# Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction

## Main Idea:

This paper out of the University of Evora, Portugal, analyzes and creates a model on the severity of accidents, defined as No Victims or With Victims. Using a two-step approach of clustering followed by a model, they found that a Random Forest generated the highest accuracy of 0.73 and precision of 0.44.

## Useful for Project?

This could be useful as it provides the methodology for a decently accurate model. The most influential factors and modeling methods are very useful for our project as we can focus on these instead of trying everything. It also models based on severity of accident, which is not the plan for our project.

## Potential Shortcomings:

There was not implementation of the model, only creating the model itself. By including pedestrian and motorcycle accidents, the data was heavily skewed to these accidents as they usually are worse than the general automotive accident.

# Inferring high-resolution traffic accident risk maps based on satellite imagery and GPS trajectories

## Main Idea:

Using non-standard data, such as satellite imagery and GPS data, this MIT-led paper generated a risk map to predict accident prone locations. Generating an end-to-end deep model, the team establishes similarity between locations to generate trends that are not present in historical data, with these trends being verified using future data. For Los Angeles, the model sees a precision of 0.4767 and RMSE of 39.11\*10^-6.

## Useful for Project?

The modeling techniques, important factors, and identification of sparsity is very important knowledge for our project. The type of data used is not as useful for us as this data is very niche and not widely available.

## Potential Shortcomings:

While this model was applied to many cities, it needs the GPS and satellite data. This is not simple to collect or implement, so this model could not easily be applied to other cities or cities without this data available.

# High-Resolution Road Vehicle Collision Prediction for the City of Montreal

## Main Idea:

In this paper from Concordia University in Montreal, the road vehicle collision prediction problem was framed as a classification problem using the measure of probability as a risk of an accident. 2.3 million negative (no accident) examples were generated to balance the positive (accident) examples. Using the Balanced Random Forest algorithm, the model predicted 85% of vehicle collisions with a false positive rate of 13%.

## Useful for Project?

This model had one of the highest fits out of any researched. The classification methodology would be worth investigated in our project due to the goodness of fit possibility. They used easily accessible datasets and algorithms. They also have important features that are useful for our project.

## Potential Shortcomings:

They only used 100 trees for their model but did not try more than 200. There may be a better solution with additional trees. They also did not include any routing or hot spot generation.