



Busan science high school

2023 Ocean ICT Festival B

2023 BOIF 33

Youtube 영상 QR

인공지능을 활용한 해양 쓰레기 분석

팀 S=S

2418 황선우
2419 황준영

1. 탐구 동기

지구과학 시간에 워커 순환에 대해 배우고 더 나아가 엘니뇨와 라니냐에 대해 배웠는데, 해수면 온도에 따라서 엘니뇨와 라니냐를 대표하는 지수 중 하나인 기압을 예측해보고 싶어서 이 주제를 선택했다.

2. 이론적 배경

ONI지수: 지구 해수면 온도 편차, 특히 NINO 3.4 구역을 나타내는 지표
SOI지수: 해수면 기압 편차, 특히 NINO 3.4 구역을 나타내는 지표
선형 회귀: 독립 변수와 그에 대한 종속 변수 간의 관계를 선형식으로 분석하는 방법

3. 코드 작동 과정

1. 데이터 입력

:미국 해양대기청(NOAA)에서 ONI 지수와 SOI 지수의 데이터를 가져와서 이를 분류 후 1차원 리스트로 만든다.

중간 과정. 데이터 분석

:입력받은 데이터를 선형 회귀 등의 방식으로 상관관계를 조사한다.

2. 사용자로부터 데이터 입력

: 사용자에게 년도와 월을 입력받는다.

3. 시각화

: 입력받은 데이터를 바탕으로 기압을 통해 구름 양을 시각화한다.

4. 코드

4-1. 데이터 추출 및 선형 회귀

```
import re
import numpy as np

def get_column_data_from_SOI(file_path, column_index):
    column_data = []
    with open(file_path, 'r') as file:
        lines = file.readlines()
        for line in lines[1:]:
            values = re.findall(r'[-+]?%d+%.*?%d+|[-+]?%d+', line)
            if len(values) > column_index:
                column_data.append(values[column_index])
    return column_data

def get_column_data_from_ONI(file_path, column_name):
    column_data = []
    with open(file_path, 'r') as file:
        lines = file.readlines()
        if lines:
            headers = lines[0].split()
            if column_name in headers:
                column_index = headers.index(column_name)
                for line in lines[1:]:
                    values = line.split()
                    if len(values) > column_index:
                        column_data.append(values[column_index])
    return column_data
```

```
SOI=[]
file_SOI = 'C:/Users/User/Desktop/부락/2학년/오선 2023/기압차(SOI).txt'
file_ONI = 'C:/Users/User/Desktop/부락/2학년/오선 2023/해수면 온도(ONI).txt' # 실제 파일 경로로 변경해야 합니다.
for i in range(1,13):
    column_index = i
    result_SOI = get_column_data_from_SOI(file_SOI, column_index)
    SOI.append(result_SOI)
SOI1 = [item for sublist in zip(*SOI) for item in sublist]
#print(SOI1)
column_name = 'ANOM'
result_ONI = get_column_data_from_ONI(file_ONI, column_name)
#print(result_ONI)
#print(len(SOI1))
#print(len(result_ONI))

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
X=result_ONI
X_resaped=np.array(X).reshape(-1, 1)
Y=SOI1
model=LinearRegression()
x_train, x_test, y_train, y_test = train_test_split(X_resaped, Y, test_size=0.3, random_state=5)
model.fit(x_train, y_train)
predicted_y_train=model.predict(x_train)
predicted_y_test=model.predict(x_test)
a=model.coef_[0]
b=model.intercept_
#print('회수 기울기: %f'%beta_0)
#print('절편: %f'%beta_2)
```

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
import numpy as np
from sklearn.metrics import r2_score
X=result_ONI
X_resaped=np.array(X).reshape(-1, 1)
Y=SOI1
model=LinearRegression()
x_train, x_test, y_train, y_test = train_test_split(X_resaped, Y, test_size=0.3, random_state=5)
model.fit(x_train, y_train)
predicted_y_train=model.predict(x_train)
mse_train=mean_squared_error(y_train,predicted_y_train)
print("MSE on train data: %f"%mse_train)
beta_0=model.coef_[0] #회수 기울기
beta_2=model.intercept_ #절편
print('회수 기울기: %f'%beta_0)
print('절편: %f'%beta_2)
predicted_y_test=model.predict(x_test)
mse_test=mean_squared_error(y_test,predicted_y_test)
print("MSE on test data: %f"%mse_test)
r2_train = r2_score(y_train, predicted_y_train)
r2_test = r2_score(y_test, predicted_y_test)
print("train의 결정계수: %f"%r2_train)
print("test의 결정계수: %f"%r2_test)
```

4-2-1. 결정 계수 분석

MSE on train data: 1.122632
회수 기울기: -1.319357
절편: 0.273216
MSE on test data: 1.144984
train의 결정계수: 0.517979
test의 결정계수: 0.565168

4-2. 데이터 분석

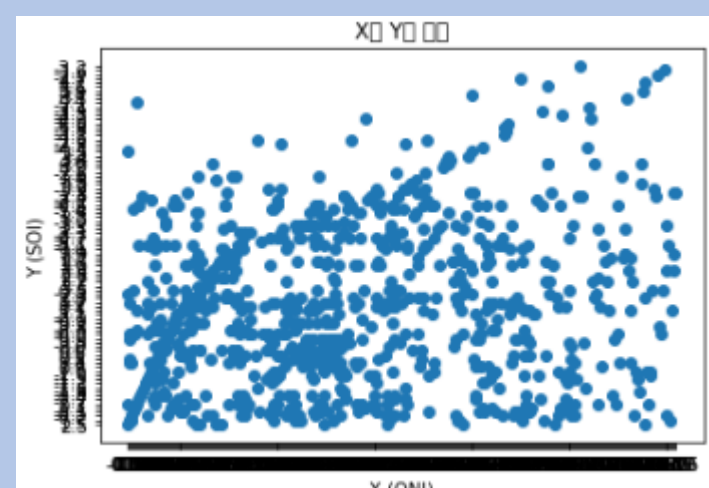
4-2-2. 산점도로 시각화

```
X = result_ONI
Y = SOI1
print(len(SOI1))
print(len(result_ONI))

data = pd.DataFrame({'X': X, 'Y': Y})

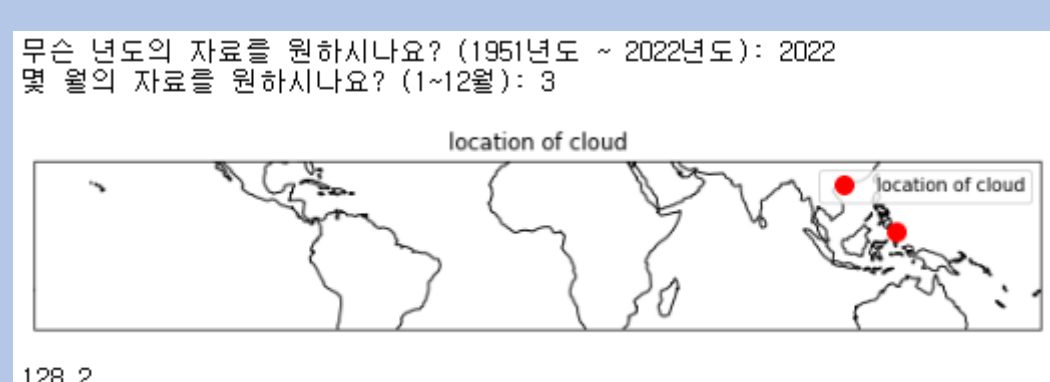
print(data)

plt.scatter(data['X'], data['Y'])
plt.xlabel('X (ONI)')
plt.ylabel('Y (SOI)')
plt.title('X와 Y의 관계')
plt.show()
```



4-3. 입력받은 값에 대한 시각화

결과



5. 관련 교과 및 고찰

우선 워커 순환과 그것의 세기인 엘니뇨, 라니냐는 대기과학으로 설명 되는데, 이것은 물리학과 지구과학에 나오는 내용이다. 이것을 바탕으로 우리는 ONI 지수와 SOI 지수를 가지고 선형 회귀를 하는 아이디어를 떠올리게 되었는데, 실제로 데이터를 통한 선형 회귀에서는 간단한 양의 상관관계를 보였으나 결정계수가 0.5~0.6으로 나타나 서로 비슷한 양상을 보이기는 하나 정밀하게 밀접한 관계를 가지고 있지는 않다는 결과를 보여주었다. 이 오차의 원인으로는 위의 지수들은 편차를 나타내는데, 결국 평균값이 작으면 절대적인 양이 줄어서 편차만으로는 설명하기 힘들기 때문이라 판단했다.

6. 기대 효과&느낀 점

워커 순환에는 많은 수치를 나타내는 데이터들이 있다. 그 중에서도 기압, 온도는 엘니뇨와 라니냐를 판단하는 직접적인 지표인데, 이 지표들의 상관관계를 밝혀냈다는 점에서 이 연구는 지구과학적인 측면에서 많은 방향으로 사용될 수 있다. 그리고 코딩을 하면서 엘니뇨, 라니냐에 대한 많은 이해를 할 뿐만이 아니라 학교에서 배웠던 선형 회귀를 이용하여 이 둘의 상관관계를 알아냄으로써 선형 회귀의 쓰임새를 알게 되어 좋았다.