

# 10

# 선박 엔진 데이터를 기반으로 하는 기계 이상 진단 시스템 개발

소속 정보컴퓨터공학부

분과 A

팀명 읍저버

참여학생 강금석, 김민수

지도교수 송길태

## 개요 및 목표

### 과제 개요

- ✓ 선박 엔진의 데이터를 사용해 선박 엔진의 부하를 분석하여 선박 엔진의 이상을 진단하는데 도움이 되는 모델을 개발하는 프로젝트
- ✓ 기존에 사용하는 Feature 외에 엔진 부하와 더 높은 상관관계를 갖는 Feature를 찾고, 선택해 더 높은 성능을 갖는 모델을 개발하는 방식으로 접근

### 과제 목표

- ✓ 기존에 사용하는 Feature 외에 엔진 부하와 더 높은 상관관계를 갖는 Feature를 찾고, 선택하기
- ✓ 더 높은 성능을 갖는 모델을 개발하기
- ✓ 데이터 수집, 가공, 시각화, 학습, 평가까지의 딥러닝 전체 프로세스 경험해보기

## 과제 내용

### 데이터 셋 준비

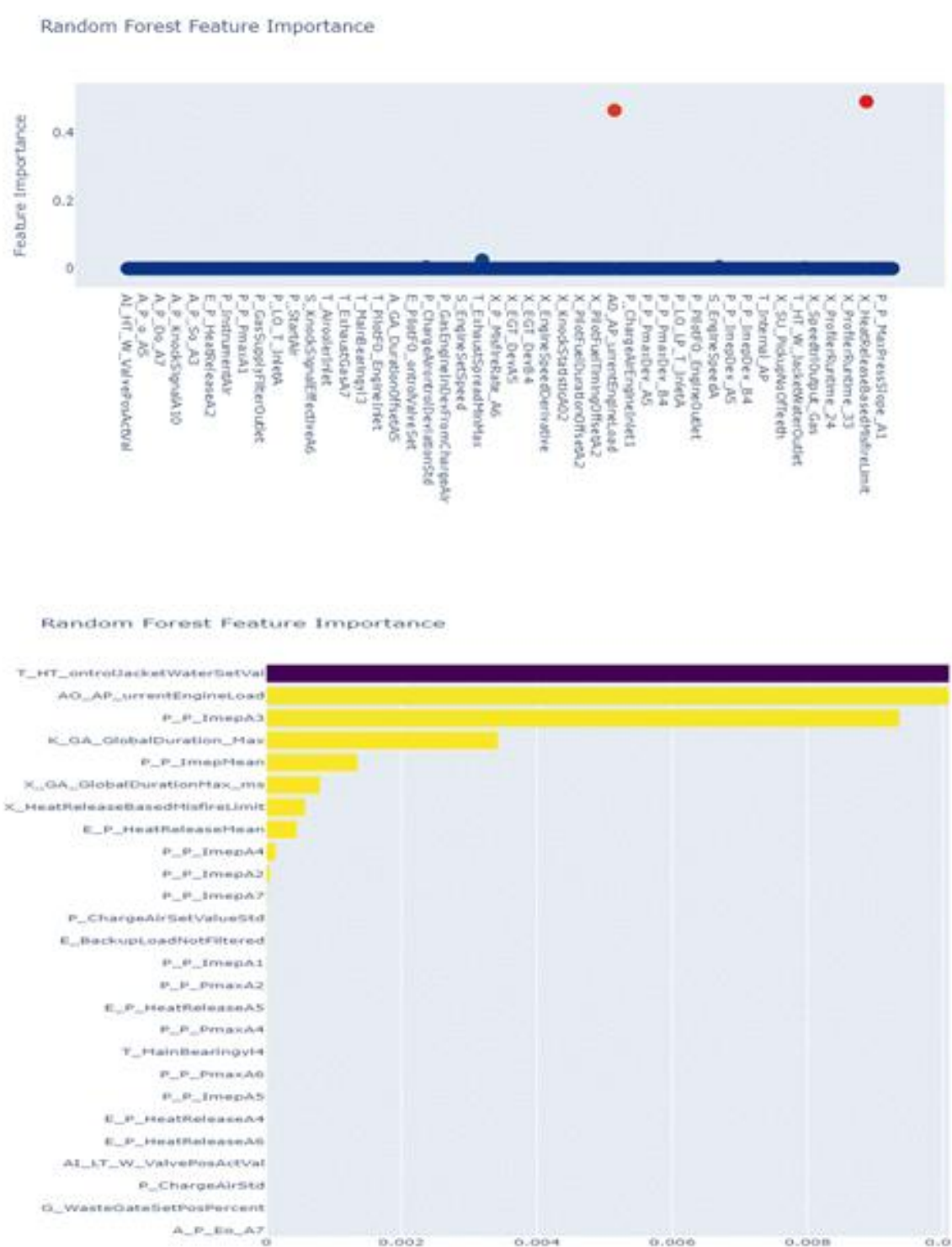
| Time stamp          | 5978 | 5979   | 5980   | 5981 | 5983 | 5984   | 5985   | 5986 | 6128  | 6867 | 6125   |
|---------------------|------|--------|--------|------|------|--------|--------|------|-------|------|--------|
| 2022-04-06 08:30:01 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:02 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:03 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:04 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:05 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:06 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:07 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:08 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:09 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:10 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:11 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:12 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:13 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:14 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:15 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:16 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |
| 2022-04-06 08:30:17 | 0.40 | 1.60 s | 0.03 s | 1.00 | 0.40 | 2.40 s | 0.03 s | 1.00 | 5.0 % | 1.00 | 15.0 % |

- ✓ H사에서 엔진 데이터를 얻을 수 있었다.

- ✓ 2022년 4월 6일 8시 30분 ~ 13시 04분의 엔진에 부하가 걸리고 꺼질 때까지 총 1524개의 Feature에 대해 1초 간격으로 얻을 수 있는 연속적인 시계열 데이터이다.

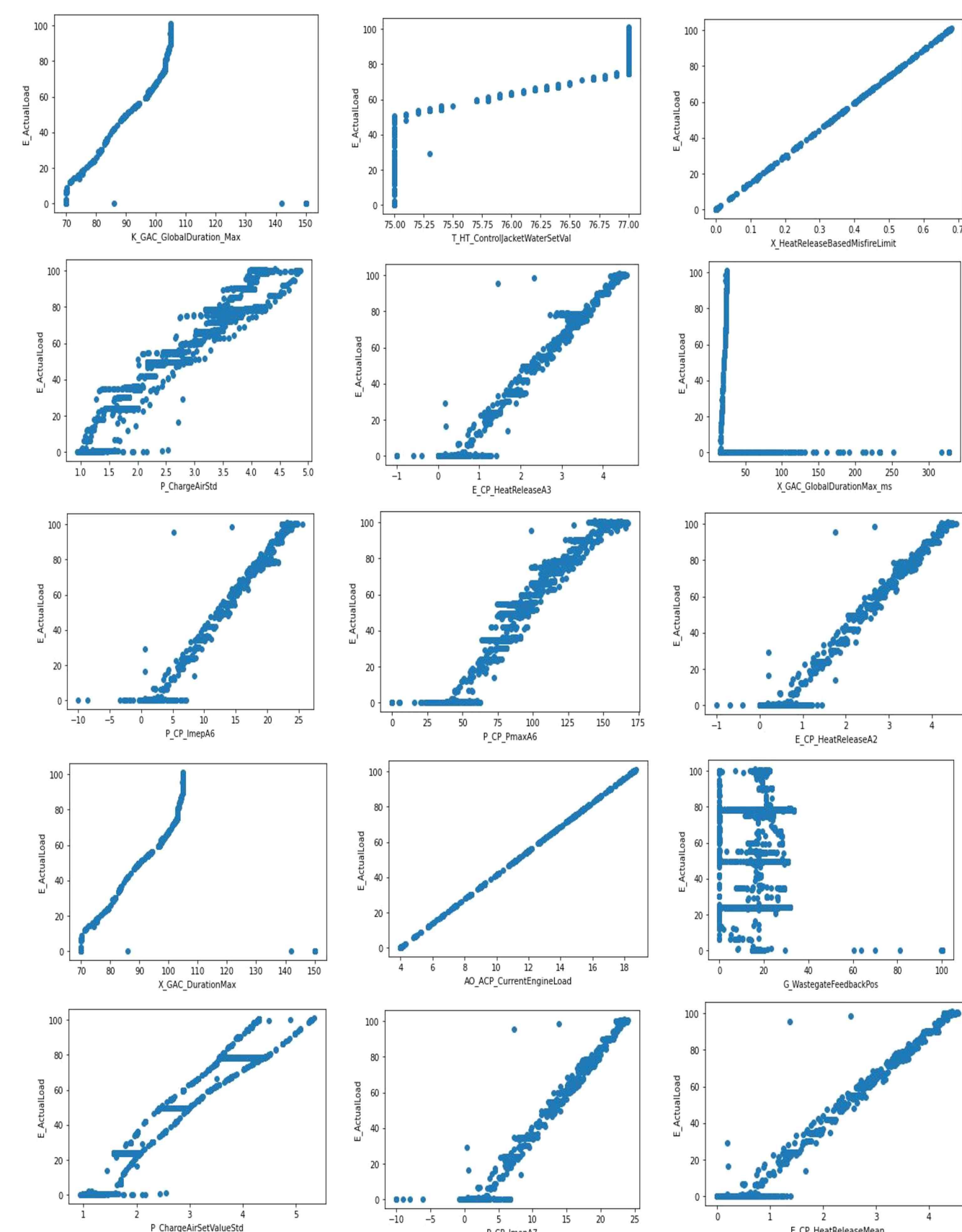
- ✓ Feature에는 Analog 값과 사용자가 직접 입력할 수 있는 Parameter값, Characteristic map값, Discrete값이 있다.

### 데이터 중요도 시각화



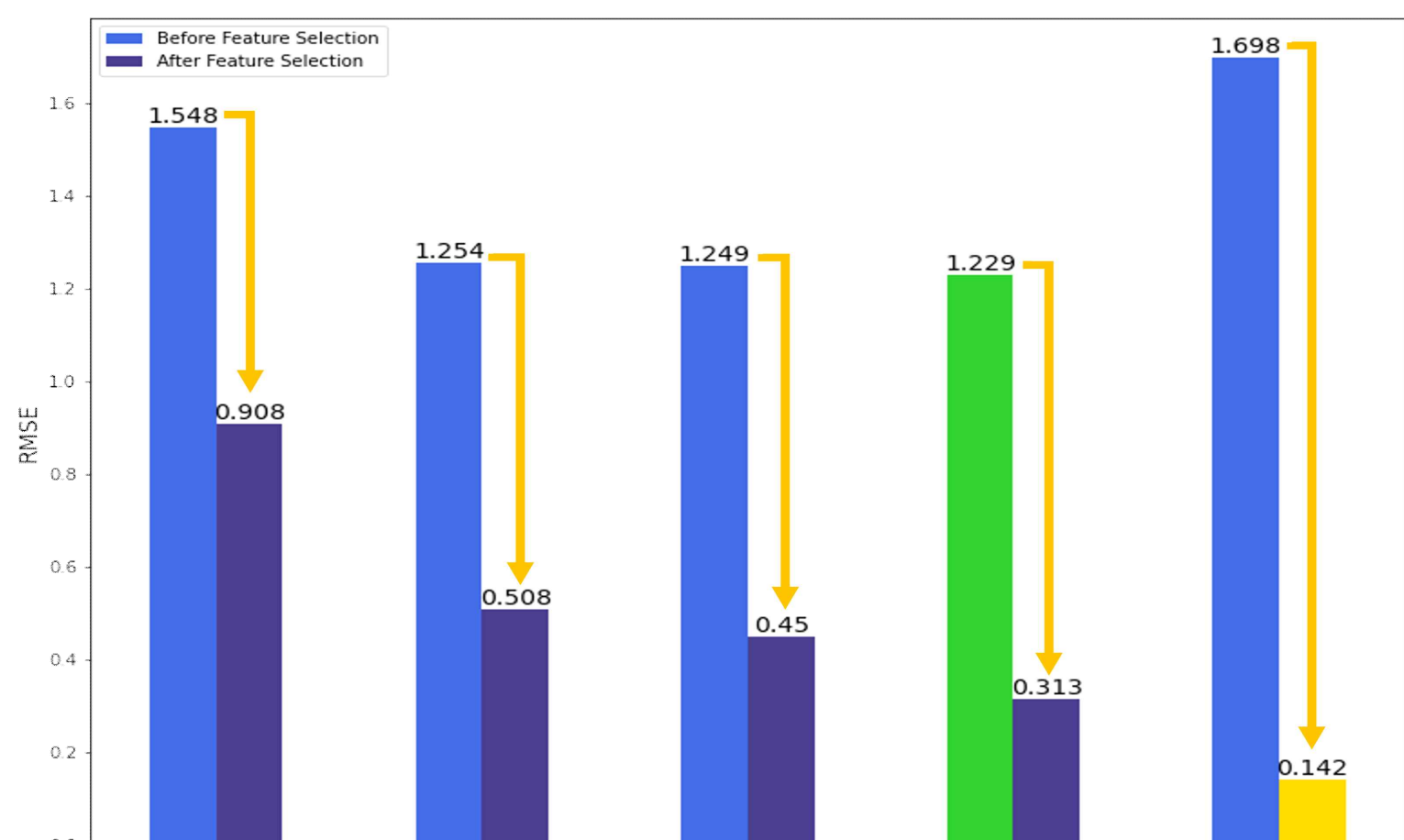
- ✓ Random Forest를 기반으로 Feature Importance를 시각화한 산점도와 그래프

### 데이터 분석 결과

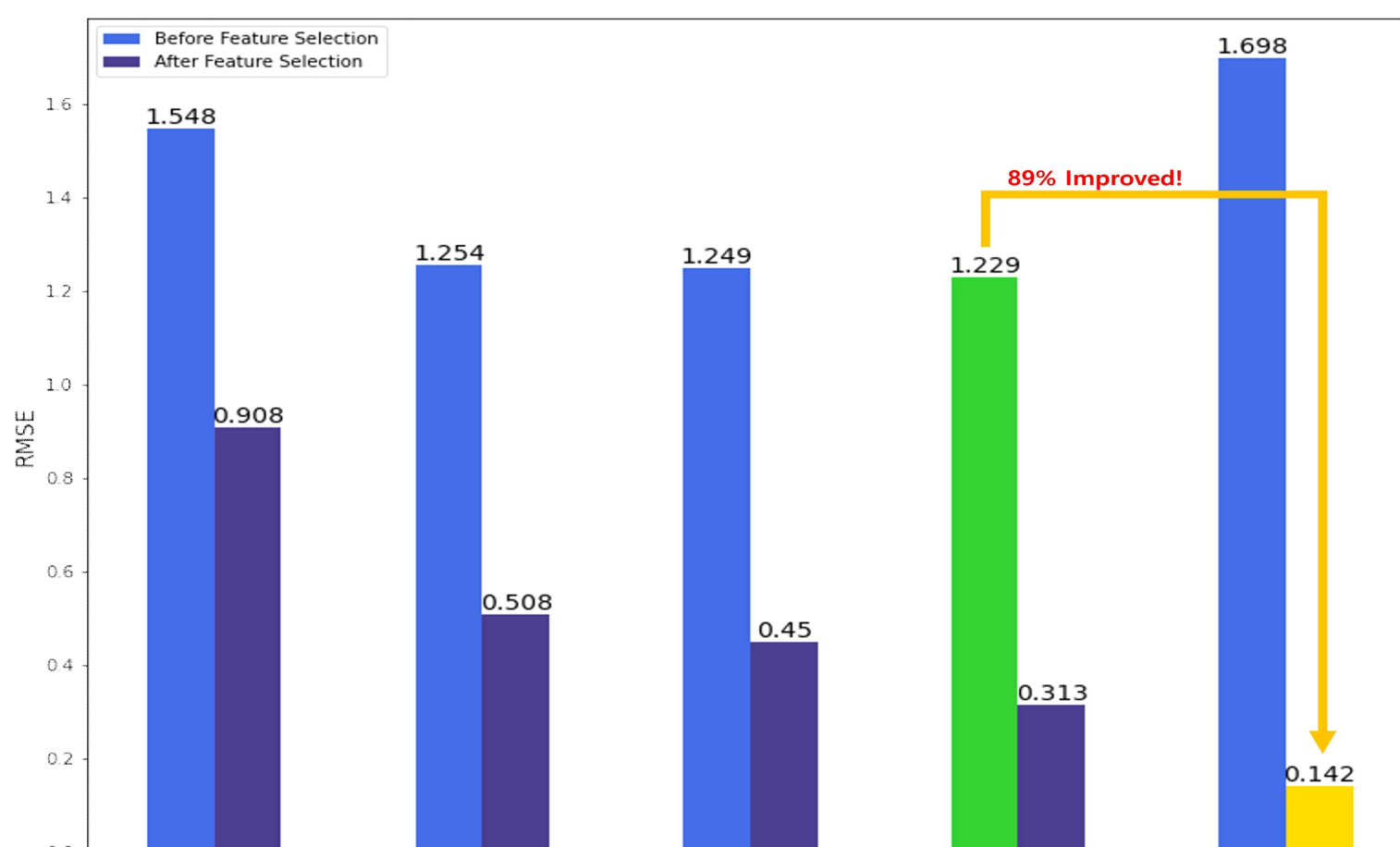


- ✓ Feature importance를 기반으로 선택된 Feature들은 Target value와 대부분 선형성을 가지고 있다.

## 결론 및 향후 효과



Feature Selection 전과 후 전체적인 모델의 RMSE 비교 그래프



Feature Selection 전과 후의 가장 좋은 성능을 가지는 모델 간의 비교

### 결론

- ✓ Feature Selection을 하고 난 후 RMSE(낮을수록 좋음)가 기존보다 많이 개선되었다.
- ✓ 가장 좋은 모델이 XgBoost에서 Linear 회귀 모델로 바뀌었다. 이는 데이터 분석 결과 해당 데이터는 Linear 모델이 더 적합한 데이터임을 알 수 있다.

### 향후효과

- ✓ TBM에서 CBM 방식의 기계 보수 방식의 근거가 될 수 있다.
- ✓ Attention기반의 Transformer등 최신 딥러닝 모델을 적용하여 더 좋은 결과가 기대된다.