Car Accident Severity Prediction

Ravinder Kumar

Business Problem

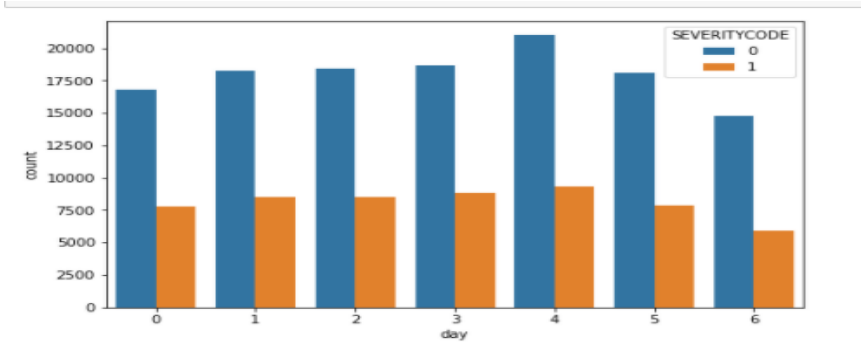
- Proactive approach to increase road safety. This can help many departments such as healthcare authorities, road transport authorities.
- Government can take proactive approach and provide ambulance near the sites where severity is very high. This will help in saving lives and reducing healthcare cost involved.
- Road transport authorities can use this model and improve the road conditions, give warnings etc to prevent accidents.
- Government can use this model to make better traffic management systems

Data

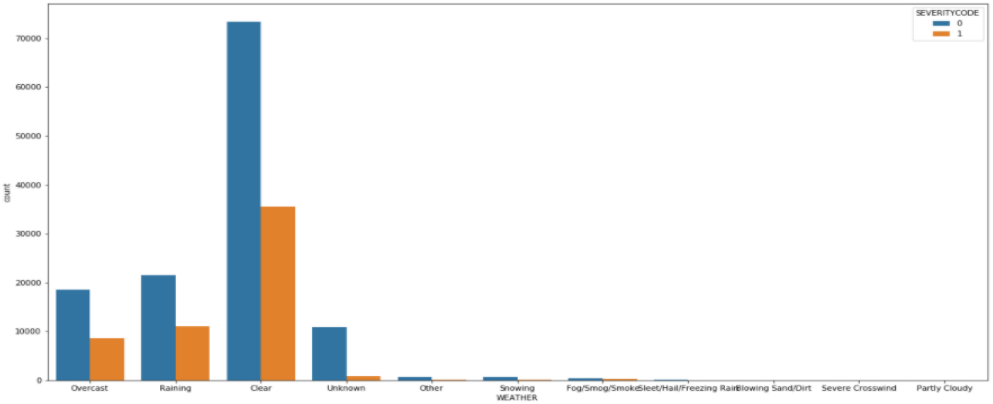
- Dataset used is provided by Coursera.
- Includes around 35 features, 194673 observation.
- Important categorical features like weather, road, light conditions, collision type, junction, address type available.
- Important numerical features like count of person involved, cars involved available.
- Data cleaning performed, such as removing unknown's, nulls etc.
- Redundant features removed.
- One hot encoding for categorical variables.

Feature Distribution

- Severity – Day of Incident

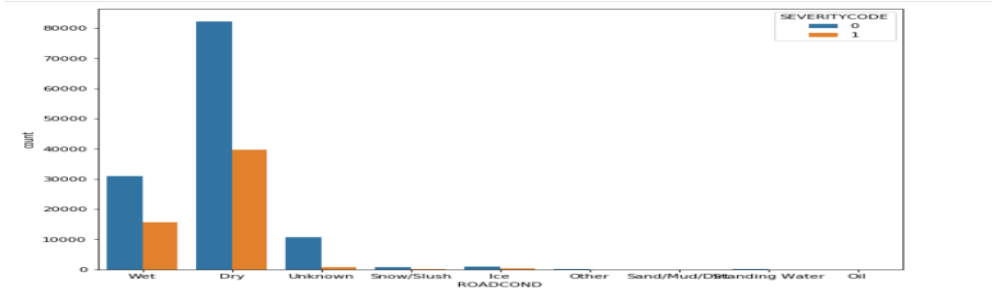


- Severity – weather

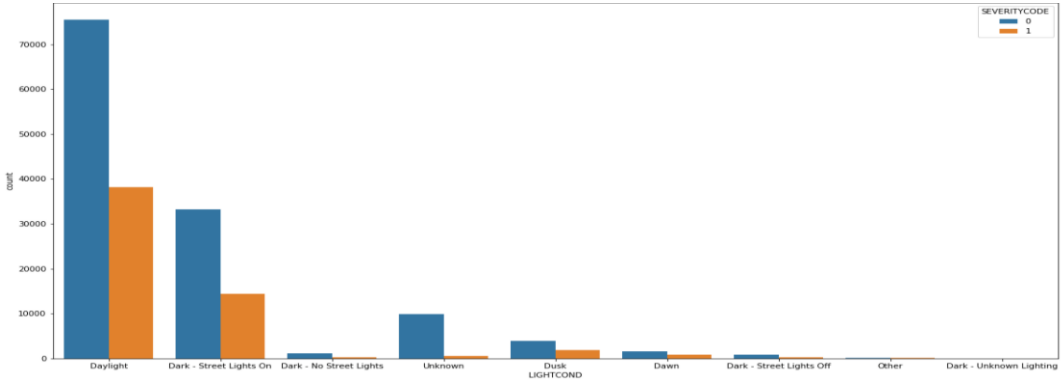


Feature Distribution…

- Severity – Road conditions:

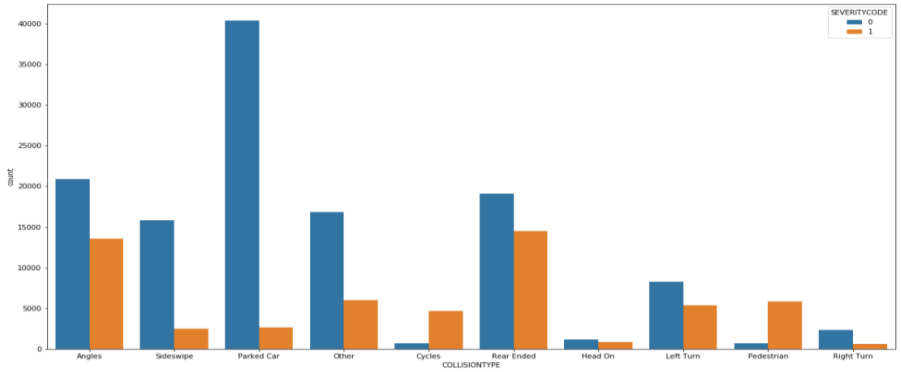


- Severity – Light Conditions:

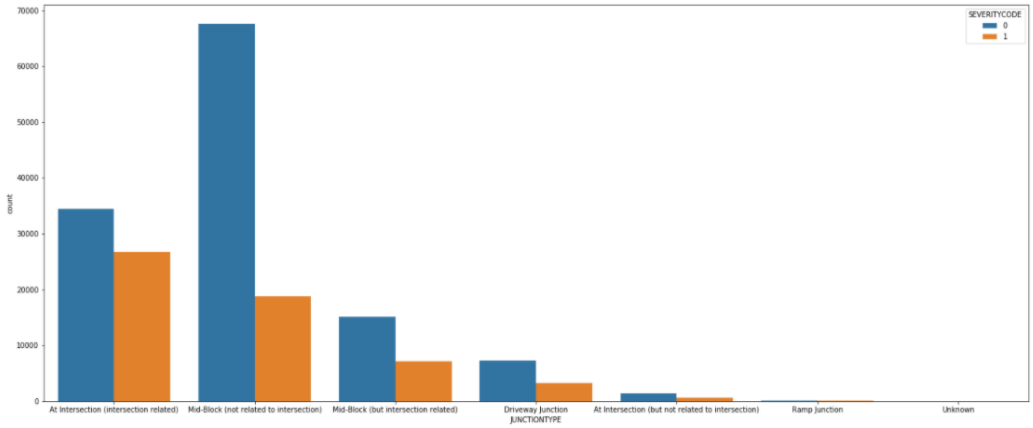


Feature Distribution…

- Severity – collision type :

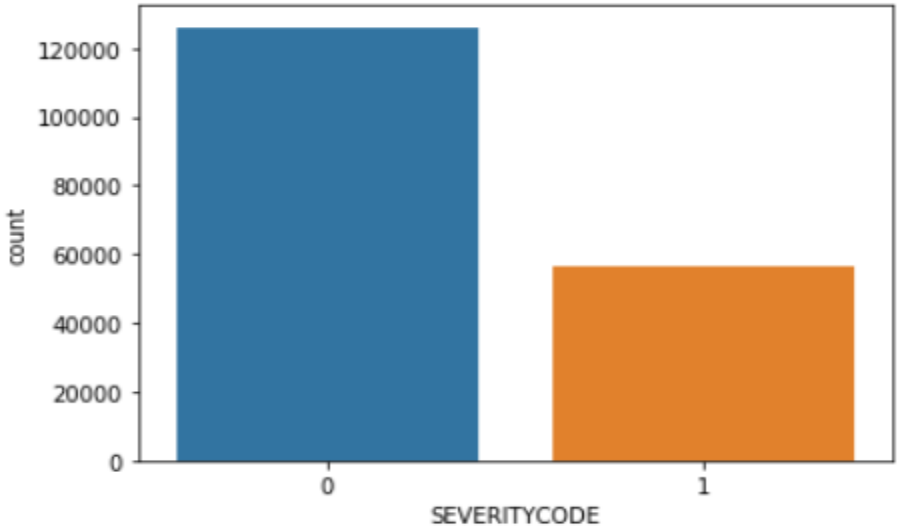


- Severity – Junction Type:



Predictive Modeling

- Label Distribution slightly imbalanced
- 0 – non serious, 1-serious involving injury



Classification without weights

- Matrix for accuracy versus algorithms:

algorithm	accuracy	f1_score	log loss
decisionTree	75%	45%	
KNN	74%	47%	
Logisitic Regression	75%	44%	49%

- **Issue ? F1_SCORE very less (due to imbalance)**

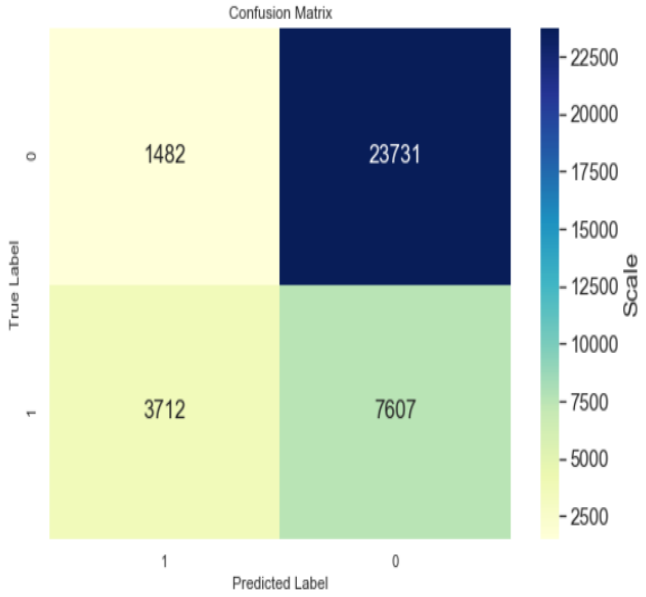
After adding weights - XGBoost

- Accuracy decreased, but f1_score increases.
- Goal is to predict more severe accidents accurately.

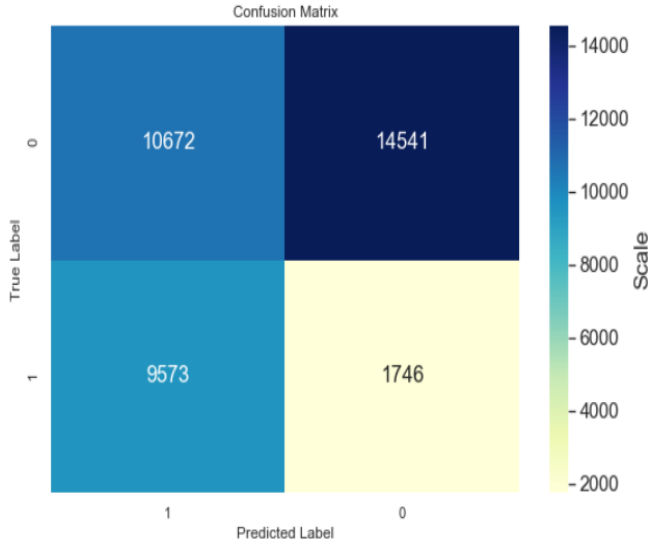
	precision	recall	f1-score	support
0	0.89	0.58	0.70	25213
1	0.47	0.85	0.61	11319
accuracy			0.66	36532
macro avg	0.68	0.71	0.65	36532
weighted avg	0.76	0.66	0.67	36532

With Weights versus Without Weights

- Without weights



With weights



Conclusion and Future Steps

- We can achieve good accuracy, but recall will be very less for severe accidents which is not the goal.
- XGBoost with hyper parameter tuning and weights has increased f1 score drastically from 45% to 61%.
- For future directions, we need more data, better data as majority values are “UNKNOWN”.
- Research on Logistic regression with weights.
- Research on Location feature, how can we do one hot encoding for so many values without slowing down the performance.