



Accidents in Seattle – A Machine Learning approach



Introduction

- In Seattle exists a lot of accidents by year.
- That include a various collision types, time of the event people involved, etc
- We need to understand which variables influence in accidents so we can take actions..
- We want to predict the "Severity" (property or injury damage) of an accident

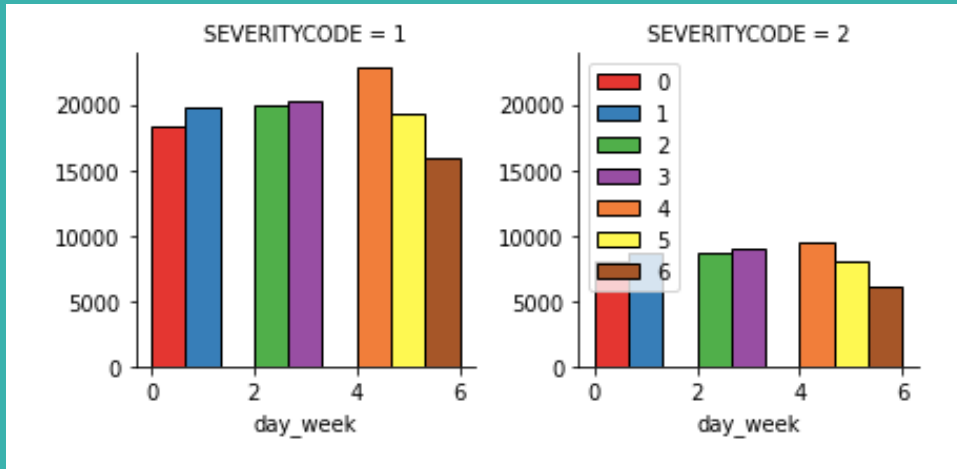


Data

- I have around 200.000 rows of accidents data in Seattle.
- I eliminated NaN values that represents the 2% of the total data
- The dataset was unbalanced (there more "Property damages" than "Injury damages"). So I balanced working with the same proportion.

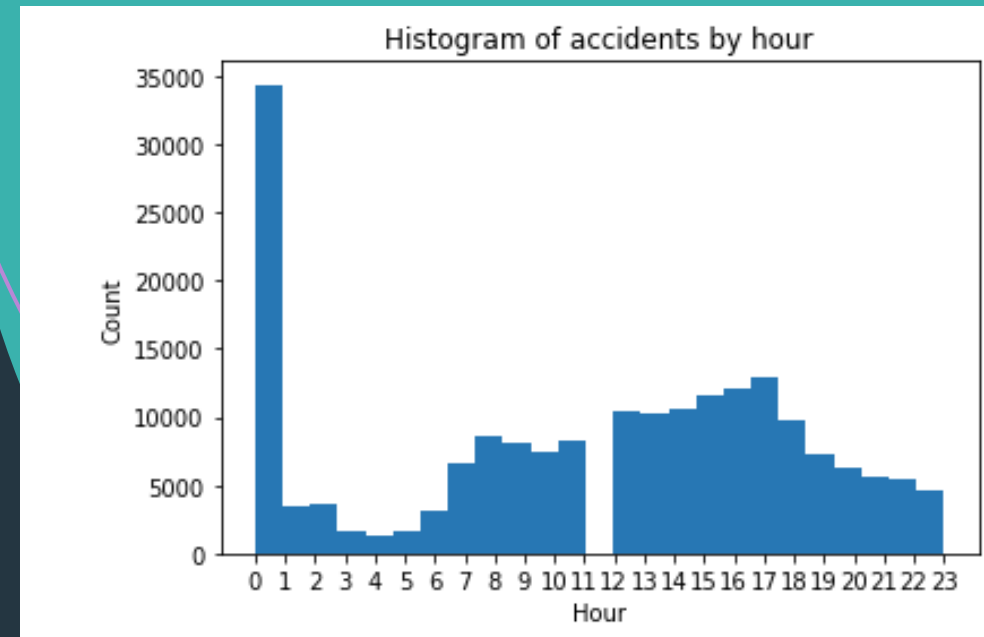
Data

- I found the behaviour of the severity types, doesn't change too much between days of the week and weekends.



Data

- Also, there are more accidents in the midnight (like four or six times more than other hours). I prefer take conservative look, because this may be weird.



Algorithms – Machine Learning

- I worked with four algorithms
 1. KNN
 2. Decision Tree
 3. Support Vector Machine
 4. Logistic Regression

Algorithms – Machine Learning

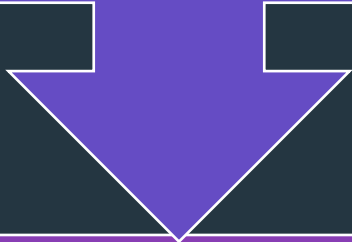
- The performance I measured it by
 1. Jaccard index
 2. F1-Score
 3. Log Loss

Results

Algorithm	Jaccard	F1-Score	Log Loss
KNN	0.65	0.64	-
Decision Tree	0.70	0.70	-
SVM	0.66	0.65	-
Logistic Regression	0.53	0.53	0.67

Conclusion

Decision tree has the best performance and is the best model



A 0.7 of performance reflects a good parameter. So we can predict with well certainty the severity of an accident

Conclusion



WE CAN MAKE DECISIONS!



IF SOMEONE DECIDE
INVEST IN THE STREET WE
CAN PREDICT THE
SEVERITY OF AN
ACCIDENT BASED IN THE
CHANGE OF THE
VARIABLE.



CONTROL AND
INSPECTION IN ROADS:
¿HOW MUCH INVEST IN
ROAD CONTROLS?
(BENEFITS OF CONVERT
INTO SEVERITY 2 TO 1
AND COMPARE WITH THE
COST OF INSPECCION).



Thanks for
your time!