**Executive Summary: Predicting Car Accident Severity in Virginia to Enhance Road Safety**

*Prepared for: Professor Devin Fensterheim*  
*Prepared by: Nasim Aalemi*  
*Class: DATA 602*  
*Date: May 14, 2025*

**Business Problem**

Car accidents in Virginia cause significant loss of life, injuries, and economic costs, with severe crashes (fatal or serious injuries) being the most critical challenge. From December 20, 2024, to February 28, 2025, a dataset of 138,813 crashes showed that only 0.7% (922 cases) were classified as "High" severity, yet these incidents cause disproportionate harm. Predicting severe crashes and understanding their key drivers are essential for prioritizing safety interventions, optimizing resource allocation, and reducing fatalities. The Virginia Department of Transportation (VDOT) and local governments seek data-driven insights to inform targeted policies to enhance road safety across urban and rural districts.

This project aims to address the following business questions:

* What factors drive severe crashes?
* Can machine learning guide proactive safety measures?

**Approach**

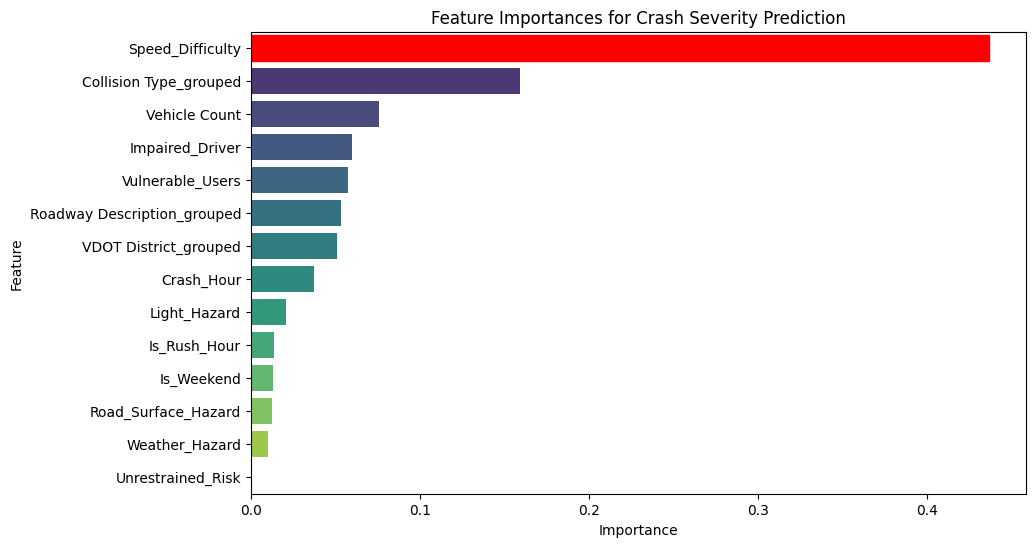
Using the CRISP-DM methodology, the project analyzed 138,813 crash records with features like max speed difference, collision type, vehicle count, vulnerable road users, roadway type, impaired driving, VDOT district, and crash hour. Key steps included:

* **Data Preparation**: Addressed class imbalance (0.7% severe, 23.5% medium, 75.8% low) using subsampling and SMOTE to balance the dataset. Grouped rare categories and applied target encoding to handle high cardinality.
* **Subsampling:** Reduced "Low" (20,000) and "Medium" (10,000) samples while retaining all 922 "High" samples to balance the training set.
* **Modeling**: Employed a BalancedRandomForestClassifier, assigning higher weight to severe crashes and optimizing recall (>0.7) and precision (>0.05).
* **Feature Engineering:** Grouped rare categories and applied target encoding to handle high cardinality.

**Key Findings**

The model identified the top drivers of severe crashes, accounting for ~79% of predictive power:

* **Speed Difference** (43.7%): Excessive speed or large speed differences between vehicles result in higher impact forces, significantly increasing severity.
* **Head-on and sideswipe collisions** (15.9%): High-impact crashes with opposing vehicles.
* **Multi-vehicle crashes** (7.6%): Complex collisions causing pileups.
* **Impaired driving** (6.0%): Impaired reaction times and vehicle control elevate crash severity.
* **Vulnerable road users** (pedestrians, cyclists, motorcyclists; 5.7%): Lack of protection increases the likelihood of severe injuries or fatalities.

Additional factors include roadway type (highways, urban arterials; 5.3%), regional differences (5.1%), and nighttime/rush-hour crashes (3.7%). Weather (0.99%) and road surface conditions (1.26%) had minimal impact in Virginia.

***Figure 1.*** *Displaying the most important features causing sever crashes in the state of Virginia*.

**Recommendations**

To reduce severe crashes, VDOT and policymakers should prioritize:

* **Speed Enforcement**: Deploy speed cameras and increase patrol operations on high-risk roads (e.g., highways, urban arterials), focusing on high-crash areas and peak risk times (rush hours, nighttime), with clear signage to deter speeding and ensure compliance.
* **Infrastructure**: Install median barriers on highways to prevent head-on collisions and enhance pedestrian crossings in urban areas.
* **Enforcement**: Increase DUI patrols at night and during rush hours on high-risk roads.
* **Public Awareness**: Launch campaigns for impaired driving and cyclist/pedestrian safety.
* **Regional Focus**: More focus on rural (undivided roads) and urban (dense traffic) districts.

**Conclusion**

This project illustrates that machine learning can effectively predict severe crash risks and inform precise safety interventions. Through targeting key factors such as speed differences, head-on collisions, impaired driving, and vulnerable road users, Virginia can substantially reduce fatalities and injuries. Ongoing data collection and model enhancement will further bolster the state’s road safety strategy, supporting its dedication to public safety.

**Reference**: Di, S. (2025, March 20). Crash Data Basic. Virginia Roads. Retrieved April 14, 2025, from <https://www.virginiaroads.org/datasets/VDOT::crash-data-1/about?layer=0>