SEATTLE PARAMEDIC DEPLOYMENT OPTIMIZATION

ENABLING QUICKER DECISIONS

TOM AMBLER

SEPTEMBER 2020

INTRODUCTION

- At a recent town hall, citizens expressed concerns about the time for Emergency Medical Services to arrive on the scene of accidents
  - The incidents cited involved serious injuries
- Rather than research individual incidents, in order to dispel their perception of arrival time, the public requested in internal review
- Further, any changes need to be reasonable
  - Cannot double/triple staff of EMS and send to every accident
  - Cannot have deployment procedures too difficult to follow
- Public safety is a top priority, so immediate action must be taken

THE GOAL

- The Mayor and Head of EMS have sanctioned a project to understand key differentiators for severe accidents
- The goal is to look for opportunities to speed the decision-making process
- Data would need to be analyzed to see if there are common circumstances exist. If so,
  - Are the reasonable
  - Are they different from existing training
  - Are they easy to implement

DATA OVERVIEW

- Seattle has been capturing robust accident data since 2004, and a copy can be found here: Seattle Collisions Data
- The metadata, found here Seattle Collisions Metadata, is well documented too
- There are 194K+ accidents recorded
  - The dataset only has two of the severity codes which could skew results
  - The dataset only has 38 of the columns listed in the metadata
- This datasets could be augmented by other sources, but it is sufficient on its own for this targeted analysis

SAMPLE DATA

- Sample data showing a severity code and potential features

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO	STATUS	ADDRTYPE	INTKEY	...	ROADCOND
0	2	-122.323148	47.703140	1	1307	1307	3502005	Matched	Intersection	37475.0	...	Wet
1	1	-122.347294	47.647172	2	52200	52200	2607959	Matched	Block	NaN	...	Wet
2	1	-122.334540	47.607871	3	26700	26700	1482393	Matched	Block	NaN	...	Dry
3	1	-122.334803	47.604803	4	1144	1144	3503937	Matched	Block	NaN	...	Dry
4	2	-122.306426	47.545739	5	17700	17700	1807429	Matched	Intersection	34387.0	...	Wet

FEATURES

- Initially, a large number variables were considered for the analysis: Address Type, Weather, Road Conditions, Light Conditions, Time of Day, Collision Type, Under Influence, Collision Code, Person Count, Pedestrian Count, Vehicle Count and Cyclist Count
- Feature Pruning
  - Ordinal variables were dropped for having skewed, low value results
  - Time of Data was removed in favor of Light Conditions for visibility
  - Junction Type was removed in favor of Address Type due to high correlation
  - Weather was removed in favor of Road Conditions due to high correlation
- Remaining five features had their empty values replaced by their most common value, and then they were converted to numeric values for analysis

ORDINAL FEATURES

- Ordinal features were considered
  - Does severity increase if the number of people involved is higher
- Sample data

	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT
0	2	0	0	2
1	2	0	0	2
2	4	0	0	3
3	3	0	0	3
4	2	0	0	2

ORDINAL CORRELATIONS

- Ordinal features show low correlation with target

	SEVERITYCODE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT
SEVERITYCODE	1.000000	0.130949	0.246338	0.214218	-0.054686
PERSONCOUNT	0.130949	1.000000	-0.023464	-0.038809	0.380523
PEDCOUNT	0.246338	-0.023464	1.000000	-0.016920	-0.261285
PEDCYLCOUNT	0.214218	-0.038809	-0.016920	1.000000	-0.253773
VEHCOUNT	-0.054686	0.380523	-0.261285	-0.253773	1.000000

CATEGORICAL FEATURES

- Features considered except Time of Day
- Sample Data

	ADDRTYPE	WEATHER	ROADCOND	LIGHTCOND	COLLISIONTYPE	JUNCTIONTYPE	UNDERINFL
0	Intersection	Overcast	Wet	Daylight	Angles	At Intersection (intersection related)	N
1	Block	Raining	Wet	Dark - Street Lights On	Sideswipe	Mid-Block (not related to intersection)	0
2	Block	Overcast	Dry	Daylight	Parked Car	Mid-Block (not related to intersection)	0
3	Block	Clear	Dry	Daylight	Other	Mid-Block (not related to intersection)	N
4	Intersection	Raining	Wet	Daylight	Angles	At Intersection (intersection related)	0

CATEGORICAL CORRELATIONS

- Weather and Road Conditions correlation is .752
- Address Type and Junction Type correlation is .951

	ADDRTYPE	WEATHER	ROADCOND	LIGHTCOND	COLLISIONTYPE	JUNCTIONTYPE	UNDERINFL
ADDRTYPE	1.000000	-0.090608	-0.030661	-0.053434	-0.466902	-0.915249	-0.041532
WEATHER	-0.090608	1.000000	0.752051	0.208585	0.026320	0.111295	-0.038970
ROADCOND	-0.030661	0.752051	1.000000	0.022630	-0.003217	0.042160	-0.008955
LIGHTCOND	-0.053434	0.208585	0.022630	1.000000	0.034426	0.057661	-0.218037
COLLISIONTYPE	-0.466902	0.026320	-0.003217	0.034426	1.000000	0.470961	0.002491
JUNCTIONTYPE	-0.915249	0.111295	0.042160	0.057661	0.470961	1.000000	0.048592
UNDERINFL	-0.041532	-0.038970	-0.008955	-0.218037	0.002491	0.048592	1.000000

TARGET / OUTCOMES

- The accident severity was clearly defined, however there were significantly more property damage incidents than severe injury collisions

Severity Description	Count
Property Damage Only Collision	136,485
Injury Collision	58,188

- When training the data, we need to ensure the data was split equally

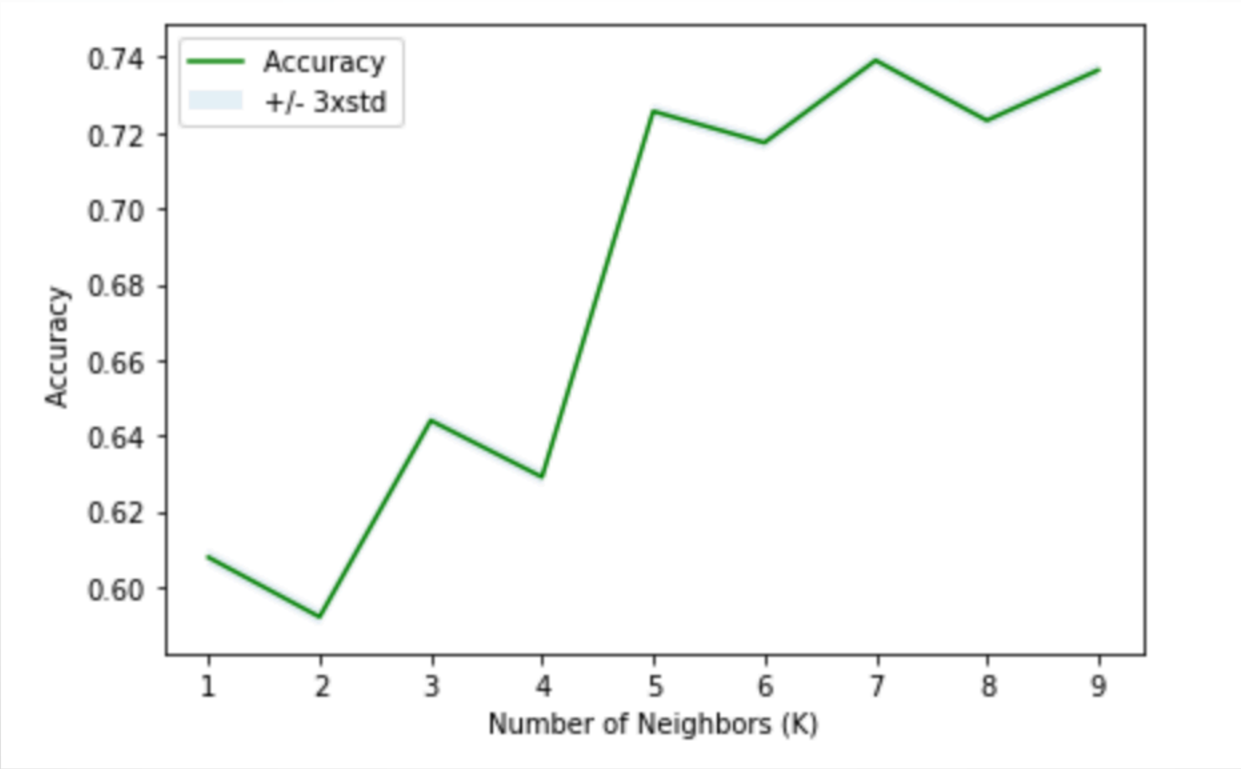
Severity Description	Train	Test	Total
Property Damage Only Collision	95,539	40,946	136,485
Injury Collision	40,732	17,456	58,188

CLASSIFICATION METHODS

- With the features selected and prepared, and the target split, the methods were tested.
- Ultimately, the Decision Tree would provide the logic that could be translated to deployment decisions
  - Other methods we used to ensure validity of the Decision Tree outcomes

Model	Training	Test
Decision Tree	0.7483543820768909	0.7507105921030102
K-Nearest Neighbor (best)	0.7390714091773011	0.7391356460395192
Logistic Regression	0.7488533877347344	0.7510530461285573

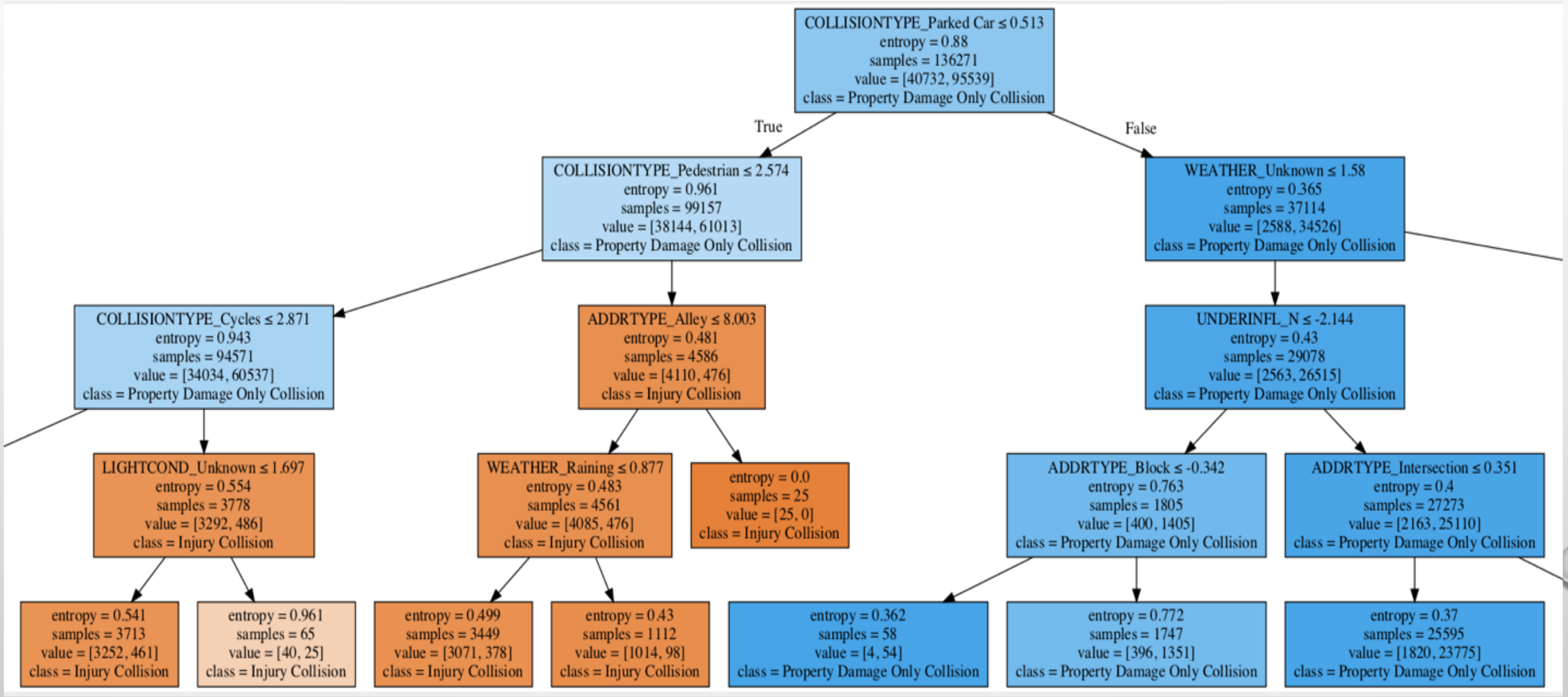
KNN MODEL RESULTS



- The best accuracy was with 0.7391356460395192 with k= 7

RESULTS

- Time to review the Decision Tree to see if there are quick wins



RESULTS

- The Decision Tree tells us two scenarios that quickly identify severe accidents
  1. Moving vehicle and a Pedestrian
  2. Moving vehicle and a Cyclist
- The underlying data shows accidents are likely to be severe 89% when involving a Pedestrian and 86% of the time when involving Cyclists
- Severe accidents did occur on other branches of the tree, of course
  - However, these deep seeded scenarios are not the type of quick wins we were trying to identify

RECOMMENDATION

- Review the existing training and procedures documents for handling reported incidents
  - How are they structured? Are they scenario based?
  - Do they already account for the findings? Trigger immediate action?
- Execute an internal reasonability check with current operators
  - Do they understand, agree and see a fit with the outcomes and proposed changes?
- Road-test the changes for a short period of time
  - Do they achieve the desired result?
- Implement and inform the public at the next town hall

PATH FORWARD

- There is plenty of room for improvement over the 75% accuracy
  - Other models and parameters
  - Additional feature engineering
  - Present for peer reviews
- Try to find a dataset that can correlate the incidents when the emergency medical services were deployed
  - Compare deployment to actual severe accidents for decision accuracy
- Try to tune staffing levels and deployment locations
  - Lack of available paramedics and departure from a distant location can contribute to arrival times