Predicting Traffic Accident Severity

Applied Data Science Capstone

Problem of Car Accidents

The Seattle Department of Transportation (SODT) objective to improve new model aims to predict the severity of car accidents and main variables that strongly related to it, in order to help decisions based to make campaigns or improvements to decrease severe accidents based on certain conditions. And detect intersection have car accidents with hight servity .

Data

- All types of collisions.
- Collisions will display at the intersection or mid-block of a segment.
- Timeframe: 2004 to Present.
- Date set not balanced for severity

Features Selected

- The target feature Severity is binary
  - 1 :Damage .
  - 2 :Injury.

- Independent variables are
  - A description of the **weather** conditions during the time of the collision.
  - The condition of the **road** during the collision.
  - The **light** conditions during the collision.

Classification Models

Data used

- Decode data
- Balance data
- Splitting data for traning and testing 0.3

Models

- K-Nearest Neighbor (KNN)

  find the best
- Decision Tree
  - Find the optimal
- Logistic Regression
  - C = 0.001

Results

- The table below report the evaluation on test data

Algorithm	Jaccard	F1-score	Logloss
KNN	0.56	0.54	NA
Decision Tree	0.56	0.49	NA
SVM	0.53	0.50	NA
Logistic Regression	0.53	0.51	0.68

Conclusion and Future work

- Predicate severity of car accident by models help to reduce it.
- Accuracy of the models need imporvments.
- I think we can add bew features like vehicle speed and the driver drink alcohol or drugs