



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

**Лабораторна робота №2**  
**Прикладні задачі машинного навчання**  
**Тема: Часові ряди**

Виконав  
студент групи ІП-11:  
Панченко С. В.

Перевірив:  
Нестерук А. О

## ЗМІСТ

<b>1 МЕТА ЛАБОРАТОРНОЇ РОБОТИ.....</b>	<b>3</b>
<b>2 ЗАВДАННЯ.....</b>	<b>4</b>
<b>3 ВИКОНАННЯ.....</b>	<b>5</b>
3.1 ЗАВАНТАЖИТИ МЕТЕОРОЛОГІЧНІ ДАНІ В 1895-2022 РОКАХ З CSV-ФАЙЛУ В DATAFRAME. ПІСЛЯ ЦЬОГО ДАНІ ВІДФОРМАТУВАТИ ДЛЯ ВИКОРИСТАННЯ.....	5
3.2 ДАНІ НА 2019, 2020, 2021 ТА 2022 РІК.....	5
3.3 ЗА ФОРМУЛОЮ, ЯКИМИ МОГЛИ Б БУТИ ПОКАЗНИКИ ДО 1895 РОКУ.....	6
3.4 ФУНКЦІЄЮ REGPLOT БІБЛІОТЕКИ SEABORN ДЛЯ ВИВЕДЕННЯ ВСІХ ТОЧОК ДАНИХ.....	7
3.5 МАСШТАБУВАННЯ ОСІ Y.....	7
3.6 ОТРИМАНИЙ ПРОГНОЗ ДЛЯ 2019, 2020, 2021 ТА ЗА 2022 РОКИ З ДАНИМИ НА NOAA «CLIMATE AT A GLANCE»: HTTPS://WWW.NCDC.NOAA.GOV/CAG/ І ЗРОБИТИ ВИСНОВОК.....	8
<b>4 ВИСНОВОК.....</b>	<b>10</b>

## 1 МЕТА ЛАБОРАТОРНОЇ РОБОТИ

Мета роботи – дослідити лінійну регресію на прикладі прогнозування січневих температур у Нью-Йорку за роками з використанням Python

## 2 ЗАВДАННЯ

1. Завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані відформатувати для використання.
2. Бібліотеку Seaborn використати для графічного представлення даних DataFrame у вигляді регресійної прямої, що представляє графік зміни обраних показників за період 1895-2018 років.
3. Спрогнозувати дані на 2019, 2020, 2021 та 2022 рік.
4. Оцінити за формулою, якими могли б бути показники до 1895 року.
5. Скористатися функцією regplot бібліотеки Seaborn для виведення всіх точок даних
6. Виконати масштабування осі y
7. Порівняти отриманий прогноз для 2019, 2020, 2021 та за 2022 роки з даними на NOAA «Climate at a Glance»: <https://www.ncdc.noaa.gov/cag/> і зробити висновок.
8. Зробити звіт про роботу

### 3 ВИКОНАННЯ

3.1 Завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані відформатувати для використання

Зчитуємо дані з CSV-файлу, використовуючи метод `read_csv`.

```
In [16]: pd.options.display.precision = 2
pd.options.display.precision

Out[16]: 2
```

Рисунок 3.5 - Зчитуємо дані з CSV-файлу, використовуючи метод `read_csv`.

Знайдемо основні статистичні показники.

```
In [17]: nyc.Temperature.describe()

Out[17]: count      124.00
mean         37.60
std           4.54
min          26.10
25%          34.58
50%          37.60
75%          40.60
max          47.60
Name: Temperature, dtype: float64
```

Рисунок 3.6 - Знайдемо основні статистичні показники.

3.2 дані на 2019, 2020, 2021 та 2022 рік

Імпортуємо модуль `stats` з пакету `scipy` та за допомогою функції `linregress` знайдемо лінійну регресію, передавши в аргументи дати та температури.

```
In [18]: from scipy import stats
linear_regression = stats.linregress(x=nyc.Date, y=nyc.Temperature)
```

Рисунок 3.7 - Імпортуємо модуль `stats` з пакету `scipy` та за допомогою функції `linregress` знайдемо лінійну регресію, передавши в аргументи дати та температури.

Розрахувавши регресію, дізнаємося про коефіцієнт нахилу.

```
In [19]: linear_regression.slope

Out[19]: 0.014771361132966163
```

Рисунок 3.8 - Розрахувавши регресію, дізнаємося про коефіцієнт нахилу.

Дізнаємося про точку перетину прямої лінії

```
In [20]: linear_regression.intercept
```

```
Out[20]: 8.694993233674289
```

Рисунок 3.9 - Дізнаємося про точку перетину прямої лінії

Спрогнозуємо дані для наступних років. Тобто підставимо роки у формулу лінійної регресії.

```
In [21]: years = np.array([2019, 2020, 2021, 2022])
values = linear_regression.slope * years + linear_regression.intercept
df_predict = pd.DataFrame({'Years': years, 'Temperature': values})
df_predict
```

```
Out[21]:
```

	Years	Temperature
0	2019	38.52
1	2020	38.53
2	2021	38.55
3	2022	38.56

Рисунок 3.10 - Спрогнозуємо дані для наступних років. Тобто підставимо роки у формулу лінійної регресії.

3.3 за формулою, якими могли б бути показники до 1895 року

Обчислимо показники до 1895 року

```
In [22]: years = np.array(range(1885, 1896))
values = linear_regression.slope * years + linear_regression.intercept
df_predict = pd.DataFrame({'Years': years, 'Temperature': values})
df_predict
```

```
Out[22]:
```

	Years	Temperature
0	1885	36.54
1	1886	36.55
2	1887	36.57
3	1888	36.58
4	1889	36.60
...	...	...
6	1891	36.63
7	1892	36.64
8	1893	36.66
9	1894	36.67
10	1895	36.69

11 rows x 2 columns

Рисунок 3.11 - Обчислимо показники до 1895 року

Як можна побачити, температура поступово зростає, і з періоду 1885 по 2023 роки спостерігається збільшення на 2 градуси.

### 3.4 функцією regplot бібліотеки Seaborn для виведення всіх точок даних

Імпортуємо Seaborn, встановлюємо стиль відображення, побудуємо графік роки-температури. Побачимо, що дані доволі розкидані.

```
In [23]: import seaborn as sns
sns.set_style('whitegrid')
axes = sns.regplot(x=nyc.Date, y=nyc.Temperature)
```

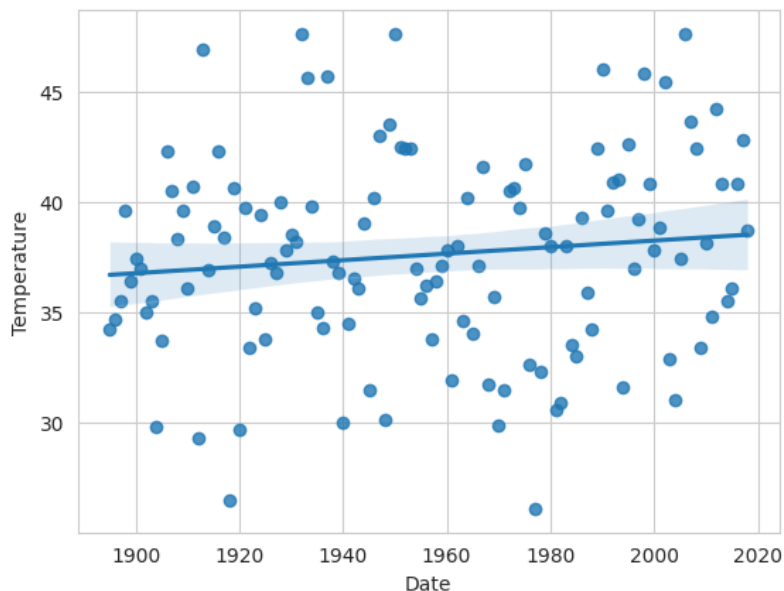


Рисунок 3.12 - Імпортуємо Seaborn, встановлюємо стиль відображення, побудуємо графік роки-температури. Побачимо, що дані доволі розкидані.

### 3.5 Масштабування осі y

За допомогою методу `set_ylim` вкажемо межі від 10 до 70 градусів.

```
In [24]: axes = sns.regplot(x=nyc.Date, y=nyc.Temperature)
axes.set_ylim(10, 70)
```

```
Out[24]: (10.0, 70.0)
```

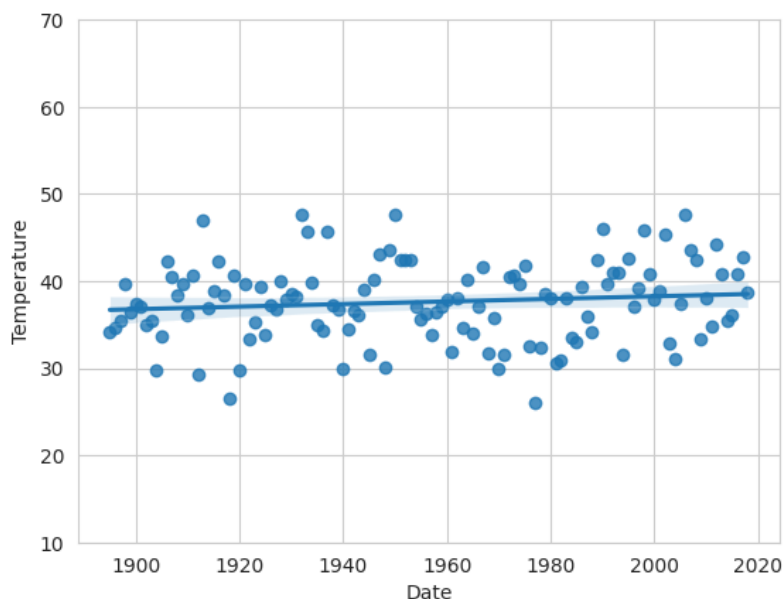


Рисунок 3.13 - За допомогою методу `set_ylim` вкажемо межі від 10 до 70 градусів.

3.6 отриманий прогноз для 2019, 2020, 2021 та за 2022 роки з даними на NOAA «Climate at a Glance»: <https://www.ncdc.noaa.gov/cag/> і зробити висновок

Подивимося на сайті дані за період 2019-2023 років. Побачимо, що фактичні дані сильно відрізняються від того, що спрогнозувала лінійна регресія. Можна зробити висновок, що треба давати їй ще якісь дані для кращого прогнозування, наприклад: кількість опадів, кількість CO2 тощо.

```
In [27]: from PIL import Image
im = Image.open('data/Screenshot from 2023-03-15 20-01-39.png')
im
```

```
Out[27]: New York, New York Average Temperature
January-December
```

Year	Average Temperature	Rank	Anomaly 1901-2000 Mean: 53.6°F
2022	56.3°F	112	2.7°F
2021	56.9°F	121	3.3°F
2020	57.3°F	125	3.7°F
2019	55.6°F	105	2.0°F
2018	55.9°F	108	2.3°F

Рисунок 3.14 — Фактичні дані середніх температур



#### 4 ВИСНОВОК

Під час виконання цієї лабораторної роботи здобув базові навички використання пакету `scipy` мови Python, досліджуючи середні температури в січня у Нью-Йорку з 1895 до 2022 років, обчисливши лінійну регресію та зробивши прогноз. У результаті спрогнозовані дані не збігалися з фактичними даними. Отже, точність низька та потрібно врахувати додаткові параметри.