

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ВНУТРІШНІХ СПРАВ**

Кафедра кібербезпеки та DATA-технологій факультету № 6

РОБОЧА ПРОГРАМА

навчальної дисципліни «**Прикладний аналіз великих даних у розслідуванні
злочинів**»

вибіркових компонент освітньої програми
першого (бакалаврського) рівня вищої освіти

072 "Фінанси, банківська справа та страхування" («Фінансова безпека та
фінансові розслідування»)

Харків 2023

ЗАТВЕРДЖЕНО

Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 30.08.2023 № 7

СХВАЛЕНО

Вченою радою факультету № 6
Протокол від 25.08.2023 № 7

ПОГОДЖЕНО

Секцією Науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від 29.08.2023 № 7

Розглянуто на засіданні кафедри кібербезпеки та DATA-технологій
факультету № 6 (*протокол від 15.08.2023 № 8*)

Розробник:

1. Професор. кафедри, к.т.н., доцент Струков В.М.

Рецензенти:

1. Провідний науковий співробітник Науково-дослідної лабораторії з проблем розвитку інформаційних технологій ХНУВС, канд.техн.наук, доцент Мордвинцев М.В.;

2. доцент кафедри програмної інженерії ХНУРЕ, кандидат технічних наук, доцент Лановий О.Ф.

1. Опис навчальної дисципліни

Найменування показників	Шифри та назви галузі знань, код та назва спеціальності, ступень вищої освіти	Характеристика навчальної дисципліни
Кількість кредитів ECTS – 3 Загальна кількість годин – 90 Кількість тем – 5	07 Управління та адміністрування; 072 – Фінанси, банківська справа та страхування перший (бакалаврський) рівень вищої освіти	Навчальний курс - 4 Семестр - 8 Види контролю: підсумковий модульний контроль – залік
Розподіл навчальної дисципліни за видами занять:		
денна форма навчання		заочна форма навчання
Лекції – <u>20</u> ; (години)		Лекції – <u>6</u> ; (години)
Семінарські заняття – _____; (години)		Семінарські заняття – _____; (години)
Практичні заняття – <u>20</u> ; (години)		Практичні заняття – <u>6</u> ; (години)
Лабораторні заняття – _____; (години)		Лабораторні заняття – _____; (години)
Самостійна робота – <u>50</u> ; (години)		Самостійна робота – <u>78</u> ; (години)

2. Мета та завдання навчальної дисципліни

Мета викладання вибіркової навчальної дисципліни «Прикладний аналіз великих даних у розслідуванні злочинів»: вивчення моделей, методів і засобів аналітичної обробки Великих Даних, а також придбання практичних навичок розв’язання задач професійної діяльності на ПЕОМ з їх використанням.

Завдання вивчення дисципліни «Прикладний аналіз великих даних у розслідуванні злочинів»:

- розвиток у студентів практичних навичок використання методів і алгоритмів Data Mining для розв’язання практичних задач за допомогою сучасних інструментальних засобів;
- підготовка студентів до ефективного розв’язання задач, що постають в подальшій професійній діяльності.

Очікувані результати навчання: у результаті вивчення навчальної дисципліни здобувач вищої освіти повинен

знати:

- специфіку проблем обробки Великих Даних,
- задачі, що виникають при обробці Великих Даних,
- етапи обробки Великих Даних;
- методи і алгоритми чіткої кластеризації;
- методи статистичного аналізу;

- інструментальні засоби обробки Великих Даних;

вміти:

- формулювати математичну модель конкретної задачі обробки Великих Даних,
- вибирати оптимальний метод для конкретної задачі обробки Великих Даних,
- виявляти тенденції і прогнозувати розвиток злочинності за допомогою стандартних інструментальних засобів;
- застосовувати інструментальні засоби для реалізації алгоритмів розв'язання задач аналізу Великих Даних у професійній діяльності.

Предметом вивчення навчальної дисципліни є вивчення моделей, методів і засобів аналітичної обробки Великих Даних, а також придбання практичних навичок розв'язання задач професійної діяльності на ПЕОМ з їх використанням.

Міждисциплінарні зв'язки: “Математика для економістів (вища математика)”, “Інформатика”, “Теорія ймовірності та математична статистика”, “Економіко-математичні методи та моделі”.

В навчальному плані для вивчення дисципліни передбачені такі організаційні форми занять як лекції і практичні заняття.

Програмні компетентності, які формуються при вивченні навчальної дисципліни:		
Інтегральна компетентність	Здатність розв'язувати складні спеціалізовані завдання та практичні проблеми у галузі фінансів, банківської справи та страхування в ході професійної діяльності або в процесі навчання, що передбачає застосування окремих методів і положень фінансової науки та характеризується невизначеністю умов і необхідністю врахування комплексу вимог здійснення професійної та навчальної діяльності.	
Загальні компетентності (ЗК)	ЗК08.	Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
Спеціальні компетентності (СК)	СК07.	Здатність складати та аналізувати фінансову звітність.
Програмні результати навчання	ПР09	Формувати і аналізувати фінансову звітність та правильно інтерпретувати отриману інформацію.

3. Програма навчальної дисципліни

Тема № 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері.

Місце, роль та задачі дисципліни. Предмет, структура і зміст дисципліни.

Технологічні передумови: Четверта технологічна революція та інформаційний вибух. Вплив технологічних факторів на правоохоронну сферу.

Актуальність і тренди аналізу Великих Даних у правоохоронній сфері: зарубіжний досвід. Прогнозування злочинності на основі Big Data. Великі Дані у протидії кіберзлочинності. Протидія фінансовим шахрайствам.

Стан аналізу Великих Даних у діяльності Національної поліції України.

Тема № 2. Задачі обробки Великих Даних.

Специфіка проблем, що виникають під час обробки Великих Даних. Задачі обробки Великих Даних: асоціація, класифікація, кластеризація, аналіз викидів.

Етапи опрацювання Великих Даних. Препроцесінг даних, аналітична обробка даних. Типи оброблюваних даних: числові, категоріальні, рангові, текстові, змішані. Оцінювання близькості точок даних у просторі. Масштабування та нормування даних.

Тема № 3. Методи і алгоритми чіткої кластеризації.

Постановка задачі кластеризації. Критерії кластеризації. Внутрішньокластерна дисперсія. Міжкластерна дисперсія. Оцінки міжкластерної відстані.

Ієрархічні алгоритми. Агломеративні алгоритми. Алгоритм K-means. Алгоритм C-means. Алгоритм J-means. Графові алгоритми.

Технологія Анакапа.

Тема № 4. Статистичне дослідження великих даних.

Основні поняття і етапи статистичного дослідження. Статистичні ряди розподілу. Ряди динаміки.

Виявлення основної тенденції і тренду зміни в часі досліджуваного явища.

Кореляція і регресія.

Основи машинного навчання.

Тема № 5. Інструментальні засоби обробки Великих Даних в поліцейській діяльності.

Стандартні загально використовувані засоби. Спеціалізовані засоби, використовувані в поліцейській діяльності: Maltego, I2, Palantir, Ricas, Watson, PredPOL. Проект Європолу ePOOLICE.

4. Структура навчальної дисципліни

4.1. Розподіл часу навчальної дисципліни за темами (денна форма навчання)

Номер та назва навчальної теми	Кількість годин, відведених на вивчення навчальної дисципліни						Вид контролю
	Всього	з них:					
		лекції	Семінарські заняття	Практичні заняття	Лабораторні заняття	Самостійна робота	
Семестр № 8							
Тема № 1: Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері.	10	4				6	к/р
Тема № 2: Задачі обробки Великих Даних.	6	2				4	
Тема № 3: Методи і алгоритми чіткої кластеризації.	12	2		4		6	
Тема № 4: Статистичне дослідження великих даних.	56	10		16		30	
Тема № 5: Інструментальні засоби обробки Великих Даних в поліцейській діяльності.	6	2				4	
Всього за семестр № 8:	90	20		20		50	залік
Всього по дисципліні	90	20		20		50	

4.3. Питання, що виносяться на самостійне опрацювання

Перелік питань до тем навчальної дисципліни			Література:
	Тема № 1: Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері.		1
		Використання систем відеофіксації в Національній поліції України.	2
		Продукти Palantir для правоохоронної сфери.	Інтернет
	Тема № 2: Задачі обробки Великих Даних.		
		Проблеми великої розмірності і трудомісткості алгоритмів обробки даних.	4,5
		Препроцесінг даних. Виділення значущих ознак.	4,5
		Препроцесінг даних. Виділення значущих ознак.	4,5
	Тема № 3: Методи і алгоритми чіткої кластеризації.		
		Агломеративні алгоритми кластеризації.	4,5
		Алгоритм K-means. Алгоритм C-means.	4,5
		Алгоритм X-means. Графові алгоритми.	4,5

	Генетичні алгоритми.	4,5
	Тема № 4. Статистичне дослідження великих даних.	
	Типи машинного навчання.	4,5
	Моделі нейронних мереж.	4,5
	Тема № 5:Інструментальні засоби обробки Великих Даних в поліцейській діяльності.	7
	Особливості аналітичних інструментів в системі І2	Інтернет
	Особливості аналітичних інструментів в системі RICAS	16
	Модуль OSINTM в системі ePOOLICE	1

5. Індивідуальні навчально-дослідні завдання

6. Методи навчання

Лекції із застосуванням електронних матеріалів; мультимедійні лекції; практичні заняття в комп'ютерному класі.

7. Перелік питань та завдань, що виносяться на підсумковий контроль

1. Технологічні передумови: Четверта технологічна революція та інформаційний вибух.
2. Вплив технологічних передумов на правоохоронну сферу.
3. Актуальність і тренди аналізу Великих Даних у правоохоронній сфері: зарубіжний досвід.
4. Прогнозування злочинності на основі Big Data.
5. Великі Дані у протидії кіберзлочинності.
6. Протидія фінансовим шахрайствам.
7. Стан аналізу Великих Даних у діяльності Національної поліції України.
8. Проблеми в задачах обробки великих масивів даних.
9. Етапи обробки великих даних.
10. Поняття графу. Вершини і ребра.
11. Матриця з'єднань і матриця відстаней.
12. Остовне дерево. Мінімальне остовне дерево.
13. Задачі і алгоритми на графах. Алгоритм Пріма.
14. Алгоритм «найближчого сусіда».
15. Технологія Анакапа.
16. Класифікація і кластеризація.
17. Препроцесінг даних.
18. Засоби відбору ознак.
19. Засоби редукції розмірності.
20. Проблема оцінки близькості об'єктів.
21. Числові метрики.
22. Метрики Мінковського.
23. Метрика Махаланобіса.
24. Категоріальна метрика.
25. Рангова метрика.

26. Масштабування значень об'єктів.
27. Нормування значень об'єктів.
28. Аналіз викидів.
29. Критерії оптимізації в задачах кластеризації.
30. Внутрішньо кластерна дисперсія.
31. Міжкластерна дисперсія.
32. Оцінки міжкластерної відстані.
33. Їєрархічні алгоритми кластеризації.
34. Агломеративні алгоритми кластеризації.
35. Алгоритм K-means.
36. Алгоритм C-means.
37. Графові алгоритми.
38. Генетичні алгоритми.
39. Етапи та методи статистичного дослідження
40. Поняття статистичного спостереження, етапи його проведення
41. Зведення та групування матеріалів статистичного спостереження
42. Поняття статистичного ряду розподілу, види рядів розподілу
43. Статистичні показники.
44. Відносні показники.
45. Середні показники.
46. Показники варіації ознаки.
47. Поняття ряду динаміки, види рядів динаміки.
48. Абсолютні та відносні показники ряду динаміки.
49. Середні показники ряду динаміки.
50. Виявлення основної тенденції зміни в часі досліджуваного явища
51. Визначення сезонних коливань рівнів ряду динаміки.
52. Прогнозування значень статистичного показника.
53. Кореляційний аналіз.
54. Регресійний аналіз.

8. Критерії та засоби оцінювання результатів навчання здобувачів

Контрольні заходи оцінювання результатів навчання включають в себе поточний та підсумковий контроль.

Засобами оцінювання результатів навчання можуть бути: екзамени (комплексні екзамени); тести; наскрізні проекти; командні проекти; аналітичні звіти, реферати, есе; розрахункові та розрахунково-графічні роботи; презентації результатів виконаних завдань та досліджень; завдання на лабораторному обладнанні, тренажерах, реальних об'єктах тощо; інші види індивідуальних та групових завдань.

Поточний контроль.

До форм поточного контролю належить оцінювання:

- рівня знань під час семінарських, практичних, лабораторних занять;
- якості виконання індивідуальної та самостійної роботи.

Поточний контроль здійснюється під час проведення семінарських,

практичних та лабораторних занять і має на меті перевірку набутих здобувачем вищої освіти (далі – здобувач) знань, умінь та інших компетентностей з навчальної дисципліни.

У ході поточного контролю проводиться систематичний вимір приросту знань, їх корекція. Результати поточного контролю заносяться викладачем до журналів обліку роботи академічної групи за національною системою оцінювання («відмінно», «добре», «задовільно», «незадовільно»).

Оцінки за самостійну та індивідуальну роботу виставляються в журналі обліку роботи академічної групи окремою графою за національною системою оцінювання («відмінно», «добре», «задовільно», «незадовільно»). Результати цієї роботи враховуються під час виставлення підсумкових оцінок.

При розрахунку успішності здобувачів в Університеті враховуються такі види робіт: навчальні заняття (семінарські, практичні, лабораторні тощо); самостійна та індивідуальна роботи (виконання домашніх завдань, ведення конспектів першоджерел та робочих зошитів, виконання розрахункових завдань, підготовка рефератів, наукових робіт, публікацій, розроблення спеціальних технічних пристроїв і приладів, моделей, комп'ютерних програм, виступи на наукових конференціях, семінарах та інше); контрольні роботи (виконання тестів, контрольних робіт у формі, передбаченій в робочою програмою навчальної дисципліни). Вони оцінюються за національною системою оцінювання («відмінно», «добре», «задовільно», «незадовільно»).

Здобувач, який отримав оцінку «незадовільно» за навчальні заняття або самостійну роботу, зобов'язаний перескласти її.

Загальна кількість балів (оцінка), отримана здобувачем за семестр перед підсумковим контролем, розраховується як середньоарифметичне значення з оцінок за навчальні заняття та самостійну роботу, та для переводу до 100-бальної системи помножується на коефіцієнт **10**.

$$\begin{array}{l} \text{Загальна кількість} \\ \text{балів (перед} \\ \text{підсумковим} \\ \text{контролем)} \end{array} = \left(\begin{array}{l} \text{Результат} \\ \text{навчальних занять} \\ \text{за семестр} \end{array} + \begin{array}{l} \text{Результат} \\ \text{самостійної} \\ \text{роботи за семестр} \end{array} \right) / 2 \cdot 10$$

Підсумковий контроль.

Підсумковий контроль проводиться з метою оцінки результатів навчання на певному ступені вищої освіти або на окремих його завершених етапах.

Для обліку результатів підсумкового контролю використовується поточно-накопичувальна інформація, яка реєструється в журналах обліку роботи академічної групи. Результати підсумкового контролю з дисциплін відображаються у відомостях обліку успішності, навчальних картках здобувачів, залікових книжках. ***Присутність здобувачів на проведенні підсумкового контролю (заліку, екзамену) обов'язкова.*** Якщо здобувач вищої освіти не з'явився на підсумковий контроль (залік, екзамен), то науково-педагогічний працівник ставить у відомість обліку успішності відмітку «не з'явився».

Підсумковий контроль (екзамен, залік) оцінюється за національною шкалою. Для переводу результатів, набраних на підсумковому контролі, з

національної системи оцінювання в 100-бальну вводиться коефіцієнт **10**, таким чином максимальна кількість балів на підсумковому контролі (екзамені, заліку), які використовуються при розрахунку успішності здобувачів, становить - **50**

Підсумкові бали з навчальної дисципліни визначаються як сума балів, отриманих здобувачем протягом семестру та балів, набраних на підсумковому контролі (екзамені, заліку).

$$\text{Підсумкові бали навчальної дисципліни} = \text{Загальна кількість балів (перед підсумковим контролем)} + \text{Кількість балів за підсумковим контролем}$$

Здобувач вищої освіти, який під час складання підсумкового контролю (екзамен, залік) отримав незадовільну оцінку, складає його повторно. Повторне складання підсумкового екзамену чи заліку допускається не більше двох разів з кожної навчальної дисципліни: один раз – викладачеві, а другий – комісії, до складу якої входить керівник відповідної кафедри та 2-3 науково-педагогічних працівника.

Якщо дисципліна вивчається протягом двох і більше семестрів з семестровим контролем у формі екзамену чи заліку, то результат вивчення дисципліни в поточному семестрі визначається як середньоарифметичне значення балів, набраних у поточному та попередньому семестрах.

$$\text{Підсумкові бали навчальної дисципліни} = \frac{\text{Підсумкові бали за поточний семестр} + \text{Підсумкові бали за попередній семестр}}{2}$$

Вимоги до здобувачів стосовно засвоєння змісту навчальної дисципліни:

Робота під час навчальних занять	Самостійна та індивідуальна робота	Підсумковий контроль
Отримати позитивні оцінки за кожну лабораторну роботу	Підготувати звіт до кожного практичного заняття, вирішити домашні завдання до практичних занять, вивчити теоретичні питання з підготовки до практичних занять.	Отримати за підсумковий контроль не менше 30 балів

9. Шкала оцінювання: національна та ECTS

Оцінка в балах	Оцінка за національною шкалою	Оцінка	
		Оцінка	Пояснення
97-100	Відмінно («зараховано»)	А	«Відмінно» – теоретичний зміст курсу засвоєний цілком , необхідні практичні навички роботи з освоєним матеріалом сформовані, усі навчальні завдання, які передбачені програмою навчання виконані в повному обсязі, відмінна робота без помилок або з однією незначною помилкою.
94-96			
90-93			

85- 89	Добре («зараховано»)	В	«Дуже добре" – теоретичний зміст курсу засвоєний цілком , потрібні практичні навички роботи з освоєним матеріалом в основному сформовані, усі навчальні завдання, які передбачені програмою навчання, виконані , якість виконання більшості з них оцінена числом балів, близьким до максимального , робота з двома-трьома незначними помилками.
80-84			
75-79		С	«Добре" – теоретичний зміст курсу засвоєний цілком , практичні навички роботи з освоєним матеріалом в основному сформовані, усі навчальні завдання, які передбачені програмою навчання, виконані , якість виконання жодного з них не оцінена мінімальним числом балів, деякі види завдань виконані з помилками , робота з декількома незначними помилками, або з однією-двома значними помилками.
70 -74	Задовільно («зараховано»)	D	«Задовільно" – теоретичний зміст курсу засвоєний частково , але прогалини не несуть істотний характер, потрібні практичні навички роботи з освоєним матеріалом в основному сформовані, більшість передбачених програмою навчання навчальних завдань виконана , деякі з виконаних завдань містять помилки , робота з трьома значними помилками.
65-69			
60-64		E	«Достатньо" – теоретичний зміст курсу освоєний частково , деякі практичні навички роботи не сформовані , частина передбачених програмою навчання навчальних завдань не виконана , або якість виконання деяких з них оцінена числом балів, близьким до мінімального , робота, що задовольняє мінімуму критеріїв оцінки.
41-59	Незадовільно («не зараховано»)	F X	«Умовно незадовільно" – теоретичний зміст курсу засвоєний частково , потрібні практичні навички роботи несформовані , більшість передбачених програмою навчання, навчальних завдань не виконано , або якість їхнього виконання оцінено числом балів, близьким до мінімального ; при додатковій самостійній роботі над матеріалом курсу можливе підвищення якості виконання навчальних завдань (з можливістю повторного складання), робота, що потребує доробки.
21-40			
1-20		F	«Безумовно незадовільно" – теоретичний зміст курсу неосвоєний , потрібні практичні навички роботи несформовані , всі виконані навчальні завдання містять грубі помилки , додаткова самостійна робота над матеріалом курсу не приведе до значного підвищення якості виконання навчальних завдань, робота, що потребує повної переробки.

10. Рекомендована література (основна, допоміжна), інформаційні ресурси в Інтернеті

Основна література.

1. Струков В.М., Узлов Д.В., Гнусов Ю.В., Колісник Т.П., Можаяев М.О. Інформаційні технології у правоохоронній діяльності. Частина 1: Високотехнологічні тренди у правоохоронній сфері зарубіжних країн. Навчальний посібник. Харків: ХНУВС, 2020. 256 с.
2. Конспект лекцій.
3. Прикладна статистика: Класифікація і зниження розмірності: Вид. [Текст] / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л.Д. Мешалкин; Під. ред. С. А. Айвазяна. - М.: Фінанси і статистика, 1989. - 607 с.

4. Чубукова, И.А. Data Mining [Текст] / И.А. Чубукова // Інтернет- університет інформаційних технологій, Біном, Лабораторія знань, 2010. - 382 с.
5. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
6. Aggarwal C.C., Reddy C.K. Data Clustering. Algorithms and Applications.- New York: CRC Press, Taylor & Francik Group, 2014. – 648p.
7. Капітонова Ю.В., Кривий С.Л., Летичевський О.А., Луцький Г.М., Печорін М.К. „Основи дискретної математики”, -Київ.: Наук.думка, 2012. – 579 с.
8. Р.М.Трохимчук. Теорія графів. Навчальний посібник для студентів факультету кібернетики - К.: РВЦ “Київський університет”, 1998. - 43 с.

Допоміжна література.

9. Шваб К. Четверта промислова революція: М.: Видавництво «Е», 2016. 206 с.
10. Шваб К. Технології четвертої промислової революції: М.: Видавництво «Ексмо», 2018. 333 с.
11. Зацеркляний М.М. Інформаційні технології у правозастосовній діяльності: Навч. посібник / М.М. Зацеркляний, В.М. Струков. : Х.: ТОВ „Східно-регіональний центр гуманітарно-освітніх ініціатив”; 2010. 332 с.
12. Узлов Д. Ю. Прикладний кримінальний аналіз на базі інформаційно-аналітичної системи «RICAS»: Методичні рекомендації щодо проведення кримінального аналізу/ Узлов Д. Ю., Дегтярьова І.В., Струков В.М. та ін.// Харків: «Юрайт», 2018. – 92 с.
13. Westphal C. Data Mining for Intelligence, Fraud and Criminal Detection. Advanced Analytic & Information Sharing Technologies / C. Westphal. – Boca Raton : CRC Press, 2009. – 426p.
14. Mena J. Investigative Data Mining for Security and Criminal Detection. – Amsterdam: Elsevier Science, 2013. – 452p.

Інформаційні ресурси в Інтернеті.

15. Третя промислова революція: сайт. URL: <https://uk.wikipedia.org/wiki/>
16. Четверта промислова революція: сайт. URL: <https://uk.wikipedia.org/wiki/>
17. Biryukov A. Neural network clustering methods to evaluate the totality of taxpayers according to their degree of creditworthiness // Artificial societies. 2017. V. 12. Issue 1-2 : сайт. URL: [http://artsoc.jes.su/s207751800000103-2-1\(circulationdate:22.07.2018\).DOI:10.18254/S0000103-2-1](http://artsoc.jes.su/s207751800000103-2-1(circulationdate:22.07.2018).DOI:10.18254/S0000103-2-1)
18. Кластерний аналіз : сайт. URL: <http://statsoft.ru/home/textbook/modules/stcluan.html>