

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-05.01/125.00.1.М/ВК-2023
	Екземпляр № 1	Арк 10 / 1

ЗАТВЕРДЖЕНО

Вченою радою факультету
факультету інформаційно-
комп'ютерних технологій

31 серпня 2023 р., протокол № 5



Голова Вченої ради

Тетяна НІКІТЧУК

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ ВК «МОВИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ»

для здобувачів вищої освіти освітнього ступеня «магістр»
спеціальності 125 «Кібербезпека та захист інформації»
освітньо-професійна програма «Кібербезпека»
факультет інформаційно-комп'ютерних технологій
кафедра комп'ютерної інженерії та кібербезпеки

Схвалено на засіданні
кафедри комп'ютерних наук
28 серпня 2023 р., протокол № 9

Завідувач кафедри

Марина ГРАФ

Гарант освітньо-

професійної програми

Володимир ВОРОТНІКОВ

Розробник: старший викладач кафедри комп'ютерних наук Галина МАРЧУК

Житомир
2023 – 2024 н.р.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф- 05.01/125.00.1.М/ВК- 2023
	Екземпляр № 1	Арк 10 / 2

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітній ступінь	Характеристика навчальної дисципліни	
		денна форма навчання	
Кількість кредитів <u>4</u>	Галузь знань 12 «Інформаційні технології»	За вибором (нормативна, за вибором)	
Модулів – 2	Спеціальність код спеціальності 125 «Кібербезпека та захист інформації»	Рік підготовки:	
Змістових модулів – 4		2023	2024
		Семестр	
Загальна кількість годин - <u>120</u>		1-й	2-й
		Лекції	
Тижневих годин для денної форми навчання: аудиторних – 48/16 самостійної роботи студента – 72/16		__ год.	16 год.
		Практичні	
	__ год.	__ год.	
	Лабораторні		
	год.	32 год.	
	Самостійна робота		
год.	72 год.		
Освітній ступінь «магістр»		Вид контролю: <u>залік</u>	

Співвідношення кількості годин аудиторних занять до самостійної та індивідуальної роботи становить:

для денної форми навчання – 66,7% аудиторних занять, _____ % самостійної та індивідуальної роботи;

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф- 05.01/125.00.1.М/ВК- 2023
	Екземпляр № 1	Арк 10 / 3

2. Мета та завдання навчальної дисципліни

Метою навчальної дисципліни є надання студентам теоретичних знань та практичних навичок з використання мови програмування Python для проведення інтелектуального аналізу даних. Вивчення методів інтелектуального аналізу даних, спрямованого на аналітичне дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень в сфері кібербезпеки.

Завданнями вивчення навчальної дисципліни є:

- використовувати сучасне програмно-апаратне забезпечення інформаційно-комунікаційних технологій;
- застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів.

Зміст навчальної дисципліни направлений на формування наступних **компетентностей**, визначених стандартом вищої освіти зі спеціальності 125 «Кібербезпека та захист інформації»:

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.

ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК8. Здатність генерувати нові ідеї (креативність).

ЗК11. Здатність приймати обґрунтовані рішення.

ЗК12. Здатність оцінювати та забезпечувати якість виконуваних робіт.

• спеціальні (фахові, предметні) компетентності

СК1. Здатність до математичного формулювання та досліджування неперервних та дискретних математичних моделей, обґрунтування вибору методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук, аналізу та інтерпретування.

СК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

СК7. Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф- 05.01/125.00.1.М/ВК- 2023
	Екземпляр № 1	Арк 10 / 4

складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів.

СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

3. Програма навчальної дисципліни

Модуль 1.

Змістовий модуль 1. Базові поняття аналізу даних

Тема 1. Інтелектуальний аналіз даних, інструменти та методи

Що таке аналіз даних. Інтелектуальний аналіз даних як процес. Інструменти інтелектуального аналізу даних. Задачі інтелектуального аналізу даних. Інтерпретатор Python. Бібліотеки Pandas, NumPy, Matplotlib, SciPy

Змістовий модуль 2. Базові методи аналізу даних

Тема 2. Розвідувальний аналіз.

Розвідувальний аналіз. Трактування результатів. Парадокс Сімпсона. Імпорт даних та дослідження даних (кількість вимірів, відображення першої та останньої частин об'єкта, імен об'єкта). Генеральна сукупність, вибірка. Репрезентативна та нерепрезентативна вибірки. Центральна тенденція (середнє значення, медіана, мода). Дисперсія та середньоквадратичне відхилення. Квартилі та інтерквартильний розмах. Коробчата діаграма. Візуальна аналітика. Бібліотеки Matplotlib і Seaborn. Класичні розподіли.

Тема 3. Статистична обробка часових рядів.

Часові ряди. Дискримінантний аналіз. Алгоритми дискримінантного аналізу.

Модуль 2.

Змістовий модуль 3. Машинне навчання

Тема 4. Класифікація інформації.

Формальний опис класифікації. Проста лінійна регресія. Метод найменших квадратів. Поліноміальна регресія. Теорема Байєса. Байєсовська логістична регресія. Класифікація на основі штучних нейронних мереж. Метод опорних векторів.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф- 05.01/125.00.1.М/ВК- 2023
	Екземпляр № 1	Арк 10 / 5

Тема 5. Кластерний аналіз

Базові алгоритми кластеризації. Об'єднання кластерів. Алгоритми ієрархічної кластеризації. Метод k-means.

Тема 6. Нейронні мережі. Практичні аспекти вирішення задач інтелектуального аналізу даних.

Визначення та характеристика нейронних мереж. Область їх застосування. Класифікація нейромережових архітектур. Моделі нейронних мереж. Класифікація методів навчання. Особливості сучасних нейронних мереж.

Тема 7. Математичні методи аналізу текстів

Обробка природної мови за допомогою пакета NLTK Python. Лексична токенізація. Стоп-слова. Частотний розподіл слів. Графік дисперсії. Генерація синтаксичного дерева. Аналіз настрою за допомогою Keras & NLTK. Схожість тексту та кластеризація за допомогою NLTK

Змістовий модуль 4. Python та OpenCV

Тема 8. Введення в обробку зображень в Python з OpenCV

Установка. Пошук деталей зображення. Поділ зображення на окремі канали. Фільтрація: Бінаризація по порозу, вибір; Класична фільтрація: Фур'є, ФНЧ, ФВ; Вейвлети; Кореляція; Фільтрації функцій; Фільтрації контурів. Логічна обробка результатів фільтрації: Морфологія; Контурний аналіз; Особливі точки. Навчання.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф- 05.01/125.00.1.М/ВК- 2023
	Екземпляр № 1	Арк 10 / 6

4. Структура (тематичний план) навчальної дисципліни

Змістові модулі і теми	Кількість годин							
	денна форма				заочна форма			
	усього	лекції	практичні	самостійна робота	усього	лекції	практичні	самостійна робота
Модуль 1								
Змістовий модуль 1. Базові поняття аналізу даних								
Тема 1 Інтелектуальний аналіз даних, інструменти та методи	4	2	2					
<i>Разом за змістовий модуль 1</i>	4	2	2					
Змістовий модуль 2. Базові методи аналізу даних								
Тема 2. Розвідувальний аналіз	9	2	2	5				
Тема 3. Статистична обробка часових рядів	11	2	4	5				
<i>Разом за змістовий модуль 2</i>	20	4	6	10				
Модуль 2								
Змістовий модуль 3. Машинне навчання								
Тема 4. Класифікація інформації	14	2	4	10				
Тема 5. Кластерний аналіз	16	2	4	10				
Тема 6. Нейронні мережі. Практичні аспекти вирішення задач інтелектуального аналізу даних.	24	2	6	16				
Тема 7. Математичні методи аналізу текстів	26	2	8	16				
<i>Разом за змістовий модуль 3</i>	80	8	20	52				
Змістовий модуль 4. Python та OpenCV								
Тема 15. Введення в обробку зображень в Python з OpenCV	16	2	4	10				
<i>Разом за змістовий модуль 4</i>	16	2	4	10				
РАЗОМ	120	16	32	72				

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-05.01/125.00.1.М/ВК-2023
	Екземпляр № 1	Арк 10 / 7

5. Теми практичних (лабораторних) занять

№ з/п	Назва теми	Кількість годин	
		денна форма	заочна форма
1	Python Pandas. Первинний аналіз даних	2	
2	Розвідувальний аналіз. Візуалізація даних. Python Pandas, Matplotlib, Seaborn	2	
3	Статистична обробка часових рядів. ARIMA	4	
4	Класифікація інформації.	2	
5	Регресія. Метод найменших квадратів	2	
6	Кластерний аналіз. Базові алгоритми кластеризації.	4	
7	Глибоке навчання в Python. Розпізнання цифр та зображень.	4	
8	Глибоке навчання в Python.	4	
9	Математичні методи аналізу текстів	4	
10	Розпізнавання обличчя на зображенні за допомогою Python та OpenCV	4	
РАЗОМ		32	

6. Завдання для самостійної роботи

Модуль 1

Тема 2. Розвідувальний аналіз

Попередній експрес-аналіз даних шляхом їх перетворення та/або представлення у зручному вигляді: графічному, табличному, схем, діаграм тощо. Одновимірний розподіл, двовимірний розподіл або крос-табуляція, міри центральної тенденції. Середнє (Mean). Медіана (Median). Мода (Mode). Нормальний розподіл (Normal distribution). Дисперсія і стандартне відхилення. Бімодальні розподіли. Квартилі. Нормальний розподіл (закон Гаусса). Крива нормального розподілу (крива Гауса). Візуалізація та її типи. Стандартні елементи візуалізації даних. Графіки та їх типи. Ефективні види графіків. Візуальний аналіз даних.

Тема 3. Статистична обробка часових рядів

Аналіз, моделювання та прогнозування часових рядів і економічних показників.

Модуль 2

Тема 4. Класифікація інформації

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-05.01/125.00.1.М/ВК-2023
	Екземпляр № 1	Арк 10 / 8

Тема 5. Кластерний аналіз

Кластеризація, або кластерний аналіз. Задача кластеризації. Два типи кластеризації. Алгоритми кластеризації. Методи кластеризації. Метод k-середніх. Ієрархічна кластеризація

Тема 6. Нейронні мережі. Практичні аспекти вирішення задач інтелектуального аналізу даних.

Згорткові нейронні мережі (Convolutional neural networks, CNN). Генеративні нейронні мережі (Generative adversarial network, GAN). Генерація зображень за допомогою нейромереж. Нейронні мережі та розпізнавання зображень. Нейромережа для розпізнавання осіб.

Тема 7. Математичні методи аналізу текстів

Текстова аналітика. Класифікація новин. Визначення мови.

Тема 8. Введення в обробку зображень в Python з OpenCV

Бібліотека комп'ютерного зору. Обробка зображень.

7. Індивідуальні завдання

Індивідуальні завдання передбачають створення кожним студентом (групою з кількох студентів) завершених програмних рішень для окремих сфер розробки програм. Детальна інформація про вимоги до створення і оформлення результатів роботи з виконання індивідуальних завдань міститься у методичних рекомендаціях для виконання курсових робіт. Орієнтована тематика курсових робіт наведена у таблиці.

8. Методи навчання

Основними видами занять, які проводяться під керівництвом викладача, є лекції та лабораторні роботи і самостійна робота.

На лекціях розглядаються загальні теоретичні положення дисципліни. Під час проведення лекцій використовуються мультимедійні засоби для інтерактивної демонстрації прикладів та графічного матеріалу. До кожної лекції студентам додається презентація основних положень.

При виконанні лабораторних робіт зміцнюються знання, отримані на лекціях.

9. Методи контролю

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф- 05.01/125.00.1.М/ВК- 2023
	Екземпляр № 1	Арк 10 / 9

Методи контролю успішності навчання: поточний тестовий контроль, усне опитування, захист лабораторних робіт, перевірка рівня засвоєння тем, винесених на самостійне опрацювання, залік.

10. Розподіл балів

Поточне тестування та самостійна робота								КР	Сума
Змістовий модуль 1			Змістовий модуль 2						
T1	T2	T3	T4	T5	T6	T7	T8		
5	5	10	10	10	10	10	10	30	100

Шкала оцінювання

За шкалою	Екзамен	Залік	Бали
A	Відмінно	Зараховано	90-100
B	Добре	Зараховано	82-89
C			74-81
D	Задовільно	Зараховано	64-73
E			60-63
FX	Незадовільно	Не зараховано	35-59
F		Не зараховано	0-34

11. Рекомендована література

Основна література

1. Adekanmbi O. Beginners' Artificial Intelligence and Python Programming: For Grades 4 to 8/ Xlytics Innovation Limited, 2019. — 104 p.
2. Agarwal S. Python re(gex)? a magical tool for text processing. New Delhi: Sundeep Agarwal, 2019. — 71 p.
3. Albrecht J., Ramachandran S., Winkler C. Blueprints for Text Analytics Using Python. O'Reilly Media, 2020. — 117 p.
4. Algoré Matt. Python For Data Analysis: The Ultimate and Definitive Manual to Learn Data Science and Coding With Python. Master The basics of Machine Learning, to Clean Code and Improve Artificial Intelligence. Independently published, 2021. — 87 p.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф- 05.01/125.00.1.М/ВК- 2023
	Екземпляр № 1	Арк 10 / 10

5. Ankan Ankur, Panda Abinash. Hands-On Markov Models with Python. Packt Publishing, 2018. — 178p.
6. Oliver Theobald Data Analytics for Absolute Beginners: A Deconstructed Guide to Data Literacy: (Introduction to Data, Data Visualization, Business Intelligence & Machine Learning) Paperback – July 21, 2019.- 150p.
7. Oliver Theobald Machine Learning with Python: A Practical Beginners' Guide (Machine Learning from Scratch) Paperback – October 15, 2019.-157p.
8. Oliver Theobald Machine Learning For Absolute Beginners: A Plain English Introduction (Machine Learning from Scratch) Paperback – January 1, 2018.- 192p.
9. Pointer Ian Programming PyTorch for Deep Learning: Creating and Deploying Deep Learning Applications. O'Reilly Media, Inc., 2019.-341p.
10. Matt Taddy Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions Hardcover – Illustrated, August 21, 2019.-352p.
11. Grus Joel Data Science from Scratch: First Principles with Python. O'Reilly Media, Inc.: 2019- 408p.
12. Mr Benjamin Smith (DATA ANALYTICS: A Comprehensive Beginner's Guide To Learn About The Realms Of Data Analytics From A-Z Paperback – April 26, 2020.- 153p.

Інформаційні ресурси в Інтернеті

1. Офіційний сайт державної служби статистики України [Електронний ресурс]. –Режим доступу: <http://www.ukrstat.gov.ua/>
2. Офіційний сайт департаменту статистики Організації Об'єднаних Націй [Електронний ресурс]. –Режим доступу: <http://unstats.un.org/unsd/default.htm>
3. The Analysis Factor [Electronic source]. –Access mode: <http://www.theanalysisfactor.com/>

*Індекс структурного підрозділу відповідно до наказу ректора «Про затвердження організаційної структури Державного університету «Житомирська політехніка» (наприклад, 22.06).

** Індекс освітньої програми відповідно до наказу ректора «Про індексацію освітніх програм Державного університету «Житомирська політехніка» (наприклад, 122.00.1/Б).

*** Шифр освітньої компоненти в освітній програмі (наприклад, ОК1).