

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

Кафедра кібербезпеки та DATA-технологій, факультет № 6

РОБОЧА ПРОГРАМА

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

вибіркових компонент освітньої програми
другого (магістерського) рівня вищої освіти

Спеціальність: 072 Фінанси, банківська справа, страхування та фондовий ринок
(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:

*Професор кафедри кібербезпеки та DATA-технологій факультету № 6, 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. Канд.техн.наук, доцент, доцент кафедри програмної інженерії ХНУРЕ
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 – 4 Загальна кількість годин – 120 Кількість тем – 7	7 Управління та адміністрування 072 Фінанси, банківська справа, страхування та фондовий ринок другий (магістерський) рівень вищої освіти	Навчальний курс - 1 Семестр - 2 Види контролю: підсумковий модульний контроль – залік
Розподіл навчальної дисципліни за видами занять:		
денна форма навчання		заочна форма навчання
Лекції – <u>24</u> ; (години) Семінарські заняття – _____; (години) Практичні заняття – <u>26</u> ; (години) Лабораторні заняття – _____; (години) Самостійна робота – <u>70</u> ; (години)		Лекції – <u>6</u> ; (години) Семінарські заняття – _____; (години) Практичні заняття – <u>6</u> ; (години) Лабораторні заняття – _____; (години) Самостійна робота – <u>108</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 ПЄОМ з їх використанням.

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

- розвиток у здобувачів практичних навичок використання методів і алгоритмів Data Mining для розв’язання практичних задач за допомогою сучасних інструментальних систем;
- підготовка здобувачів 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 кластеризації текстових документів,
- засоби інструментальної мови R для аналітичної обробки Великих Даних;

- системи та інструментальні засоби обробки Великих Даних, які застосовуються в правоохоронних органах для протидії організованих злочинності.

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 засоби інструментальної мови R для аналітичної обробки Великих Даних;
- 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 заняття.

Програмні компетентності, які формуються при вивченні навчальної дисципліни:		
Інтегральна компетентність	Здатність розв'язувати складні задачі і проблеми в професійної діяльності або у процесі навчання у сфері фінансів, банківської справи та страхування, що передбачає проведення досліджень та/або здійснення інновацій та характеризується невизначеністю умов і вимог.	
Загальні компетентності (ЗК)	ЗКЗ	Здатність проведення досліджень на відповідному рівні.
Спеціальні компетентності (СК)	СК9	Здатність розробляти технічні завдання для проектування інформаційних систем у сфері фінансів, банківської справи та страхування.
Програмні результати навчання	ПР10	Здійснювати діагностику і моделювання фінансової діяльності суб'єктів господарювання.

3. Програма навчальної дисципліни

ТЕМА № 1. Технологічні тренди використання Великих Даних у кримінальній фінансовій сфері.

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

Основні напрямки використання штучного інтелекту кримінальними співтовариствами у фінансовій сфері.

Фінансові і торгові роботи. Протидія фінансовим шахрайствам в епоху Великих Даних. Боротьба з біткойн-злочинністю та використання технології блокчейн для попередження злочинів.

. 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. Підходи, методи і алгоритми кластеризації текстових документів.

Форми представлення текстових документів. Оцінки близькості текстових документів. Проблема «прокляття великої розмірності». Поняття латентного семантичного аналізу. Методи і алгоритми кластеризації текстових документів.

ТЕМА № 5. Основи нейронних мереж і машинного навчання.

Елементи нейронних мереж. Моделі нейронних мереж.

Навчання нейронних мереж. Типи систем машинного навчання. Навчання з вчителем. Навчання без вчителя. Часткове навчання. Навчання з підкріпленням. Пакетне і динамічне навчання. Навчання на основі зразків або на основі моделей. Глибоке навчання (Deep Learning).

ТЕМА № 6. Інструментальні засоби аналітичної обробки Великих Даних: мова R.

Вступ до мови R. Типи даних мови R. Опрацювання матриць і фреймів. Імпорт та експорт даних. Графічні можливості мови R.

ТЕМА № 7. Система Європолу ePOOLICE протидії організований злочинності і тероризму.

Загальна характеристика. Сканування відкритих даних в системі ePOOLICE. Розпізнавання загроз ОЗ в системі ePOOLICE. «Слабкі сигнали» для прогнозу ОЗ в системі ePOOLICE. Аналіз соціальних мереж в системі ePOOLICE. Поліцейська розвідка за відкритими джерелами в системі 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 № 2							
Тема № 1. Технологічні тренди використання Великих Даних у кримінальній фінансовій сфері.	6	2				4	к/р
Тема № 2: Задачі і етапи опрацювання великих даних.	6	2				4	
Тема № 3: Елементи теорії множин і теорії графів.	20	4		4		12	к/р
Тема № 4. Підходи, методи і алгоритми кластеризації текстових документів.	10	2		2		6	
Тема № 5: Основи нейронних мереж і машинного навчання.	12	4				8	
Тема № 6: Інструментальні засоби аналітичної обробки Великих Даних: мова R.	60	8		20		32	
Тема № 7: Система Європолу ePOOLICE протидії організованій злочинності і тероризму.	6	2				4	
Всього за семестр № 1:	120	24		26		70	залік
Всього по дисципліні	120	24		26		70	

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 № 2							

Тема № 1. Технологічні тренди використання Великих Даних у кримінальній фінансовій сфері.	6	1			5	к/р
Тема № 2: Задачі і етапи опрацювання великих даних.	6	1			5	
Тема № 3: Елементи теорії множин і теорії графів.	20			2	18	к/р
Тема № 4. Підходи, методи і алгоритми кластеризації текстових документів.	10	1			9	
Тема № 5: Основи нейронних мереж і машинного навчання.	12	1			11	
Тема № 6: Інструментальні засоби аналітичної обробки Великих Даних: мова R.	60	1		4	55	
Тема № 7: Система Європолу ePOOLICE протидії організований злочинності і тероризму.	6	1			5	
Всього за семестр № 1:	120	6		6	108	залік
Всього по дисципліні	120	6		6	108	

4.3. Питання, що виносяться на самостійне опрацювання

Перелік питань до тем навчальної дисципліни		Література:
Тема № 1: Технологічні тренди використання Великих Даних у кримінальній фінансовій сфері.	Тема № 1: Технологічні тренди використання Великих Даних у кримінальній фінансовій сфері.	1
	Проблеми великої розмірності і трудомісткості алгоритмів обробки даних.	1
	Різнотипність ознак і неповнота даних.	1
	Тема № 2: Етапи опрацювання великих даних.	2
	Препроцесінг даних. Виділення значущих ознак.	2
	Аналітична обробка даних.	2
	Редукція розмірності.	2
	Критерії оптимізації і відстань між об'єктами.	2
Тема № 3: Елементи теорії множин і теорії графів.	Тема № 3: Елементи теорії множин і теорії графів.	2
	Операції над множинами.	2
	Алгоритм Пріма.	2
	Алгоритм Краскала.	2
	Максимально-зв'язні компоненти.	2
	Тема № 4: Підходи, методи і алгоритми кластеризації текстових документів.	2
	Латентний семантичний аналіз.	2
	Частотні оцінки близькості текстів.	2
Тема № 5: Основи нейронних мереж і машинного навчання.	Семантичні і множинні оцінки близькості текстів.	2
	Тема № 5: Основи нейронних мереж і машинного навчання.	3
	Нейромережеві алгоритми.	3
	Карти Кохонена.	3
	Навчання мережі Кохонена.	3
	Тема № 6: Інструментальні засоби аналітичної обробки Великих Даних: мова R.	3
	Функції роботи з матрицями і фреймами.	3
	Впорядкування та пошук у масивах.	3

	Злиття наборів даних — merge.	3
	Тема № 7: Система Європолу ePOOLICE протидії організованим злочинності і тероризму.	1
	Розпізнавання загроз ОЗ в системі ePOOLICE.	1
	«Слабкі сигнали» для прогнозу ОЗ 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. Основні напрямки використання штучного інтелекту кримінальними співтовариствами у фінансовій сфері.
2. Фінансові і торгові роботи.
3. Протидія фінансовим шахрайствам в епоху Великих Даних.
4. Боротьба з біткойн-злочинністю та використання технології блокчейн для попередження злочинів.
5. Вплив технологічних передумов на правоохоронну сферу.
6. Актуальність і тренди аналізу Великих Даних у правоохоронній сфері: зарубіжний досвід.
7. Поняття графу. Вершини і ребра.
8. Матриця з'єднань і матриця відстаней.
9. Остовне дерево. Мінімальне остовне дерево.
10. Задачі і алгоритми на графах. Алгоритм Пріма.
11. Алгоритм «найближчого сусіда».
12. Проблеми в задачах обробки великих масивів даних.
13. Етапи обробки великих даних.
14. Класифікація і кластеризація.
15. Препроцесінг даних.
16. Засоби відбору ознак.
17. Засоби редукції розмірності.
18. Проблема оцінки близькості об'єктів.
19. Числові метрики.
20. Метрики Мінковського.
21. Метрика Махаланобіса.
22. Категоріальна метрика.
23. Рангова метрика.
24. Масштабування значень об'єктів.
25. Нормування значень об'єктів.
26. Аналіз викидів.
27. Форми представлення текстових документів.
28. Векторна форма представлення текстових документів.

29. Метрика, побудована на коефіцієнті Жаккарда.
30. Підходи до скорочення простору ознак текстових документів.
31. Застосування частотних порогів.
32. Латентний семантичний аналіз.
33. Метод головних компонент.
34. Критерії оптимізації в задачах кластеризації.
35. Внутрішньокластерна дисперсія.
36. Міжкластерна дисперсія.
37. Оцінки міжкластерної відстані.
38. Їєрархічні алгоритми кластеризації.
39. Агломеративні алгоритми кластеризації.
40. Алгоритм K-means.
41. Алгоритм C-means.
42. Базові поняття мови R.
43. Типи даних, операції та елементарні функції мови R.
44. Матриці.
45. Фрейми.
46. Векторні і матричні функції. Функція apply. Пропущені значення.
47. Функції роботи з матрицями і фреймами.
48. Впорядкування та пошук у масивах.
49. Злиття наборів даних — merge.
50. Експорт та імпорт даних у внутрішньому форматі.
51. Експорт та імпорт текстових таблиць з даними.
52. Стовпцеві та кругові діаграми.
53. Точки та лінії на площині.
54. Елементи тривимірної графіки.
55. Географічні карти.
56. Загальна характеристика системи ePOOLICE.
57. Сканування відкритих даних в системі ePOOLICE.
58. Розпізнавання загроз ОЗ в системі ePOOLICE.
59. «Слабкі сигнали» для прогнозу ОЗ в системі ePOOLICE. Аналіз соціальних мереж в системі ePOOLICE.
60. Поліцейська розвідка за відкритими джерелами в системі 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).

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

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 кожної теми 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 за шкалою ECTS	
		Оцінка	Пояснення

90 – 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.
80 – 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.
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.
65 – 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.
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.
21–59	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
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. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с.
3. Майборода Р. Є. Комп'ютерна статистика : підручник / Р. Є. Майборода. – К. : ВПЦ "Київський університет", 2019. – 589 с.
4. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
5. Aggarwal C.C., Reddy C.K. Data Clustering. Algorithms and Applications.- New York: CRC Press, Taylor & Francis Group, 2014. – 648p.
6. Капітонова Ю.В., Кривий С.Л., Летичевський О.А., Луцький Г.М., Печорін М.К. „Основи дискретної математики”, -Київ.: Наук.думка, 2012. – 579 с.
7. Р.М.Трохимчук. Теорія графів. Навчальний посібник для студентів факультету кібернетики - К.: РВЦ “Київський університет”, 1998. - 43 с.
8. Конспект лекцій.

Допоміжна література.

9. Han L., Kamber M. Data Mining: Concepts and Techniques. – Amsterdam: Morgan Kaufman Publ., 2006. – 754p.
10. Westphal C. Data Mining for Intelligence, Fraud and Criminal Detection. Advanced Analytic & Information Sharing Technologies / C. Westphal. – Boca Raton : CRC Press, 2015. – 426p.
11. Mena J. Investigative Data Mining for Security and Criminal Detection. – Amsterdam: Elsevier Science, 2013. – 452p.

Інформаційні ресурси в Інтернеті

12. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.
13. 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)
14. Кластерний аналіз [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/modules/stcluan.html>