

BEAR 빅데이터 경진대회

설비 이상 판단 모델

16102203 위성률

19102078 강인영

목차

- ① 문제 정의
- ② 데이터 수집
- ③ 데이터 전처리 및 EDA
- ④ 데이터 모델링
- ⑤ 결과 분석
- ⑥ 기대효과

문제 정의

설비 이상을 가장 정확하게 판단하고 예측할 수 있는 방법은 무엇일까?

문제 정의

설비 이상을 가장 정확하게 판단하고 예측할 수 있는 방법은 무엇일까?

- 전력 설비를 위해 여러 측정 데이터들이 존재
- 설비 이상을 정확하게 판단하여 물품 생산 지연 등의 문제 발생 감소
- 설비 이상 예측을 통해 문제 발생 사전 방지

문제 정의

Q. 기존 설비 이상 진단을 위해 설비에서 측정한 전류 고조파 평균, 전압 고조파, 역률 평균을 고려했던데 그 이유에 대해 알 수 있을까요?

고조파 전류가 많은 기기는 역률이 낮기 때문에 실제로 소비되는 전력보다 피상 전력이 크기때문에 입력 전류가 많이 흐르기 때문에 전력 설비에 여유가 필요합니다.

전원은 발전소에서 송전선이나 변전소 등을 통해 각 가정에 접속되어 있는데 그 전원 라인에는 임피던스가 존재하고, 고조파 전류가 흐르면 그 임피던스에 의해 전압 강하가 발생합니다. 결과적으로 전원 전압 파형이 고조파를 포함한 파형이 됩니다.

전원 라인에 접속된 기기나 배전 설비에 악영향을 미친 예를 들어 보면, 진상 콘덴서에 고조파 전류가 흘러 소손 되거나 변압기에서 소음이 나거나 전압의 피크 치가 내려 스위칭 전원이 정상 동작 하지 않게 된 사례가 있습니다.

=> 전기를 사용하는 기기는 전류 고조파, 전압 고조파, 역률 수치에 따라 영향을 받는다.



문제 정의

Q. 설비 이상을 3가지 요소마다 개별적으로 진단하는 것에서 나아가 3가지 요소의 데이터를 합쳐 경향을 파악하고 설비 이상을 판단하고 예측한다면 보다 의미 있는 결과라고 말할 수 있을까요?

네 맞습니다. 특히 고조파 문제는 일주일 아니면 보름간 등 일정 기간에 해당하는 데이터를 모아서 판단 합니다. 고조파는 재현성이 없어서 특히 더 그렇습니다. 예를 들어, 오전 조업 중에는 괜찮다가 오후가 되면 문제가 많이 발생하기도 하기 때문입니다. 또 요일 마다 달라질 수도 있습니다.

=> 3가지의 개별 진단을 하나로 합쳐 보다 의미 있는 모델을 만들 수 있지 않을까?



문제 정의

Q. 전류 고조파, 전압 고조파, 역률 데이터들이 모두 T상, R상, S상 정보들이 따로 있는데요. 예를 들어 T상 전류고조파라는 데이터가 있는데 해당 값은 제외하고 전류 고조파 만을 고려해도 되는 것인가요?

이왕이면 측정 신뢰성을 높이기 위해서 3상 다 하는 것이 좋겠습니다.

⇒ 전류 고조파, 전압 고조파, 역률 데이터의 평균은 3상을 통해 구해진 값이므로 평균을 제외한 3상 개별 데이터를 활용해보자!

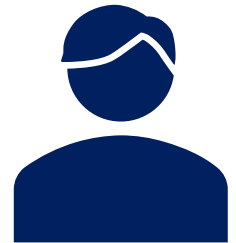


문제 정의

Q. 전류 고조파, 전압 고조파, 역률 외에도 누적 전력량, 선간 전압평균, 무효전력평균, 전류평균, 주파수, 온도 등의 데이터도 있는데 저 3가지의 정보만을 고려해도 설비 이상 판단 모델을 만드는 것이 괜찮을까요?

네 그렇게 하면 될 것 같습니다. 누적 전력량, 선간 전압평균 등을 알아 보는 것은 좋지만.... 괜히 프로그램 작성 등에 시간만 더 걸릴 수 있으니 간편하게 3가지만을 고려한 모델을 만드는 방식이 좋을 것 같네요.

- ⇒ 우선적으로는 전류 고조파, 전압 고조파, 역률 데이터만 고려해서 모델을 만들자!
- ⇒ 활용 기회가 있다면 다른 측정 데이터도 활용하면 좋을 것이다!



문제 정의

“ 기존에 제공하는 개별 데이터에
대한 진단이 아닌 통합 데이터를
활용해 보다 정확도 높은 설비
이상 판단 모델을 구축해보자. ”

데이터 수집

- 데이터 소개

- 데이터셋명 : 전력 설비 에너지 패턴 및 고장 분석 센서
- 제공 : AI Hub 개방 데이터
- 구축 데이터량 : 162만 / 구축년도 : 2020년
- Mobile Energy Meter를 이용한 전기 에너지 사용 패턴 및 고장 진단 분석을 위한 전력 품질 데이터



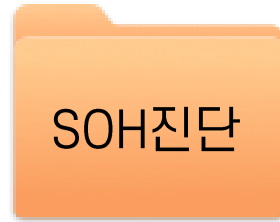
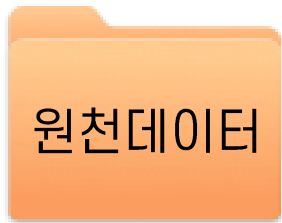
Mobile Energy Meter

데이터 수집

- 전체 데이터 구조



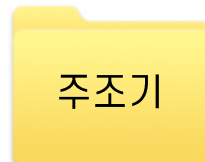
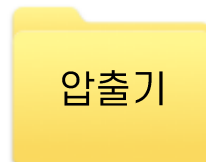
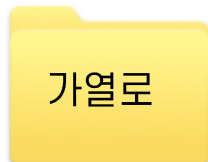
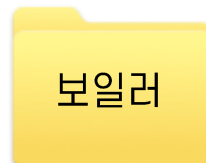
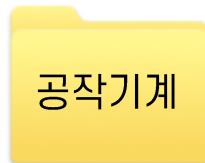
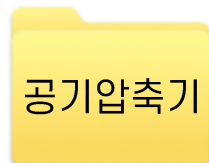
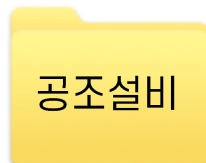
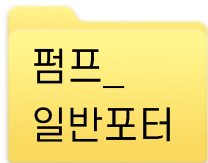
훈련 데이터 / 검증 데이터



원천데이터: 라벨 X 데이터

기동패턴: Loading, Unloading, Stop 라벨 데이터

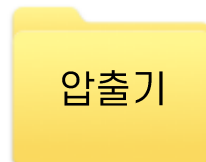
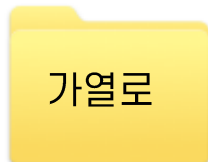
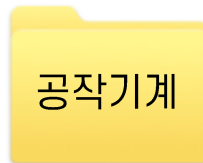
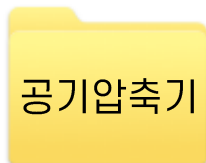
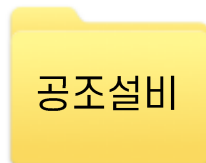
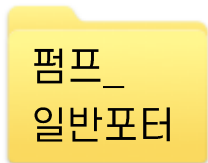
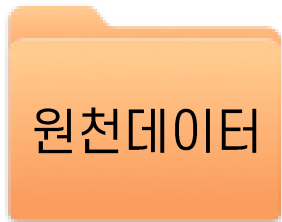
SOH진단: 정상, 주의, 경고 라벨 데이터



10개의 설비 개별 데이터

데이터 수집

- 데이터 선정



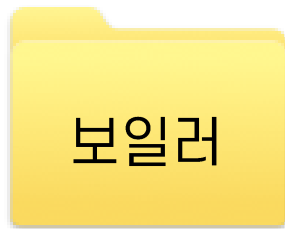
모델 생성을 위해 Training 데이터 활용

기존 설비 이상 진단에 따른 Label 데이터를 포함하고 있는 SOH진단 데이터 사용

데이터 수집

- 데이터 선정

■ 보일러



보일러 기기 14대

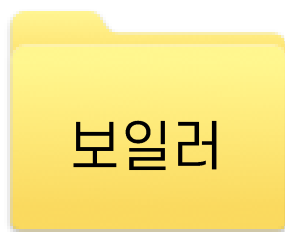
98, 211, 347, 359, 360,,,

이유) 공장 내 에너지 다소비 범용 설비로,
설비를 사용하는 업종에 따라 영향을 받지
않으므로 해당 보일러 데이터를 선정.

데이터 수집

- 데이터 선정

■ 보일러



보일러 기기 14대

98, 211, 347, 359, 360,,,

-> 임의의 2개 기기의 데이터(98, 347) 사용

Combined_LabelledData_98_역률평균	2021-02-21 오전 1:55	JSON 파일	263,788KB		
Combined_LabelledData_211_역률평균	2021-02-21 오전 2:30	JSON 파일	265,832KB		
Combined_LabelledData_347_역률평균	2021-02-21 오후 12:01	JSON 파일	177,120KB		
Combined_LabelledData_359_역률평균	2021-02-21 오후 12:01	JSON 파일	176,952KB		
Combined_LabelledData_3	Combined_LabelledData_98_전류고조...	2021-02-21 오전 1:56	JSON 파일	263,788KB	
Combined_LabelledData_3	Combined_LabelledData_211_전류고조...	2021-02-21 오전 2:31	JSON 파일	265,832KB	
Combined_LabelledData_3	Combined_LabelledData_347_전류고조...	2021-02-21 오후 12:01	JSON 파일	177,120KB	
Combined_LabelledData_4	Combined_LabelledData_359_전류고조...	2021-02-21 오후 12:01	JSON 파일	176,952KB	
Combined_LabelledData_4	Combined_LabelledData_3	Combined_LabelledData_98_전압고조...	2021-02-21 오전 1:55	JSON 파일	263,788KB
Combined_LabelledData_4	Combined_LabelledData_3	Combined_LabelledData_211_전압고조...	2021-02-21 오전 2:31	JSON 파일	265,832KB
Combined_LabelledData_5	Combined_LabelledData_3	Combined_LabelledData_347_전압고조...	2021-02-21 오후 12:01	JSON 파일	177,120KB
Combined_LabelledData_5	Combined_LabelledData_4	Combined_LabelledData_359_전압고조...	2021-02-21 오후 12:01	JSON 파일	176,952KB
Combined_LabelledData_5	Combined_LabelledData_4	Combined_LabelledData_360_전압고조...	2021-02-21 오후 12:01	JSON 파일	176,454KB
Combined_LabelledData_5	Combined_LabelledData_4	Combined_LabelledData_364_전압고조...	2021-02-21 오후 12:01	JSON 파일	163,571KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_391_전압고조...	2021-02-21 오후 12:01	JSON 파일	149,916KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_443_전압고조...	2021-02-21 오후 12:00	JSON 파일	176,245KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_445_전압고조...	2021-02-21 오후 12:00	JSON 파일	150,735KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_452_전압고조...	2021-02-21 오전 11:58	JSON 파일	124,179KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_531_전압고조...	2021-02-21 오후 12:01	JSON 파일	176,949KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_542_전압고조...	2021-02-21 오전 4:56	JSON 파일	150,506KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_543_전압고조...	2021-02-21 오전 4:58	JSON 파일	26,756KB
Combined_LabelledData_5	Combined_LabelledData_5	Combined_LabelledData_546_전압고조...	2021-02-21 오전 5:01	JSON 파일	150,353KB

역률평균, 전류고조파평균, 전압고조파평균 기준 Label 데이터
=> 총 6개 파일의 데이터 사용

데이터 전처리 및 EDA

- 데이터 불러오기

```
1 powerfactor = pd.read_json('347_역률평균.json')
2 powerfactor
3
4 # 데이터 불러옴
```

```
1 electric = pd.read_json('347_전류고조파평균.json')
2 electric
3
4 # 데이터 불러옴
```

```
1 voltage = pd.read_json('347_전압고조파평균.json')
2 voltage
3
4 # 데이터 불러옴
```

- 설비 측정 데이터 선별

SYN_NAME	BASE_ITEM	DEVICE_ID	DEVICE_ID_ID	COMPANY_NAME	ESTABLISH_YEAR	LOCATION	COMPANY_TYPE	COMPANY_CAT	MAJOR_PRODUCT	CONTRACT_POWER	BILL_OPTION	KEPCO_INFO	facility_name	facility_type_name	facility_vendor	facility_year	facility_capacity	facility_vo	data	
0	SOH	역률평균	5764	1	CO-313	1963	대구 달서구	제조	염색가공	염유	980	산업용(용)고압A 선택H	05-*****	아이소프트1	보일러	명인 인력선	2008	59	300	(ITEM_NAME: '누락전력량', ITEM_VALUE: 6.7485, TIM...
1	SOH	역률평균	5764	1	CO-313	1963	대구 달서구	제조	염색가공	염유	980	산업용(용)고압A 선택H	05-*****	아이소프트1	보일러	명인 인력선	2008	59	300	(ITEM_NAME: '순도', ITEM_VALUE: 31.873, TIM...
2	SOH	역률평균	5764	1	CO-313	1963	대구 달서구	제조	염색가공	염유	980	산업용(용)고압A 선택H	05-*****	아이소프트1	보일러	명인 인력선	2008	59	300	(ITEM_NAME: '양전압평균', ITEM_VALUE: 131.41667...
3	SOH	역률평균	5764	1	CO-313	1963	대구 달서구	제조	염색가공	염유	980	산업용(용)고압A 선택H	05-*****	아이소프트1	보일러	명인 인력선	2008	59	300	(ITEM_NAME: '신간정압평균', ITEM_VALUE: 227.6666...
4	SOH	역률평균	5764	1	CO-313	1963	대구 달서구	제조	염색가공	염유	980	산업용(용)고압A 선택H	05-*****	아이소프트1	보일러	명인 인력선	2008	59	300	(ITEM_NAME: '전압고조파평균', ITEM_VALUE: 2.50651...

```
powerlist = []
def powerapply(col):
    powerlist.append(list(col.values()))
    return list(col.values())

power_temp = powerfactor['data'].apply(powerapply)
column_name = ['ITEM_NAME', 'ITEM_VALUE', 'LABEL_NAME', 'TIMESTAMP']

power_label = pd.DataFrame.from_records(powerlist, columns = column_name)

powerfactor = pd.concat([powerfactor, power_label], axis = 1)
powerfactor.drop('data', axis = 1, inplace = True)

powerfactor.rename(columns = {'LABEL_NAME' : 'TIMESTAMP', 'TIMESTAMP' : 'LABEL_NAME'}, inplace = True)
powerfactor

# 역률, 전류고조파평균, 전압고조파 평균 data에 dict형태로 되어있는 걸 꺼내서 column으로 추가해 줌
```

json 형태의 전체 데이터에서 data라는 key에 들어 있는 설비의 측정 데이터만을 가져온다. (6개 파일 모두 적용)

데이터 전처리 및 EDA

- 데이터 추출

SYD_NAME	BASE_ITEM	DEVICE_ID	DEVICE_ID	COMPANY_NAME	ESTABLISH_YEAR	LOCATION	COMPANY_TYPE	COMPANY_CAT	MAJOR_PRODUCT	CONTRACT_POWER	BILL_OPTION	KEPCO_INFO	facility_name	facility_type_name
0	SOH	전압고조파 평균	5764	1	CO-313	1963	대구 달서구	제조	열대기공	정유	980	산전형(중고압) 선형	아이소소프트1	보일러
1	SOH	전압고조파 평균	5764	1	CO-313	1963	대구 달서구	제조	열대기공	정유	980	산전형(중고압) 선형	아이소소프트1	보일러
2	SOH	전압고조파 평균	5764	1	CO-313	1963	대구 달서구	제조	열대기공	정유	980	산전형(중고압) 선형	아이소소프트1	보일러
3	SOH	전압고조파 평균	5764	1	CO-313	1963	대구 달서구	제조	열대기공	정유	980	산전형(중고압) 선형	아이소소프트1	보일러
4	SOH	전압고조파 평균	5764	1	CO-313	1963	대구 달서구	제조	열대기공	정유	980	산전형(중고압) 선형	아이소소프트1	보일러
...
1007995	SOH	전압고조파 평균	5764	1	CO-313	1963	대구 달서구	제조	열대기공	정유	980	산전형(중고압) 선형	아이소소프트1	보일러

facility_vendor	facility_year	facility_capacity	facility_volt	ITEM_NAME	ITEM_VALUE	TIMESTAMP	LABEL_NAME
명인 인덕션	2008	59	380	누적전력량	0.000000	2021-01-16 00:00:13	경고
명인 인덕션	2008	59	380	온도	31.875000	2021-01-16 00:00:13	경고
명인 인덕션	2008	59	380	상전압평균	131.416672	2021-01-16 00:00:13	경고
명인 인덕션	2008	59	380	선간전압평균	227.666672	2021-01-16 00:00:13	경고
명인 인덕션	2008	59	380	전압고조파 평균	2.506511	2021-01-16 00:00:13	경고
...
명인 인덕션	2008	59	380	R상전압고조파	3.125000	2021-02-04 23:59:06	경고

```
1 powerfactor = powerfactor[['BASE_ITEM', 'ITEM_NAME', 'ITEM_VALUE', 'TIMESTAMP', 'LABEL_NAME']]
2 electric = electric[['BASE_ITEM', 'ITEM_NAME', 'ITEM_VALUE', 'TIMESTAMP', 'LABEL_NAME']]
3 voltage = voltage[['BASE_ITEM', 'ITEM_NAME', 'ITEM_VALUE', 'TIMESTAMP', 'LABEL_NAME']]
4
5 # 일단 BASE_ITEM, ITEM_NAME, ITEM_VALUE, TIMESTAMP, LABEL_NAME만 꺼내 줌
6 # LABEL_NAME이 타겟값임
```

실제 설비의 측정 데이터와 라벨 데이터를 위해서 ITEMNAME, ITEM_VALUE, TIMESTAMP, LABEL_NAME, BASE_ITEM 열만 추출

	BASE_ITEM	ITEM_NAME	ITEM_VALUE	TIMESTAMP	LABEL_NAME
0	역률평균	누적전력량	0.000000	2021-01-16 00:00:13	경고
1	역률평균	온도	31.875000	2021-01-16 00:00:13	경고
2	역률평균	상전압평균	131.416672	2021-01-16 00:00:13	경고
3	역률평균	선간전압평균	227.666672	2021-01-16 00:00:13	경고
4	역률평균	전압고조파평균	2.506511	2021-01-16 00:00:13	경고
...

- 데이터 정리

```
1 def second(col):
2     return col[:16]
3
4 powerfactor['TIMESTAMP'] = powerfactor['TIMESTAMP'].apply(second)
5 electric['TIMESTAMP'] = electric['TIMESTAMP'].apply(second)
6 voltage['TIMESTAMP'] = voltage['TIMESTAMP'].apply(second)
7
8 # 00:00:13에서 초를 없애 줌
```

TIMESTAMP 열에 있는 시간 데이터에서 초를 제거

	BASE_ITEM	ITEM_NAME	ITEM_VALUE	TIMESTAMP	LABEL_NAME
0	역률평균	누적전력량	0.000000	2021-01-16 00:00	경고
1	역률평균	온도	31.875000	2021-01-16 00:00	경고
2	역률평균	상전압평균	131.416672	2021-01-16 00:00	경고
3	역률평균	선간전압평균	227.666672	2021-01-16 00:00	경고
4	역률평균	전압고조파평균	2.506511	2021-01-16 00:00	경고
...
1007995	역률평균	R상전압고조파	3.125000	2021-02-04 23:59	경고
1007996	역률평균	R상전류고조파	0.000000	2021-02-04 23:59	경고
1007997	역률평균	S상무효전력	0.000000	2021-02-04 23:59	경고
1007998	역률평균	S상전류	0.000000	2021-02-04 23:59	경고
1007999	역률평균	R상전류	0.000000	2021-02-04 23:59	경고

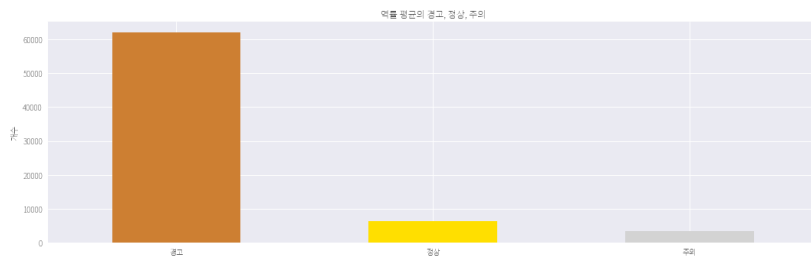
데이터 전처리 및 EDA

- Label 데이터 시각화

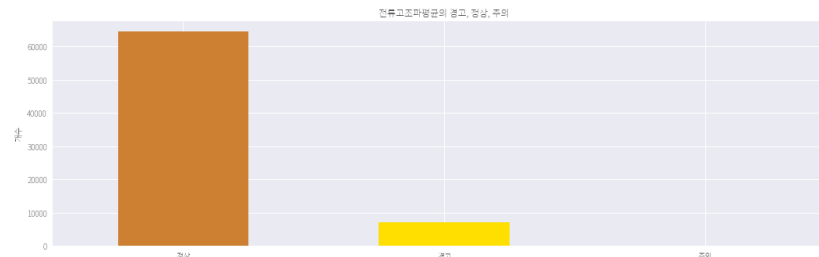
```
1 fig, ax = plt.subplots(3,1,figsize = (18,18))
2
3
4 powerfactor['LABEL_NAME'].value_counts().plot.bar(color=['#CD7F32','#FFDF00','#D3D3D3'], ax=ax[0], rot = 0)
5 ax[0].set_title('역률 평균의 경고, 정상, 주의')
6 ax[0].set_ylabel('개수')
7 plt.xticks(rotation = 90)
8 electric['LABEL_NAME'].value_counts().plot.bar(color=['#CD7F32','#FFDF00','#D3D3D3'], ax=ax[1],rot = 0)
9 ax[1].set_title('전류고조파평균의 경고, 정상, 주의')
10 ax[1].set_ylabel('개수')
11 voltage['LABEL_NAME'].value_counts().plot.bar(color=['#CD7F32','#FFDF00','#D3D3D3'], ax=ax[2], rot = 0)
12 ax[2].set_title('전압고조파평균의 경고, 정상, 주의')
13 ax[2].set_ylabel('개수')
```

SOH 진단을 위한 정상, 주의, 경고 label 데이터의 개수를 막대 그래프로 시각화

< 시각화 결과 >



역률평균 데이터의 경고, 정상, 주의



전류고조파평균 데이터의 정상, 경고, 주의



전압고조파평균 데이터의 주의, 정상, 경고

데이터 전처리 및 EDA

- 데이터 정리

```
1 makecolumn = powerfactor['ITEM_NAME'].unique()
2
3 data = pd.DataFrame(columns = makecolumn)
4
5 data['TIMESTAMP'] = 0
6 data['LABEL_NAME'] = 0
7
8 time = powerfactor['TIMESTAMP'].unique()
9 data['TIMESTAMP'] = time
10
11 datacol = powerfactor[powerfactor['TIMESTAMP'] == '2021-02-05 00:00'].sort_values(by = 'ITEM_NAME')['ITEM_NAME']
12 datacol = list(datacol)
13
14 datacol.append('TIMESTAMP')
15 datacol.append('LABEL_NAME')
16
17 data.columns = datacol
18 data
```

행으로 존재하는 설비의 측정 데이터를 활용보다 잘 활용하기 위해 먼저 ITEM_NAME을 열로 이동

```
timecols = powerfactor['TIMESTAMP'].unique()
len(timecols)

for time, col in enumerate(timecols):
    if time % 1000 == 0:
        print(time)
    temp = powerfactor[powerfactor['TIMESTAMP'] == col].sort_values(by = 'ITEM_NAME')
    temp = temp.drop_duplicates()
    temp2 = temp['ITEM_VALUE']
    data.iloc[time,:35] = list(temp2)
    data.loc[time,'LABEL_NAME'] = temp['LABEL_NAME'].unique()[0]
```

열로 이동한 ITEM_NAME에 대한 값으로 ITEM_VALUE를 적절하게 찾아 넣어 준다.

데이터 전처리 및 EDA

- 데이터 정리

변환 전 :

	BASE_ITEM	ITEM_NAME	ITEM_VALUE	TIMESTAMP	LABEL_NAME
0	역률평균	누적전력량	0.000000	2021-01-16 00:00:13	경고
1	역률평균	온도	31.875000	2021-01-16 00:00:13	경고
2	역률평균	상전압평균	131.416672	2021-01-16 00:00:13	경고
3	역률평균	선간전압평균	227.666672	2021-01-16 00:00:13	경고
4	역률평균	전압고조파평균	2.506511	2021-01-16 00:00:13	경고
...
1007995	역률평균	R상전압고조파	3.125000	2021-02-04 23:59:06	경고
1007996	역률평균	R상전류고조파	0.000000	2021-02-04 23:59:06	경고
1007997	역률평균	S상무효전력	0.000000	2021-02-04 23:59:06	경고
1007998	역률평균	S상전류	0.000000	2021-02-04 23:59:06	경고
1007999	역률평균	R상전류	0.000000	2021-02-04 23:59:06	경고

1008000 rows x 5 columns

1분마다 35가지에 대한 측정 데이터로 인해 35개의 열로 구성된 데이터를 1분 당 하나의 열 안에 측정 데이터가 들어갈 수 있도록 정리하여 모델 구축에 해당 설비의 측정 데이터를 유용하게 사용할 수 있다.

변환 후 :

	R상누계전력	R상선간전압	R상선대	R상누계전력	R상전류	R상전류고조파	R상전압	R상전압고조파	S상누계전력	S상선간전압	S상선대	S상누계전력	S상전류	S상전류고조파	S상전압	S상전압고조파	T상누계전력	T상선간전압	T상선대	T상누계전력	T상전류	T상전류고조파	T상전압	T상전압고조파	누적전력량	무효전력(과전)	상전압평균	선간전압평균	합계전압	교인	유효전력(과전)	전압고조파평균	전류고조파평균	전압고조파평균	주파수	TIMESTAMP	LABEL_NAME	
0	0.0	227.50	0.0	0.0	0.0	0.0	131.00	2.539063	0.0	228.25	0.0	0.0	0.0	0.0	131.75	2.343750	0.0	227.25	0.0	0.0	0.0	0.0	131.50	2.636719	0.000	0.0	131.416672	227.666672	0.0	31.875	0.0	0.0	0.0	2.506511	59.938175	2021-01-16 00:00	경고	
1	0.0	226.75	0.0	0.0	0.0	0.0	130.50	2.539063	0.0	227.75	0.0	0.0	0.0	0.0	131.25	2.343750	0.0	226.50	0.0	0.0	0.0	0.0	131.25	2.636719	0.000	0.0	131.000000	227.000000	0.0	33.125	0.0	0.0	0.0	2.506511	59.956898	2021-01-16 00:01	경고	
2	0.0	227.75	0.0	0.0	0.0	0.0	131.25	2.539063	0.0	228.50	0.0	0.0	0.0	0.0	132.00	2.539063	0.0	227.25	0.0	0.0	0.0	0.0	131.75	2.539063	0.000	0.0	131.666672	227.833328	0.0	32.500	0.0	0.0	0.0	2.539063	59.910122	2021-01-16 00:02	경고	
3	0.0	227.00	0.0	0.0	0.0	0.0	130.75	2.539063	0.0	228.25	0.0	0.0	0.0	0.0	131.75	2.343750	0.0	227.00	0.0	0.0	0.0	0.0	131.50	2.636719	0.000	0.0	131.333328	227.416672	0.0	31.875	0.0	0.0	0.0	2.506511	59.947537	2021-01-16 00:03	경고	
4	0.0	228.00	0.0	0.0	0.0	0.0	131.50	2.734375	0.0	229.00	0.0	0.0	0.0	0.0	132.00	2.441406	0.0	227.75	0.0	0.0	0.0	0.0	131.75	2.636719	0.000	0.0	131.750000	228.250000	0.0	32.500	0.0	0.0	0.0	2.604167	59.928820	2021-01-16 00:04	경고	
...
28795	0.0	224.50	0.0	0.0	0.0	0.0	129.50	3.125000	0.0	225.50	0.0	0.0	0.0	0.0	130.00	2.832031	0.0	224.25	0.0	0.0	0.0	0.0	130.00	2.636719	170257.266	0.0	129.833328	224.750000	0.0	22.500	0.0	0.0	0.0	2.864583	59.844750	2021-02-04 23:55	경고	
28796	0.0	225.25	0.0	0.0	0.0	0.0	129.75	3.027344	0.0	226.25	0.0	0.0	0.0	0.0	130.50	2.636719	0.0	225.00	0.0	0.0	0.0	0.0	130.50	2.636719	170257.266	0.0	130.250000	225.500000	0.0	21.875	0.0	0.0	0.0	2.766927	59.910122	2021-02-04 23:56	경고	

데이터 전처리 및 EDA

- 데이터 저장

```
data.to_csv('new역률평균', index = False)
```

역률평균, 전류고조파평균, 전압고조파평균 파일 속 데이터를 모두 정리하여 다시 csv 파일로 저장

- 데이터 불러오기

```
1 powerfactor = pd.read_csv('new역률평균')
2 electric = pd.read_csv('new전류평균')
3 voltage = pd.read_csv('new전압평균')
```

직전까지의 전처리를 통해 저장해둔 csv 파일을 불러와서 사용

- 데이터 합치기

역률 평균, 전류고조파평균, 전압고조파평균 데이터에 존재하는 Label들을 모아 하나의 Dataframe으로 생성

	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전압	상전압 상전압 상전
--	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	------------------

71783 rows x 31 columns

데이터 모델링

- Label 값이 모두 동일한 경우

```
1 # 다 같은거 뽑고
2 same_label = data[(data['POWER_LABEL'] == data['ELECTRIC_LABEL']) & (data['POWER_LABEL'] == data['VOLTAGE_LABEL'])]
3 same_label['FINAL_LABEL'] = same_label['POWER_LABEL']
4 display(same_label)
5 same_label.drop(['POWER_LABEL', 'ELECTRIC_LABEL', 'VOLTAGE_LABEL'], axis = 1, inplace = True)
```

S상당 표전력	S상전 압	S상당 표전력	S상전류 고조파	S상전류 고조파	S상전 압	S상전압 고조파	T상당 표전력	T상전 압	T상당 표전력	T상전류 고조파	T상전류 고조파	T상전 압	T상전압 고조파	누적전력 량	온도	주파수	TIMESTAMP	POWER_LABEL	ELECTRIC_LABEL	VOLTAGE_LABEL	FINAL_LABEL		
392.75	220.00	0.648438	336.00	4.06250	47.070313	127.25	3.710938	478.00	219.25	0.656250	418.25	5.03125	39.746094	126.75	3.808594	106472.211	25.000	59.872740	2021-01-19 13:21	주의	주의	주의	주의
0.00	221.75	0.000000	0.00	0.00000	0.000000	128.00	2.832031	0.00	221.00	0.000000	0.00	0.00000	0.000000	128.00	2.636719	194950.875	30.625	59.863422	2021-01-20 16:10	정상	정상	정상	정상
0.00	221.75	0.000000	0.00	0.00000	0.000000	127.75	2.734375	0.00	220.50	0.000000	0.00	0.00000	0.000000	127.75	2.539063	194950.875	30.000	59.872740	2021-01-20 16:11	정상	정상	정상	정상
0.00	221.25	0.000000	0.00	0.00000	0.000000	127.75	2.832031	0.00	220.00	0.000000	0.00	0.00000	0.000000	127.25	2.539063	194950.875	30.000	59.882076	2021-01-20 16:12	정상	정상	정상	정상
270.00	223.25	0.542969	176.50	2.43750	57.910156	128.75	3.320313	318.00	222.50	0.578125	225.75	3.09375	51.953125	128.75	3.515625	274251.156	33.125	59.891422	2021-01-21 16:05	주의	주의	주의	주의
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
420.00	363.75	0.777344	1761.00	11.12500	5.078125	210.75	2.636719	1577.50	364.00	0.683594	1491.00	10.75000	5.664063	208.75	2.929688	141677.563	15.625	59.947537	2020-11-17 10:38	정상	정상	정상	정상
437.25	364.25	0.832031	2166.00	12.53125	4.785156	210.75	2.636719	1588.50	364.50	0.761719	1880.00	11.93750	5.761719	208.75	2.832031	141770.766	14.375	59.985000	2020-11-17 10:39	정상	정상	정상	정상
494.75	365.25	0.480469	826.25	8.00000	5.273438	211.50	2.734375	1613.00	365.50	0.316406	540.50	8.03125	5.566406	209.25	2.929688	141839.016	15.625	59.956898	2020-11-17 10:40	정상	정상	정상	정상
497.75	365.00	0.871094	2661.50	14.12500	4.687500	211.25	2.636719	1644.75	365.25	0.820313	2363.50	13.46875	5.664063	209.25	2.929688	141926.750	15.000	59.938175	2020-11-17 10:41	정상	정상	정상	정상
457.25	364.50	0.484375	815.00	7.87500	4.882813	211.00	2.734375	1581.50	364.75	0.320313	541.75	8.00000	5.566406	209.00	2.929688	142023.672	15.625	59.928820	2020-11-17 10:42	정상	정상	정상	정상

역률 평균, 전류고조파평균, 전압고조파평균을 통해 진단한 label이 모두 같은 경우, 해당 일시의 설비 상태는 1개로 판단할 수 있다.

Ex)

역률평균label	전류고조파평균label	전압고조파평균label
정상	정상	정상

=> “정상” 으로 판단

데이터 모델링

- 2개의 Label 값이 다른 경우

```
1 # 2 : 1 뽑는다
2 # 1 : 1 : 1, 다 같은 데이터를 빼줘서 2대 1을 얻음
3
4 total_index = same_label_index + all_diff_label_index
5
6 two_one_label = data.drop(total_index,axis = 0)
7
8 two_one_label= two_one_label.reset_index(drop=True)
```

역률 평균, 전류고조파평균, 전압고조파평균을 통해
진단한 label 중 2개의 label은 같으나 1개의 label이 다른
경우(3가지 경우의 수 존재), 2개 존재하는 label이 해당
설비의 상태라고 판단할 수 있다.

R상전압	R상전압고조파	S상전압표준편차	S상전압고조파	T상전압표준편차	T상전압고조파	I상전압표준편차	I상전압고조파	노력전력량	온도	주파수	TIMESTAMP	POWER_LABEL	ELECTRIC_LABEL	VOLTAGE_LABEL	FINAL_LABEL								
131.00	2.539063	0.0	228.25	0.0	0.0	0.0	131.75	2.343750	0.0	227.25	0.0	0.0	0.0	131.50	2.636719	0.000	31.875	59.938175	2021-01-16 00:00	경고	정상	정상	정상
130.50	2.539063	0.0	227.75	0.0	0.0	0.0	131.25	2.343750	0.0	226.50	0.0	0.0	0.0	131.25	2.636719	0.000	33.125	59.956898	2021-01-16 00:01	경고	정상	정상	정상
131.25	2.539063	0.0	228.50	0.0	0.0	0.0	132.00	2.539063	0.0	227.25	0.0	0.0	0.0	131.75	2.539063	0.000	32.500	59.910122	2021-01-16 00:02	경고	정상	정상	정상
130.75	2.539063	0.0	228.25	0.0	0.0	0.0	131.75	2.343750	0.0	227.00	0.0	0.0	0.0	131.50	2.636719	0.000	31.875	59.947537	2021-01-16 00:03	경고	정상	정상	정상
131.50	2.734375	0.0	229.00	0.0	0.0	0.0	132.00	2.441406	0.0	227.75	0.0	0.0	0.0	131.75	2.636719	0.000	32.500	59.928820	2021-01-16 00:04	경고	정상	정상	정상
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
219.25	5.078125	0.0	379.25	0.0	0.0	0.0	220.50	5.566406	0.0	379.75	0.0	0.0	0.0	217.75	5.566406	154200.156	10.000	59.938183	2020-11-22 23:44	경고	정상	경고	경고
219.00	4.980469	0.0	378.75	0.0	0.0	0.0	220.25	5.273438	0.0	379.00	0.0	0.0	0.0	217.50	5.468750	154200.156	9.375	59.938210	2020-11-22 23:45	경고	정상	경고	경고
219.75	5.078125	0.0	380.00	0.0	0.0	0.0	221.00	5.273438	0.0	380.25	0.0	0.0	0.0	218.00	5.371094	154200.156	9.375	59.891422	2020-11-22 23:46	경고	정상	경고	경고
219.00	5.175781	0.0	378.75	0.0	0.0	0.0	220.25	5.468750	0.0	379.25	0.0	0.0	0.0	217.50	5.761719	154200.156	10.000	59.919464	2020-11-22 23:47	경고	정상	경고	경고
220.00	5.078125	0.0	380.75	0.0	0.0	0.0	221.50	5.371094	0.0	381.00	0.0	0.0	0.0	218.50	5.664063	154200.156	10.000	59.891422	2020-11-22 23:48	경고	정상	경고	경고

Ex)

역률평균label	전류고조파평균label	전압고조파평균label
주의	주의	정상

=> “주의” 로 판단

데이터 모델링

- Label 값이 모두 다른 경우

```
1 # 1 : 1 : 1 문제다
2 all_diff_label = data[(data['POWER_LABEL'] != data['ELECTRIC_LABEL']) & (data['POWER_LABEL'] != data['VOLTAGE_LABEL']) & (data['ELECTRIC_LABEL'] != data['VOLTAGE_LABEL'])]
3 display(all_diff_label)
4 all_diff_label_index = list(all_diff_label.index)
5 all_diff_label_index
6 all_diff_label = all_diff_label.reset_index(drop=True)
```

S상전압	S상전류	S상전압고조파	S상전류고조파	S상전압고조파	S상전류고조파	T상전압	T상전류	T상전압고조파	T상전류고조파	T상전압고조파	T상전류고조파	누적전압강	온도	주파수	TIMESTAMP	POWER_LABEL	ELECTRIC_LABEL	VOLTAGE_LABEL	FINAL_LABEL			
227.75	-0.738281	3803.0	39.21875	123.730469	131.50	3.515625	3505.0	226.50	0.753906	4044.0	40.62500	111.523438	131.25	3.222656	63.132298	26.250	59.863403	2021-01-16 06:49	정상	경고	주의	0
227.25	-0.738281	3802.0	39.15625	94.921875	131.25	4.101563	3540.0	225.50	0.750000	4036.0	41.43750	88.671875	130.75	3.906250	250.490417	25.625	59.844757	2021-01-16 06:50	정상	경고	주의	0
227.50	0.808594	4766.0	45.25000	77.148440	131.25	4.003906	3557.0	226.00	0.808594	4932.0	46.78125	73.339840	131.00	4.003906	479.197235	25.000	59.802084	2021-01-16 06:51	정상	경고	주의	0
227.00	0.804688	4556.0	43.31250	73.046875	131.00	4.199219	3476.0	226.25	0.804688	4761.0	45.43750	70.019530	131.00	4.003906	711.311500	26.250	59.891422	2021-01-16 06:52	정상	경고	주의	0
227.25	-0.726563	3601.0	38.50000	86.523440	131.25	4.199219	3470.0	225.50	0.746094	3914.0	40.18750	81.250000	130.75	3.808594	911.526800	26.250	59.882103	2021-01-16 06:53	정상	경고	주의	0
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--			
381.25	0.000000	0.0	0.00000	0.000000	221.75	4.296875	0.0	381.50	0.000000	0.0	0.00000	0.000000	219.00	4.980469	154200.156000	10.000	59.938183	2020-11-22 23:55	경고	정상	주의	0
380.50	0.000000	0.0	0.00000	0.000000	221.50	4.199219	0.0	380.75	0.000000	0.0	0.00000	0.000000	218.25	4.980469	154200.156000	10.000	59.910122	2020-11-22 23:56	경고	정상	주의	0
380.25	0.000000	0.0	0.00000	0.000000	221.25	4.296875	0.0	380.75	0.000000	0.0	0.00000	0.000000	218.25	4.980469	154200.156000	9.375	59.928820	2020-11-22 23:57	경고	정상	주의	0
380.25	0.000000	0.0	0.00000	0.000000	221.25	4.199219	0.0	380.75	0.000000	0.0	0.00000	0.000000	218.25	4.980469	154200.156000	9.375	59.928820	2020-11-22 23:58	경고	정상	주의	0
380.00	0.000000	0.0	0.00000	0.000000	221.00	4.394531	0.0	380.25	0.000000	0.0	0.00000	0.000000	218.25	5.078125	154200.156000	10.000	59.947537	2020-11-22 23:59	경고	정상	주의	0

역률 평균, 전류고조파평균, 전압고조파평균을 통해 진단한 label이 모두 다른 경우, 어떠한 label로도 설비의 이상을 판단할 수 없다. -> label 값을 예측해보자.

Ex)

역률평균label	전류고조파평균label	전압고조파평균label
주의	경고	정상
=> 판단 불가능.		

데이터 모델링

- Model Training

■ Data

R상전압 고조파	S상무 효전력	S상선 간전압	S상역률	S상 유효 전력	S상전 류	S상전류 고조파	S상전 압	S상전압 고조파	T상 무효 전력	T상선 간전압	T상역률	T상 유효 전력	T상전 류	T상전류 고조파	T상전 압	T상전압 고조파	누적전력 량	온도	주파수	TIMESTAMP	FINAL_LABEL	
3.320313	392.75	220.00	0.648438	336.0	4.0625	47.070313	127.25	3.710938	478.0	219.25	0.656250	418.25	5.03125	39.746094	126.75	3.808594	106472.211	25.000	59.872740	2021-01-19 13:21	주의	
2.734375	0.00	221.75	0.000000	0.0	0.0000	0.000000	128.00	2.832031	0.0	221.00	0.000000	0.00	0.00000	0.000000	128.00	2.636719	194950.875	30.625	59.863422	2021-01-20 16:10	정상	
2.636719	0.00	221.75	0.000000	0.0	0.0000	0.000000	127.75	2.734375	0.0	220.50	0.000000	0.00	0.00000	0.000000	127.75	2.539063	194950.875	30.000	59.872740	2021-01-20 16:11	정상	
2.832031	0.00	221.25	0.000000	0.0	0.0000	0.000000	127.75	2.832031	0.0	220.00	0.000000	0.00	0.00000	0.000000	127.25	2.539063	194950.875	30.000	59.882076	2021-01-20 16:12	정상	
3.515625	270.00	223.25	0.542969	176.5	2.4375	57.910156	128.75	3.320313	318.0	222.50	0.578125	225.75	3.09375	51.953125	128.75	3.515625	274251.156	33.125	59.891422	2021-01-21 16:05	주의	
...
5.078125	0.00	379.25	0.000000	0.0	0.0000	0.000000	220.50	5.566406	0.0	379.75	0.000000	0.00	0.00000	0.000000	217.75	5.566406	154200.156	10.000	59.938183	2020-11-22 23:44	경고	
4.980469	0.00	378.75	0.000000	0.0	0.0000	0.000000	220.25	5.273438	0.0	379.00	0.000000	0.00	0.00000	0.000000	217.50	5.468750	154200.156	9.375	59.938210	2020-11-22 23:45	경고	
5.078125	0.00	380.00	0.000000	0.0	0.0000	0.000000	221.00	5.273438	0.0	380.25	0.000000	0.00	0.00000	0.000000	218.00	5.371094	154200.156	9.375	59.891422	2020-11-22 23:46	경고	
5.175781	0.00	378.75	0.000000	0.0	0.0000	0.000000	220.25	5.468750	0.0	379.25	0.000000	0.00	0.00000	0.000000	217.50	5.761719	154200.156	10.000	59.919464	2020-11-22 23:47	경고	
5.078125	0.00	380.75	0.000000	0.0	0.0000	0.000000	221.50	5.371094	0.0	381.00	0.000000	0.00	0.00000	0.000000	218.50	5.664063	154200.156	10.000	59.891422	2020-11-22 23:48	경고	

라벨이 모두 같은 경우와
2개의 라벨이 같은 경우에서
판단한 최종 판단 라벨이
포함된 데이터를 활용하여
라벨이 모두 다른 경우의
설비 이상 라벨을 예측한다.

데이터 모델링

- Model Training

■ Model Code

```
1 from sklearn.metrics import accuracy_score
2 from imblearn.over_sampling import SMOTE
3 from sklearn.model_selection import GridSearchCV
4 from sklearn.ensemble import RandomForestClassifier
5 from sklearn.model_selection import train_test_split
6 from sklearn.model_selection import cross_val_score
7 from sklearn.preprocessing import LabelEncoder
8
9 encoder = LabelEncoder()
10 model_data['FINAL_LABEL'] = encoder.fit_transform(model_data['FINAL_LABEL'])
11 model_data['FINAL_LABEL'] = model_data['FINAL_LABEL'].astype('category')
12
13
14 x_train, x_test, y_train, y_test = train_test_split(x_data, y_data, random_state = 0, stratify = y_data) # y_data에 비율에 맞춰서 모델을 만들어줌
15
16 params = {
17     'n_estimators': [500],
18     'max_depth': [6, 7, 8],
19     'min_samples_leaf': [6, 7, 8],
20     'min_samples_split': [8, 10, 12]
21 }
22
23 smote = SMOTE(random_state = 0)
24
25 x_train_over, y_train_over = smote.fit_sample(x_train, y_train) # SMOTE해서 불균형 데이터 없애줌
26
27
28 grid_cv = GridSearchCV(model, param_grid = params, cv = 3, n_jobs = -1)
29 grid_cv.fit(x_train, y_train)
30
31 print('최적 하이퍼 파라미터: ', grid_cv.best_params_)
32 print('최고 예측 정확도: {:.4f}'.format(grid_cv.best_score_))
33
```

정확도 : 0.9960641840750832

=> 높은 정확도로 인해 overfitting이 의심된다.

최적 하이퍼 파라미터: {'max_depth': 7, 'min_samples_leaf': 6, 'min_samples_split': 8, 'n_estimators': 500} 최고 예측 정확도: 0.9956

결과 분석

■ 검증 데이터 생성

```
3 power_val = pd.read_csv('new역률평균_val')
4 elec_val = pd.read_csv('new전류평균_val')
5 volt_val = pd.read_csv('new전압평균_val')
6
7 power_val.drop(['무효전력평균', '상전압평균', '선간전압평균', '역률평균', '유효전력평균', '전류고조파평균', '전류평균', '전압고조파평균'], axis = 1, inplace = True)
8 elec_val.drop(['무효전력평균', '상전압평균', '선간전압평균', '역률평균', '유효전력평균', '전류고조파평균', '전류평균', '전압고조파평균'], axis = 1, inplace = True)
9 volt_val.drop(['무효전력평균', '상전압평균', '선간전압평균', '역률평균', '유효전력평균', '전류고조파평균', '전류평균', '전압고조파평균'], axis = 1, inplace = True)
```

■ 데이터 가공 후 예측

```
1 from sklearn.metrics import f1_score, confusion_matrix
2 from sklearn import metrics
3
4 prediction = model.predict(x_val)
5 print('정확도 : \n', accuracy_score(prediction, y_val))
6
```

정확도 :
0.6103395061728395

결과 분석

```
1 final_x_data=all_diff_label.drop(['TIMESTAMP','POWER_LABEL','ELECTRIC_LABEL','VOLTAGE_LABEL','FINAL_LABEL'],axis = 1)
2 prediction2 = model.predict(final_x_data)
3 all_diff_label['FINAL_LABEL'] = prediction2
```

온도	주파수	TIMESTAMP	POWER_LABEL	ELECTRIC_LABEL	VOLTAGE_LABEL	FINAL_LABEL
10.000	59.947544	2020-10-24 01:24	경고	정상	주의	0
11.875	59.938198	2020-10-24 01:25	경고	정상	주의	0
10.000	59.910110	2020-10-24 02:54	경고	정상	주의	0
8.750	59.966270	2020-10-24 05:21	경고	정상	주의	0
8.125	59.919510	2020-10-24 05:22	경고	정상	주의	0
...
25.000	59.882084	2021-02-04 22:32	경고	정상	주의	1
23.750	59.863403	2021-02-04 22:33	경고	정상	주의	1
23.750	59.882084	2021-02-04 22:34	경고	정상	주의	1
24.375	59.910122	2021-02-04 22:35	경고	정상	주의	1
24.375	59.872750	2021-02-04 22:36	경고	정상	주의	1

결과 분석

■ 맵핑

```
1 label_dict = {0:'경고',1:'정상',2:'주의'}  
2 all_diff_label['FINAL_LABEL'] = all_diff_label['FINAL_LABEL'].map(label_dict)
```

TIMESTAMP	POWER_LABEL	ELECTRIC_LABEL	VOLTAGE_LABEL	FINAL_LABEL
2020-10-24 01:24	경고	정상	주의	경고
2020-10-24 01:25	경고	정상	주의	경고
2020-10-24 02:54	경고	정상	주의	경고
2020-10-24 05:21	경고	정상	주의	경고
2020-10-24 05:22	경고	정상	주의	경고
...
2021-02-04 22:32	경고	정상	주의	정상
2021-02-04 22:33	경고	정상	주의	정상
2021-02-04 22:34	경고	정상	주의	정상
2021-02-04 22:35	경고	정상	주의	정상
2021-02-04 22:36	경고	정상	주의	정상

결과 분석

■ 최종

```
1 elec_count += all_diff_label[all_diff_label['ELECTRIC_LABEL'] == all_diff_label['FINAL_LABEL']].shape[0]
2 volt_count += all_diff_label[all_diff_label['VOLTAGE_LABEL'] == all_diff_label['FINAL_LABEL']].shape[0]
3 power_count += all_diff_label[all_diff_label['POWER_LABEL'] == all_diff_label['FINAL_LABEL']].shape[0]
4
5 sum = elec_count + volt_count + power_count
6 print(f'전류고조파 비율 : {elec_count / sum * 100 :.2f} 전압고조파 비율 : {volt_count / sum * 100 :.2f} 역률 비율 : {power_count / sum * 100 :.2f}')
```

전류고조파 비율 : 35.64 전압고조파 비율 : 39.72 역률 비율 : 24.64

설비 이상 판단에 있어 중요한 요소의 순위를 파악해 볼 수 있다.

전압고조파 > 전류고조파 > 역률 순으로 가중치를 부여하여 다음 설비 이상 판단에 유용하게 사용할 수 있을 것이라고 기대한다.

기대효과

- 1) 개별 값이 아닌 하나의 값으로 판단되는 설비 이상 판단이 가능하다.
- 2) 다른 설비 데이터를 활용하여 설비 종류별 정밀한 이상 감지가 가능하다.
- 3) 판단에서 나아가 예측 모델을 통해 전기료 절감 등의 솔루션을 제안할 수 있다.
- 4) 추가 연구를 통해 설비 이상에 영향을 줄 것이라고 판단되는 날씨 같은 외부 데이터를 추가하여 판단의 정확도를 높일 수 있다.

감사합니다.