

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет радіофізики, електроніки та комп'ютерних систем

Кафедра комп'ютерної інженерії

«ЗАТВЕРДЖУЮ»

Заступник декана з навчальної роботи

_____ Наталія ГОРБОВЦОВА

« ____ » _____ 2023 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Аналіз даних мовою Python

для студентів

галузь знань	12 Інформаційні технології
спеціальність	123 Комп'ютерна інженерія
рівень вищої освіти	перший (бакалаврський)
освітньо-професійна програма	Інженерія комп'ютерних систем і мереж
вид дисципліни	обов'язкова

Форма навчання	Денна
Навчальний рік	2023/2024
Семестр	3
Кількість кредитів ECTS	5
Мова викладання	українська
Форма заключного контролю	залік

Викладач:

Андрій КОНОВАЛОВ, кандидат фіз.-мат. наук, асистент кафедри комп'ютерної інженерії

Пролонговано: на 20__/20__ н. р. _____ (_____) « ____ » _____ 20__ р.

на 20__/20__ н. р. _____ (_____) « ____ » _____ 20__ р.

Розробник:

Андрій КОНОВАЛОВ, кандидат фіз.-мат. наук, асистент кафедри комп'ютерної інженерії

«ЗАТВЕРДЖЕНО»

Завідувач кафедри комп'ютерної інженерії

_____ **Юрій БОЙКО**

Протокол № ____ від « ____ » _____ 2023 р.

Схвалено науково-методичною комісією факультету радіофізики, електроніки та комп'ютерних систем

Протокол № ____ від « ____ » _____ 2023 р.

Голова науково-методичної комісії

Сергій РАДЧЕНКО

« ____ » _____ 2023 року.

ВСТУП

1. Мета дисципліни – опанування студентами основних методів і програмних інструментів аналізу даних на основі мови Python.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

Навчальна дисципліна «Аналіз даних мовою Python» базується на дисциплінах «Програмування», «Вища математика», «Теорія ймовірності та математична статистика», «Алгоритми і методи обчислень».

3. Анотація навчальної дисципліни:

У курсі «Аналіз даних мовою Python» розглядаються методи і програмний інструментарій мови Python, які використовуються під час проведення аналізу даних, зокрема: стандартна методологія аналізу даних CRISP-DM, середовища інтерактивної розробки програм (Jupyter Notebook, Google Colab), способи оцінювання часу обчислень виразів, структури Python для представлення наборів даних, numpy-масиви, бібліотека pandas для роботи з табличними даними, графічна візуалізація даних, методи виявлення викидів, обробка пропусків даних, кореляційний аналіз даних, методи декореляції і пониження розмірності даних, робота з множинами, початкові відомості про машинне навчання.

4. Завдання навчальної дисципліни (навчальні цілі):

1. Надати основні відомості із загальної методології проведення аналізу даних, методів розвідувального аналізу даних і обробки даних.
2. Розвинути у студентів навички інтерактивної роботи у Jupyter-подібних середовищах розробки програм мовою Python.
3. Ознайомити студентів із сучасними бібліотеками аналізу даних, що використовуються під час розробки програм мовою Python.
4. Дисципліна спрямована на формування програмних компетентностей:
 - ЗК3. Здатність застосовувати знання у практичних ситуаціях.
 - ЗК7. Вміння виявляти, ставити та вирішувати проблеми.
 - ФК2. Здатність використовувати сучасні методи і мови програмування для розроблення алгоритмічного та програмного забезпечення.
 - ФК13. Здатність вирішувати проблеми у галузі комп'ютерних та інформаційних технологій, визначати обмеження цих технологій.
 - ФК15. Здатність аргументувати вибір методів розв'язування спеціалізованих задач, критично оцінювати отримані результати, обґрунтовувати та захищати прийняті рішення.

5. Результати навчання за дисципліною:

Результат навчання (1, знати; 2, вміти; 3, комунікація; 4, автономність та відповідальність)		Форми (та/або методи і технології) викладання і навчання	Методи оцінювання та пороговий критерій оцінювання (за необхідності)	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання			
1	знати:	лекційні заняття	модульне тестування та залік	до 40
1.1	Структури Python для представлення наборів даних	лекція	модульне тестування та залік	10
1.2	Методи кореляційного аналізу даних	лекція	модульне тестування та залік	10
1.3	Методи декореляції і пониження розмірності даних	лекція	модульне тестування та залік	10
1.4	Принципи машинного навчання	лекція	модульне тестування та залік	10
2	вміти:	Лабораторні заняття	захист лабораторних робіт	до 40
2.1	Використовувати numpy-масиви для обробки даних	лабораторні роботи	захист лабораторних робіт	10

2.2	Проводити розвідувальний аналіз даних з використанням бібліотеки pandas	лабораторні роботи	захист лабораторних робіт	10
2.3	Понижувати розмірність вхідних даних методом PCA з використанням бібліотеки scikit-learn	лабораторні роботи	захист лабораторних робіт	10
2.4	Виявляти викиди у вхідному наборі даних та проводити обробку пропусків даних	лабораторні роботи	захист лабораторних робіт	10
3	комунікація:	лекційні та лабораторні заняття	захист лабораторних робіт	до 10
3.1	Здатність вербально пояснювати обраний метод і програмний інструмент для розв'язання задач аналізу даних	лабораторні роботи	захист лабораторних робіт	5
3.2	Здатність представляти результати аналізу даних у графічному вигляді	лабораторні роботи	захист лабораторних робіт	5
4	автономність та відповідальність:	лабораторні заняття	захист лабораторних робіт	до 10
4.1	Здатність самостійно розв'язувати задачі аналізу даних з використанням програмних інструментів і технічної документації	лабораторні роботи	захист лабораторних робіт	10

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Результати навчання дисципліни Програмні результати навчання (назва)	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	4.1
	ПРН1. Знати і розуміти наукові положення, що лежать в основі функціонування комп'ютерних засобів, систем та мереж.		+	+	+						
ПРН2. Мати навички проведення експериментів, збирання даних та моделювання в комп'ютерних системах.	+					+	+	+			
ПРН6. Вміти застосовувати знання для ідентифікації, формулювання і розв'язування технічних задач спеціальності, використовуючи методи, що є найбільш придатними для досягнення поставлених цілей.							+				+
ПРН10. Вміти розробляти програмне забезпечення для вбудованих і розподілених застосувань, мобільних і гібридних систем, розраховувати, експлуатувати типове для спеціальності обладнання.					+						
ПРН18. Використовувати інформаційні технології та для ефективного спілкування на професійному та соціальному рівнях.									+	+	
ПРН21. Якісно виконувати роботу та досягати поставленої мети з дотриманням вимог професійної етики.									+		+

7. Схема формування оцінки

7.1. Форми оцінювання

Рівень досягнення всіх запланованих результатів навчання визначається за результатами захисту лабораторних робіт, модульних тестувань та заліку. Внесок результатів навчання у підсумкову оцінку, за умови їх опанування на належному рівні:

- результати навчання 1.1 – 1.4 [знання] – до 40 %;
- результат навчання 2.1 – 2.4 [вміння] – до 40%;
- результат навчання 3.1 – 3.2 [комунікація] – до 10%;
- результат навчання 4.1 [автономність та відповідальність] – до 10%;

Форми оцінювання:

- **семестрове оцінювання:** Навчальний семестр має три змістовні модулі. Перші два змістовних модуля оцінюються після завершення лекційних тем №7 та №13 шляхом проведення електронних тестувань. Третій змістовний модуль оцінюється за результатами складання (захисту) 6 лабораторних робіт.
- **підсумкове оцінювання (у формі заліку):** форма заліку – електронне тестування (за технічної можливості) або письмова контрольна робота. Всього за залік можна отримати від 0 до 40 балів. Умовою досягнення позитивної оцінки за дисципліну є складання (захист) всіх 6 лабораторних робіт і отримання не менш ніж 60 балів підсумкової оцінки.
- **умови допуску до заліку:** умовою допуску до заліку є отримання студентом сумарно не менше, ніж *критично-розрахунковий мінімум* за семестр, рівний 20 балів. Студенти, які протягом семестру сумарно набрали меншу кількість балів, для одержання допуску до заліку обов'язково повинні добрати бали шляхом перескладання відповідних елементів семестрового оцінювання.

У випадку відсутності студента з поважних причин відпрацювання та перескладання елементів семестрового оцінювання здійснюються у відповідності до «Положення про організацію освітнього процесу у Київському національному університеті».

7.2. Організація оцінювання;

Оцінювання за формами контролю:

Семестрова робота	Кількість балів	
	Min – 20	Max –60
Модульне тестування 1	5	15
Модульна тестування 2	5	15
Захист лабораторних робіт	10	30

Орієнтований графік оцінювання:

Форма оцінювання	Орієнтовний період для здійснення відповідної форми оцінювання
Модульне тестування 1	жовтень
Модульне тестування 2	листопад
Захист лабораторних робіт	з вересня по грудень
Добір балів	грудень
Залік	грудень

Розрахунок балів, які отримують при успішній здачі заліку:

Значення	Протягом семестру	Залік	Підсумкова оцінка
Мінімум	20	10	60
Максимум	60	40	100

7.3. Шкала відповідності оцінок

Оцінка (за національною шкалою) / National grade	Рівень досягнень, % / Marks, %
Зараховано / Passed	60 – 100%
Не зараховано / Fail	0 – 59%

8. Структура навчальної дисципліни. Тематичний план лекційних занять

№ з/п	Назва теми	Кількість годин		
		Лекційні заняття	Лабораторні заняття	Самостійна робота
Змістовний модуль 1				
1	Вступ. Стандартна методологія аналізу даних. Jupyter-подібні середовища для проведення аналізу даних в інтерактивному режимі.	2		2
2	Оцінювання часу обчислень виразів.	2		2
3	Набір даних. Таблиці «об'єкти-ознаки». Типи ознак. Структури Python для представлення наборів даних.	2		2
4	NumPy-масиви. Масиви однотипних елементів.	2		2
5	Структуровані масиви та масиви записів.	2		2
6	Бібліотека pandas для роботи з табличними даними. Робота з об'єктами типу Series.	2		2
7	Робота з об'єктами типу DataFrame.	2		2
	Модульне тестування	1		2
Змістовний модуль 2				
8	Обробка даних за схемою split-apply-combine в pandas.	2		2
9	Графічна візуалізація даних. Стовпчикова діаграма, гістограма, ядра оцінка густини ймовірності (KDE).	2		2
10	Діаграма розмаху (box plot), діаграма розсіювання, матриця діаграм розсіювання.	2		2
11	Кореляційний аналіз. Наукові положення, що лежать в основі функціонування програмного інструментарію для проведення кореляційного аналізу.	2		2
12	Декореляція і пониження розмірності даних. Наукові положення, що лежать в основі комп'ютерних методів декореляції і пониження розмірності даних.	2		2
13	Робота з множинами.	2		2
14	Початкові відомості про машинне навчання. Наукові положення, що лежать в основі функціонування комп'ютерних систем з використанням технологій машинного навчання.	2		2
	Модульне тестування	1		2

Змістовний модуль 3				
15	Методи оцінювання часу обчислень виразів для прогнозування часу розв'язування технічних задач.		8	8
16	Використання numpy-масивів для розробки програмного забезпечення мовою Python.		7	7
17	Розвідувальний аналіз даних, отриманих в результаті проведення експериментів. Графічні технології представлення результатів проведеного аналізу на професійному рівні.		8	8
18	Обробка пропусків даних.		7	7
19	Виявлення викидів у реальних наборах експериментальних даних.		7	7
20	Метод PCA для пониження розмірності даних під час побудови моделей машинного навчання у комп'ютерних системах.		7	7
Всього			30	44
			76	

Загальний обсяг **150** год., в тому числі:
 Лекції **30** год.
 Лабораторні заняття **44** год.
 Самостійна робота **76** год.

9. Рекомендована література:

Основні джерела:

1. Wes McKinney. Python for Data Analysis. 3rd Edition. O'Reilly Media, 2022. – 989 pages.
2. Jake VanderPlas. Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, 2017. – 546 pages.
3. Michael Heydt. Learning pandas - Second Edition: High performance data manipulation and analysis using Python. 2nd Edition. Packt Publishing, 2017. – 703 pages.
4. В.Є. Бахрушин. Методи аналізу даних : навчальний посібник для студентів. – Запоріжжя: КПУ, 2011. – 268 с.

Додаткові джерела:

1. J. Brownlee. Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python. Machine Learning Mastery, 2020. – 398 pages.
2. D. Sarkar, R. Bali, T. Sharma. Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems. 1st Edition. Apress. 2017. – 555 pages.