

Predicting accident severity

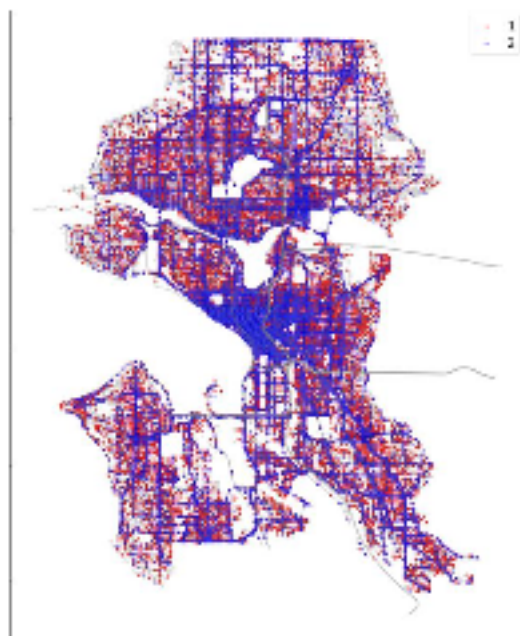
Accident severity prediction is useful for the city.

- We can see the causes of an accident being severe or not.
- We can take countermeasures to prevent these accidents from happening
- We can also look at conditions of the road and maybe use that to give drivers warnings or err to the side of caution

Data information

- Main CSV file from seattle's open data website
URL : <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>
- Seattle collision information
http://data-seattlecitygis.opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0
- Geodata and shape of Seattle

Map of Seattle and its accidents



Red represents lower severity, and blue represents higher severity.

While there are a lot of accidents throughout the entire city, a lot of blue seems to be concentrated on the center area of Seattle.

We can note that the dense area contains a lot of intersections, which probably causes a larger amount of severe accidents.

Also, most of the red dots seem to be very sporadic in where they are. All over the city there are red, and blue is mainly where roads are.

Feature checking: Weather vs severity

Clear	78858
Raining	20000
Overcast	17600
Snowing	048
Fog/Smog/Smoke	353
Other	101
Sleet/Hail/Freezing Rain	88
Blowing Sand/Dirt	38
Severe Crosswind	17
Partly Cloudy	2

Severity 1

Clear	34954
Raining	18740
Overcast	8409
Fog/Smog/Smoke	188
Snowing	162
Other	74
Sleet/Hail/Freezing Rain	27
Blowing Sand/Dirt	12
Severe Crosswind	7
Partly Cloudy	3

Severity 2

Weather data doesn't show
too many differences between
each severity level.

Feature checking: Road condition vs severity

Dry	79286
Wet	29553
Ice	885
Snow/Slush	672
Standing Water	89
Sand/Mud/Dirt	35
Oil	27

Severity 1

Dry	39883
Wet	15153
Ice	250
Snow/Slush	155
Standing Water	26
Sand/Mud/Dirt	21
Oil	28

Severity 2

The ratio of dry:wet road conditions in both are around the same. We cannot say that the road conditions have a direct impact towards one severity or the other.

Feature checking: Address type vs severity

0	76182	0 = Block	0	27988
1	34205	1= Intersection	1	26729
Severity 1			Severity 2	

Address type is where it gets really interesting. Just as proven with the map, the location of where the accident occurs has quite a bit of correlation with the level of severity. While accidents on blocks also occur in severity 2, it's much closer to a 50:50 ratio.

Modeling: imbalanced data

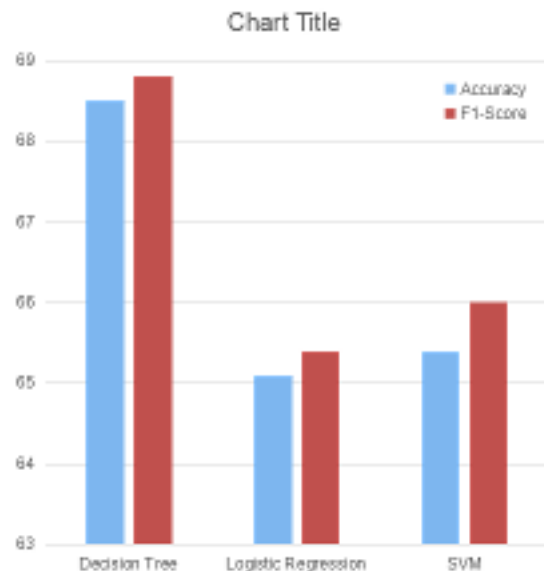
- Dataset for severity
0 is 77290
- For 1 is 38268
- We will apply
SMOTE technique
to get equal
datapoints.

{1: 77290, 2: 38268}

{1: 77290, 2: 77290}

Model predictions

Of the three models, Decision tree has the highest of both accuracy and F1-score. However, since it's only around 70%, we should look at ways to improve its accuracy.



Conclusion

- Our model isn't too bad in predicting the level of severity in an accident.
- The accuracy can use some work though.
- Some things that can improve accuracy : Better recording of date/time. Recording the driver's age or whether or not they have any pre-existing conditions or illness.
- Some ways to reduce traffic is maybe adding some extra safety measures on intersections. Some kind of way that allows you to see incoming traffic on your blind side either through mirrors or lights.