Classification of Accident Severity

A Metrics Based Approach

Introduction

- Road safety is a primary concern for drivers since traffic accidents can be fatal.
- Without data, drivers rely on their intuitive perception of the environment to determine the level of cautiousness when they drive, which can be inaccurate at times.
- For example, drivers may underestimate the danger of speeding on a wet road even after the rain has stopped, or overestimate the risks of driving with light fog.

Introduction

Aim: Classify severity of traffic accidents using relevant features

- Therefore, it is important to educate drivers with correct road safety knowledge that is backed-up by empirical evidence, to boost driving safety and confidence.
- A multi-class model that classifies accidents of various degrees of severity according to different road conditions can inform drivers of the relevant risks so that they can adjust their level of cautiousness when driving.

Data

Shared dataset: Seattle city collisions

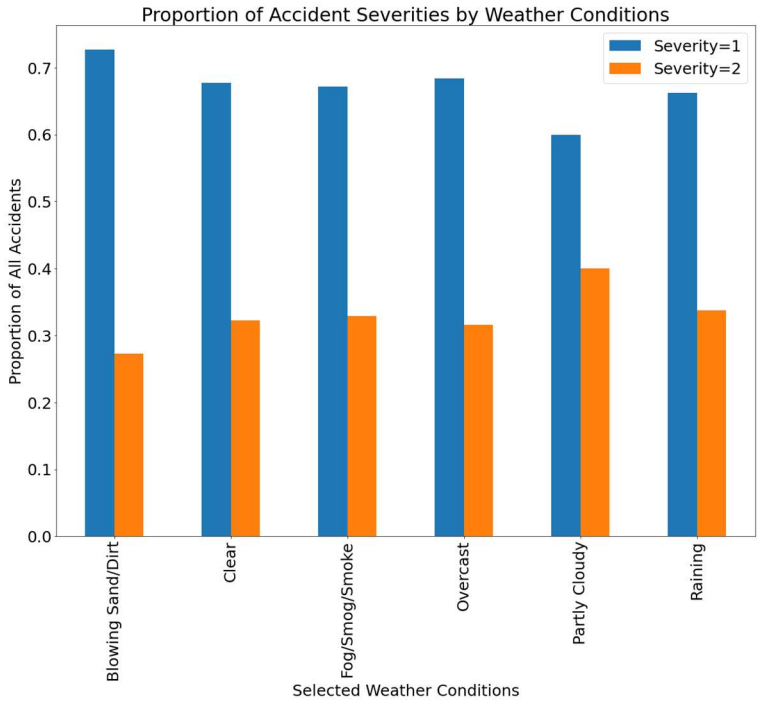
- A multi-class model can be used to classify the various degrees of severity. The features used to classify would be variables such as road condition, light condition, weather, speeding (or not), and the lane the driver is in.
- For example, consider a driver who is driving under 'wet', 'dark, streetlights on', 'not speeding', and on a particular lane.

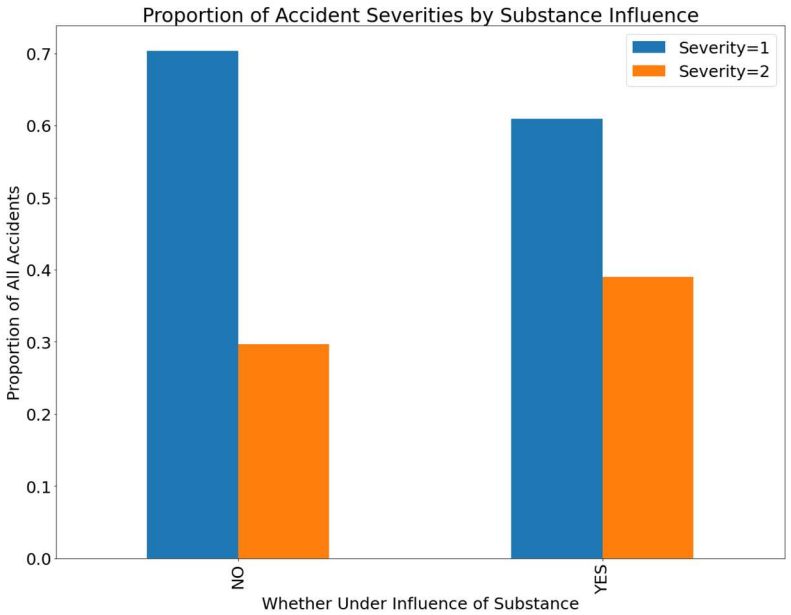
Data

- A trained model, when assuming that an accident does occur, can classify the severity of the accident. This would provide the driver with the 'worst-case scenario.
- This can still have the effect of inducing an appropriate level of cautiousness in the driver.
- Human features: Speeding, whether accident due to inattention, and whether driver under substance influence
- Environmental features: Light condition, road condition, and weather condition

Methodology

- Visual evidence of varying severity by conditions





Methodology

- I take a metric-based approach in selecting the best classification algorithm to generate insights.
- The two candidate models are: Decision Tree and Logistic Regression.
- The two metrics I rely on for model selection are “accuracy score” and “F1-score”.

Methodology

Decision Tree selection

- We can see that the scores are not significantly different. Thus, we should be indifferent to the value of the Maximum Depth parameter. The value of 5 is arbitrarily chosen to compete against the Logistic Regression model.

	MaxDepth	AccScore	F1Score
5	6	0.700380	0.578062
7	8	0.700275	0.578107
6	7	0.700248	0.578095
8	9	0.700248	0.578191
4	5	0.700195	0.576851
0	1	0.700169	0.576692
1	2	0.700169	0.576692
2	3	0.700169	0.576692
3	4	0.700143	0.576679
9	10	0.699958	0.578242

Methodology

Logistic Regression selection

- We can see that the scores are not significantly different. Thus, we should be indifferent to the value the type of solver used. The default solver is arbitrarily chosen to compete against the Decision Tree model.

	Solver	AccScore	F1Score	LogLoss
0	newton-cg	0.700169	0.576741	0.587845
3	sag	0.700169	0.576741	0.587846
4	saga	0.700169	0.576741	0.587846
1	lbfgs	0.700169	0.576741	0.587846
2	liblinear	0.700169	0.576741	0.587892

Methodology

Final Model selection

- Once again, the scores do not differ significantly. I chose the Logistic Regression model since it has the capability to predict probabilities of a severity instead of merely categorizing them.
- This can be justified as probabilities give a clearer picture to drivers by telling them how likely they will encounter a severe accident.

	Algorithm	F1-score	Accuracy
0	Decision Tree	0.576851	0.700195
1	LogisticRegression	0.576741	0.700169

Results

- Notice that regardless of the combination of environmental factors, the determining factor of severity is the human condition of drug influence while driving.
- Otherwise, we should see variation in severity categorization when the combination changes.

Human Condition: Drug Influence 'Yes' vs. 'No'	Environmental Condition: Light 'Dark' vs. 'Daylight'	Environmental Condition: Road 'Wet' vs. 'Dry'	Environmental Condition: Weather 'Raining' vs. 'Clear'	Classification of Severity 'Level 1' vs. 'Level 2'
No	Dark	Wet	Raining	1
No	Daylight	Wet	Raining	1
No	Dark	Dry	Clear	1
No	Dark	Wet	Clear	1
No	Daylight	Dry	Clear	1
Yes	Dark	Wet	Raining	2
Yes	Daylight	Wet	Raining	2
Yes	Dark	Dry	Clear	2
Yes	Dark	Wet	Clear	2
Yes	Daylight	Dry	Clear	2

Conclusion

- Empirical evidence that driving while under substance abuse increases the severity of an accident if it happens.
- Environmental factors such as weather, road, and lighting conditions do not impact severity of accident.
- Potential bias by not accounting for the driver's natural cautiousness under different environmental conditions.
- Deployment in a real-time setting where it warns a driver of his risks given inputs from sensors about his driving environments.

END