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

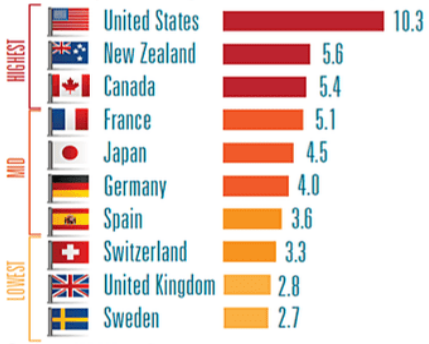
Adithya Narayanan
Adit Bhargav Modi
Akhil Gaddipati

University at Buffalo The State University of New York

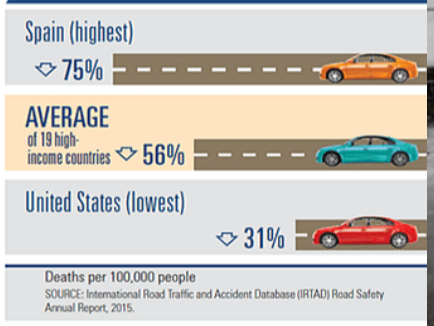
MOTIVATION

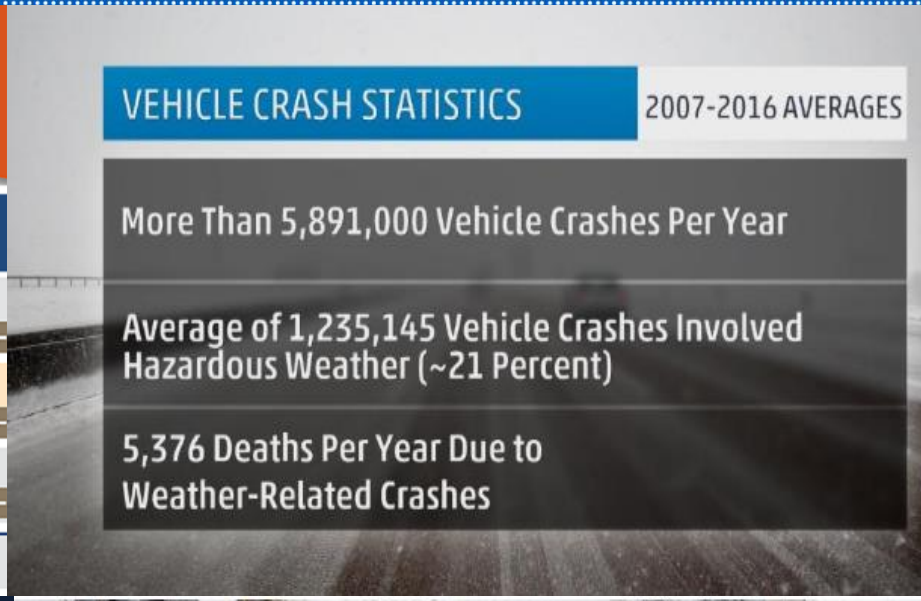
Road traffic deaths in the US and other high-income countries.

Motor vehicle crash deaths
in 10 comparison high-income countries, 2013



Countries with the highest and lowest reductions in crash deaths, 2000-2013





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

University at Buffalo The State University of New York

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

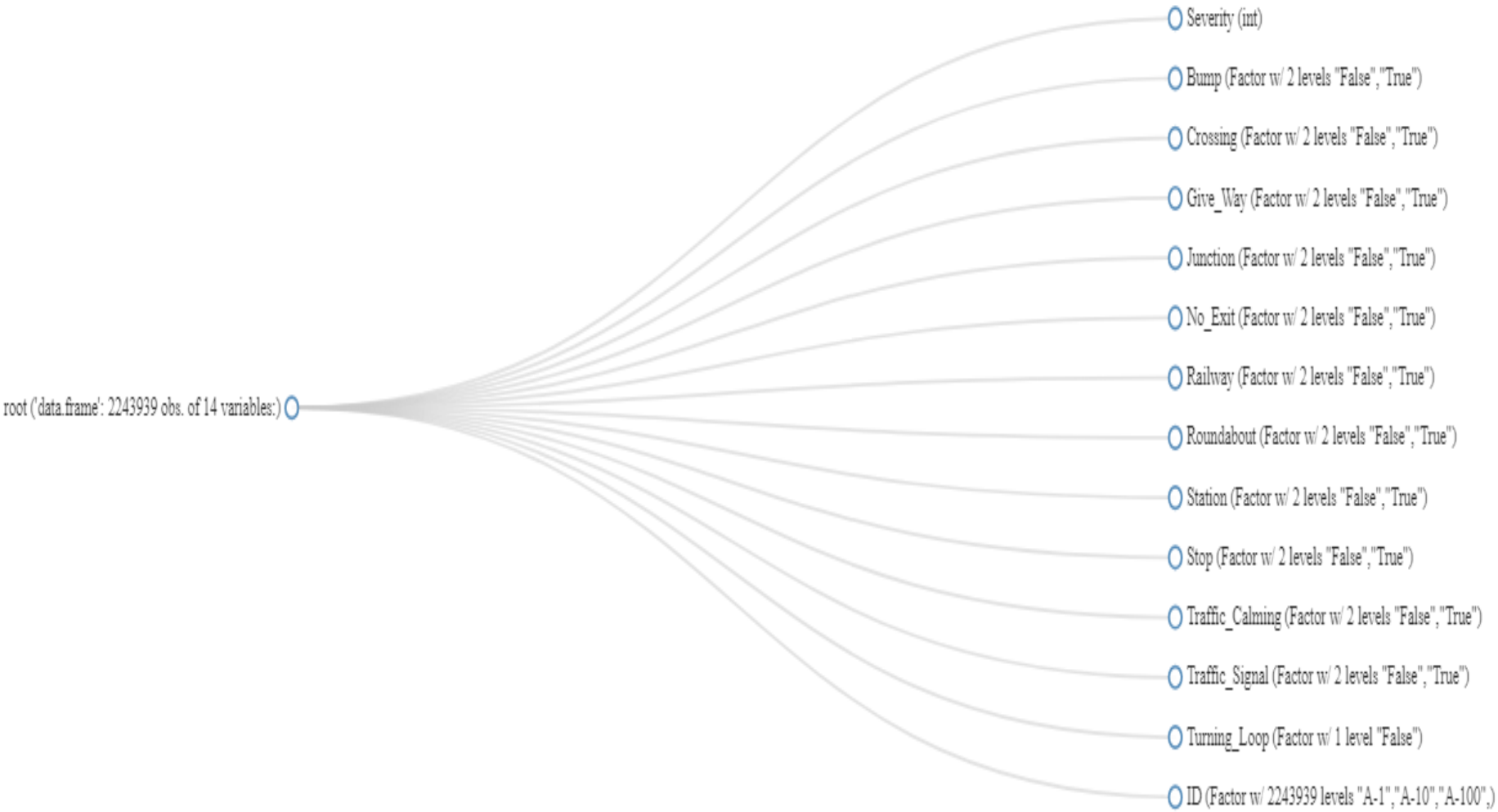
Initial Data : US Accidents

- 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.

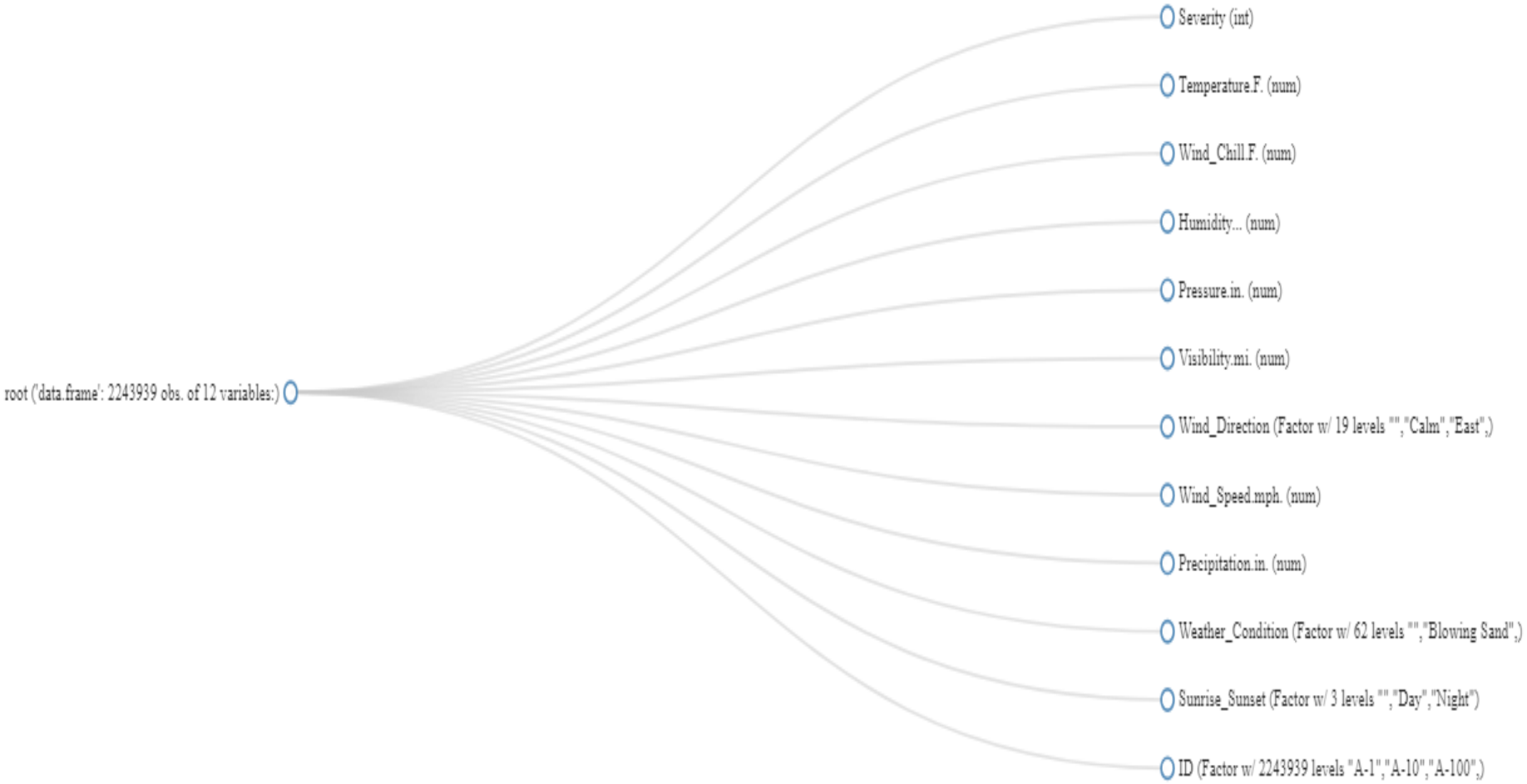
- Variable of Interest: Severity
  - *Level 1 : 2 Minutes and 30 Seconds (814)*
  - *Level 2 : 3 Minutes and 15 Seconds (1455524)*
  - *Level 3: 8 Minutes (715582)*
  - **Level 4: 18 Minutes (72002)**

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",)

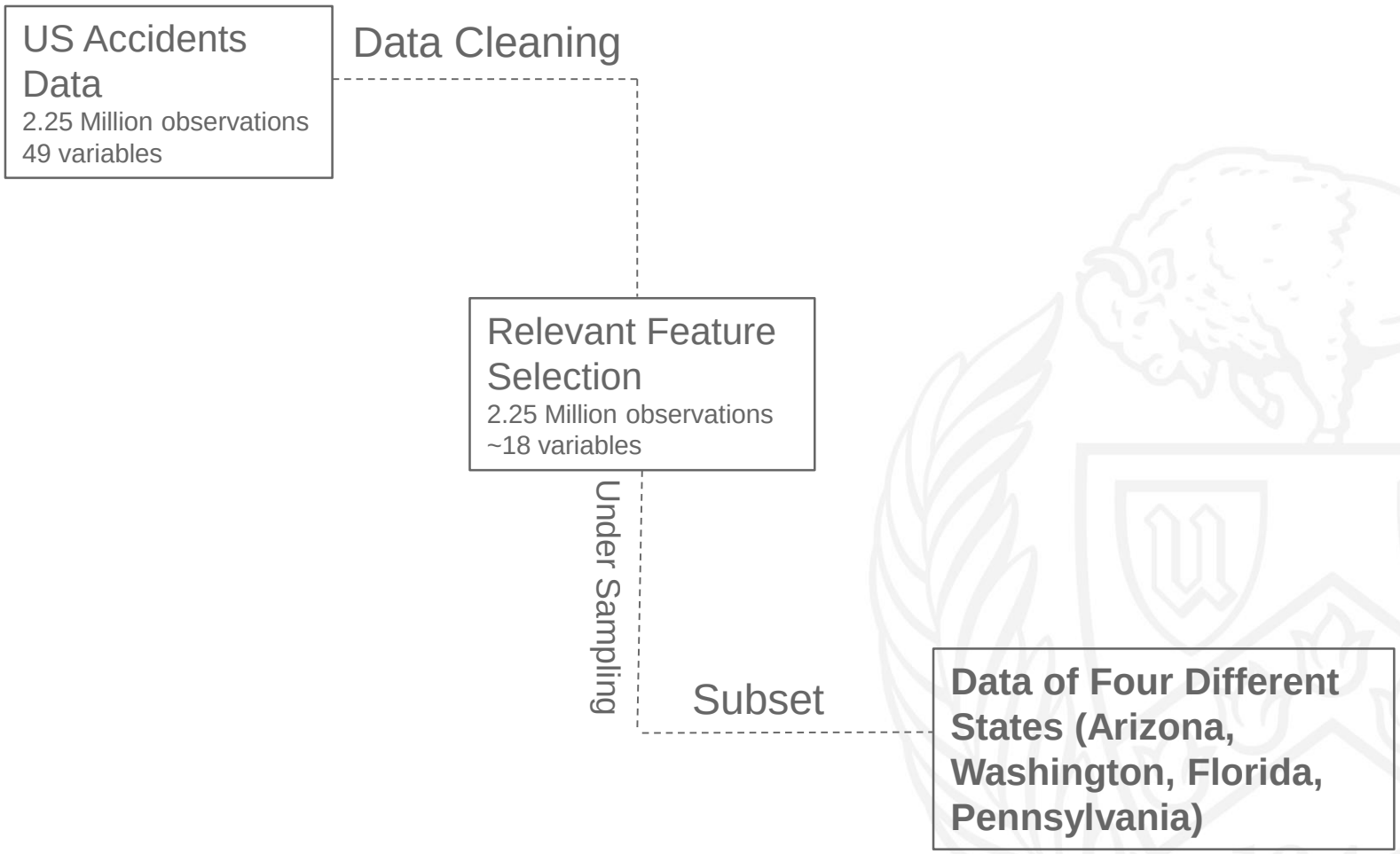
Weather Variables



University at Buffalo The State University of New York

DATA PREPROCESSING AND CLEANING

Data Preprocessing and Cleaning



Variables Considered

WEATHER VARIABLES

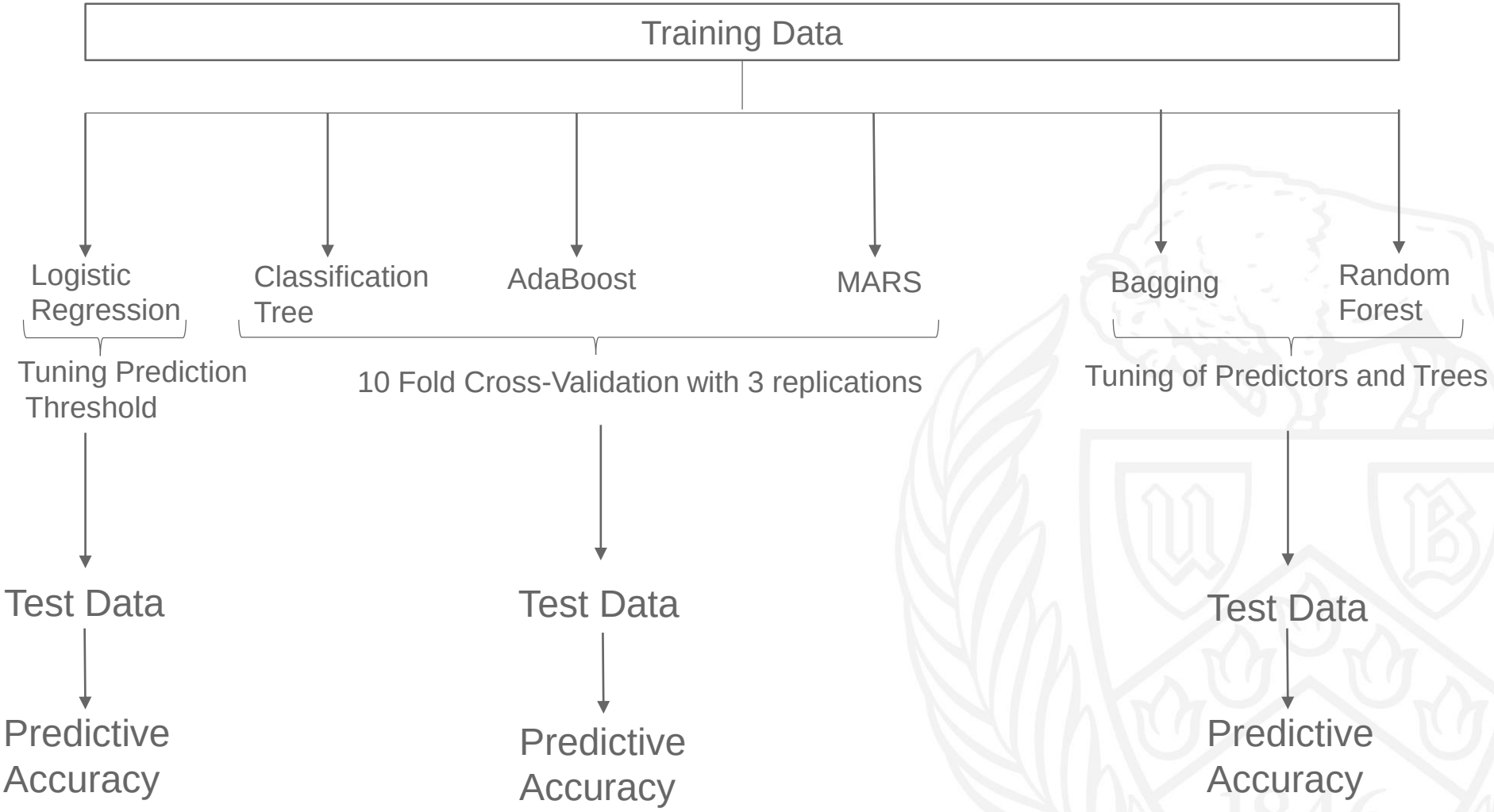
Temperature
Wind Direction
Wind Speed
Humidity
Pressure
Visibility

ROAD POIs~

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

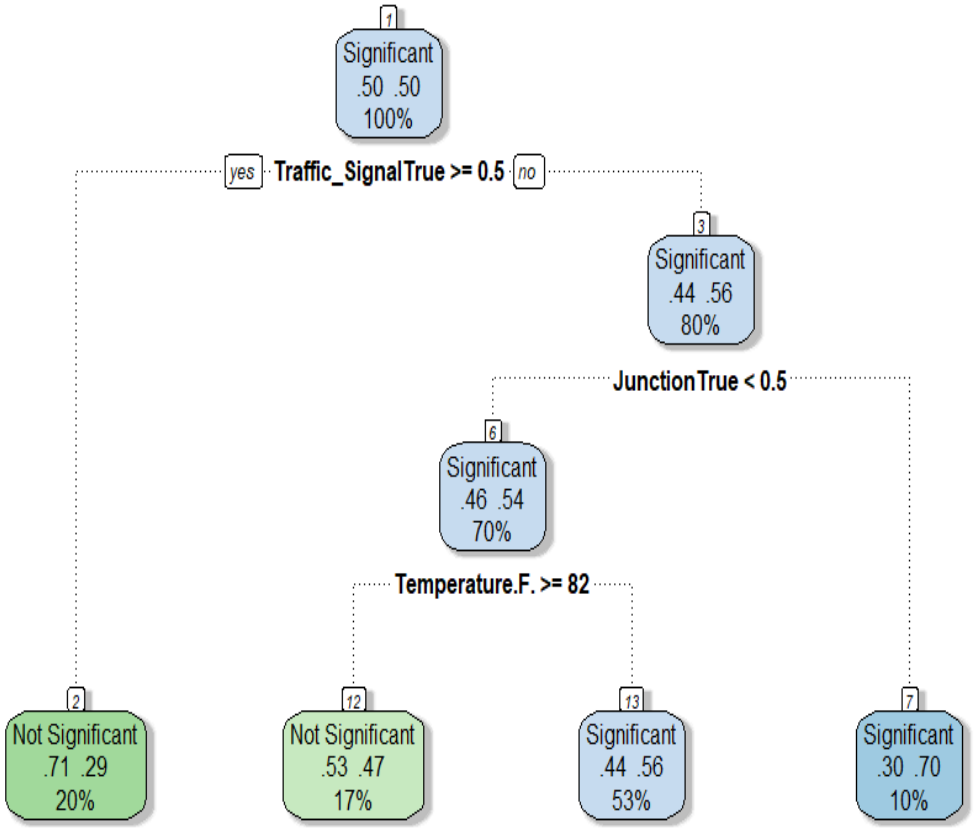
METHODOLOGY

State Subsets

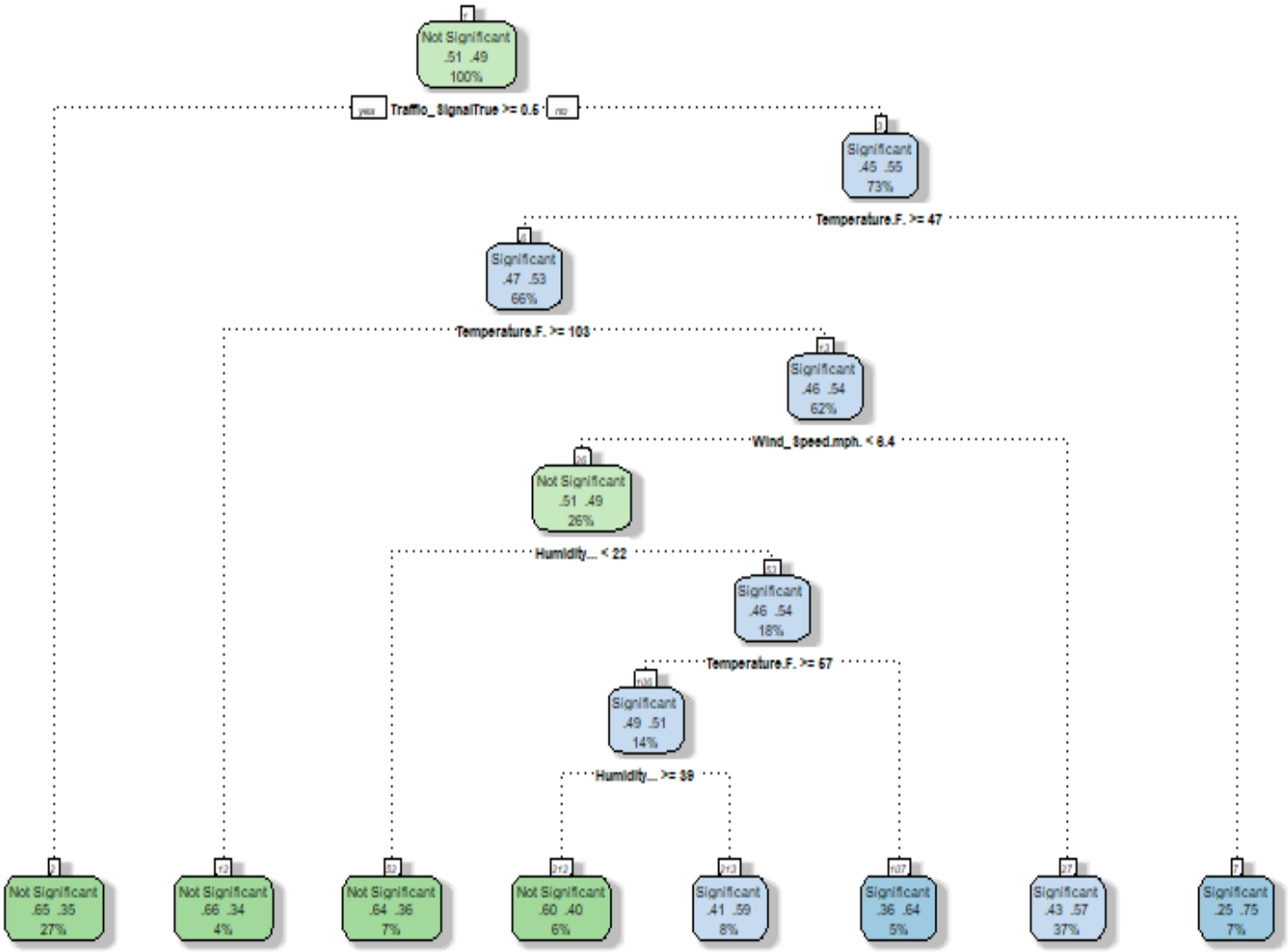


RESULTS

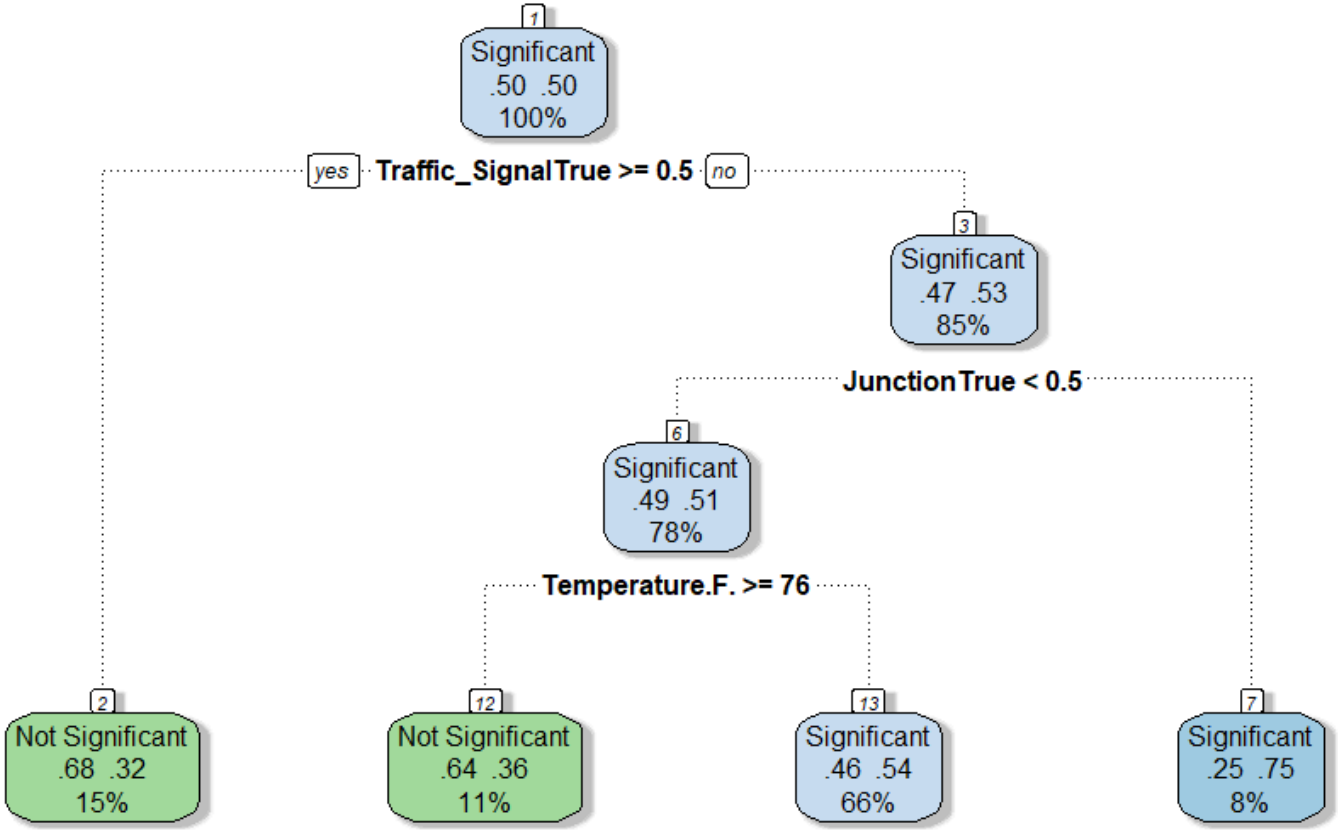
Classification Trees- Florida



Classification Trees-Arizona

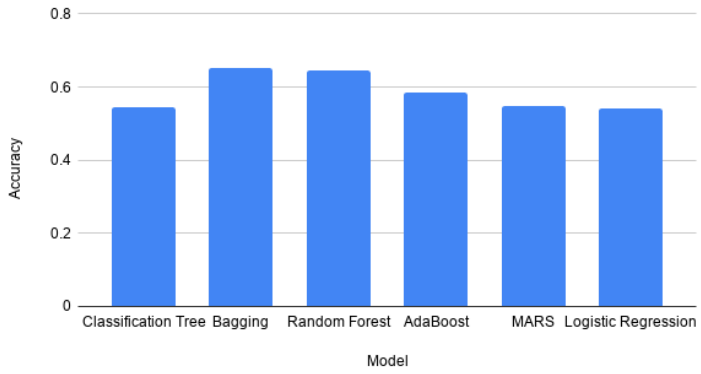


Classification Trees-Pennsylvania

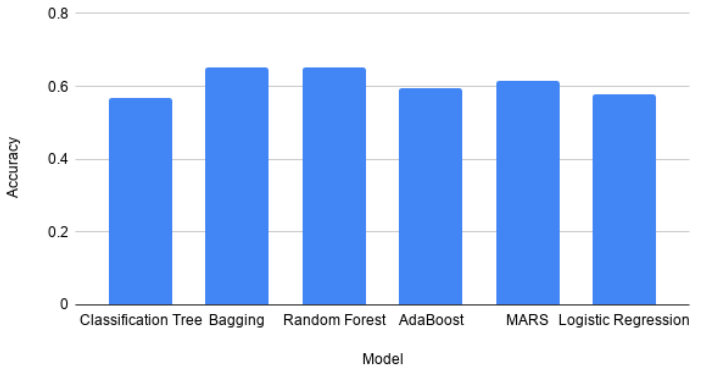


Predictive Accuracy

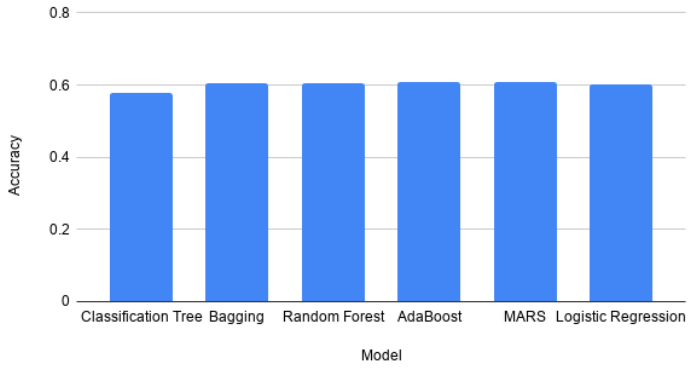
Washington



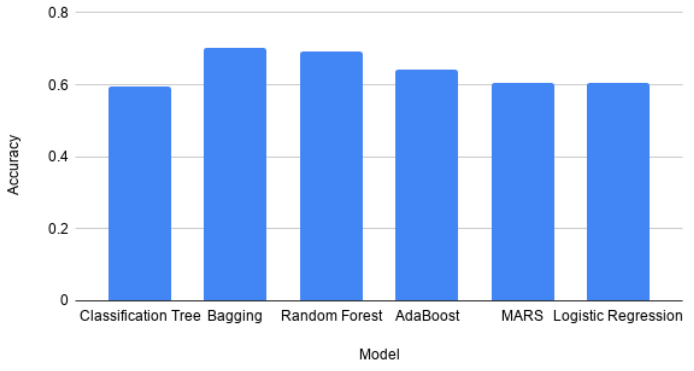
Arizona



Florida

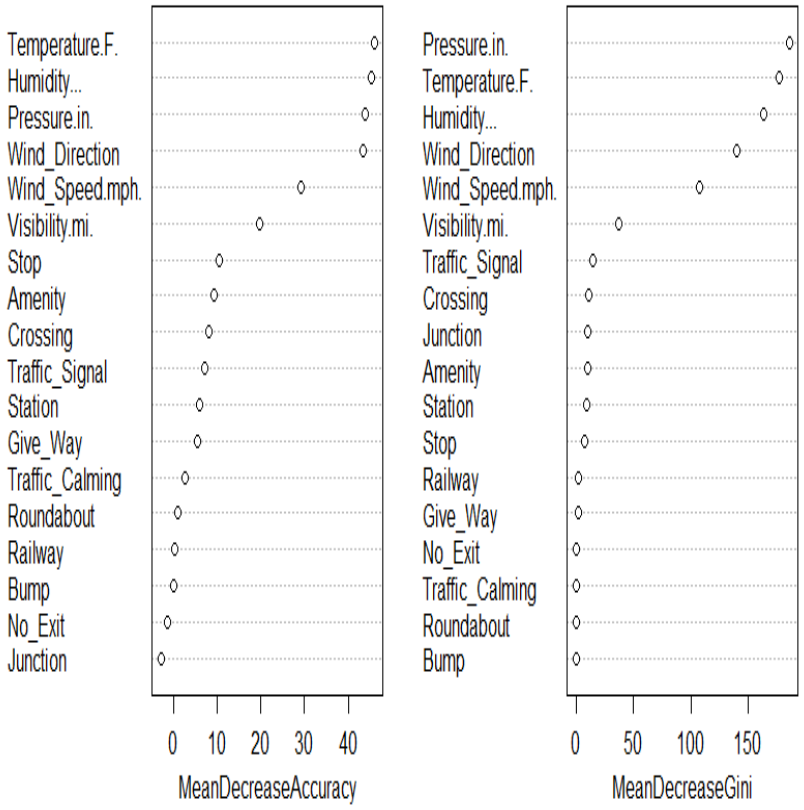


Pennsylvania

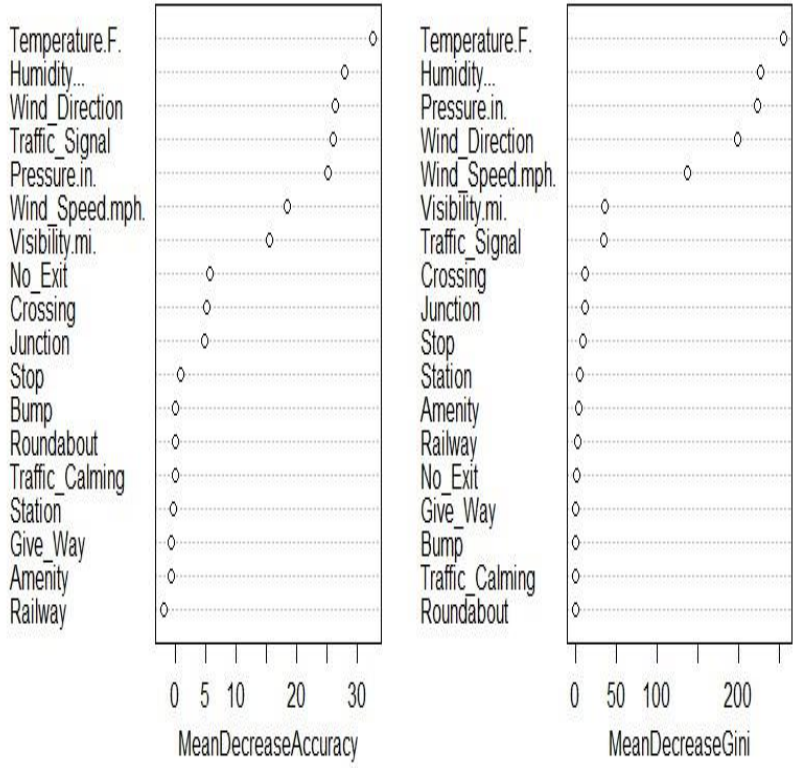


Classification Trees

Washington

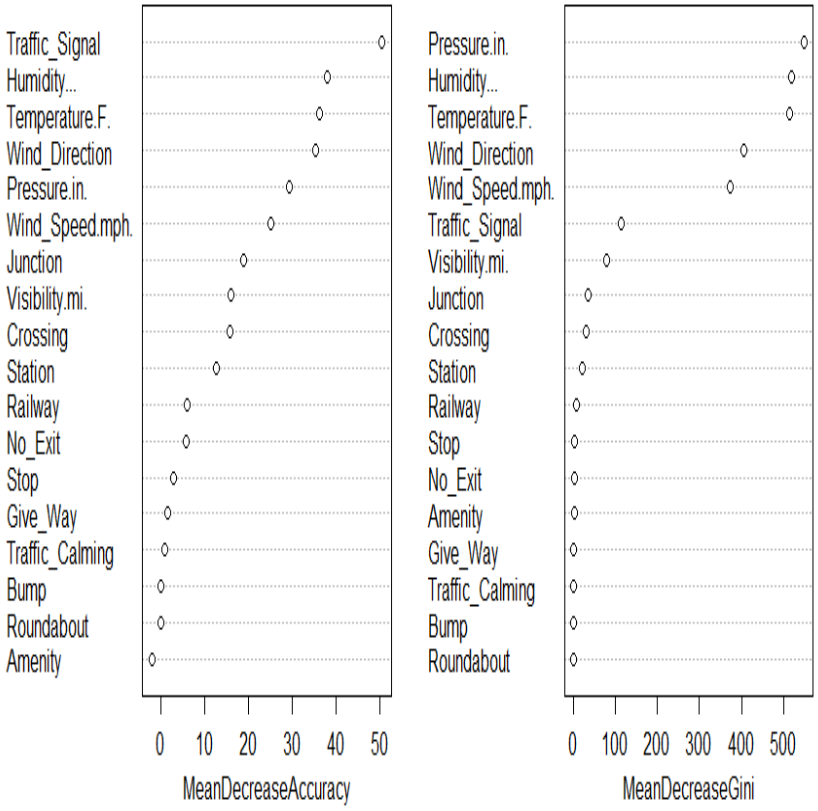


Arizona

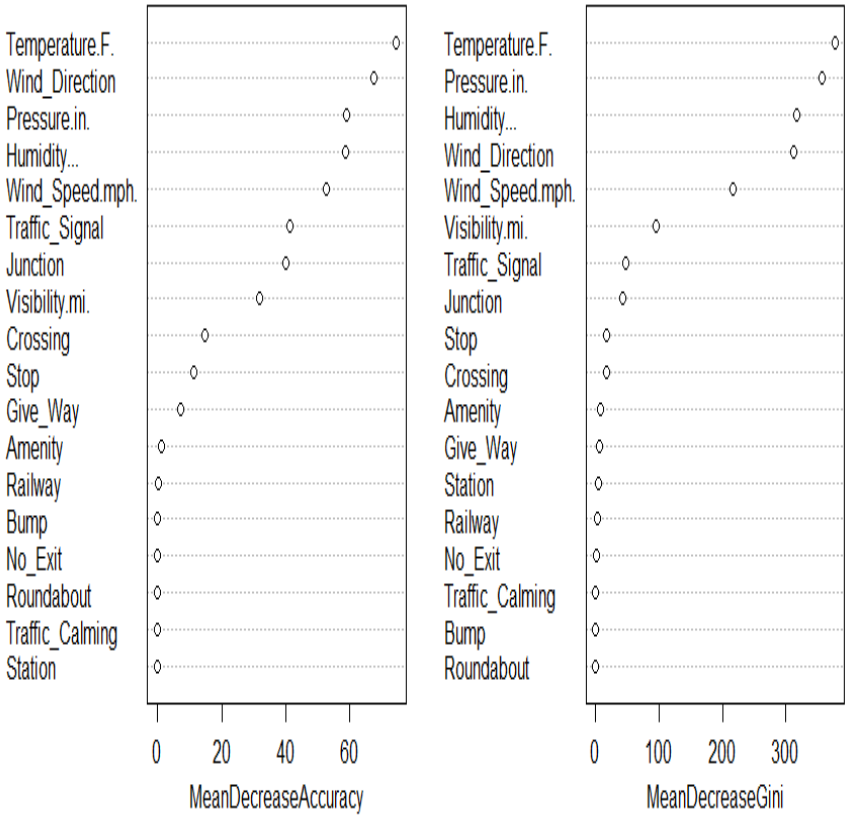


Classification Trees

Florida



Pennsylvania



Predictive Accuracy

State	Model	Accuracy	Error	State	Model	Accuracy	Error
	Classification Tree	0.54414481	0.45585519		Classification Tree	0.5783299	0.4216701
	Bagging	0.6505771	0.3494229		Bagging	0.6066253	0.3933747
	Random Forest	0.6463799	0.3536201		Random Forest	0.6045549	0.3954451
	AdaBoost	0.5844701	0.4155299		AdaBoost	0.6090407	0.3909593
	MARS	0.5466946	0.4533054		MARS	0.6080055	0.3919945
Washington	Logistic Regression	0.5414481	0.4585519	Florida	Logistic Regression	0.6011042	0.3988958
	Classification Tree	0.5671168	0.4328832		Classification Tree	0.5958873	0.4041127
	Bagging	0.6536934	0.3463066		Bagging	0.7027269	0.2972731
	Random Forest	0.6521048	0.3478952		Random Forest	0.6937863	0.3062137
	AdaBoost	0.5949166	0.4050834		AdaBoost	0.6428252	0.3571748
	MARS	0.6155679	0.3844321		MARS	0.606616	0.393384
Arizona	Logistic Regression	0.5782367	0.4217633	Pennsylvania	Logistic Regression	0.606169	0.393831

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