Analyzing of Seattle road collisions and predicting their severity

Tristan MOUYNA-HAINRY
September 19, 2020

1. Introduction

1.1. Background

Accidents and collisions are the burdens of transportation, even if they tend to diminish in time – with the development of new technologies – they still exist and can lead to major problems of property destruction to serious injuries and death.

For example, in the city of Seattle (population of 609k in 2010), 483k people were involved in a collision from 2004 and 2019, i.e. 4 inhabitants in 5 (if people are not involved twice and population does not change). During this period, around 190 collisions a week happened, of which 3 led to serious injuries or deaths.

Road accidents are thus an important problem for public authorities, people, and other stakeholders.

1.2. Problem

The goal of this project is to analyze Seattle road collisions and try to predict their severity. The data that might contribute to determine this might include the location of the collision, its type, whether it involves pedestrians, cyclists, other vehicles, when it occurred, and other driving conditions.

1.3. Interest

Public authorities (especially the City of Seattle) would definitely be interested in understanding in what kind of circumstances theses collisions happened, to better target how to prevent them – especially deadly ones – in a time of scarce resources. Others who can be also interested are insurers, using these insights to modify premiums accordingly, or simple citizens to be more aware of the most dangerous circumstances.

2. Data acquisition and cleaning

2.1. Data sources

I used the data from the City of Seattle website (https://data.seattle.gov/Land-Base/Collisions/9kas-rb8d). They are quite thorough and include most of the information sought.

2.2. Data cleaning

Since the objective is to predict the severity of a collision, I first removed all the instances for which the severity code was missing or equal to 0, as I was not able to label them.

Then I created a test set with a stratified shuffle split on the severity code, as the data were very imbalanced. I implemented the following data cleaning and feature selection in a pipeline so that I was able to repeat the process on the test set.

First, I replaced some values for specific features. Indeed, the ***under-influence*** feature had *'Y', '1', 'N', '0'* values. I thus replaced the *'Y'* by *'1'* and the *'N'* by *'0'*. Similarly, I replaced the *'Y', 'N'* values of the ***hit-a-parked*-*car*** feature by *'1',* and *'0'*.

Second, I binned the continuous features. These features are the coordinates ***X*** and ***Y*** of a collision. I binned them so that the collision could be located in a 100m^2 square (10m x 10m).

Third, I created new features.

Mostly, the new features are a grouping of different values for an existing feature.
Thereby, I created a ***good-weather*** feature dividing the ***weather*** feature into *'1':* good visibility (for *'Clear', 'Overcast', 'Severe Crosswind', 'Partly Cloudy'*) and *'0'*: bad visibility (for *'Raining', 'Snowing', 'Fog/Smog/Smoke', 'Sleet/Hail/Freezing Rain', 'Blowing Sand/Dirt', 'Blowing Snow'*), and *NaN* (for *NaN* and *'Other'*). I put *'Severe Crosswind'* in ***good-weather*** because it does not impaired visibility and, as there are only 22 values of them, it is not statistically important.
I did the same for the ***road condition*** feature, for which I created a ***good-road*** feature (with a *'1'* when *'Dry'*, a *'0'* when *'Wet', 'Ice', 'Snow/Slush', 'Standing Water', 'Sand/Mud/Dirt', 'Oil'*, and a *NaN* when *NaN, 'Unknown', 'Other'*).
I created a new ***light-condition*** feature in which I regrouped some of the old one's values *('Dusk'* and *'Dawn'*; *'Dark - No Street Lights'* and *'Dark - Street Lights Off'*; *'Dark - Street Lights On'* and *'Dark - Unknown Lighting'*).
I created a ***with-pedestrian*** feature, using the ***pedestrian-count*** feature, and taking the value *'0'* when no pedestrian is involved in the collision, and *'1'* otherwise.
I did the same with the ***with-cyclist*** feature.
I did similarly with the ***with-vehicle*** feature, except that this new feature is taking a *'0'* when no vehicle is involved, a *'1'* when only 1 vehicle is involved, a *'2'* when only 2 vehicles are involved and a *'3'* otherwise.
I also created a ***time***, ***month*** and ***day*** features, respectively giving the hour of the collision (from *'0'*h to *'23'*h), its month (from *'1'* to *'12'*) and the day of the week it happened (from *'0'*: Monday to *'6'*: Sunday).

Fourth, I dealt with missing values.

The feature with the most important number of missing values was ***time***. As it was more than 15% of the values, I decided to drop all the missing values for that feature.
As it was quite similar for ***good-weather***, ***good-road***, and ***light-condition***, I choose to drop all their missing values.
The ***under-influence*** feature also had missing values. As it was less than 3% and well distributed among the different severity-type collisions, I choose to replace them by the most common value.
As it was the same thing for ***X*** and ***Y*** binned, ***junction-type***, ***collision-type*** features, I choose to replace their missing values by their median or their most common values.
For the ***with-pedestrian***, ***with-cyclist*** and ***with-vehicle*** feature, the missing values are when the three of them equal 0. In that case, I choose the most common values to replace them (i.e. ***with-vehicle*** equals 2, and the others two equal 0).
For ***good-weather***, ***good-road***, and ***light-condition*** features, the missing value rate is ~15% for type-1 collisions and ~3% for the other collisions types. And, as for type-1 collisions, their most common value is overwhelmingly represented (frequency > 60%), I choose to replace their missing values by the most common value.

Finally, I created dummies variables for ***light-condition, with-vehicle, junction-type and collision-type*** features.

2.3. Feature selection

The data were initially composed of 40 features. Some of them were collisions identification numbers for the city or the state (***OBJECTID***, ***INCKEY***, ***COLDETKEY***, ***REPORTNO***, ***SDOTCOLNUM***), or meant nothing or were empty (***STATUS***, ***EXCEPTRSNCODE***, ***EXCEPTRSNDESC***), or were composed of too many missing values (***INTKEY***: >67%, ***INATTENTIONIND***: >86%, ***SPEEDING***: >95%, ***PEDROWNOTGRNT***: >97%, ***CROSSWALKKEY***: >98%, ***SEGLANEKEY***: >98%). I, therefore, dropped all these features.

The rest of the features could be split into 6 different groups: space-location, time-location, type of collision, exterior conditions, intrinsic conditions, and of course severity features.

Obviously, the severity-related features (***SEVERITYCODE***, ***SEVERITYDESC***, ***INJURIES***, ***SERIOUSINJURIES***, ***FATALITIES***) were only kept as a target feature in the following computation, and I only used the ***SEVERITYCODE*** feature for that.

Thereby, the features I kept for the analysis are the following:

- Space-Location: ***X, Y***, ***junction-type***
- Time-Location: ***time***, ***month***, ***day***
- Type of collision: ***collision-type***, ***with-pedestrian***, ***with-cyclist***, ***with-vehicle***, ***hit-a-parked-car***
- Exterior conditions: ***good-weather***, ***good-road, light-condition***
- Intrinsic conditions: ***under-influence***

I dropped ***ST_COLCODE***, ***ST_COLDESC*** and ***SDOT_COLCODE, ST_COLDESC*** because they gave the same information as the type of collision's ones, had more missing values and were not completely coherent with one another.

I also dropped the ***ADDRTYPE*** and ***LOCATION*** features because they gave the same information as the space-location's ones.

I finally did not use ***PERSONCOUNT*** because it is mixing many things in it and it is difficult to determine a policy based on this feature.

3. Exploratory data analysis

3.1. The target data and their consistence

The target data (***SEVERITYCODE*** feature) are composed of 4 different classes: 1, 2, 2b, and 3, for *Property Damage Only Collision*, *Injury Collision*, *Serious Injury Collision*, and *Fatality Collision* respectively. Thereafter, I would only refer to type-1, type-2, type-2b and type-3 collisions.

SEVERITYCODE	Repartition of collisions by type
1	68.9%
2	29.4%
2b	1.6%
3	0.2%

Table 1 - Repartition of collisions by severity-type

The data are very imbalanced (Table 1). Nevertheless, they roughly keep the same distribution through the years (even if we can see a decline over years), enabling us to use all these data for our analyses.

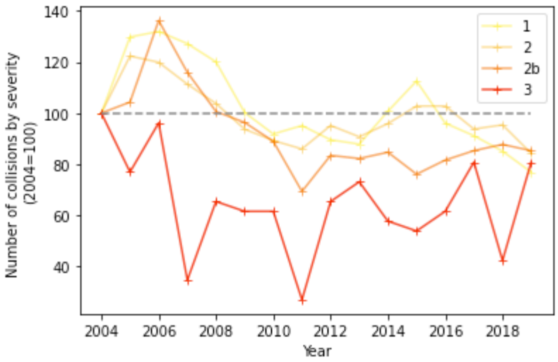
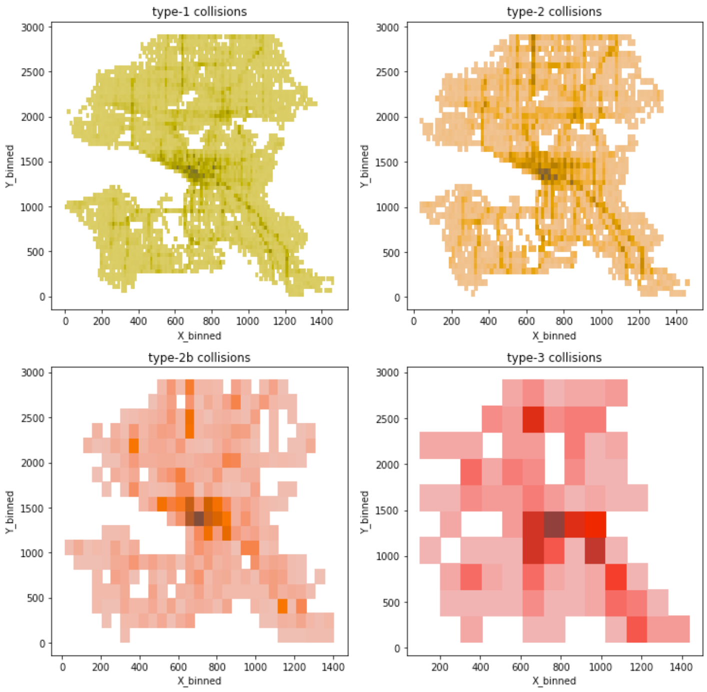


Figure 1 - Evolution, by severity, of the number of collisions since 2004 (2004=100)

3.2. As we might think, collisions happen where there is traffic

3.2.1. Because of the place

If we plot the two-dimensional histogram (with ***X_binned*** and ***Y_binned***) of the collisions, we can see that they happen nearly everywhere, but that the main streets, big avenues, highest dense locations (like downtown) are more susceptible to present more collisions (cf. Figures 2 and 3). We can also see that the repartition is quite similar for all severity-types.



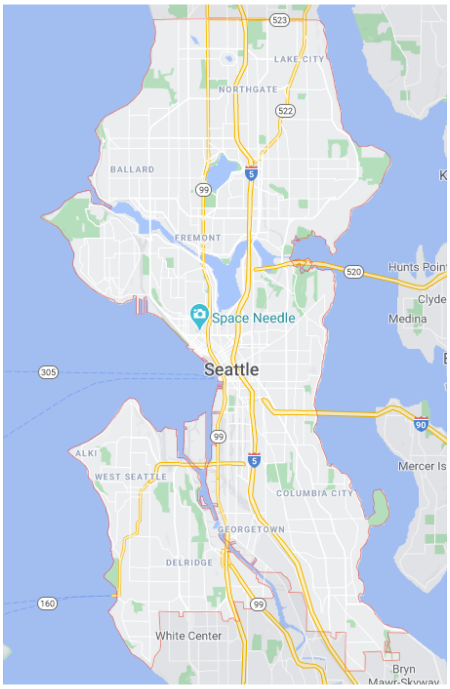


Figure 2 - Repartition, by severity, of the collisions according to X and Y (the darker, the more)

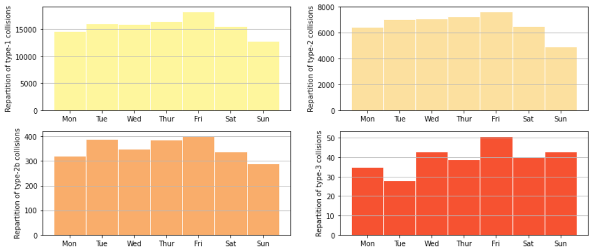
Figure 3 - Google map of Seattle

3.2.2. Or because of the moment

If we look for the collisions by day and time, we can see that the moment matters.

Indeed, there are less collisions on the weekend (Figure 4), because people don't go to work and thus are not moving roughly at the same time. The fact that the distribution is not uniform is confirmed by a chi-squared test for all the severity-types, except for the type-3 (due to the small number of data). Moreover, we can see that the distribution is roughly alike between the different severity-types: this feature is probably a poor predictor for severity.

Similarly, the number of collisions depends of the hour of the day (Figure 5). When there is more traffic (from 6h to 17h) it grows, then drops quickly to lower levels (from 20h to 3h) when most people are at home, and to a low peak from 3h to 5h in the morning when very few people are out. We can see that the distribution is not the same between type-1/2 and type-2b/3, for which there are peaks around midnight and 21h.



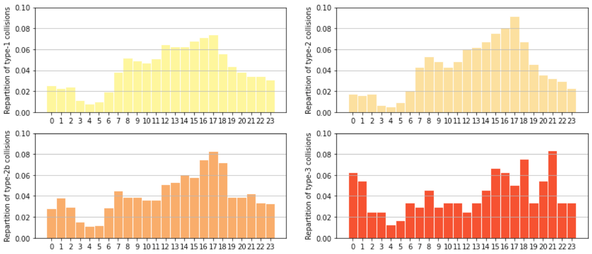


Figure 4 - Repartition, by severity, of the number of collisions by day

Figure 5 - Repartition, by severity, of the collisions by hour (0h to 23h)

3.3. Collisions happen also when people pay less attention

3.3.1. Because they are tired or under influence

If we look again to collisions by day (Figure 4), the Friday-peak can also be interpreted by the fact that people are tired of the working-week and more susceptible to have accidents. That would explain the 21h-peak (Figure 5) for type-2b/3, for late (and thus tired) workers coming home.

Likewise, driving under influence certainly increases the chance of collisions and the severity of these ones (Table 2). Indeed, in 2014, 11.1% of the US population had driven at least once under influence[1]. People certainly are not under influence each time they drive. If we assume that they are driving less than one in two time under influence, it means that the rate of rides under influence is certainly under 5%, which is the rate of collisions under influence.

SEVERITYCODE	Collisions rate under influence
1	4.2%
2	6.1%
2b	13.6%
3	28.9%

Table 2 - Collisions rate, by severity-type, under influence

3.3.2. Or, more surprisingly, because driving conditions are good

If we look at driving conditions, particularly weather and road conditions, the better they are, the more collisions (Table 3). Indeed, collisions happened mostly during good weather (and with good road conditions, as they are linked). As it is raining on average 152 days a year in Seattle[2] (~42% of the time), the difference cannot be explained by less frequent bad weather.

SEVERITYCODE	Collisions rate with good weather	Collisions rate with good road condition
1	80.2%	71.5%
2	79.3%	71.0%
2b	82.1%	74.4%
3	84.0%	79.2%

Table 3 - Collisions rate, by severity-type, with good weather and good road conditions

SEVERITYCODE	Rate of missing values for **GOOD_WEATHER**	Rate of missing values for **GOOD_ROAD**
1	13.8%	13.4%
2	3.4%	3.1%
2b	3.3%	2.8%
3	3.9%	3.6%

Table 4 - Rate of missing values, by severity-type, for the features GOOD_WEATHER and GOOD_ROAD

Furthermore, most severe collisions happen more frequently with good road conditions. It is less clear though with good weather, because the rates of missing values (Table 4) for ***GOOD_WEATHER*** are of the same order of magnitude as the difference in collisions rates.

[1] cf. https://www.samhsa.gov/data/sites/default/files/report_2688/ShortReport-2688.html

[2] cf. https://weather.com/science/weather-explainers/news/seattle-rainy-reputation

The same results are seen with light conditions (Figure 6). Collisions mostly happen during daylight. It is not due to the fact that there is less nighttime than daytime in Seattle, as for six months sun rises after 7h and for 4 months sun sets before 18h[3]. So, this situation does not explain the difference in the distributions. Conversely very few collisions happen during dark with no lights. But this situation is probably very infrequent in a city like Seattle, so the few collisions are most certainly due to the rare situation.

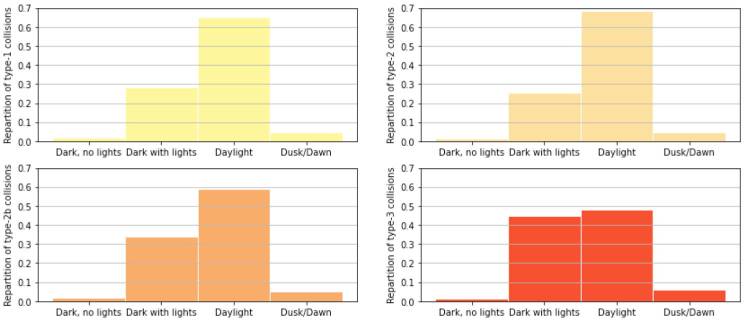


Figure 6 - Repartition, by severity, of the collisions by light conditions

Nevertheless, when there are poor driving conditions, people tend to be more careful, to pay more attention, to drive slower, or even to don't move (e.g. in a blizzard), leading to the fact that collisions happen more while driving conditions are good.

3.4. Other features that explain the collisions and their severity

3.4.1. A monthly look

If we look at the monthly weighted[4] repartition of the collisions (Figure 7), we can see a peak of collisions during June and October, and a low in December. This statement still holds when looking by severity types (except for type-3). I performed a chi-squared test to test if the weighted repartitions were uniform or not, to know if the peaks and lows were statistically significant. For types-1/2/2b the repartitions were not uniform (at 95%), for type-3 the test was not conclusive due to the few numbers of data.

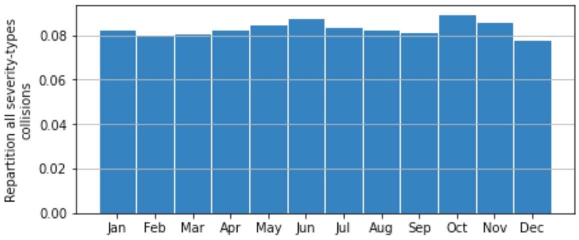


Figure 7 - Weighted repartition (all severity-types) of the collisions by months

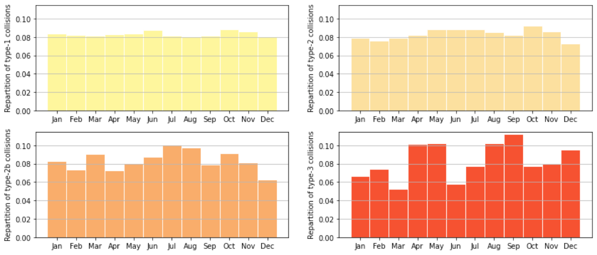


Figure 8 - Weighted repartition, by severity, of the collisions by months

[3] Cf. https://www.sunrise-and-sunset.com/en/sun/united-states/seattle

[4] to take into account the number of days in each month

3.4.2. Severity predictors

We have already seen some features that might be good predictors. Others are essentially the kind of collision.

The number of pedestrians involved in a collision is a very good predictor for its severity. Indeed, in type-1 collisions, mostly no pedestrian is involved, whereas more than 40% of type-3 collisions involved 1 or more pedestrians (Figure 9).

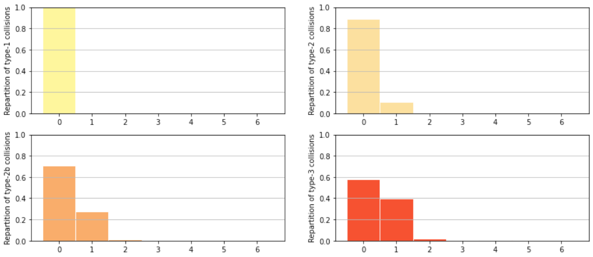


Figure 9 - Repartition, by severity, of the collisions by number of pedestrians involved

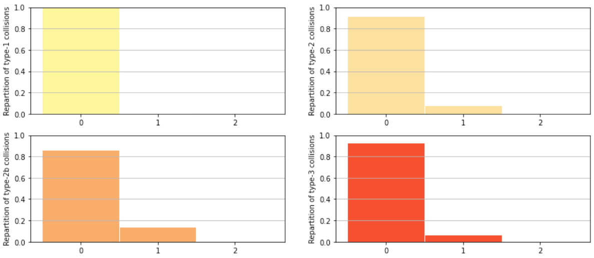


Figure 10 - Repartition, by severity, of the collisions by number of cyclists involved

In mostly 100% of type-1 collisions, no cyclist is involved, but it changes with the severity (Figure 10).

Besides, 80% of type-1 collisions involve only 2 vehicles and less than 10% of them involve 1. With increase in severity, it progressively shifts to the ratio 65%/20% with 1 to 2 vehicles involved (Figure 11).

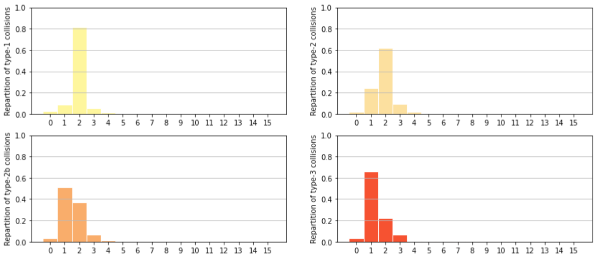


Figure 11 - Repartition, by severity, of the collisions by number of vehicles involved

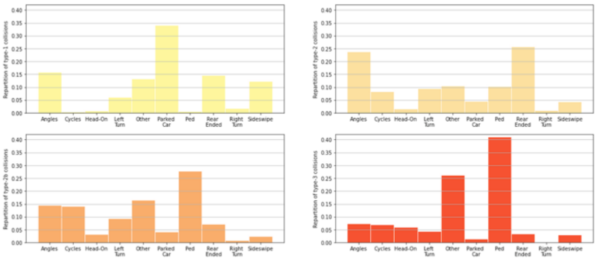


Figure 12 - Repartition, by severity, of the collisions by its nature

Similarly, the nature of the collisions is also a good predictor of its severity. The most common type-1 collisions happen with parked car (35% of them), whereas parked-car collisions are less than 5% in types-2/2b/3. But with severity increases the proportion of pedestrians involved (Figure 12).

In order to have more insights on severity, and to try to prevent the most severe collisions, I built models to predict it.

4. Predicting the severity of collisions

4.1. Methodology

One of the main issues I had was that the data were imbalanced. And the class I wanted to predict (type-3 collisions) did not have enough data in comparison to types-1/2.

So, I first choose to aggregate type-2b and type-3 collisions in a new class: a new type-3. It is not very problematic, because we want to be able to find patterns to prevent the most severe collisions, and that includes both fatalities and severe injuries. Thereby, I only worked on three classes of severity-type collisions: type-1 (only property damages), type-2 (collisions with injuries), and type-3 (collisions with serious injuries and deaths).

Second, I rebalanced training data, by randomly choosing within the data of classes 1 and 3 the same number of data than class-2 data. Thus, I increased the number of class-3 data, and I reduced the number of class-1 data.

Finally, I had to create a new cross-validation function, which could rebalance the training data, but keep the validation data imbalanced.

As my main objective was to predict – in order to prevent – type-3 collisions with some global accuracy, I chose two metrics to evaluate my models. The first metric was the weighted F1-score, because it is a good metrics for classification, especially with imbalanced data. The second metric was the recall-score for the class-3, as I mostly want to predict all the type-3 collisions, but do not really care to wrongly predict this outcome for other classes. Naturally, my objective was to have the highest possible values for the two scores.

4.2. Results

I have built 8 different models and tuned them with my cross-validation function to find the best hyperparameter (cf. Appendix I).

	cross-validation step			**generalization step**	
Model (hyperparameter)	**Best value of the hyperparameter**	**Weighted F1-score**	**Recall-score for the class-3**	**Weighted F1-score**	**Recall-score for the class-3**
SGDClassifier (alpha)	0.16	62.2%	64.3%	62.3%	63.0%
DecisionTree (max_depth)	7	61.5%	55.2%		
(max_leaf_nodes)	27	61.7%	60.7%	61.0%	57.2%
RandomForest (max_leaf_nodes)	31	61.1%	63.2%		
ExtraTreesClassifier (max_leaf_nodes)	26	62.2%	63.1%		
LogisticRegression (C)	0.22	61.1%	62.8%		
KNNeighbors (n_neighbors)	24	60.5%	53.9%		
VotingClassifier (voting)	hard	61.2%	64.2%		

Table 5 - Scores for the different models built

Then I computed the corresponding weighted F1-score and Recall-score for the class-3. The results are shown in Table 5.

The VotingClassifier model was built with the SGDClassifier, the ExtraTreesClassifier (the best model for the ‘tree’ family) and the LogisticRegression models.

We can see that the best model (with both highest scores for the two metrics considered) is the SGDClassifier. Moreover, it generalizes well when tested on the test set, because the two metrics stay roughly the same (Table 5).

Nevertheless, as the DecisionTree is a very easy to understand model, I would also pick it to explain the severity of collision. Thereby, it generalizes not badly, except for the recall-score for the class-3, which is poorer on the test set than on the train set.

4.3. Discussion

4.3.1. With the best model (SGDClassifier)

We can determine the impact of each features on the outcome – in that case: class-3 prediction – and discuss their magnitude (Figure 13).

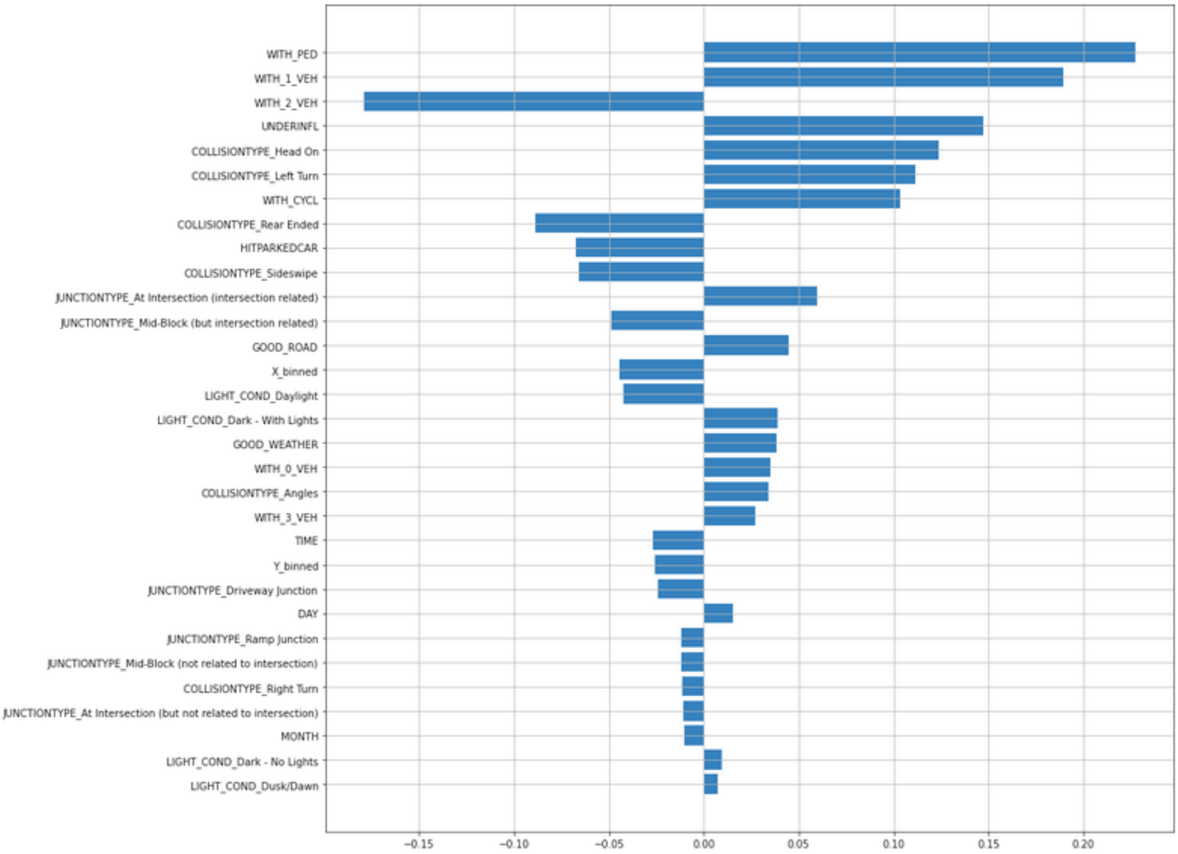


Figure 13 - Impact of each features on the class-3 prediction, sorted by magnitude impact, with the SDGClassifier

A positive magnitude means that the feature has a positive relation with the class-3 prediction. Conversely, a negative magnitude means that the feature has a negative relation with the class-3 prediction.

We see that the most impacting feature is whether the collision involves pedestrians or not, then if 1 vehicle is involved and then if 2 vehicles are involved (in that case it reduces the chance for a type-3 prediction). Besides, if a collision involves a cyclist, it also increases its chance of being severe.

Then, we see that if a collision happens head-on or left turn, it increases the chance of being severe. Conversely, if a collision happens rear-end or sideswipe, it decreases the chance of being severe.

Furthermore, we see that if a collision happens on the intersection and is intersection related, it increases the chance of being severe. Conversely, if a collision happens mid-block but is intersection related, it decreases the chance of being severe. Probably because, on the second case, it would only involve cars and not pedestrians, and would be due to a lack of visibility.

Thus, to prevent severe collision, it would be important to protect pedestrians and cyclists, by improving the safety of sidewalks, creating specific cycle paths. Second, preventing head-on collisions would also prevent most severe collisions. As they are probably due to high speed, putting speed radars could be helpful. Similarly, to improve the safety of intersections, it might be impactful to put radars there.

4.3.2. With the DecisionTree

As decision trees are good models to elaborate policies, I will discuss this model, even if it is not the best I trained.

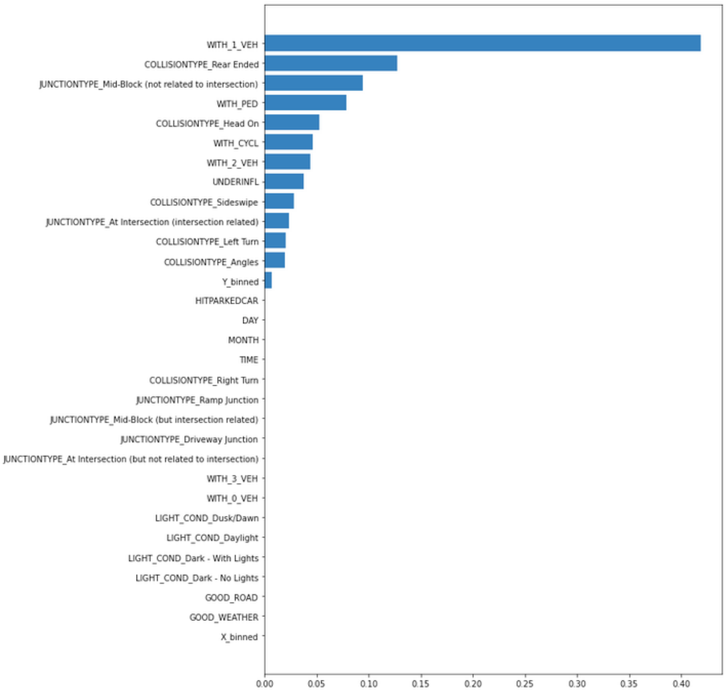


Figure 14 - Importance of each features on the class-3 prediction, with the DecisionTree

Looking at the importance of features (Figure 14) and the corresponding tree of decisions (Appendix II), we see that they are coherent with the results we found with the SGDClassifier. Indeed, we see that a collision involving only one vehicle is a good predictor for a severe one.

Then, we see that, when a pedestrian or a cyclist is involved, a severe collision is more likely. Same thing when the collision is head-on.

The thing that the decision tree brings – compared to the SGDClassifier – is that we can see that the serious head-on collisions happen mostly in Mid-block (not related to intersection), leading to the proposition of better separating the two ways.

5. Conclusion

In this project, I studied the road collisions in Seattle and tried to predict their severity, in order to propose policies to prevent them the most possible and, above all, the most severe ones.

We saw that collisions happen: • naturally, where and when there is more traffic: downtown, on highways, during the week and less during the weekend, between 8h and 19h; • when people are tired: on Friday, at 17h; • when there are good driving conditions: good visibility weather, good road conditions, enough light; • at some time in the year: on June, October and November.

We can thus propose the road police to favor those moments and locations, rather than small roads, weekends or December, because their presence would be more impacting.

We also saw that collisions are the most severe when: • pedestrians or cyclists are involved, • when only one vehicle is involved, • with more than one vehicle, when it is head-on, • when it happens in intersections.

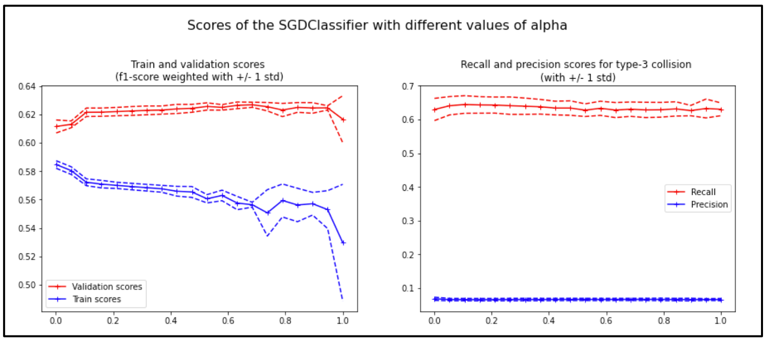
We can thus propose for the police to focus more on these situations, or/and, to create or develop urban amenities: • improve the safety of sidewalks, • create specific cycle paths, • develop speed radars on head-on collision-prone blocks, • improve the safety of intersections (e.g. with radars).

6. Future directions

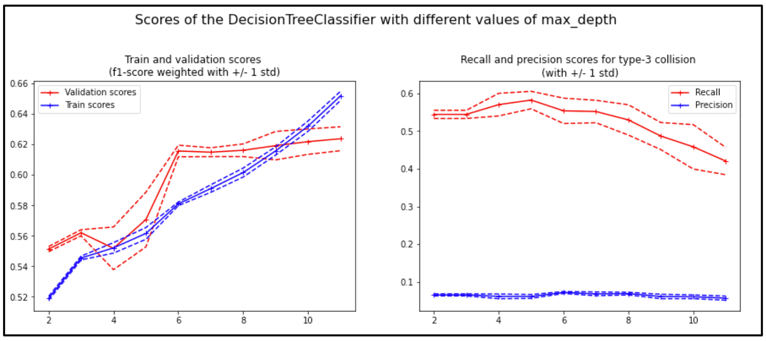
Even my best model did not predict all the severity cases (with a score of 62%). That means that more features could help to better explain the severity of a collision. We could think about the characteristics of the cars involved (old and small cars probably lead to more severe collisions), characteristics of the people involved, the traffic situation…

Furthermore, it could be interesting to do the study with other cities to have more insights on data.

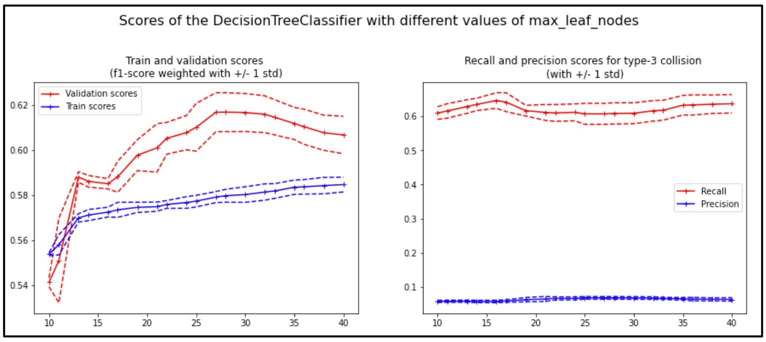
Appendix I – fine-tuning of hyperparameters for the different models



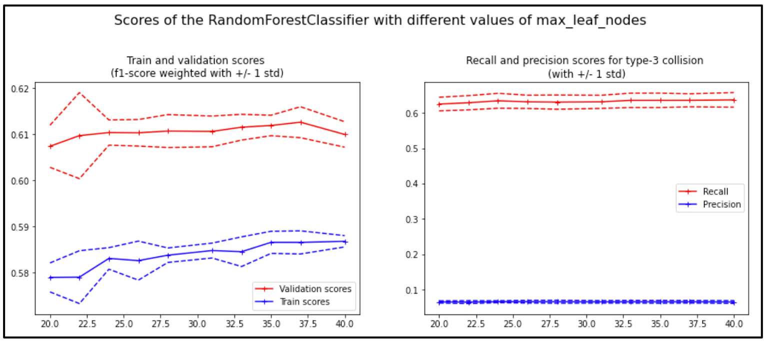
Scores of the SGDClassifier with different values of alpha



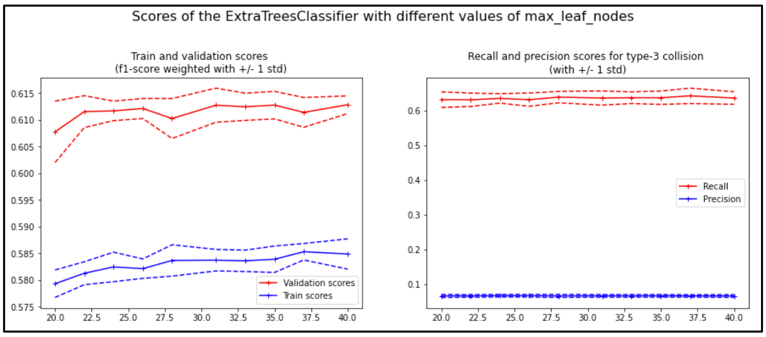
Scores of the DecisionTreeClassifier with different values of max_depth

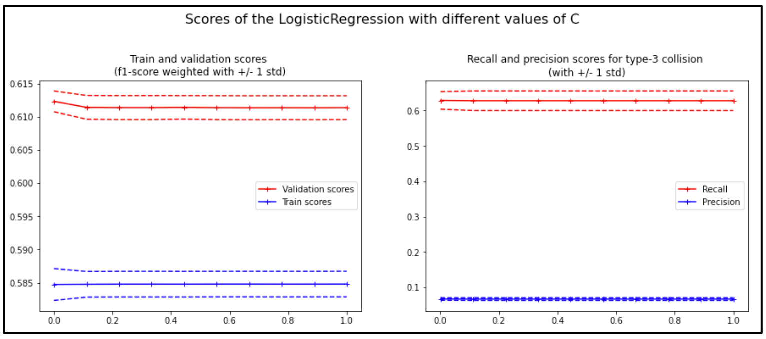


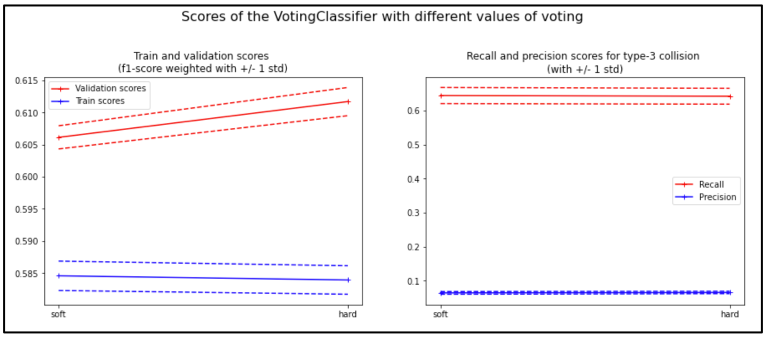
Scores of the DecisionTreeClassifier with different values of max_leaf_nodes



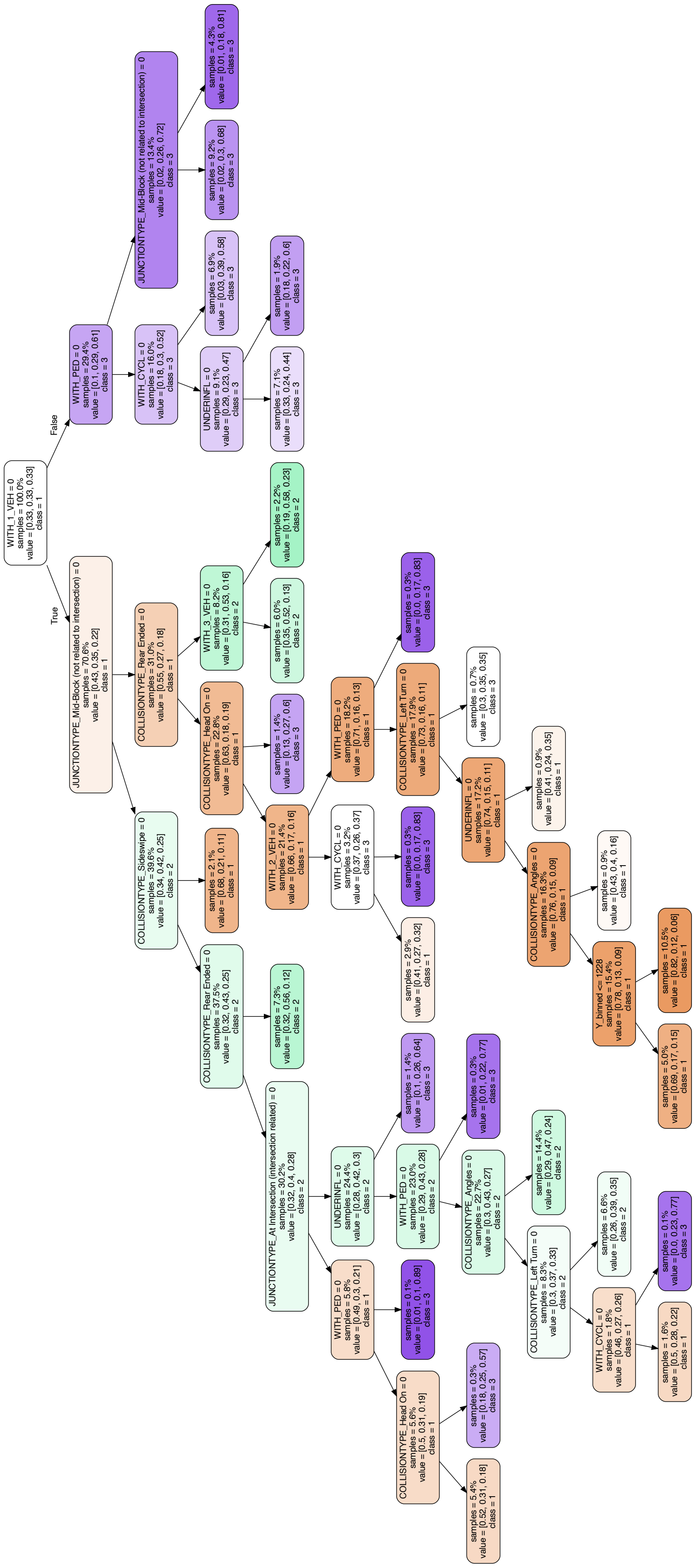
Scores of the RandomForestClassifier with different values of max_leaf_nodes


Scores of the ExtraTreesClassifier with different values of max_leaf_nodes
Train and validation scores
(f1-score weighted with +/- 1 std)
0.615
0.610
0.605
0.600
0.595
0.590
0.585
0.580
0.575
20.0
22.5
25.0
27.5
30.0
32.5
35.0
37.5
40.0
Validation scores
Train scores
Recall and precision scores for type-3 collision
(with +/- 1 std)
0.6
0.5
0.4
0.3
0.2
0.1
Recall
Precision


Scores of the LogisticRegression with different values of C
Train and validation scores
(f1-score weighted with +/- 1 std)
0.615
0.610
0.605
0.600
0.595
0.590
0.585
0.0
0.2
0.4
0.6
0.8
1.0
Validation scores
Train scores
Recall and precision scores for type-3 collision
(with +/- 1 std)
0.6
0.5
0.4
0.3
0.2
0.1
Recall
Precision


Scores of the VotingClassifier with different values of voting
Train and validation scores
(f1-score weighted with +/- 1 std)
0.615
0.610
0.605
0.600
0.595
0.590
0.585
soft
hard
Validation scores
Train scores
Recall and precision scores for type-3 collision
(with +/- 1 std)
0.6
0.5
0.4
0.3
0.2
0.1
Recall
Precision

Appendix II: the tree of decisions


WITH_1_VEH = 0
samples = 100.0%
value = [0.33, 0.33, 0.33]
class = 1
True
False
JUNCTIONTYPE_Mid-Block (not related to intersection) = 0
samples = 70.6%
value = [0.43, 0.35, 0.22]
class = 1
COLLISIONTYPE_Rear Ended = 0
samples = 31.0%
value = [0.55, 0.27, 0.18]
class = 1
WITH_3_VEH = 0
samples = 8.2%
value = [0.31, 0.53, 0.16]
class = 2
samples = 2.2%
value = [0.19, 0.58, 0.23]
class = 2
samples = 6.0%
value = [0.35, 0.52, 0.13]
class = 2
COLLISIONTYPE_Head On = 0
samples = 22.8%
value = [0.63, 0.18, 0.19]
class = 1
samples = 1.4%
value = [0.13, 0.27, 0.6]
class = 3
WITH_2_VEH = 0
samples = 21.4%
value = [0.66, 0.17, 0.16]
class = 1
WITH_PED = 0
samples = 18.2%
value = [0.71, 0.16, 0.13]
class = 1
samples = 0.3%
value = [0.0, 0.17, 0.83]
class = 3
COLLISIONTYPE_Left Turn = 0
samples = 17.9%
value = [0.73, 0.16, 0.11]
class = 1
samples = 0.7%
value = [0.3, 0.35, 0.35]
class = 3
UNDERINFL = 0
samples = 17.2%
value = [0.74, 0.15, 0.11]
class = 1
samples = 0.9%
value = [0.41, 0.24, 0.35]
class = 1
COLLISIONTYPE_Angles = 0
samples = 16.3%
value = [0.76, 0.15, 0.09]
class = 1
samples = 0.9%
value = [0.43, 0.4, 0.16]
class = 1
Y_binned <= 1228
samples = 15.4%
value = [0.78, 0.13, 0.09]
class = 1
samples = 10.5%
value = [0.82, 0.12, 0.06]
class = 1
samples = 5.0%
value = [0.69, 0.17, 0.15]
class = 1
WITH_CYCL = 0
samples = 3.2%
value = [0.37, 0.26, 0.37]
class = 3
samples = 0.3%
value = [0.0, 0.17, 0.83]
class = 3
samples = 2.9%
value = [0.41, 0.27, 0.32]
class = 1
COLLISIONTYPE_Sideswipe = 0
samples = 39.6%
value = [0.34, 0.42, 0.25]
class = 2
samples = 2.1%
value = [0.68, 0.21, 0.11]
class = 1
COLLISIONTYPE_Rear Ended = 0
samples = 37.5%
value = [0.32, 0.43, 0.25]
class = 2
samples = 7.3%
value = [0.32, 0.56, 0.12]
class = 2
JUNCTIONTYPE_At Intersection (intersection related) = 0
samples = 30.2%
value = [0.32, 0.4, 0.28]
class = 2
UNDERINFL = 0
samples = 24.4%
value = [0.28, 0.42, 0.3]
class = 2
samples = 1.4%
value = [0.1, 0.26, 0.64]
class = 3
WITH_PED = 0
samples = 23.0%
value = [0.29, 0.43, 0.28]
class = 2
samples = 0.3%
value = [0.01, 0.22, 0.77]
class = 3
COLLISIONTYPE_Angles = 0
samples = 22.7%
value = [0.3, 0.43, 0.27]
class = 2
samples = 14.4%
value = [0.29, 0.47, 0.24]
class = 2
COLLISIONTYPE_Left Turn = 0
samples = 8.3%
value = [0.3, 0.37, 0.33]
class = 2
samples = 6.6%
value = [0.26, 0.39, 0.35]
class = 2
WITH_CYCL = 0
samples = 1.8%
value = [0.46, 0.27, 0.26]
class = 1
samples = 0.1%
value = [0.0, 0.23, 0.77]
class = 3
samples = 1.6%
value = [0.5, 0.28, 0.22]
class = 1
WITH_PED = 0
samples = 5.8%
value = [0.49, 0.3, 0.21]
class = 1
samples = 0.1%
value = [0.01, 0.1, 0.89]
class = 3
COLLISIONTYPE_Head On = 0
samples = 5.6%
value = [0.5, 0.31, 0.19]
class = 1
samples = 0.3%
value = [0.18, 0.25, 0.57]
class = 3
samples = 5.4%
value = [0.52, 0.31, 0.18]
class = 1
WITH_PED = 0
samples = 29.4%
value = [0.1, 0.29, 0.61]
class = 3
JUNCTIONTYPE_Mid-Block (not related to intersection) = 0
samples = 13.4%
value = [0.02, 0.26, 0.72]
class = 3
samples = 9.2%
value = [0.02, 0.3, 0.68]
class = 3
samples = 4.3%
value = [0.01, 0.18, 0.81]
class = 3
WITH_CYCL = 0
samples = 16.0%
value = [0.18, 0.3, 0.52]
class = 3
samples = 6.9%
value = [0.03, 0.39, 0.58]
class = 3
UNDERINFL = 0
samples = 9.1%
value = [0.29, 0.23, 0.47]
class = 3
samples = 1.9%
value = [0.18, 0.22, 0.6]
class = 3
samples = 7.1%
value = [0.33, 0.24, 0.44]
class = 3