EFFECT OF WEATHER STIMULI AND ROAD POINTS OF INTEREST IN CLASSIFYING TRAFFIC ACCIDENT SEVERITY

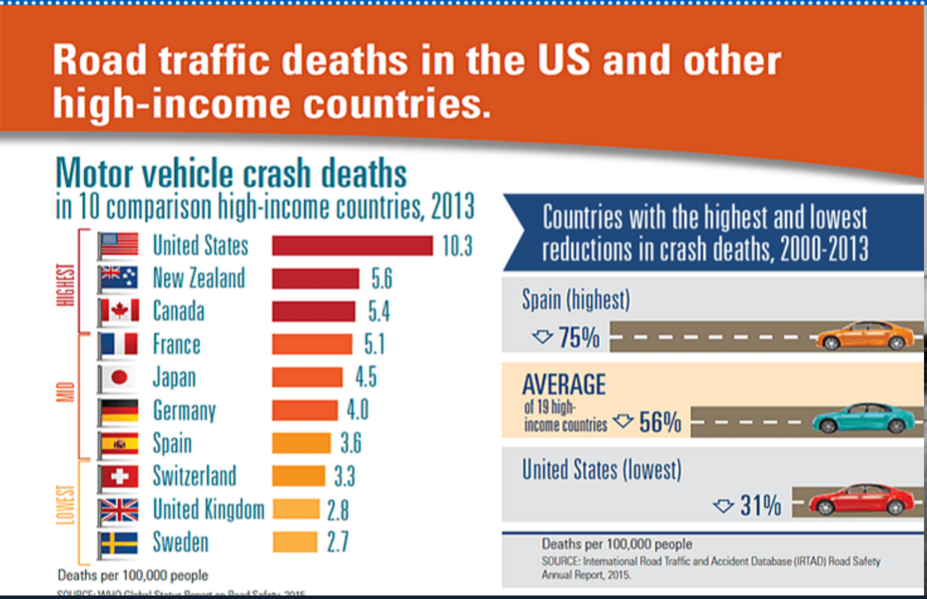
Adithya Narayanan

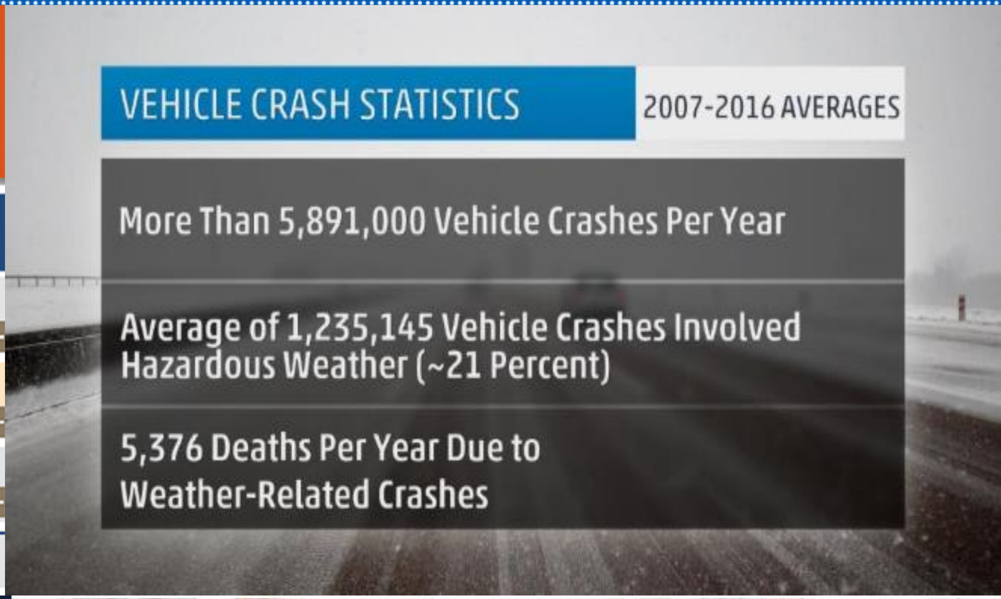
Adit Bhargav Modi

Akhil Gaddipati

University at Buffalo The State University of New York

MOTIVATION





Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk

The automated car lacked "the capability to classify an object as a pedestrian unless that object was near a crosswalk," an NTSB report said.

https://weather.com/safety/winter/news/weather-fatalities-car-crashes-accidents-united-states
http://www.prism.engineering/traffic-facts-international-by-country.html
https://weather.com/safety/winter/news/weather-fatalities-car-crashes-accidents-united-states
https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281

Problem Statement

Analyzing and modeling the effect that weather stimuli and road points-of-interest (POIs) bear on accident severity

LITERATURE REVIEW

Past Work

- Malin et al. (2019) analyzed accidents in Finland across all main road networks using Palm Probabilites by also accounting for driver behavior and found that both**, road type and weather conditions, were found to have significant impact on accident severity.**
- Sherretz et al. (1978) analyzed the data from seven southern Illinois cities and found a **linear relationship between rainfall and occurrence of traffic accidents.**
- The Montella et al. (2012) study analyzed two wheeler crashes in Italy. **Alignment subsets such as intersection, curved roads and so on, were observed to be related to crashes. A lot of times, good weather did not bear any impact on crashes.**
- Moosavi et al. (2019) used deep neural networks to predict accident risks in real time- up to 15 minute durations within a precision of 5sq. Km radius using the same dataset.
- Theofilatos, A. (2017) **did not find a strong relationship between weather and accident severity.**

Methodologies from Past Work

- Bayesian Logistic Regression
- CART
- Random Forest
- SVMs, Neural Networks, Hybrid Decision Tree-NN methods
- Multiple Objective Particle Swarm Optimization (MOPSO)

Pitfalls: Niche Data, leading to high bias and low variance across regions or climatic conditions- doesn't capture all conditions well

DATA DESCRIPTION

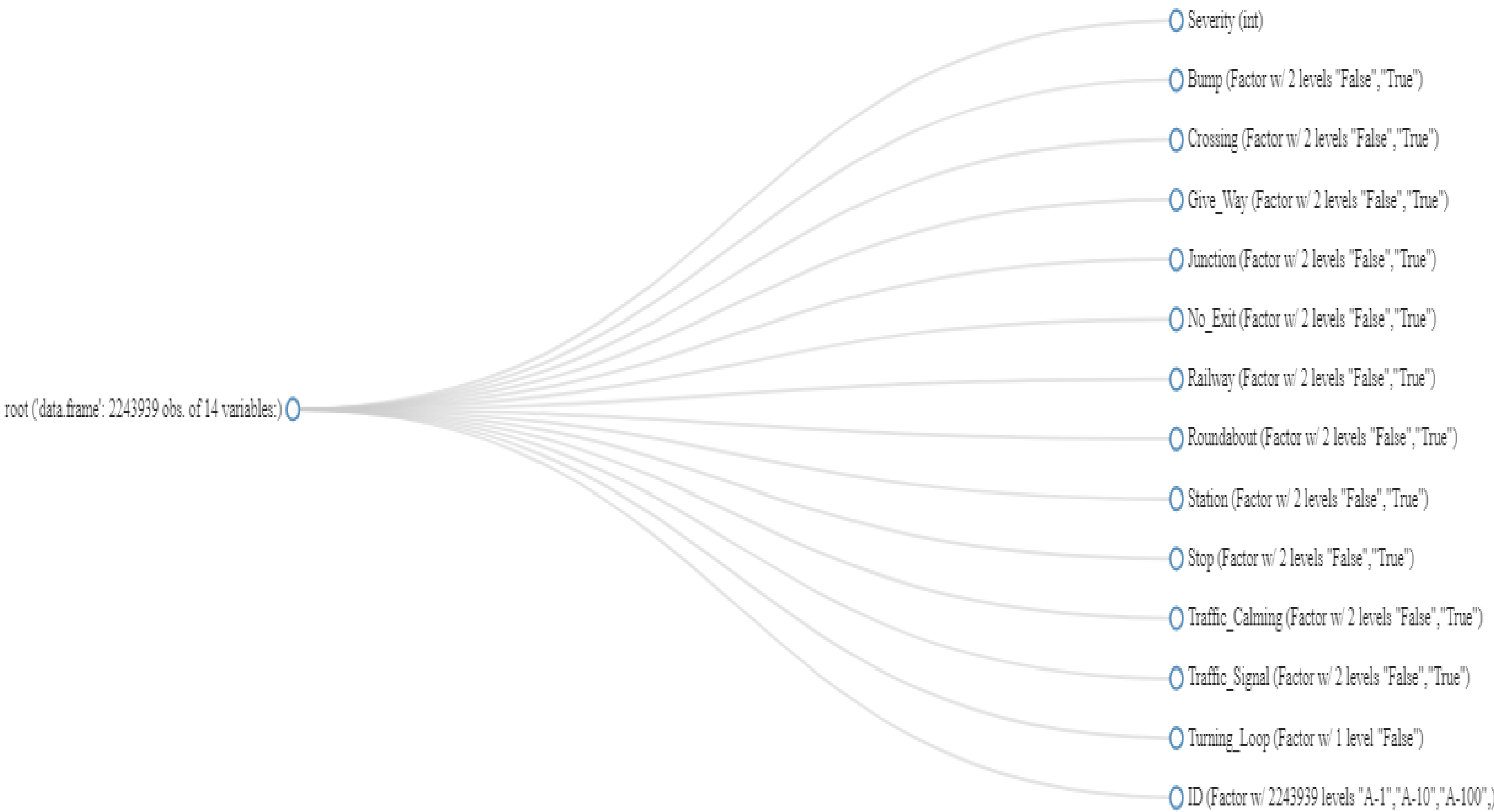
Initial Data : US Accidents

- 2.25 Million Observations (obtained using APIs that provide traffic event data)
- Spanning 49 states
- Data collected from February 2016 to March 2019
- 49 columns including weather data and Road POIs
- Certain variables such as Source, TMC, Start and End Times, Start and End Latitudes and Longitudes, and some other geographical variables were removed since they were not included in our primary scope of interest.
- Variables containing more than 80% of missing values were also removed.

- Variable of Interest: Severity
    - *Level 1 : 2 Minutes and 30 Seconds (814)*
    - *Level 2 : 3 Minutes and 15 Seconds (1455524)*
    - *Level 3: 8 Minutes (715582)*
    - **Level 4: 18 Minutes (72002)**

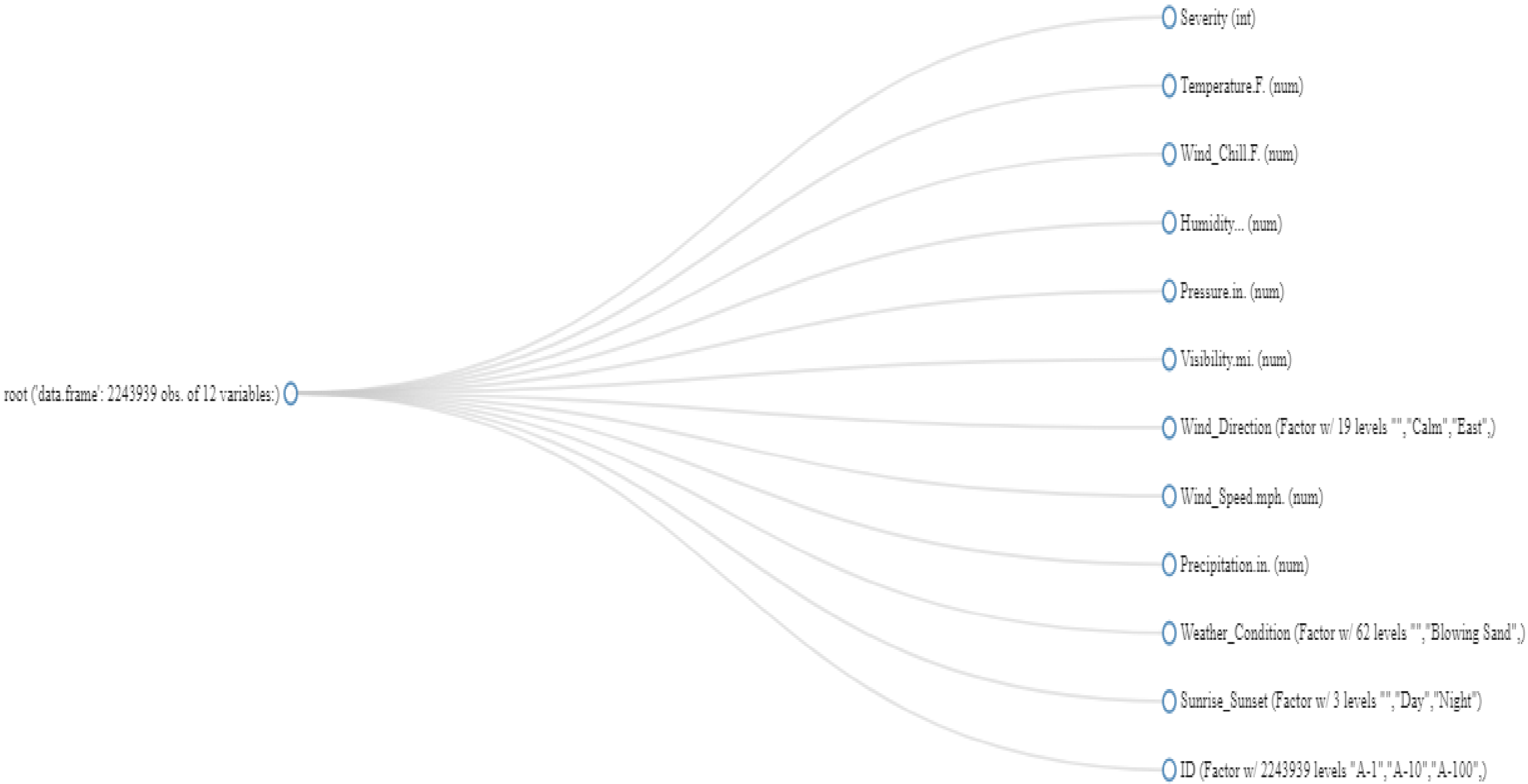
Moosavi et al.(2019)

Road POIs



root ('data.frame': 2243939 obs. of 14 variables:)
Severity (int)
Bump (Factor w/ 2 levels "False","True")
Crossing (Factor w/ 2 levels "False","True")
Give_Way (Factor w/ 2 levels "False","True")
Junction (Factor w/ 2 levels "False","True")
No_Exit (Factor w/ 2 levels "False","True")
Railway (Factor w/ 2 levels "False","True")
Roundabout (Factor w/ 2 levels "False","True")
Station (Factor w/ 2 levels "False","True")
Stop (Factor w/ 2 levels "False","True")
Traffic_Calming (Factor w/ 2 levels "False","True")
Traffic_Signal (Factor w/ 2 levels "False","True")
Turning_Loop (Factor w/ 1 level "False")
ID (Factor w/ 2243939 levels "A-1","A-10","A-100",)

Weather Variables



DATA PREPROCESSING AND CLEANING

Data Preprocessing and Cleaning

US Accidents Data
2.25 Million observations
49 variables

Data Cleaning

Relevant Feature Selection
2.25 Million observations
~18 variables

Under Sampling

Subset

Data of Four Different States (Arizona, Washington, Florida, Pennsylvania)

Variables Considered

WEATHER VARIABLES

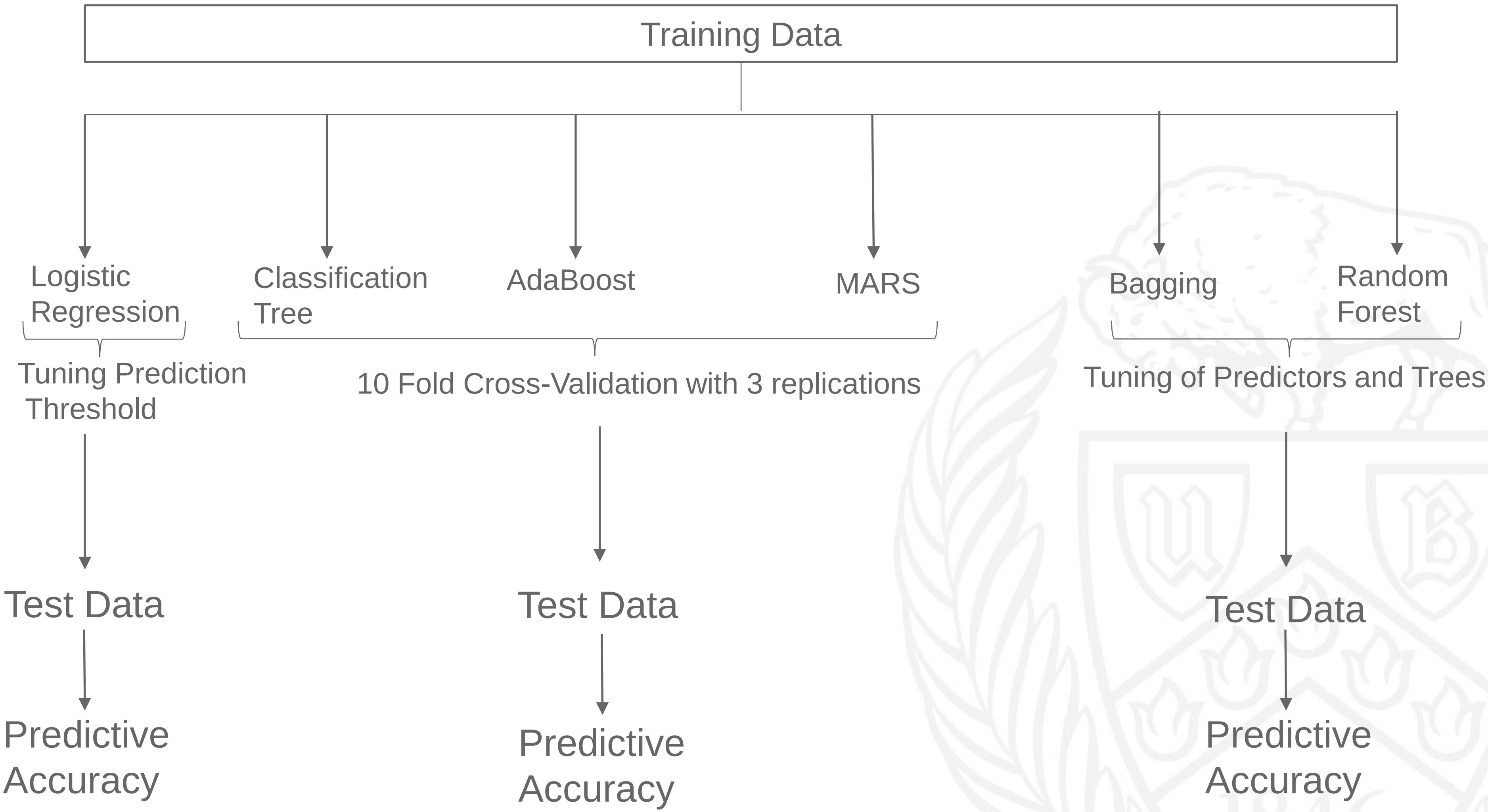
Temperature
Wind Direction
Wind Speed
Humidity
Pressure
Visibility

ROAD POIs~

Amenity
Crossing
Station
Stop
Give Way
Traffic Calming
Junction
No Exit
Traffic signal
Railway Crossing
Bump
Roundabout

METHODOLOGY

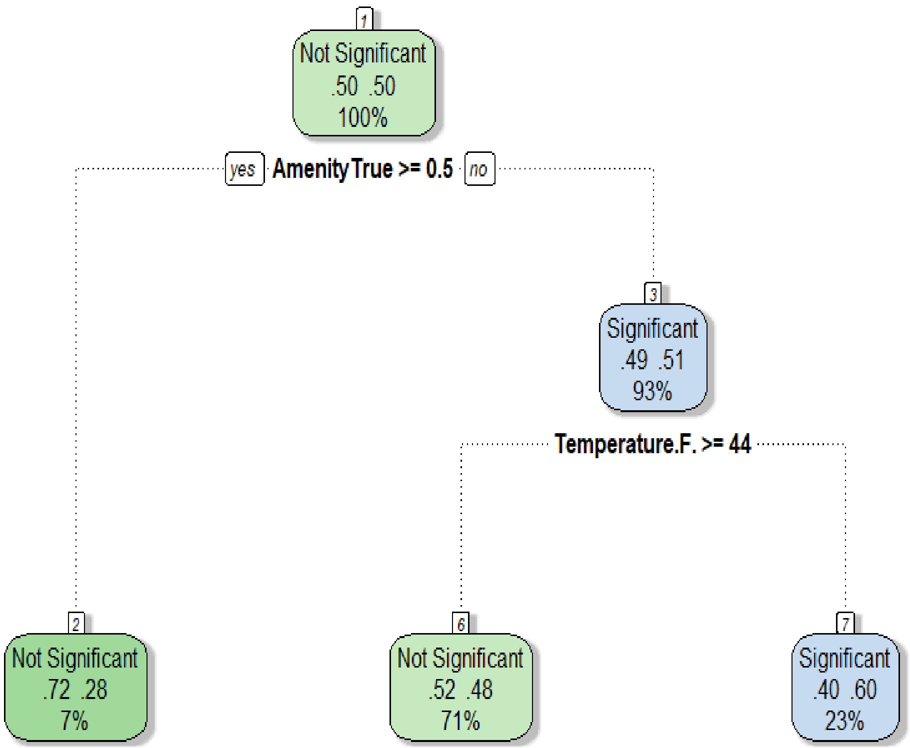
State Subsets



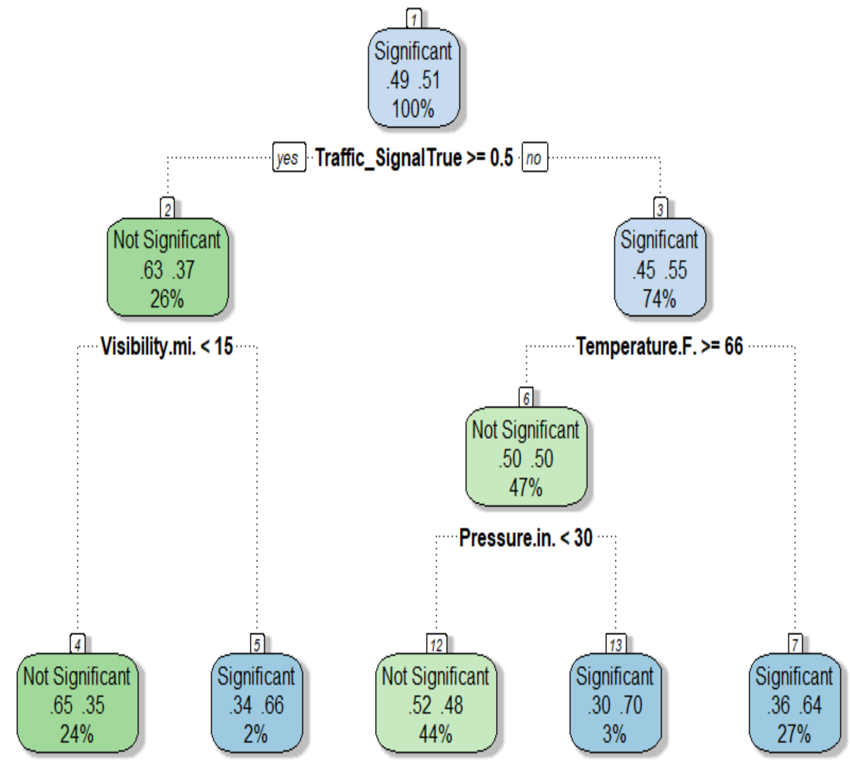
RESULTS

Classification Trees

Washington

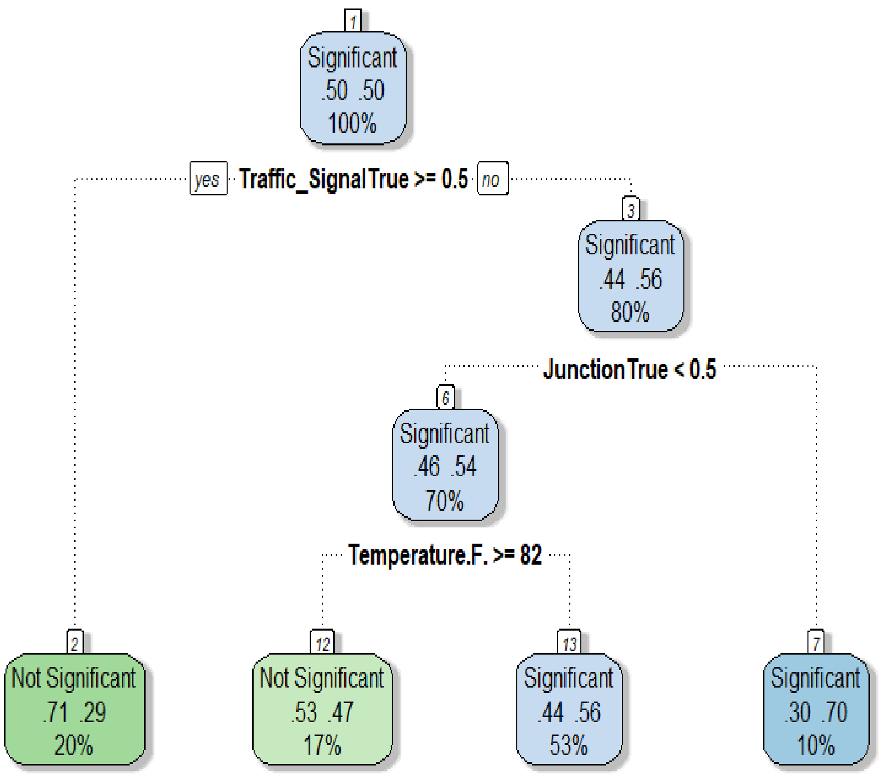


Arizona

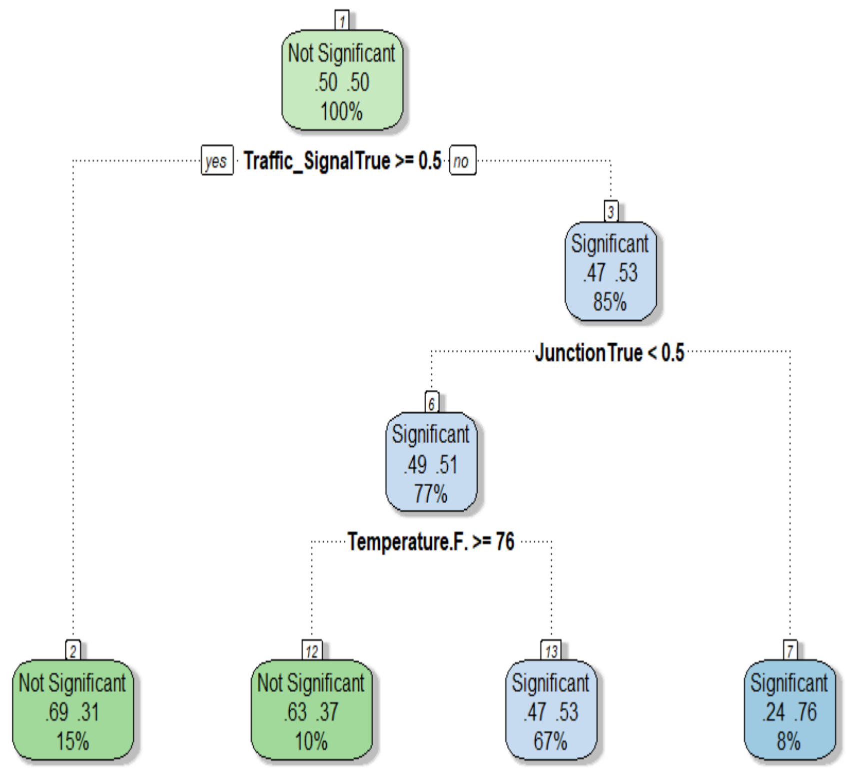


Classification Trees

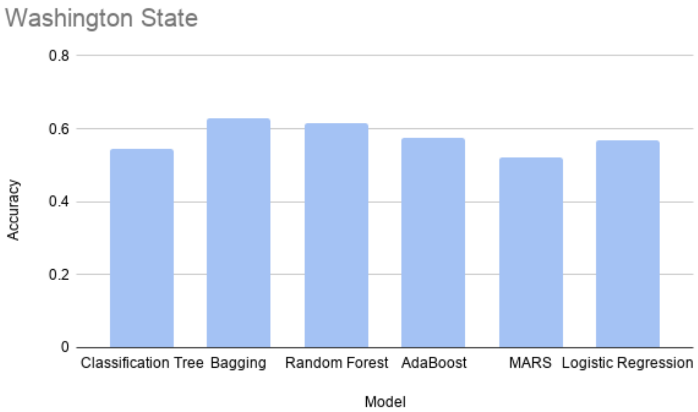
Florida

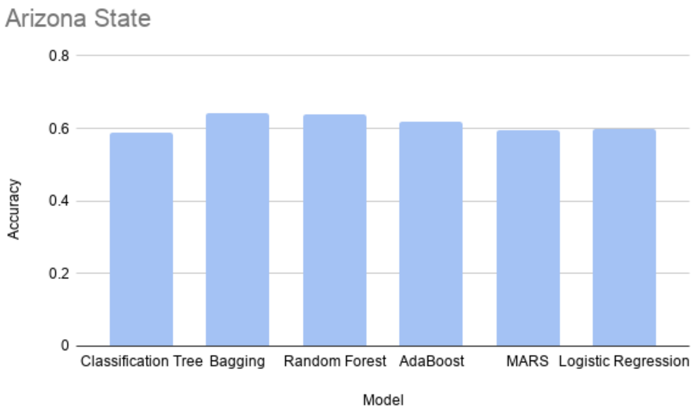


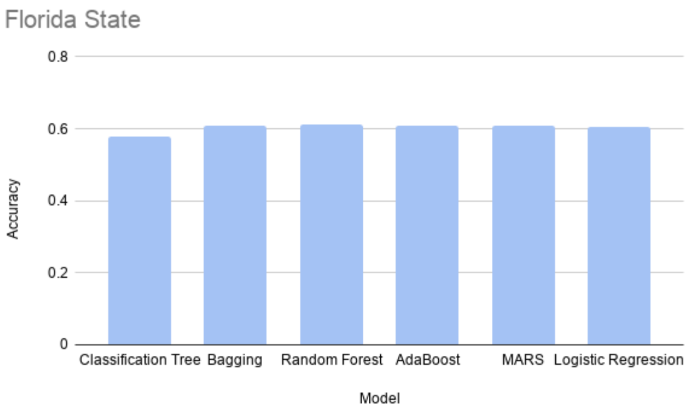
Pennsylvania

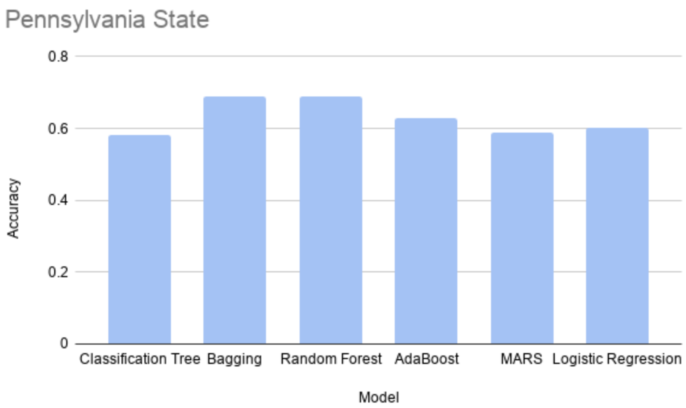


Predictive Accuracy



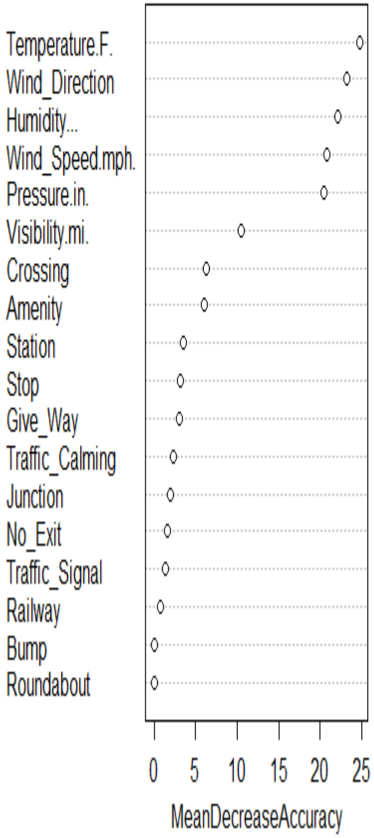


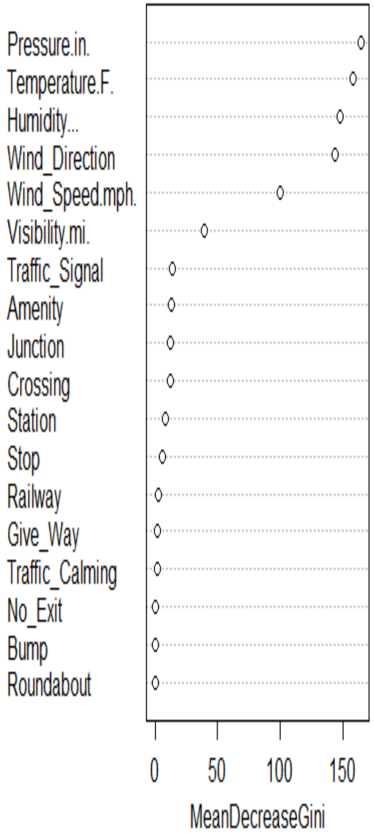




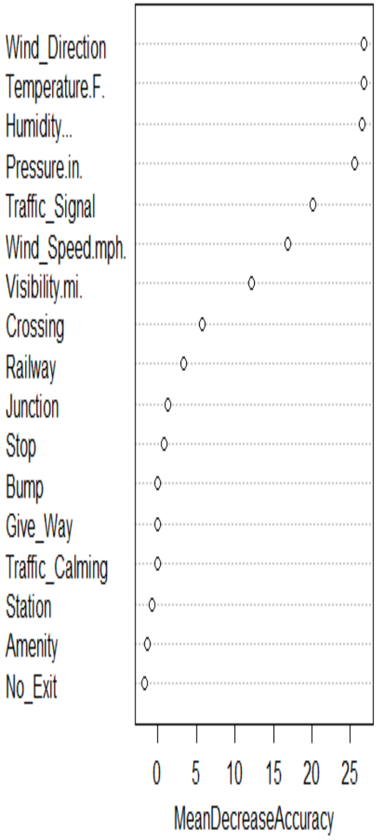
Classification Trees

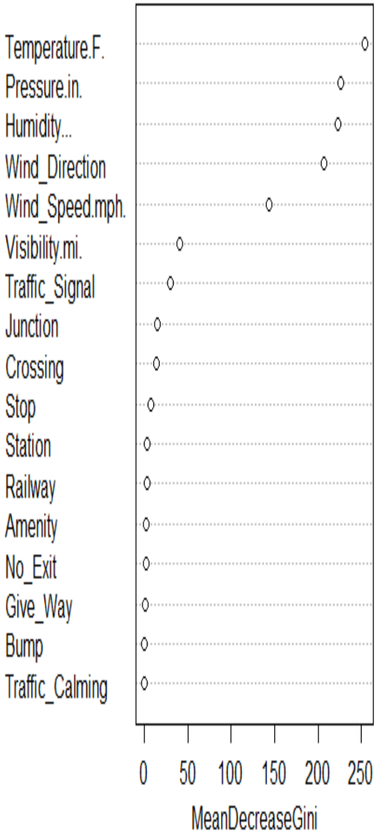
Washington





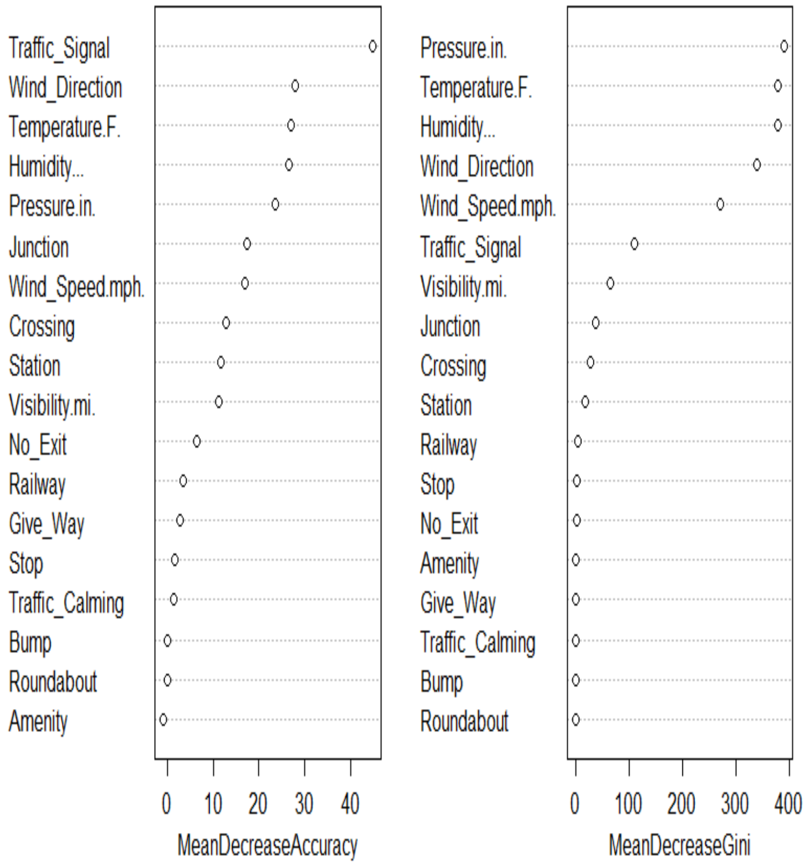
Arizona



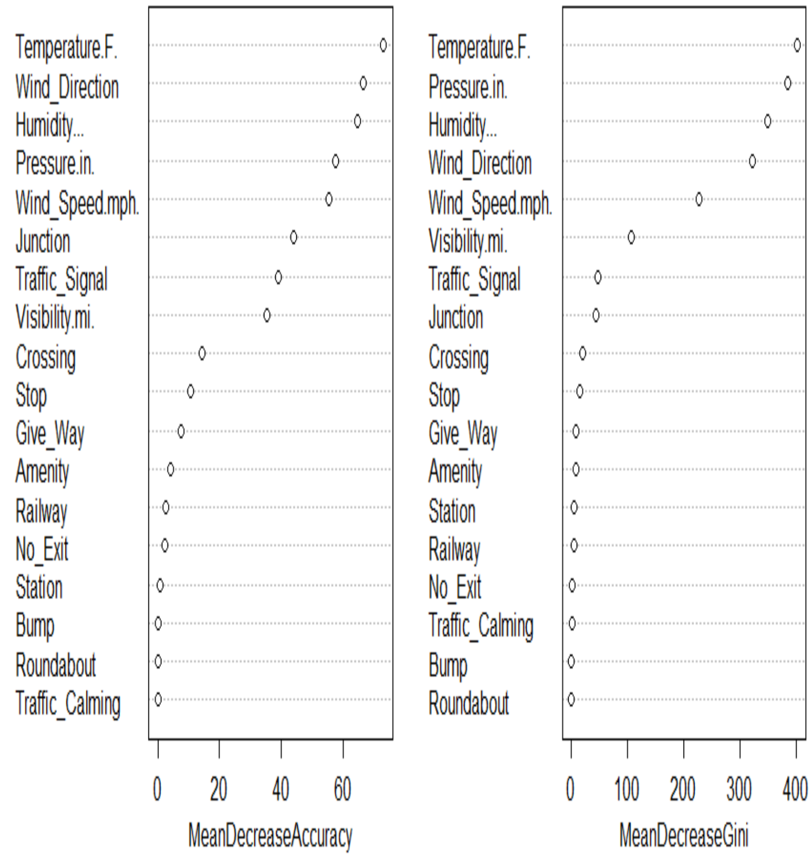


Classification Trees

Florida



Pennsylvania



Predictive Accuracy

State	Model	Accuracy	Error
	Classification Tree	0.5433884	0.4566116
	Bagging	**0.6301653**	**0.3698347**
	Random Forest	**0.6136364**	**0.3863636**
	AdaBoost	0.5743802	0.4256198
	MARS	0.5206612	0.4793388
Washington	Logistic Regression	0.5671488	0.4328512
	Classification Tree	0.5900234	0.4099766
	Bagging	**0.6430242**	**0.3569758**
	Random Forest	**0.6383476**	**0.3616524**
	AdaBoost	0.6196415	0.3803585
	MARS	0.5962588	0.4037412
Arizona	Logistic Regression	0.6001559	0.3998441

State	Model	Accuracy	Error
	Classification Tree	0.5783299	0.4216701
	Bagging	0.6083506	0.3916494
	Random Forest	**0.6114562**	**0.3885438**
	AdaBoost	**0.6100759**	**0.3899241**
	MARS	0.6080055	0.3919945
Florida	Logistic Regression	0.605245	0.394755
	Classification Tree	0.5818831	0.4181169
	Bagging	**0.68747**	**0.31253**
	Random Forest	**0.6894244**	**0.3105756**
	AdaBoost	0.6282909	0.3717091
	MARS	0.5885765	0.4114235
Pennsylvania	Logistic Regression	0.6024096	0.3975904

Conclusion and Future Work

- A combination of Weather and Road POIs was found to be a very poor predictor of accident severity- contrary to several studies. Despite Weather variables dominating, accuracy remains fairly low.
- Weather related variables were found to be the most important predictors across different states and Road POIs performed poorly in terms of variable importance (could be due to their sparsity).
- Bagging and Random Forest ensemble were found to perform marginally better out-of-sample in this data, despite the use of models such as MARS and AdaBoost – they handle variance better.
- Future direction – Studying driver behavior, eliminating driver behavior by studying autonomous vehicle crashes, topic modeling across insurance statements to understand driver/passenger understanding of the situation etc.

THANK YOU