

Project Proposal :

Autonomous driving technology is on the rise but even those who owns the auto driving car is not widely using this feature. Here we would like to study traffic accident data and come up with variables that contribute the traffic accident which we can feed the autonomous driving algorithm to improve safety of the driver. Also we will use this findings so local police and high way patrol can utilize the data by allocating appropriate resources to prevent fatal traffic accident.

Problem Statement :

With more people driving in US, the traffic is getting worse and more traffic accident is happening than before. The car makers started autonomous driving technology development but also it is not 100% safe. We will review the US Accident data ([US Accidents (4.2 million records) | Kaggle](https://www.kaggle.com/sobhanmoosavi/us-accidents)) and identify the variables that contribute the traffic accident.

Criteria for Success:

Successfully identify the variables that contribute the traffic accident and make a model to predict where the traffic accident is likely to happen, so the police/ highway patrol can allocate their resources properly and they can work on traffic accident prevention

Scope of solution space:

1) California Only

2) Identify the variables

3) Successfully come up with a model that can predict the place of the traffic accident

Constraints:

With Covid 19, the driving pattern has been changed in that there are not many drivers on the road. This data has up until 2020 and won’t reflect as of now status.

Stakeholders:

Autonomous algorithm modeler, Local Police, Highway Patrol

Data Sources:

[US Accidents (4.2 million records) | Kaggle](https://www.kaggle.com/sobhanmoosavi/us-accidents)

* “Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. “[A Countrywide Traffic Accident Dataset](https://arxiv.org/abs/1906.05409).”, 2019.
* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. ["Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights."](https://arxiv.org/abs/1909.09638) In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.”