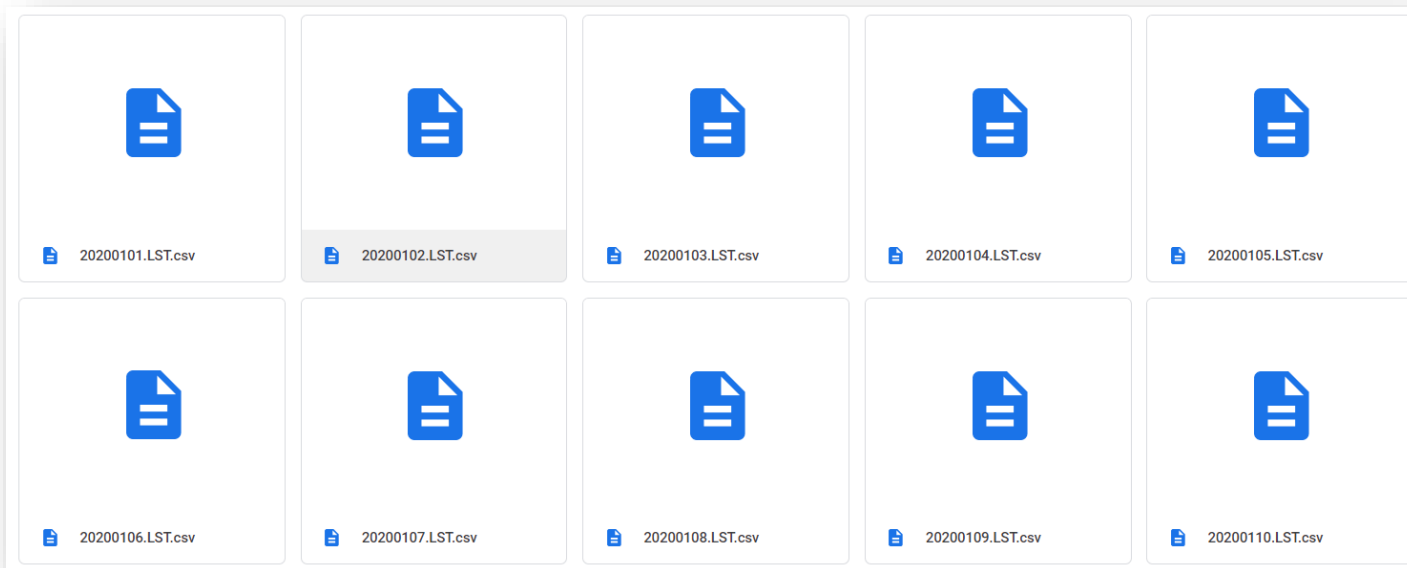


## 01

## 데이터 설명 (미래예측 알고리즘)

- ▶ 2020, 2021년도 24개월 (730일) 데이터.
- ▶ 하루에 726개 관측 지점 별 10분 단위로 구성.
- ▶ 하루에(한 csv에) 약 10만개, 한 지점에 144개 데이터로 구성.
- ▶ 총 44개의 변수. (2개 종속변수)



### 1. 다운샘플링

- 종속변수 지면온도와 지상온도 모두 90583개 결측치 존재.
- 결측치를 처리하기 위해 기상청 데이터 값으로 직접 채워넣음.
- 대체할 기상청 데이터는 정각마다의 데이터만 존재하기 때문에 10분 단위가 아닌 1시간 단위로 downsampling 진행.

### 2. 19개 독립 변수 지정

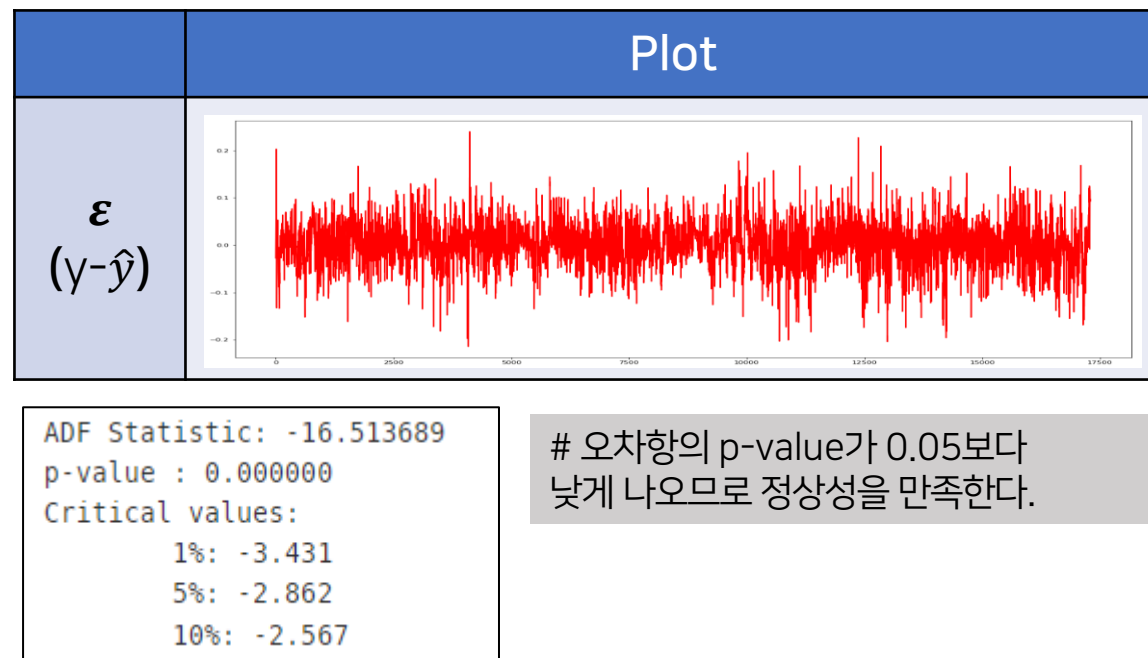
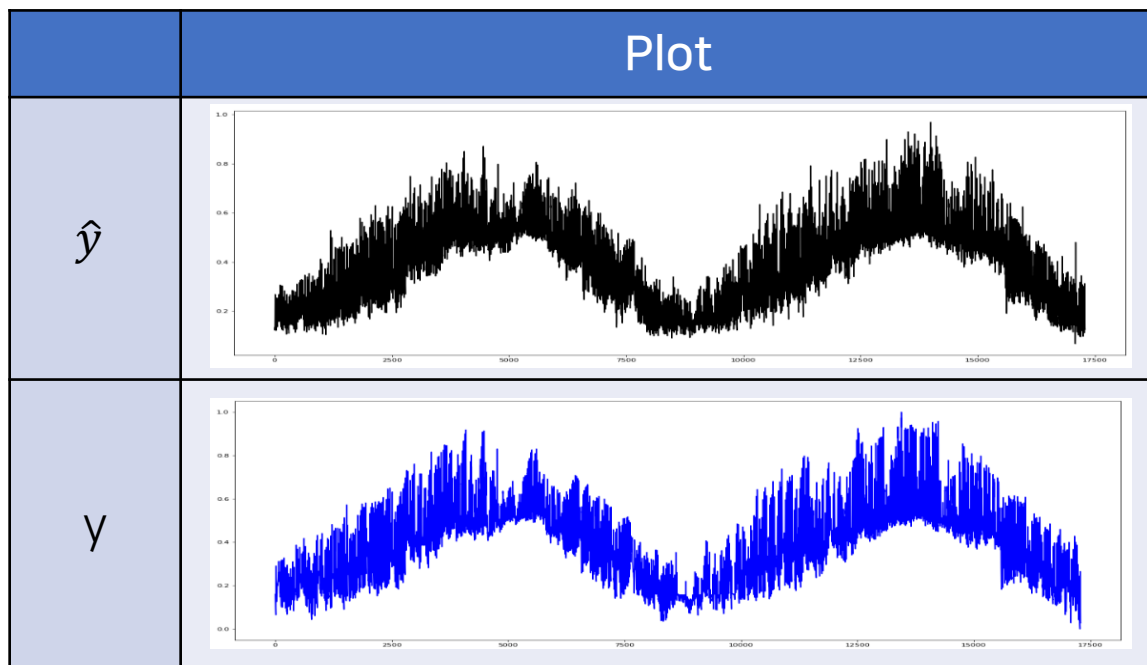
- 기상청에서 관측 가능한 데이터를 제외하고, 기상위성에서 얻은 자료로만 결정.

# 시계열 데이터 정상성 검정

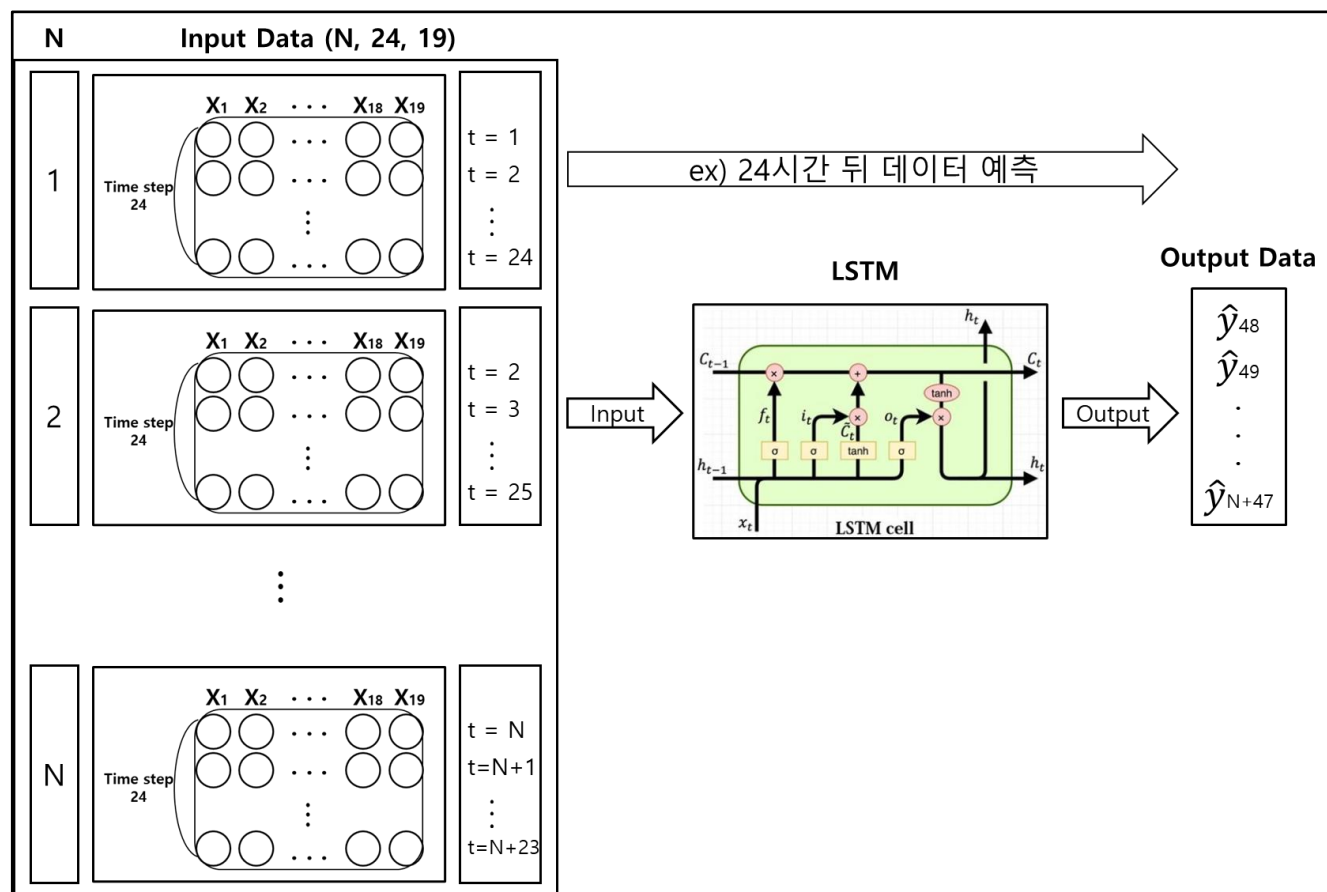
# ADF Test(단위근 검정) 사용: "단위근을 포함하는지 = 비정상 시계열 인지" 판단

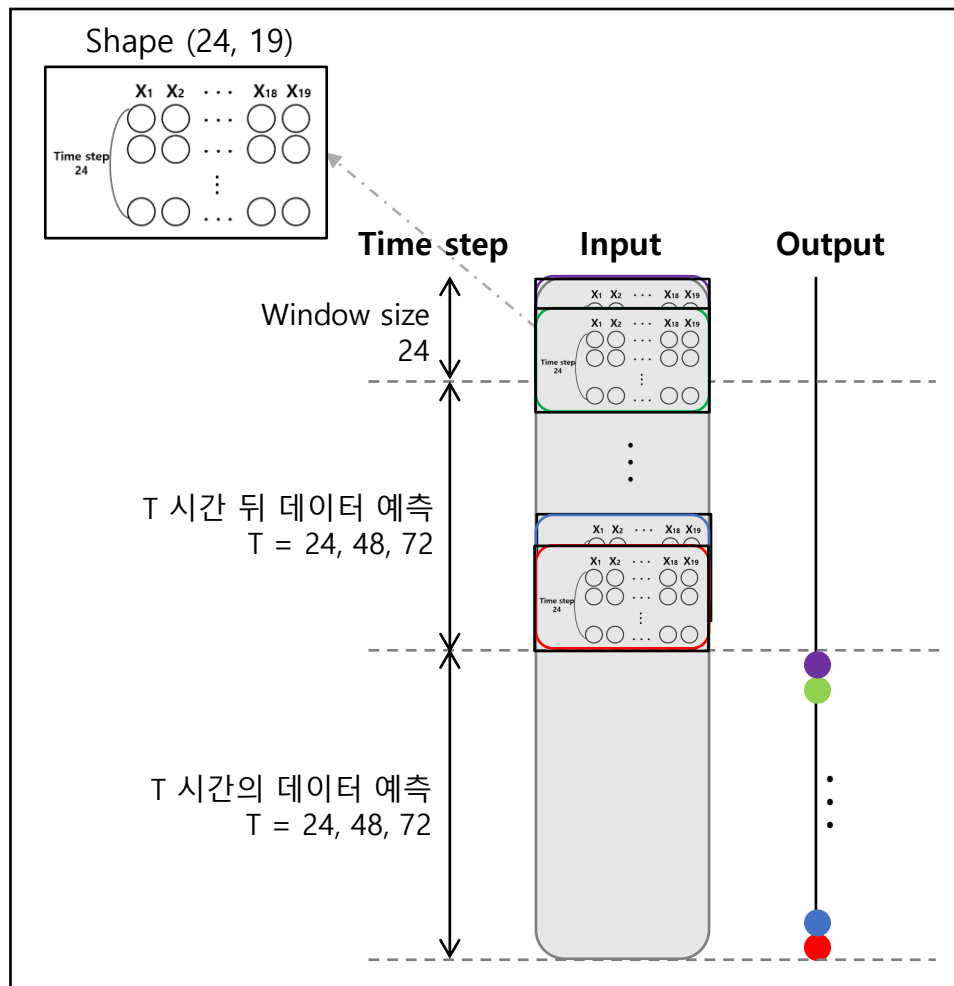
# 귀무가설  $H_0 = \alpha = 1$ 로 설정

# 방법: 예측값  $\hat{y}$ 과  $y$ 의 차이인 오차항을 이용



# All Shape (17483, 24, 19)





```
def make_dataset_24(data, label, window_size):
    feature_list = []
    label_list = []
    for i in range(len(data) - window_size):
        feature_list.append(np.array(data.iloc[i:i + window_size]))

    for i in range(len(data) - window_size - 24):
        label_list.append(np.array(label.iloc[i+24 + window_size]))

    return np.array(feature_list), np.array(label_list)
```

**i + 24**

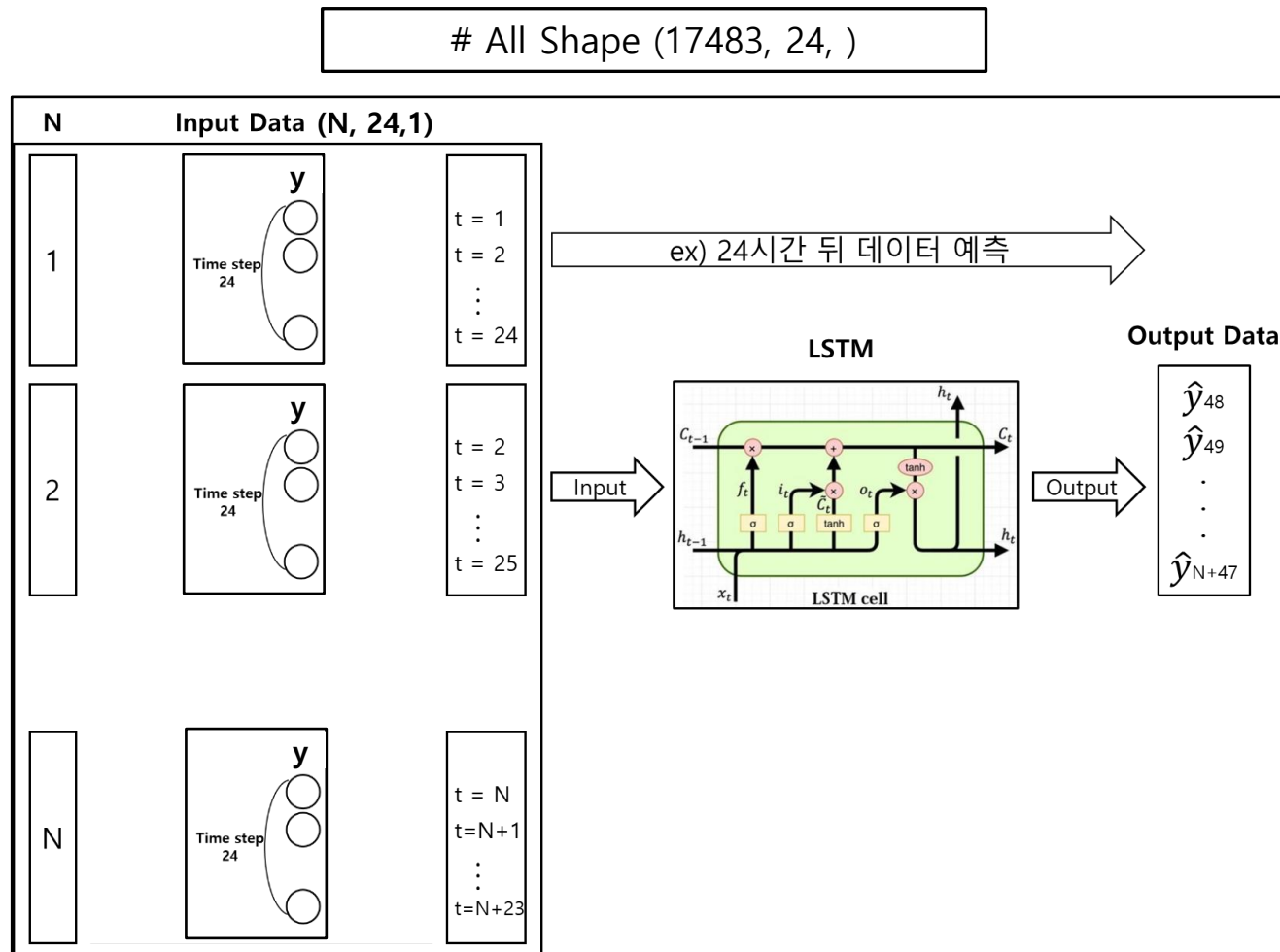
```
print(x_train.shape, y_train.shape, x_test.shape, y_test.shape)
(17411, 24, 19) (17411,) (24, 24, 19) (24,)
```

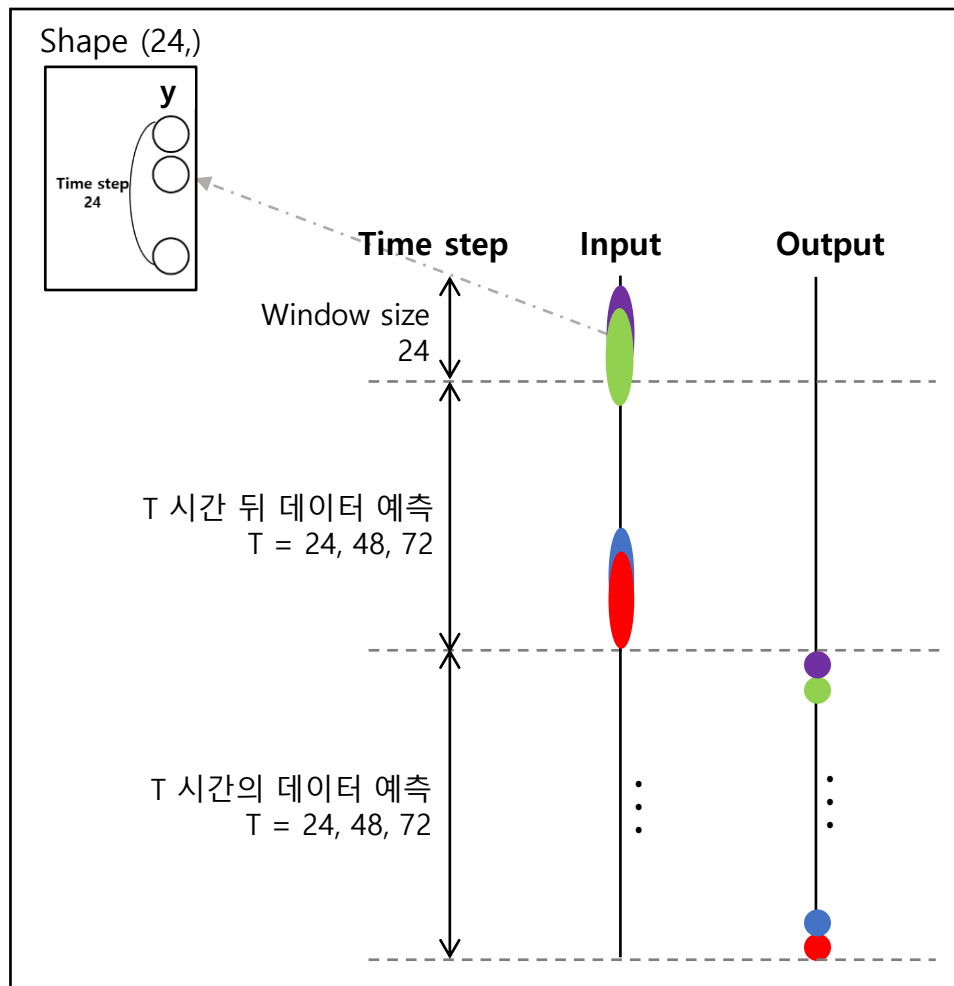
**i + 48**

```
print(x_train.shape, y_train.shape, x_test.shape, y_test.shape)
(17363, 24, 19) (17363,) (48, 24, 19) (48,)
```

**i + 72**

```
print(x_train.shape, y_train.shape, x_test.shape, y_test.shape)
(17315, 24, 19) (17315,) (72, 24, 19) (72,)
```





```
def windowed_dataset_24(series, window_size, batch_size):
    series = tf.expand_dims(series, axis=-1)
    ds = tf.data.Dataset.from_tensor_slices(series)
    ds = ds.window(window_size + 24, shift = 1, drop_remainder = True)
    ds = ds.flat_map(lambda w: w.batch(window_size + 24)) # window_size 시간의 데이터로 24시간 뒤 예측

    ds = ds.map(lambda w: (w[-24], w[-1]))
    print(ds)

    return ds.batch(batch_size).prefetch(1)
```

# 예측 모델이 최선의 모델인가?

- "위성 자료"만을 이용하여 분석 -> 종속변수와의 상관성이 낮은 독립변수들을 활용
- Training Data -> 20년, 21년 2년의 제공받은 데이터만 사용 가능
- 날씨의 변동성이 크기 때문에 완벽하게 예측하기에 한계가 있음



## # 결론

- 다양한 방법으로 기상 자료(지면온도)를 수집하고 있지만 결측값이 많이 존재
- 프로젝트에서 개발된 예측 모델을 통해 실제값에 실제 지면 온도에 근접한 추정 값을 산출

## # 기대효과

- 지면 온도 예측으로 가뭄, 태풍 등 자연재해의 사전적인 예방
- 기후변화에 대한 환경 정책을 마련, 일상생활에서의 이용 등 다양한 분야에서 폭넓게 사용
- 예측된 지면 온도를 사용하여 체감온도, 폭염 및 열대야 발생 여부와 발생 확률을 예측하는 시스템을 다른 지표들과 결합하여 개발할 수 있으며 해당 예측시스템을 통해 각각의 분야에서 선제적으로 인명피해 및 재산피해를 최소화할 수 있을 것