

북극 빙하 예측

Arctic Glacier Prediction

데이터사이언스프로그래밍 3조

박성철, 이수민, 강정원, 최연우, 홍지윤, 선우정인

목차

INDEX

서론

1. 소개
2. 모델 설명
3. 모델 사용

본론

1. 북극 빙하량(1990-2023)
2. 해빙 면적 예측
 - 1) CO2 변화
 - 2) 기온 변화
 - 3) 오존층 파괴 물질 배출량 변화

결론



지구 온난화 가속화



해빙 현상

“ 파이썬을 통해 북극 해빙 면적을 예측해본다. ”

알려진 다른 관련 데이터 값을 사용하여 알 수 없는 데이터의 값을 예측하는 데이터 분석 기법

선형 회귀 모델

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

scikit-learn을 사용한 선형 회귀

statsmodels를 사용한 선형 회귀

선형 회귀 모델

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

scikit-learn을 사용한 선형 회귀

선형 회귀 모델을 포함하여 다양한 머신러닝 알고리즘을 쉽게 사용할 수 있도록 해주는 라이브러리

선형 회귀 모델

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

statsmodels를 사용한 선형 회귀

통계 모델, 통계 테스트, 데이터 탐색 및 시각화를 위한 라이브러리

Q. 두 라이브러리를 함께 사용하는 이유?



예측과 분석의 조합



상호 보완적 기능



다양한 분석 관점 제공

모델 사용

scikit-learn을 사용한 선형 회귀

독립 변수: CO2 배출량, 온도, 오존 농도
종속 변수: 빙하량

1. 단순 선형 회귀 모델을 사용하여 빙하량 예측
2. 모델을 학습하고, 2011년부터 2023년까지의 데이터를 예측
3. 예측 결과를 시각화하여 실제 데이터와 비교

모델 사용

statsmodels을 사용한 선형 회귀

독립 변수: CO2 배출량, 온도, 오존 농도
종속 변수: 빙하량

1. OLS(Ordinary Least Squares) 회귀 모델을 사용하여 빙하량을 예측
2. 모델 요약 결과를 통해 회귀 계수, p-value, R-squared 값 등을 확인
3. 독립 변수와 종속 변수 간의 관계를 개별적으로 시각화

북극 빙하량 (1990-2023)

Year	Extent(million km2)	Area(million km2)
1990	6.14	4.55
1991	6.47	4.51
1992	7.47	5.43
1993	6.4	4.58
1994	7.14	5.13
1995	6.08	4.43
1996	7.58	5.62
1997	6.69	4.89
1998	6.54	4.3
1999	6.12	4.29

북극 빙하량 (1990-2023)

Year	Extent(million km2)	Area(million km2)
2000	6.25	4.35
2001	6.73	4.59
2002	5.83	4.03
2003	6.12	4.05
2004	5.98	4.39
2005	5.5	4.07
2006	5.86	4.01
2007	4.27	2.82
2008	4.69	3.26
2009	5.26	3.76

북극 빙하량 (1990-2023)

Year	Extent(million km2)	Area(million km2)
2010	4.87	3.34
2011	4.56	3.21
2012	3.57	2.41
2013	5.21	3.78
2014	5.22	3.74
2015	4.62	3.42
2016	4.53	2.91
2017	4.82	3.35
2018	4.79	3.35
2019	4.36	3.17

북극 빙하량 (1990-2023)

Year	Extent(million km2)	Area(million km2)
2020	4	2.83
2021	4.95	3.47
2022	4.87	3.43
2023	4.88	3.44



Year	Extent(million km2)	Area(million km2)
1990	6.14	4.55
2023	4.88	3.44

해빙 면적 예측-CO2 변화

Year	CO2 emissions (Mt of CO2)
1990	20540.491
1991	20668.477
1992	20608.612
1993	20721.527
1994	20823.243
1995	21399.839
1996	21849.473
1997	22248.307
1998	22423.479
1999	22568.253

해빙 면적 예측-CO2 변화

Year	CO2 emissions (Mt of CO2)
2000	23266.324
2001	23607.387
2002	23940.343
2003	24969.723
2004	26146.401
2005	27105.957
2006	27972.763
2007	29021.145
2008	29237.161
2009	28818.584

해빙 면적 예측-CO2 변화

Year	CO2 emissions (Mt of CO2)
2010	30616.252
2011	31498.463
2012	31837.169
2013	32449.021
2014	32515.959
2015	32377.913
2016	32427.489
2017	32953.72
2018	33740.007
2019	33667.851

해빙 면적 예측-CO2 변화

Year	CO2 emissions (Mt of CO2)
2020	31739.551
2021	33572.108
2022	36800.001
2023	37400.001



Year	CO2 emissions (Mt of CO2)
1990	20540.491
2023	37400.001

해빙 면적 예측-CO2 변화

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
import seaborn as sns

# CSV 파일 읽기
data = pd.read_csv('/arctic.csv')

# 데이터 확인
print(data.head())

# 독립 변수와 종속 변수 설정
X = data[['CO2 emissions', 'temperature', 'ozone']]
y = data['glacial volume']

# 상수 항 추가
X = sm.add_constant(X)

# 모델 학습
model = sm.OLS(y, X).fit()

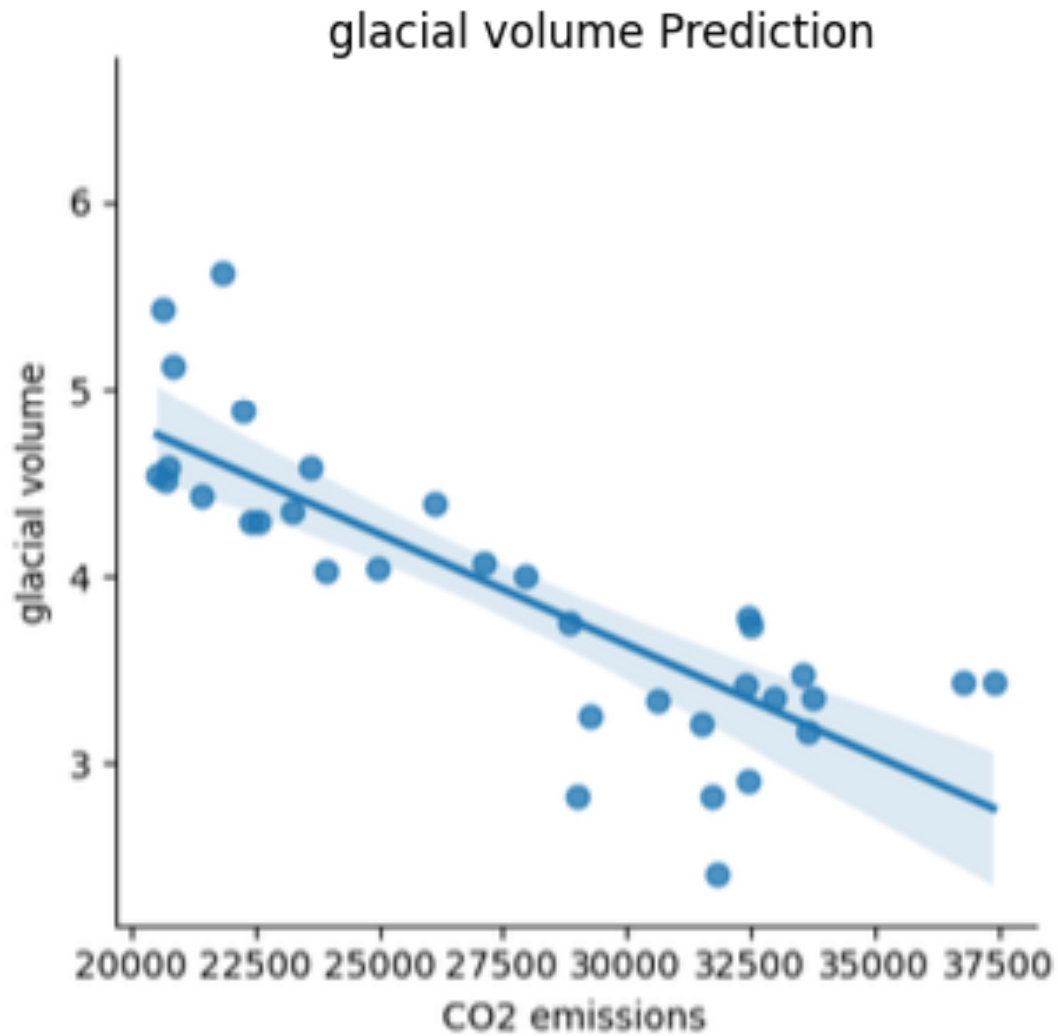
# 각 독립 변수와 종속 변수 간의 관계를 개별적으로 시각화
sns.pairplot(data, x_vars=['CO2 emissions', 'temperature', 'ozone'], y_vars='glacial volume', height=4, aspect=1, kind='reg')
plt.suptitle('glacial volume Prediction')
plt.show()

# 모델 요약 결과 출력
print(model.summary())
```



	date	temperature	CO2 emissions	ozone	glacial volume
0	1990	-24	20540.491	977580	4.55
1	1991	-23	20668.477	892329	4.51
2	1992	-33	20608.612	857490	5.43
3	1993	-30	20721.527	765773	4.58
4	1994	-32	20823.243	490092	5.13

해빙 면적 예측-CO2 변화



해빙 면적 예측-기온 변화(1월 기준)

Year	Temperature (°C)
1990	-24
1991	-23
1992	-33
1993	-30
1994	-32
1995	-33
1996	-34
1997	-34
1998	-23
1999	-33

해빙 면적 예측-기온 변화(1월 기준)

Year	Temperature (°C)
2000	-13
2001	-28
2002	-28
2003	-18
2004	-23
2005	-24
2006	-29
2007	-30
2008	-25
2009	-23

해빙 면적 예측-기온 변화(1월 기준)

Year	Temperature (°C)
2010	-23
2011	-29
2012	-25
2013	-19
2014	-28
2015	-24
2016	-20
2017	-21
2018	-23
2019	-24

해빙 면적 예측-기온 변화(1월 기준)

Year	Temperature (°C)
2020	-28
2021	-14
2022	-26
2023	-23



Year	Temperature(°C)
1990	-24
2023	-23

해빙 면적 예측-기온 변화(1월 기준)

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
import seaborn as sns

# 독립 변수와 종속 변수 설정
X = data[['CO2 emissions', 'temperature', 'ozone']]
y = data['glacial volume']


# 상수 항 추가
X = sm.add_constant(X)

# CSV 파일 읽기
data = pd.read_csv('/arctic.csv')

# 모델 학습
model = sm.OLS(y, X).fit()

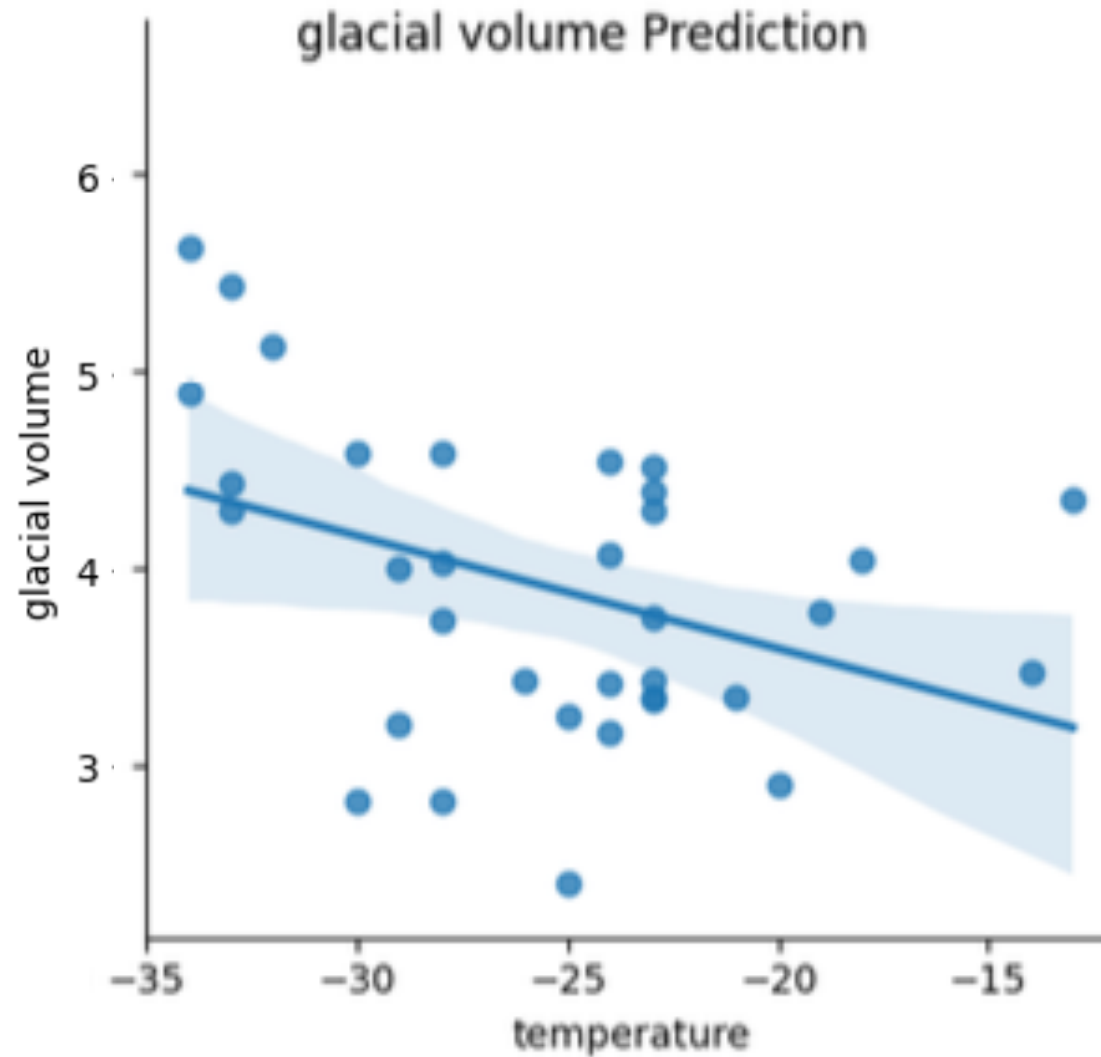
# 각 독립 변수와 종속 변수 간의 관계를 개별적으로 시각화
sns.pairplot(data, x_vars=['CO2 emissions', 'temperature', 'ozone'], y_vars='glacial volume', height=4, aspect=1, kind='reg')
plt.suptitle('glacial volume Prediction')
plt.show()

# 모델 요약 결과 출력
print(model.summary())
```



	date	temperature	CO2 emissions	ozone	glacial volume
0	1990	-24	20540.491	977580	4.55
1	1991	-23	20668.477	892329	4.51
2	1992	-33	20608.612	857490	5.43
3	1993	-30	20721.527	765773	4.58
4	1994	-32	20823.243	490092	5.13

해빙 면적 예측-기온 변화(1월 기준)



해빙 면적 예측-오존층 파괴 물질 배출량 변화

Year	Emissions of Ozone-depleting substances (ODP ton)
1990	977,580
1991	892,329
1992	857,490
1993	765,773
1994	490,092
1995	406,651
1996	292,720
1997	269,777
1998	338,117
1999	287,720

해빙 면적 예측-오존층 파괴 물질 배출량 변화

Year	Emissions of Ozone-depleting substances (ODP ton)
2000	263,120
2001	207,931
2002	162,687
2003	170,636
2004	127,920
2005	92,858
2006	86,518
2007	61,600
2008	43,797
2009	49,241

해빙 면적 예측-오존층 파괴 물질 배출량 변화

Year	Emissions of Ozone-depleting substances (ODP ton)
2010	43,698
2011	40,165
2012	43,923
2013	29,230
2014	30,795
2015	23,040
2016	21,508
2017	20,565
2018	20,667
2019	21,287

해빙 면적 예측-오존층 파괴 물질 배출량 변화

Year	Emissions of Ozone-depleting substances (ODP ton)
2020	9,538
2021	5,763
2022	8,892
2023	7,326



Year	Emissions of Ozone-depleting substances (ODP ton)
1990	977,580
2023	7,326

해빙 면적 예측-오존층 파괴 물질 배출량 변화

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
import seaborn as sns

# 독립 변수와 종속 변수 설정
X = data[['CO2 emissions', 'temperature', 'ozone']]
y = data['glacial volume']


# 상수 항 추가
X = sm.add_constant(X)

# CSV 파일 읽기
data = pd.read_csv('/arctic.csv')

# 모델 학습
model = sm.OLS(y, X).fit()

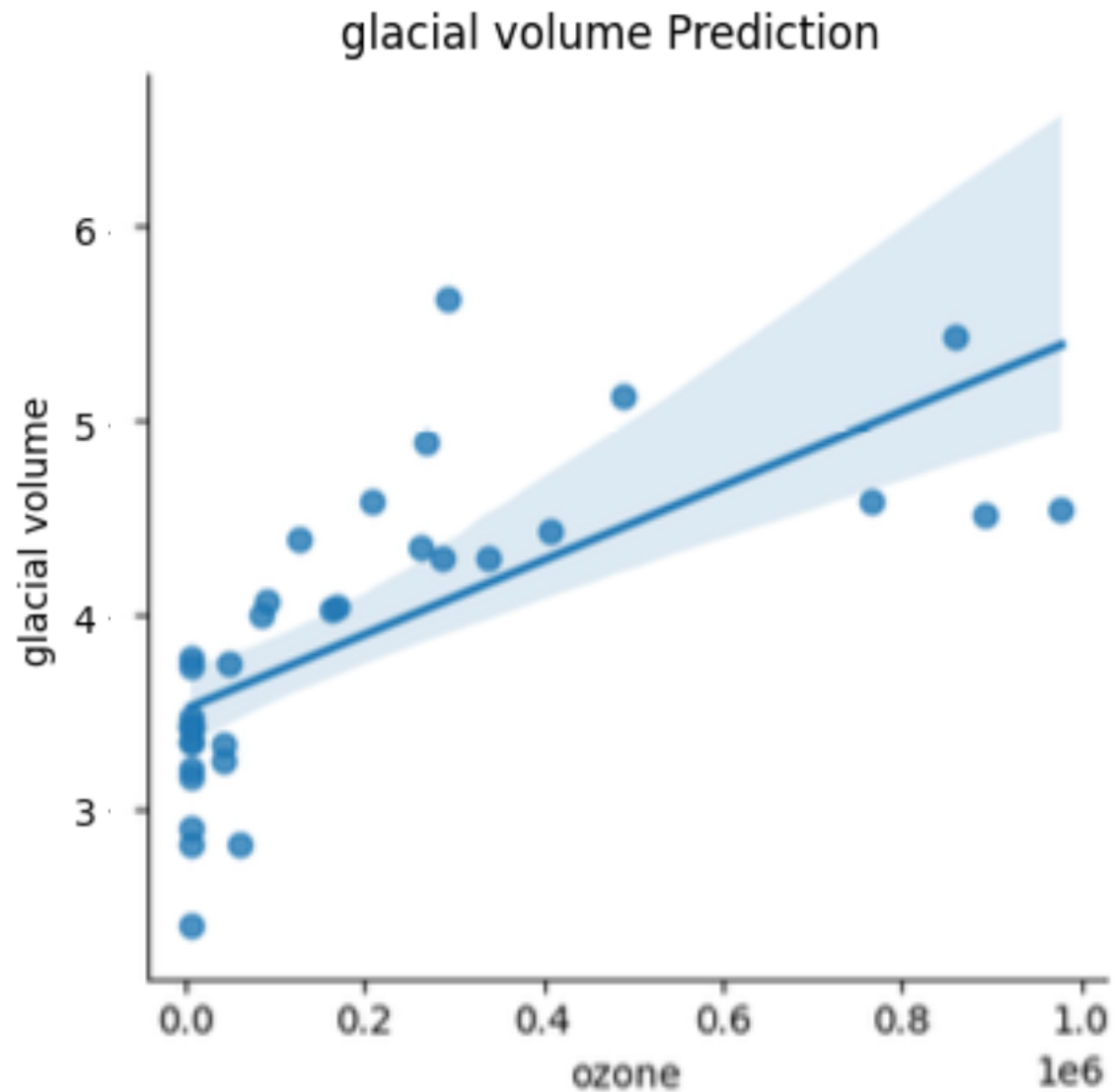
# 각 독립 변수와 종속 변수 간의 관계를 개별적으로 시각화
sns.pairplot(data, x_vars=['CO2 emissions', 'temperature', 'ozone'], y_vars='glacial volume', height=4, aspect=1, kind='reg')
plt.suptitle('glacial volume Prediction')
plt.show()

# 모델 요약 결과 출력
print(model.summary())
```



	date	temperature	CO2 emissions	ozone	glacial volume
0	1990	-24	20540.491	977580	4.55
1	1991	-23	20668.477	892329	4.51
2	1992	-33	20608.612	857490	5.43
3	1993	-30	20721.527	765773	4.58
4	1994	-32	20823.243	490092	5.13

해빙 면적 예측-오존층 파괴 물질 배출량 변화



미래의 북극 해빙 면적 예측

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# CSV 파일 읽기
data = pd.read_csv('/arctic.csv')

# 데이터 확인
print(data.head())

# 모델 학습을 위한 데이터 준비
X = data[['CO2 emissions', 'temperature', 'ozone']]
y = data['glacial volume']

# 모델 초기화 및 학습
model = LinearRegression()
model.fit(X, y)

# 2011년부터 2023년까지의 데이터를 예측
years_predict = np.arange(2011, 2024, 1).reshape(-1, 1)
predicted_data = model.predict(data[['CO2 emissions', 'temperature', 'ozone']][-13:])
```

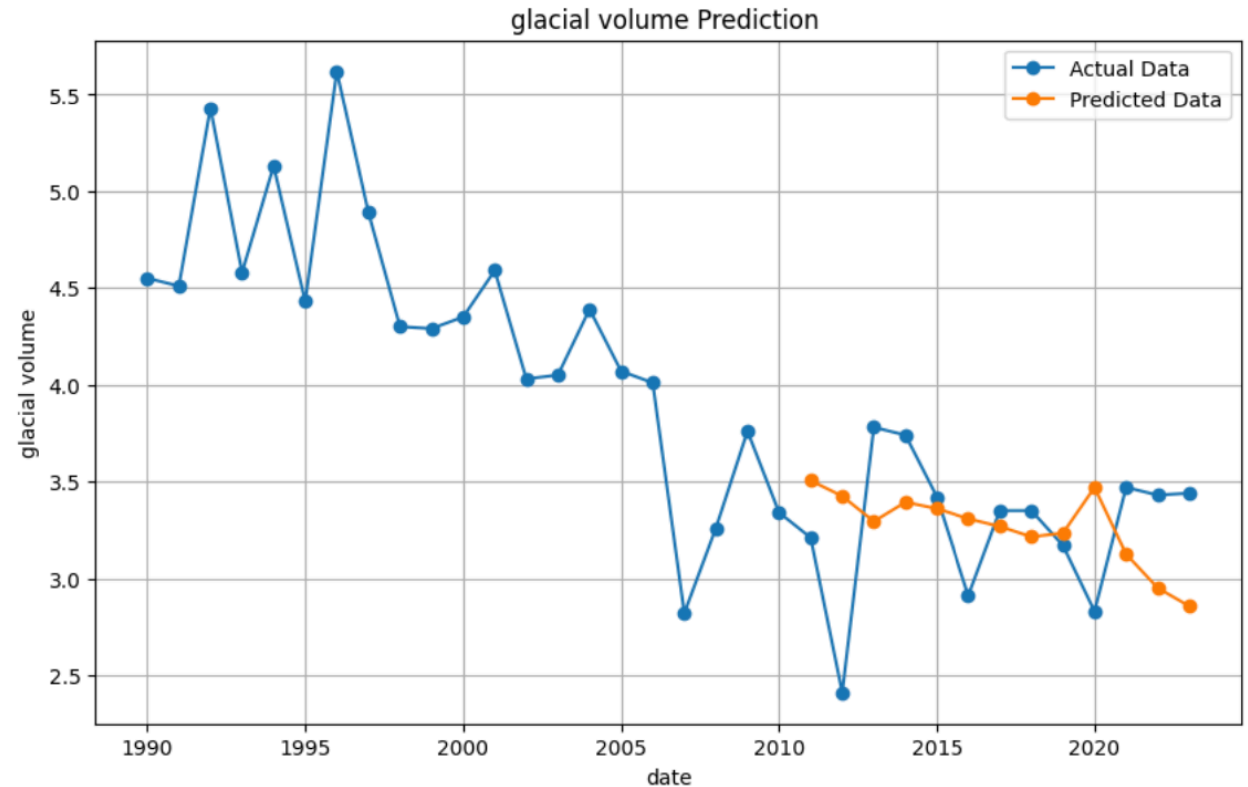

북극 해빙 면적 예측

그래프 그리기

```
plt.figure(figsize=(10, 6))
plt.plot(data['date'], data['glacial volume'], label='Actual Data', marker='o')
plt.plot(np.arange(2011, 2024, 1), predicted_data, label='Predicted Data', marker='o')
plt.xlabel('date')
plt.ylabel('glacial volume')
plt.title('glacial volume Prediction ')
plt.legend()
plt.grid(True)
plt.show()
```



	date	temperature	CO2 emissions	ozone	glacial volume
0	1990	-24	20540.491	977580	4.55
1	1991	-23	20668.477	892329	4.51
2	1992	-33	20608.612	857490	5.43
3	1993	-30	20721.527	765773	4.58
4	1994	-32	20823.243	490092	5.13



“ 북극 해빙 면적 예측 결과가 실제
와 일치하지 않았다. ”

북극 해빙 면적 예측

OLS Regression Results

Dep. Variable:	glacial volume	R-squared:	0.682
Model:	OLS	Adj. R-squared:	0.650
Method:	Least Squares	F-statistic:	21.42
Date:	Sat, 01 Jun 2024	Prob (F-statistic):	1.31e-07
Time:	12:18:47	Log-Likelihood:	-19.241
No. Observations:	34	AIC:	46.48
Df Residuals:	30	BIC:	52.59
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	6.2450	1.000	6.244	0.000	4.202	8.288
CO2 emissions	-9.793e-05	2.61e-05	-3.754	0.001	-0.000	-4.46e-05
temperature	-0.0118	0.016	-0.726	0.474	-0.045	0.021
ozone	3.669e-07	4.7e-07	0.780	0.442	-5.94e-07	1.33e-06

Omnibus:	1.059	Durbin-Watson:	1.859
Prob(Omnibus):	0.589	Jarque-Bera (JB):	0.433
Skew:	-0.259	Prob(JB):	0.805
Kurtosis:	3.194	Cond. No.	4.44e+06

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 4.44e+06. This might indicate that there are strong multicollinearity or other numerical problems.

“모델은 통계적으로 유의미하며 R-squared가 68.2%로 비교적 높습니다.”

“CO2의 배출량은 빙하 부피에 통계적으로 유의미한 영향을 미치지만 기온과 오존층 파괴 물질 배출량은 빙하 부피에 유의미한 영향을 미치지 않는 것으로 나타났습니다.”

감사합니다.