

차량 주행 데이터 기반 도난 탐지

- 정보보호 R&D 데이터 챌린지 2018

전인태, 김민정

2018.11.30

Approach - Driver Behavior Analysis

Driver

1. 운전자의 습관이 가장 잘 나타나는 도로는 어디인가
2. 운전자의 습관을 나타내는 주요한 Feature는 무엇인가

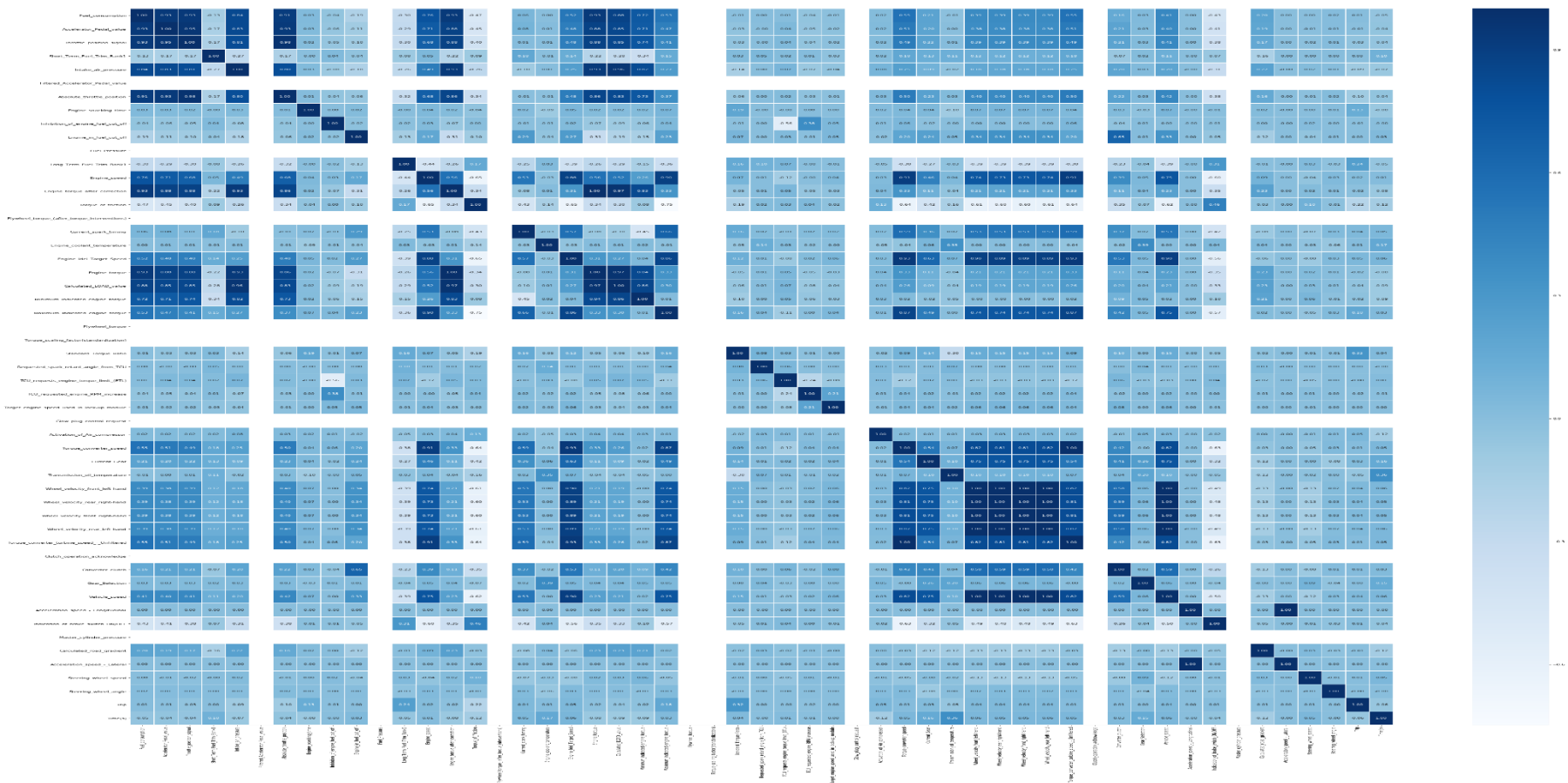
Data

1. 데이터는 충분한가 – Deep / Machine
2. 데이터의 아웃라이어는 없는가
3. 아웃라이어가 있다면 제거해줘야 할 것인가

Approach - Driver Behavior Analysis

- **Feature Selection**
 1. Driver 별 Feature 분석
 2. OBD data 별 Hitmap 분석
- **Feature Processing**
 1. Robust Scaling
 2. Sliding
- **Machine Learning**
 1. ML Selection
 2. Hyperparameter Tuning

Feature Selection - EDA



[illegible]

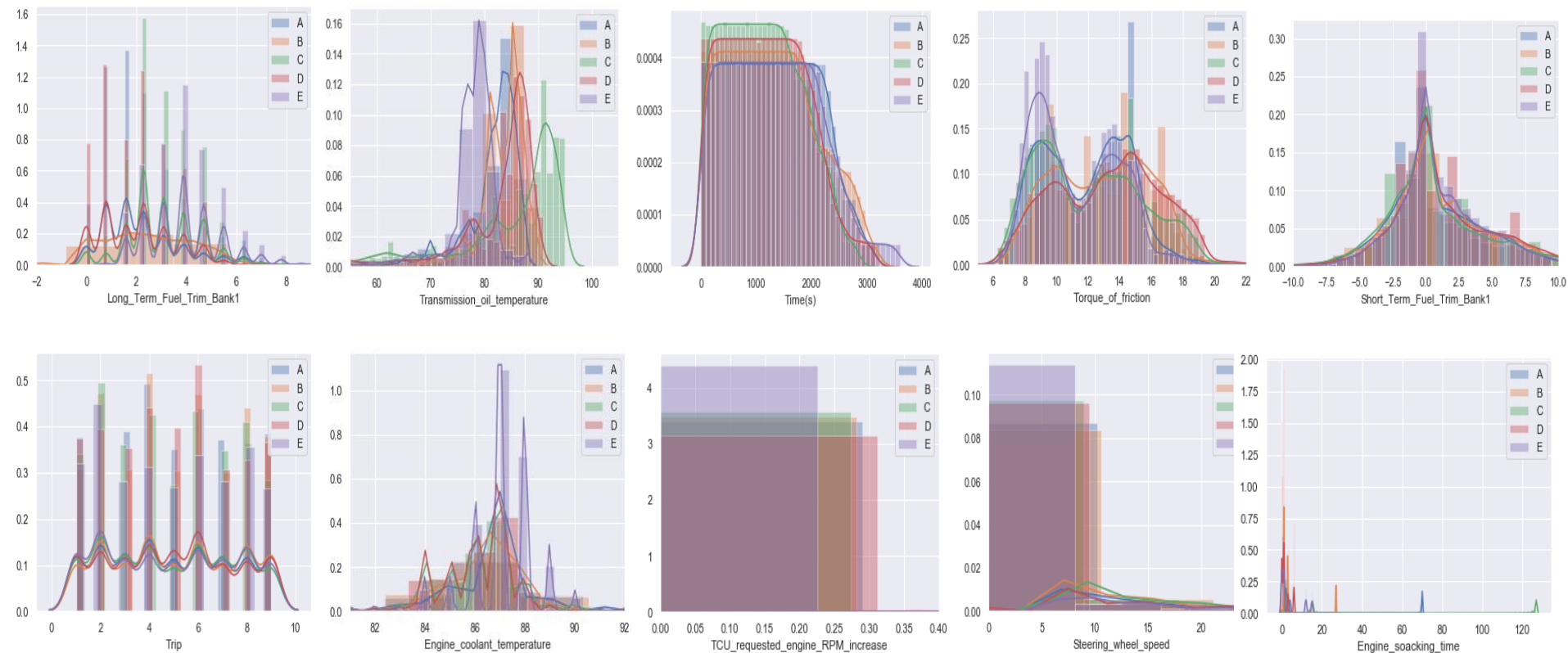
[illegible]

0인 Feature 제거

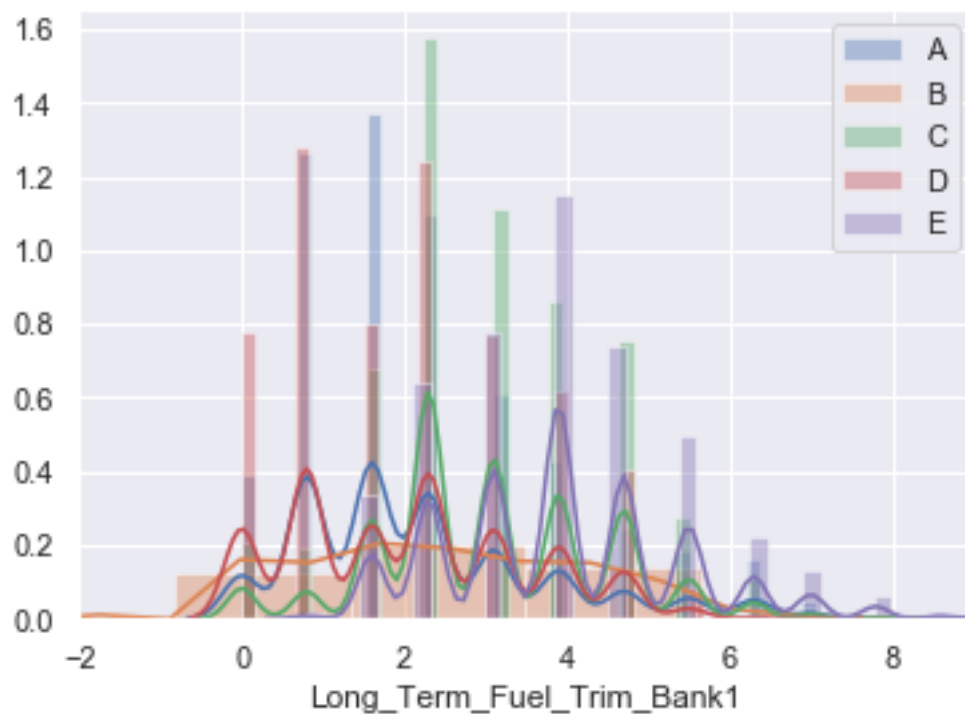
상관 관계 높은 Feature 제거

[illegible]

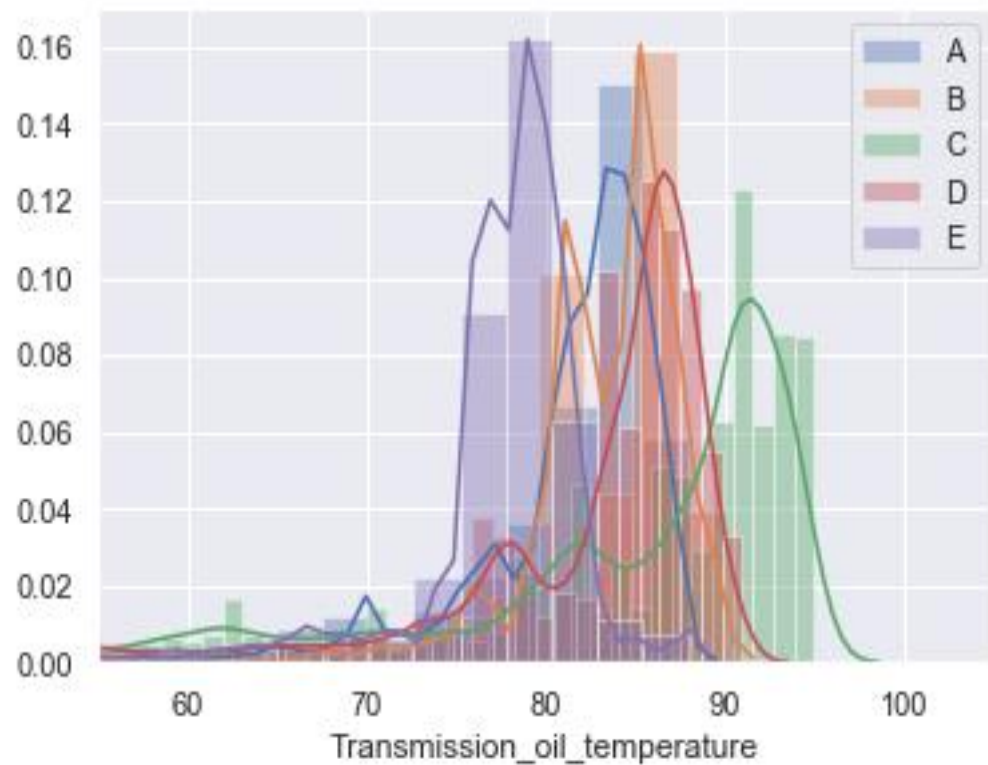
Feature Selection - EDA



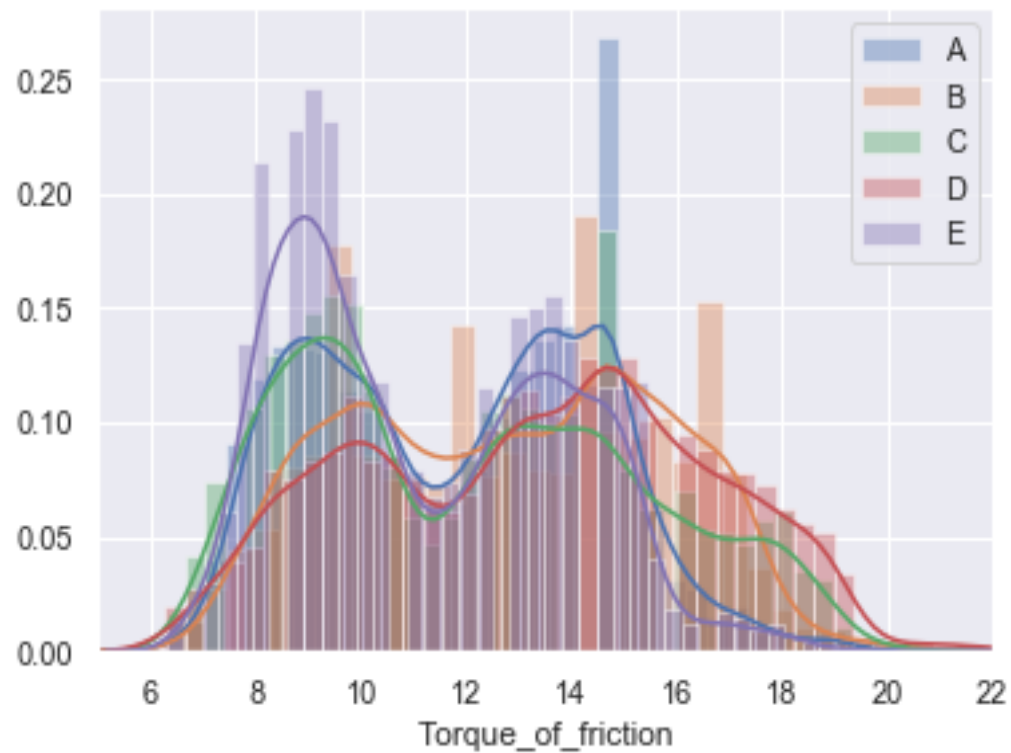
Feature Selection - EDA



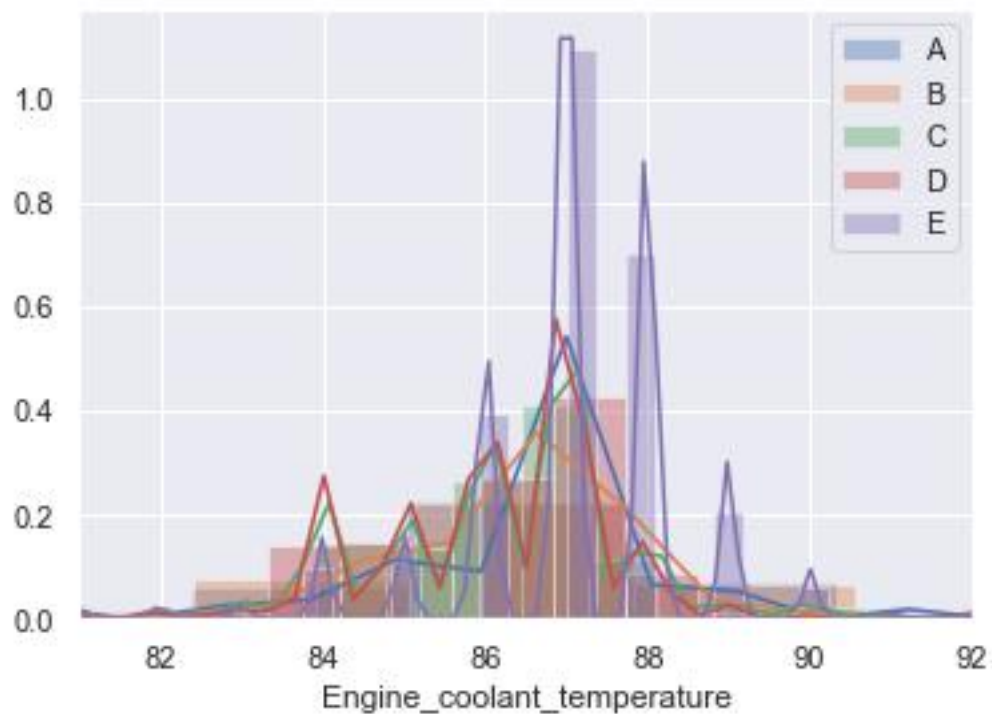
Feature Selection - EDA



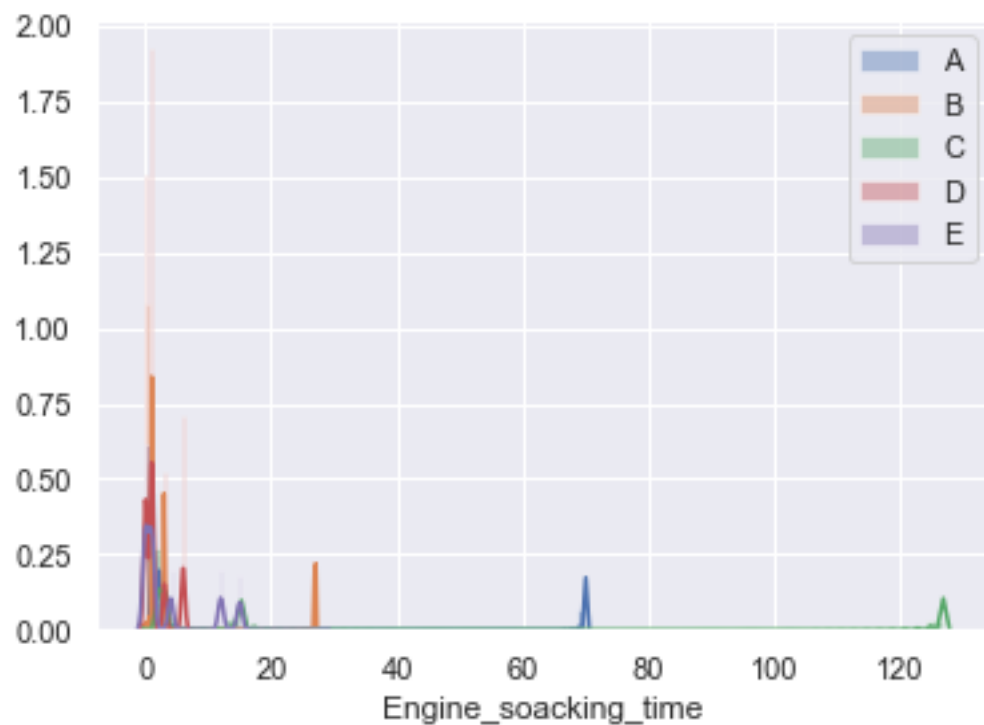
Feature Selection - EDA



Feature Selection - EDA



Feature Selection - EDA



Feature Selection – Feature Importance

	features_name	importance
11	Long_Term_Fuel_Trim_Bank1	0.063489
34	Transmission_oil_temperature	0.061769
52	Time(s)	0.057834
14	Torque_of_friction	0.053234
3	Short_Term_Fuel_Trim_Bank1	0.046676
51	Trip	0.045398
17	Engine_coolant_temperature	0.040475
28	TCU_requested_engine_RPM_increase	0.039640
49	Steering_wheel_speed	0.038806
7	Engine_soaking_time	0.038516

Feature Processing

Window Sliding – 30, 60, 90, 120



Selection : 60

Robust Scaling – 아웃라이어 영향 최소화

Machine Learning – ML Selection

Train dataset

- Random Forest
 - 0.9999692004435136
- XGBoost
 - 0.9999384008870272

Selection : Random Forest

Grid Search CV

Hyperparameter Tuning

Learning_rate

max_delta_step

Subsample

colsample_bytree

colsample_bylevel

reg_alpha

reg_lambda

Machine Learning – ML Selection

Train dataset

- Random Forest
 - 0.9999692004435136
- XGBoost
 - 1

Selection : XGBoost

THANK YOU