Alaska Automobile Crash Analysis

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Moro Bamber, Youji Seto, Nemed Aleman
Department of Civil Engineering, Department of Computer Science and Engineering

Client/Advisor : Dr. Vinod Vasudevan
Advisor: Dr. Frank Witmer
Date: 12/6/2024

Project Summary

Our mission was to find valuable insights into automobile crash data in Alaska.

Guiding Research Questions:
- What can we find in this data that tells us more about why a crash happened?

- How do crashes differentiate based on if they are in an urban or rural setting?

We employed advanced Machine Learning Techniques to enhance our research.

We also developed a web application to aid in further research with this data.

Data Pre-Processing

The crash data consisted of tables of 50,000 - 100,000+ rows and 30+ columns. The data was broken down into 3 levels of increasing detail about each crash..

We used a combination of SQL and Python to build the datasets we used in our ML models.

Using Google Earth Engine and python, we classified each crash into an urban, suburban and rural category.

Crash Level
High Level Characteristics about Crash and Environment

Driver-Vehicle Level
Includes a Row for all vehicles/units invloved & Characteristics pertaining to the Driver of the Unit

Person Level
Includes a row for every person involved. Says who was injured, where they sat etc.

Three Levels of Crash Data
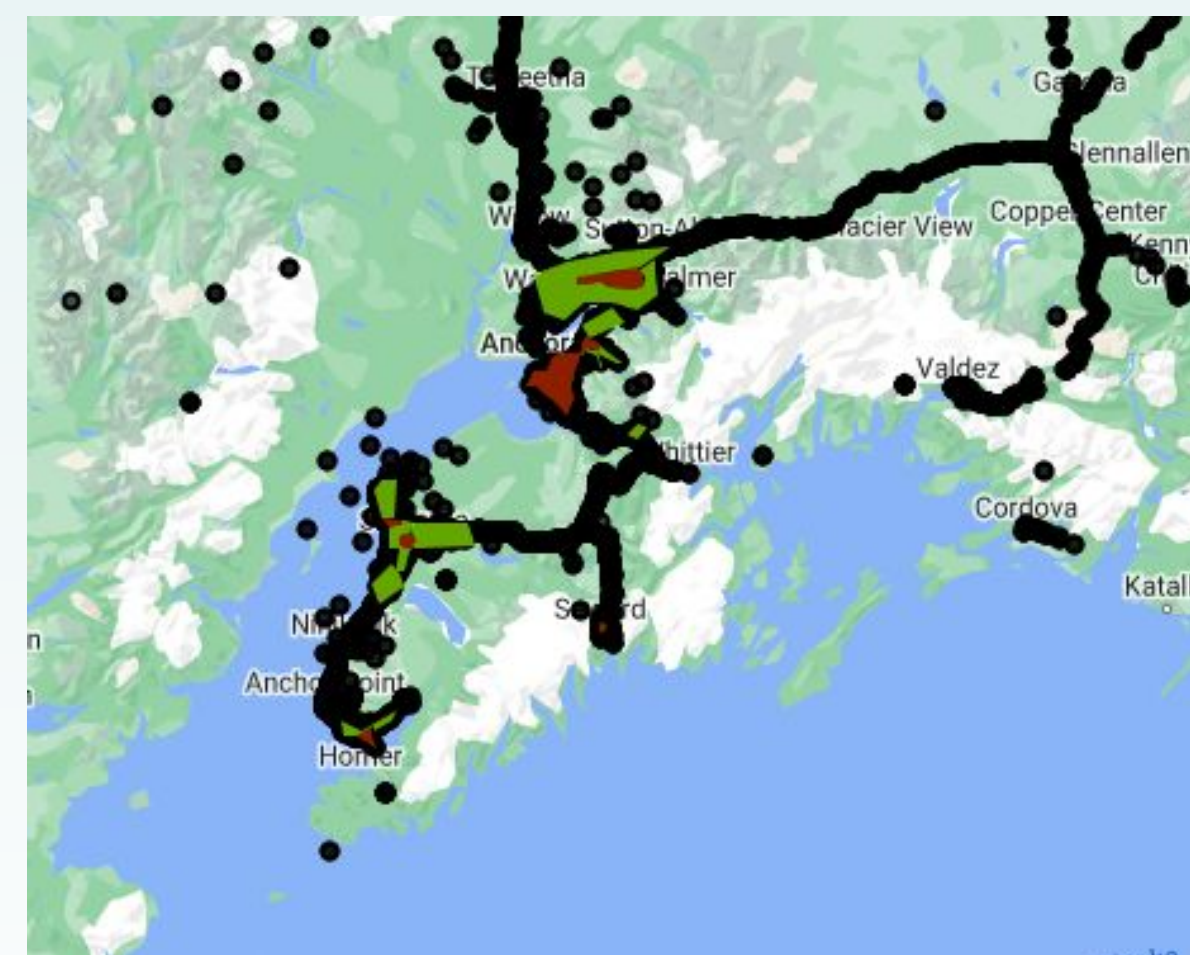
Black Dot: Crash
Green Area: Suburban Area
Red Area: Urban Area

Areas and Crashes in Google Earth Engine
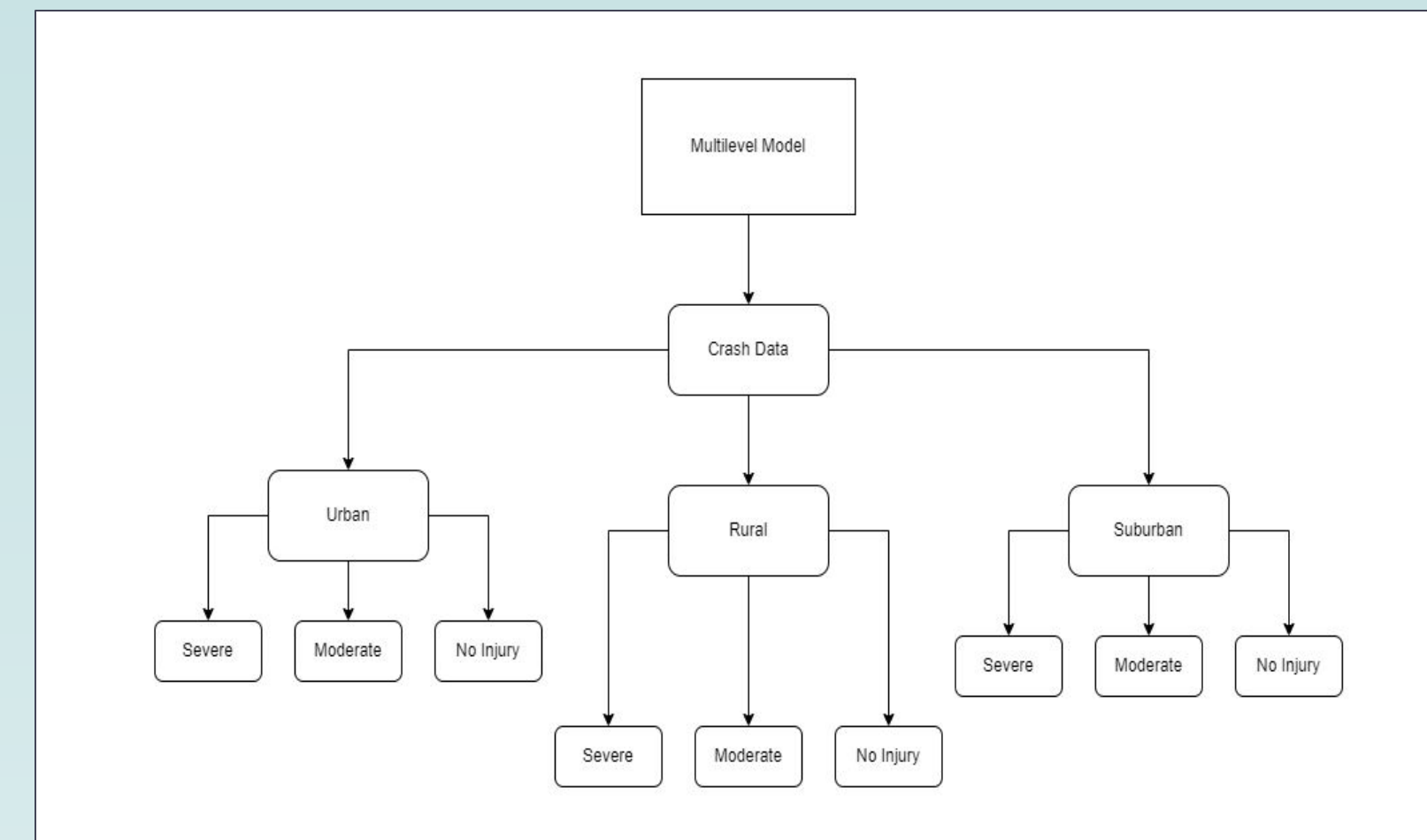
Machine Learning

Why use the Multilevel Random Forest Model?

Enables Advanced Analysis:
- Analyzes data in a nested structure for area-specific insights.

Key Question:
- "Which contributing factor correlates most with a specific crash type based on area (urban, rural, suburban)?"

Multilevel Model Structure

Multilevel Random Forest Model Evaluation Metrics

Moderate Injury
- Rural: 63.3% accuracy
- Suburban: 62.5% accuracy
- Urban: 62.8% accuracy

No Injury
- Rural: 67.7% accuracy
- Suburban: 62.1% accuracy
- Urban: 65.7% accuracy

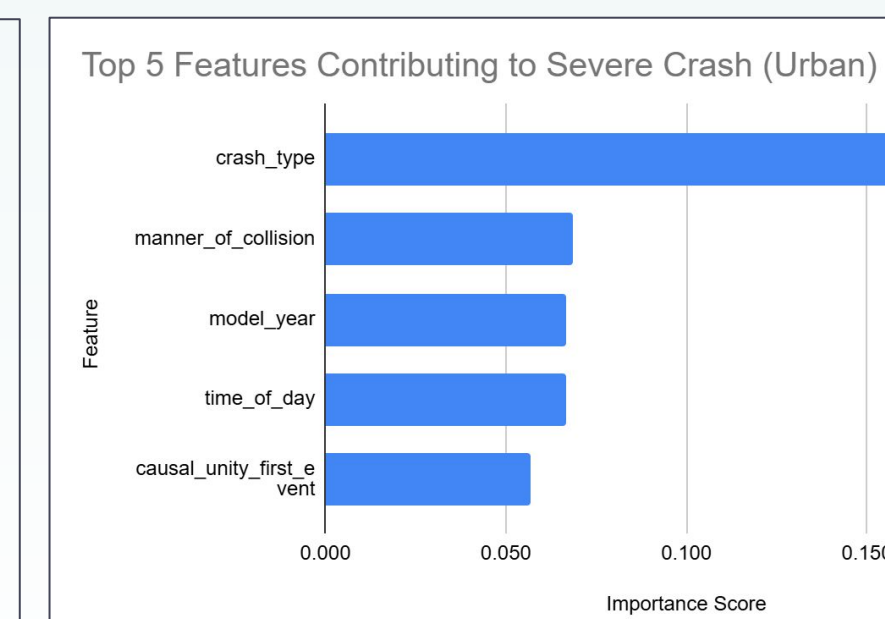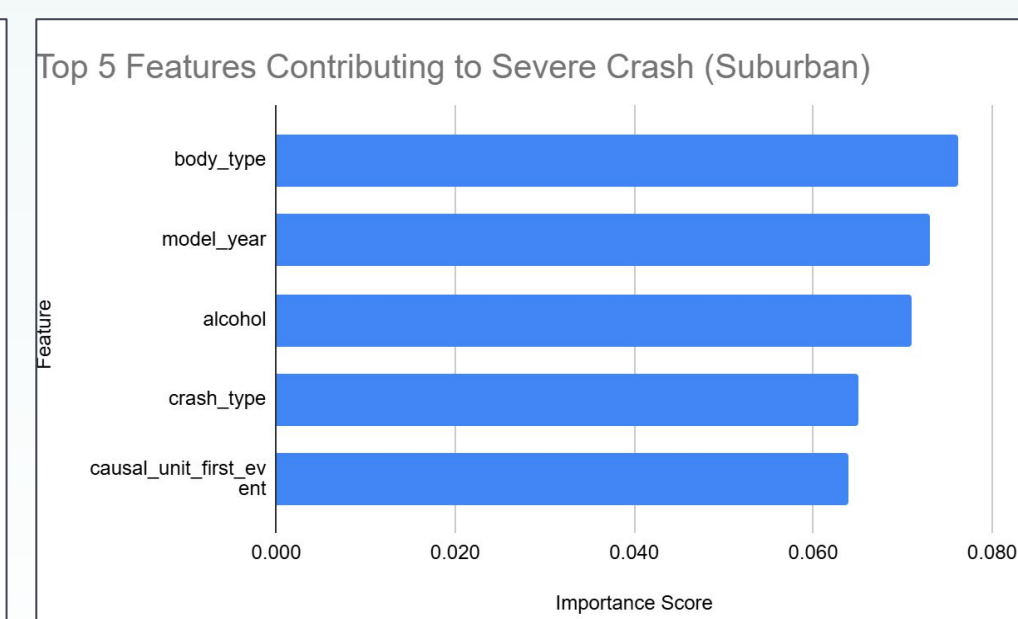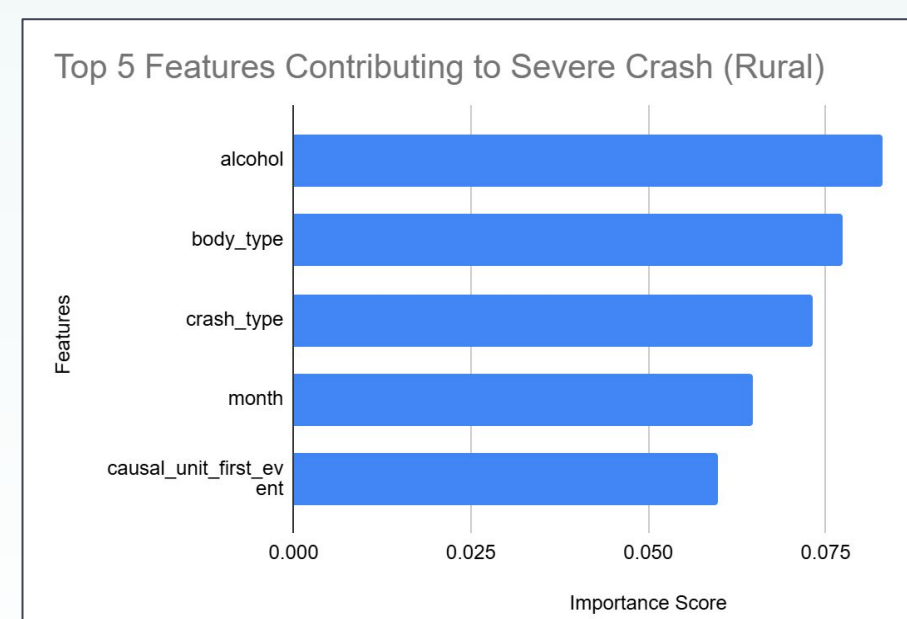| Severe Crash Type Evaluation Metrics			
	Rural	Urban	Suburban
Accuracy	0.77	0.73	0.75
Precision	0.78	0.75	0.76
Recall	0.75	0.71	0.73
F1	0.77	0.73	0.75

Severe Performance Metrics

Rural Areas
- Vehicle characteristics play a crucial role in rural severe crashes
- Seasonal patterns are more important in rural areas
- Alcohol is a more significant factor compared to other areas

Urban Areas
- Specific crash types dominate urban severe accidents
- Traffic patterns and vehicle interactions are crucial
- Time of day is more important than in rural areas

Suburban Areas
- Shows a hybrid pattern between urban and rural
- More balanced distribution of important factors
- Vehicle characteristics remain important

Top 5 Features Contributing to Severe Crash (Rural)

Top 5 Features Contributing to Severe Crash (Suburban)

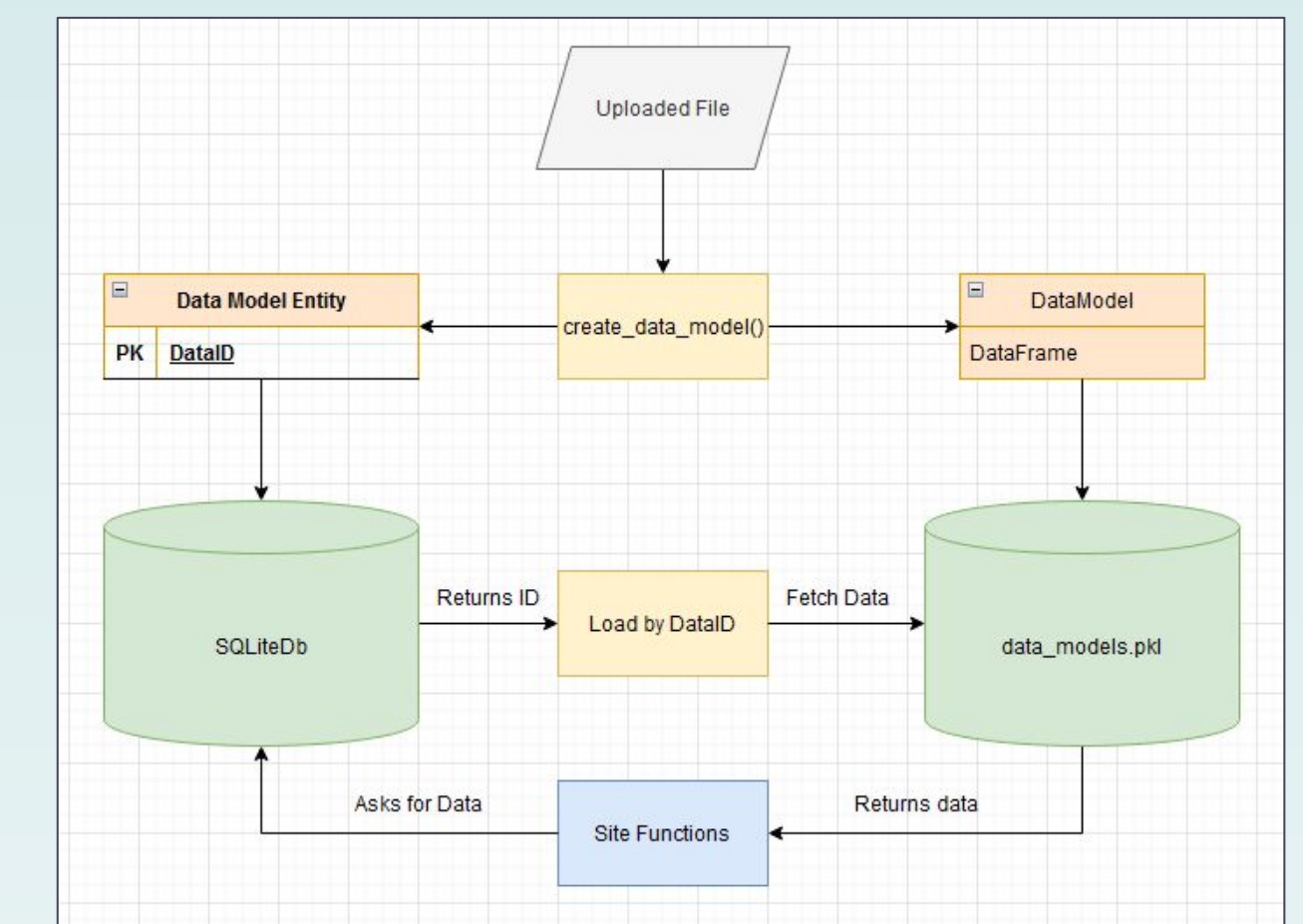Top 5 Features Contributing to Severe Crash (Urban)

Alaska Crash Data Analysis Tool

The application allows users to:
1. Upload and access large datasets
2. Filter Data to specified preferences
3. Run Machine Learning models on specified data
4. Output meaningful and easy to understand results

Tech Stack

Backend Data Handling

Machine Learning Results

Classification Report:

Accuracy: 80.17%

Value - Encoding Table

Value	Encoding
Fatal	0
Incapacitating	1
Possible	2

Label	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	47.0
1	0.55	0.32	0.41	268.0
2	0.83	0.93	0.88	956.0
macro avg	0.79	0.75	0.76	1271.0
weighted avg	0.78	0.80	0.78	1271.0

Machine Learning Results Page

Future Work

1. Expand dataset to include more accurate weather data.
2. Improve ML models with more experimentation.
3. Improve web tool interface.