

Міністерство освіти і науки України
Національний технічний університет України «Київський політехнічний
інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

Звіт

з лабораторної роботи № 2 з дисципліни
«Прикладні задачі машинного навчання»

«Часові ряди і проста лінійна регресія»

Варіант 7

Виконав студент: ПІ-12 Васильєв Єгор Костянтинович

Перевірив: Нестерук Андрій Олександрович

Київ 2022

Лабораторна робота №2

Тема: Часові ряди і проста лінійна регресія

Постановка завдання:

- 1) завантажити метеорологічні дані;
- 2) виконати первинну обробку даних;
- 3) побудувати регресійну пряму зміни температури;
- 4) спрогнозувати дані на 2019-2023 рік;
- 5) оцінити, якою могли б бути показники до 1895;
- 6) будувати діаграму розкиду даних та налаштувати графік;
- 7) порівняти отриманий прогноз для 2019-2023 роки з фактичними вимірами.

Хід роботи:

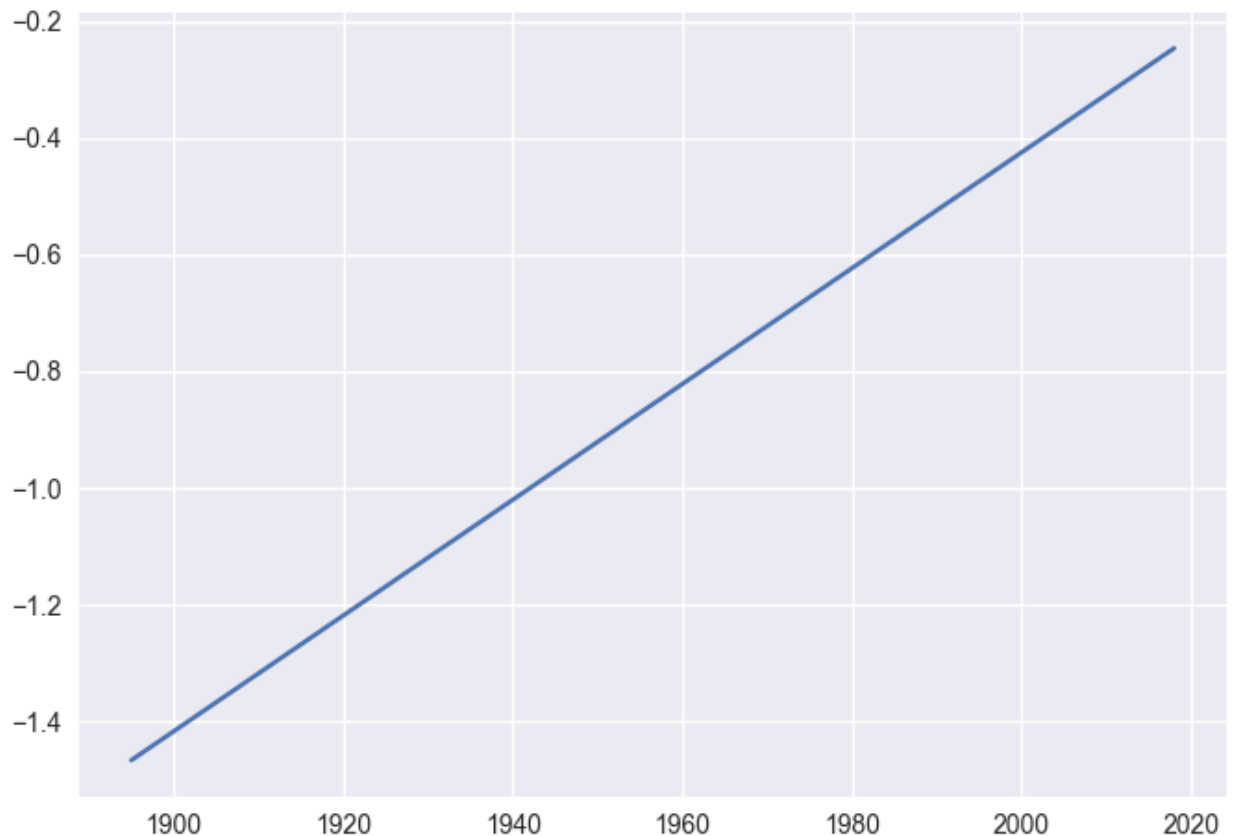
- Завантаження необхідних даних:

	A	B	C	D	E
1	Contiguous U.S., Average Temperature, January				
2	Units: Degrees Fahrenheit				
3	Base Period: 1901-2000				
4	Missing: -99				
5	Date, Value, Anomaly				
6	189501	26.69	-3.43		
7	189601	31.48	1.36		
8	189701	28.17	-1.95		
9	189801	30.67	0.55		
10	189901	29.68	-0.44		
11	190001	34.11	3.99		
12	190101	31.62	1.50		
13	190201	30.06	-0.06		
14	190301	30.88	0.76		
15	190401	27.55	-2.57		
16	190501	27.10	-3.02		
17	190601	32.88	2.76		
18	190701	30.69	0.57		
19	190801	32.45	2.33		
20	190901	32.31	2.19		
21	191001	29.61	-0.51		
22	191101	32.90	2.78		
23	191201	24.76	-5.36		
24	191301	30.47	0.35		
25	191401	34.50	4.38		
26	191501	28.44	-1.68		
27	191601	28.22	-1.90		
28	191701	28.27	-1.85		
29	191801	23.86	-6.26		
30	191901	31.82	1.70		
31	192001	29.23	-0.89		
32	192101	34.36	4.24		

- Первинна обробка даних

Contiguous U.S.			Average Temperature	January	Date			Temperature	Anomaly
0	Units: Degrees Fahrenheit		nan	nan	0	1895		-2.95000	-3.43
1	Base Period: 1901-2000		nan	nan	1	1896		-0.28889	1.36
2	Missing: -99		nan	nan	2	1897		-2.12778	-1.95
3	Date		Value	Anomaly	3	1898		-0.73889	0.55
4	189501		26.69	-3.43	4	1899		-1.28889	-0.44
5	189601		31.48	1.36	5	1900		1.17222	3.99
6	189701		28.17	-1.95	6	1901		-0.21111	1.50
7	189801		30.67	0.55	7	1902		-1.07778	-0.06
8	189901		29.68	-0.44	8	1903		-0.62222	0.76
9	190001		34.11	3.99	9	1904		-2.47222	-2.57
10	190101		31.62	1.50	10	1905		-2.72222	-3.02
11	190201		30.06	-0.06	11	1906		0.48889	2.76
12	190301		30.88	0.76					
13	190401		27.55	-2.57					

- Графічне представлення даних у вигляді регресійної прямої, що представляє графік зміни температури за період 1895-2018 років



- Прогнозування даних на 2019, 2020, 2021, 2022 та 2023 роки

```

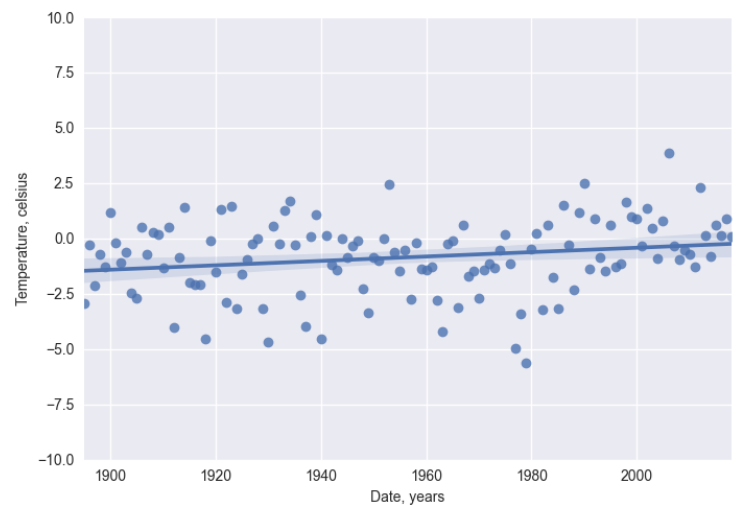
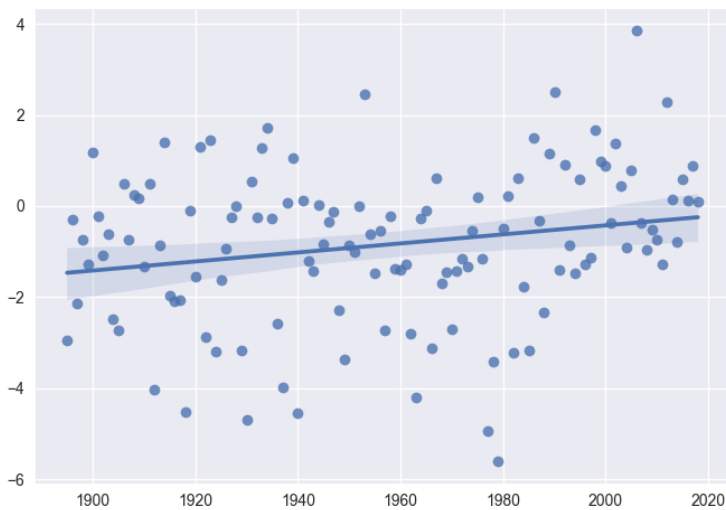
Projected temperature of 2019 year: -0.24
Projected temperature of 2020 year: -0.23
Projected temperature of 2021 year: -0.22
Projected temperature of 2022 year: -0.21
Projected temperature of 2023 year: -0.20

```

- Оцінка можливих показників температури до 1895 року

```
Projected temperature of 1890 year: -1.52
Projected temperature of 1891 year: -1.51
Projected temperature of 1892 year: -1.50
Projected temperature of 1893 year: -1.49
Projected temperature of 1894 year: -1.48
```

- Побудова діаграми розкиду даних за допомогою методу reglot бібліотеки Seaborn та її налаштування



- Порівняння отриманого прогнозу температури 2019-2023 з фактичними даними

Прогнозована температура, F	Фактична температура, F	Різниця, %
31,57	32,56	3,14
31,59	35,55	12,54
31,61	34,59	9,43
31,63	31,17	1,45
31,65	35,17	11,12
Середнє значення:		7,536

Вихідний код

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

pd.set_option('display.precision', 2)
plt.style.use('seaborn-v0_8')

initial_df = pd.read_csv(r'F:\Егор\Уроки\Машинне навчання\Ла62\1895-2018.csv')
df = initial_df.copy()
df.columns = ['Date', 'Temperature', 'Anomaly']
df.drop(index=range(4), inplace=True)
df.reset_index(drop=True, inplace=True)
df = df.astype({'Date': 'int64', 'Temperature': 'float'})
# df.loc[:, 'Temperature'] = df.loc[:, 'Temperature'].apply(lambda x: 5 / 9 * (x - 32))
df.loc[:, 'Date'] = df.loc[:, 'Date'].floordiv(100)
print(df.head())
dates = df.loc[:, 'Date'].values
temps = df.loc[:, 'Temperature'].values
linear_regression = stats.linregress(dates, temps)

plt.plot(dates, linear_regression.slope * dates + linear_regression.intercept)
plt.show()

def temperature_forecast(regression, years):
    for year in years:
        print(f'Projected temperature of {year} year: {regression.slope * year + regression.intercept:.2f}')
    print()

temperature_forecast(linear_regression, range(2019, 2024))
temperature_forecast(linear_regression, range(1890, 1895))

sns.regplot(x=dates, y=temps)
plt.xlabel('Date, years')
plt.ylabel('Temperature, celsius')
plt.xlim(1895, 2018)
plt.ylim(-10, 10)
plt.show()
```

Висновок

Було досліджено можливості бібліотек pandas, matplotlib, seaborn, scify та використано деякі з них на практиці для візуалізації зміни середньої температури січня континентальної Америки; було вивчено основні методи аналізу та прогнозування часових рядів; було досліджено принцип побудови простої лінійної регресії та методи її візуалізації; таким чином було закріплено вивчену тему «Часові ряди і проста лінійна регресія»