Reducing the Frequency of Fatal Road Accidents in the City of Toronto

Sam Khonji

MIE1624H Course Project

Summary:

This report summarises the development of a machine learning tool that aims to identify, for a given number of intersections, which ones will have fatal accidents. The tool is intended to be useful in prioritising specific intersections in response to the question: where should resources be deployed to reduce fatalities by motor vehicles in urban centres? This report focuses on the tool's potential to reduce the frequency of fatal road accidents in Toronto.

1 INTRODUCTION

Road traffic accidents have a large impact on societies worldwide. In densely populated urban centres with complex infrastructure like Toronto, the effects are pronounced; the city has recorded 4,400 accidents between 2007 and 2017. A total of 580 of these lead to fatalities, resulting in 593 deaths [1]. Notably, 190 pedestrians and 16 cyclists were killed in collisions with vehicles in the last 5 years [1]. There is a marked upward trend in the number of severe accidents that have occurred each year over the last decade, as shown by Figure 1.

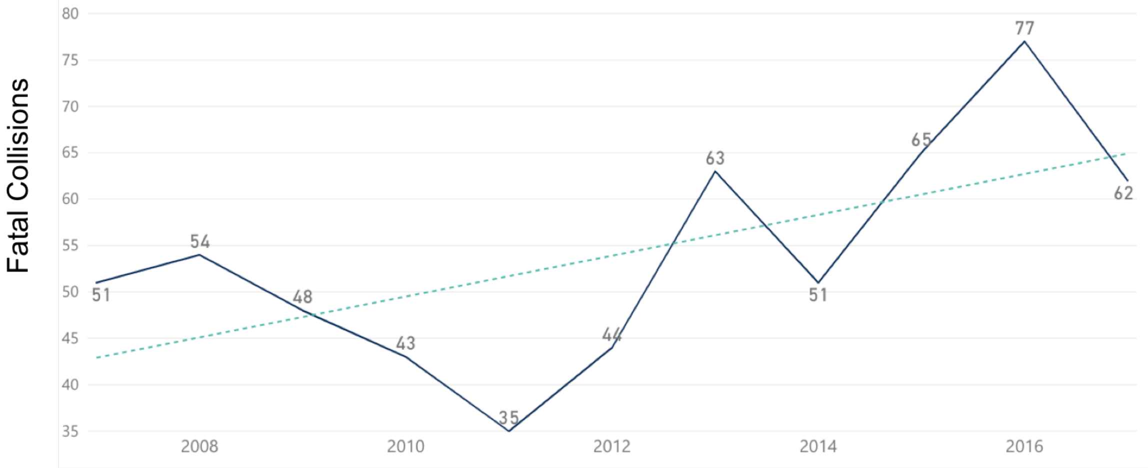

FIGURE 1: Graph showing a general increase in the number of traffic fatalities from 2007 to 2017 [1].

Fatal accidents are not only devastating to any individuals involved and those who know them, they are also a significant economic loss to the city. Some of the components of the social cost of a fatal/severe accident include: loss of life, loss of life quality, loss of economic output from those affected, medical costs, legal costs, and repair costs. While the costs of life-changing injuries and loss of life are difficult to quantify, estimates for the average social cost per fatal road accident are greater than C$ 3.36 Million (using 2008 prices) [3]. This amount sums to a cost of around C$ 240 Million in 2017 alone (adjusted for inflation).

A significant number of fatal accidents in Toronto occur at road intersections [2], making these areas of critical importance for the city to address its road safety issues. There is a multitude of causes of accidents at intersections, but there is also much that can be done to reduce the risk and severity of these events through – both long-term and short term – preventative measures. For example, this might involve construction/maintenance work to improve road quality, or the application of road salt to tackle road surface issues. Additionally, even causes related to driver negligence may be addressed through the implementation of various speed limit enforcement methods. Another strategy to reduce the impact of severe collisions, and reduce the possibility of fatal accidents, is to enhance reactive measures. One way to do this is to station paramedic services in closer proximity to certain intersections, in order to reduce post-crash response times (the length of time it takes for paramedic services to arrive at an emergency scene).

The time and resources available to tackle these issues are limited. With more than 2280 intersections in Toronto, the municipal government is faced with the task of identifying which intersections should be prioritised for the application of these measures. This a dynamic process – the factors that render certain intersections prone to fatal accidents are constantly changing, meaning that there is no final solution to the problem. As such, there is significant value to be gained from a tool that is able to make data-driven predictions to pinpoint which intersections are most likely to have fatal accidents at a given moment in time. In being able to identify these 'hotspots', preventative and reactive measures can be periodically targeted to specific locations, which is particularly helpful for short-term safety measures. Ultimately, these insights have the capacity to decrease both the likelihood and impact of severe and fatal accidents in the city. The following sections outline the development and capabilities of such a tool.

2 DATA

The first step in the development of this tool was to identify the factors that affect (fatal) road accidents at intersections. This process determined the data that was collected to enable the model to make predictions. While the potential number of factors is extremely large, only those factors deemed relatively important, and for which data was obtainable, were of interest. Table 1 summarises the data that the tool uses. The datasets were merged together and manipulated such that they corresponded to each of the intersections that were analysed.

TABLE 1: The different data sets that were used to build the model.

Data Used	Source
Peak 8 hour vehicular volume	City of Toronto Open Data – Transportation
Peak 8 hour pedestrian volume	City of Toronto Open Data – Transportation
Traffic intensity (rush hour and normal hours)	City of Toronto Open Data – Transportation
Historical accident counts	City of Toronto Open Data – Transportation
Historical incidents of aggressive driving	City of Toronto Open Data – Transportation
Historical incidents of speeding	City of Toronto Open Data – Transportation
Presence of a streetcar routes	City of Toronto Open Data – Transportation
Presence and characteristics of bike lanes	City of Toronto Open Data – Transportation
Presence of red light cameras	City of Toronto Open Data – Transportation
Presence of red light cameras	City of Toronto Open Data – Transportation
Season (weather conditions)	Toronto Police Service Data – Killed or Seriously Injured
Day of the week (weekday/weekend)	Toronto Police Service Data – Killed or Seriously Injured
Time of the day	Toronto Police Service Data – Killed or Seriously Injured
Road class (e.g. major, minor)	Toronto Police Service Data – Killed or Seriously Injured
Visibility and light levels	Toronto Police Service Data – Killed or Seriously Injured
Median age of population	Statistics Canada – Census of Population Data
Population density	Statistics Canada – Census of Population Data
Presence of attractions (distractions)	City of Toronto Open Data – Attractions

3 RESULTS & DISCUSSION

The tool uses a Random Forest algorithm to classify the severity of potential accidents at different intersections . When applied to a set a intersections, it classifies them into:

1. Intersections where, if accidents were to occur, they would be fatal (severe)
2. Intersections where, if accidents were to occur, they would be non-fatal

In doing this, it is returns a subset of intersections that are most at risk of having fatal accidents. Figure 2 visualises the output of the tool, demonstrating how it can be used to prioritise locations for the implementation of preventative and reactive safety measures.

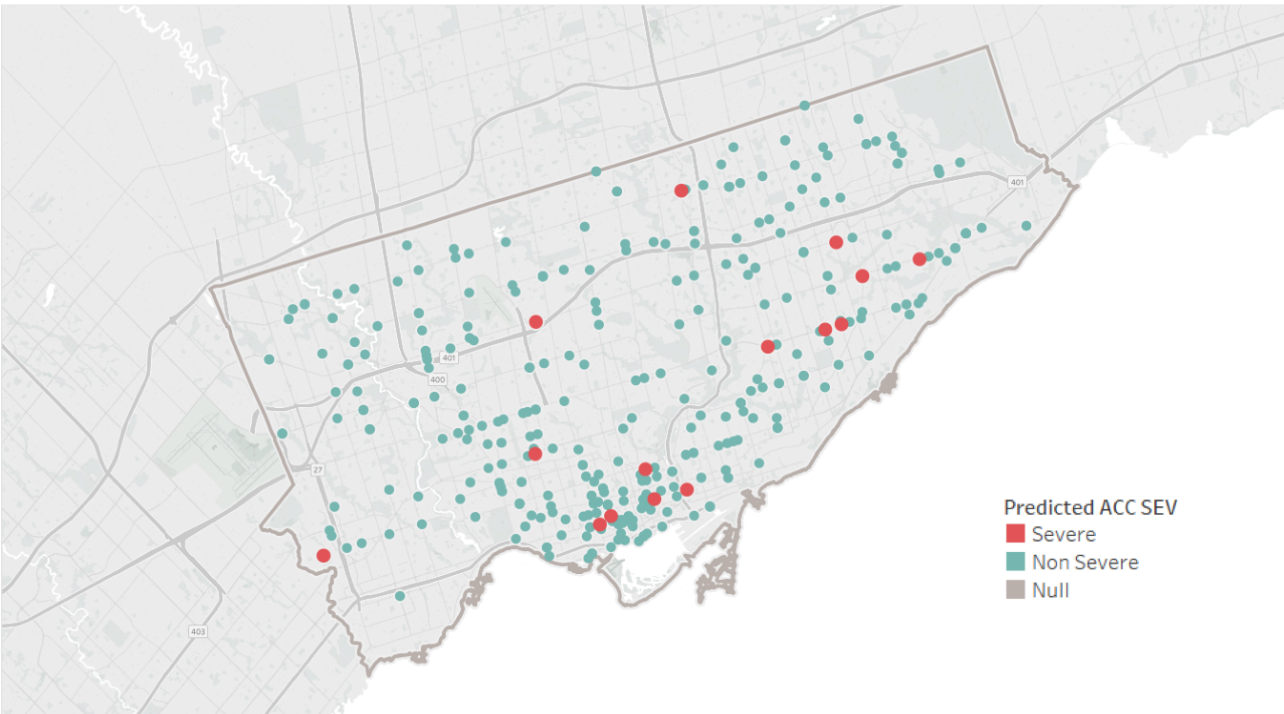

FIGURE 2: An example of the tool's output – red dots indicate intersections where road accidents are likely to be fatal.

The tool is able to pinpoint which intersections will have fatal accidents with a precision of 88%. This value was obtained by evaluating the tool's predictions against actual historical fatalities that took place at different intersections; 88% of the predicted fatal accidents were correctly identified. The parameters of the model were calibrated to maximise the tool's precision, thereby minimising the number of intersections that are falsely identified (false positives). This was done because its primary goal is to specify a set of intersections prone to fatal accidents, that decision makers can be confident in targeting with preventative and reactive measures, to reduce the number of fatalities in the city.

Figure 3 illustrates the relative importance of the factors in allowing the tool to classify the intersections. Note that some of the items shown in Table 1 consisted of multiple datasets, and some new features were created from the existing data as new feeds to increase the model's performance. For example, 'Pedestrians per car' was derived by dividing vehicular volume by pedestrian volume. As a result, a total of 35 features (shown in Figure 3) are actually used by the model. It is apparent that the recorded incidents of aggressive driving gives the strongest indication of whether an accident at an intersection will be fatal or not. Other strong indicators are vehicular volume, light levels, season (weather conditions), time of the day, and demographic data.

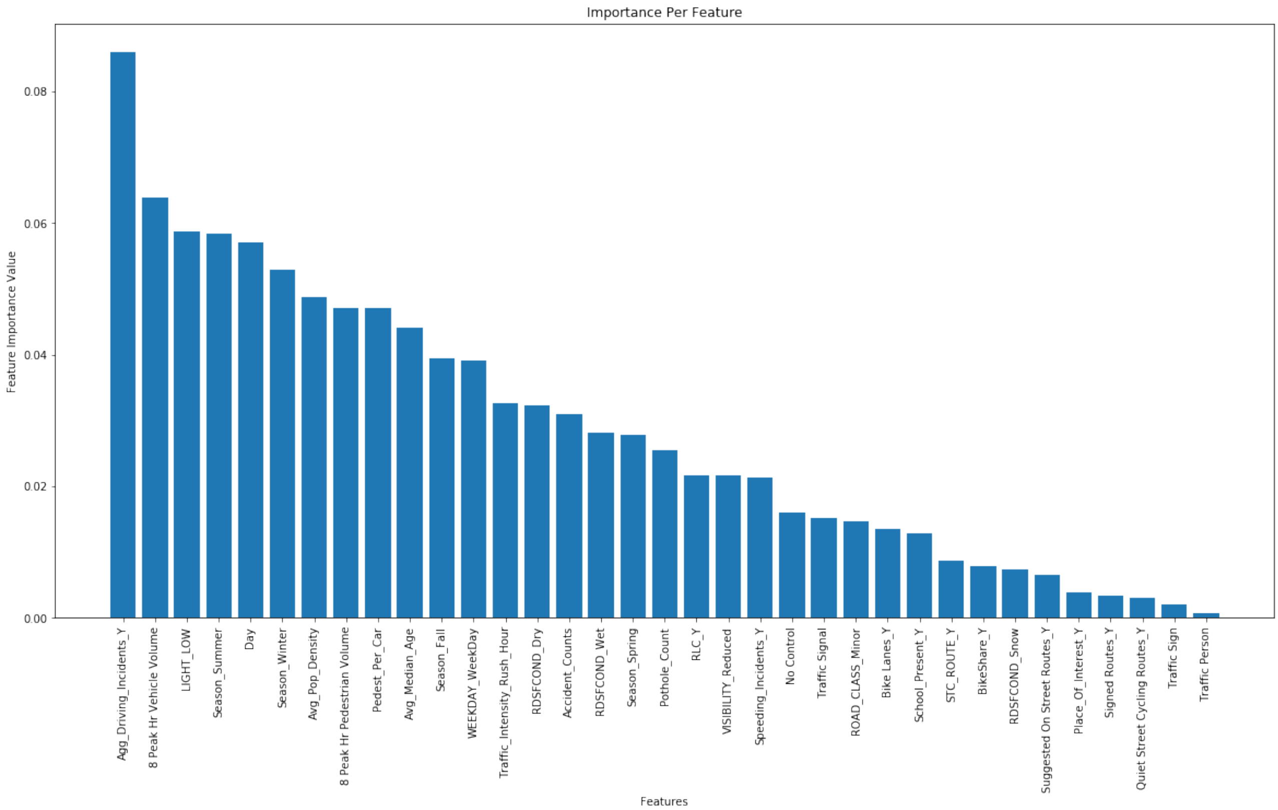

FIGURE 3: Graph showing the relative importance of the various inputs (features) of used by the tool. Refer to the appendix on page 6 for a key that explains each feature.

It is worth pointing out some limitations of the model. There are certain factors that are highly significant in predicting fatal accidents, that cannot be be accounted for using publicly available data. Some examples include whether passengers are wearing seat belts, and the makes/models (and associated safety ratings) of cars driven through the intersections. Without using these factors, there is a limit to the degree of accuracy that is attainable. The model can, however, be improved by increasing the quantity of data it uses by deepening the level of detail of the existing features, and increasing the frequency of data collected as inputs for the model.

4 CONCLUSIONS & RECOMMENDATIONS

The City of Toronto has set a number of safety targets to be met, and implemented various safety measures to achieve them, as part of its Vision Zero Road Safety Plan [4]. However, there is currently no indication that the capabilities of machine learning are being utilised. This report has presented a machine learning tool that has the potential to assist the city in meeting its goals.

It is recommended that the tool is implemented as a means to improve decision-making in the application of preventative and reactive safety measures. In particular, it is proposed that the subset of at-risk intersections provided by the model are addressed with respect to the top 5 most important *controllable* factors from Figure 3. These factors, and some possible preventative measures that be can used to tackle them, are shown in Table 2.

TABLE 2: The 5 most important factors in identifying which intersections will have fatal accidents, and the preventative safety measures that can be taken to address them.

Factor	Safety Measure	Cost Per Unit (C$)
Aggressive driving	Speed traps, bumps, cameras, sensors, police patrol	3,000–150,000
Low light levels	Enhance lighting, build street lights	5,000–10,000
Road surface condition	Road salt/sand (when appropriate)	30–200
Presence of potholes	Road maintenance work (if present)	50–400
No traffic control	Appropriate control (e.g. signs, signals)	200–700,000

In relation to paramedic response times, at present, paramedic services are situated such that they have a response time of 11.5 minutes to all locations in the city [5]. This means means that the city is relying on random chance that an accident will occur in the deployment of emergency services, and therefore a precision of greater than 50% for the model is valuable in this context; the tool fulfils this requirement. To maximise the usefulness of the tool, it is recommended that the model is provided with access to live dynamic inputs such as weather data, and data from videos and traffic signals to provide live vehicle counts and accident counts. The model can be ran on a periodic basis (hourly/daily/weekly/monthly) to identify which intersections should be targeted in the allocation of certain resources.

References

[1] Toronto Police Service Public Safety Data Portal. Traffic Collision Fatalities dashboard (KSI) `http://data.torontopolice.on.ca/pages/fatalities`

[2] Toronto Police Service Public Safety Data Portal. Killed or Seriously Injured (KSI) Data. `http://data.torontopolice.on.ca/datasets/9f05c21dea4c40458264cb3f1e2362b8_0`

[3] Victoria Transport Policy Institute. Transportation Cost and benefit Analysis II – Safety and Health Costs. `hhttp://www.vtpi.org/tca/tca0503.pdf`

[4] City of Toronto. Transportation Services Vision Zero Road Safety Plan 2017. https://www.toronto.ca/wp-content/uploads/2017/11/990f-2017-Vision-Zero-Road-Safety-Plan_June1.pdf

[5] City of Toronto. Toronto Paramedic Services. 2017 Annual Report. https://www.toronto.ca/wp-content/uploads/2018/03/9730-Toronto-Paramedic-Services-Annual-Report-2017-sm.pdf

APPENDIX

The following dictionary is required to understand the meaning of each feature in Figure 3:

1) Main: the main intersection
2) Side 1 Route: second intersection
3) 8 Peak Hr Vehicle Volume: the number of vehicle that pass by the intersection in 8hrs
4) 8 Peak Hr Pedestrian Volume: the number of pedestrians that pass by the intersection in 8hrs
5) Pothole_Count: number of potholes around each intersection
6) Avg_Pop_Density: average Population Density by ward (44 Ward Model)
7) Avg_Median_Age: median age by ward (44 Ward Model)
8) Pedest_Per_Car: number of pedestrians by vehicle
9) Accident_Severity: target variable; represents severity of accidents (fatal vs. non-fatal)
10) No Control: 0 = intersection has no traffic control (such as traffic signal or red light camera) and 1 = intersection has at least one type of control
11) Traffic Person: 1 = Police directing traffic, 0 = No Police directing traffic
12) Traffic Sign: 1 = there is a stop and yield sign, 0 = there is no stop and yield sign
13) Traffic Signal: 1= there is a traffic light, 0 = there is no traffic light
14) Day: 1 = day time, 0 = night time
15) Traffic_Intensity_Rush_Hours: 1 = during rush hour, 0 = no during rush hour
16) Season_Fall: 1= accidents during Fall, 0 = another season
17) Season_Spring: 1= accidents during Spring, 0 = another season
18) Season_Summer: 1= accidents during Summer, 0 = another season
19) Season_Winter: 1= accidents during Winter, 0 = another season
20) BikeShare_Y: 1 = bike share present around the intersection, 0 = bike share is not present around the the intersection
21) RLC_Y: 1= Redlight camera present at the intersection, 0 = Redlight camera is not present at the intersection
22) Agg_Driving_Incident: 1= aggressive driving involved, 0 = aggressive driving is not involved
23) Speeding_Incidents_Y: 1= accident caused by speeding, 0 = accident was not caused by speeding
24) RDSFCOND_Dry: 1 = Dry Road surface condition, 0 = other type of road surface condition
25) RDSFCOND_Snow: 1 = Snow Road surface condition, 0 = other type of road surface condition
26) RDSFCOND_Wet: 1 = Wet Road surface condition, 0 = other type of road surface condition
27) LIGHT_LOW: 1= Low light, 0 = high light

28) VISIBILITY_Reduced: 1 = reduced visibility, 0= good visibility

29) ROAD_CLASS_Minor: 1 minor road, 0 = is not a minor road

30) WEEKDAY_WeekDay: 1 = weekday, 0 = weekend

31) STC_ROUTE_Y: 1= there is a Street Car Route by the the intersection,
0 = there is not a street car route by the the intersection

32) School_Present_Y: 1 = if there is a school around the intersection,
0 = if there is not a school around the intersection

33) Suggested On Street: 1 = suggested on street bike route, 1 = no suggested on street bike route

34) Signed Routes_Y: 1 = signed bike route, 0 = no signed bike route

35) Bike Lanes_Y: 1 if there is a bike line at the intersection and 0 if there is not a bike line at the intersection

36) Quiet Street Cycling: 1 = quiet street route