

МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ

Харківський національний університет внутрішніх справ

Кафедра інформаційних технологій та кібербезпеки факультету № 4

РОБОЧА ПРОГРАМА

**навчальної дисципліни «Моделі, методи та засоби аналітичної обробки
великих масивів даних»**

вибіркових компонент освітньої програми
другого (магістерського) рівня вищої освіти

Спеціальність: 125 "Кібербезпека («Безпека інформаційних та комунікаційних
систем»)

м. Харків 2020 р.

ЗАТВЕРДЖЕНО

Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 23.09.2020 № 9

СХВАЛЕНО

Вченою радою факультету № 4
Протокол від 16.09.2020 № 5

ПОГОДЖЕНО

Секцією Науково-методичної ради
ХНУВС з технічних дисциплін

Протокол від 18.09.2020 № 5

Розглянуто на засіданні кафедри інформаційних технологій та кібербезпеки
(протокол від 15.09.2020 № 16)

Розробники:

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

Рецензенти:

1. д.т.н., професор Зацеркляний М.М.,
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 – 6 Загальна кількість годин – 180 Кількість тем – 5	12 Інформаційні технології; 125 – Кібербезпека другий (магістерський) рівень вищої освіти	Навчальний курс - 1 Семестр - 1 Види контролю: підсумковий модульний контроль – екзамен
Розподіл навчальної дисципліни за видами занять:		
денна форма навчання		заочна форма навчання
Лекції – <u>28</u> ; (години)		Лекції – <u>10</u> ; (години)
Семінарські заняття – <u>16</u> ; (години)		Семінарські заняття – <u>6</u> ; (години)
Практичні заняття – <u>16</u> ; (години)		Практичні заняття – <u>4</u> ; (години)
Лабораторні заняття – <u>-</u> ; (години)		Лабораторні заняття – <u>-</u> ; (години)
Самостійна робота – <u>120</u> ; (години)		Самостійна робота – <u>160</u> ; (години)

2. Мета та завдання навчальної дисципліни

Мета викладання вибіркової навчальної дисципліни «Моделі, методи та засоби аналітичної обробки великих масивів даних»: вивчення моделей, методів і засобів аналітичної обробки великих масивів даних, а також придбання практичних навичок розв’язання задач професійної діяльності на ПЕОМ з їх використанням.

Завдання вивчення дисципліни „Моделі, методи та засоби аналітичної обробки великих масивів даних”:

- розвиток у студентів практичних навичок використання методів і алгоритмів Data Mining для розв’язання практичних задач за допомогою сучасних інструментальних систем;
- 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 і алгоритми нечіткої кластеризації,

- методи і алгоритми кластеризації текстових документів,
- інструментальні засоби обробки Великих Даних в поліцейській діяльності.

вміти:

- формулювати математичну модель конкретної задачі обробки великих даних,
- вибирати оптимальний метод кластеризації для конкретної задачі обробки великих даних,
- застосовувати інструментальні засоби для реалізації алгоритмів кластеризації;
- 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 масивів даних, а також придбання практичних навичок розв'язання задач професійної діяльності на ПЕОМ з їх використанням.

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

Програмні компетентності, які формуються при вивченні навчальної дисципліни:		
Інтегральна компетентність	Здатність самостійно досліджувати і розроблювати комплексні системи забезпечення кібербезпеки, викладати і здійснювати аналітичну діяльність в області кібербезпеки.	
Загальні компетентності (КЗ)	КЗ.1.	Здатність до абстрактного, логічного, критичного мислення та встановлення взаємозв'язків між явищами та процесами.
	КЗ.2.	Здатність приймати управлінські й обґрунтовані рішення в складних і непередбачуваних умовах.
	КЗ.6.	Набуття гнучкого способу мислення, який дає можливість зрозуміти й розв'язати проблеми та задачі, зберігаючи при цьому критичне відношення до усталених наукових концепцій .
	КЗ.7.	Здатність провести усну презентацію та написати зрозумілу статтю за результатами проведених досліджень, а також щодо сучасних загроз і

		концепцій у забезпечення кібербезпеки для загальної публіки (не фахівців).
Фахові компетентності (КФ)	КФ.2.	Здатність формулювати, аналізувати та синтезувати рішення наукових проблем на абстрактному рівні шляхом декомпозиції їх на складові, які можна дослідити окремо в їх більш та менш важливих аспектах
	КФ.3.	Здатність будувати відповідні моделі сценаріїв кіберризиків, досліджувати їх для отримання нових висновків та поглиблення розуміння природи кіберзагроз
	КФ.6.	Здатність формулювати (роблячи презентації, або представляючи звіти) нові гіпотези та наукові задачі в області кібербезпеки, вибирати належні напрями і відповідні методи для їх розв'язку, беручи до уваги наявні ресурси
Програмні результати навчання	ПРН 1	Випускники будуть здатні проводити збір даних для виконання досліджень шляхом планування, виконання та аналізу експериментів.
	ПРН 2	Випускники будуть здатні аналізувати отримані результати в контексті існуючих теорій, робити відповідні висновки (включаючи ступінь невизначеності).
	ПРН 4	Випускники будуть володіти достатніми науковими навичками, принаймні в одній області кібербезпеки, для того, щоб успішно проводити наукові дослідження під наглядом наставника.
	ПРН 10	Випускники здобудуть розуміння різних інструментів та стратегій, що мають відношення до аналізу різних типів масивів великих даних на рівні, що надасть здатність ефективно використовувати на практиці теоретичні концепції системного аналізу в області кібербезпеки.

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 і етапи опрацювання великих даних.

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

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. Елементи теорії множин і теорії графів.

Поняття графу. Вершини і ребра. Матриця суміжності. Матриця інцидентності. Матриця відстаней. Орієнтовані і неорієнтовані графи. Зв'язність графів. Остовне дерево. Мінімальне остовне дерево. Задачі і алгоритми на графах. Алгоритм Пріма. Алгоритм «найближчого сусіда».

ТЕМА № 4. Статистичне дослідження великих даних.

Етапи та методи статистичного дослідження. Поняття статистичного ряду розподілу, види рядів розподілу. Статистичні показники. Статистичне вивчення динаміки правових явищ. Виявлення основної тенденції зміни в часі досліджуваного явища. Визначення сезонних коливань рівнів ряду динаміки. Прогнозування значень статистичного показника.

ТЕМА № 5. Алгоритми ієрархічної кластеризації.

Агломеративні алгоритми ієрархічної кластеризації.

Дивизимні алгоритми ієрархічної кластеризації.

ТЕМА № 6. Методи і алгоритми чіткої кластеризації.

Алгоритм K-means. Алгоритм C-means. Алгоритм J-means. Графові алгоритми. Генетичні алгоритми.

ТЕМА № 7. Інструментальні засоби обробки Великих Даних у правоохоронній 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 № 1							
Тема № 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері Проблема обробки великих масивів даних.	18	4	4			10	к/р
Тема № 2: Задачі і етапи опрацювання великих даних.	12	2	2			8	
Тема № 3: Елементи теорії множин і теорії графів.	18	4	2			12	к/р
Тема № 4.Статистичне дослідження великих даних.	36	4		8		24	
Тема № 5: Алгоритми ієрархічної кластеризації.	30	4	2	4		20	
Тема № 6: Методи і алгоритми чіткої кластеризації.	34	4	2	4		24	
Тема № 7: Інструментальні засоби обробки Великих Даних у правоохоронній діяльності.	32	6	4			22	
Всього за семестр № 1:	180	28	16	16		120	екзамен
Всього по дисципліні	180	28	16	16		120	

4.2. Розподіл часу навчальної дисципліни за темами (заочна 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 № 1							
Тема № 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері Проблема обробки великих масивів даних.	18	4		2		10	к/р
Тема № 2: Задачі і етапи опрацювання великих даних.	12	2		1		8	
Тема № 3: Елементи теорії множин і теорії графів.	18	4		1		12	к/р
Тема № 4.Статистичне дослідження великих даних.	36	4	2			24	
Тема № 5: Алгоритми ієрархічної кластеризації.	30	4	1	1		20	
Тема № 6: Методи і алгоритми чіткої кластеризації.	34	4	1	1		24	
Тема № 7: Інструментальні засоби обробки Великих Даних у правоохоронній діяльності.	32	6				22	
Всього за семестр № 1:	180	10	4	6		160	екзамен
Всього по дисципліні	180	10	4	6		160	

4.3. Питання, що виносяться на самостійне опрацювання

Перелік питань до тем навчальної дисципліни		Література:
Тема № 1: Проблема обробки великих масивів даних.	Проблеми великої розмірності і трудомісткості алгоритмів обробки даних.	1
	Різноманітність ознак і неповнота даних.	1
	Тема № 2: Етапи опрацювання великих даних.	1
	Препроцесінг даних. Виділення значущих ознак.	1
	Аналітична обробка даних.	1
	Редукція розмірності.	1
	Критерії оптимізації і відстань між об'єктами.	
	Тема № 3: Методи і алгоритми чіткої кластеризації.	2
Тема № 3: Методи і алгоритми чіткої кластеризації.	Агломеративні алгоритми кластеризації.	2
	Алгоритм K-means.	2
	Алгоритм C-means.	2
	Алгоритм X-means.	2
	Графові алгоритми.	2
	Генетичні алгоритми.	2
	Тема № 4: Методи і алгоритми нечіткої кластеризації.	2
	Нечіткий алгоритм K-means.	2
Тема № 4: Методи і алгоритми нечіткої кластеризації.	Нечіткий алгоритм C-means.	2
	Алгоритм з використанням мап Кохонена.	2
	Нейромережеві алгоритми.	2
	Тема № 5: Підходи, методи і алгоритми кластеризації текстових документів.	3
	Латентний семантичний аналіз.	3
Тема № 5: Підходи, методи і алгоритми кластеризації текстових документів.	Частотні оцінки близькості текстів.	3
	Семантичні і множинні оцінки близькості текстів.	3

5. Індивідуальні навчально-дослідні завдання

5.1. Розрахунково-графічне завдання № 1 (1 семестр).

Тема РГЗ – розробка алгоритмів обробки великих масивів даних.

6. Методи навчання

Лекції із застосуванням електронних матеріалів; мультимедійні лекції; практичні та лабораторні 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. Поняття графу. Вершини і ребра.
5. Матриця з'єднань і матриця відстаней.
6. Остовне дерево. Мінімальне остовне дерево.
7. Задачі і алгоритми на графах. Алгоритм Пріма.
8. Алгоритм «найближчого сусіда».
9. Проблеми в задачах обробки великих масивів даних.
10. Етапи обробки великих даних.
11. Класифікація і кластеризація.
12. Препроцесінг даних.
13. Засоби відбору ознак.
14. Засоби редукції розмірності.
15. Проблема оцінки близькості об'єктів.
16. Числові метрики.
17. Метрики Мінковського.
18. Метрика Махаланобіса.
19. Категоріальна метрика.
20. Рангова метрика.
21. Масштабування значень об'єктів.
22. Нормування значень об'єктів.
23. Аналіз викидів.
24. Форми представлення текстових документів.
25. Векторна форма представлення текстових документів.
26. Метрика, побудована на коефіцієнті Жаккарда.
27. Підходи до скорочення простору ознак текстових документів.
28. Застосування частотних порогів.
29. Латентний семантичний аналіз.
30. Метод головних компонент.

31. Нечіткі моделі подання даних.
32. Критерії оптимізації в задачах кластеризації.
33. Внутрішньокластерна дисперсія.
34. Міжкластерна дисперсія.
35. Оцінки міжкластерної відстані.
36. Їєрархічні алгоритми кластеризації.
37. Агломеративні алгоритми кластеризації.
38. Алгоритм K-means.
39. Алгоритм C-means.
40. Алгоритм X-means.
41. Графові алгоритми.
42. Генетичні алгоритми.
43. Приклади застосування алгоритмів кластеризації.
44. Сучасні інструментальні системи кластеризації та тенденції їх розвитку.
45. Спеціалізовані засоби, використовувані в поліцейській діяльності: Maltego, I2, Palantir, Ricas, Watson, PredPOL.
46. Проект Європолу ePOOLICE.

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).

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

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 кожної лабораторної роботи, вирішити індивідуальне завдання до кожної лабораторної роботи, 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 за шкалою ECTS	
			Оцінка	Пояснення
12	97 – 100	Відмінно ("зараховано")	A	„Відмінно” – теоретичний зміст курсу освоєний цілком , необхідні практичні навички роботи з освоєним матеріалом сформовані, всі навчальні завдання, які передбачені програмою навчання, виконані в повному обсязі, відмінна робота без помилок або з однією незначною помилкою.
11	94-96			
10	90-93			
9	85 – 89	Добре ("зараховано")	B	„Дуже добре” – теоретичний зміст курсу освоєний цілком , необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, всі навчальні завдання, які передбачені програмою навчання, виконані , якість виконання більшості з них оцінено числом балів, близьким до максимального , робота з двома-трьома незначними помилками.
8	80-84			
7	75 – 79		C	„Добре” – теоретичний зміст курсу освоєний цілком , практичні навички роботи з освоєним матеріалом в основному сформовані, всі навчальні завдання, які передбачені програмою навчання, виконані , якість виконання жодного з них не оцінено мінімальним числом балів, деякі види завдань виконані з помилками , робота з декількома незначними помилками або з однією–двома значними помилками.

6	70–74	Задовільно (“зараховано”)	D	„Задовільно” – теоретичний зміст курсу освоєний неповністю , але прогалини не носять істотного характеру, необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, більшість передбачених програмою навчання навчальних завдань виконано , деякі з виконаних завдань містять помилки , робота з трьома значними помилками.
5	65–69			
4	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 оцінено числом балів, близьким до мінімального , робота, що задовольняє мінімуму критеріїв оцінки.
3	40–59	Незадовільно („не зараховано”)	FX	„Умовно незадовільно” – теоретичний зміст курсу освоєний частково , необхідні практичні навички роботи не сформовані , більшість передбачених програм 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
2	21–40			
1	1–20		F	„Безумовно незадовільно” – теоретичний зміст курсу не освоєно , необхідні практичні навички роботи не сформовані , всі виконані навчальні завдання містять грубі помилки , додаткова самостійна робота над матеріалом курсу не приведе до значимого 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., Можаяєв М.О. Інформаційні технології у правоохоронній діяльності. Частина 1: Високотехнологічні тренди у правоохоронній сфері зарубіжних країн. Навчальний посібник. Харків: ХНУВС, 2020. 256 с.
2. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
3. Aggarwal C.C., Reddy C.K. Data Clustering. Algorithms and Applications.- New York: CRC Press, Tailor & Francik Group, 2014. – 648p.
4. Конспект лекцій.

Допоміжна література.

5. Han L., Kamber M. Data Mining: Concepts and Techniques. – Amsterdam: Morgan Kaufman Publ., 2006. – 754p.
6. Westphal C. Data Mining for Intelligence, Fraud and Criminal Detection. Advanced Analytic & Information Sharing Technologies / C. Westphal. – Boca Raton : CRC Press, 2009. – 426p.
7. Mena J. Investigative Data Mining for Security and Criminal Detection. – Amsterdam: Elsevier Science, 2003. – 452p.

Інформаційні ресурси в Інтернеті

8. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.
9. 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 [Electronic resource]. Access for registered users. 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)
10. Кластерний аналіз [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/modules/stcluan.html>