

Лабораторна робота №_1

ПІДГОТОВКА ТА АНАЛІЗ ДАНИХ ДЛЯ СТАТИСТИЧНОГО НАВЧАННЯ

Мета роботи:

Виявити дослідити та узагальнити особливості застосування методів статистичного навчання для задач визначення статистичних характеристик вхідного потоку даних з використанням спеціалізованих пакетів мови програмування Python.

I. SKILLS, які прокачуємо.

1. Моделювання статистичної вибірки із заданими властивостями.
3. Моделювання випадкових величин із заданими законами розподілу.
2. Парсинг файлів.
3. Парсинг даних.
4. Дослідження даних: визначення статистичних характеристик.
5. Візуалізація результатів розрахунків.
6. Верифікація розроблених скриптових реалізацій.
7. Опанування функціоналу бібліотек: Numpy, pandas, matplotlib, bs4, requests, json.

II. Корисні ресурси.

Матеріали Лекцій №1,2 курсу «Вступ до технології Data Science»

Навчально-методичний комплекс дисципліни:

<https://drive.google.com/drive/folders/1UqgoeFhzKQQm5108CsFYEKvrxZkXuS9V?usp=sharing>
<https://classroom.google.com/c/NjE4NjE4OTEwMTcx?cjc=zlumu3r>

Література:

William McKinney Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.

Plas J. Wander. Python Data Science.

Sebastian Raska, Vahid Mirjalili. Python and machine learning.

Plas J. Wander. Python Data Science.

Prateek Joshi Artificial Intelligence applications with Python.

Корисні ресурси / бібліотеки:

<https://www.kaggle.com/>

<https://github.com/PacktPublishing/Artificial-Intelligence-with-Python>

<https://scapy.net/>

<https://developers.google.com/optimization>

<https://www.tensorflow.org/>

<https://scikit-learn.org/stable/modules/sgd.html#regression>

<https://keras.io/>

<https://opencv.org/>

Статистичне навчання (МНК / регресія / поліном) в пакетах:

<https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html>

<https://www.statsmodels.org/stable/examples/notebooks/generated/ols.html>

<https://scikit-learn.org/stable/modules/sgd.html#regression>

III. Завдання.

Лабораторія провідної IT-компанії реалізує масштабний проект розробки універсальної платформи з обробки Big Data масиву статистичних даних поточного спостереження для виявлення закономірностей і прогнозування розвитку контрольованого процесу. Платформа передбачає розташування back-end компоненти на власному хмарному сервері з наданням повноважень користувачам заздалегідь адаптованого front-end функціоналу універсальної платформи.

Замовниками ресурсів платформи є: державні та комерційні компанії валютного трейдингу для прогнозування динаміки зміни курсу валют та ціни інших товарів; метеорологічні служби для прогнозування параметрів метеоумов; департаменти охорони здоров'я для прогнозування зміни показників епідеміологічних ситуацій тощо.

Вам, як Data Science Engineer поставлено завдання.

Завдання I, II рівнів складності – максимально 7, 8 балів відповідно.

Розробити програмний скрипт мовою Python що забезпечує аналіз властивостей і характеристик вихідних даних відповідно до етапів:

1. Модель генерації випадкової величини за заданим у табл.1 додатку 1 закону розподілу;
2. Модель зміни (ідеальний тренд) досліджуваного процесу за заданим у табл.1 додатку 1 законом;
3. Адитивна модель статистичної вибірки відповідно до синтезованих в п.1,2 моделей випадкової (стохастична) і не випадкової складових. Параметри закону розподілу та закону зміни досліджуваного процесу обрати самостійно.
4. Визначення статистичних (числових) характеристик сформованих в п.1,3 вибірок (дисперсія, середньоквадратичне відхилення, математичне очікування, гістограма закону розподілу).
5. Визначення статистичних характеристик реальних даних, заданих файлом Oschadbank (USD).xls за умов табл. 1 додатку 1.
6. Провести аналіз отриманих результатів та верифікацію розробленого скрипта.

Завдання III рівня – максимально 9 балів.

1. Провести парсинг самостійно обраного сайту. Вміст даних, що підлягають парсингу – обрати самостійно.
2. Результати парсингу зберегти у файлі. Тип файлу обрати самостійно.
3. Оцінити динаміку тренду реальних даних.
4. Здійснити визначення статистичних характеристик результатів парсингу.
5. Синтезувати та верифікувати модель даних, аналогічних за трендом і статистичними характеристиками реальним даним, які є результатом парсингу.
6. Провести аналіз отриманих результатів.

Приклади реалізації завдань див. матеріали Лекцій 1,2.

VI. Порядок виконання завдання лабораторної роботи.

- 4.1. Обрати завдання на лабораторну роботу за рівнем складності та відповідно до вказаного варіанту технічного завдання.
- 4.2. Реалізувати етап вибору / розробки / синтезу математичної моделі за якими здійснюватимуться обробка даних програмного скрипта.
- 4.3. Реалізувати етап архітектурного проектування (структурна схема /або/ діаграма класів /або/ блок-схема алгоритму). Здійснити опис функціонування результатів архітектурного проектування.
- 4.4. Розробити програму, що втілює розроблений алгоритм.

- 4.5. Провести тестування та верифікацію роботи програми
- 4.6. Реалізувати дослідження, що вказані в меті лабораторної роботи та сформулювати висновки.
- 4.7. Оформити звіт з лабораторної роботи та своєчасно представити його викладачеві.

V. Структура звіту з лабораторної роботи (див. Додаток 2).

- 5.1. Титульний аркуш, що містить інформацію: номер, тема, навчальна дисципліна, виконавець роботи, роботу прийняв.
- 5.2. Мета і завдання лабораторної роботи.
- 5.3. Результати виконання лабораторної роботи:
 - 5.3.1. Синтезована математична модель;
 - 5.3.2. Результати архітектурного проектування та їх опис;
 - 5.3.3. Опис структури проекту програми;
 - 5.3.4. Результати роботи програми відповідно до завдання (допускається у формі скріншотів);
 - 5.3.5. Програмний код, що забезпечує отримання результату (допускається у формі скріншотів).
- 5.4. Висновки.
- 5.5. Підпис виконавця, викладача, що прийняв роботу.
- 5.6. Звіт з лабораторної роботи оформлюється відповідно до вимог 3008:2015 «ЗВІТИ У СФЕРІ НАУКИ І ТЕХНІКИ. СТРУКТУРА ТА ПРАВИЛА ОФОРМЛЕННЯ».

Технічні вимоги до звіту: аркуш формату А4 шрифтом Times New Roman 12 pt через 1,0 інтервал. Поля: зверху - 2 см, знизу - 2 см, справа - 2 см, зліва - 2,5 см, абзац - 1,25 см.

VI. Звітність за лабораторну роботу.

Результатом виконання лабораторної роботи є:

- 6.1. Звіт з лабораторної роботи в електронному вигляді. Файл звіту кодується за формою:

Прізвище_Ім'я_(укр.)_номер групи_номер лр.*

- 6.2. Проект програми, що реалізує завдання лабораторної роботи, якій надається в формі архіву, як невід'ємний додаток звіту.

6.3. Оформлений звіт надається викладачеві в електронному вигляді кожним виконавцем індивідуально !

Своєчасним вважається надання звіту до початку заняття з наступної лабораторної роботи.

Оформлені звітні матеріали надсилаються за адресою:

agd015979@gmail.com

VII. Порядок оцінювання та захисту лабораторної роботи.

Максимальна кількість балів за лабораторні роботи (РЛ) за високим рівнем складає 81 бал, за середнім рівнем - 63 балів.

Загальний рейтинг за дисципліною

Звітність	Лр 1	Лр 2	Лр 3	Лр 4	Лр 5	Лр 6	Лр 7	Лр 8	Лр 9	М К	СУ МА	Зал ік	Сумма+з алік
Високий рівень	9	9	9	9	9	9	9	9	9	9	90	10	100
Середній рівень	7	7	7	7	7	7	7	7	7	9	72	10	82

Розподіл балів за виконання лабораторних робіт.

7.1. Якість / повнота оформлення протоколу з лабораторної роботи – 1 бал.

7.2. Своєчасний захист роботи – 1 бал.

7.3. Повнота аналізу отриманих результатів – 1 бал.

7.4. Якість та повнота виконання технічних умов завдання, функціональність розробленої технічної продукції (програмного скрипта) -4 бали.

7.5. Рівень теоретичної підготовки – 2 бали.

**** Для умов дистанційного навчання бали за теоретичну підготовленість (п.7.4) можуть нараховуватись за результатами аналізу вмісту протоколу з лабораторної роботи.*

**** Для умов військового стану – своєчасність захисту лабораторної роботи (п.7.2) – не застосовується а додається до п.7.4.*

професор кафедри

О. Писарчук

ДОДАТОК 1

Таблиця 1

Завдання I, II рівнів складності

Варіант (місяць народження)	I рівень складності 7 балів	II рівень складності 8 балів
1	Закон зміни похибки – рівномірний; Закон зміни досліджуваного процесу (тренду) – постійна величина. Реальні дані – 1 показник на вибір.	Закон зміни похибки – рівномірний, нормальний; Закон зміни досліджуваного процесу (тренду) – постійна, квадратичний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
2	Закон зміни похибки – нормальний; Закон зміни досліджуваного процесу (тренду) – постійна величина. Реальні дані – 1 показник на вибір.	Закон зміни похибки – нормальний, експоненційний; Закон зміни досліджуваного процесу (тренду) – постійна, квадратичний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
3	Закон зміни похибки – експонентційний; Закон зміни досліджуваного процесу (тренду) – постійна величина. Реальні дані – 1 показник на вибір.	Закон зміни похибки – експонентційний, нормальний; Закон зміни досліджуваного процесу (тренду) – постійна, квадратичний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
4	Закон зміни похибки – рівномірний; Закон зміни досліджуваного процесу (тренду) – лінійний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – χ^2 -квадрат, рівномірний; Закон зміни досліджуваного процесу – лінійний, квадратичний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
5	Закон зміни похибки – нормальний; Закон зміни досліджуваного процесу – лінійний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – нормальний, рівномірний; Закон зміни досліджуваного процесу (тренду) – лінійний, квадратичний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
6	Закон зміни похибки – експонентційний; Закон зміни досліджуваного процесу (тренду) – лінійний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – експонентційний, нормальний; Закон зміни досліджуваного процесу (тренду) – лінійний, квадратичний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
7	Закон зміни похибки – χ^2 -квадрат; Закон зміни досліджуваного процесу (тренду) – квадратичний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – рівномірний, нормальний; Закон зміни досліджуваного процесу (тренду) – квадратичний, лінійний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
8	Закон зміни похибки – нормальний; Закон зміни досліджуваного процесу (тренду) – квадратичний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – нормальний, експоненційний; Закон зміни досліджуваного процесу (тренду) – квадратичний, лінійний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
9	Закон зміни похибки – експонентційний; Закон зміни досліджуваного процесу (тренду) – квадратичний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – експонентційний, рівномірний; Закон зміни досліджуваного процесу (тренду) – квадратичний, лінійний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.

10	Закон зміни похибки – χ^2 -квадрат; Закон зміни досліджуваного процесу (тренду) – кубічний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – рівномірний, нормальний; Закон зміни досліджуваного процесу (тренду) – кубічний, лінійний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
11	Закон зміни похибки – нормальний; Закон зміни досліджуваного процесу (тренду) – кубічний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – нормальний, рівномірний; Закон зміни досліджуваного процесу (тренду) – кубічний, лінійний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.
12	Закон зміни похибки – експонентційний; Закон зміни досліджуваного процесу (тренду) – кубічний. Реальні дані – 1 показник на вибір.	Закон зміни похибки – експонентційний, нормальний; Закон зміни досліджуваного процесу (тренду) – кубічний, лінійний. Комбінаторика похибка / тренд – довільна. Реальні дані – 3 показники.

СТРУКТУРА
звіту з лабораторної роботи

Міністерство освіти і науки України
Національний технічний університет України «КПІ» імені Ігоря Сікорського
Кафедра обчислювальної техніки ФІОТ

ЗВІТ
з лабораторної роботи №1
з навчальної дисципліни «Технології Computer Vision»

Тема:

ДОСЛІДЖЕННЯ

Виконав

Перевірив

Київ 2023

I. Мета:

II. Завдання:

III. Результати виконання лабораторної роботи.

- 3.1. Синтезована математична модель;
- 3.2. Результати архітектурного проектування та їх опис;
- 3.3. Опис структури проекту програми;
- 3.4. Результати роботи програми відповідно до завдання (допускається у формі скриншотів);
- 3.5. Програмний код, що забезпечує отримання результату (допускається у формі скриншотів).

IV. Висновки.

Виконав: