Alaskan Car Crash Analysis

UAA College of Engineering
University of Alaska Anchorage

Department of Civil Engineering, Department of Computer Science and Engineering

Client/Advisor: Dr. Vinod Vasudevan
Advisor: Dr. Shawn Butler
Author 1: Lydia Stark
Author 2: Cale Cornichuck
Date: 4/23/2022

1 Project Summary

Determine driver, environmental, temporal, and roadway risk factors for Alaskan crashes.

Research Questions

1. What are the risk factors for injury in Alaskan crashes?
2. When is each type of crash likely to occur?
3. Are there factors beyond the original dataset that impact crash rates?

2 Data Used

The data we used as in our analysis is listed below.

Meta Data: Crash Severity, Crash Type, Number of Vehicles Involved, Number of Injuries

Road Data: Road Surface, Pavement, Functional Class, Intersection

Environmental Data: Lighting, Temperature, Weather,

Temporal Data: Date, Time, Month, Year

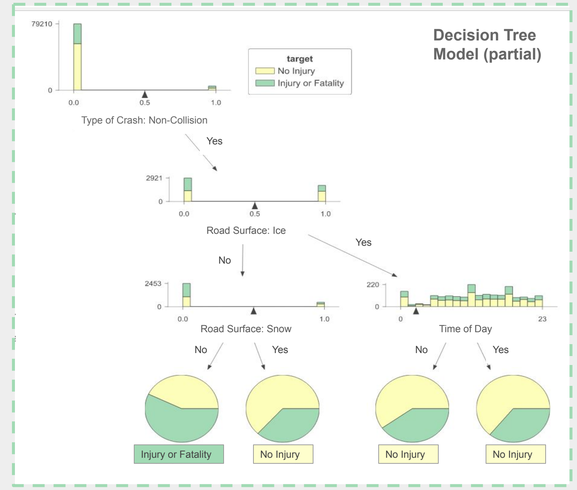
Driver Data: Age, Gender

3 Question 1: Risk Factors for Injury in Crashes

We trained four machine learning models to predict whether injury would occur in a crash.

<u>Model Accuracy</u>
- XGBoost - 70%
- Decision Tree - 70%
- Support Vector Machine - 69%
- Linear Regression - 62%

<u>Risk Factors</u>
- Type of Crash
- Road Conditions
- Time of Day



Key Takeaway:
Crashes that occur on icy/snowy roads are less likely to result in injury than crashes that occur on dry roads.

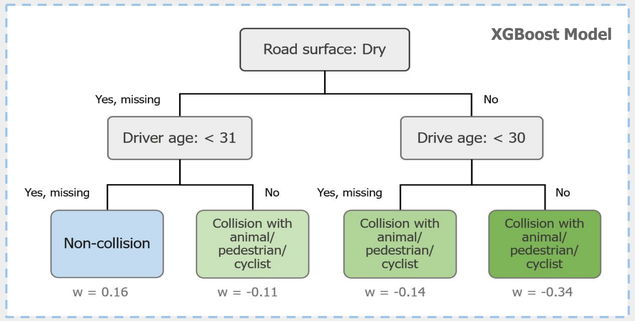
4 Question 2 - Likeliness of Each Crash Type Occurring

We trained two models to predict whether a <u>non-collision</u> or <u>collision with an animal/pedestrian/cyclist</u> would occur.

Note: in a non-collision crash, the vehicle does not collide with another object (for example, a rollover).

<u>Model Accuracy</u>
- XGBoost - 70%
- Decision Tree - 66%

<u>Risk Factors</u>
- Road Surface
- Driver Age



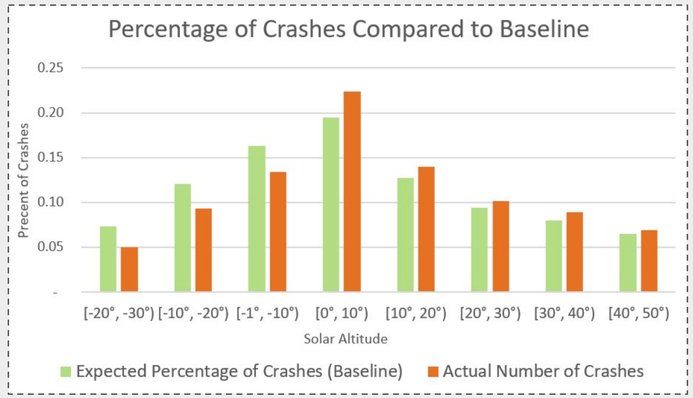
Key Takeaways:
1. **A collision with a pedestrian/animal/cyclist is most likely to happen when the driver is 30 or older and the roads are dry.**
2. **Conversely, a non-collision is most likely to occur when the driver is 29 or younger and the roads are not dry.**

5 Question 3 - Factors Beyond our Dataset That Impact Crash Rates

Step 1: Calculated solar altitude (the angle of the sun in the sky) for each crash using latitude and longitude.

Step 2: Determined baseline solar altitude tables for each month in Anchorage, weighted to account for rush hour.

Step 3: Compared the number of crashes in Anchorage to the baseline.

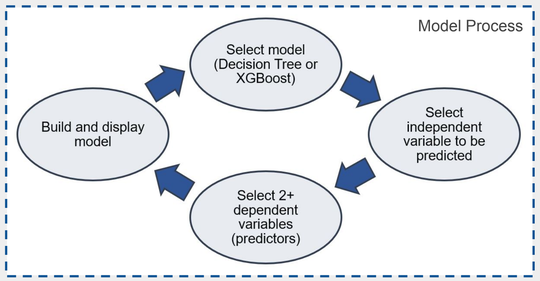


Key Takeaways:
1. **Solar altitude is significantly correlated with crash rates (p = 0.02).**
2. **The highest crash rates occur at twilight (0° - 9°).**

6 Data Analysis Tool

We developed a website that allows users to utilize our machine learning models.

The models can be used to predict any data attribute in our dataset.



7 Future Research

1. Expand our dataset to include more data: vehicle information, driver actions, rush hour data, snowfall, etc.
2. Compare crashes in rural areas to urban areas to see if the risk factors change.

We would like to thank our advisors, Dr. Vinod Vasudevan and Dr. Shawn Butler, for their help and insights on this project!