



# CAPSTONE PROJECT

## CAR ACCIDENT SEVERITY MODELLING

COURSERA – IBM DATA SCIENCE COURSE

# INTRODUCTION OF BUSINESS PROBLEM

The ultimate purpose of this project is to prevent avoidable car accidents by alerting drivers and relevant public functions with forecasted severity of car accidents. The estimation can be used as a good reference to remind people to be more careful in critical situations.

Some car accidents are caused by lacking of attention during driving, abusing drugs and alcohol or over-speed driving. Majority of these accidents can be prevented by setting harsher regulations and implementing properly. However, there are also other uncontrollable factors like weather, visibility, road conditions significantly increase the probability of car accidents. Therefore revealing the underlying pattern in historical data and sending timely warnings to the drivers and public functions would be helpful in preventing avoidable car accidents and better allocating of rescue efforts.

The project should benefit individual drivers, local government, police, rescue groups, and car insurance institutes as well. The model and its results are going to provide some advice for these target audience to make insightful decisions for reducing the number of accidents and injuries.



# DESCRIPTION OF DATA

The data, collected since 2004, consists of 37 independent variables and 194,673 rows. The dependent variable, “SEVERITYCODE”, contains numbers corresponding to different levels of severity caused by an accident from 0 to 4.

Severity codes are as follows:

- 0: Little to no Probability (Clear Conditions)**
- 1: Very Low Probability — Chance or Property Damage**
- 2: Low Probability — Chance of Injury**
- 3: Mild Probability — Chance of Serious Injury**
- 4: High Probability — Chance of Fatality**





# METHODOLOGY



**K Nearest  
Neighbour  
(KNN)**

**Logistic Regression**

**Decision Tree**

# RESULTS AND CONCLUSION

Below table is the summary of three models. The best one is Logistic Regression model. Of course there are plenty of other modelling methods we can apply, and the choice of different features will result in different results. Therefore there are many ways to improve the model.

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.56	0.47	NA
Decision Tree	0.56	0.47	NA
LogisticRegression	0.55	0.55	0.67

We tested three machine learning models designed to predict car accident severity based on given dataset. We focus on three factors - weather, road, and light conditions. Based on the results, we can conclude that particular conditions have clear impact on the chance getting injury. Proper alerts would possibly reduce the possibility and severity of car accidents.



The background is a gradient of purple and blue, filled with numerous out-of-focus circular light spots (bokeh) in various shades. Overlaid on this are several faint, white geometric patterns. On the left side, there is a large circular arc with a degree scale ranging from 140 to 260. Other smaller circular patterns, some with arrows indicating direction, are scattered across the left and center. The overall aesthetic is modern and technical.

# THANK YOU!