

●○ 전력 설비 에너지 품질 과제

전력 설비 에너지 품질



●○ 개요: 전력설비 에너지 품질 데이터셋이란?

에너지 AI Solution 개발 확산을 목표로 스마트 그리드, 스마트 팩토리용 AI 데이터 구축을 통한 장비별 에너지 효율화·설비별 이상 감지 서비스 개발을 위한 전력 설비 에너지 품질 AI 학습용 데이터를 구축하였다. 설비별 35종 전력 품질 원시 데이터를 10종 산업 대표 설비로 분류하고, 설비 정보 메타데이터, 업체 정보 메타데이터와 병합한 '에너지 사용패턴' 학습용 AI 데이터 (세부 10종) 64.8만건, '설비 SOH 진단' 학습용 AI 데이터 (세부 10종) 97.2만건, 총 162만건의 데이터로 구성되어 있다.

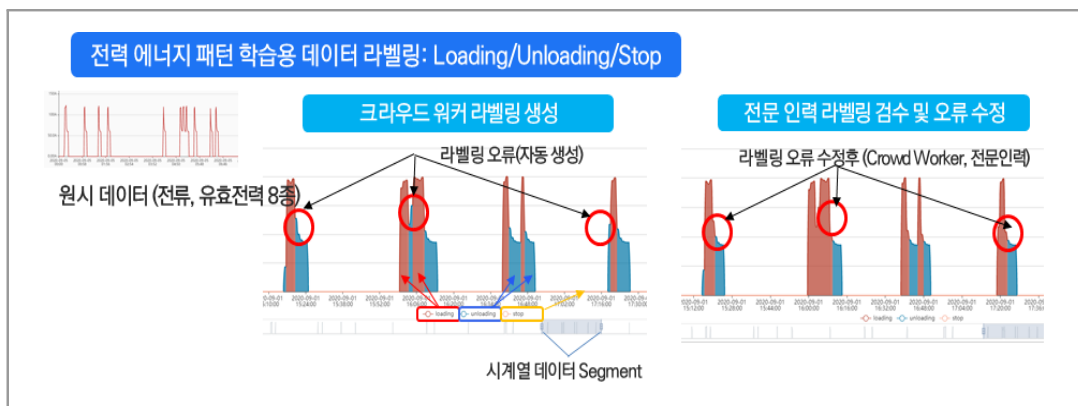
데이터 이름	전력 설비 에너지 품질 AI데이터
활용 분야	스마트그리드 및 스마트팩토리 분야 에너지 효율화 및 설비 이상 진단
데이터 요약	설비별 35종 전력 품질 원시 데이터를 10종 산업 대표 설비로 분류하고, 설비 정보 메타데이터, 업체 정보 메타데이터와 병합한 '에너지 사용패턴'과 '설비 SOH 진단' 학습용 AI 데이터 세부 2종 162만건 구축
데이터 출처	35종 전력 품질 원시 데이터는 컨소시엄에서 제공하는 Mobile Energy Meter를 이용하여 설비별로 직접 수집 및 가공

●○ 데이터셋의 구성

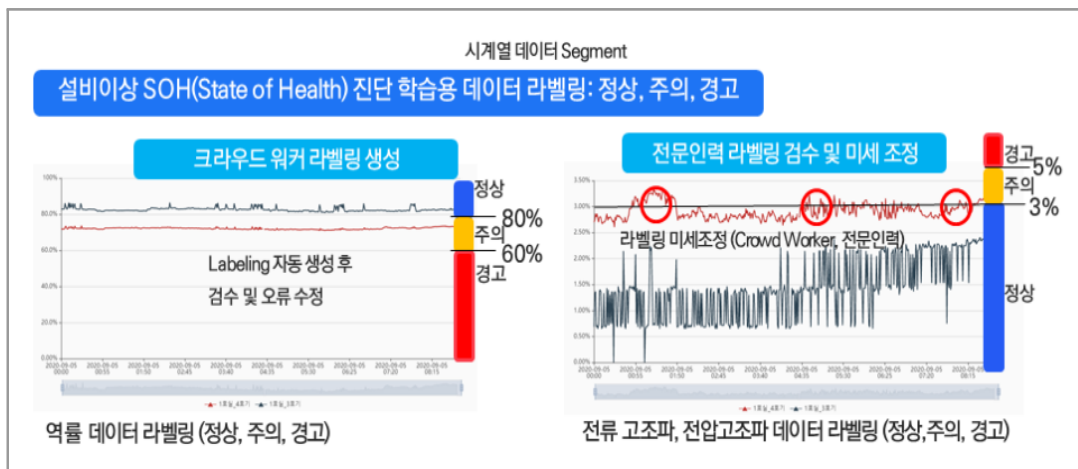
본 데이터셋은 2종의 AI 학습 데이터 (전력 에너지 패턴 학습용 데이터, 설비 이상 SOH(State of Health) 진단 학습용 데이터)로 포함하며, 전력 설비 에너지 품질 AI 학습을 위한 5개 항목에 대해 계측된 162만건의 데이터와 AI 알고리즘의 Constraint 설정을 위한 메타 데이터로 설비정보 2700건과 업체정보 4050건으로 구성되어 있다.

AI 학습 데이터는 원시데이터와 라벨링 데이터를 포함하는 json 포맷으로 구성되어 있다.

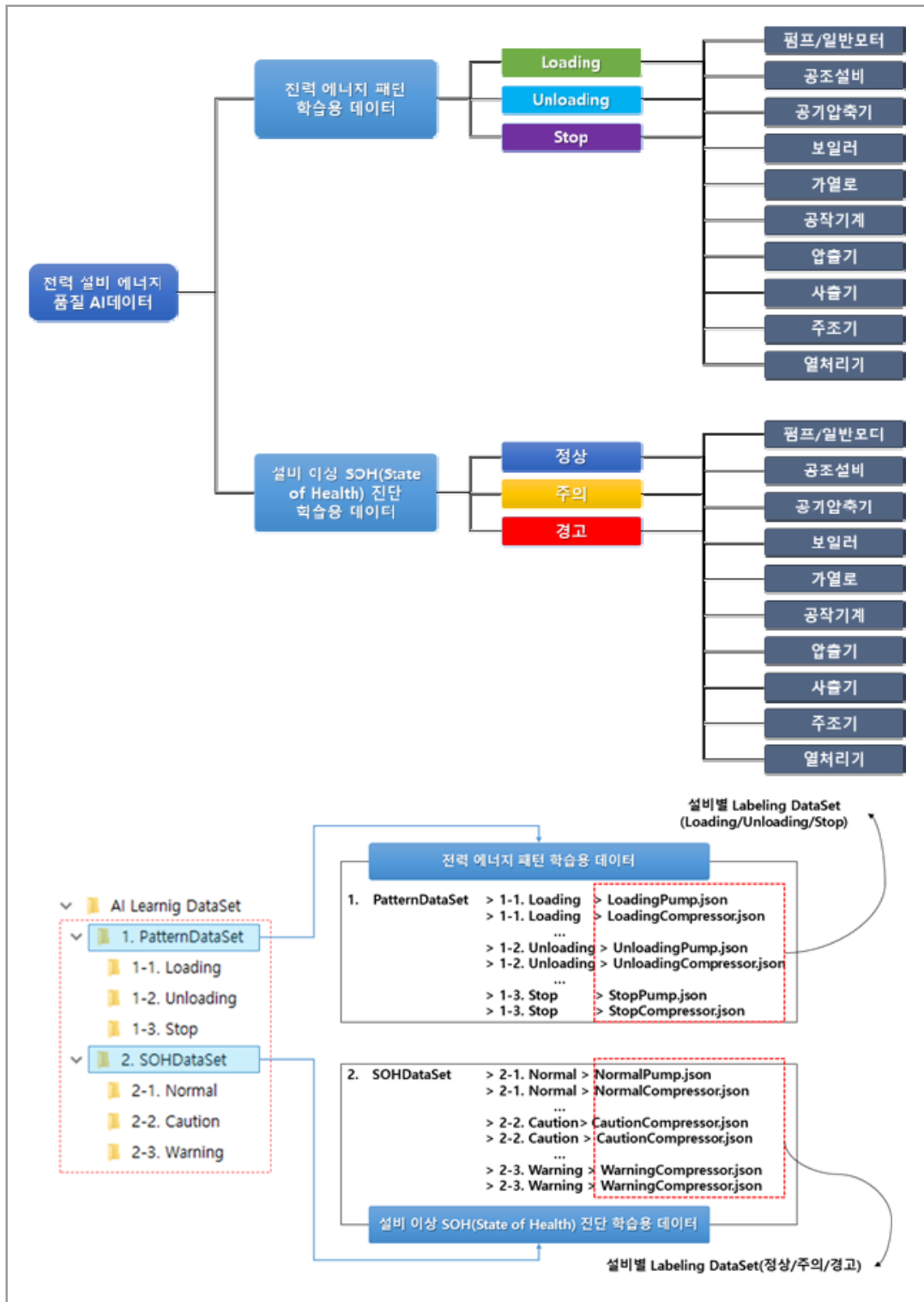
데이터 종류	포함 내용	데이터 구축량	제공 방식
전력 설비 에너지 품질 AI 학습 데이터 (5계측항목)	전력 에너지 패턴 학습용(10종), 설비 SOH 진단 학습용(10종) AI 알고리즘 개발을 위한 시계열 데이터 및 라벨링 데이터	162만건 (450대x24시간x30일x5계측항목)	JSON 포맷 파일
설비 정보 (6종)	AI 알고리즘의 Constraint 설정을 위한 메타 데이터	2,700건 (450대x6종)	JSON 포맷 파일
업체 정보 (9종)		4,050건 (450대*9종,가공)	JSON 포맷 파일



[전력 에너지 패턴 학습용 데이터]



[설비 이상 SOH(State of Health) 진단 학습용 데이터]



●○ 데이터셋의 설계 기준과 분포

1. 데이터 설계 기준 및 구축 규모

- 최종 인공지능 데이터: 2종 AI 서비스에 대한 AI 데이터 셋 162만건
 - 에너지 사용 패턴 AI 데이터셋(전류 평균, 유효전력 평균): 32.4만건 × 2 = 64.8만건
 - 450대설비 × 24시간 × 30일 = 32.4만건 (전류 평균)
 - 450대설비 × 24시간 × 30일 = 32.4만건 (유효전력 평균)
- 각 데이터셋은 1분 단위로 수집한 전력 품질 원시 데이터를 1시간 단위로 Segmentation 하여 전류 평균과 유효전력 평균 각각에 대한 에너지 사용패턴 라벨링 (Loading, Unloading, Stop 3종) 수행후 설비별 메타데이터(6종), 업체별 메타데이터(9종)를 포함한 데이터 셋으로 구성
 - 설비 SOH 진단 AI 데이터셋: 32.4만건 × 3 = 97.2만건
 - 450대설비 × 24시간 × 30일 = 32.4만건 (역률 평균)
 - 450대설비 × 24시간 × 30일 = 32.4만건 (전압고조파 평균)
 - 450대설비 × 24시간 × 30일 = 32.4만건 (전류고조파 평균)
- 각 데이터셋은 1분 단위로 수집한 전력 품질 원시 데이터를 1시간 단위로 Segmentation 하여 역률 평균, 전압고조파 평균, 전류고조파 평균 각각에 대해 설비별 SOH 진단 라벨링 (정상, 주의, 경고 3종) 수행후 설비별 메타데이터(6종), 업체별 메타데이터(9종)를 포함한 데이터 셋으로 구성
 - 메타 데이터: 설비별 메타데이터(6종) 2,700건, 업체별 메타데이터(9종) 4,050건
 - 설비별 메타데이터: 450대설비 × 6종
 - 업체별 메타데이터: 450대설비 × 9종

표 | 구축 AI 학습 데이터(요약)

원천 데이터	설비 종류 분류 (각 10종)		AI 서비스 (2종)	라벨링 대상	라벨링	구축 데이터량 (AI 데이터 셋)
35종 전력 품질 데이터 (450대 설비)	펌프/일반모터	x (곱)	에너지 사용패턴 분석 서비스	전류 평균, 유효전력 평균	Loading, Unloading, Stop	648,000건 (450대*24시간*30일* 2계측항목)
	공조설비					
	공기압축기					
	보일러					
	가열로					
	공작기계		설비 SOH 스코어 산출 서비스	역률 평균, 전압고조파 평균, 전류고조파 평균	정상, 주의, 경고	972,000건 (450대*24시간*30일* 3계측항목)
	압출기					
	사출기					
	주조기					
	열처리					

표 | 구축 데이터(요약)

데이터 셋	필요성 및 속성	데이터 수집 방법	데이터 구축량
(데이터#1) 전력 설비 에너지 품질 AI 학습 데이터 (5계측항목)	전력 에너지 패턴 학습용(10종), 설비 SOH 진단 학습용(10종) AI 알고리즘 개발을 위한 시계열 데이터 및 라벨링 데이터	Mobile Energy Meter 클라우드 소싱 방식 원시 데이터 수집 클라우드 소싱 방식 라벨링	162만건 (450대x24시간x30일x5계측항목)
(데이터#2) 설비 정보 (6종)	AI 알고리즘의 Constraint 설정을 위한 메타 데이터	실사를 통한 직접 수집	2,700건 (450대x6종)
(데이터#3) 업체 정보 (9종)		실사를 통한 직접 수집	4,050건 (450대*9종,가공)

2. 설비 대표성 및 특수성

- 10종 설비 선정은 사전에 전문가 검토를 받아 선정
 - 대표설비 10종은 에너지 다소비 설비 6종과 범용 설비 4종으로 구성
 - 에너지 다소비 설비는 공장내 에너지 다소비 범용 설비(보일러, 가열로) 업종에 따라 대표성 있는 설비(압출기, 사출기, 주조기, 열처리)를 계측 대상으로 선정하였음. 현재 선정된 계측 대상 설비는 이에 대한 체계적인 통계가 없어 수행기관 및 참여기업의 EMS 구축 사업의 경험을 토대로 선정하였으며, 설비 대표성에 대해서는 사업기간 중 전문가 자문을 통한 검토 및 대상 설비 조정 가능 (사업기간내 총 2회에 걸쳐 전문가 검토예정. (1차: 150개 설비에 대한 설비 메타데이터 수집 결과 검토, 2차: 300개 설비에 대한 설비 메타데이터 수집 결과 검토. 검토 내용에는 사전 검토 결과에 조사된 산업용 기계분류 중 공조설비의 비율이 높은 것에 대한 타당성 조사 포함)
 - 범용 설비는 산업용 설비 통계를 바탕으로 선정되었으며, 설비 대표성에 대해서는 사업기간 중 전문가 자문을 통한 검토 및 대상 설비 조정 가능
 - 본 사업에서 선정된 10개 설비군은 업종에 따른 특수 설비를 제외한 공장 내 대부분의 설비를 포함하는 분류체계임
 - 원시 데이터 확보 및 가공 후 설비별 업종별 특이성은 별도의 연구로 진행할 예정

분류	계측 대상 설비
에너지 다소비 설비 (6종)	보일러, 가열로, 압출기, 사출기, 주조기, 열처리
범용 설비 (4종)	펌프/일반모터, 공조설비, 공기압축기, 공작기계

3. 데이터 분포

- 450개의 설비에 대한 전력 품질 데이터를 산업 10종 설비 분류에 대한 구축 예상 분포는 하기와 같음
 - 산업용 대표 설비 10종 분류와 AI 데이터셋 2종(전력 에너지 패턴, 설비 SOH 진단)으로 총 20종의 AI 데이터 셋 생성
 - 설비 분포는 10종으로 분류하여 각각 AI 데이터셋 구성

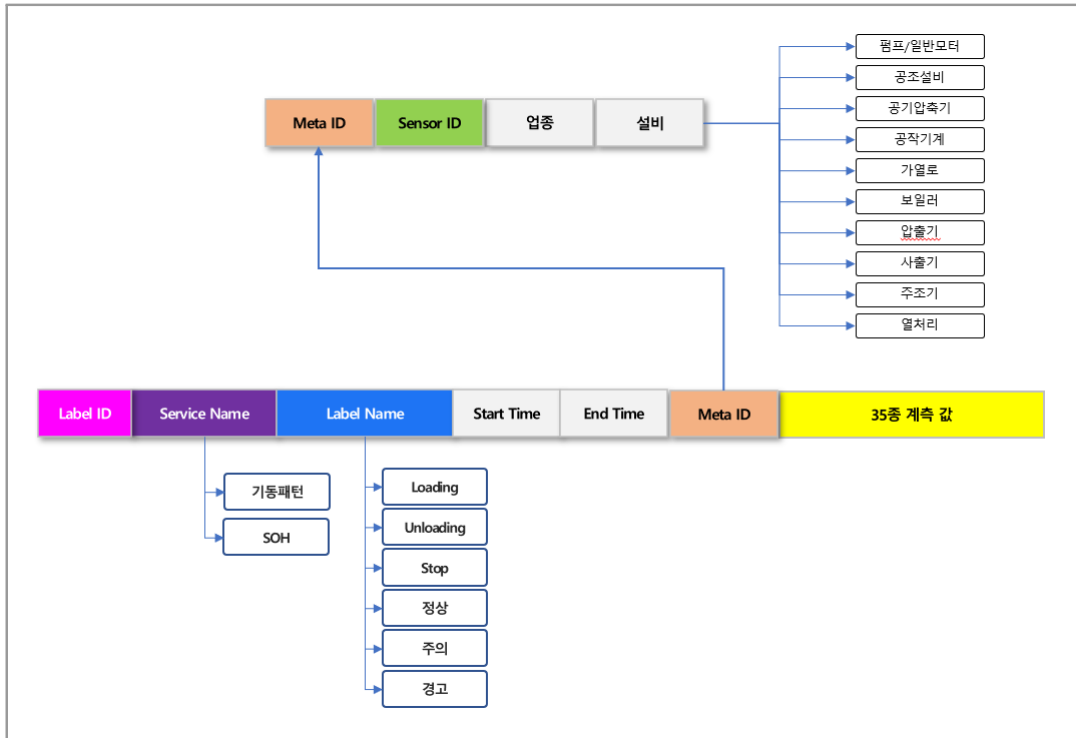
구축 데이터	설비 종류(각 10종)	구축 설비수(예상)	라벨
전력 에너지 패턴 학습용 데이터	펌프/일반모터	50	Loading, Unloading, Stop
	공조설비	60	
	공기압축기	120	
	보일러	10	
	가열로	50	
	공작기계	100	
	압출기	10	
	사출기	10	
	주조기	20	
	열처리	20	
설비 SOH (State of Health) 진단 학습용 데이터	펌프/일반모터	50	정상, 주의, 경고
	공조설비	60	
	공기압축기	120	
	보일러	10	
	가열로	50	
	공작기계	100	
	압출기	10	
	사출기	10	
	주조기	20	
	열처리	20	
20종 (2종 AI 서비스 x 10종 설비 분류) 전력 품질 AI 학습 데이터 구축			

- 450개의 설비에 대한 전력 품질 데이터를 산업 10종 설비 분류에 대한 구축 예상 분포는 하기와 같음

구축 데이터	라벨	예상 분포
전력 에너지 패턴 AI 데이터	Loading	60
	Unloading	30
	Stop	10
설비 SOH 진단 AI 데이터	정상	70
	주의	20
	경고	10
20종 (2종 AI 서비스 x 10종 설비 분류) 전력 품질 AI 학습 데이터 구축		

●○ 데이터 구조

데이터셋에 따른 항목과 해당 값은 아래와 같다.



[AI Data 포맷: 원시데이터, 라벨링, 메타데이터 통합 구조]

표 | 설비 정보

	데이터 중분류	데이터 소분류	내 용	타입	필수여부	비고
1	설비별	설비명	설비의 이름	Text		선택 공개 (일반,특수)
2		설비종류	설비의 종류	Text	Y	공개
3		제조사	설비 제조사	Text		선택 공개 (일반,특수)
4		제작연도	설비 제작연도 또는 설치 연도	Value(Int)	Y	공개
5		정격용량	설비가 사용하는 최대 전력 정격	Value(Int)	Y	공개
6		정격전압	설비가 사용하는 전압 정격	Value(Int)	Y	공개

표 | 업체 정보

	데이터 중분류	데이터 소분류	내 용	유형	필수여부	비고
1	업체별	업체명	설비가 설치된 업체명, 공장, 빌딩명	Text		비공개
2		설립연도	업체, 공장, 빌딩 설립 연도	Value(Int)	Y	공개
3		주소	업체, 공장, 빌딩 주소	Text		일부 공개(지역)
4		업종	업체, 공장, 빌딩의 업종	Text	Y	공개
5		업태	업체, 공장, 빌딩의 업종	Text	Y	공개
6		주생산물	업체, 공장, 빌딩의 주 생산물	Text		선택 공개 (일반, 특수)
7		계약전력	업체, 공장, 빌딩의 수전 용량에 따른 한전과의 계약 전력	Value(Int)	Y	공개
8		선택요금 제	업체, 공장, 빌딩이 한전과 계약한 선택 요금제 형태	Text	Y	공개
9		한전 정보	한전 과금 정보 조회용 업체별 정보	Text		비공개

●○ 데이터 예시

이 데이터는 메타 데이터와 AI 학습용 전력 패턴데이터 중 “Loading”으로 라벨링된 데이터의 예시이다.

```
{
  "META_ID": 210,
  "sensor_id": 2734,
  "dsid": 2291,
  "distbdid": 1,
  "dsmtrser": "985DAD57E220",
  "sensor_name": "985DAD57E220-1",
  "sensor_descr": "",
  "company_id": 263,
  "company_name": "데폴로레스",
  "C_TYPE": "",
  "MAJOR_PRODUCT": "",
  "valid_start": "2020-09-16",
  "valid_end": "2020-12-31",
  "facility_name": "일신 인발기 메인 판넬",
}
```



```

    "facility_type": 4,
    "facility_type_name": "공작기계",
    "facility_vendor": "신기산업",
    "facility_age": 220,
    "facility_capacity": 0,
    "facility_volt": 132
  }
  {
    "LABEL_ID": 173726,
    "SVC_NAME": "기동패턴",
    "DSITEM_NAME": "전류평균",
    "LABEL_NAME": "Loading",
    "RAW_DATA_FROM": "2020-11-16 15:55:58",
    "RAW_DATA_TO": "2020-11-16 17:04:58",
    "META_ID": 210,
    "RAW_DATA": [
      {
        "item_id": 3,
        "item_val": 232884.453000,
        "gen_time": "2020-11-16 15:55:58"
      },
      {
        "item_id": 3,
        "item_val": 232943.484000,
        "gen_time": "2020-11-16 15:56:58"
      },
      ...
      {
        "item_id": 37,
        "item_val": 236876.359000,
        "gen_time": "2020-11-16 17:04:58"
      }
    ]
  }
}

```

●○ 데이터 구축 과정

전력 품질 AI 학습 데이터(20종) 구축은 Mobile Energy Meter를 이용한 원시데이터 수집(설비별 전력 에너지 및 전력 품질 데이터) 및 설비 메타데이터 및 업체 메타데이터 수집 과정과 데이터 분류 및 가공을 통한 라벨링 그리고 검수 과정을 포함하며, 아래와 같은 데이터 수집 및 가공 프로세스를 거침

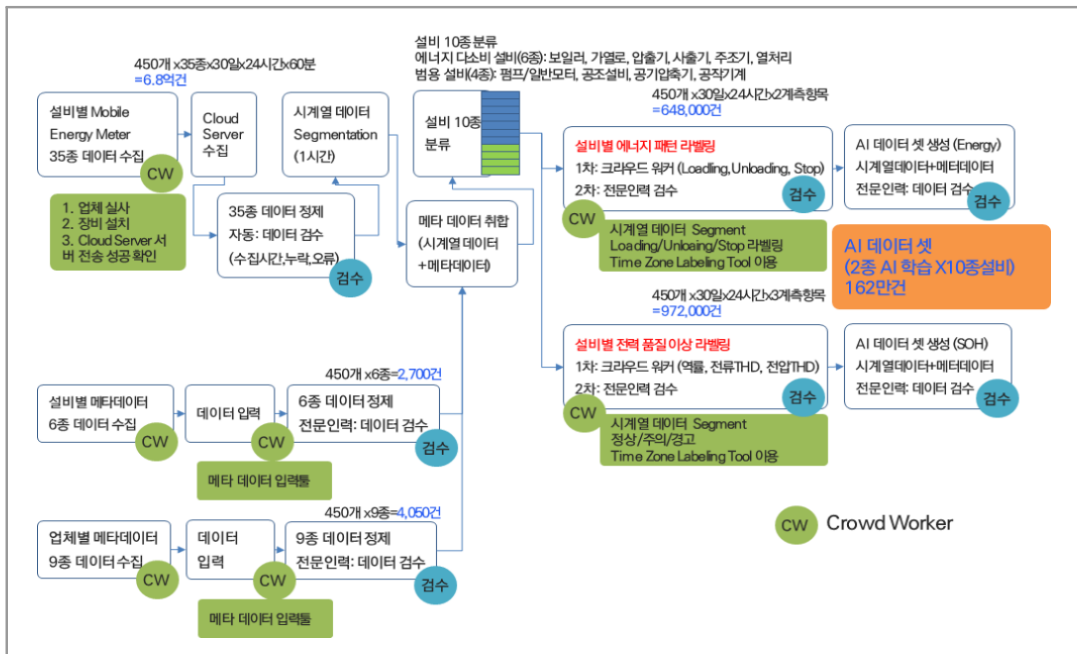


그림 | 인공지능 학습 데이터 구축을 위한 데이터 수집 및 데이터 가공 프로세스

1. 데이터 획득/정제 과정

- Mobile Energy Meter를 이용한 35종의 원시데이터는 총 70개 업체로부터 450개의 설비에 대한 1개월 데이터를 수집하며, 누적전력량, 유효전력, 역률 등 과금과 관련된 데이터 3종 이외에 전력 품질 데이터 35종 6.8억건을 획득함
 - 데이터 수집 설비 개수: 450개 설비
 - 데이터 수집 주기: 1분
 - 데이터 수집 기간: 1개월

설비개수	데이터 종류	데이터 저장 기간(일)	데이터 수집 주기(분)	일간 데이터 개수(건/일)	총 데이터 개수(건)	데이터량
450	35	30	1	1440	680,400,000	2.71 TB

- 상기 원시 데이터의 항목 및 대표성은 1.6절에서 상세히 기술하였으며, 실제 데이터 구축 및 수집 배포에 법률적인 문제가 없음을 확인하였음
- 설비별 메타데이터 및 업체별 메타데이터는 450개 설비 및 70개 업체에 대해 수집하며 하기와 같은 정보를 포함하며, AI 학습을 위한 데이터 이외의 민감한 정보는 비공개할 수 있음

표 | 설비별 정보 데이터 요약표

	데이터 중분류	데이터 소분류	내 용	유형	비고
1	설비별	설비명	설비의 이름	Text	선택 공개 (일반,특수)
2		설비종류	설비의 종류	Text	공개
3		제조사	설비 제조사	Text	선택 공개 (일반,특수)
4		제작연도	설비 제작연도 또는 설치 연도	Value(Int)	공개
5		정격용량	설비가 사용하는 최대 전력 정격	Value(Int)	공개
6		정격전압	설비가 사용하는 전압 정격	Value(Int)	공개

표 | 업체 정보 데이터 요약표

	데이터 중분류	데이터 소분류	내 용	유형	비고
1	업체별	업체명	설비가 설치된 업체명, 공장, 빌딩명	Text	비공개
2		설립연도	업체, 공장, 빌딩 설립 연도	Value(Int)	공개
3		주소	업체, 공장, 빌딩 주소	Text	일부 공개 (지역)
4		업종	업체, 공장, 빌딩의 업종	Text	공개
5		업태	업체, 공장, 빌딩의 업종	Text	공개
6		주생산물	업체, 공장, 빌딩의 주 생산물	Text	선택 공개 (일반,특수)
7		계약전력	업체, 공장, 빌딩의 수전 용량에 따른 한전과의 계약 전력	Value(Int)	공개
8		선택요금제	업체, 공장, 빌딩이 한전과 계약한 선택 요금제 형태	Text	공개
9		한전 정보	한전 과금 정보 조회용 업체별 정보	Text	비공개

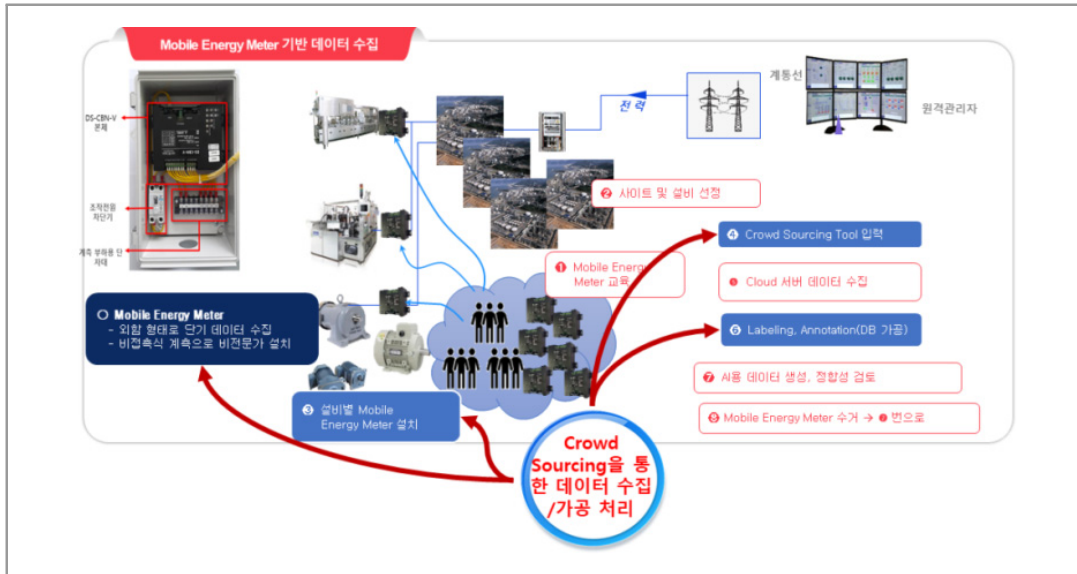


그림 | Mobile Energy Meter를 이용한 원시 데이터 수집

2. 획득·정제 기준

- AI학습용 원시데이터는 Mobile Energy Meter를 이용한 계측 데이터로, 1.6절의 계측 데이터를 수집
- 원시데이터는 데이터 수집 서버에서는 수집된 데이터에 대한 자동 필터링을 통한 정제 과정을 거침
- 라벨링 데이터는 2.4.2절의 라벨링 기준을 어노테이션 및 검수 과정을 통해 정제

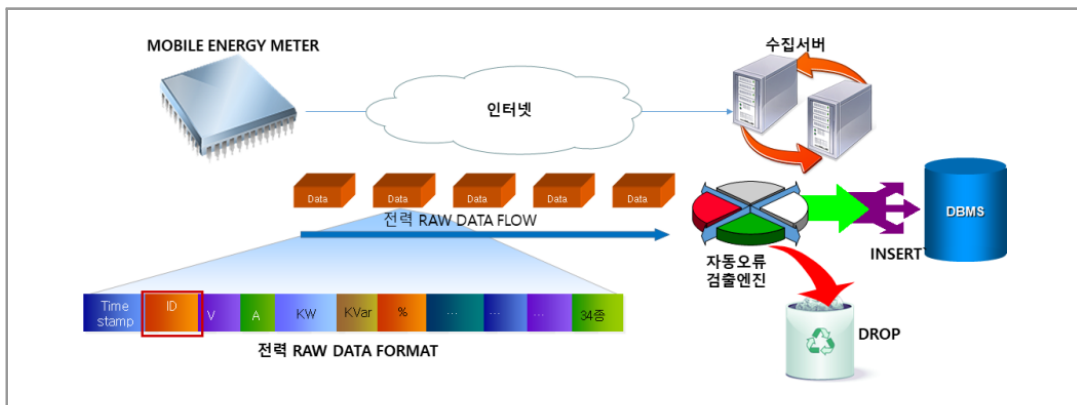


그림 | 원천 데이터 수집 및 오류 데이터 필터링

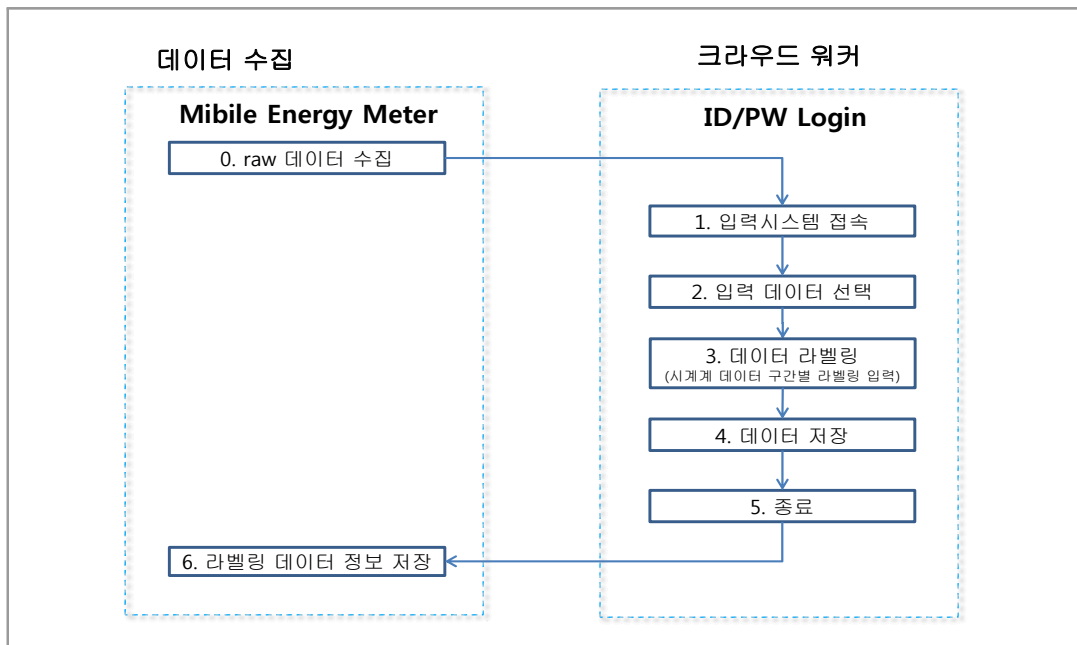
- 클라우드 워커 라벨링 및 전문가 검수
 - 에너지 패턴 학습용 데이터의 라벨링을 프로그램을 이용하여 자동생성할 경우 Loading/Unloading/Stop의 레벨을 사전에 설정할 수 있어야 하나, 설비별로 매우 상이한 패턴 및 레벨을

갖게 됨. Loading/Unloading/Stop 라벨링은 AI학습의 대상으로 설비간 동작 사이의 Glitch 및 Transient를 제거한 라벨링이 필요함

- 하기 그림은 실제 데이터를 캡처하여 에너지 사용 패턴을 클라우드 워커 또는 자동 라벨링을 생성한 라벨링 오류의 사례로, Glitch 및 Transient를 잘못 인식하여 라벨링을 한 사례임

3. 어노테이션/라벨링 절차

- 어노테이션 및 라벨링 절차와 기준을 이용하여 클라우드 워커의 라벨링 및 전문 인력의 검수 과정을 거쳐 라벨링 수행



4. 어노테이션/라벨링 기준

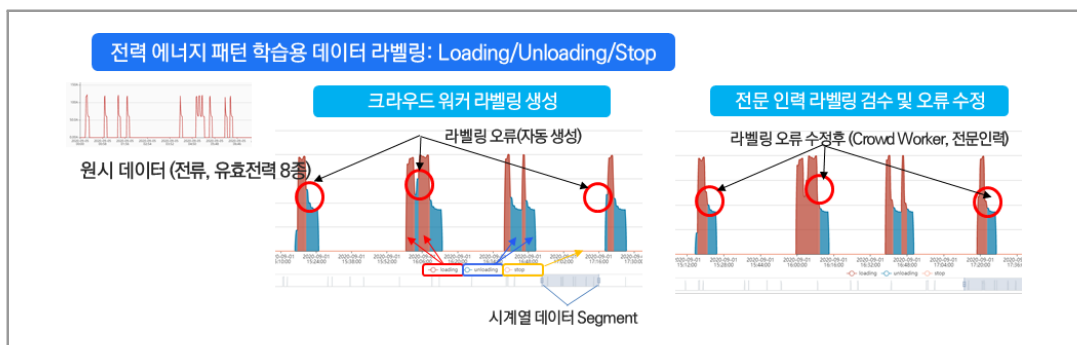


표 | 에너지 사용 패턴 라벨링 및 검수 기준

구축 데이터	라벨	의미	라벨링 기준 및 라벨링 방법
전력 에너지 패턴 학습용 데이터	Loading	설비의 정상 가동	기준: 전력 계측 데이터 중 전류 및 유효 전력 데이터의 패턴 중 최대 정격의 80% 이상의 값을 갖는 경우. 단 매우 짧은 유희 기간이나 (transient) 초기 Stop상태에서의 기동으로 인해 기준값 이하의 경우는 Loading으로 간주함 라벨링 방법: 1. 클라우드 워커 라벨링 2. 전문인력 검수 및 오류 수정
	Unloading	설비의 유희 및 대기 상태	기준: 전력 계측 데이터 중 전류 및 유효 전력 데이터의 패턴 중 최대 정격의 80% 이하 10% 이상의 값을 갖는 경우. 단 Loading상태에서 감소하는 경우(transient)는 Unloading상태가 Stable한 상태로 진입한 이후부터 Unloading으로 라벨링 라벨링 방법: 1. 클라우드 워커 라벨링 2. 전문인력 검수 및 오류 수정
	Stop	설비의 비가동	기준: 전력 계측 데이터 중 전류 및 유효 전력 데이터의 패턴 중 최대 정격의 10% 이하의 값을 갖는 경우. 단 Stop상태에서 Loading상태로 진입하는 경우에는 Loading으로 라벨링 라벨링 방법: 1. 클라우드 워커 라벨링 2. 전문인력 검수 및 오류 수정

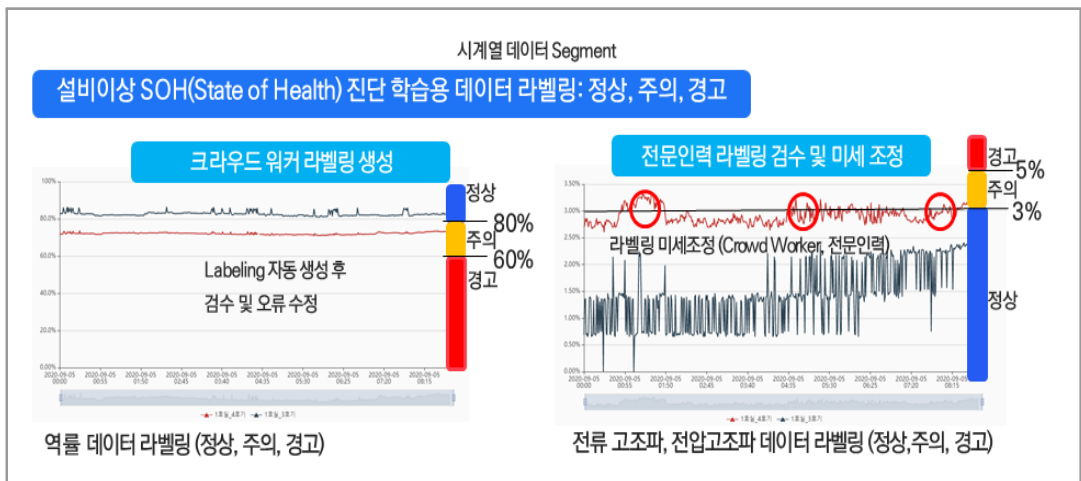
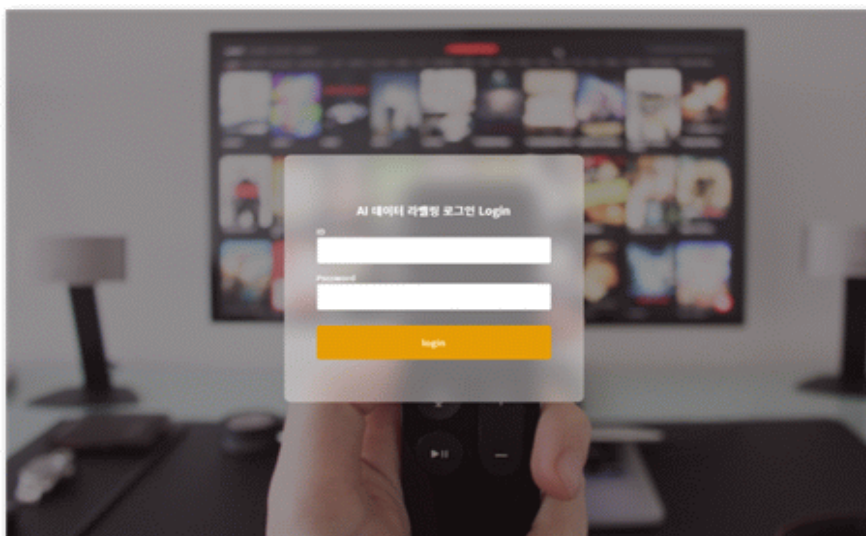


표 | 설비 SOH 라벨링 및 검수 기준

구축 데이터	라벨	의미	라벨링 기준 및 라벨링 방법
설비 SOH (State of Health) 진단 학습용 데이터	정상	전력 품질 데이터가 정상인 상태	기준: 역률: 80% 이상, 전압 고조파: 3% 이하 라벨링 방법: 1. 클라우드 워커 라벨링 2. 전문인력 검수 및 오류 수정
	주의	전력 품질데이터가 한전 기준 이상으로 높아지거나 낮아질 경우	기준: 역률: 60%~80%, 전압 고조파: 3%~5% 사이 라벨링 방법: 1. 클라우드 워커 라벨링 2. 전문인력 검수 및 오류 수정
	경고	전력 품질데이터가 한전 기준 보다 과도하게 높아지거나 낮아질 경우	기준: 역률: 60% 이하, 전압 고조파: 5% 이상 라벨링 방법: 1. 클라우드 워커 라벨링 2. 전문인력 검수 및 오류 수정

5. 어노테이션/라벨링 도구

- TimeZone Labeling tool을 활용(웹페이지 접속, work.aidl.mbz.kr)하여 수집된 시계열 원천 데이터에 대한 annotation (labeling + 메타 데이터)을 수행하고, 수집된 데이터의 정합성 검토한다.



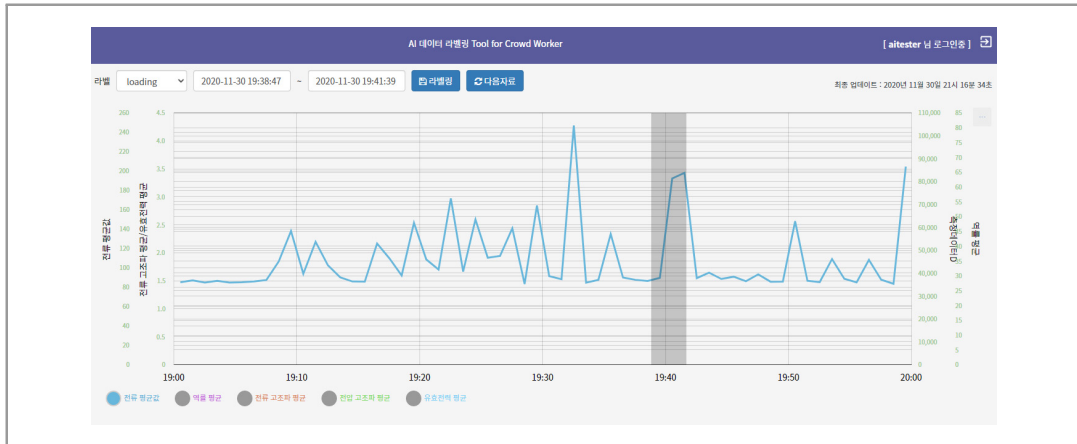


그림 | AI 학습용 데이터 구축용 TimeZone 라벨링 툴 (work.aidl.mbz.kr)

●○ 검수와 품질 확보

원천 데이터 획득/정제 과정에서의 전문가 검수를 통해 원천 데이터에 대한 1차 검수를 진행하고 크라우드 워커가 라벨링 한 결과물에 대해 라벨링 툴을 통해 전문가가 2차 검수를 진행함으로써, 최종 구축 AI 데이터셋에 대한 품질을 확보한다.

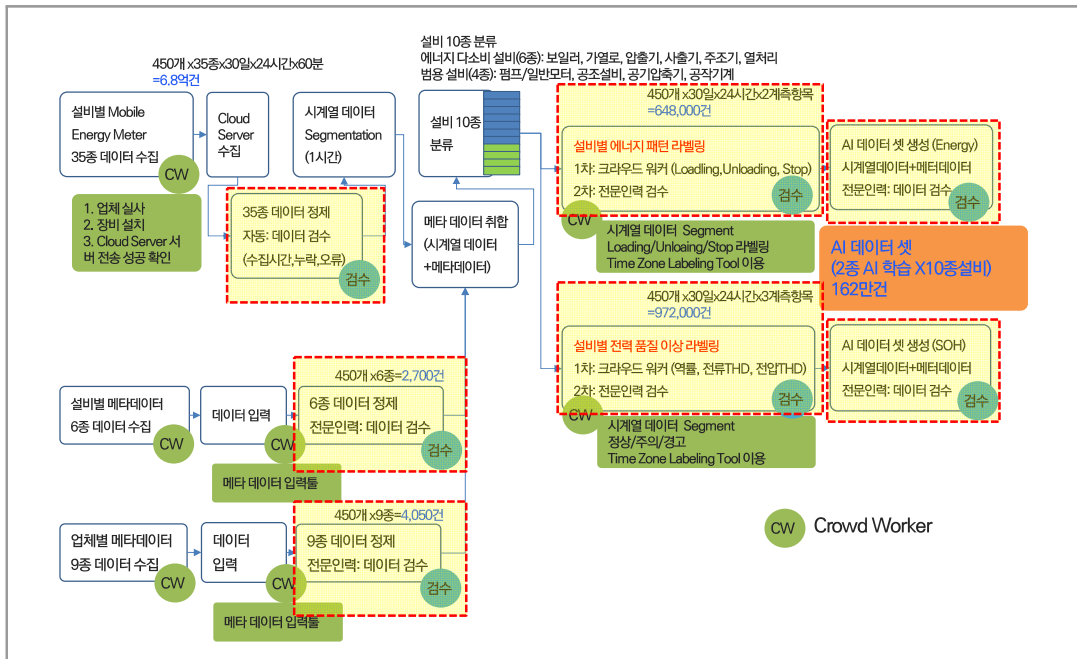


그림 | 전체 AI 학습 데이터 구축 과정 단계별 검수 Process

검수 단계	설명
설비별 에너지 패턴 데이터 수집/정제 검수	<ul style="list-style-type: none"> 수집 원천 데이터(35종)에 대한 수집시간, 누락, 오류 검수
설비별 메타데이터 수집/정제 검수	<ul style="list-style-type: none"> 업체별 실사 후 주요 설비 선정 리스트 확인(450개 설비) 설비별 메타데이터 정합도 검수(정합률 99%)
업체별 메타데이터 수집/정제 검수	<ul style="list-style-type: none"> 대상 업체 선정 리스트 확인(전국 70개 업체) 업체별 메타데이터 정합도 검수(정합률 99%)
설비별 에너지 패턴 라벨링 데이터 검수	<ul style="list-style-type: none"> 에너지 사용 패턴 라벨링 데이터(Loading/Unloading/Stop)에 대한 전문인력 검수 과정을 통해 오류 수정
설비별 전력 품질 이상 SOH 라벨링 데이터 검수	<ul style="list-style-type: none"> 설비별 SOH 라벨링 데이터(Loading/Unloading/Stop)에 대한 전문인력 검수 과정을 통해 오류 수정
AI 데이터 셋 생성(Energy) 검수	<ul style="list-style-type: none"> 사용패턴 분석 정확도 검수 (Accuracy 95% 이상)
AI 데이터 셋 생성(SOH) 검수	<ul style="list-style-type: none"> 고장진단 정확도 검수 (Accuracy 95% 이상)

