



# Predicting the Severity of Car Accidents in Seattle

Presentation by Ryan Bruyninckx

For the Coursera IBM Data Science Capstone Project

Presented on: October 11, 2020



# Introduction

- Seattle, Washington, has a notorious reputation as being a wet and rainy city.
- Seattle experiences an average of 152 rainy days a year\* -- more than most North American cities.
- Are there more car accidents during bad weather?
- And how can we predict the severity of these car accidents?

\*"Seattle's Rainy Reputation Is Well-Deserved", The Weather Channel, October 14, 2016;  
<https://weather.com/science/weather-explainers/news/seattle-rainy-reputation>





# The Questions

## **Is There A Correlation between Weather and Car Accidents?**

Using extensive accident data gathered from the Seattle Department of Transportation, we will predict the severity of such accidents under various weather conditions and make a determination if such factors as weather, road and light condition, have an effect on the frequency of car accidents, as well as the severity of them.

## **Who Would Be Interested in These Findings?**

The assessment and predictions to be performed here in this project shall be of interest to motorists of Seattle streets. A motorist could use these predictions and analyses to plan their travels around certain weather, road, and lighting conditions.



# The Data

We will be using a data set published by the Seattle Department of Transportation (SDOT) titled "Collisions - All Years" that contains 194,673 collisions that occurred between January 1, 2004, and May 20, 2020. The collisions were provided by the Seattle Police Department and recorded by the SDOT's Traffic Records group. A sample of the data is below:

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





# The Data

We will be focusing on the following fields to find out our answers:

- Severity Code = Ranked 1 for no injury to 4 for fatality
- Weather = The main weather phenomenon at the time of the accident
- Road Condition = Was the road wet or dry at the time?
- Light Condition = Was it dark or was there daylight?

	SEVERITYCODE	WEATHER	ROADCOND	LIGHTCOND
0	2	Overcast	Wet	Daylight
1	1	Raining	Wet	Dark - Street Lights On
2	1	Overcast	Dry	Daylight
3	1	Clear	Dry	Daylight
4	2	Raining	Wet	Daylight



# Data Methodology

Some Highlights of Our Data Processing:

- 194,673 accidents occurred between January 1, 2004, and May 20, 2020, as published by Seattle Dept of Transportation.
- 24,716 records had incomplete or missing data, so those were purged from our analysis.
- We kept our focus on three Weather Condition types: Clear, Overcast, and Raining – this accounted for 99% of all accidents.
- We combined smaller, similar categories under Light Conditions to all fall under either Daylight or Dark skies.
- We combined smaller, similar categories under Road Conditions to all fall under either Dry or Wet road conditions.
- We transformed our categorical data into a numerical representation to complete our modeling.





# Data Modeling

We tested our data to make predictions on severity by using several data model algorithms:

- Decision Tree
- Support Vector Model (SVM)
- Logistic Regression
- K-Nearest Neighbor (KNN)



# Data Results

We tested the accuracy of our models, fine-tuned them, and calculated the maximum accuracy of our predictions by comparing the test results to our actual results on car accident severity and using :

- Jaccard Similarity Index
- F1-Score
- Logistic Loss





# Data Results

Algorithm	Jaccard	F1-Score	Log Loss
Logistic Regression	0.6712	0.5391	0.6328
KNN	0.6396	0.5714	N/A
Decision Tree	0.6715	0.5395	N/A
SVM	0.6712	0.5391	N/A



# Discussion

Based on our modeling, we were only able to predict the severity of car accidents accurately up to 67% of the time using either Decision Tree, Support Vector Model, or Logistic Regression. KNN modeling had a 64% accuracy.

Based on the analysis we completed, the current weather, road conditions, and lighting conditions don't successfully determine the severity of the car accidents.

However, we could consider for future analysis a look at other data fields provided, such as location or speeding, to determine if severity can be predicted using those values.





# Conclusion

- There was no definitive correlation predicting the severity of a car accident when factoring in weather, road and lighting conditions.
- 100% of recorded car accidents in the city of Seattle resulted in either no injuries (severity code = 1) or very minor injuries (severity code = 2). No accidents with substantial injuries or fatalities were recorded.
- More accidents occurred during clear weather conditions than in conditions where precipitation was falling.
- 64.6% of all car accidents in the 16-year period of data occurred during clear weather.
- 19.4% of all accidents occurred during rainy weather.
- The likelihood of getting into an accident on a clear day in Seattle is 3.3 times greater than that on a rainy day.
- Despite the inconclusive analysis, always drive safe, obey the speed limit, and keep your distance between yourself and other vehicles.



Thank You