

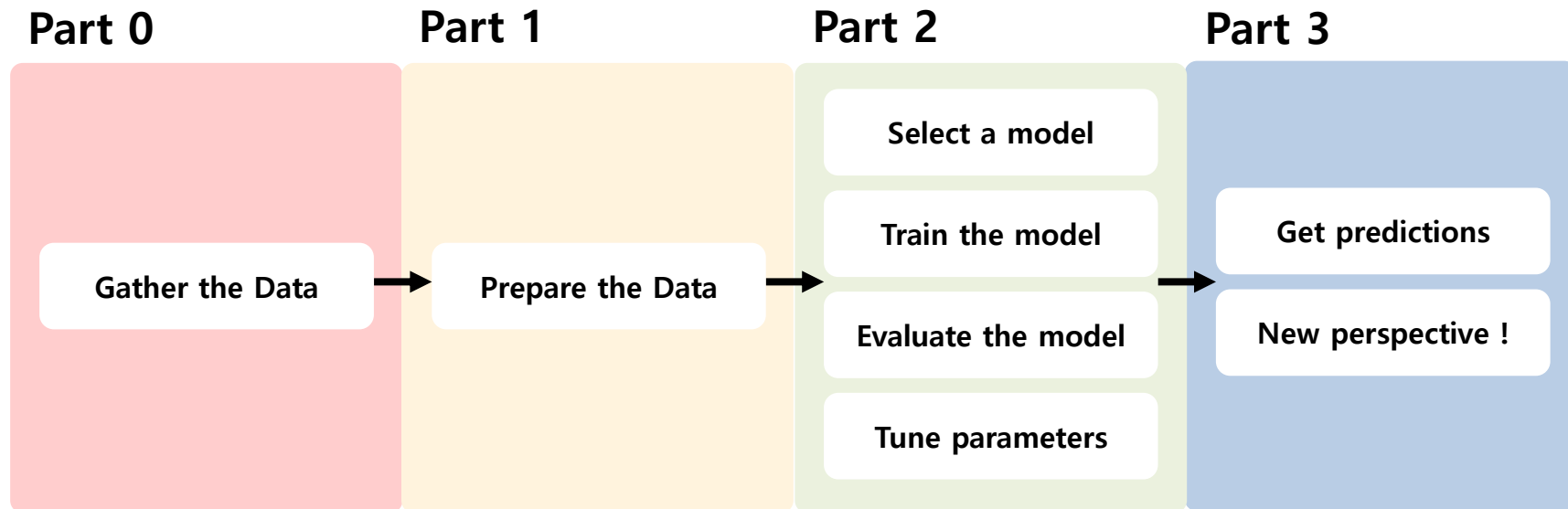
SoC-NAND Validation 데이터 기반 UI예측 Model 개발

2021. 07

솔루션 개발 / SoC / SV & SoC PE / 전해성 인턴

Executive Summary : 7+1 steps of Machine Learning¹⁾

- NAND SI Validation의 ① 목적과 과정 이해 ② I/O Parameter의 통계적 분석 ③ Feature selection 통해 데이터의 특성 확인 및 Model 생성할 수 있는 형태로 전처리
- UI를 예측할 수 있는 Regression model의 ① fitting 및 성능 평가의 반복 ② Grid search로 Hyper-parameter tuning하여 최적화
- 최적화 된 Model의 UI 예측 결과 분석 및 자동화 process와 관련한 향후 발전 방향 모색



[1] ["The 7 Steps of Machine Learning, Towards Data Science"](#), *Towards Data Science*, Yufeng G., last modified 01.11.2017., accessed 22.07.2021.

0 **NAND SI Validation**

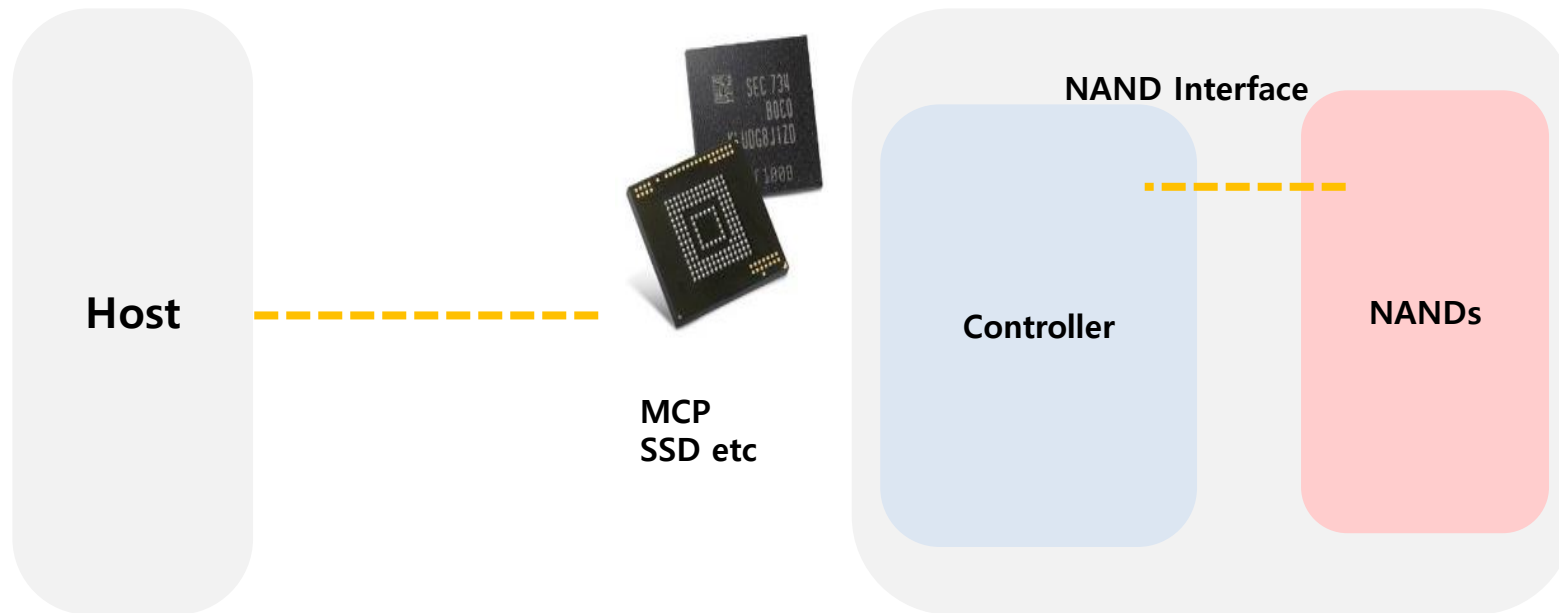
1 **Data Preprocessing**

2 **Model Fitting and Evaluation**

3 **Utilizing Model Prediction**

0. NAND SI Validation – 목표와 과정

- 온도, 전압, Process에 따른 NAND SI Validation.
- P.V.T (Process/Voltage/Temperature) 조건에서 테스트.
- 각 조건에서 Test 결과가 동작에 문제 없는 spec in 수준인지 검증.
- 목표 : 데이터를 활용하여 추가적인 Test 대신 ML model로 ui의 양산성 예측 및 검증



1-1. Data Preprocessing

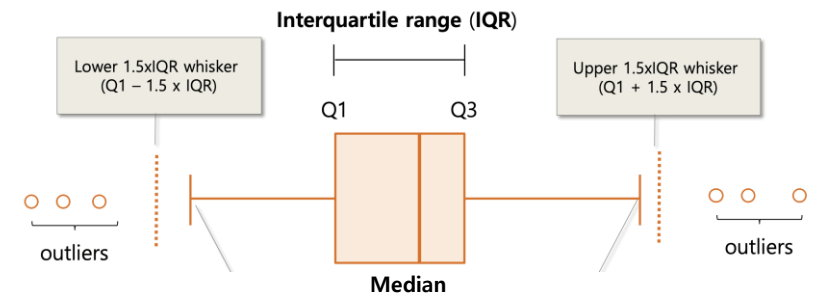
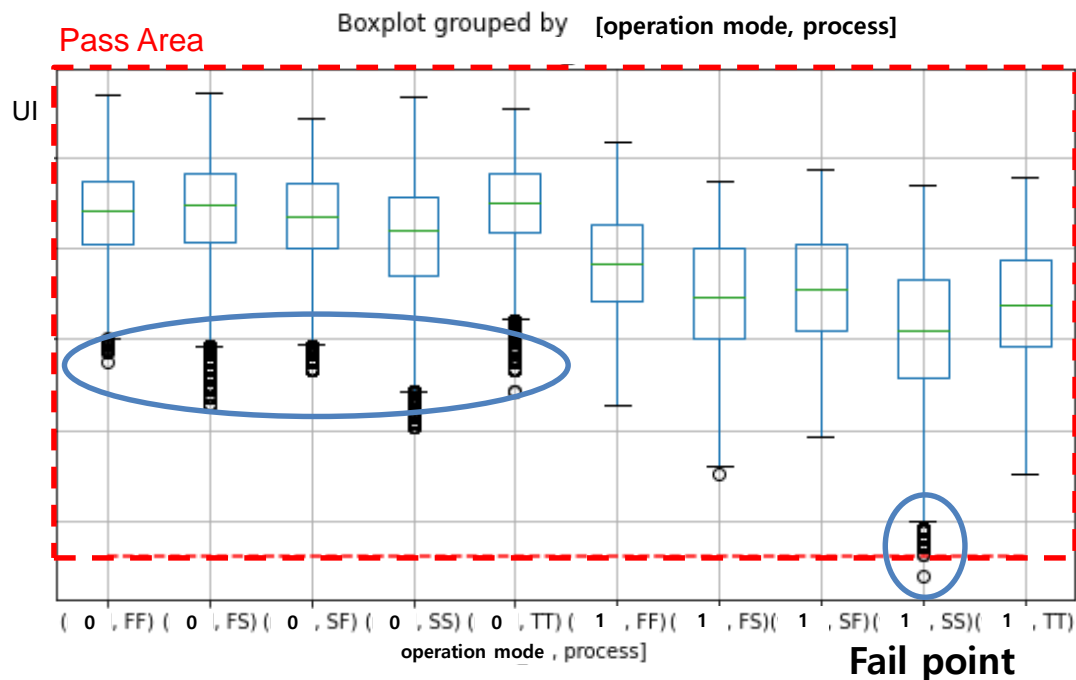
- 변인으로 작용할 수 있는, 2가지 이상의 항목을 갖는 Parameter 확인
- 이들의 의미와 종속성 조사 (기본적으로 Test는 완전요인 Test)
- Version 와 sn는 비독립적.

Parameter	(종류) 항목	의미
process	(5) FF, FS, SF, SS, TT	반도체 process corner
version	(2) Ver0, Ver1	SoC version
Operation	(2) operation_mode : 0, 1 (4) operation_name : A0, ..., A3	NAND operation
Temperature(°C)	(13) -30, -20, ..., 90	온도
driver_strength_1	(3) DS_1-1, DS_1-2, DS_1-3	Drive strength
driver_strength_2	(2) DS_onoff : on, off (6, 4) DS_value : on_1, ..., on_6, off_1, ..., off_4	
Input voltage	(7) in_vol_1, 2, 3~7	SoC 공급 전원
NAND speed	(4) speed_1, 2, 3, 4	동작 frequency
sn	(3) 1, 2, 3	sample numbering
iter	(2) 0, 1	실험 반복 횟수

Input Parameters and each values

1-2. Data Preprocessing- Outlier check

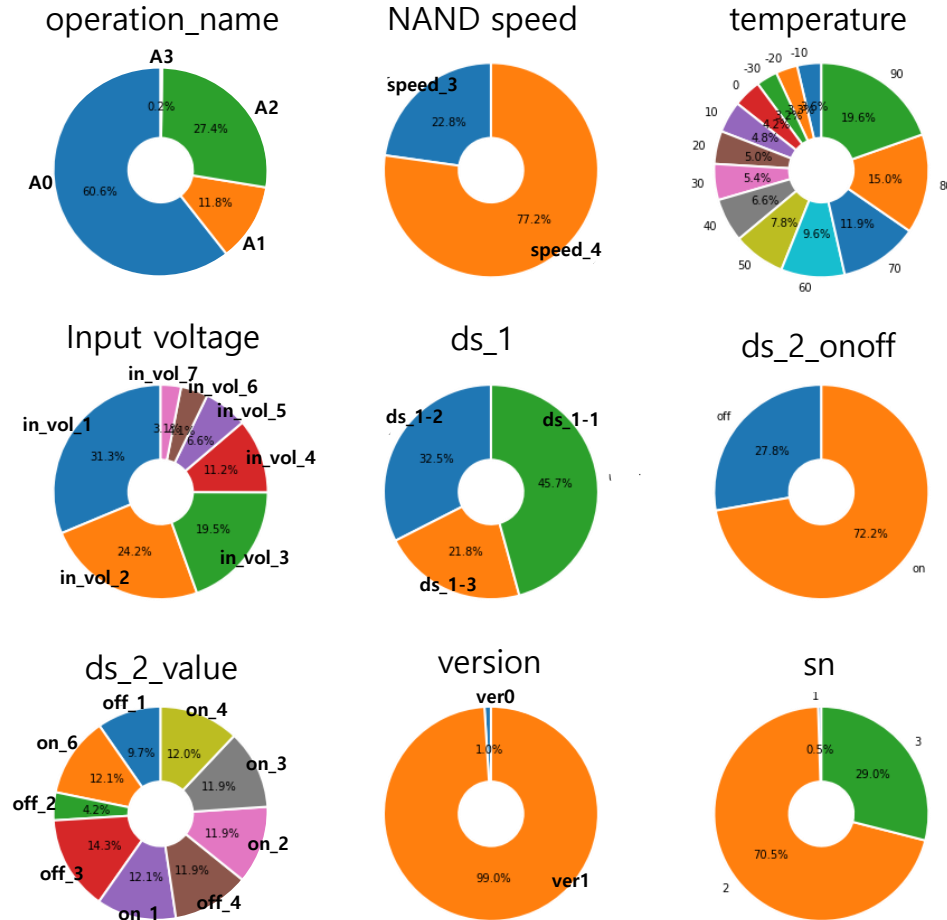
- Operation mode과 Process 별 Boxplot graph에서 Outlier points check 결과
- (1, SS)의 경우 최저점에 분포, Fail에 해당하는 outlier point 검출
- (1, SS)의 Outlier는 Input voltage=in_vol_1일 때 발생 (1개의 Fail point 포함)



Boxplot graph²⁾의 해석

1-2. Data Preprocessing- Outlier check

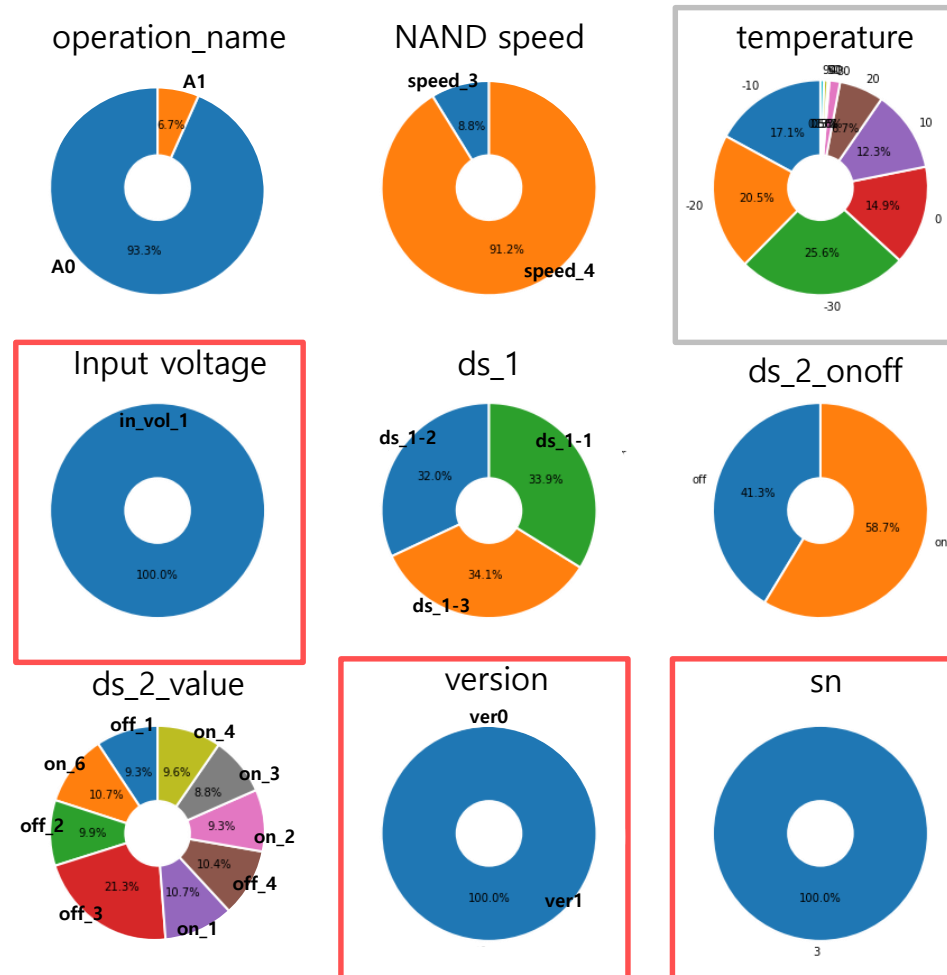
○ Operation mode = 0 의 경우 Outlier는 여러 parameter의 영향을 받음



(0,)의 Outlier points의 parameter 별 value 비율

1-2. Data Preprocessing- Outlier check

- Temperature 낮을수록 UI가 증가하나, outlier는 더 많이 발생 (편차가 커짐)
- (1, SS)의 Outlier는 Input voltage=in_vol_1, version=Ver1, sn=3의 SoC에서 발생

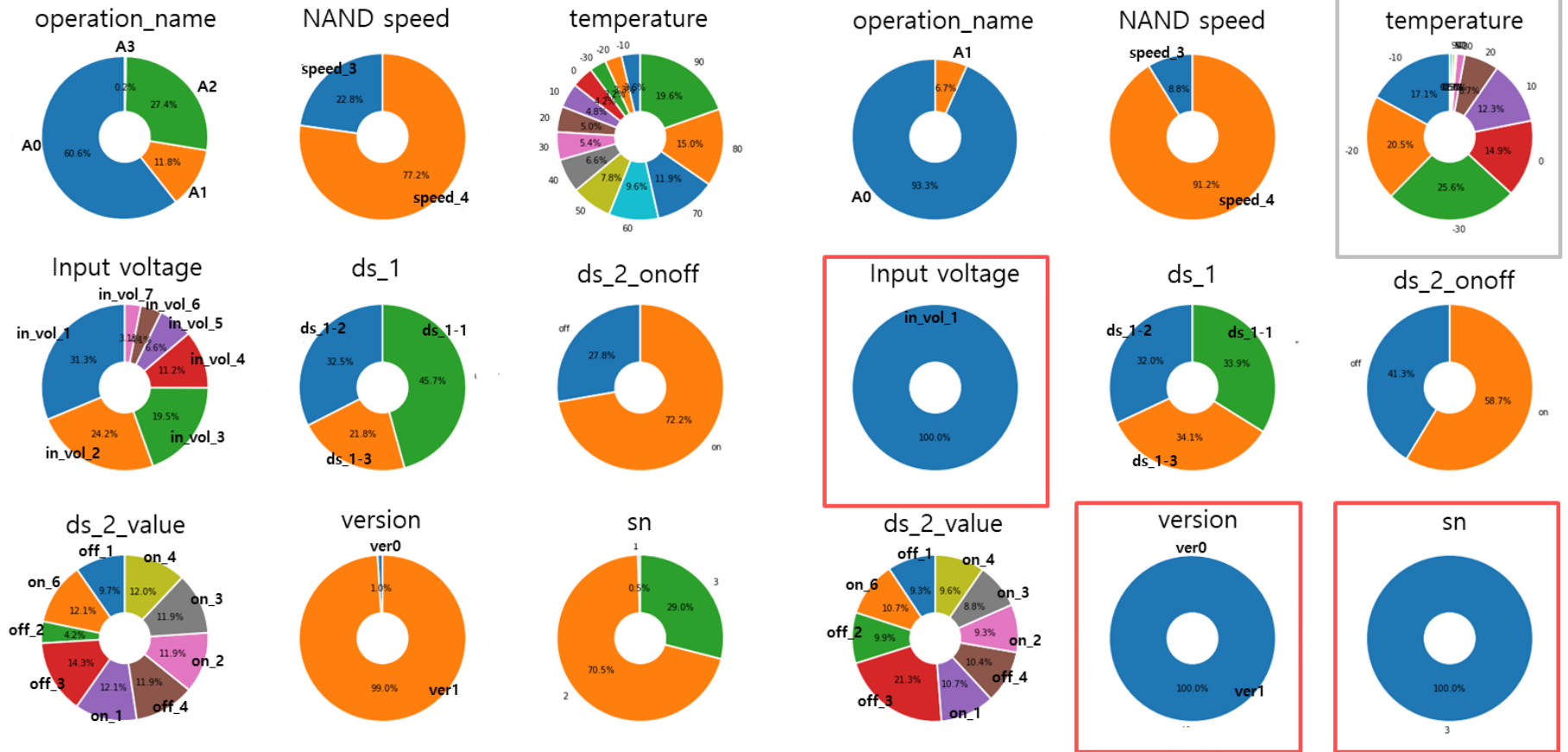


(1, SS)의 Outlier points의 parameter 별 value 비율

1-2. Data Preprocessing- Outlier check

○ SPEC은 Input voltage=in_vol_7으로 주어지므로 $\pm 10\%$ margin 고려하여

Input voltage=in_vol_1 < in_vol_7 $\times 0.9$ 인 dataset 제외 결정

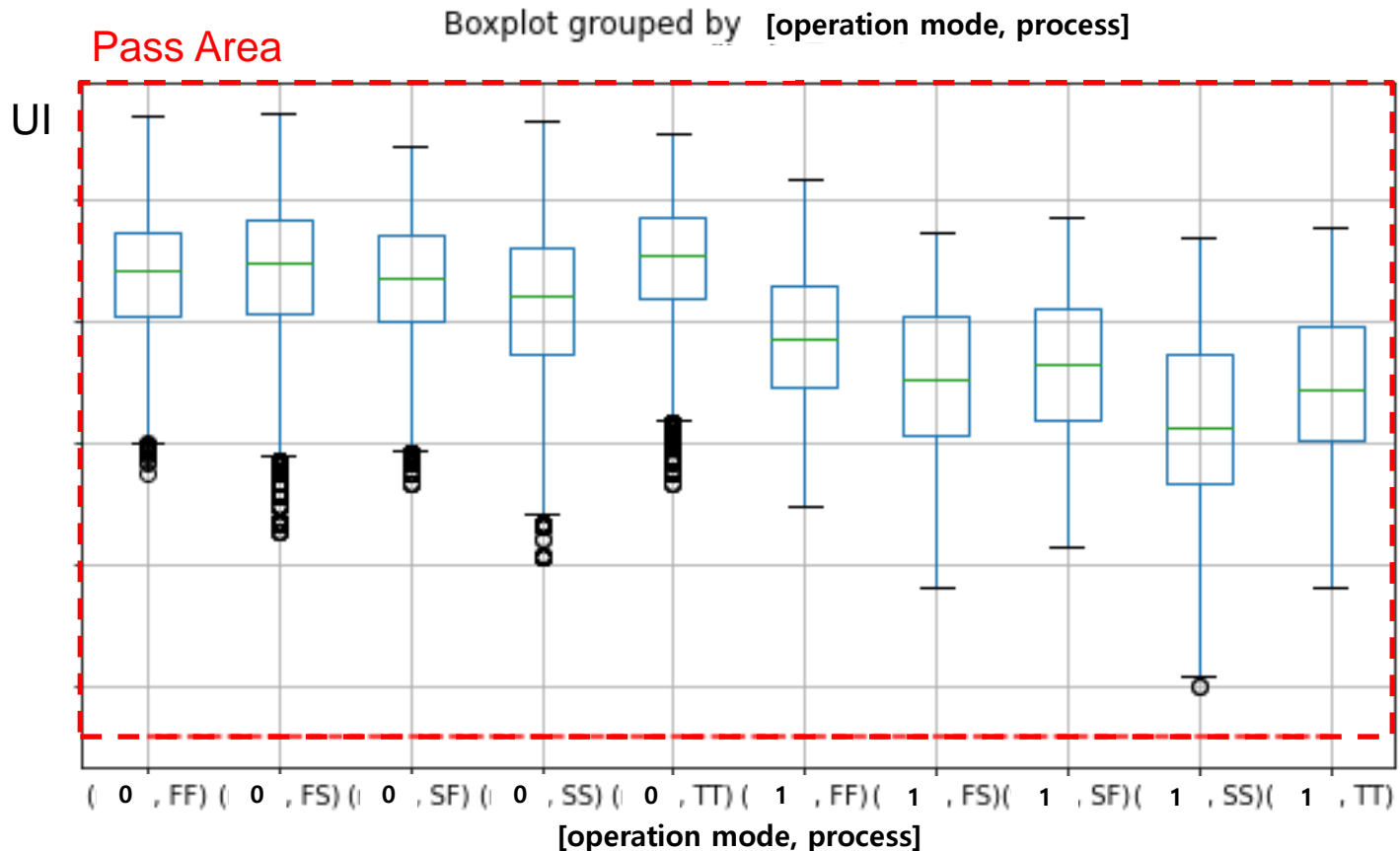


(0)의 Outlier points의 parameter 별 value 비율

(1, SS)의 Outlier points의 parameter 별 value 비율

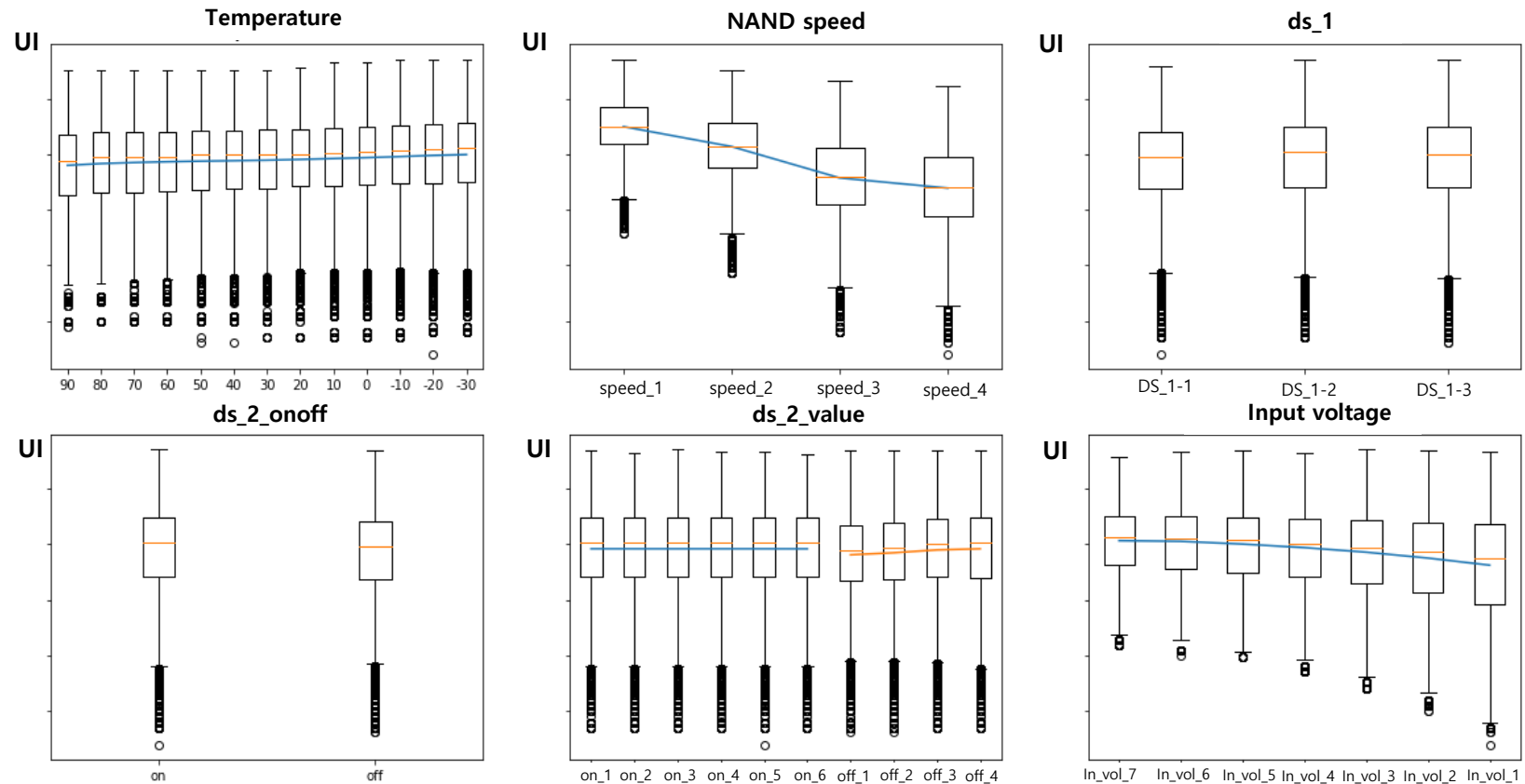
1-2. Data Preprocessing- Outlier check

- Input voltage=in_vol_1 인 dataset 제외한 결과 분포는 아래와 같음
- Outlier point 수는 줄어들고 Fail point 사라짐



1-3. Data Preprocessing – 경향성 파악

- 각 Parameter의 value에 따라 선형적인 UI 감소/증가 경향 존재
- 각 value 별 평균값들의 표준편차를 Feature selection의 지표로 사용 가능
- 중복되는 Parameter 통일, 2가지 Parameter 기록된 column은 2개로 분리



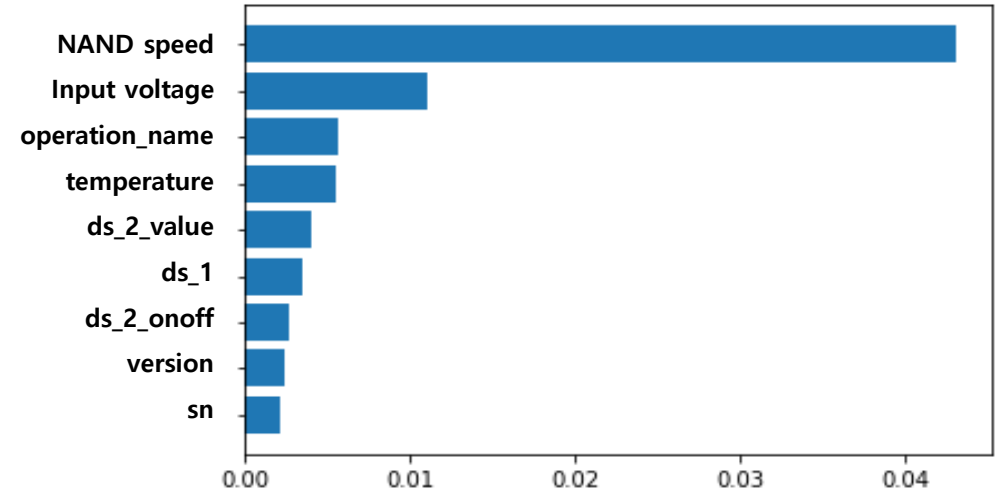
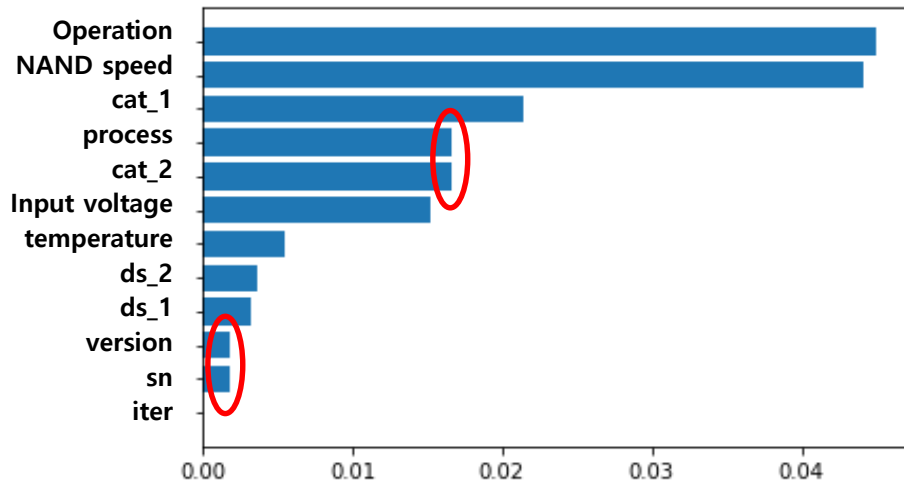
각 parameter의 value 별 평균값을 line graph로 그려보았을 때 선형적인 증감이 나타남

1-3. Data Preprocessing – Feature Selection

○ Model training에 필요한 parameter만을 선택하기 위하여 importance 지표 설정

① Value별 평균의 표준편차

① operation_mode > NAND speed > process > Input voltage > operation_name > temperature > ds_2_value > ds_1 > ds_2_onoff > version > sn

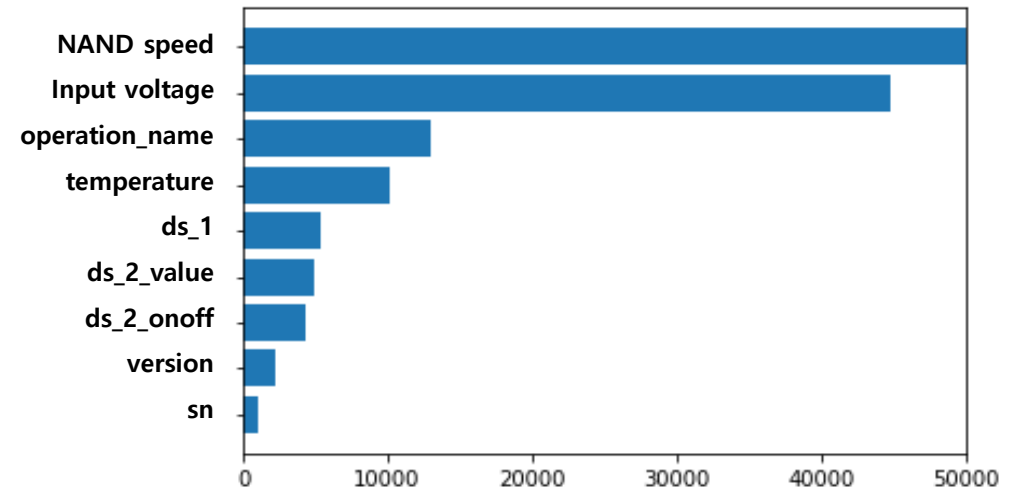
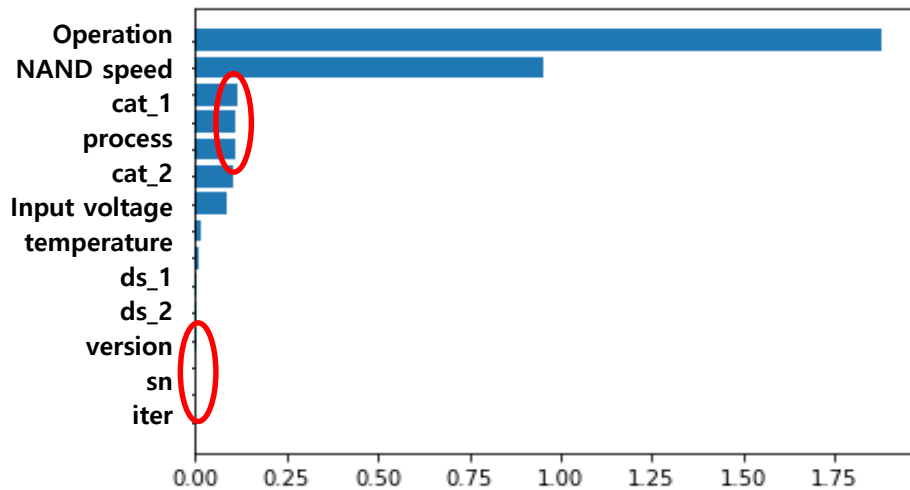


- process와 cat_2가 중복되는 parameter임과 version와 sn의 종속성 확인됨
- 이를 제외한 parameter들의 importance를 확인하였을 때, NAND speed의 영향이 가장 큼

1-3. Data Preprocessing – Feature Selection

- Model training에 필요한 parameter만을 선택하기 위하여 importance 지표 설정
- ② Scikit-learn의 F-test score

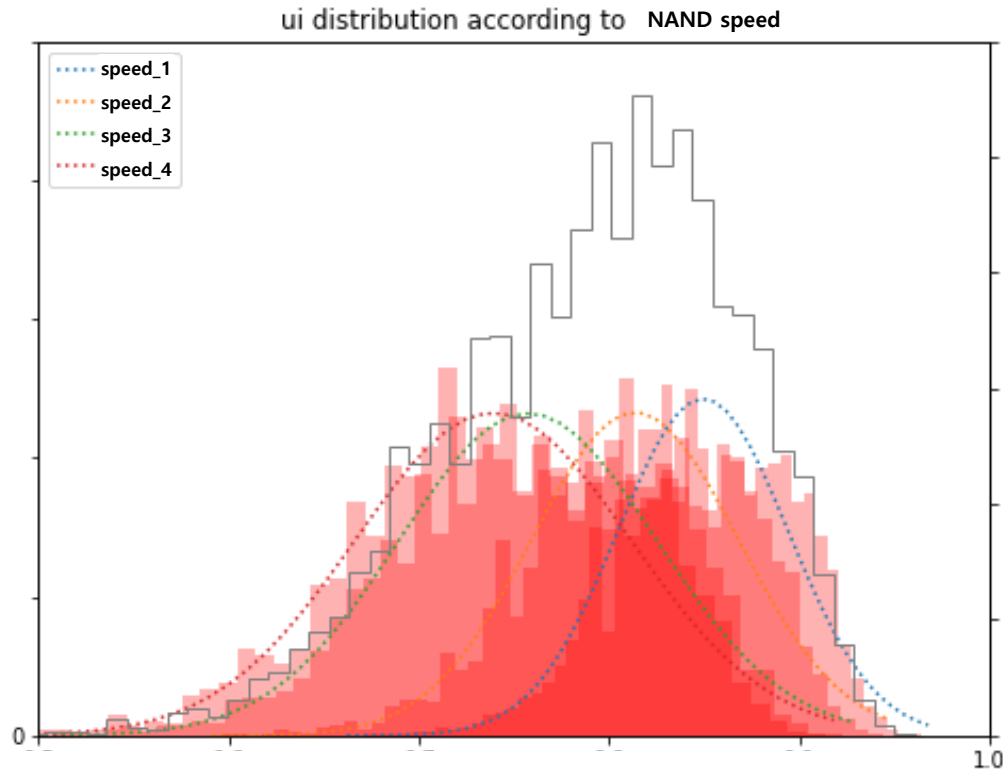
② operation_mode > NAND speed > process > Input voltage > operation_name > temperature > ds_1 > ds_2_value > ds_2_onoff > version > sn



- process와 cat_2의 중복은 확인되나 version와 sn의 종속성 확인 어려움
- 이를 제외한 parameter들의 importance를 확인하였을 때, NAND speed의 영향이 가장 큼
- ① Value별 평균의 표준편차 방법의 합리성 검증됨

1-3. Data Preprocessing – Feature Selection

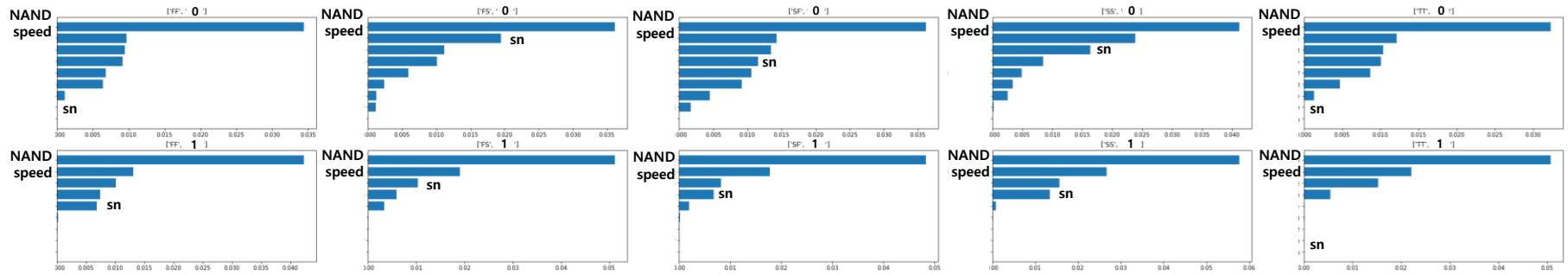
- Operation과 process별로 보았을 때 NAND speed가 UI에 가장 큰 영향을 주며, 각 NAND speed에 따라 히스토그램을 plot하였을 때 그 편차가 아래와 같이 나타남



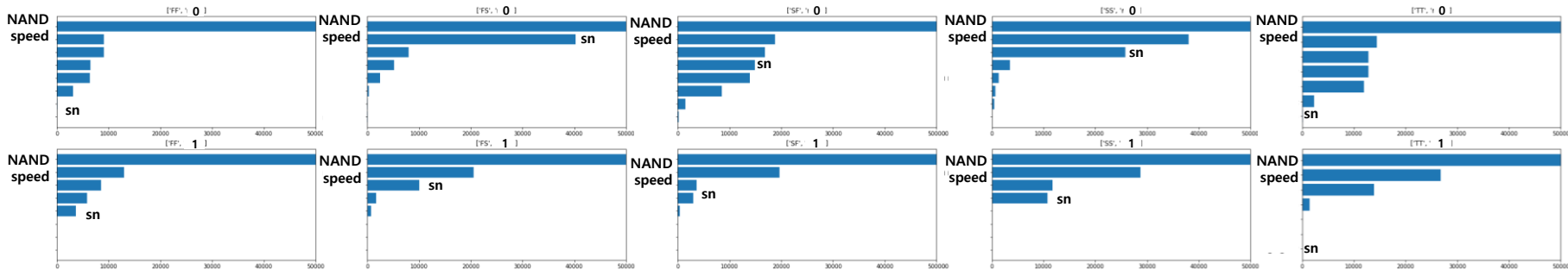
NAND speed의 각 value에 대한 UI 분포

1-3. Data Preprocessing – Feature Selection

- Operation mode와 Process 별 지표의 importance ranking이 전체 ranking과 다름
- (0, FS)의 경우 등에서는 SoC 샘플에 따라 편차가 크게 나타남



① Value별 평균의 표준편차

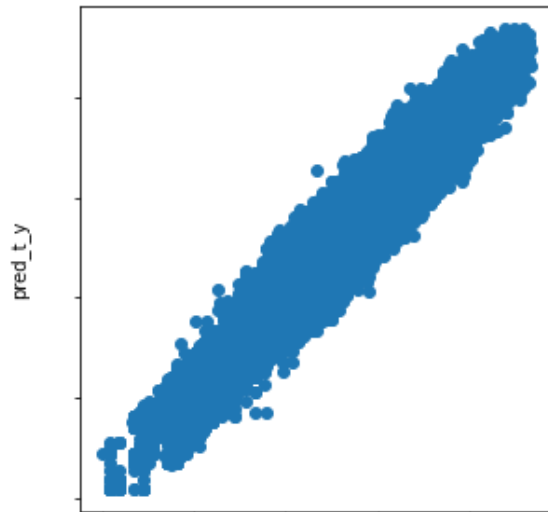


② Scikit-learn의 F-test score

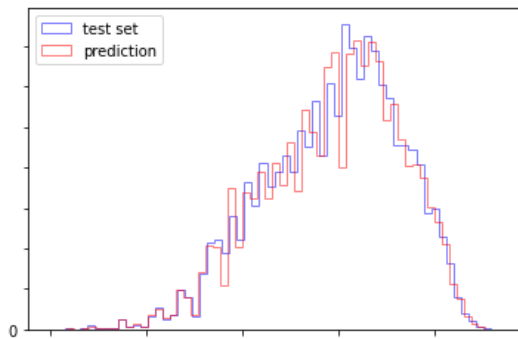
2-1. Model Fitting and Evaluation

- Decision tree, Random forest, XGBoost models (Metric : MAE)
- RF와 XGB는 hyper parameter tuning으로 optimize 가능

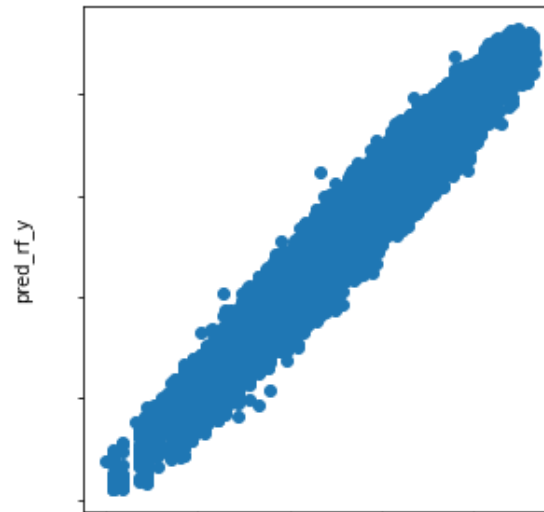
DT MAE : 0.00857



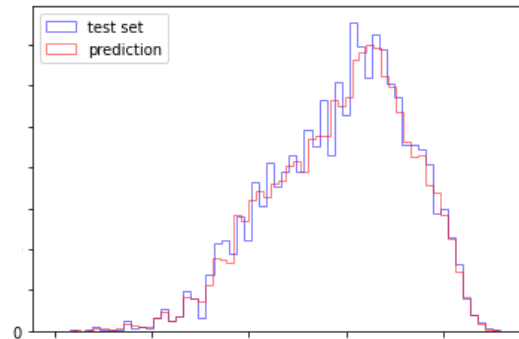
test_y



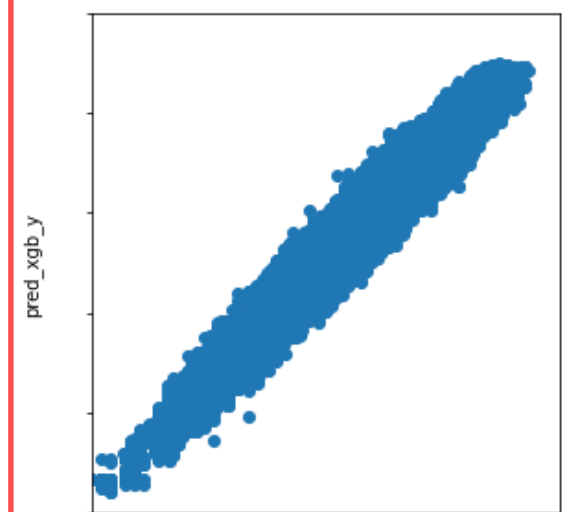
RF MAE : 0.00783



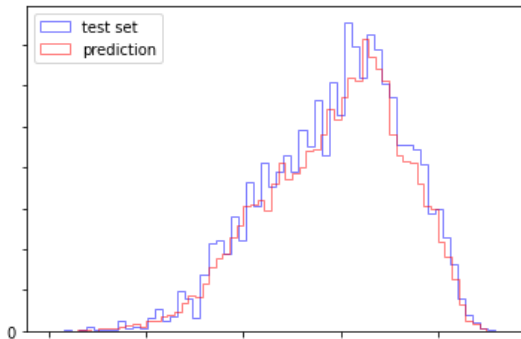
test_y



XGB MAE : 0.00939



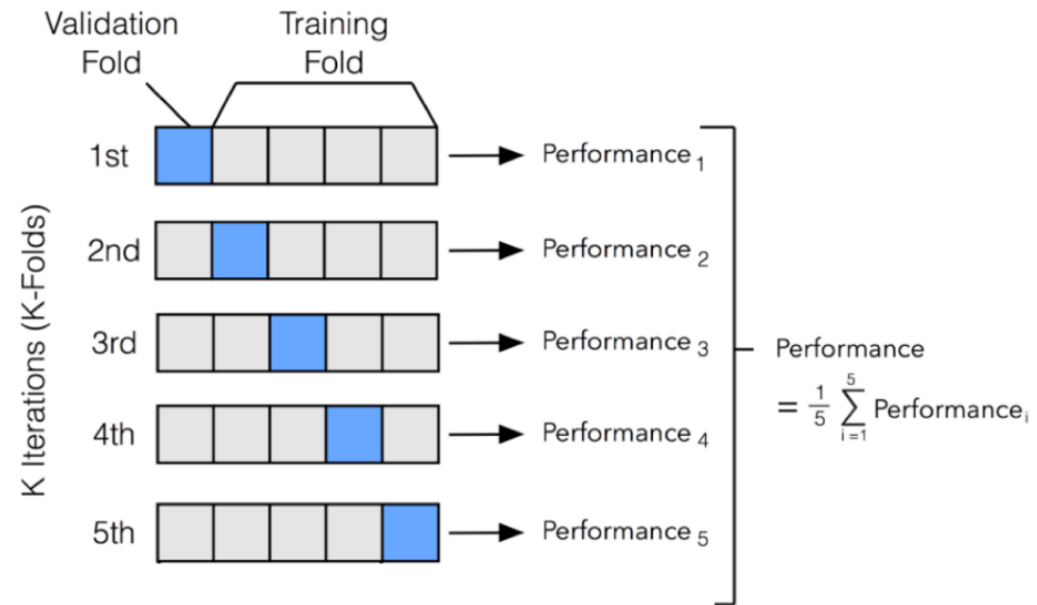
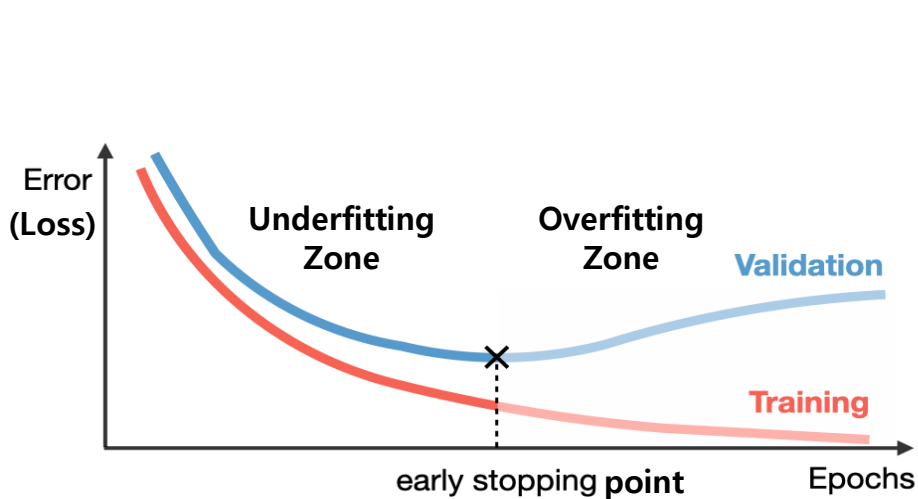
test_y



2-2. Model Fitting and Evaluation – Optimization methods

① Implementation of Early stopping, Cross validation : Overfitting 방지

- Error가 k번 이상 iterate 되어도 개선되지 않으면 training을 멈추도록 하는 것
- Error 계산 시 fold 수 만큼 계산하여 이들의 평균치를 metric으로 하는 것



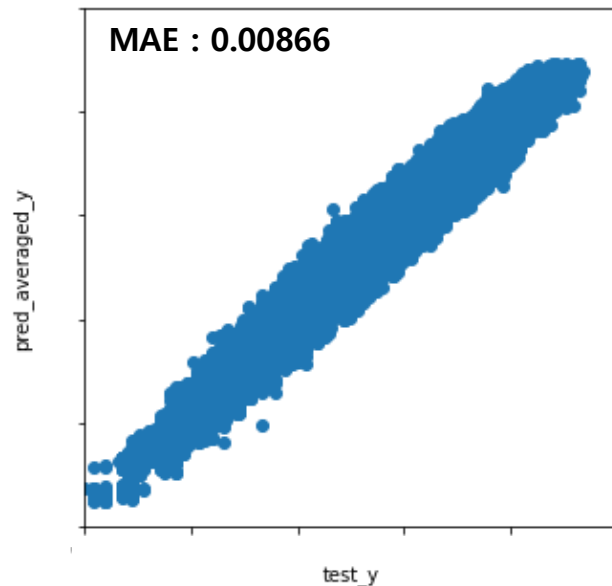
[2] "[Deep Learning Tips and Tricks cheatsheet](#)", Stanford Edu CS230-Deep Learning, Afshine Amidi and Shervine Amidi,, accessed 26.07.2021.

[3] "[Model Selection](#)", <http://ethen8181.github.io>, accessed 26.07.2021.

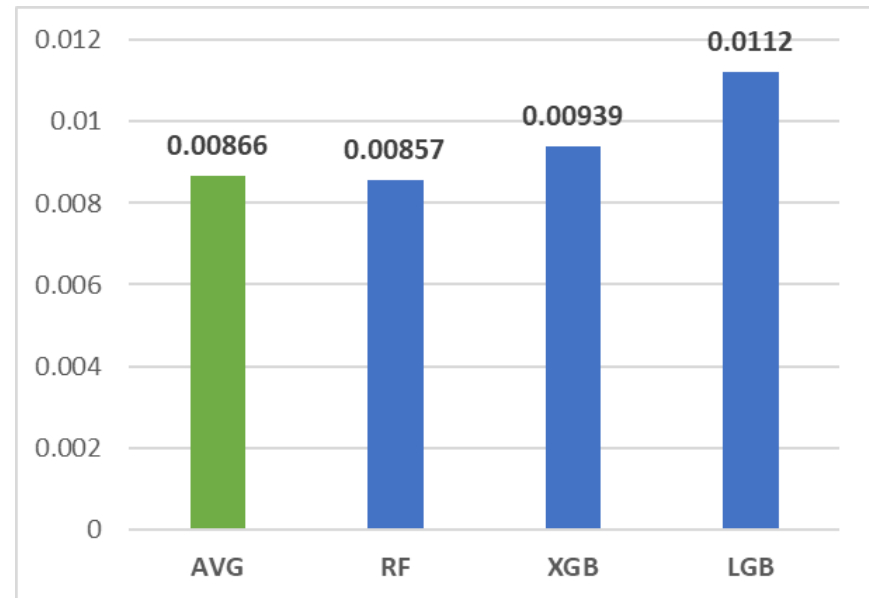
2-2. Model Fitting and Evaluation – Optimization methods

② Idea of using Averaged model

- RF, XGB, LGB model의 prediction의 평균치를 도출하는 AVG model은 세 model의 오류의 평균치보다 작은 오류를 가지는 synergy effect 확인됨



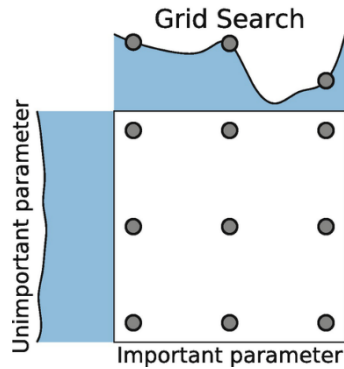
AVG model prediction



MAE error of Base models(RF, XGB, LGB) & Averaged model

2-3. Model Fitting and Evaluation – Hyper-parameter Tuning

- Grid search on epoch number, tree depth, learning rate
- RF model은 런타임이 매우 길어 XGB model에 대해서만 결과 확인
- 1st Best model과 2nd Best model의 Averaged model은 이들보다 작은 Error 가짐



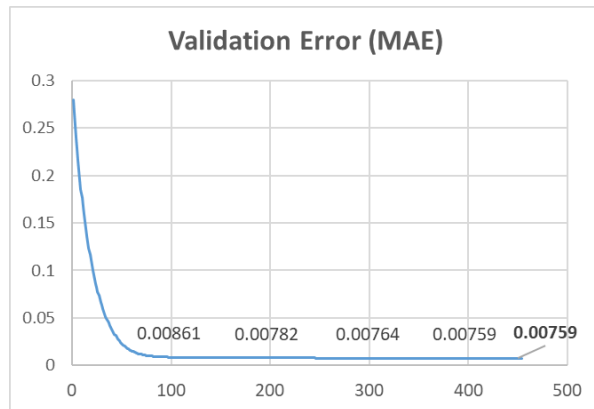
: 1st Best model's hyper parameter

```

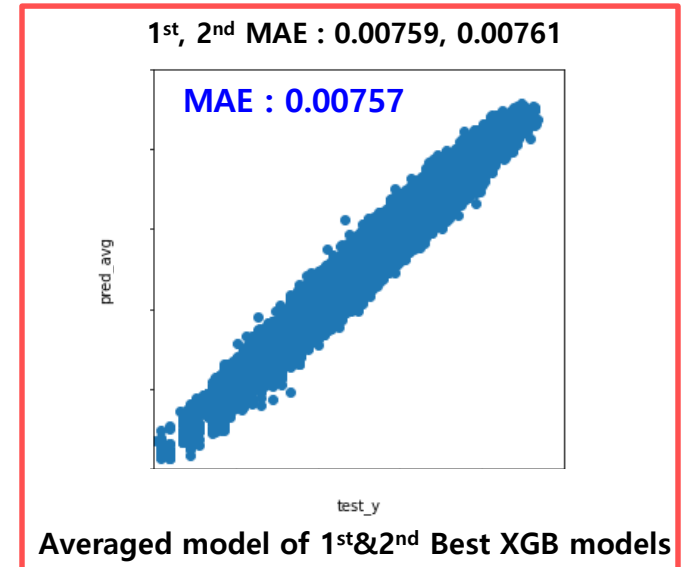
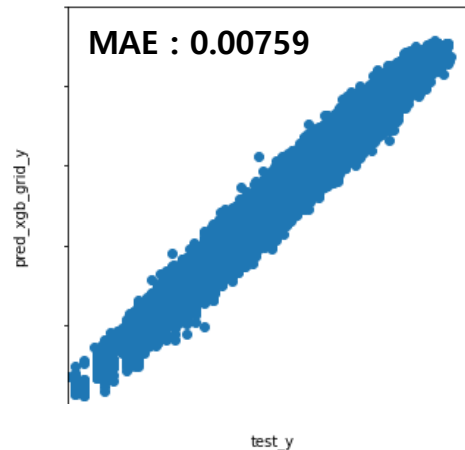
xgb_param_grid = {
  'n_estimators' : [1600, 2000, 2400, 3000],
  'learning_rate' : [0.02, 0.05, 0.1, 0.2],
  'max_depth' : [6, 8, 10, 12]
}
    
```

```

rf_param_grid={
  'max_depth':[6,8, 10, 12],
  'n_estimators':[100, 200, 500, 1000],
  'min_samples_leaf':[1,2],
  'min_samples_split':[2,3,4],
  'criterion':['mae']
}
    
```

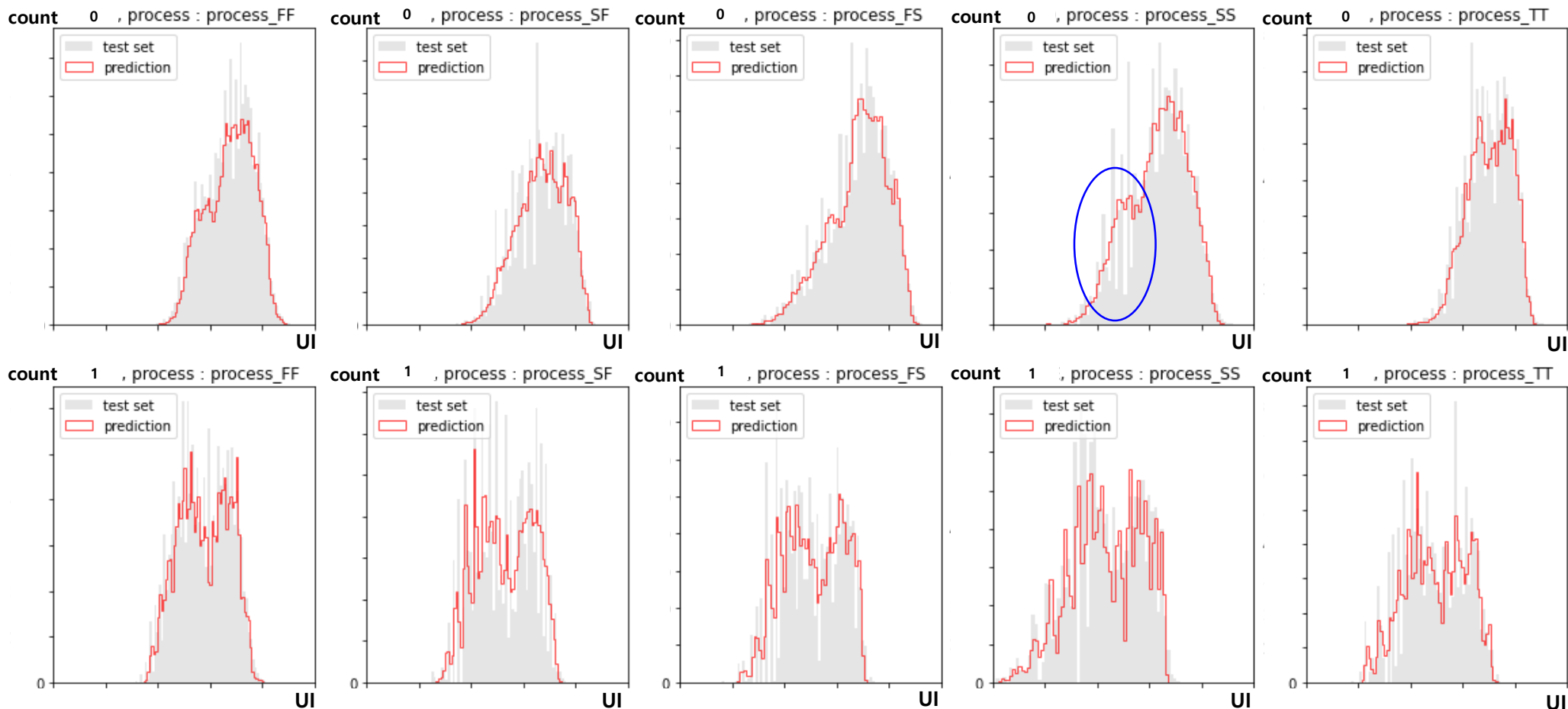


XGB grid search result : 1st Best model



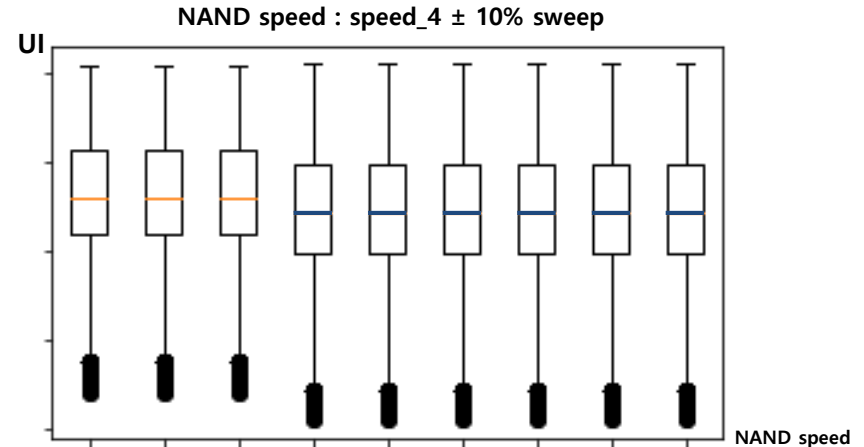
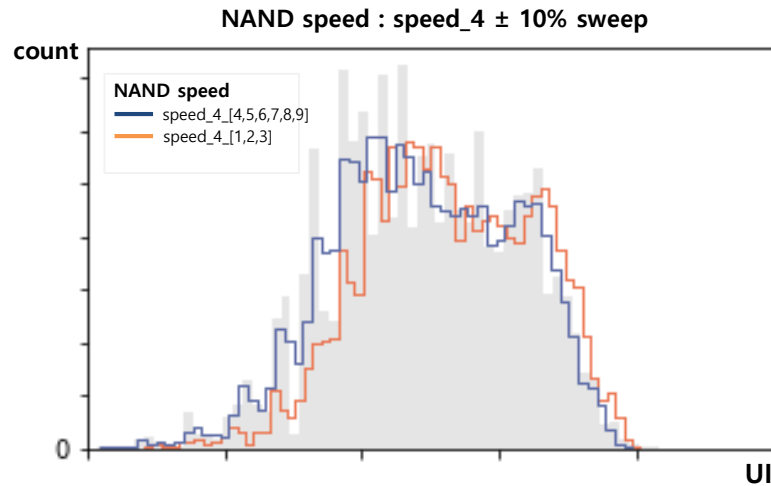
3-1. Utilizing Model Prediction – Input sweep에 따른 Margin check

- Operation mode와 Process 별 분포는 정규분포 가정 시 유의수준 99.2%이상 일치
- MAE error 0.006-0.008 미만이며 (0, SS)에서 최대 오차 발생 : 쌍봉 예측 오류



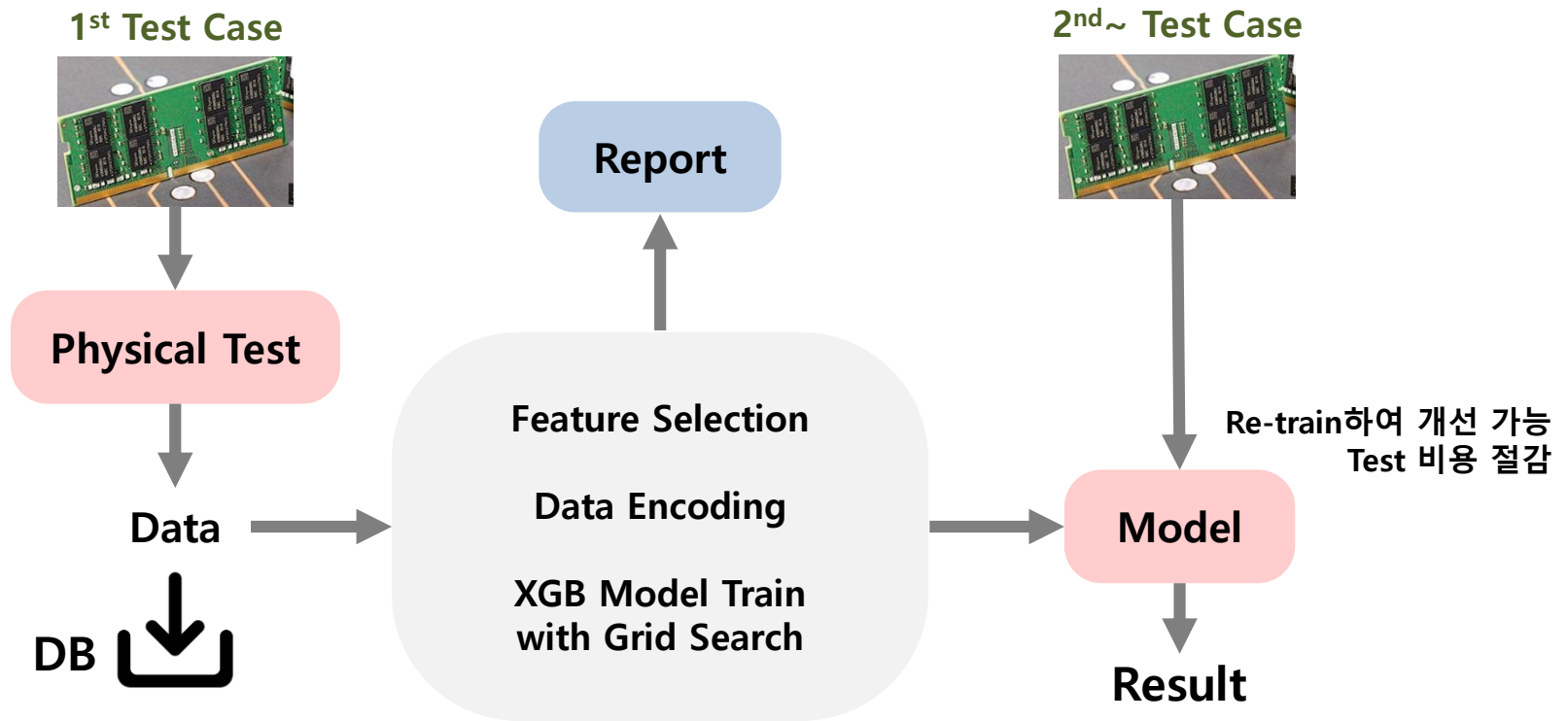
3-1. Utilizing Model Prediction – Input sweep에 따른 Margin check

- Datasheet에서 항목 별 표준편차가 MAE error보다 큰 parameter인 NAND speed에 대해 input을 $\pm 10\%$ sweep하였을 때 UI의 분포 확인
- NAND speed = speed_4에서 $\pm 10\%$ sweep하면 UI의 범위는 실험 결과와 같음
- ML model로서 실험을 대체하는 프로젝트 목표와 부합하는 결과 도출됨



3-2. Utilizing Model Prediction – Align with Automation Process

- Score return test에는 Regression model, Pass/Fail test는 Classification model 적용 가능하므로, Test result datasheet 및 test 종류 입력 시 ML model을 자동 생성
- NAND 포함 다양한 제품의 Validation process automation의 한 과정이 될 수 있음.



Validation의 ML model 생성 Automation 파이프라인

End of Document