# PREDICTING CAR ACCIDENT SEVERITY

IBM DATA SCIENCE CAPSTONE

---

George Abou Assi
https://github.com/Georgeab9

# CASE UNDERSTANDING

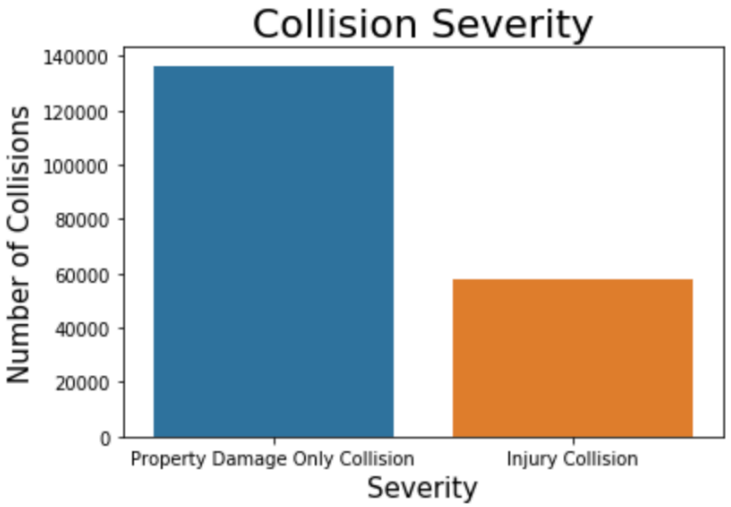
Car accidents occur all over the world and cause multiple injuries and death

- Estimated more than 1 million deaths
- Most death involves young adults

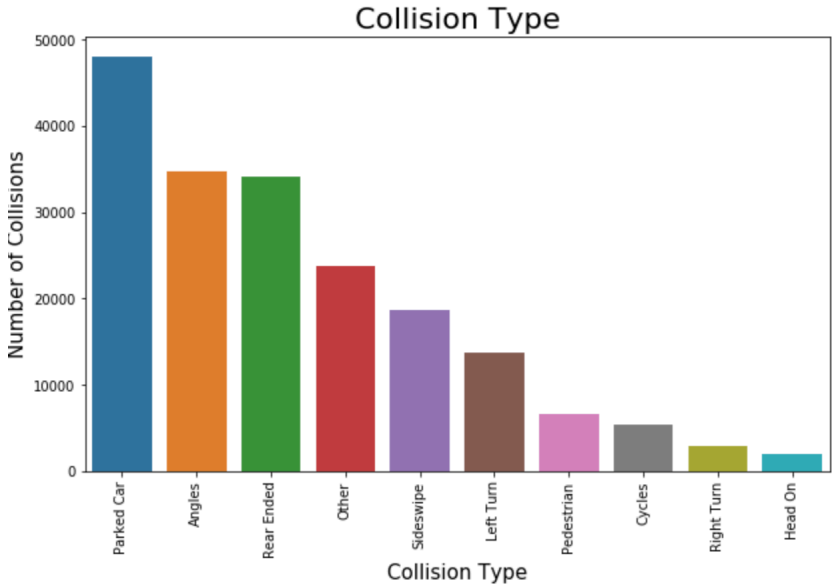
Predicting accident severity may be applied in a mobile application that sends notifications to drivers, policemen, resuce teams, ambulances, car insurance companies, civil defence... to mobilize help immediately, protect people and themselves, avoid further impact in the car crash and many more benefits.

# DATA

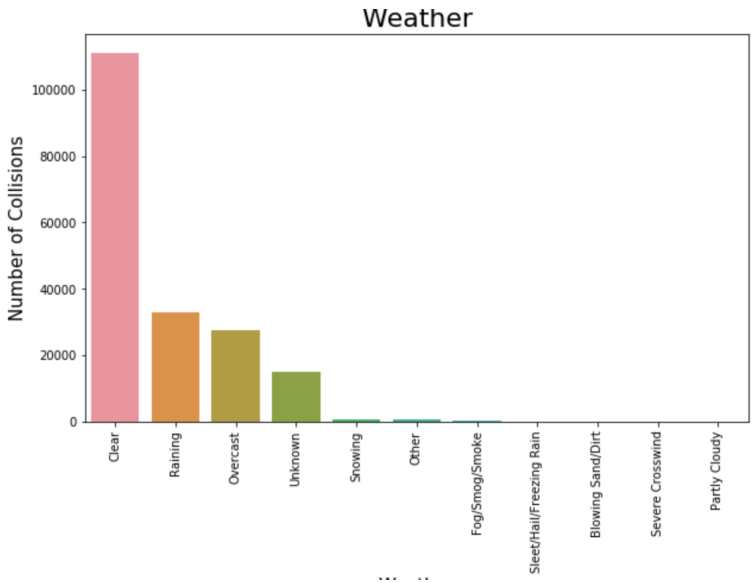
- Data was extracted from Seattle Government ( Police Department)' https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv'
- Data includes 194673 observations and 38 features
- Dataset involved several missing values and wrong formats that required cleaning
- Target variable contained inbalanced data which required undersampling / balancing



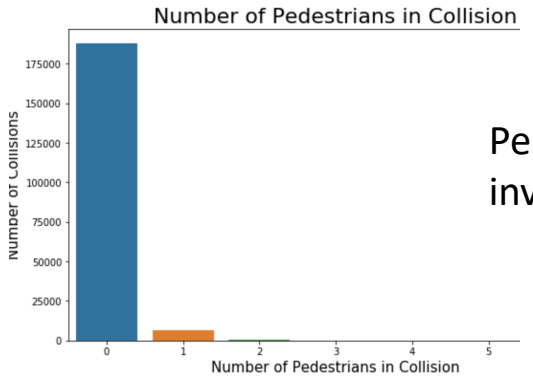

Most accidents involve property rather than injuries



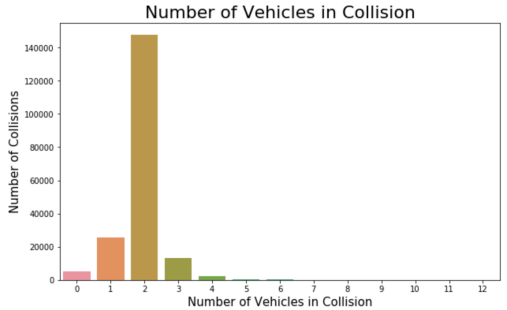

Parked cars are the biggest reason of car accidents



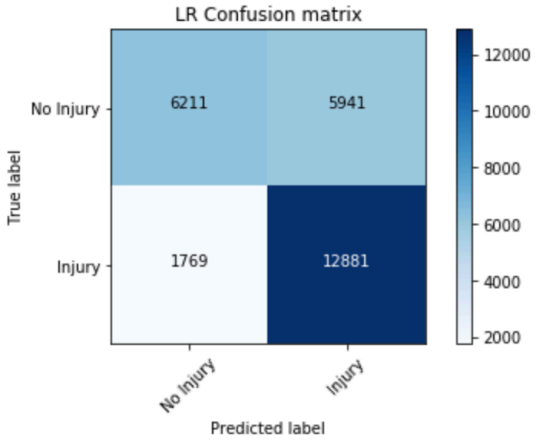

Weather does not seem to impact car crashes



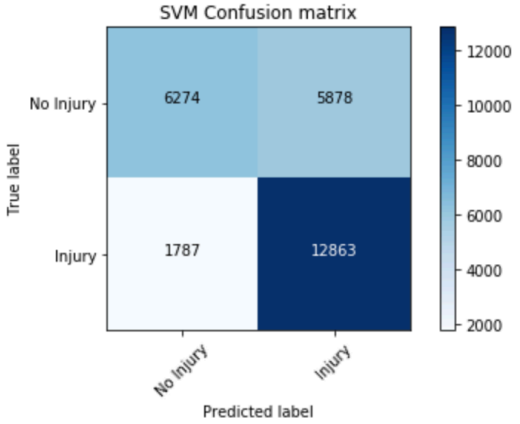

Pedestrians are rarely involved in car crashes



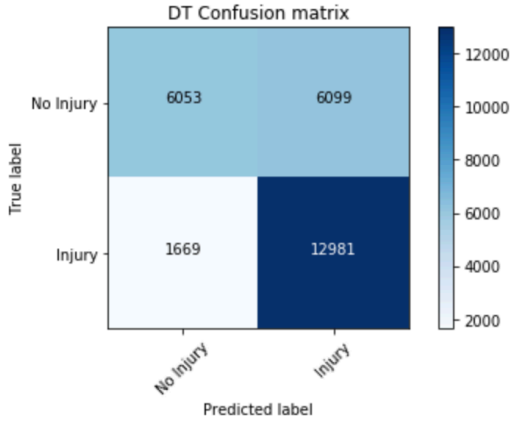

Most of the time, 2 vehicles are involved in accidents



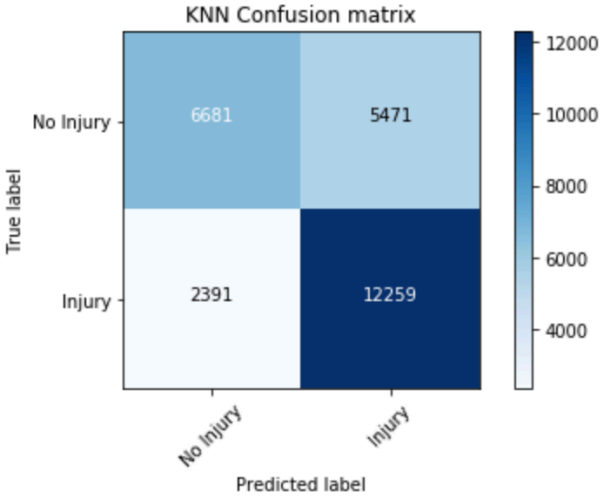

Logistic Regression



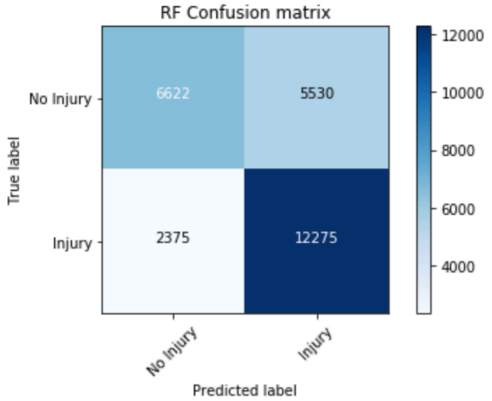

SVM




Decision Tree




KNN




Random Forest

SVM is the best classifier

| | Decision Tree | K-Nearest Neighbors | Logistic Regression | Random Forest | Support Vector Machine |
|---|---|---|---|---|---|
| **Jaccard Score** | 0.7102 | 0.7067 | 0.7123 | 0.7051 | 0.7140 |
| **F1 Score** | 0.6969 | 0.6993 | 0.7005 | 0.6974 | 0.7026 |
| **Injury Class F1 Score** | 0.7697 | 0.7572 | 0.7697 | 0.7564 | 0.7704 |
| **Injury Class Recall Score** | 0.8861 | 0.8368 | 0.8792 | 0.8379 | 0.8780 |

Parked Car and Number of Vehicles are the most important features

| | Feature | Importance Score |
|---|---|---|
| **0** | Parked Car | 0.39 |
| **1** | VEHCOUNT | 0.28 |
| **2** | Sideswipe | 0.13 |
| **3** | PERSONCOUNT | 0.11 |
| **4** | UNDERINFL | 0.02 |
| **5** | PEDCOUNT | 0.01 |

# CONCLUSION

After cleaning and analyzing the data set, the number of vehicles and if a car is parked are the most important features for the accident severity predicition.
This shows that for the application to be most effective, an immediate report must be done by the police (or volunteers in the area) and communicated in the app.

With the advance of sensor technology, more data can be extracted and used to build more precise models.