UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Alaska Car Crash Analysis

By Moro Bamber, Youji Seto, and Nemed Aleman

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Our Mission

1. Discover new insights into car crashes in Alaska
2. Create web application that allows users to select variables for use in ML models that produce results the user can easily interpret
3. Allow new data to be added for better accuracy

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Project Background

- Building on previous capstone project
- Crash data is provided by the Alaska DoT
- Client: Dr. Visudevan from UAA Civil Engineering
- Exploring new ways to look at the data and adding new variables
  - like an urban vs. rural classification

TRANSPORTATION & PUBLIC FACILITIES
STATE OF ALASKA

Dr. Visudevan

UAA College of Engineering
University of Alaska Anchorage

Societal Relevance

Web application

- Easier for traffic engineers to filter and analyze data
- Making ML tool accessible to people with limited experience with ML
- Making data accessible to more people

Research

- Rural vs. Urban Study
- Identified crash characteristics were different among different areas
- Agencies can allocate resources according to challenges faced in different areas
- Allocate resources for safety protocols

Design Requirements - ML

- Develop models that identify risk factors contributing to car crashes in rural versus urban Alaska.
- The models should be able to accurately classify these factors based on the data provided, drawing meaningful insights from the comparison

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Design Requirements - Web Application

- Use Django to serve as the framework for the machine learning model.
- Allow users to interact with the ML model, showcasing the findings through an easy-to-navigate UI.
- Provide functionality for users to upload new car crash data, ensuring specific fields are present for model processing.
- Display data in a user-friendly format, making the insights clear and easy to interpret.
- Implement data security measures, such as password protection, for accessing sensitive data and model results.
- Ensure the web app supports scalability for potential future expansions or additional data integrations

ML Models

Classification Models

- Decision Trees - Effective for multi-class classification problems
- Gradient Boosting Machines (XGBoost) - Often achieve high accuracy
- Random Forest
- Support Vector Machine (SVM) - Useful for high dimensional data (too many features)
- Multilevel Models - Effective for grouped data

"What factors are most strongly associated with crashes in urban areas?"

Decision Trees/Random Forest for identifying key features.

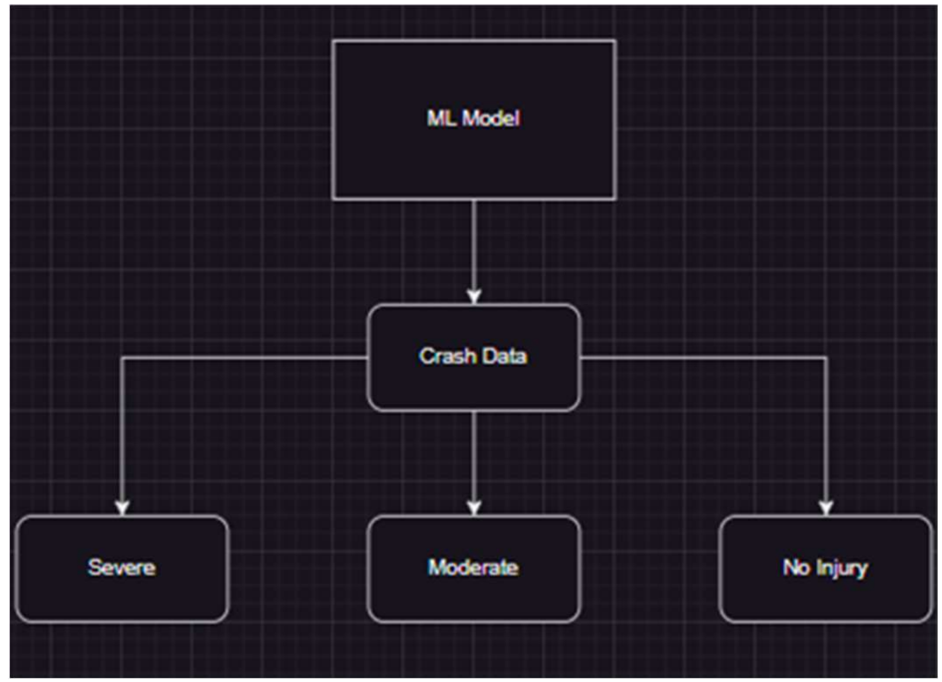
ML Models

What is the ML model doing?

- trying to predict which crash severity type a car crash is based on the features (alcohol, time, car body type, car year, etc.)

How does it do this?

- The model first identifies patterns that help it to recognize which combination of features makes a car crash a certain severity type based off of known data



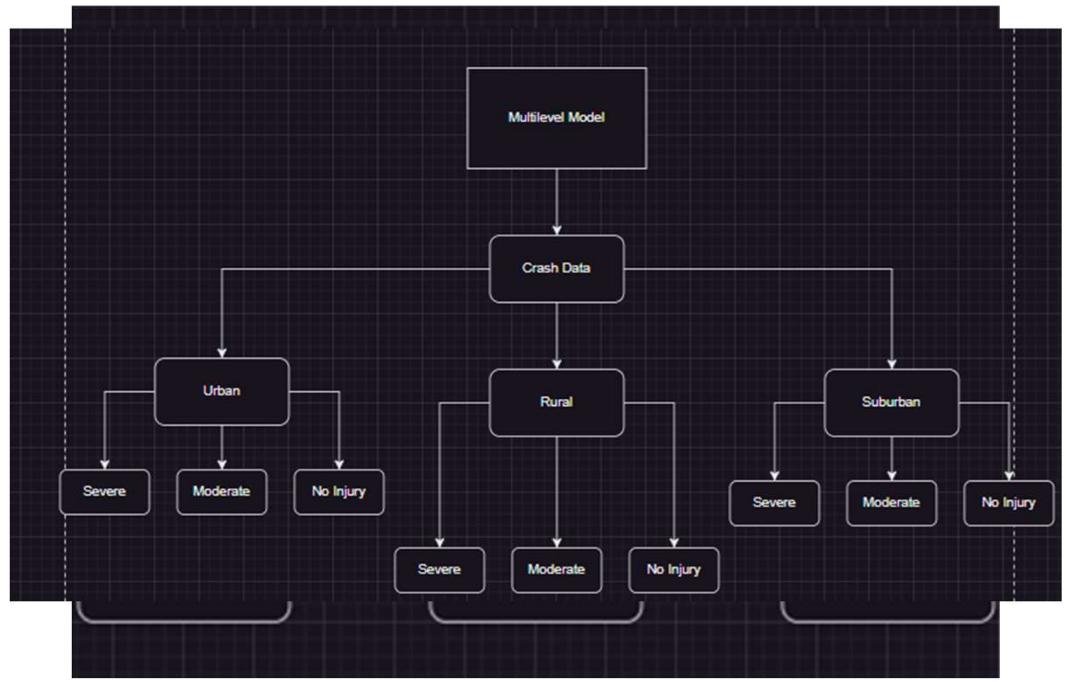
Why Use the Multilevel Model?

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)?"



UAA College of Engineering
University of Alaska Anchorage

Multilevel Model Using Random Forest

No Injury			
	Rural	Urban	Suburban
Accuracy	0.677176	0.621314	0.657425
Precision	0.665501	0.622782	0.645201
Recall	0.708802	0.61599	0.699416
F1	0.686278	0.619257	0.671076
ROC_AUC	0.746404	0.672016	0.709973

Moderate			
	Rural	Urban	Suburban
Accuracy	0.633249	0.625315	0.628689
Precision	0.622902	0.622092	0.62212
Recall	0.677163	0.638882	0.655841
F1	0.648419	0.630216	0.638142
ROC_AUC	0.688476	0.678145	0.676658

Severe			
	Rural	Urban	Suburban
Accuracy	0.772269	0.734445	0.751527
Precision	0.787462	0.750433	0.764069
Recall	0.745533	0.705639	0.731666
F1	0.765655	0.726248	0.74572
ROC_AUC	0.843473	0.806915	0.811955

Multilevel Model Using Random Forest

Best Performance Category

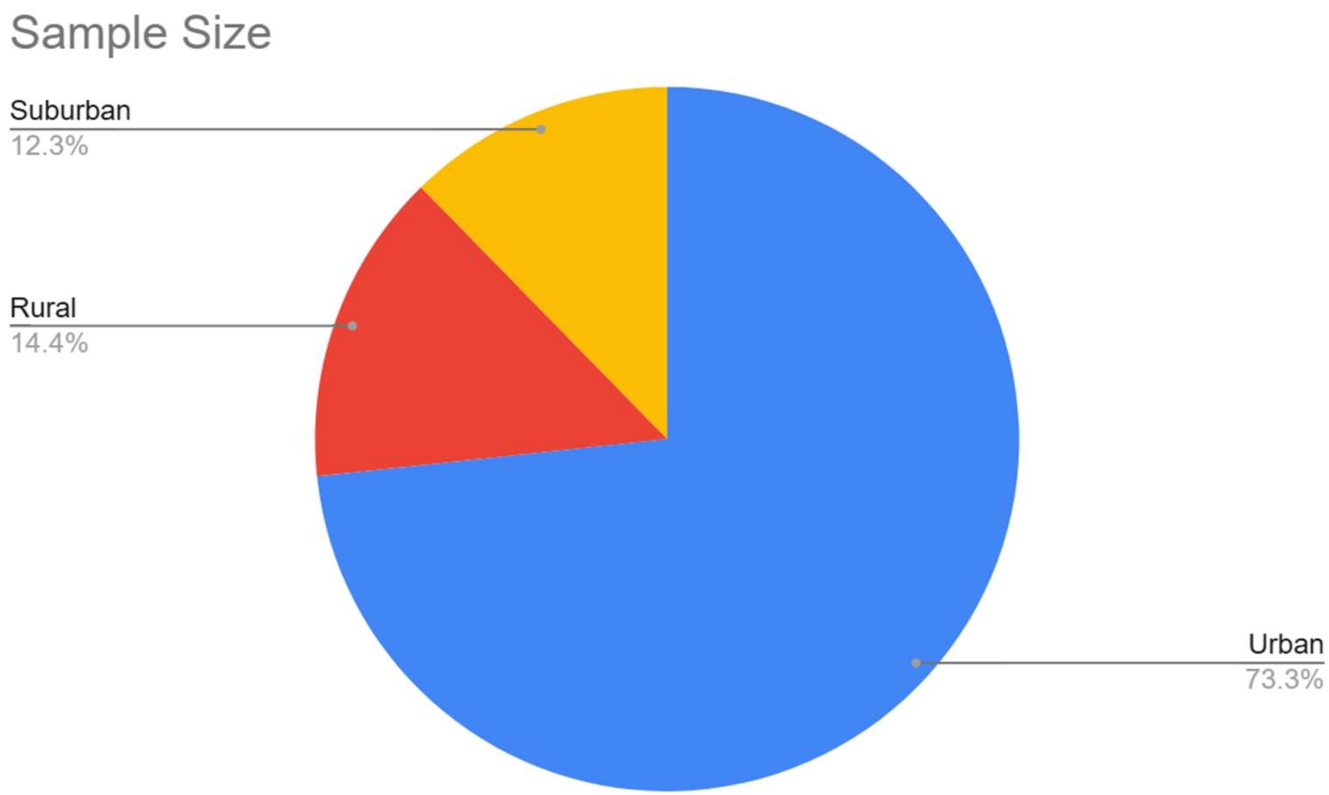
- Rural areas: 77.2% accuracy (84.3% ROC-AUC)
- Suburban areas: 75.2% accuracy (81.2% ROC-AUC)
- Urban areas: 73.4% accuracy (80.7% ROC-AUC)

Most reliable predictions are for **severe accidents**, especially in rural areas.

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

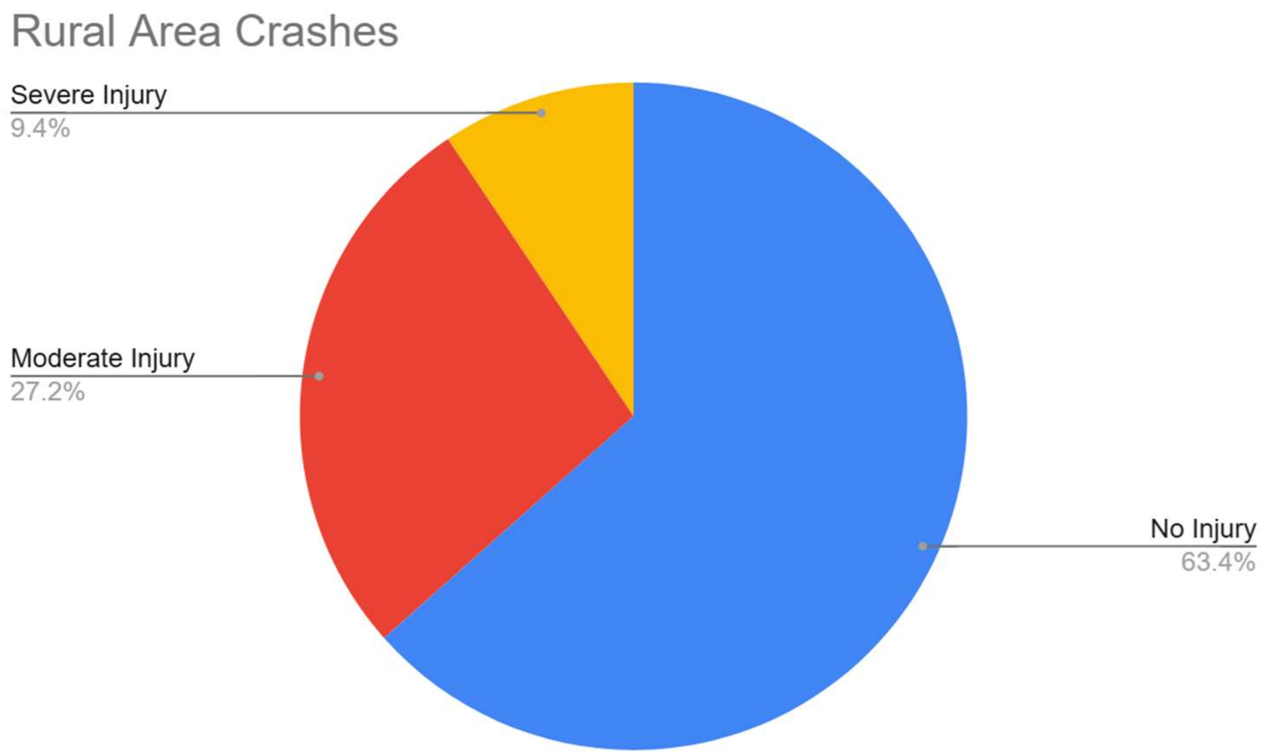
Dataset

	Sample Size
Urban	26,386
Rural	5170
Suburban	4434



Dataset

Rural Area Crashes	
No Injury	3,343
Moderate Injury	1,432
Severe Injury	495



UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Multilevel Model Using Random Forest

```
Top 10 most important features:
                          Feature  Importance
12                   alcohol_freq    0.083098
19                 body_type_freq    0.077582
3                 Crash_Type_freq    0.073130
18                     Month_freq    0.064813
20                model_year_freq    0.059957
10  Causal_Unit_First_Event_freq    0.059696
2         Causal_Unit_Action_freq    0.058836
17               time_of_day_freq    0.051590
1         First_Harmful_Event_freq    0.049454
22                distracted_freq    0.047551
```

Rural

Multilevel Model Using Random Forest

Top 10 most important features:

	Feature	Importance
3	Crash_Type_freq	0.159266
8	Manner_Of_Collision_freq	0.068681
20	model_year_freq	0.066952
17	time_of_day_freq	0.066510
10	Causal_Unit_First_Event_freq	0.057000
19	body_type_freq	0.055865
18	Month_freq	0.054768
2	Causal_Unit_Action_freq	0.049168
12	alcohol_freq	0.046615
15	Surface_freq	0.041305

Urban

Multilevel Model Using Random Forest

```
Top 10 most important features:
                          Feature  Importance
19                 body_type_freq    0.076251
20                model_year_freq    0.073478
12                   alcohol_freq    0.071119
3                 Crash_Type_freq    0.065130
10  Causal_Unit_First_Event_freq    0.064205
18                     Month_freq    0.062674
17               time_of_day_freq    0.061466
1         First_Harmful_Event_freq    0.054303
2          Causal_Unit_Action_freq    0.051677
6                    Weather_freq    0.042204
```

Suburban

Multilevel Model Using Random Forest

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

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

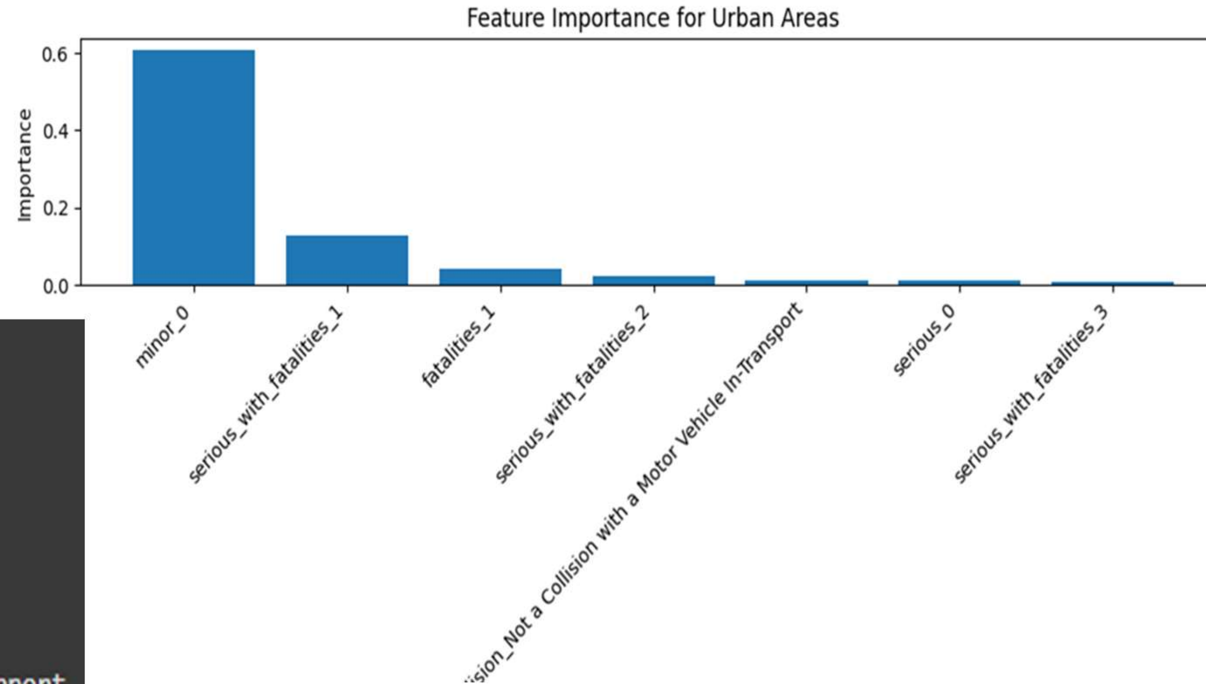
Multilevel Model Using Random Forest

Prevention Strategies

- Rural: Target alcohol and vehicle-specific measures
  - Stronger alcohol prevention measures
- Urban: Focus on traffic management and crash prevention
  - Time-based traffic management
- Suburban: Implement hybrid approaches
  - Mix of both urban and rural

Decision Tree - Urban

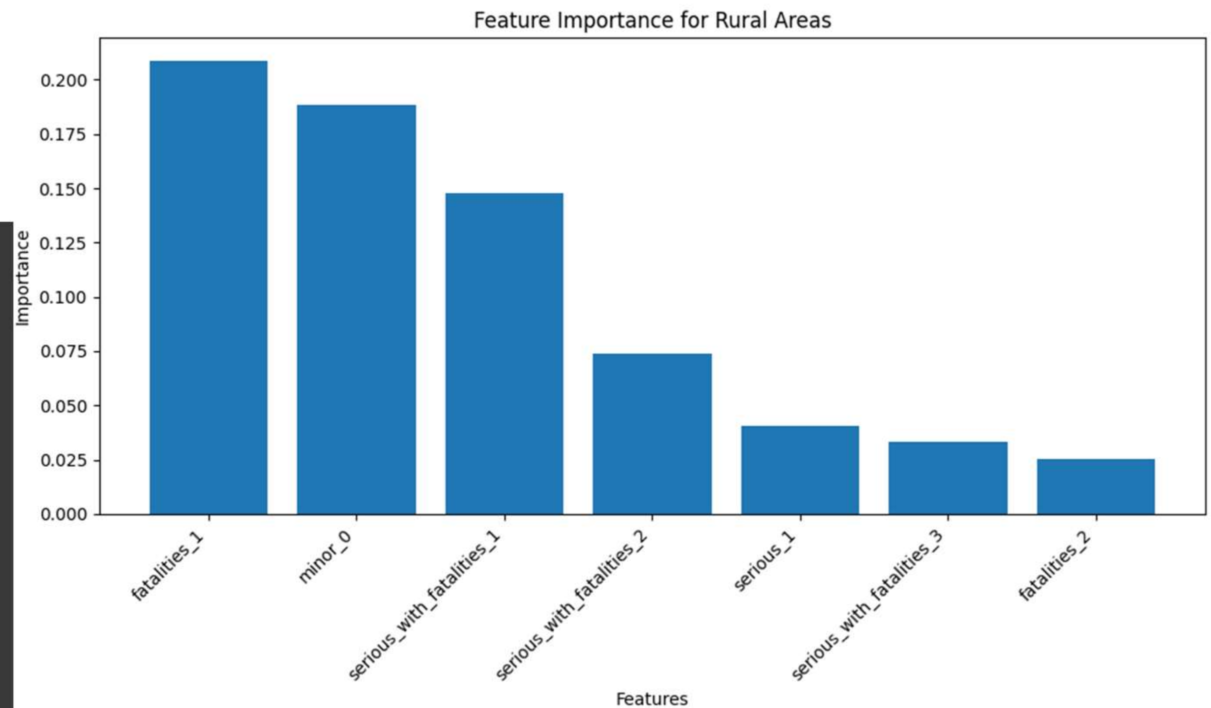
```
Accuracy for Urban: 87.42%

Confusion Matrix for Urban:
 [[  31    0    0]
 [   1 7426 1100]
 [   0  204 1615]]

Classification Report for Urban:
                   precision    recall  f1-score   support

    High Severity       0.97      1.00      0.98        31
     Low Severity       0.97      0.87      0.92      8527
Moderate Severity       0.59      0.89      0.71      1819

         accuracy                           0.87     10377
        macro avg       0.85      0.92      0.87     10377
     weighted avg       0.91      0.87      0.88     10377
```



UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Decision Tree - Rural

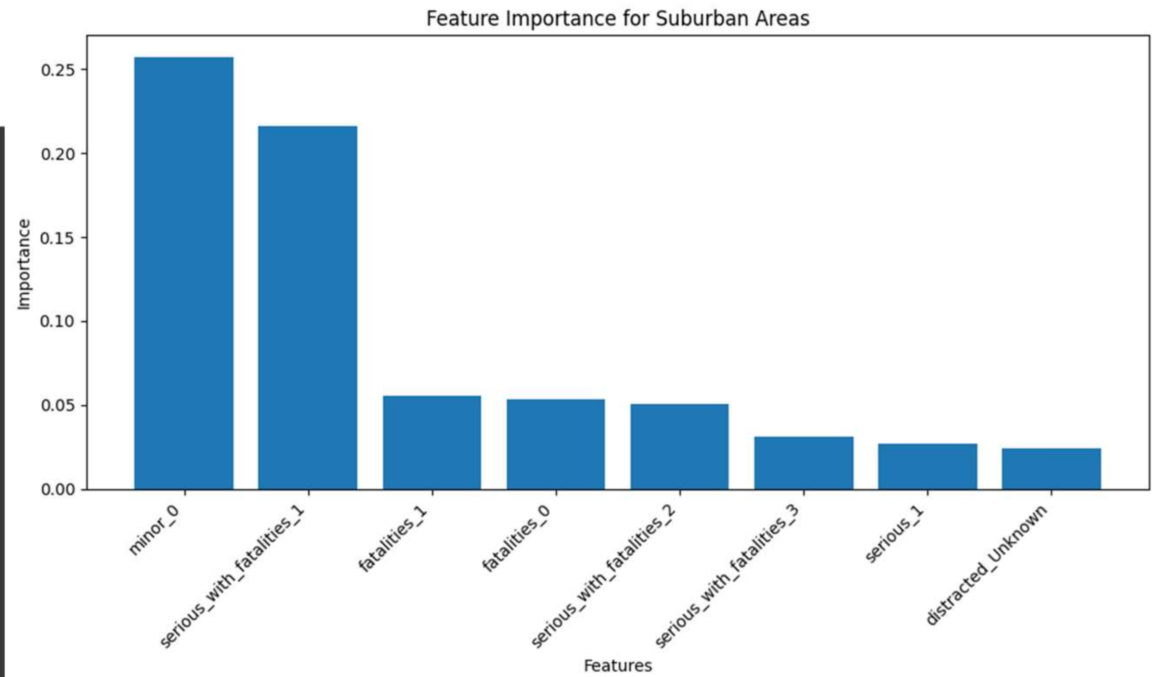
```
Accuracy for Rural: 90.84%

Confusion Matrix for Rural:
 [[  35    4    0]
 [   0 1512   40]
 [   2  126  158]]

Classification Report for Rural:
                   precision    recall  f1-score   support

    High Severity       0.95      0.90      0.92        39
     Low Severity       0.92      0.97      0.95      1552
Moderate Severity       0.80      0.55      0.65       286

         accuracy                           0.91      1877
        macro avg       0.89      0.81      0.84      1877
     weighted avg       0.90      0.91      0.90      1877
```



UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

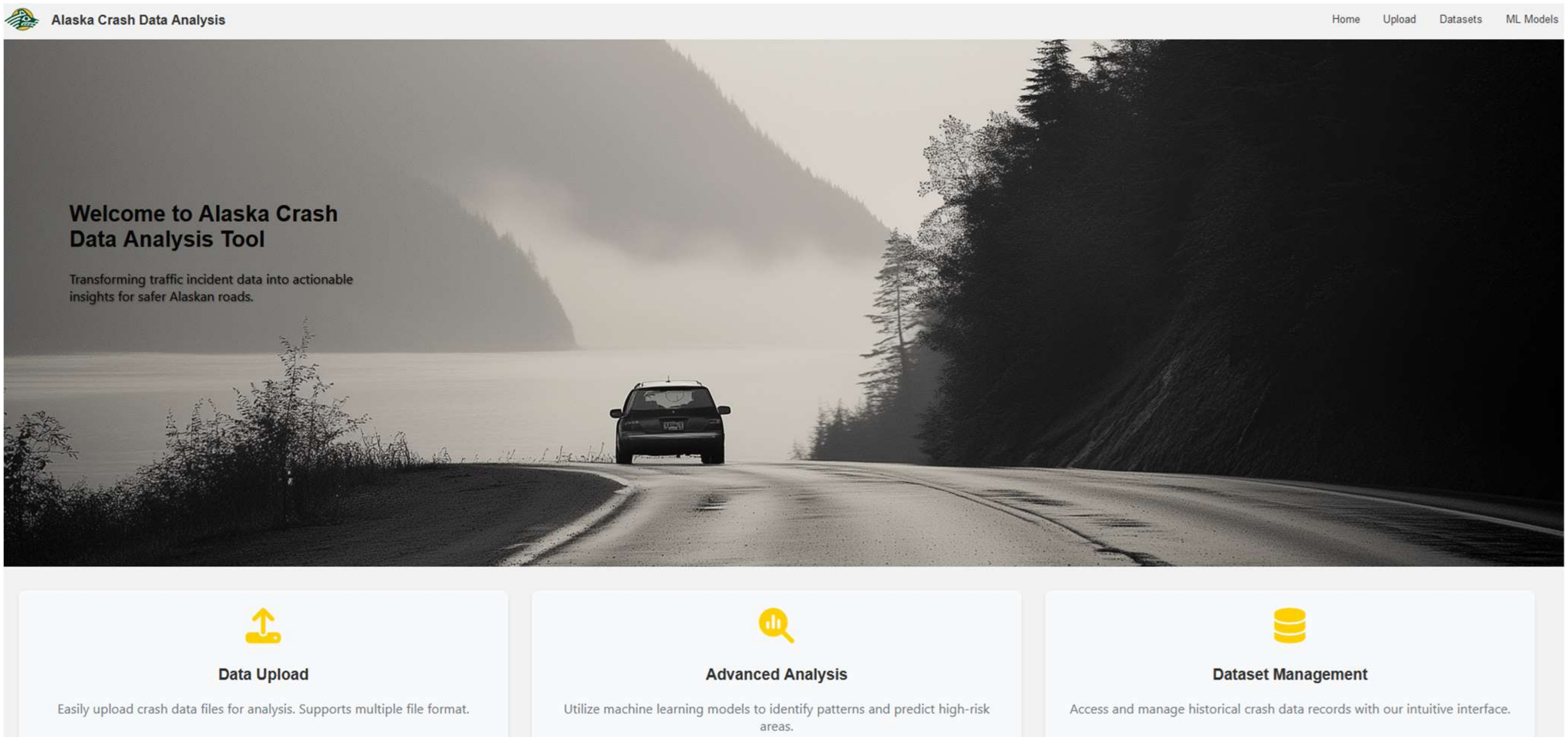
Decision Tree - Suburban

```
Accuracy for Suburban: 85.33%

Confusion Matrix for Suburban:
 [[  16    1    2]
 [   0 1195  194]
 [   1   45  203]]

Classification Report for Suburban:
                   precision    recall  f1-score   support

    High Severity       0.94      0.84      0.89        19
     Low Severity       0.96      0.86      0.91      1389
Moderate Severity       0.51      0.82      0.63       249

         accuracy                           0.85      1657
        macro avg       0.80      0.84      0.81      1657
     weighted avg       0.89      0.85      0.87      1657
```



UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Web Application



UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Handling Data

Alaska Crash Data Analysis

Upload Parameters

Start Year:
2004

End Year:
2004

Header Size (Optional):
Enter header size

Click for more info

Rename Dataset (Optional):

Select File:
Browse... No file selected.

Upload File

Alaska Crash Data Analysis

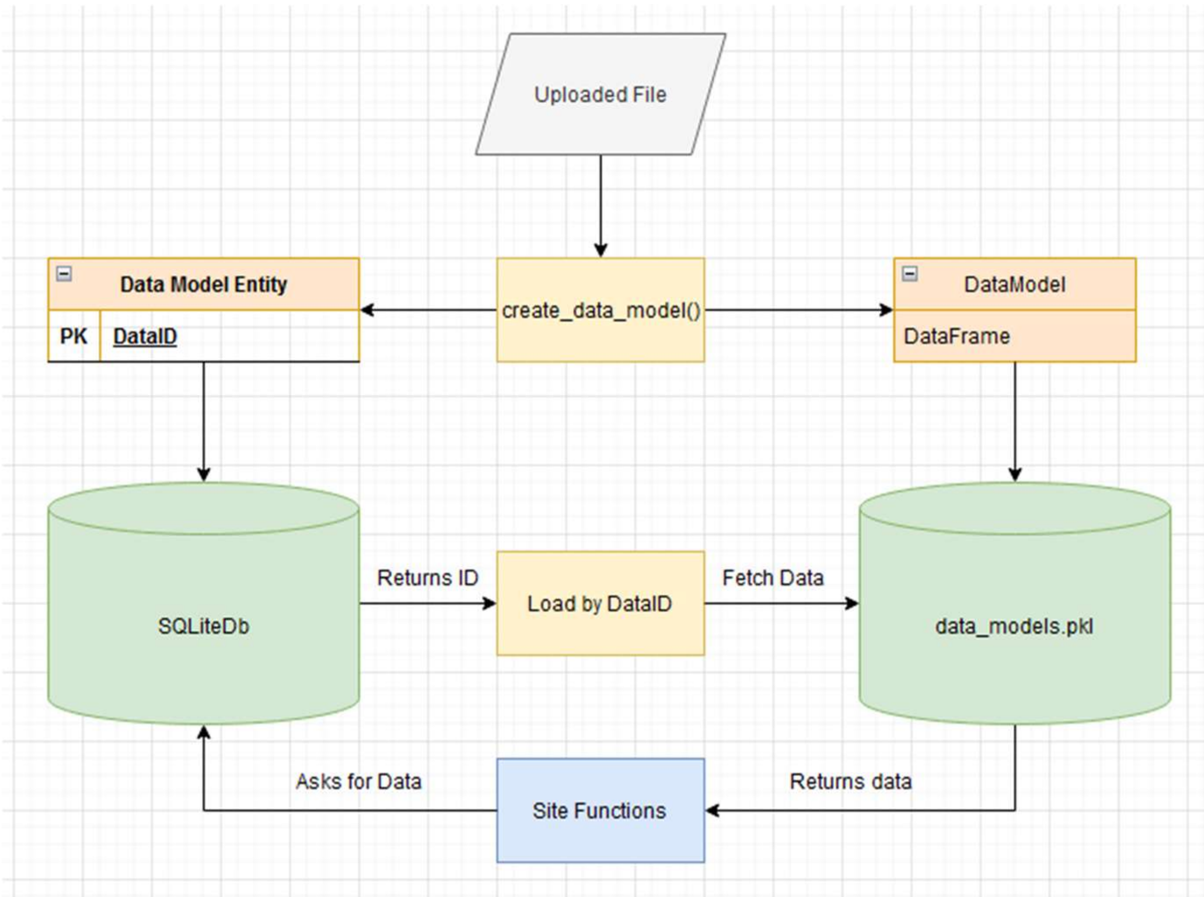
Datasets

Data ID	Name	Date Uploaded	Start Year	End Year
crashdata_2013_2017	crashdata	Nov. 22, 2024, 12:20 p.m.	2013	2017
newtest0912_2009_2012	newtest0912	Nov. 22, 2024, 12:21 p.m.	2009	2012
rural_suburban_winter_2013_2017	rural_suburban_winter	Nov. 24, 2024, 4:31 p.m.	2013	2017
all_winter_test_2013_2017	all_winter_test	Nov. 24, 2024, 5:30 p.m.	2013	2017
DoT_test_2013_2017	DoT_test	Nov. 25, 2024, 11:50 a.m.	2013	2017

Alaska Crash Data Analysis Tool 2024 - University of Alaska Anchorage College of Engineering

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Backend Data Pipeline



UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Machine Learning Results

Alaska Crash Data Analysis

Machine Learning Results

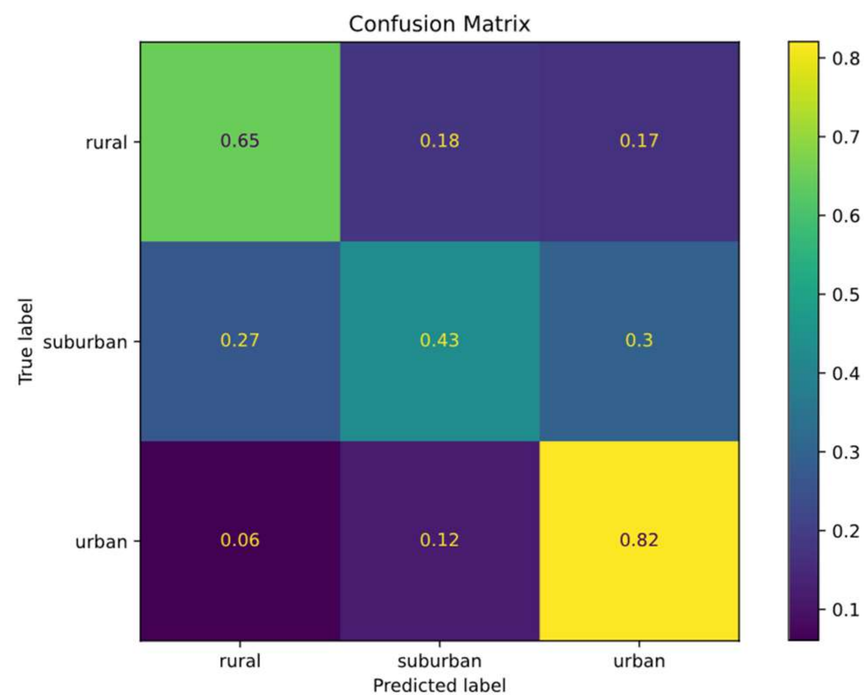
Classification Report:

Accuracy: 75.67%

Value - Encoding Table

Value	Encoding
rural	0
suburban	1
urban	2

Label	Precision	Recall	F1-score	Support
0	0.48	0.65	0.55	664.0
1	0.32	0.43	0.37	747.0
2	0.92	0.82	0.87	4668.0
macro avg	0.57	0.63	0.60	6079.0
weighted avg	0.80	0.76	0.77	6079.0



What did you learn?

Moro: How to build a django web application. How to structure code for Machine Learning models in a web application. How to handle and parse large datasets efficiently using SQL and Python.

Youji: Techniques involving balancing data (SMOTE, weights, removing excess data), data preprocessing: identifying which features or columns are necessary, story behind what the model is predicting.

Nemed: Selecting appropriate machine learning models for a real-world classification problem. Evaluation metrics and techniques like cross-validation, grid search, statistical analysis(Chi-Square test), and Cramér's V. Software development cycles.

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Questions?