



J 공장 장비 유지 보수 해결안

AI_Section2_Project

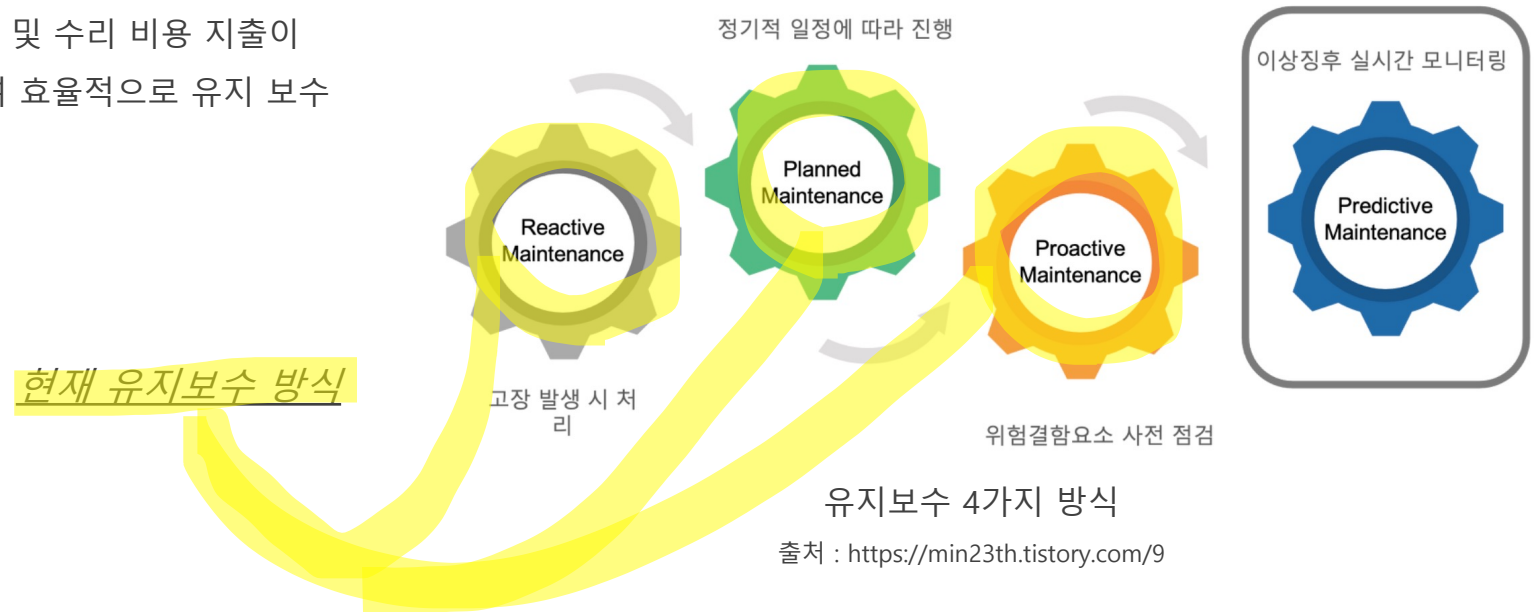
AI_17_김준혁

Project 목표

AI_Section2_Project

- 현재 상황 -

J 공장의 회전체 장비의 유지 보수 및 수리 비용 지출이 많은 상황, 장비의 RUL을 예측하여 효율적으로 유지 보수하기를 원함



- RUL : (Remaining Useful Life) 기계 또는 장비의 잔여 유효 수명

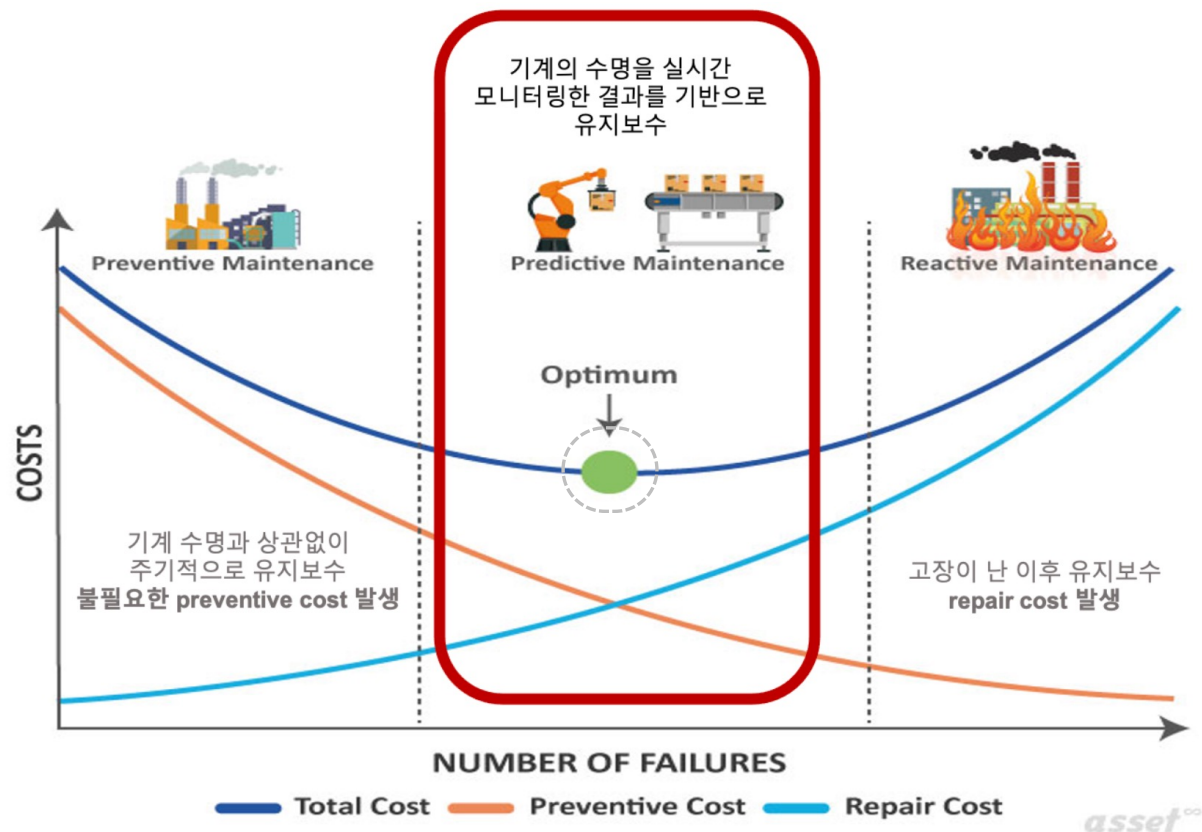
Project 목표

AI_Section2_Project

- 목표 -

장비의 이상 감지를 위해 장착되어 있는 원격 센서를 통한 전압, 회전력, 압력, 진동 관측값과, 열화 지수 등의 데이터를 사용하여 장비의 (RUL)을 예측을 통해 유지 보수 필요 시점을 관제에 알림.

이를 통해 유지 보수 및 수리 비용을 절감하고, 더욱 효율적으로 공장이 가동될 것으로 예상



예지보전

출처 : <https://min23th.tistory.com/9>

- RUL : (Remaining Useful Life) 기계 도는 장비의 잔여 유효 수명

목차

I. Data 분석 및 정제

- 1) Dataset 설명
- 2) Data EDA, Feature Engineering
- 3) Data 분석
- 4) ML 문제 정의
- 5) ML모델 학습 전, Data 전처리

II. ML모델 학습 및 검증

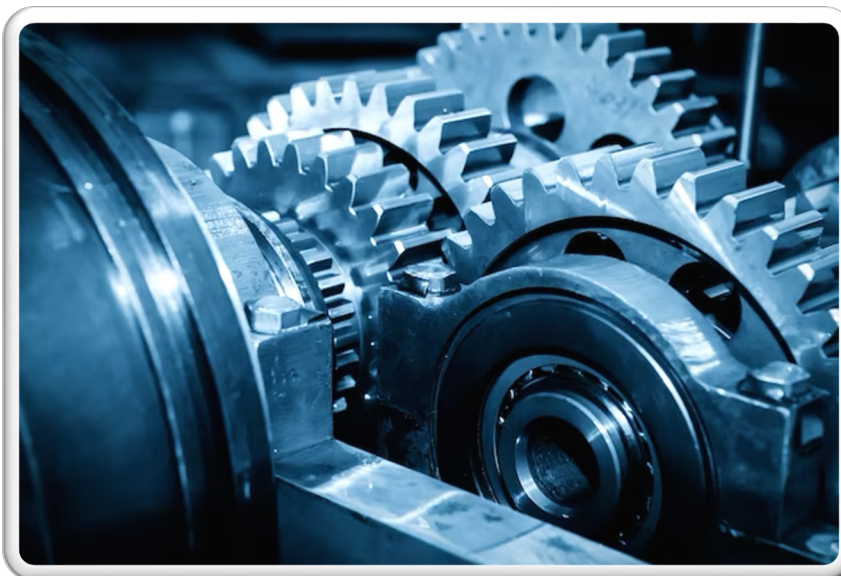
- 1) 기준, 최종 모델 및 평가 지표
- 2) 최종 ML모델 학습 및 검증
- 3) 기준 모델과 최종 모델 비교

III. 최종 ML모델 해석

- 1) 중요 특성 확인
- 2) 특성 값에 따른 변화

IV. 결론

- 1) 콘텐츠에 대한 내용을 적어요



I. Data 분석 및 정제

1) Dataset 설명

AI_Section2_Project

장비가 고장 날 때까지의 Count, 24시간 가동을 기준으로 한 원격 센서로 측정 값이 들어있는 Dataset 사용 (28056개)

Kaggle사이트의 **Predictive Useful Life based into telemetry**

출처 : <https://www.kaggle.com/datasets/tiagotgoz/predictive-useful-life-based-into-telemetry>

Columns info (column 수가 많아 압축해서 설명)

- 장비 ID (1~100)
- 정상작동 Count (유지 보수 및 수리 시 초기화)
- (주요 특성) 전압, 회전력, 압력, 진동 센서 값. 지난 24시간 (평균, 표준편차), 지난 5일 (평균, 표준편차)
- (주요 특성) DI (열화 지수)
- Error 여부 (5종류)
- 구성요소별 유지 보수 주기 (4종류)
- 장비 나이 (장비가 작동한 년(Year))
- 작동 실패 여부
- **RUL (Target)**
- 장비 모델 (4종류) 등

	count	mean	std	min	25%	50%	75%	max
machineID	28056.0	50.557243	28.629237	1.000000	26.000000	51.000000	75.000000	100.000000
time_in_cycles	28056.0	166.031865	94.886557	1.000000	84.000000	164.000000	245.000000	363.000000
voltmean_24h	28056.0	170.760430	4.706852	157.745683	168.055649	170.192722	172.475352	218.265191
rotatemean_24h	28056.0	446.784770	17.716433	271.246607	441.536424	449.192198	456.338450	493.381312
pressuremean_24h	28056.0	100.806369	4.626168	90.973896	98.660468	100.103726	101.586085	152.314600
vibrationmean_24h	28056.0	40.377697	2.038066	36.163723	39.368793	40.069101	40.828250	61.113082
voltsd_24h	28056.0	14.900058	2.256319	7.612362	13.331030	14.835504	16.413618	26.725027
rotatesd_24h	28056.0	49.933053	7.683843	25.169282	44.685678	49.554386	54.856157	101.437628
pressuresd_24h	28056.0	10.047270	1.705857	4.513850	8.922390	9.915881	10.986283	21.138799
vibrationsd_24h	28056.0	5.007331	0.801517	2.433576	4.466938	4.964931	5.489964	10.060271
voltmean_5d	28056.0	170.750530	2.640913	164.258868	169.227681	170.224214	171.373839	193.359322
rotatemean_5d	28056.0	446.840126	10.369640	354.269938	445.222480	449.195954	452.593356	468.068473
pressuremean_5d	28056.0	100.791378	2.831350	96.254848	99.439580	100.103044	100.857990	125.078212
vibrationmean_5d	28056.0	40.375005	1.225228	38.266496	39.745411	40.075951	40.467239	52.143560
voltsd_5d	28056.0	3.606863	2.362023	0.145179	2.239491	3.009315	3.947089	22.722184
rotatesd_5d	28056.0	12.720902	9.630125	0.742933	7.481958	10.010344	13.200899	90.129985
pressuresd_5d	28056.0	2.819368	2.897263	0.147009	1.471053	1.971084	2.594406	26.565238
vibrationsd_5d	28056.0	1.348784	1.205935	0.073920	0.736431	0.993293	1.320537	10.310603
error1	28056.0	0.027409	0.165445	0.000000	0.000000	0.000000	0.000000	2.000000
error2	28056.0	0.026661	0.163727	0.000000	0.000000	0.000000	0.000000	2.000000
error3	28056.0	0.022776	0.150854	0.000000	0.000000	0.000000	0.000000	2.000000
error4	28056.0	0.018855	0.136800	0.000000	0.000000	0.000000	0.000000	2.000000
error5	28056.0	0.009089	0.095278	0.000000	0.000000	0.000000	0.000000	2.000000
comp1	28056.0	1.023097	9.558970	0.000000	0.000000	0.000000	0.000000	270.000000
comp2	28056.0	1.030546	9.030819	0.000000	0.000000	0.000000	0.000000	270.000000
comp3	28056.0	0.934381	8.837160	0.000000	0.000000	0.000000	0.000000	270.000000
comp4	28056.0	0.979398	8.887214	0.000000	0.000000	0.000000	0.000000	270.000000
age	28056.0	11.200064	5.904624	0.000000	6.000000	11.000000	16.000000	20.000000
DI	28056.0	1.353352	0.687165	0.000000	0.787464	1.312205	1.907928	2.701757
RULWeek	28056.0	6.429855	5.779276	1.000000	3.000000	5.000000	8.000000	39.000000
failed	28056.0	0.017002	0.129280	0.000000	0.000000	0.000000	0.000000	1.000000
RUL	28056.0	41.924187	40.474382	1.000000	15.000000	30.000000	55.000000	270.000000
RUL_I	28056.0	40.924187	40.474382	0.000000	14.000000	29.000000	54.000000	269.000000

	count	unique	top	freq
datetime	28056	361	2015-02-21	93
model	28056	4	model3	9461
failure	28056	5	none	27579

I. Data 분석 및 정제

2) Data EDA, Feature Engineering

AI_Section2_Project

EDA, Feature Engineering

- Dataset의 결측치, 중복치, 이상치는 없는 것을 확인
- 장비 또는 원격 센서를 통해 입력받기 어려운 column Drop
- 각 센서의 24시간 평균값에 표준편차를 계산해 새로운 column 생성 (Max 값, Min 값)
- 원활한 코딩 작업을 위해 Column 명 변경

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
ID	28056.0	NaN	NaN	NaN	50.557243	28.629237	1.0	26.0	51.0	75.0	100.0
Datetime	28056	361	2015-02-21	93	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Cycle	28056.0	NaN	NaN	NaN	166.031865	94.886557	1.0	84.0	164.0	245.0	363.0
Voltage_mean	28056.0	NaN	NaN	NaN	170.76043	4.706852	157.745683	168.055649	170.192722	172.475352	218.265191
Rotate_mean	28056.0	NaN	NaN	NaN	446.78477	17.716433	271.246607	441.536424	449.192198	456.33845	493.381312
Pressure_mean	28056.0	NaN	NaN	NaN	100.806369	4.626168	90.973896	98.660468	100.103726	101.586085	152.3146
Vibration_mean	28056.0	NaN	NaN	NaN	40.377697	2.038066	36.163723	39.368793	40.069101	40.82825	61.113082
Model	28056	4	model3	9461	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Age	28056.0	NaN	NaN	NaN	11.200064	5.904624	0.0	6.0	11.0	16.0	20.0
DI	28056.0	NaN	NaN	NaN	1.353352	0.687165	0.0	0.787464	1.312205	1.907928	2.701757
Failed	28056.0	NaN	NaN	NaN	0.017002	0.12928	0.0	0.0	0.0	0.0	1.0
RUL	28056.0	NaN	NaN	NaN	40.924187	40.474382	0.0	14.0	29.0	54.0	269.0
Voltage_max	28056.0	NaN	NaN	NaN	5.228891	0.028101	5.151849	5.211564	5.225952	5.240877	5.455335
Voltage_min	28056.0	NaN	NaN	NaN	5.05485	0.031586	4.944471	5.0354	5.052501	5.069994	5.329
Rotate_max	28056.0	NaN	NaN	NaN	3.90034	0.339312	0.0	3.72416	3.920474	4.090162	5.401796
Rotate_min	28056.0	NaN	NaN	NaN	3.977567	0.326485	0.0	3.793925	3.983021	4.149518	5.476551
Pressure_max	28056.0	NaN	NaN	NaN	4.716172	0.044156	4.607287	4.69366	4.709602	4.726138	5.108472
Pressure_min	28056.0	NaN	NaN	NaN	4.518054	0.04609	4.363132	4.493877	4.513007	4.532546	4.969868
Vibration_max	28056.0	NaN	NaN	NaN	3.835831	0.046946	3.708456	3.810086	3.829435	3.849867	4.216893
Vibration_min	28056.0	NaN	NaN	NaN	3.592288	0.053037	3.433535	3.562511	3.586824	3.612081	4.055861

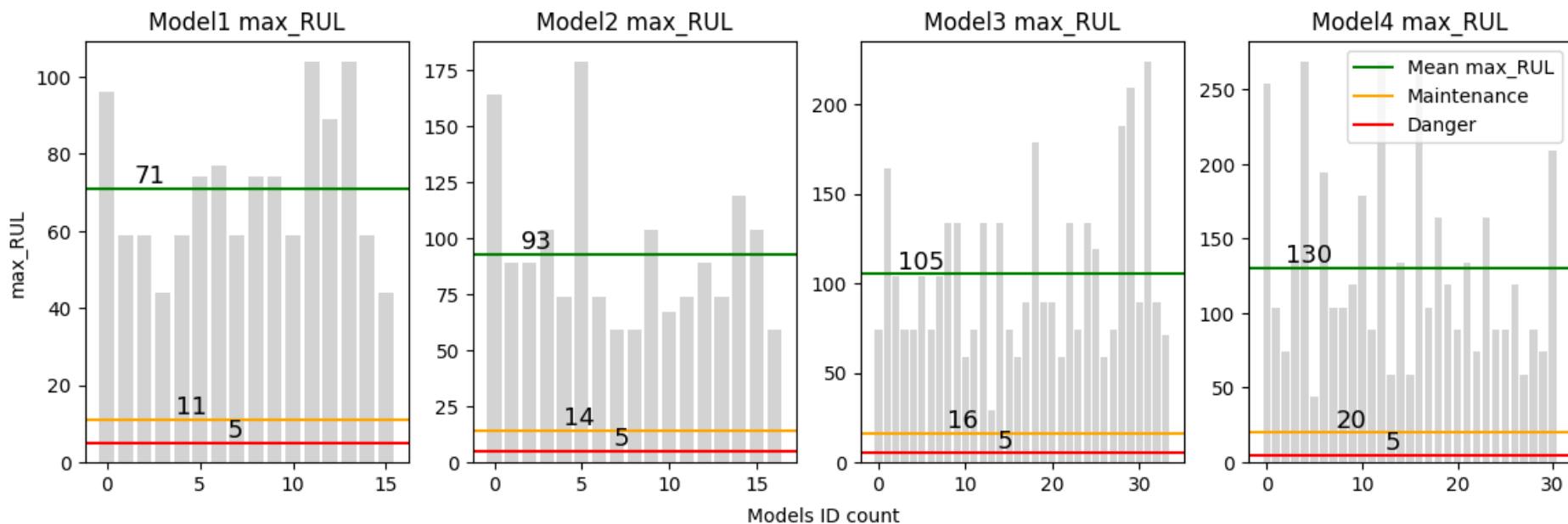
I. Data 분석 및 정제

3) Data 분석

AI_Section2_Project

RUL

- 각 장비 모델의 ID 별 RUL 시각화
- 모델 3, 4는 ID 별 RUL 값의 차이가 많이 나는 것을 확인
- 유지 보수 임계점 설정 **예시** 추가 (평균 RUL * 0.15)



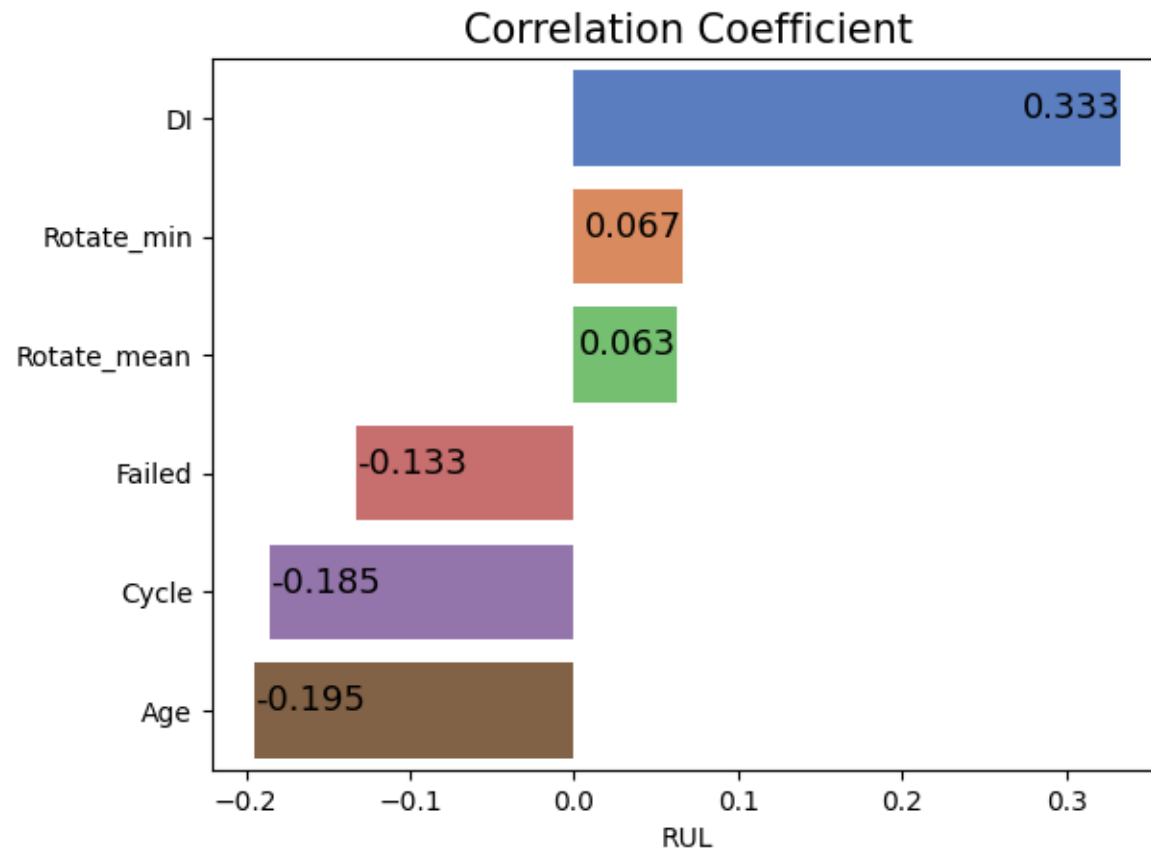
I. Data 분석 및 정제

3) Data 분석

AI_Section2_Project

Target(RUL)과 다른 특성들의 상관관계

- 양의 상관관계 Top3
 - DI : 열화 지수
 - Rotate_min : 지난 24시간 관측한 회전력의 최솟값
 - Rotate_mean : 지난 24시간 관측한 회전력의 평균값
- 음의 상관관계 Top3
 - Age : 장비 나이
 - Cycle : 정상 작동 Count
 - Failed : 작동 실패 여부



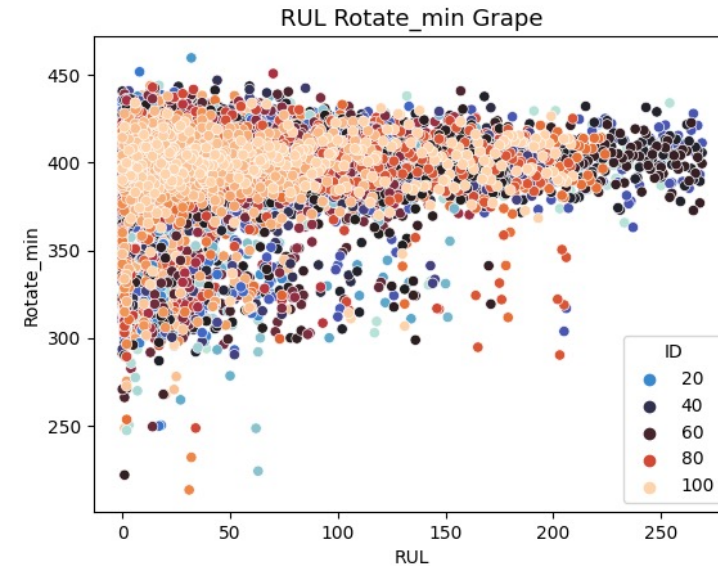
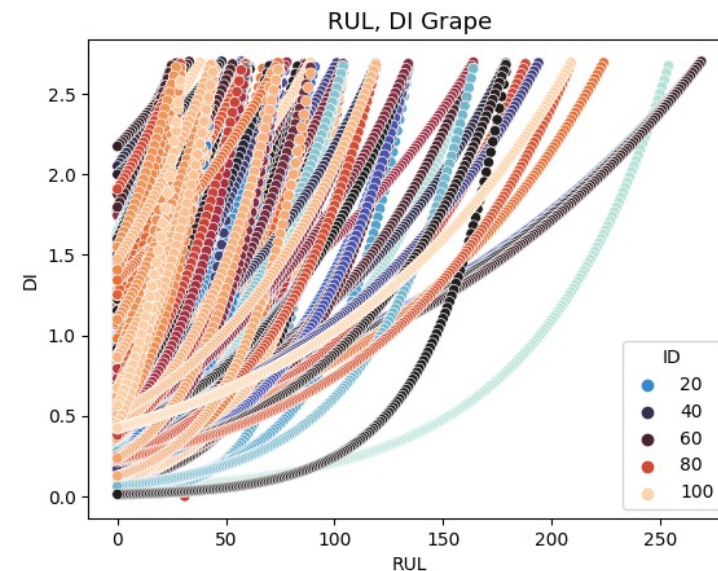
I. Data 분석 및 정제

3) Data 분석

AI_Section2_Project

Target(RUL)과 다른 특성들의 상관관계

- 주요 특성에 포함되어 있는 DI, Rotate_min과 RUL 관계 시각화
- DI : 비선형 관계인 것을 확인
- Rotate_min : 상관계수와 같이 참고했을 때, RUL 과는 상관이 거의 없는 특성
- 이외 주요 특성(원격 센서 관측 값)들도 DI를 제외하면 높은 상관관계에 있는 특성이 없음
- 추가로 Cycle은 선형 관계





I. Data 분석 및 정제

4) ML 문제 정의

AI_Section2_Project

Project 목적은 유지 보수를 더 효율적으로 할 수 있도록 장비 모델 별 RUL, 유지 보수 필요 시점을 예측하는 것이므로, 회귀, 분류 문제 둘 다 사용 가능

예)

- 회귀일 경우 직접 장비의 RUL을 예측하고 상황에 따라 유연하게 대처할 수 있다.
- 분류일 경우 RUL 값에 따라 임곗값을 정해서 상태(Status)(정상, 유지 보수 필요, 매우 위험) 구간을 설정해서 예측할 수 있다.

학습 모델 선택

유지 보수 필요 구간에만 집중한다면 분류 문제가 좋겠지만, RUL 값을 직접 예측하여 필요에 따라 임곗값을 바꿀 수 있고, 관리자가 유연하게 행동할 수 있도록 하는 것이 더 효율적이라고 판단.

Data 분석 결과 선형성 및 비선형성을 띤 특성이 많이 없다고 판단되어 트리 모델이 적합하다고 판단.

따라서 트리 회귀 모델을 사용할 것

(가설)

- 관측된 주요 특성(전압, 회전력, 압력, 진동, 열화 지수) 값들이 변하면서 RUL에 영향을 주는지?

I. Data 분석 및 정제

5) ML모델 학습 전, Data 전처리

AI_Section2_Project

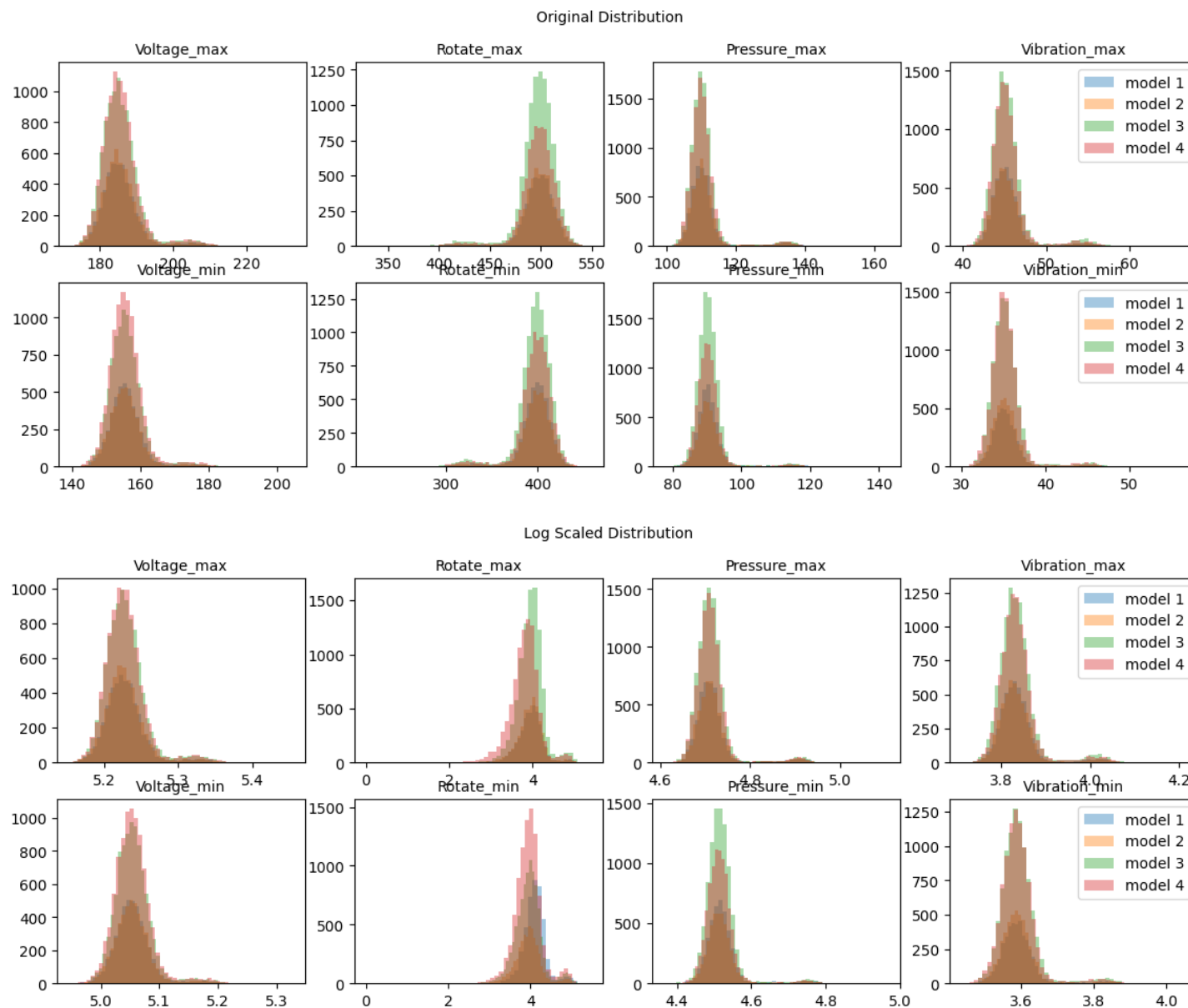
Data 전처리

트리 모델이라 전처리가 많이 필요하지는 않지만, 전처리 유무에 따른 성능 차이를 확인했으므로, 전처리 진행

- 전압, 회전력, 압력, 진동 값의 분포 시각화 (왜도(Skewness)의 꼬리가 많이 치우쳐 있는 것을 확인할 수 있다.)
- Log scaling을 통해 왜도 값을 낮춤 (Grape 상으로 큰 변화는 없는 것으로 보이지만, 이후 모델 학습 시 성능 향상을 보였다.)

Training, Test set 분리 후

- 표준화(StandardScaler),
- 다항 특성(PolynomialFeatures) 적용





II. ML모델 학습 및 검증

1) 기준, 최종 모델 및 평가 지표

AI_Section2_Project

기준 ML모델 : DecisionTreeRegressor

최종 ML모델 : ComponentwiseGradientBoostingSurvivalAnalysis

최종 모델 설명

장비의 수명을 예측하기 위해 생존 분석 모델 사용

위 모델은 지도학습 방식이고, Target(label)에는 event(Failed), time(RUL)이 필요. time 또는 event가 발생할 확률을 예측.

GradientBoosting이라는 Ensemble 기법을 사용하며, 트리 모델 한계상 외삽 Data는 예측 불가

평가 지표

RUL의 실제 값과 예측 값의 오차가 줄어들어야 하기 때문에

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)

위 3가지를 평가 지표로 사용



II. ML모델 학습 및 검증

2) 최종 ML모델 학습 및 검증

AI_Section2_Project

Hyperparameter 설정

- 최적의 Hyperparameter를 설정하기 위해 Bayesian Search 사용 (평가 지표는 MAE 사용)

```
best params : {'learning_rate': [0.18258899382089233], 'n_estimators': [140.0], 'subsample': [0.9068277038201439]}
best score : 25.43696380450882
```

사용한 Hyperparameter

- learning_rate : 학습률
- n_estimators : 수행할 Boosting 단계 수
- subsample : 개별 학습 시 사용할 샘플의 양

모델 검증

- 전체 데이터는 28056개이지만, 데이터에 있는 장비 Model이 4종류이고, 각각의 평균 RUL 차이를 고려하여 교차 검증으로 진행
- 교차 검증을 사용 할 것이기 때문에, Dataset을 Training set, Test set로 분리
(Test set은 각 장비 모델에 속한 3개의 장비ID 랜덤 추출)
- 위의 최적 Hyperparameter를 사용해 학습
- 최종 모델이 기준 모델보다 검증 MAE 값이 낮은 것을 확인

기준 모델

CV average	31.449869
------------	-----------

최종 모델

Train	25.230243
CV average	25.436964

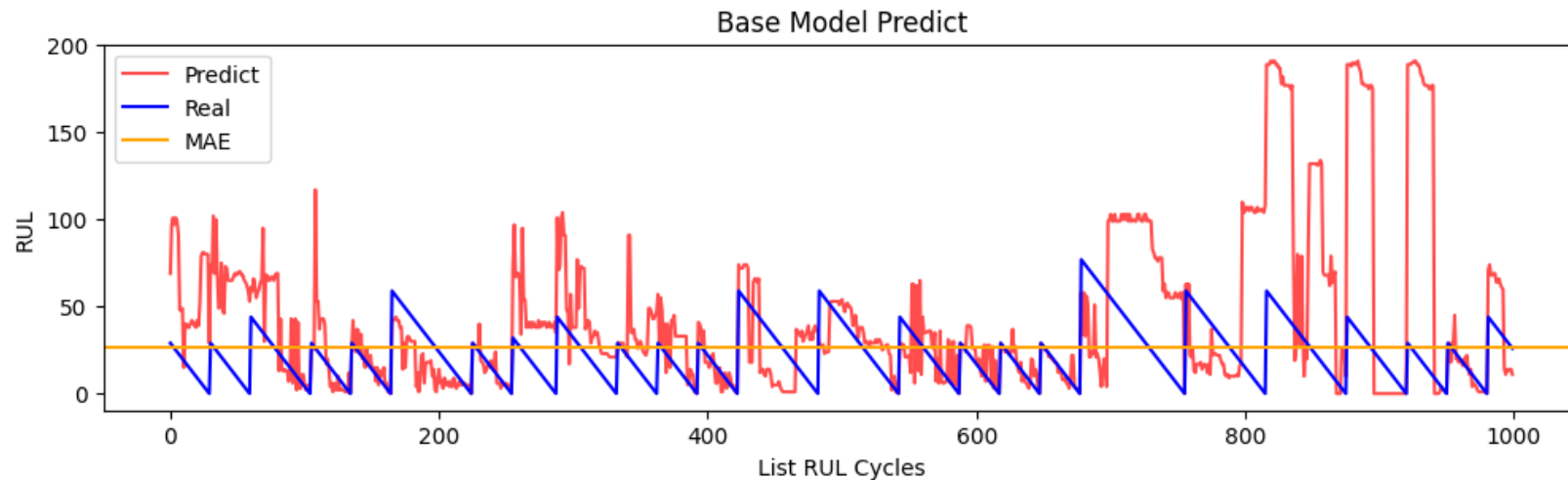
II. ML모델 학습 및 검증

3) 기준 모델과 최종 모델 비교

AI_Section2_Project

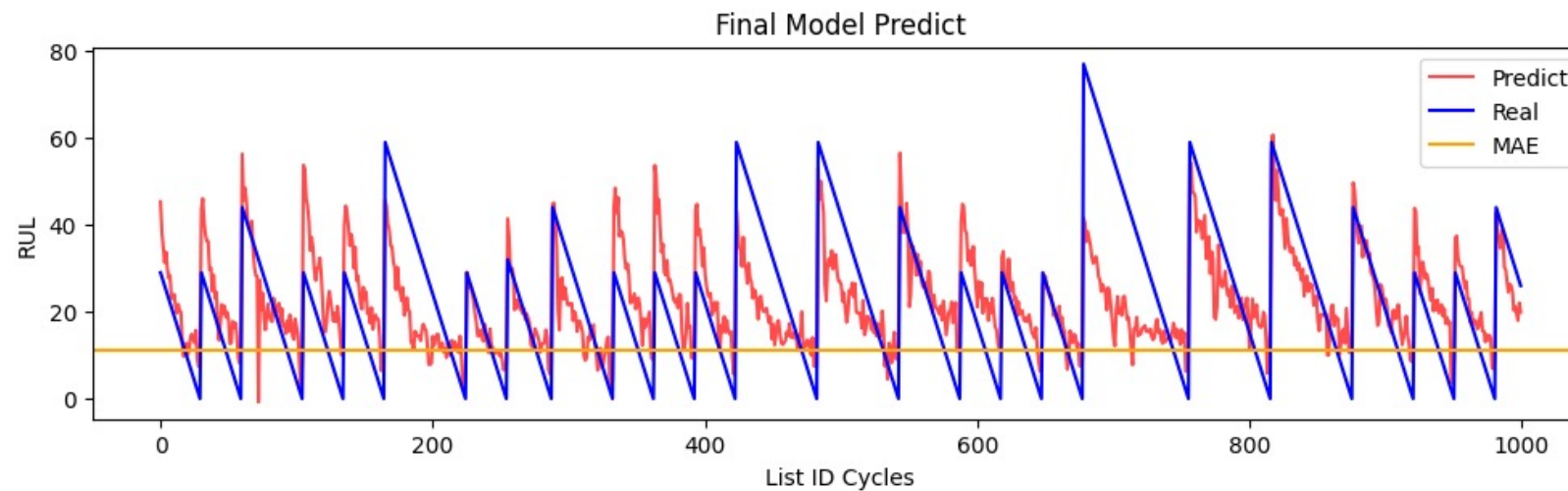
기준 모델 : DecisionTreeRegressor

MSE	1849.884952
RMSE	43.010289
MAE	27.072000



최종 모델 : ComponentwiseGradientBoostingSurvivalAnalysis

MSE	189.326754
RMSE	13.759606
MAE	10.900833



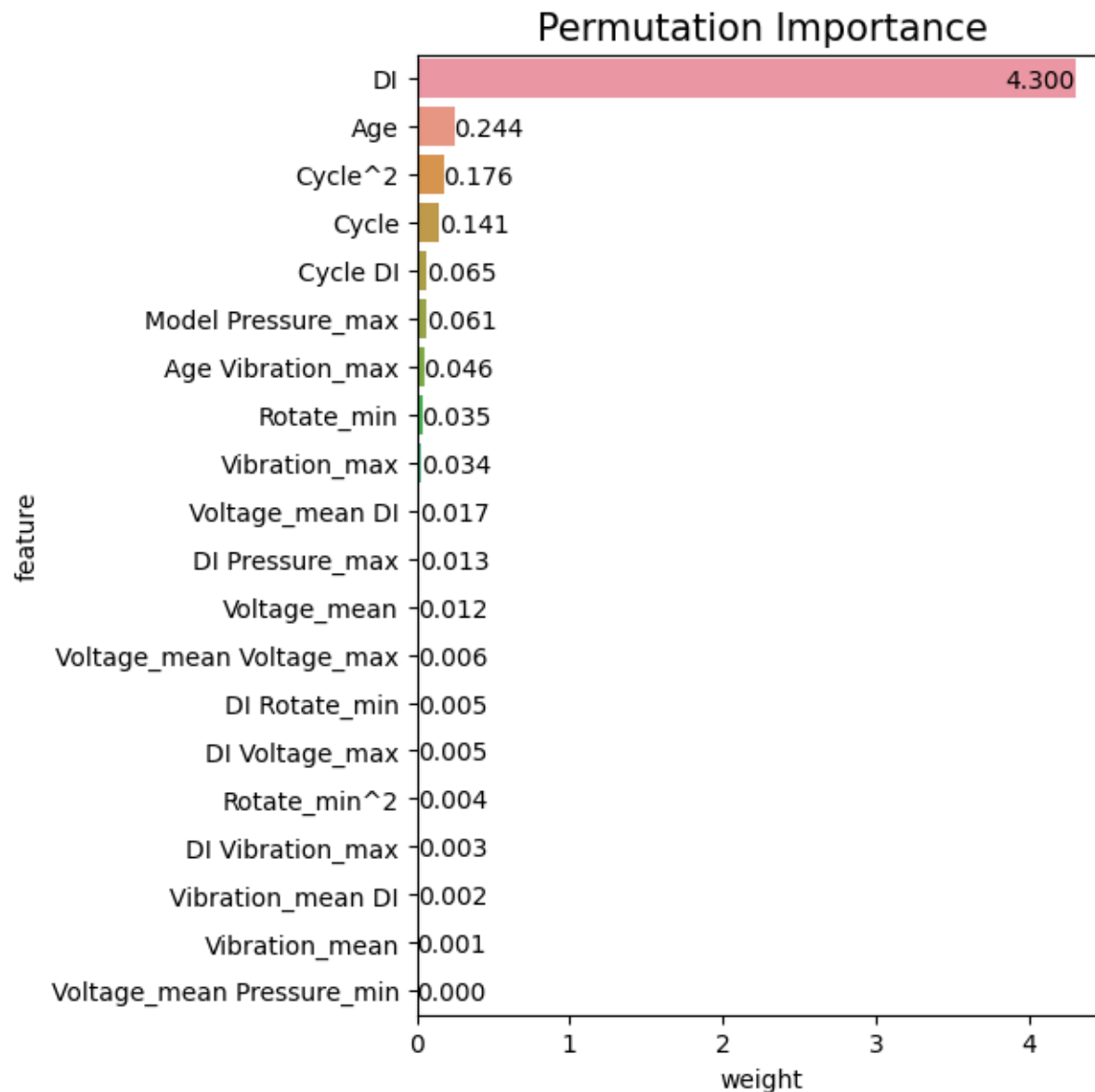
III. 최종 ML모델 해석

1) 중요 특성 확인

AI_Section2_Project

순열 중요도 TOP 20

- 다항 특성은 크게 눈에 띄는 부분이 없으므로 제외하고 해석
- DI(열화 지수) 특성이 모델 학습 시 가장 많은 영향을 주었다는 것을 알 수 있다.
- 다음으로 Age, Cycle 특성이 영향을 많이 주었고, 데이터 분석에서 구한 상관계수와 비교했을 때 실제로 영향을 주는 관계라는 것을 알 수 있다.
- 주요 특성 값은 DI를 제외하고는 매우 낮은 영향력을 보이고 있다.



III. 최종 ML모델 해석

2) 특성 값에 따른 변화

AI_Section2_Project

PDP를 통해 해석

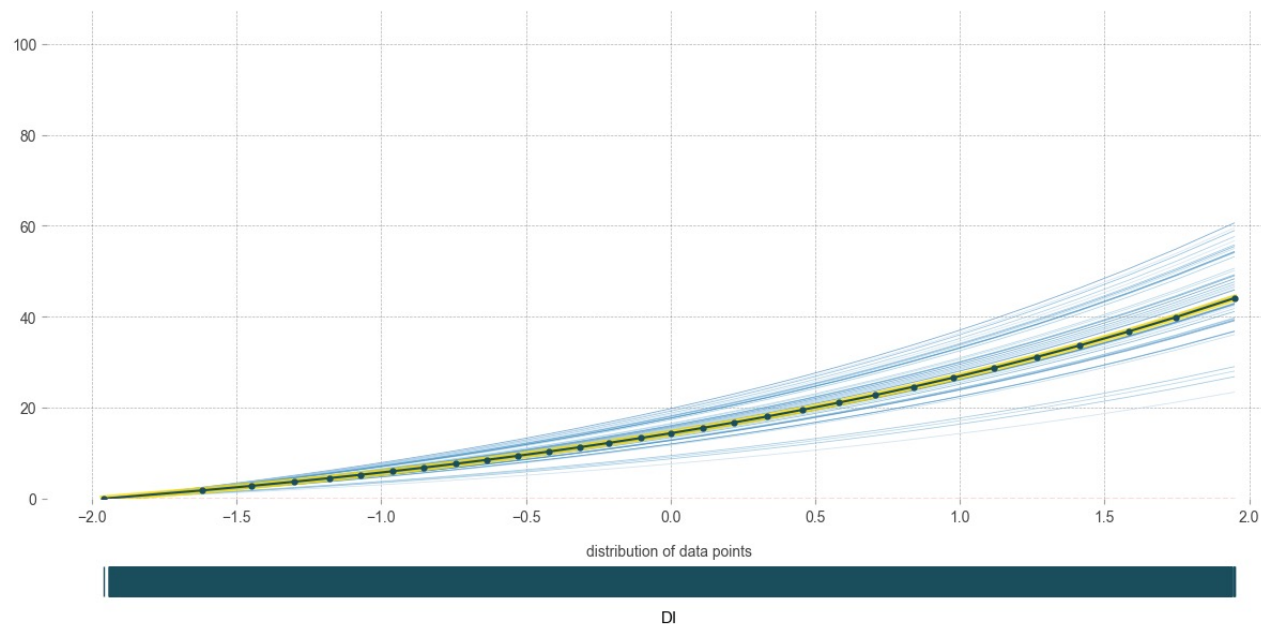
Scaled Data로 분석하기 때문에 값의 변화만 확인

DI 값에 따른 RUL 변화

- DI(열화 지수) 값이 감소할수록, RUL도 감소하는 것을 확인
- 열화란 절연체가 내외부의 영향에 의해 상태가 나빠지는 것
- RUL이 높을수록 장비의 마모가 덜 되었다는 뜻이고, 이에 따라 장비의 가동이 활발할 때 열화 지수 또한 높다는 것으로 추측

PDP for feature "DI"

Number of unique grid points: 30



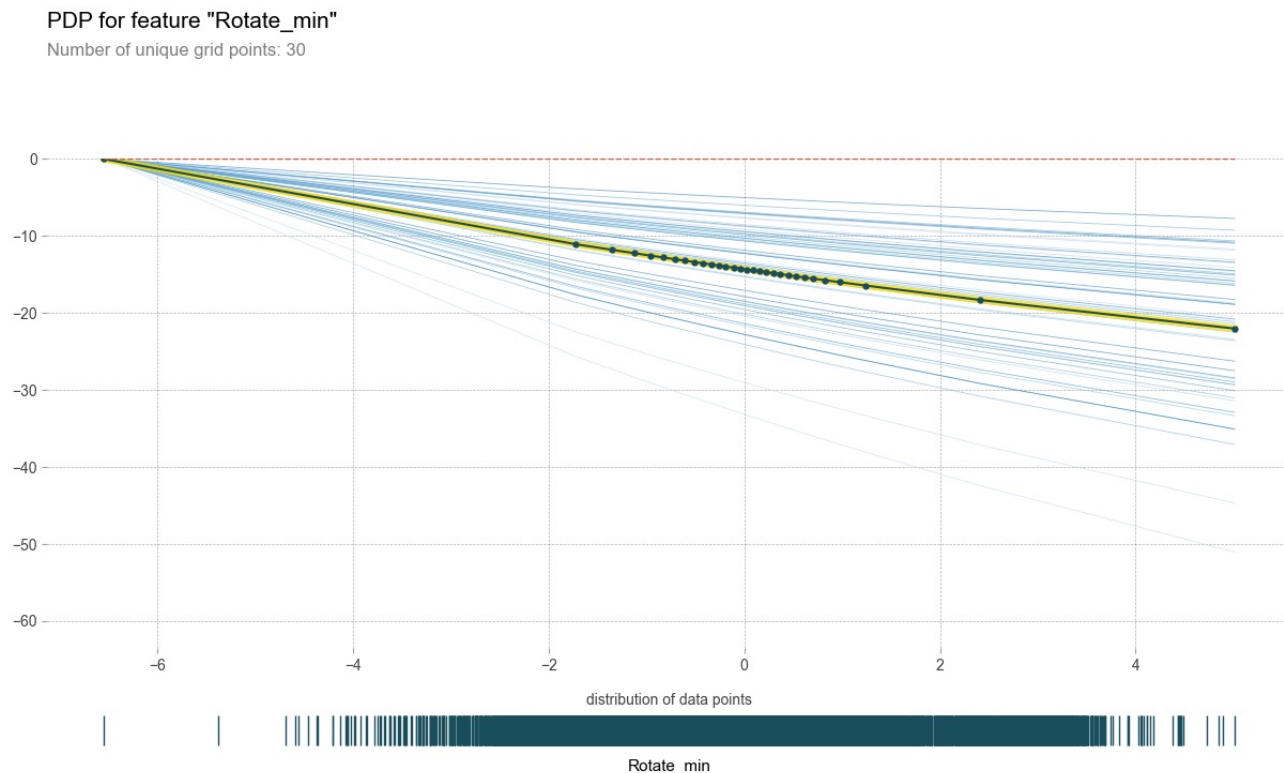
III. 최종 ML모델 해석

2) 특성 값에 따른 변화

AI_Section2_Project

Rotate_min 값에 따른 RUL 변화

- Rotate_min 값이 증가할수록 RUL은 감소
- Torque Sensor를 이용해 값을 받은 것으로 추측
- Torque Sensor는 기계의 회전 축에 가해지는 뒤틀림 (Torque)을 검출하는 방식, 값이 클수록 이상 작동 따라서 값이 높아질수록 RUL이 감소
- Rotate_min의 대략 1분위수 이하 값과 3분위수 이상 값은 분포가 적어 신뢰할 수 없다.



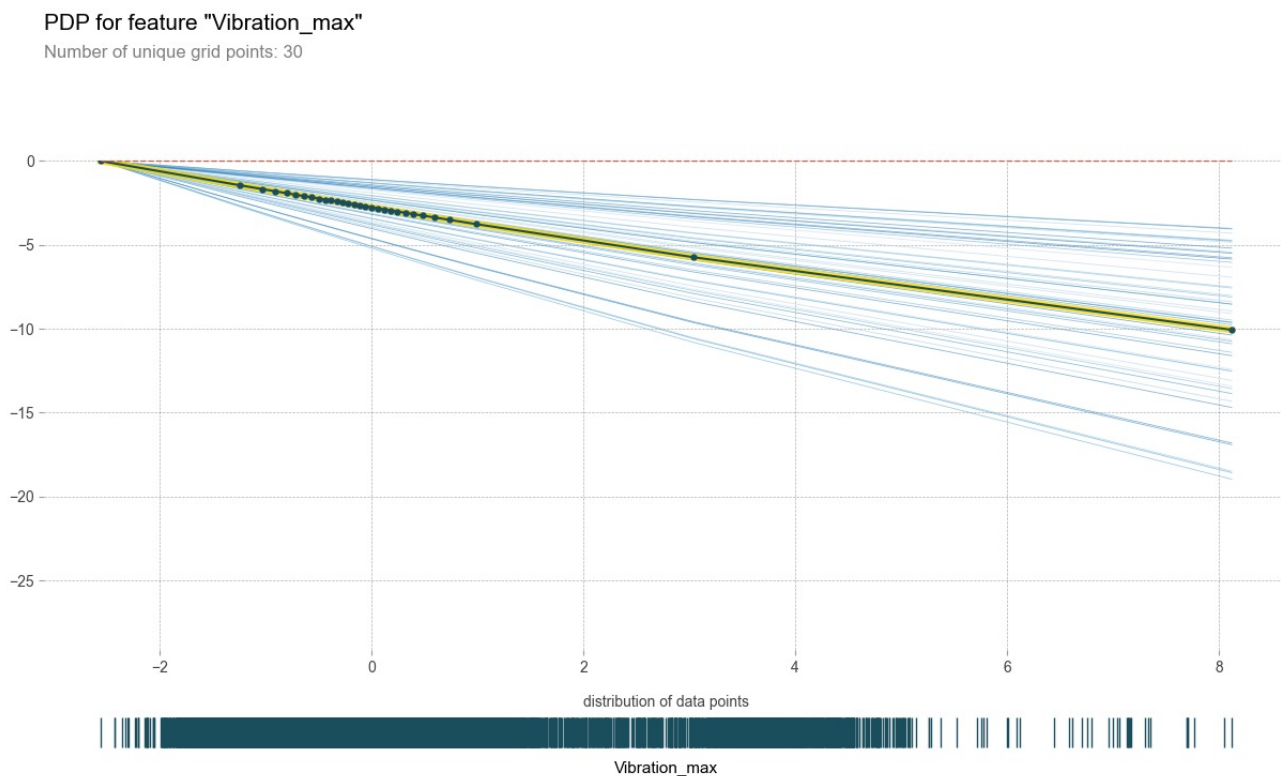
III. 최종 ML모델 해석

2) 특성 값에 따른 변화

AI_Section2_Project

Vibration_max 값에 따른 RUL 변화

- Vibration_max 값이 증가할수록 RUL은 감소
- 진동 값이 커질수록 장비에 이상이 생긴다고 추측. 따라서 RUL이 감소
- 3분위수 이상 값은 분포가 적어서 신뢰할 수 없다.



IV. 결론

AI_Section2_Project

(가설) 관측된 주요 특성(전압, 회전력, 압력, 진동, 열화 지수) 값들이 변하면서 RUL에 영향을 주는지?

- 앞서 진행한 PDP 해석을 통해 압력을 제외한 주요 특성의 값들은 장비의 RUL에 영향을 주는 것을 확인
(압력은 특성 중요도에서 제외되어 측정 불가 = RUL 예측에 영향을 주지 못함)
- 다만 주요 특성 값의 분포가 일정하지 않아서 ML모델이 정확한 RUL을 예측하기 어렵다고 판단.

= 결론 =

- 최종 ML모델을 장비 100개에 대해서 사용 가능하다고 판단됨
실제 값, 예측값 비교 그래프를 확인해 보면 임곗값(유지 보수 필요 시점)보다 아래 있어야할 예측값의 오차가 적기 때문
- 장비의 임곗값을 설정할 때 MAE(오차 절댓값 평균) 값을 고려해서 설정해야 함
(최종 모델 MAE ≈ 11 , 장비 가동이 멈추면 안되므로 더 많은 유예기간 필요)

주의할 점 : 장비를 추가하게 되면 외삽 데이터가 생길 수 있다.

최종 ML모델이 Tree 기반 모델이므로, 외삽 데이터가 입력되면 부정확하게 예측할 가능성이 높아진다.
따라서 현재 장비에만 적용하는 것을 추천

- 추후 계획 -

- 현재 최종 ML모델은 과소 적합이라고 판단됨
- 추후 장비의 Data를 추가하여, Data 분포 문제를 해결하고 ML모델 재학습

THANK YOU!!

김준혁 kjh970104@gmail.com

2023. 02. 08

