Predicting Severe Traffic Outcomes: Recommendations for Proactive and Reactive Traffic Accident Mitigation

Stefan Banducci
September 2020

I. Introduction / Business Problem

Road traffic accidents are a serious problem in the United States and worldwide. The most severe traffic accidents each year are responsible for serious injuries as well as loss of life. According to the National Safety Council, in 2019, an estimated 38,800 people lost their lives to car crashes in the United States. In the same year, an additional ~4.4 million people in the US were injured seriously enough to require medical attention in crashes. The cost associated with a traffic accident death is often immeasurable - a lost love one, a child who grows up without a loving mom or dad. Additionally, those that suffer serious injuries in traffic accidents often bear significant costs as well. Traffic-accident-driven serious injuries can lead to job losses, financial hardships, as well as psychological impacts for both the injured party as well as his/her family. Given the high societal, family, and individual costs associated with accidents, addressing traffic accident fatalities and serious injuries represents a significant opportunity.

At the highest level, severe traffic accidents can be addressed proactively and/or reactively. A proactive approach involves trying to stop severe accidents from happening before they occur. This approach necessarily involves a solid understanding of the variables that contribute to severe accident outcomes. Knowing what variables drive severe accident outcomes (e.g., speeding, drunk driving, weather, etc.) can enable development of specific strategies for accident prevention. For example, if speeding is found to be a key driver of severe traffic accidents in an area, this understanding can inform appropriate speed limits and enforcement strategies.

In addition to proactive strategies, it's also possible to mitigate the cost of traffic accidents by improving the way that society reacts when accidents are first reported. Today, emergency response resources are not always deployed optimally when accidents occur. Being able to predict whether a newly-reported accident is likely to be relatively severe vs. minor could be very helpful in ensuring that severe accidents are responded to quickly and appropriately. This could in turn save lives and/or reduce the severity of injuries sustained in accidents.

This study takes an analytical, machine-learning approach to:

(1) Understand the key drivers of fatal and serious injury traffic accidents so as to inform proactive strategies for mitigating accidents before they happen, and

(2) Develop a predictive model to gauge whether or not recently reported accidents are severe vs. minor so that emergency response resources can be more appropriately deployed.

II. Data Overview

This study will leverage historical collisions data made available by the City of Seattle via their open data portal (https://data-seattlecitygis.opendata.arcgis.com/). The dataset contains information on ~220,000 accidents that occurred in the City of Seattle between 10/06/2003 and 9/5/2020. The data is collected by the Seattle Police Department (SPD) and is updated weekly. Accident category / severity data is available for ~200,000 of the accidents in the database. Accidents within the data set are classified into one of 4 severity categories:

Table1: Seattle GeoData Collisions Data

Severity Category	Description	% of Accidents
1	Property Damage Only Collision	68.9%
2	Injury Collision	29.4%
2b	Serious Injury Collison	1.6%
3	Fatality Collision	0.2%

Given the focus of this study on mitigating serious and fatal collisions, ‘Serious Injury Collisions’ and ‘Fatality Collisions’ will be grouped together as ‘Severe Accidents’. Additionally, ‘Property Damage Only Collisions’ and ‘Injury Collisions’ will be analyzed collectively as ‘Minor/Less Severe Accidents’.

The City of Seattle collisions database also contains a rich number of additional fields that describe each accident. Amongst other data, information on accident date, location, and type (head on, sideswipe, etc.) are available. Additionally, environmental information relevant to each accident is also provided including weather and road conditions. A summary of select key fields included in the data is provided in Table 2.

Table2: Seattle GeoData Collisions Data – Select Data Fields

Field	Description
INCDATE	Date of the accident
INCDTTM	Time of the accident
ADDRTYPE	Collision Address Type (Block, Intersection, or Alley)
LOCATION	Description of the general location of accident
COLLISIONTYPE	Type of collision e.g., Parked Car, Sideswipe, etc.
JUNCTIONTYPE	Category of junction at which incident took place
UNDERINFL	Whether or not driver was under influence of alcohol/drugs
WEATHER	Weather conditions during the accident
ROADCOND	Road conditions during the accident
LIGHTCOND	Light conditions during the accident
SPEEDING	Whether or not speeding was a factor
HITPARKEDCAR	Whether or not collision involved a parked car