We used Decision Tree Classifier to Classify the Severity of a US Road Accident on the basis of 50 features

We explored the different classification and sampling methods such as:
1.) Over sampling
2.) Logistic Regression
3.) Decision Tree Classifier
4.) Neural Network

Precision vs Sensitivity (Recall):
For this project, Sensitivity (Recall) is more important than Precision. Specially for higher severity values, since a false negative will have a worse impact than a false positive.

We chose Decision Tree Classifier (DTC) for Machine Learning due to better Accuracy and better Recall

Model	Pros	Cons
Logistic Regression	Fast Precise L2 Recall, Decent L2 Precision Acc = 67%	Heavily influenced by large L2 population so other accident level metrics = ~0
Overfitting	Best L1 & L4 Recall & L2 Precision	Acc = 41% 10-20% L2 & L3 Recall Model pushes predictions to the boundaries
Decision Tree Classifier	Acc = 81% (highest) Strong Precision & Recall for L2 & L3	Heavily influenced by large L2 & L3 population
Neural Network	Acc = 75% Best Precision/Recall for L2 Satisfactory P/R for L3 (57%/63%)	Lower P/R performance for L4 that DTC (29%/47%)

DTC Confusion Matrix

	Predicted 1	Predicted 2	Predicted 3	Predicted 4
Actual 1	1	166	31	0
Acutal 2	162	364399	50038	6070
Actual 3	27	48356	131901	3729
Actual 4	3	5742	3752	9184

Accuracy Score : 0.810642
Classification Report

	precision	recall	f1-score	support
1	0.01	0.01	0.01	198
2	0.87	0.87	0.87	420669
3	0.71	0.72	0.71	184013
4	0.48	0.49	0.49	18681
accuracy			0.81	623561
macro avg	0.52	0.52	0.52	623561
weighted avg	0.81	0.81	0.81	623561