

현재 생각하고 있는 모델 : 출강 여부(0,1 이진 분류) 예측하는 모델

주어진 데이터 : 현재 분 단위로 측정된 여러 변수들의 값(전력 사용량 ~ 여러 온도)이 주어지고, 각 시점에서 **출강여부**(0 or 1)의 타겟값이 주어진 상태입니다.

히트NO. = 하나의 출강 과정에 대한 ID값입니다.
총 unique 히트NO.는 950개입니다. 즉, 950번의 출강이 이루어진 것입니다.

히트 NO.	date	time	전력 사용량	로제냉각수유량	로제 입구온도	로제 출구온도	천정 냉각수 유량	천정 입구온도	천정 출구온도	수냉덕트 입구온도	수냉덕트 출구온도	공랭덕트입구온도	로제 입출구 온도차이	천정 입출구 온도차이	수냉덕트 입출구온도차이	출강 여부	
0	0	2021-01-08	16:24:00	0.0	640.0	21	21.0	428.0	21	26.0	21	23.0	NaN	0.0	5.0	2.0	0
1	0	2021-01-08	16:25:00	0.0	640.0	21	21.0	428.0	21	26.0	21	23.0	175.833951	0.0	5.0	2.0	0
2	0	2021-01-08	16:26:00	0.0	640.0	21	21.0	427.0	21	25.0	21	23.0	162.490734	0.0	4.0	2.0	0
3	0	2021-01-08	16:27:00	0.0	639.0	21	21.0	427.0	21	25.0	21	23.0	148.851001	0.0	4.0	2.0	0
4	0	2021-01-08	16:28:00	0.0	639.0	21	21.0	425.0	21	24.0	21	23.0	134.914752	0.0	3.0	2.0	0
...
61	0	2021-01-08	17:25:00	45.8	640.0	22	24.0	431.0	22	28.0	22	26.0	277.538918	2.0	6.0	4.0	0
62	0	2021-01-08	17:26:00	46.9	639.0	22	32.0	440.0	22	28.0	22	26.0	276.649370	10.0	6.0	4.0	0
63	0	2021-01-08	17:27:00	47.3	640.0	22	28.0	428.0	22	28.0	22	26.0	287.027428	6.0	6.0	4.0	1
64	0	2021-01-08	17:28:00	47.3	640.0	22	24.0	431.0	22	28.0	22	26.0	284.951816	2.0	6.0	4.0	1
65	0	2021-01-08	17:29:00	47.3	640.0	22	23.0	430.0	22	27.0	22	26.0	272.201631	1.0	5.0	4.0	1

66 rows x 17 columns

출강여부 : 타겟값입니다.
모든 경우에서 마지막 3분이 출강 가능 시점(1)이고, 나머지는 0(출강 불가능)입니다.

히트 NO.	date	time	전력 사용량	로제냉각수유량	로제 입구온도	로제 출구온도	천정 냉각수 유량	천정 입구온도	천정 출구온도	수냉덕트 입구온도	수냉덕트 출구온도	공랭덕트입구온도	로제 입출구온도차이	천정 입출구온도차이	수냉덕트 입출구온도차이	출강 여부	
132	2	2021-01-08	19:32:00	0.0	640.0	23	24.0	430.0	23	29.0	23	27.0	200.444774	1.0	6.0	4.0	0
133	2	2021-01-08	19:33:00	0.0	640.0	23	24.0	430.0	23	29.0	23	27.0	192.438844	1.0	6.0	4.0	0
134	2	2021-01-08	19:34:00	0.0	640.0	23	23.0	429.0	23	29.0	23	27.0	181.467754	0.0	6.0	4.0	0
135	2	2021-01-08	19:35:00	0.0	640.0	23	23.0	424.0	23	28.0	23	26.0	165.455893	0.0	5.0	3.0	0
136	2	2021-01-08	19:36:00	0.0	638.0	23	23.0	423.0	23	27.0	23	26.0	151.519644	0.0	4.0	3.0	0
...
202	2	2021-01-08	20:42:00	51.9	636.0	25	34.0	432.0	25	36.0	25	30.0	412.750185	9.0	11.0	5.0	0
203	2	2021-01-08	20:43:00	52.4	639.0	25	32.0	430.0	25	35.0	25	30.0	423.128243	7.0	10.0	5.0	0
204	2	2021-01-08	20:44:00	52.4	638.0	25	27.0	427.0	25	36.0	25	30.0	400.593032	2.0	11.0	5.0	1
205	2	2021-01-08	20:45:00	52.4	639.0	25	27.0	428.0	25	36.0	25	29.0	365.011119	2.0	11.0	4.0	1
206	2	2021-01-08	20:46:00	52.4	639.0	25	26.0	425.0	25	35.0	25	29.0	329.132691	1.0	10.0	4.0	1

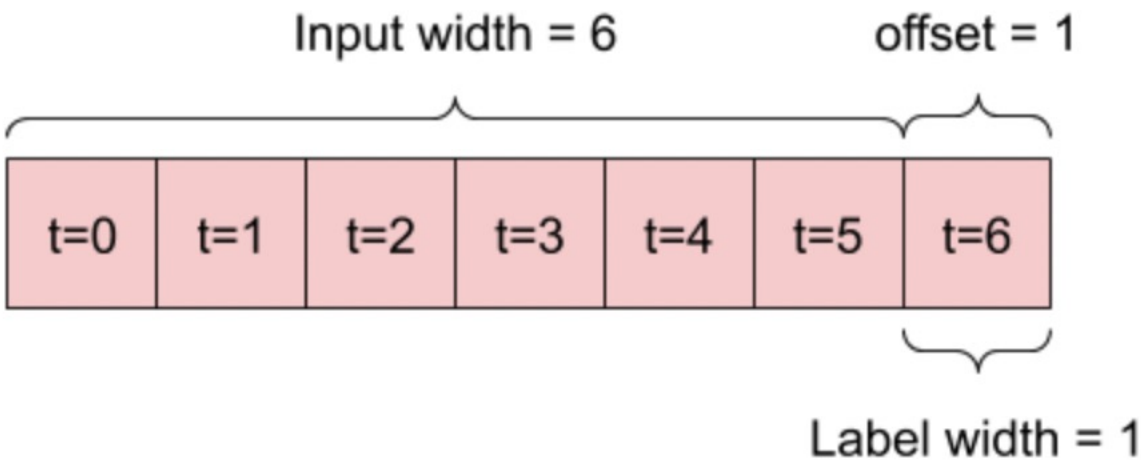
75 rows x 17 columns

- 위에 샘플로 가져온 두 개의 히트NO.를 보면, 각각 총 66분, 75분의 데이터가 축적된 것입니다.
- 이러한 히트NO.가 총 950개 있습니다. (데이터는 계속 추가될 예정입니다.)

하나의 **히트** 과정이란, 처음 철강물들을 투입하고 녹이는 과정입니다.
출강 여부는, 현업에서 일하시는 분들이 녹여진 내용물의 온도를 측정해 1600도 정도이면 출강(1), 아니면 계속 진행(0)으로 판단합니다.

이러한 하나 하나 **히트NO.**의 시계열 패턴을 학습하여,
새로운 **히트NO.**의 데이터가 들어왔을 때 출강 가능한 시점인지에 대한 모델링을 고민하고 있습니다.

예를 들어 아래 그림처럼 window size를 6으로 두어, 6분의 데이터가 주어졌을 때 7분째에 0 or 1 예측을 해볼 수 있지 않을까 싶어서요!



- 문제는,
- 1. 각각의 히트NO.가 독립적이라는 것입니다. 이런 경우 각각의 패턴을 어떻게 하나의 모델이 학습할 수 있을지 계속 찾아보고 있습니다. (히트NO.가 바뀌는 row를 경계로 padding을 넣는 것을 생각해 봤습니다.)
 - 2. 모든 히트NO. 경우에서 0000...000111 이러한 양상입니다.
 - 3. 1의 비율 전체 데이터 중 5% 정도밖에 안됩니다.

```
data_pp[ '출강여부' ].value_counts()
0      47784
1       2808
Name: 출강여부, dtype: int64
```

모델 훈련시

각각의 히트NO.의 패턴을 학습한다.

히트 NO.	date	time	전역 사용량	로제넙각 수량량	로제 입구온도	로제 출구온도	현정 냉각 수량량	현정 냉각구온도	현정 출구온도	수냉액트 입구온도	수냉액트 출구온도	공행액트입구온도	로제 입출구온도차이	현정 입출구온도차이	수냉액트 입출구온도차이	출감여부	
0	0	2021-01-08	16:24:00	0.0	640.0	21	21.0	428.0	21	26.0	21	23.0	NaN	0.0	5.0	2.0	0
1	0	2021-01-08	16:25:00	0.0	640.0	21	21.0	428.0	21	26.0	21	23.0	175.833951	0.0	5.0	2.0	0
2	0	2021-01-08	16:26:00	0.0	640.0	21	21.0	427.0	21	25.0	21	23.0	162.490734	0.0	4.0	2.0	0
3	0	2021-01-08	16:27:00	0.0	639.0	21	21.0	427.0	21	25.0	21	23.0	148.851001	0.0	4.0	2.0	0
4	0	2021-01-08	16:28:00	0.0	639.0	21	21.0	425.0	21	24.0	21	23.0	134.914752	0.0	3.0	2.0	0
...
61	0	2021-01-08	17:25:00	45.8	640.0	22	24.0	431.0	22	28.0	22	26.0	277.538918	2.0	6.0	4.0	0
62	0	2021-01-08	17:26:00	46.9	639.0	22	32.0	440.0	22	28.0	22	26.0	276.649370	10.0	6.0	4.0	0
63	0	2021-01-08	17:27:00	47.3	640.0	22	28.0	428.0	22	28.0	22	26.0	287.027428	6.0	6.0	4.0	1
64	0	2021-01-08	17:28:00	47.3	640.0	22	24.0	431.0	22	28.0	22	26.0	284.951816	2.0	6.0	4.0	1
65	0	2021-01-08	17:29:00	47.3	640.0	22	23.0	430.0	22	27.0	22	26.0	272.201631	1.0	5.0	4.0	1

66 rows × 17 columns

input

히트 NO.	date	time	전역 사용량	로제넙각 수량량	로제 입구온도	로제 출구온도	현정 냉각 수량량	현정 입구온도	현정 출구온도	수냉액트 입구온도	수냉액트 출구온도	공행액트입구온도	로제 입출구온도차이	현정 입출구온도차이	수냉액트 입출구온도차이	출감여부	
66	1	2021-01-08	18:26:00	0.1	639.0	22	22.0	428.0	22	23.0	22	22.0	40.326168	0.0	1.0	0.0	0
67	1	2021-01-08	18:27:00	0.1	640.0	22	22.0	430.0	22	23.0	22	23.0	41.512231	0.0	1.0	1.0	0
68	1	2021-01-08	18:28:00	0.1	640.0	22	22.0	426.0	22	23.0	22	23.0	41.215715	0.0	1.0	1.0	0
69	1	2021-01-08	18:29:00	0.1	638.0	22	22.0	428.0	22	23.0	22	23.0	38.843588	0.0	1.0	1.0	0
70	1	2021-01-08	18:30:00	1.2	640.0	22	23.0	436.0	22	23.0	22	23.0	43.587843	1.0	1.0	1.0	0
...
127	1	2021-01-08	19:27:00	47.2	639.0	23	26.0	438.0	23	33.0	23	26.0	182.357302	3.0	10.0	3.0	0
128	1	2021-01-08	19:28:00	48.4	640.0	23	27.0	437.0	23	33.0	23	27.0	186.508525	4.0	10.0	4.0	0
129	1	2021-01-08	19:29:00	49.6	640.0	23	28.0	438.0	23	32.0	23	30.0	196.293551	5.0	9.0	7.0	1
130	1	2021-01-08	19:30:00	49.8	640.0	23	25.0	428.0	23	31.0	23	29.0	206.968125	2.0	8.0	6.0	1
131	1	2021-01-08	19:31:00	49.8	639.0	23	24.0	429.0	23	30.0	23	28.0	209.043736	1.0	7.0	5.0	1

66 rows × 17 columns

input

	히트 NO.	date	time	전역 사용량	로제넙각 수량량	로제 입구온도	로제 출구온도	현정 냉각 수량량	현정 입구온도	현정 출구온도	수냉액트 입구온도	수냉액트 출구온도	공행액트입구온도	로제 입출구온도차이	현정 입출구온도차이	수냉액트 입출구온도차이	출감 여부
132	2	2021-01-08	19:32:00	0.0	640.0	23	24.0	430.0	23	29.0	23	27.0	200.444774	1.0	6.0	4.0	0
133	2	2021-01-08	19:33:00	0.0	640.0	23	24.0	430.0	23	29.0	23	27.0	192.438844	1.0	6.0	4.0	0
134	2	2021-01-08	19:34:00	0.0	640.0	23	23.0	429.0	23	29.0	23	27.0	181.467754	0.0	6.0	4.0	0
135	2	2021-01-08	19:35:00	0.0	640.0	23	23.0	424.0	23	28.0	23	26.0	165.456893	0.0	5.0	3.0	0
136	2	2021-01-08	19:36:00	0.0	638.0	23	23.0	423.0	23	27.0	23	26.0	151.519644	0.0	4.0	3.0	0
...
202	2	2021-01-08	20:42:00	51.9	636.0	25	34.0	432.0	25	36.0	25	30.0	412.750185	9.0	11.0	5.0	0
203	2	2021-01-08	20:43:00	52.4	639.0	25	32.0	430.0	25	35.0	25	30.0	423.128243	7.0	10.0	5.0	0
204	2	2021-01-08	20:44:00	52.4	638.0	25	27.0	427.0	25	36.0	25	30.0	400.593032	2.0	11.0	5.0	1
205	2	2021-01-08	20:45:00	52.4	639.0	25	27.0	428.0	25	36.0	25	29.0	365.011119	2.0	11.0	4.0	1
206	2	2021-01-08	20:46:00	52.4	639.0	25	26.0	425.0	25	35.0	25	29.0	329.132691	1.0	10.0	4.0	1

75 rows × 17 columns

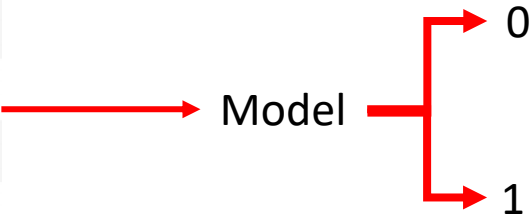
input

Train Model

모델 예측시

예를 들어 5개의 시계열 데이터가 들어오면,
출강 여부를 예측한다.

히트 NO.	date	time	전력 사용량	로체냉각 수유량	로체 입 구온도	로체 출 구온도	천정 냉각 수 유량	천정 입 구온도	천정 출 구온도	수냉덕트 입구온도	수냉덕트 출구온도	공랭덕트입구 온도	로체 입출 구온도차이	천정 입출 구온도차이	수냉덕트 입출 구온도차이
900	2021-02-27	01:12:00	46.6	633.0	31	31.0	410.0	31	36.0	31	34.0	318.458117	0.0	5.0	3.0
900	2021-02-27	01:13:00	0.0	633.0	31	31.0	409.0	31	36.0	31	34.0	289.696071	0.0	5.0	3.0
900	2021-02-27	01:14:00	0.0	633.0	31	31.0	409.0	31	35.0	31	34.0	260.934025	0.0	4.0	3.0
900	2021-02-27	01:15:00	0.0	631.0	31	31.0	407.0	31	36.0	31	33.0	238.398814	0.0	5.0	2.0
900	2021-02-27	01:16:00	0.1	632.0	31	31.0	412.0	31	35.0	31	33.0	217.642698	0.0	4.0	2.0



예측할 데이터