

UK Traffic Accidents Analysis

DATS 6202 Machine Learning I
Final Project – Group 1

Mengjie Zhang, Xinning Wang, Yilin Wang

INTRODUCTION

- Background:
 - Increasingly automobile traffic volume
 - Annual average daily traffic (AADT): the total volume of vehicle traffic of a road for one year, or by date.
- Our goal:
 - Predicting the severity of accident based on AADT data.

DATASET

- Two datasets were all from Kaggle:
 - 1.6 million UK traffic accidents:
 - 400,000 observations
 - Over 30 features (location, accident severity, weather conditions and date, etc.)
 - <https://www.kaggle.com/yesterdog/eda-of-1-6-mil-traffic-accidents-in-london/data>
 - Predicting car count:
 - 10,000 observations
 - Over 25 variables
 - <https://www.kaggle.com/coolcoder001/predicting-car-count/data>

DATASET

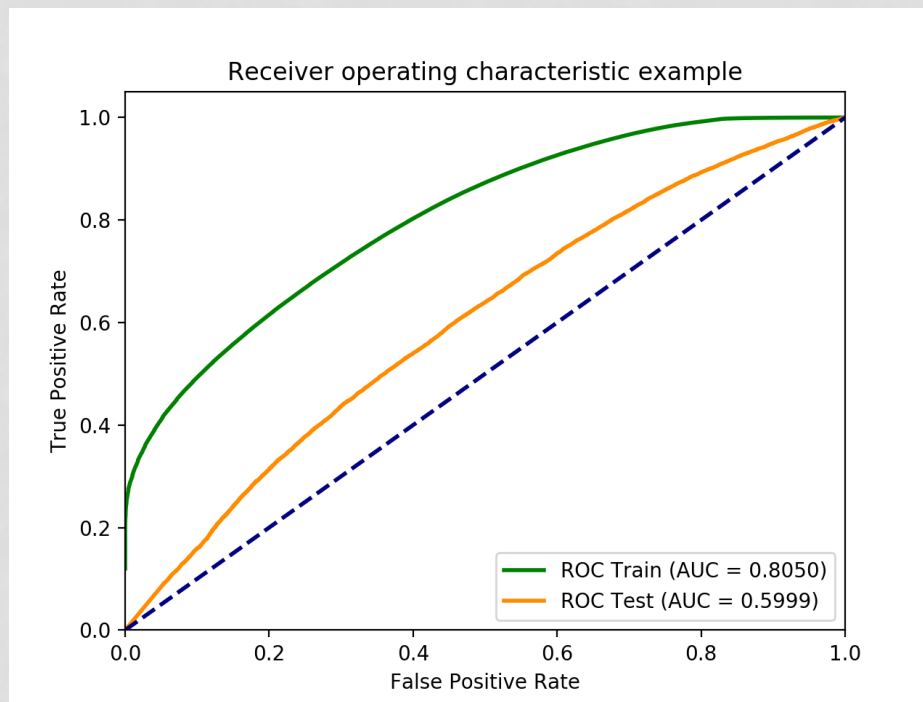
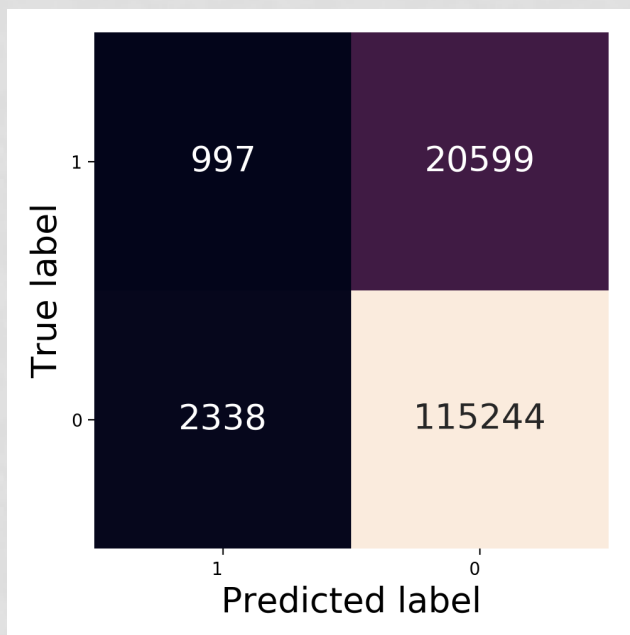
- Target:
 - label (Accident Severity): binary
- Predictors:
 - Day of Week, Road Type, Speed limit, Pedestrian Crossing Human Control, Pedestrian Crossing Physical Facilities, Light Conditions, Weather Conditions, Road Surface Conditions, Special Conditions at Site, Carriageway Hazards, Urban or Rural Area, Did Police Officer Attend Scene of Accident, Hour, Peak

METHODS

- Using Python to conduct:
 - Data preprocessing
 - Machine learning Network:
 - Random Forest
 - Neural Network
 - Naïve Bayes
 - Support Vector Machine (SVM)

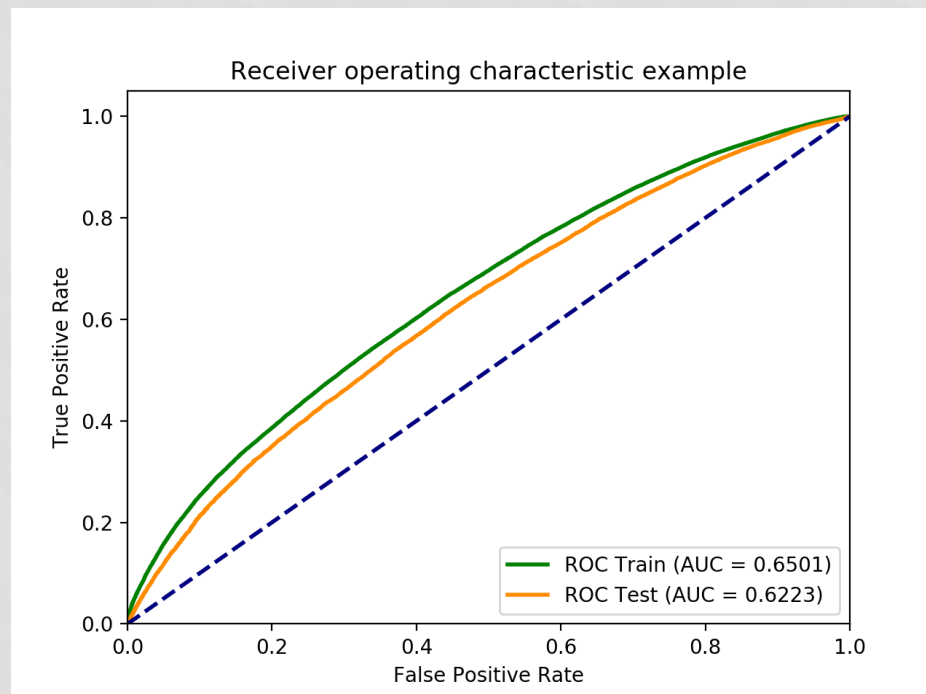
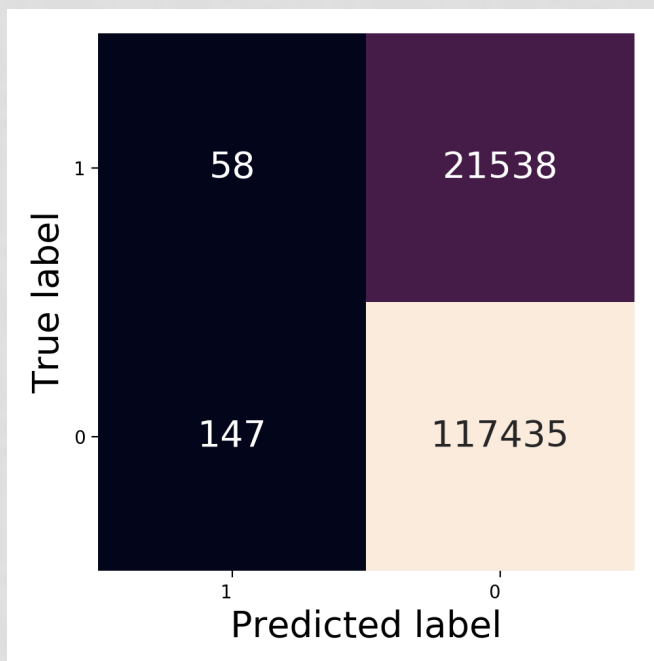
RESULTS & ANALYSIS

- Random Forest



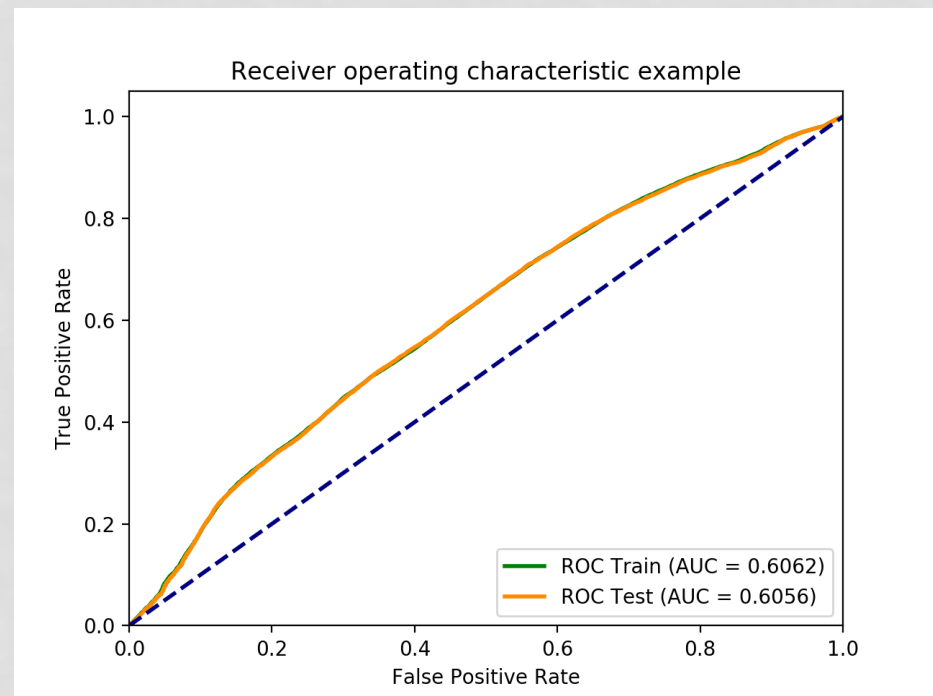
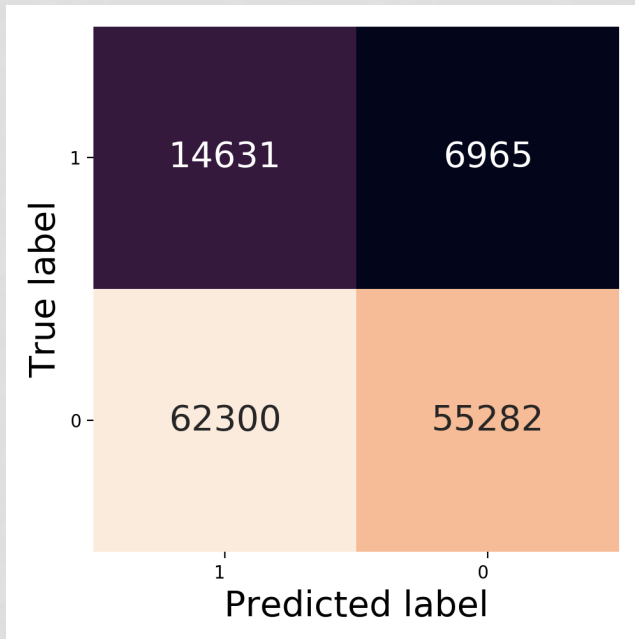
RESULTS & ANALYSIS

- Neural Network



RESULTS & ANALYSIS

- Naïve Bayes



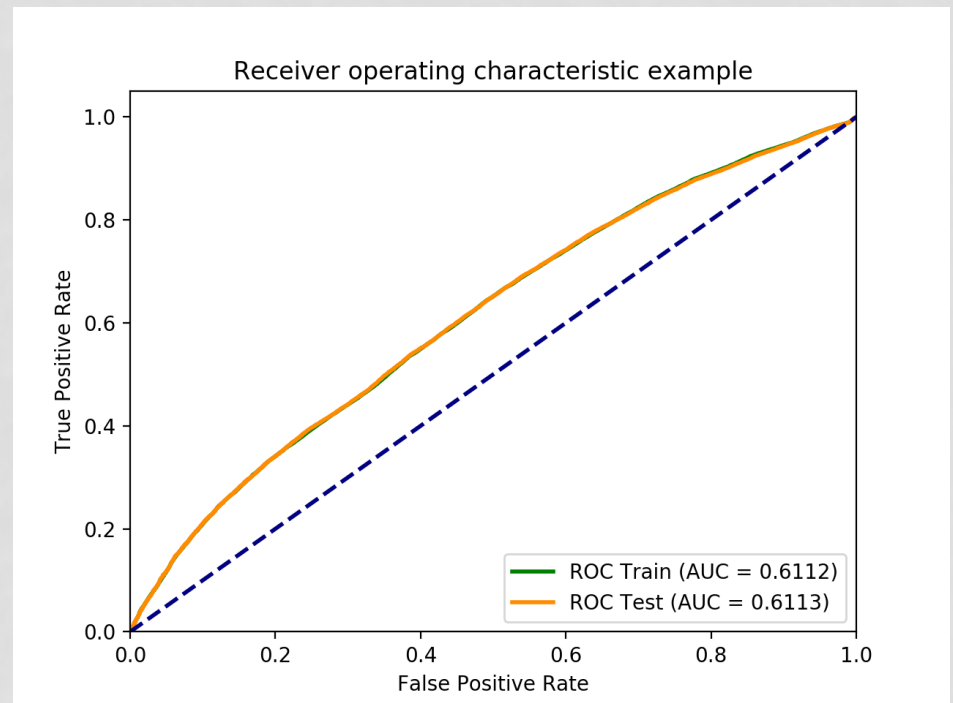
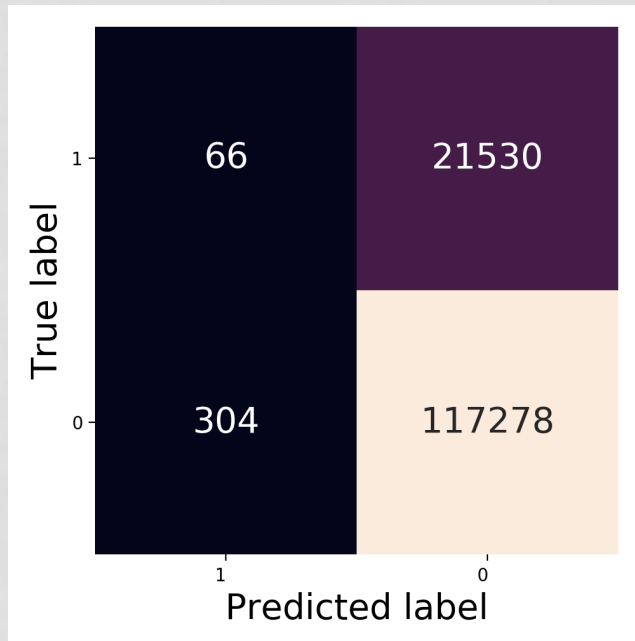
RESULTS & ANALYSIS

- SVM
 - The classification report for linear SVC

Class	Precision	Recall	F1-score	support
0	0.25	0.01	0.01	21596
1	0.85	1.00	0.91	117582
Avg/total	0.75	0.84	0.77	13978

RESULTS & ANALYSIS

- SVM



SUMMARY & CONCLUSION

- Performance metrics for all machine learning methods

Method	AUC	Accuracy	Precision	Recall	F1-score
Random forest	0.60	0.84	0.76	0.84	0.78
Neural network	0.62	0.85	0.76	0.84	0.77
Naïve Bayes	0.60	0.51	0.78	0.51	0.57
Support vector machine	0.62	0.84	0.75	0.84	0.77

REFERENCE

- Hoang Nguyen , Chen Cai, Fang Chen(2017). Automatic classification of traffic incidents' severity using machine learning approaches. *IET Intell. Transp. Syst.*, Vol. 11 Iss. 10, pp. 615-623.
- UK Department of transport
<https://www.dft.gov.uk/traffic-counts/index.php>
- Kaggle
<https://www.kaggle.com/daveianhickey/2000-16-traffic-flow-england-scotland-wales>
- Neural Network Design (2nd Ed), by Martin T Hagan

THE END

Thanks!