



# Predicting Car Accident Injuries in New York State

DANIELO BENNETT

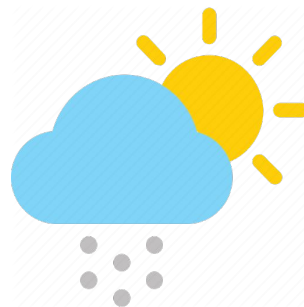
# Problem Statement

- According to the New York State DMV, the yearly average number of accidents statewide is ~300,000. Many of those result in injuries.
- Ambulance/EMT response for accidents where injuries occur are crucial for survival of the victims.
- **Solution:** Assist 911 operators with dispatching additional ambulances to the scene of an accident if they reasonably suspect that there are injuries. This decision would be made based on information collected from 911 calls.

# Methodology

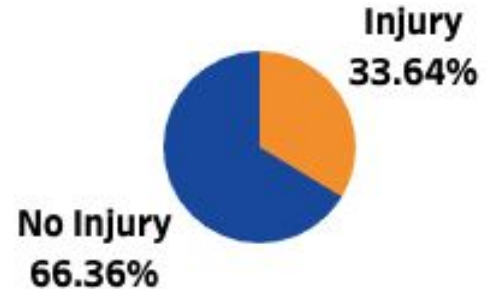


timeanddate.com



# Methodology - Data set

- **560K** obs of accidents in NYS (Excluding NYC) from 2014-2016
- **2-to-1** class imbalance
- 59 Features, some of which are:
  - Road Curvature & Slope
  - Object the vehicle collided with(vehicle, animal, etc)
  - County
  - Whether the accident occurred at a traffic device
  - Weather conditions

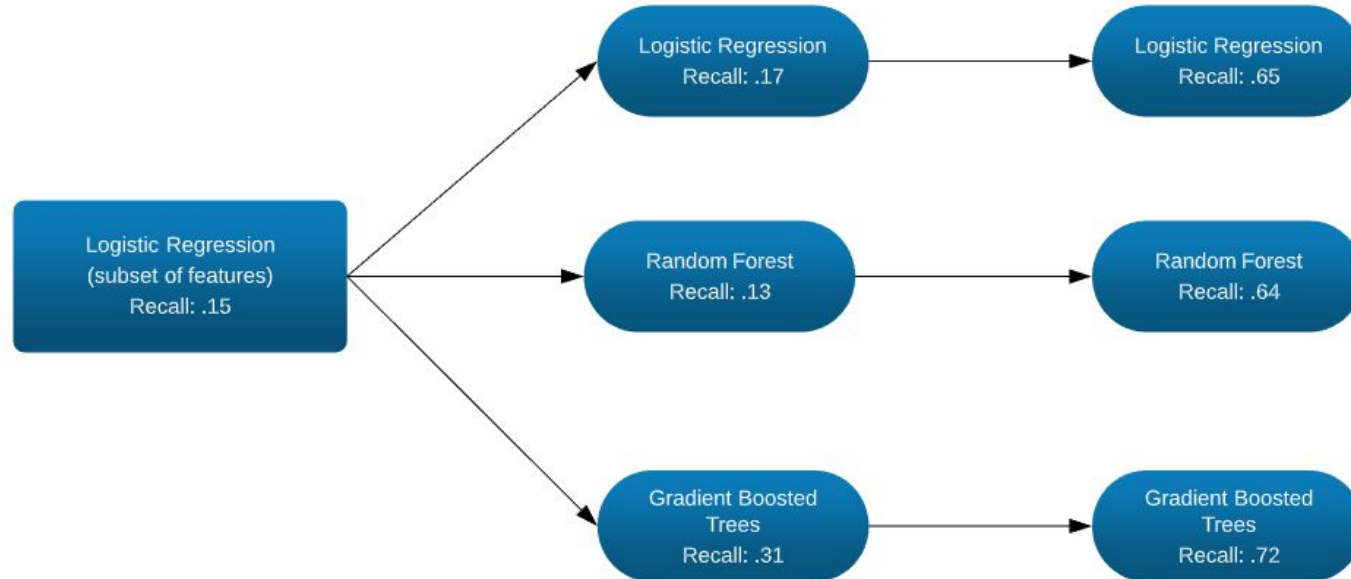


# Results - Model Testing

Selected Classification Metric:  
Recall

50% Threshold

32.5% Threshold



# Results - Best Models

- Gradient Boosted Tree: Recall = **0.72**
- Logistic Regression: Recall = **0.65**
- Logistic Regression Feature Importance:
  - Collisions with pedestrians: **0.908**
  - Collisions with bicycles: **0.486**
  - Collisions with Animals: **-0.484**
  - Number of Vehicles Involved: **0.421**

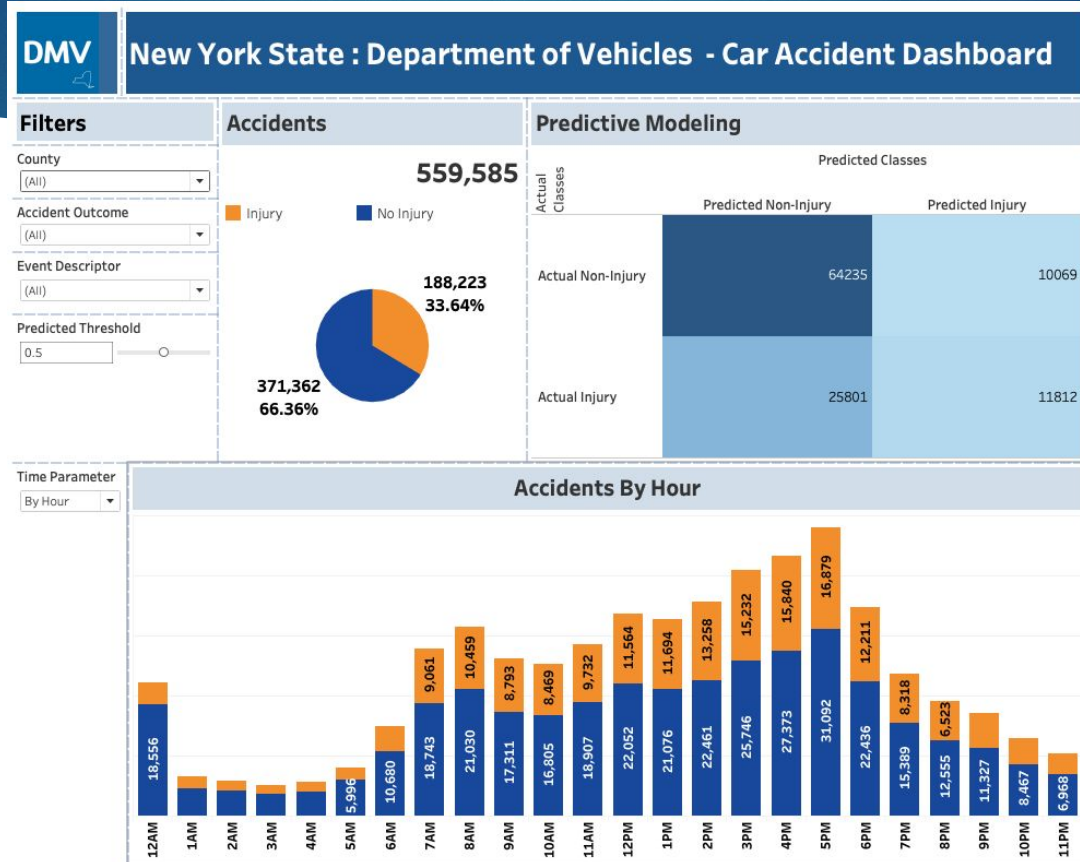
Gradient Boosted Tree confusion matrix

Actual outcome	Predicted outcome	
	No Injury	Injury
No Injury	36152	38152
Injury	10419	27194

# Takeaways

- Certain types of object that the car collides with tend to have strong weight in whether or not an injury occurs
- Boosted trees is the best model for prediction and Logistic Regression for Interpretation

# EDA Dashboard







Thank You!