What is a Crash Prediction Model? Look into this.

* Especially for Toronto
* Has there been a crash prediction model developed for Toronto? Ans. Doesn’t look like it.

Do I have to combine different data sets into one dataset before conducting analysis?

Datasets

* Toronto KSI
* Toronto All Collisions
* Toronto Road Network
* Road Features (Speed limit etc.)
* Toronto Traffic Speeds
* Toronto Traffic Volumes
* Toronto Neighborhood Demographics
* Toronto Traffic Control
* Toronto Bike Lanes
* Road Infrastructure Data (Sidewalks, Number of Lanes etc.)
* Land Use and Zoning Map

Models

* Gradient Boosting
* Random Forest
* Neural Networks (Try Keras NN as well)

Questions to Model

* What is the probability that a collision will occur on this intersection at a given day?
* How many collisions are predicted to occur at this intersection monthly/annually?
* If a collision occurs at this intersection, what is the probability that it will be fatal?
* If a collision occurs, what is the predicted number of fatalities/injuries likely to happen?

\*\*I will have to define the problem as a classification or a regression problem.

* David said classification might be easier to work with.

\*\*You might have to change your objective to “if a collision occurs at this intersection, what is the probability that it will be fatal?”

* This will enable you to avoid having to create negative examples.
* Could also make it a classification problem: major/fatal or minor. Or 3-stage classification – minor, major, fatal.

There are two ways I can model the collision prediction:

1. Based on intersections (I have intersection names. Use fuzzy search)
2. Based on gps coordinates (with less significant digits) to match locations.
3. There seems to be grid models that convert latitude and longitude values to grids (e.g. Uber H3).

Decide on which metrics to compare the performance on the models on.

* The research paper says accuracy might not be a good choice and recall might be better.

Also have to figure out what time resolution to use: per hour, day, month etc.

Could also do time series analysis of the data.

The project could also be just predicting bike collisions or pedestrian collisions, does not have to be all collisions.

**Project Steps**

1. Finalize project objectives.
2. Gather datasets.
3. Merge all datasets together.
4. Conduct exploratory analysis for modeling. (make sure to include traffic volumes).
5. Data pre-processing.
6. Conduct descriptive analytics (maybe build a dashboard)
7. Feature Engineering.
8. Modeling
9. Performance review and hyper-parameter tuning.
10. Create risk map of roads/intersections.
11. Interpretation and conclusion.