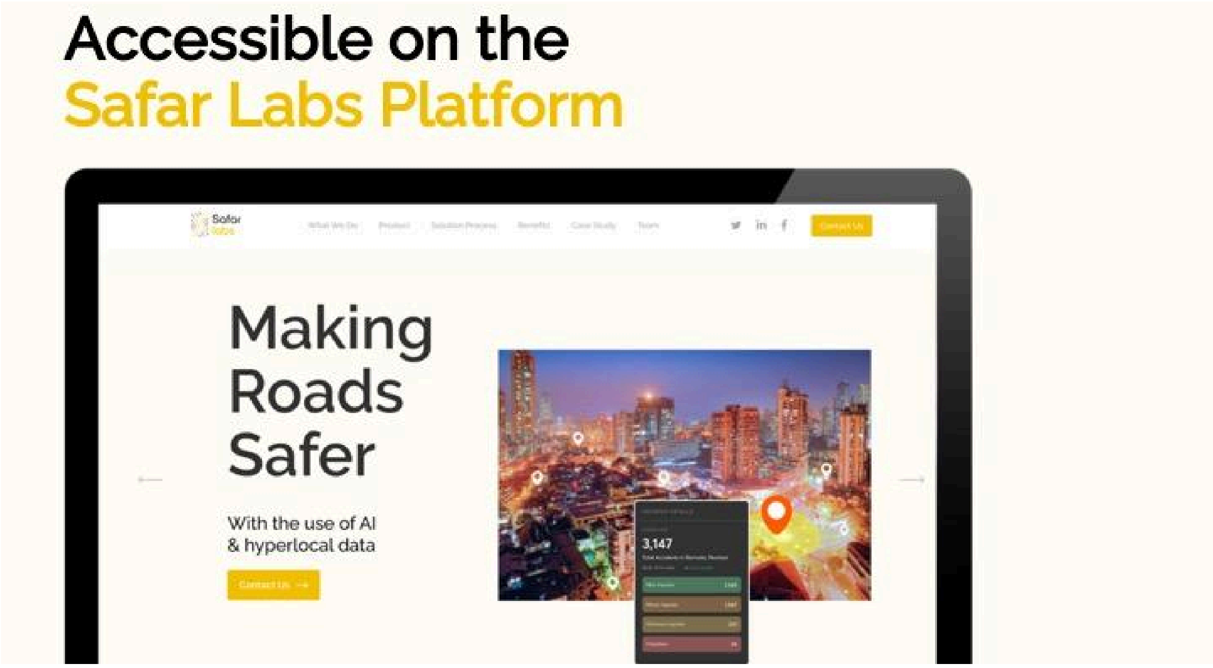
Safar Labs harnesses the power of big data and AI to make road travel safer. Our proprietary AI/ML engine conducts root-cause analysis at the hyperlocal level (street, neighbourhood levels), and recommends hyperlocal policy and regulatory action. SafarLabs dashboard has three modules as follows.

Safar View: The first module - a largely map-based interface - visualises road accidents at a street/hyperlocal level, and identifies black spots, plots causes of accidents or accidents by vehicle type and other parameters.

Safar Cause: The second module identifies causes of different kinds of accidents and predicts the nature of injuries and fatalities in different scenarios. Simple machine learning-based analysis such as clustering of incidents and classification using Decision Trees or Random Forests are used.

Safar SIM: The third and final module is a traffic simulation module which measures the impact of any intervention planned. The user can choose multiple parameters and understand which intervention will give the best result.

Using the aforementioned approach, a pilot demonstration of Safar Labs was conducted in Purba Bardhaman district of West Bengal, India. Data of road accidents spanning 2018-21 was analysed. Results of Safar View - a visual analysis of road accidents - was published in August 2022. The case study report can be accessed here.



Following Part 1 in the Case Study series, we bring to you an interim Feasibility Report*. This Feasibility Report evaluates if the modules of Safar Labs can be built, and if so, specifies the associated requirements such as data, and technology.*

Data Sources

1. Road Accident data in Purba Bardhaman District, West Bengal during 2018-21, shared by Mr. Kamanasish Sen, Superintendent of Police (SP), Purba Bardhaman;
2. Road measurements and video data captured from two accident-prone spots in the same district, during a fieldwork carried out in August 2022;
3. Publicly available road accident datasets from South Korea, Australia, and New Zealand.

Feasibility Report

Safar View: Building interactive dashboard of accidents in Purba Bardhaman District

This work was carried out in May-July 2022. The interactive dashboard allows the user to apply various filters to categorise the accidents according to factors such as type of area, type of vehicles/pedestrians involved, weather condition, hour of the day etc, and visualise them using bar-charts, histograms, time-series etc., and also identify hot-spots.

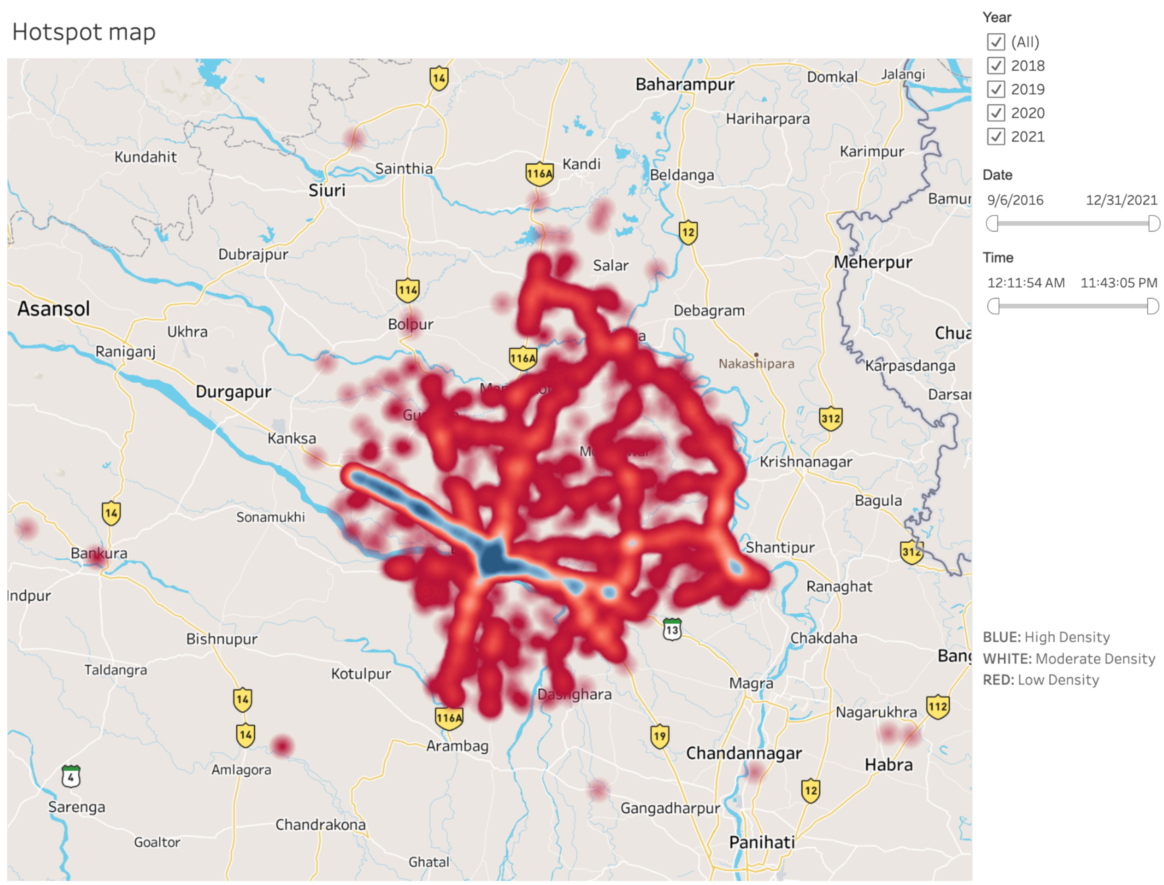


Fig. 1: Map showing hotspot areas of accidents with the colour coding, Blue - high density of accidents, White - medium density of accidents, Red - low density of accidents.

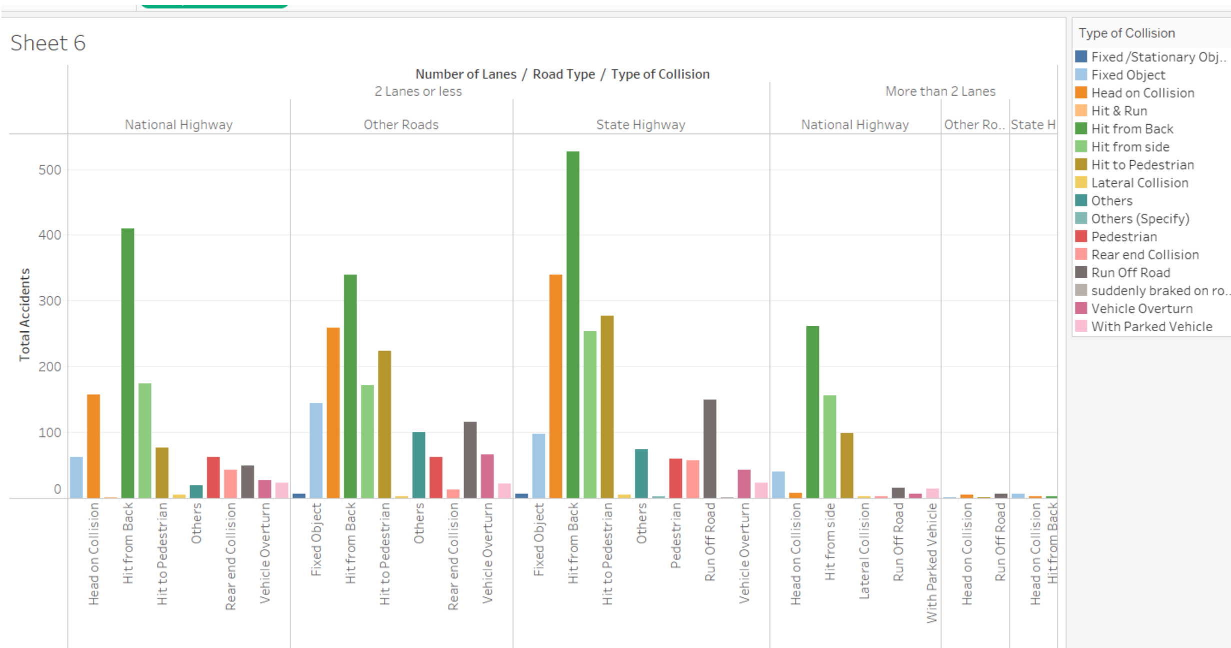


Fig. 2: Histogram representing the number of accidents in 2018-21 by number of lanes, road type, and accident type

Comment on feasibility

Safar View is entirely feasible, given a dataset with appropriate details (*see below*).

Data requirement

Detailed dataset of individual accidents with precise location (lat-lon) and date/time, speeds of vehicles involved, type of vehicles/pedestrians/infrastructure involved, weather conditions, traffic conditions, road condition/nature, number of nature of injuries/fatalities etc.

Technology requirement

Python, any database (eg. MySQL) and web technology (eg. Django)

Safar Cause: Identification of Causes of Individual Accidents, and predicting Accident Risks

This work was carried out during Jan-May 2023 in two parts. The first part focussed on quantitative attribution of individual accidents to different associated conditions (weather, speed limit, type of vehicles, weather and lighting conditions etc.) while the second part was predicting accident risk under different associated conditions. Both parts were based on a Machine Learning (ML) model which was trained using the kind of accident dataset as was mentioned for *Safar View*.

In part 1, a Machine Learning algorithm called Gradient Boosting was used to predict the severity of an accident given the conditions. The importance of the different features (associated conditions) was quantified by “Shapley Values”, a Game Theoretic concept used to quantify the contributions by different players in a cooperative game. Each associated condition present in the dataset can be given an importance score, both for the overall dataset and for individual incidents. Higher the score, more important is its contribution to the accidents.

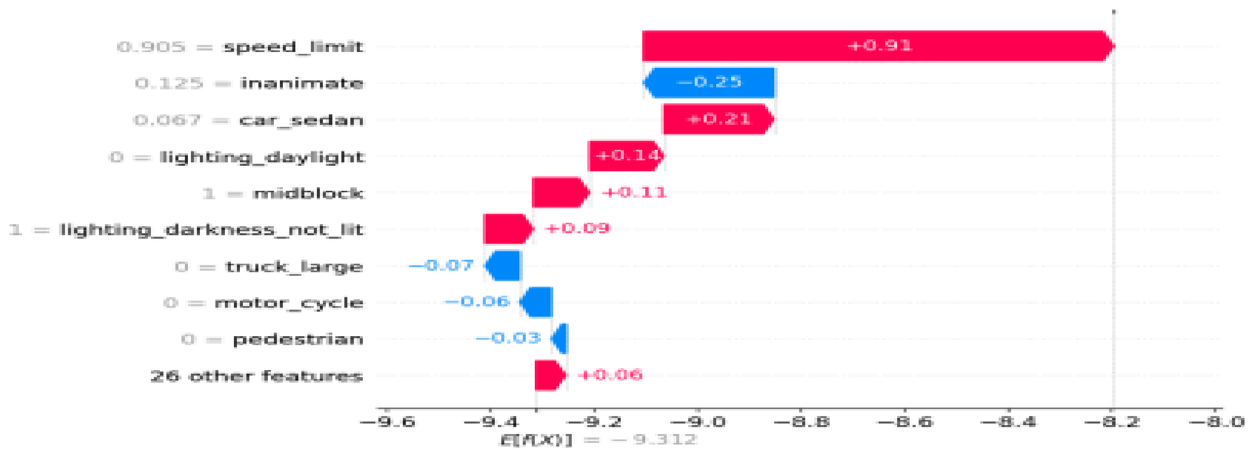


Fig. 3: Waterfall plot for a particular instance

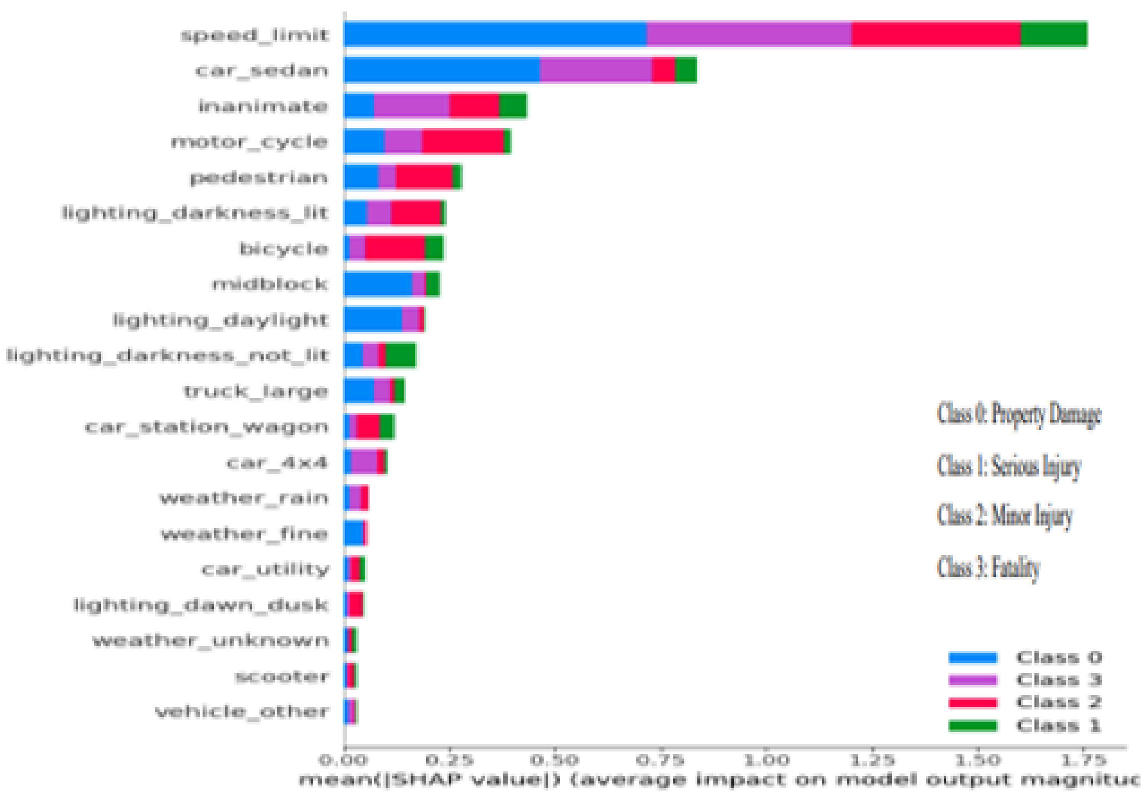


Fig. 4: Plot of mean SHAP values

The second part of the study focussed on predicting accident risk score based on unsafe driving coupled with various external/environmental factors. Gross statistics of "unsafe driving" were identified and correlated with similar statistics of accident risk. Next, ML models were trained to predict accident risk scores (expected number of accidents in a fixed space-time window) using dangerous driving statistics as well as static and dynamic environmental factors.

Comment on feasibility

Safar Cause is reasonably feasible, given a dataset with appropriate details (*see below*).

Data requirement

Same as *Safar View*; can work better with extra information regarding driving and pedestrian behaviour. Such behaviour can be obtained from vehicle log data or traffic cameras, as mentioned in the last section.

Technology requirement

Machine Learning, Deep Learning (usually implemented in Python)

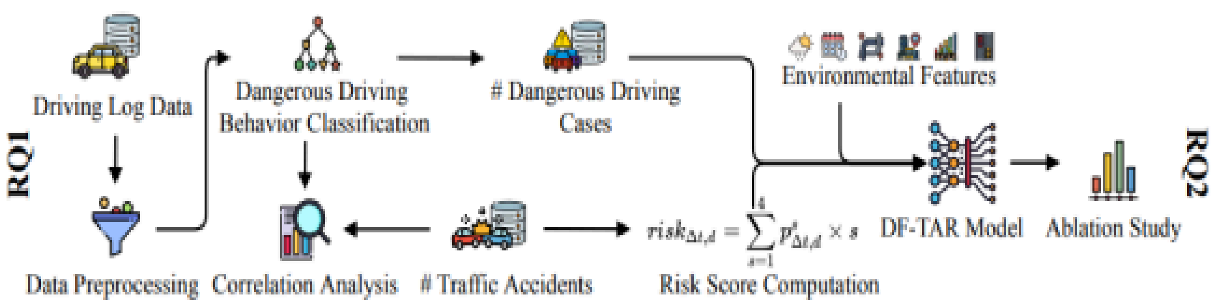


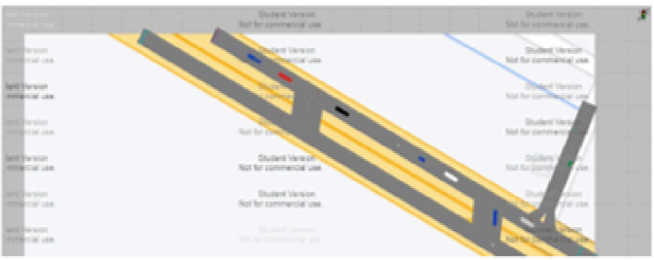
Fig. 5: Overview of the methodology used

Safar SIM: Generation of Microsimulation-based Hyperlocal Policy Recommendations

The third module of *Safar Labs*, namely *Safar SIM*, aims to build a computer model (digital twin) for a junction and find out risks under potential intervention scenarios through simulations. For the feasibility study, we addressed the objective of this module as follows.

We used the software, PTV VisSim. For the case study, we focussed on two locations on National Highway Two (NH2) – near Galsi and Palsit. The measurements of the roads were entered manually into the model based on field measurements, while traffic statistics were entered based on video analytics. These statistics include types of vehicles including pedestrians, their speed distributions, directions etc. The simulation was allowed to run, and data containing positions and velocities of all vehicles and pedestrians at each time-point were logged. By analysing the logs, the number of "risky situations" was calculated. Risky situations were of three types: two vehicles very close to each other, a vehicle too close to a pedestrian, and a vehicle too close to the road edge. This gave the baseline risk, i.e. frequency of potentially dangerous situations under no intervention.

Next, some restrictions were imposed as potential interventions, such as i) imposing speed limit of vehicles, ii) restricting certain types of vehicles (eg. truck), iii) regulating pedestrian movement (they can cross the road at specific points only), and iv) restricting vehicles from making certain turns. In each case, the simulation was allowed to run and the frequency of "risky situations" was calculated from the logged data. Interventions which result in maximum reduction of risky situations were identified.



Count: 5	VehType	DesSpeedDistr	RelFlow
1	100: Car	80: 80 km/h	0.256
2	200: HGV	40: 40 km/h	0.300
3	300: Bus	40: 40 km/h	0.010
4	610: Bike Man	80: 80 km/h	0.220
5	620: Bike Woman	80: 80 km/h	0.040

Fig. 6: The screenshot of the simulation network designed

Data Collection Results

Count: 17	SimRun	TimeInt	DataCollectionMeasurement	Acceleration(All)	Dist(All)	Length(All)	Vehs(All)	Pers(All)	QueueDelay(All)	SpeedAvgArith(All)	SpeedAvgHarm(All)	OccupRate(All)
1	63	0-3600	1	-0.23	169.02	4.03	26	26	0.00	72.81	70.07	14.91 %
2	64	0-3600	1	-0.63	178.53	4.85	94	94	0.00	57.11	51.08	25.82 %
3	65	0-3600	1	-0.56	177.79	4.81	99	99	0.00	56.88	51.13	26.36 %
4	66	0-3600	1	-0.24	169.64	3.87	24	24	0.00	73.36	70.83	13.47 %
5	67	0-3600	1	0.23	178.77	5.68	575	575	10.95	37.95	22.81	60.04 %
6	68	0-3600	1	0.09	179.72	5.64	601	601	10.64	35.23	24.34	59.78 %
7	69	0-3600	1	0.02	178.04	5.65	564	564	12.09	34.16	22.42	63.52 %
8	70	0-3600	1	0.02	178.04	5.65	564	564	12.09	34.16	22.42	63.52 %

Fig. 7: The data collected after running a number of simulations

Further, the correlation between number of "risky situations", and different attributes such as number of vehicles, length of vehicles, average speed, and acceleration etc. were also studied. These help us to get an idea of which factors are most important in creating "risky situations" in the given location, and these point to the possibilities of other intervention measures.

Comment on feasibility

Feasible, subject to data availability.

Data requirement

Detailed measurements and traffic flow statistics are needed. These might be difficult to obtain at each and every junction.

Software requirement

PTV VisSim (for simulation), Python (for analytics)

Additional comment

There is a potential research direction where AI/ML may be used to predict the results based on partial inputs without actually running the simulations. This may alleviate the requirement of detailed data for every location.

Extracting Traffic Statistics from Video Feeds

This part of the project concerns analysing traffic video feeds to extract traffic statistics. Although this is not part of the aims of *Safar Labs*, this can help both *Safar Cause* and *Safar Sim*. Here, we can use various Deep Learning models used for Computer Vision (object detection from images/videos) by adapting them to detect vehicles/pedestrians, classify them according to their types, track their movements, and calculate their speeds, directions etc. Risky behaviour such as abrupt acceleration and turns can also be detected.

Fig. 8a: Detected vehicles in a frame of video 00121

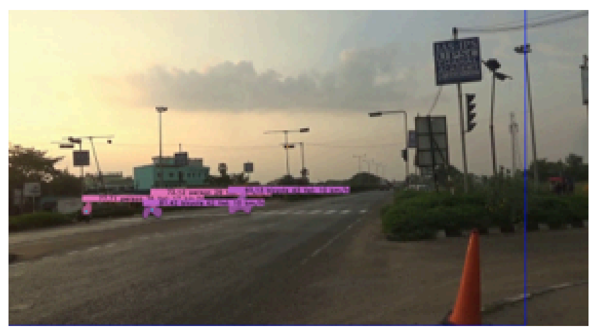


Fig. 8b: Detected vehicles in a frame of video 00130

Proposed Pipeline

Input sources: camera feeds from different traffic junctions, detailed road measurements at the different traffic junctions, detailed accident dataset over the region.

The video feeds will be analysed using Deep Learning-based Computer Vision models to identify traffic statistics, as well as statistics of risky traffic movements.

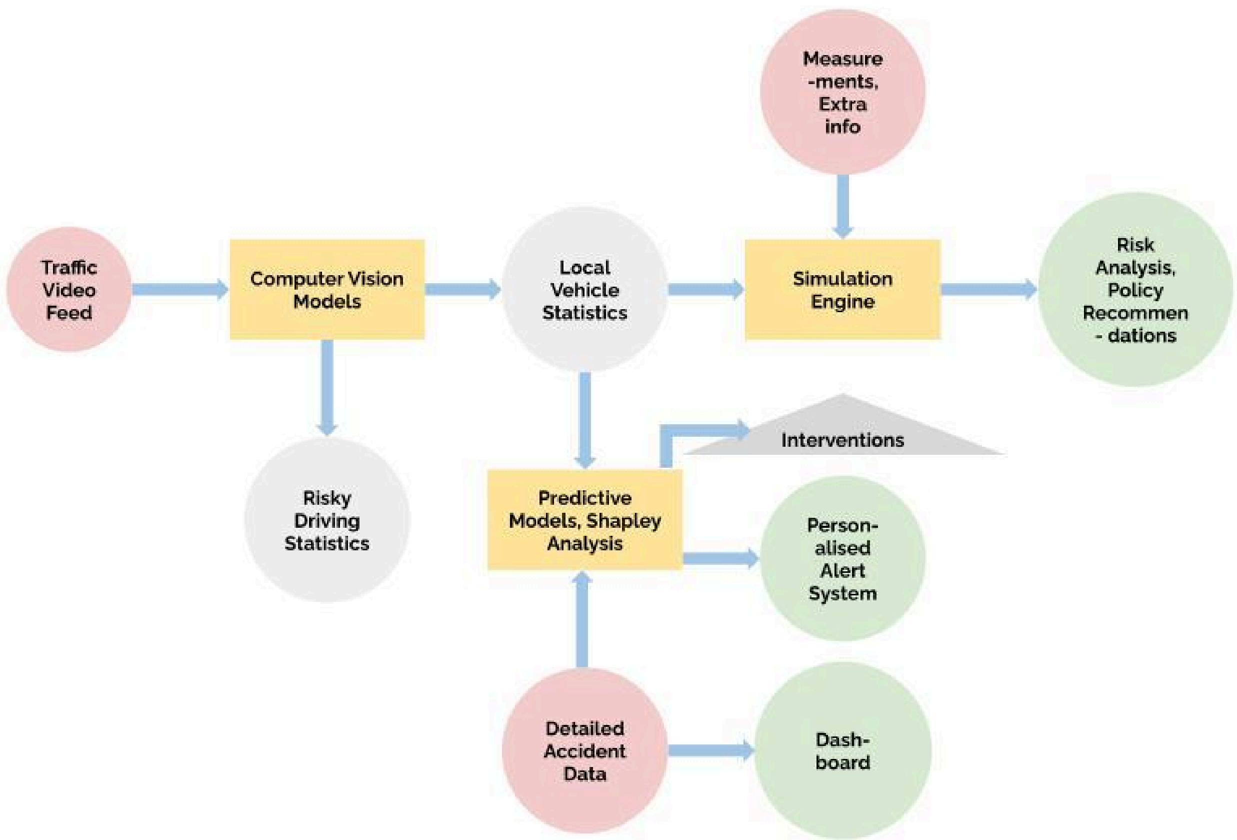
These statistics, in addition to the detailed accident dataset, will be provided as inputs to train Machine Learning-based predictive models. Such models will be able to predict the risk of accidents under the given conditions at any region, and this information can be relayed to traffic control rooms. Shapley Value analysis can be carried out to identify the most likely factors contributing to the risk, and this information can be conveyed to road users through personalised alert message systems, such as “You are entering a high-risk region. Maintain a speed limit of 40 kmph." (for drivers) or "You are entering a high-risk region. Please cross roads cautiously." (for pedestrians).

The traffic statistics calculated from the video feed, and detailed road measurements will be provided as input to the microsimulator. Simulations will be run with adequate duration to calculate the risk. Further simulations will be run with different interventions, and the risk increase or reduction will be calculated in each case. Accordingly, the merits of different intervention measures will be declared for policy-makers to decide.

Deliverables

1. Interactive dashboard of accident records and traffic statistics
2. Alerting system based on risk prediction, including personalised alert with recommendation
3. Microsimulation-based hyperlocal intervention suggestion for each target location

Finally, as a research direction, it can be explored if AI/ML can be used to prevent the need of having to tune the microsimulation for each target location with detailed measurements, and if information from a few select locations can be reused for other locations to predict simulation results without actually running the simulations. Another research direction is to explore if synthetic data from driving simulators like Carla can be used to complement the risky driving statistics, and the detailed accident data.



Safarlabs

About Safar Labs

safarlabsglobal@gmail.com

Safar Labs is a 21st-century initiative which harnesses the power of big data and AI to make road travel safer. The initiative is supported by Salzburg Global Seminar and the Japan India Transformative Technology Network (JITTN). Salzburg Global Seminar is an independent non-profit organisation founded in 1947 with a mission to challenge current and future leaders to shape a better world. JITTN, launched in 2020 by Salzburg Global Seminar and The Nippon Foundation, connects tech entrepreneurs from India and Japan to foster collaborations and surface creative ideas to use tech and artificial intelligence as a force for good, solving some of the pressing challenges of today: mobility, equity and access, economic development. Safar Labs founders, Aishwarya and Adway, are Salzburg Global Fellows participating in the JITTN fellowship programme.

Safar Labs Team

Aishwarya Raman
Executive Director, OMI Foundation; Member - Global Future Council on Urban Mobility Transitions, WEF; Salzburg Global Fellow

Dr. Adway Mitra
Asst Professor, Centre of Excellence in Artificial Intelligence, IIT Kharagpur; Salzburg Global Fellow

Dr. Akhilesh Srivastava
Project Lead - Road Safety 2.0, World Economic Forum, Former Board Member National Highways Authority of India

Acknowledgements

Advisors: Prof. (Dr.) Arkopal Goswami (IIT Kharagpur), Dr. Saddam Hussain (ex-IIT Kharagpur), and Dr. Mahesh Devnani (IPGMER Chandigarh)
Data Access and Guidance: Mr. Kamanasish Sen (SP, Purba Bardhaman), and Mr. Atanu Banerjee (DSP Traffic, Purba Bardhaman), Purba Bardhaman Police.
Students of IIT Kharagpur: Prince Raj, Naman Goyal, Saptarshi Mondal, Pratik Kumar, Tanmay Nagpal, Sajjad Ansari, Abhishek Shukla, Sreyan Biswas, Priyanka Isha, Sadhvika Kadari, Charvi Jain, Deepayan Chakraborty, Sumanta Mishra, Anjali Raj, Sudip Kumar Bhattacharya, Asmita Nandy, Priyanka, and Raktima Barman.

Disclaimer

No party associated with this report will be liable for any loss or damage incurred by the use of this report.