Car Accident Severity Analysis: Seattle, Washington

APPLIED DATA SCIENCE CAPSTONE - REPORT

Topics

- Introduction
- Data
- Methodology
- Results - F1-score and Accuracy from 3 Methods
- Discussion / Recommendation
- Conclusion

Introduction

- Car accidents occurs everywhere worldwide and is one of the leading causes for people between the ages 5-29 years.
- More than half of road traffic deaths involve users such as pedestrians, cyclists, and motorcyclists.
- The goal is to identity relevant factors and derive insight on what events lead to these car accident and what severity is depended on
- Problems and Factors
  - What attributes to car accidents?
  - What can be done to remedy these?

Source: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Data

- Car accident data from Data contains ~250,000 records from 2004 to 2020 based in Seattle, Washington
- Severity code (1 = prop damage, 2 = injury) is assigned to every car accident with additional information such as:
  - Location
  - Speeding Involved
  - Road Condition
  - Lighting Condition
  - Weather Condition
  - .. And etc

Source data: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv
Metadata: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf

Methodology

- To be able to perform exploratory analysis first, some data cleansing is performed to categorically assign values for road condition and speeding to derive insight
- Methods used are:
  - K-Nearest Neighbor (KNN)
  - Logistic Rgression
  - Decision Trees

Results - F1-score and Accuracy from 3 Methods

- Judging from the F1-score and accuracy shown on the right, the results are very similar. Perhaps more features should be considered for more visible differences to surface
- Logistic regression comparison results show some correlation between the road conditions and whether driver in accidents are speeding or not

F1-score and Accuracy - Results

	Analysis Method	F1-score	Accuracy
0	KNN	0.591438	0.697085
1	Logistic Regression	0.576051	0.699679
2	Decision Tree	0.576051	0.699679

Logistic Regression – Comparison Result

	Intercept	SPEEDING	ROADCOND
0	-0.857229	0.090932	0.149966

Discussion / Recommendation

- To Seattle state authority/government:
  - Install additional safety signs (such as slow-down and speed limit) in areas where most accidents occur to raise general awareness
  - Consider inspecting road conditions in these areas and make repairment as necessary
- To drivers and pedestrians/cyclists:
  - Be extra cautious near highways as they are hot spots for car acciedents
  - Stay alert during adverse weather and be aware of road and lighting conditions. Slow down and stay under speed limit

Conclusion

- Road conditions as well as whether or not if the driver is speeding contribute a lot to how the severity of traffic accidents can occur
- Other factors such as lighting conditions and locations also attribute to these car incidents