CAPSTONE PROJECT – SEVERITY PREDICTION

Sep 24, 2020

INTRODUCTION

- It is important to make driving experience as safe as possible
- Wet road, bad weather, and other conditions may affect driving
- Explore car accidents data to find potential reasons for accidents
- Give suggestions to avoid danger.

DATA

- Downloaded from link
- Contains 194674 car accident cases with description of location, time, weather condition, road condition, light condition, other features and severity.
- Six features are selected as independent variable, they are: 'LIGHTCOND', 'ROADCOND','WEATHER', 'JUNCTIONTYPE', 'COLLISIONTYPE' and 'ADDRTYPE'.
- The dependent variable is "SEVERITYCODE" with two values "1" and "2".

ENCODING

- An encoder employed to transform data from string to numbers.
- After encoding, all strings are converted to calculatable integer format.

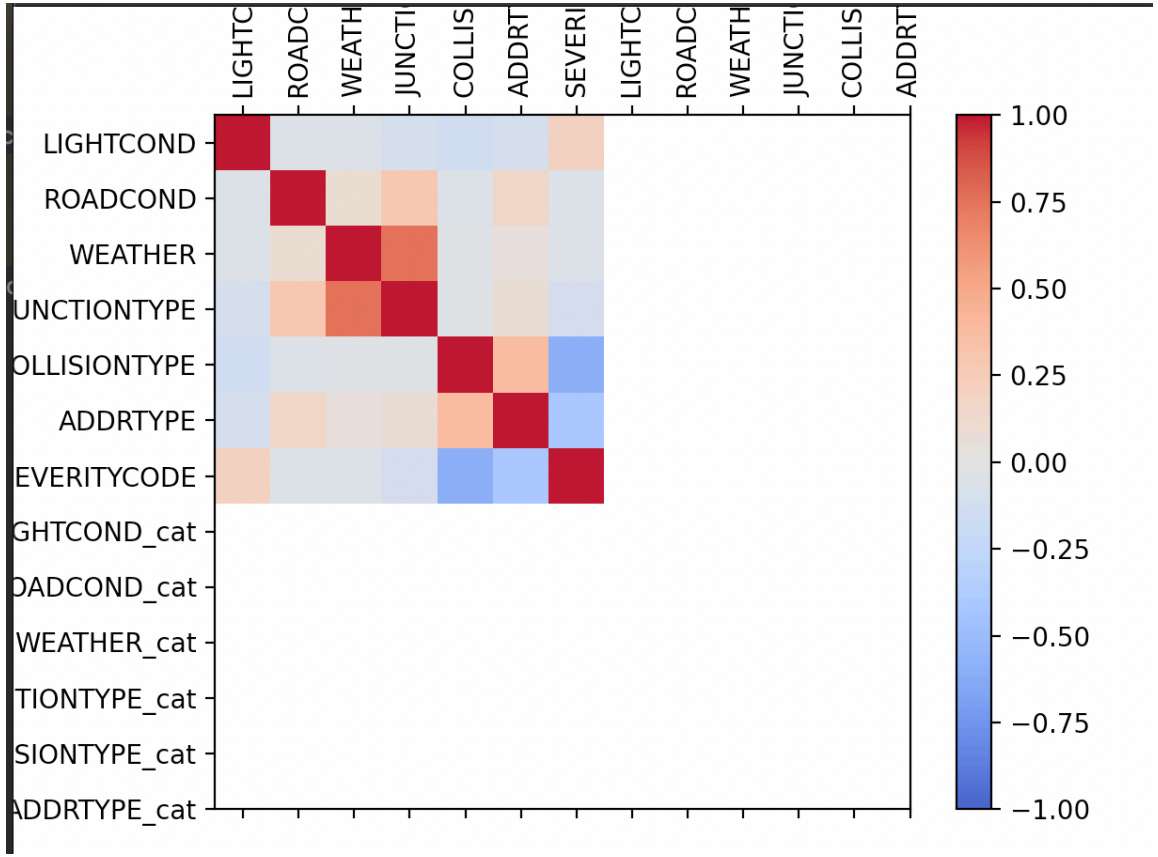
SPLIT TRAINING SET AND TEST SET

- Training set takes 80% of the whole set,
- Test set takes 20% of the whole set.

ALGORITHM

- Choose one from Logistic Regression, Naïve Bayes and Support Vector machine, are employed.
- The performance are compared, and the best one of algorithms chosen by the performance.

CORRELATION ANALYSIS



LOGISTIC REGRESSION

- Confusion matrix

	Actually positive	Actually negative
Predicted positive	27030	395
Predicted negative	11317	193

- F1 score: 0.588
- Accuracy rate: 0.699

NAIVE BAYES

- Confusion matrix

	Actually positive	Actually negative
Predicted positive	21461	5964
Predicted negative	7006	4504

- F1 score: 0.662
- Accuracy rate: 0.667

SUPPORT VECTOR MACHINE

- Confusion matrix

	Actually positive	Actually negative
Predicted positive	26463	962
Predicted negative	9948	1562

- F1 score: 0.650
- Accuracy rate: 0.720

DISCUSSION

- Support Vector Machine has the best performance to predict. It achieves 0.650 f1 score, and 0.720 accuracy rate.

- Light condition, road condition, weather condition have potentiality to make driving more dangerous.