НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

**Лабораторна робота № 1**

з дисципліни «Прикладні задачі машинного навчання»

Тема: «Часові ряди і прості лінійна регресія»

|  |  |
| --- | --- |
| **Прийняв:** | **Виконав:**  **студент групи ІП-13**  **Недельчев Є.О.** |

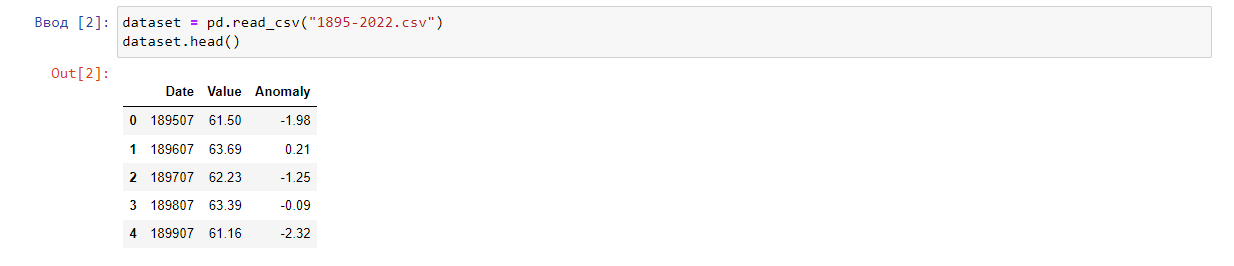
Київ – 2023

**Завдання:**

1. **В даній лабораторній роботі Вам треба завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані треба буде 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. **Спрогнозуйте дані на 2019, 2020, 2021 та 2022 рік.**
4. **Оцініть за формулою, якою могли б бути показники до 1895 року. Наприклад, оцінка середньої температури за січень 1890 року може бути отримана наступним чином.**
5. **Скористайтесь функцією regplot бібліотеки Seaborn для виведення всіх точок даних; дати представляються на осі x, а показники на осі y. Функція regplot будує діаграму розкиду даних, на якій точки представляють показники за заданий рік, а пряма лінія - регресійну пряму.**
6. **Виконайте масштабування осі *у* від (приклад від 10 до 70 градусів):**
7. **Порівняйте отриманий прогноз для 2019, 2020, 2021 та за 2022 роки з даними на NOAA «Climate at a Glance»:** [**https://www.ncdc.noaa.gov/cag/**](https://www.ncdc.noaa.gov/cag/) **і зробити висновок.**

**Виконання**

**1.** Завантажимо дані та продивимося їх структуру:

****

Перейменуємо назви стовпчиків на більш інтуїтивно зрозумілі:

**Изображение выглядит как текст

Автоматически созданное описание**

Оскільки будемо обробляти тільки січневі дані, мітки осі x будуть краще читатися без позначення 01 (для січня); видалимо місяць з Date:

**Изображение выглядит как текст

Автоматически созданное описание**

**2.** Використаємо бібліотеку seaborn для графічного представлення даних DataFrame у вигляді регресійної прямої, що представляє графік зміни обраних показників за період 1895-2018 років.

**Изображение выглядит как диаграмма

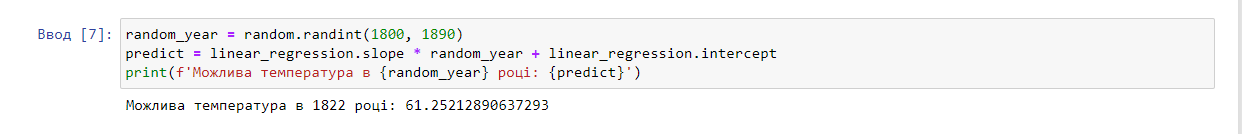
Автоматически созданное описание**

**3.** Спрогнозуємо дані на 2019, 2020, 2021 та 2022 рік.

Изображение выглядит как текст

Автоматически созданное описание

**4.** Оцінимо, які могли б бути дані у випадковий рік на проміжку з 1800 по 1890.



**5 & 6.** Скористаємось функцією regplot бібліотеки Seaborn для виведення всіх точок даних. Також виконаємо масштабування осі у від 58 до 69 для кращої візуалізації:

Изображение выглядит как диаграмма

Автоматически созданное описание

**7.** Порівняємо отримані в результаті лабораторної роботи прогнози із реальними даними.

|  |  |  |  |
| --- | --- | --- | --- |
| Рік | NOAA «Climate at a Glance» | Дані, отримані в ході роботи | Різниця |
| 2019 | 62.78 | 64.46 | +1.68 |
| 2020 | 65.34 | 64.49 | -0.85 |
| 2021 | 64.82 | 64.50 | -0.32 |
| 2022 | 65.19 | 64.52 | -0.67 |

Досить легко помітити, що похибка є досить суттєвою, оскільки таких підхід прогнозування даних ігнорує безліч чинників. Такий спосіб прогнозування дозволяє лише приблизно оцінити необхідні дані.

**Висновок**

Виконуючи цю лабораторну роботу я ознайомився з бібліотекою seaborn та використав на практиці графік лінійної регресії. Також були спрогнозовані дані на підставі старих даних та було зроблено висновок, що прогнозування подій із використанням лінійної регресії є досить неточним і дозволяє лише приблизно оцінити необхідні нам дані.