Predict the Severity of Car Accident

Introduction

Benefits

➔ Make drivers to drive more carefully or warn them to change their travel method.
Reduce the economic and societal impact of car accidents

Objective

➔ Recognize key factors affecting the accident severity.
Develop a model to predict accident severity based on the sophisticated traffic accident dataset.

Data acquisition and cleaning

- ➔ “Road Safety Data – Accidents 2018” downloaded from Open Data Platform UK.
- ➔ 32 columns and 122635 rows in raw dataset

- ➔ Missing value are deleted
- ➔ Unrelated variables are deleted
- ➔ Convert categorical variables to binary variables
- ➔ Data standard scaled
- ➔ Cleaned data contains 67037 rows and 14 variables
- ➔ Unbalanced data

Exploratory data analysis

- ➔ There is no big difference in the severity of accident when the weather conditions are different
- ➔ “Serious” accidents are more likely to happen when the light condition is not good.
- ➔ There is no big difference in the severity of accident when the accident happens at a junction or not.

Accident_Severity	Weather_Conditions	
serious	fine	0.859769
	not_fine	0.140231
slight	fine	0.845962
	not_fine	0.154038

Accident_Severity	Light_Conditions	
serious	daylight	0.699421
	not_daylight	0.300579
slight	daylight	0.763166
	not_daylight	0.236834

Accident_Severity	Junction_Detail	
serious	at_junctino	0.505073
	not_at_junction	0.494927
slight	at_junctino	0.516705
	not_at_junction	0.483295

Predictive modeling

- Dataset spilting
  - 70% training dataset
  - 30% testing dataset

- 5 classification models are used
  - Logistic regression
  - K-Nearest Neighbors
  - Decision Tree
  - Random Forest
  - Support Vector Machine

Model performance

- ➔ Accuracy
  - ◆ 64.45 to 67.80 between 5 models
- ➔ F1 - Score
  - ◆ 57.50 to 62.13 between 5 models
- ➔ SVM performed best

	Model	Accuracy	F1 - Score	Jaccard - Score
0	Support Vector Machines	67.80	62.13	16.27
2	Logistic Regression	67.01	57.50	7.23
4	Decision Tree	66.48	61.71	17.06
3	Random Forest	66.46	62.11	18.12
1	KNN	64.45	61.28	19.13

Conclusion and future decision

- ➔ Built useful models to predict whether a car accident would be serious or slight
- ➔ Accuracy of the models has room to improve
- ➔ Ideas
  - ◆ Make the dataset balanced
  - ◆ Get more variables and data, such as drivers' background information and vehicle conditions


Thank You

Reference

Road Safety Data – Accidents 2018
Retrieved from https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data