

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ**

**Харківський національний університет внутрішніх справ**

**Кафедра інформаційних технологій та кібербезпеки факультету № 4**

**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

## **ЗАТВЕРДЖЕНО**

Науково-методичною радою  
Харківського національного  
університету внутрішніх справ  
Протокол від 24.10.2019 № 9

## **СХВАЛЕНО**

Вченою радою факультету №  
Протокол від 16.10.2019 № 9

## **ПОГОДЖЕНО**

Секцією Науково-методичної ради  
ХНУВС з технічних дисциплін

Протокол від 17.10.2019 № 10

Розглянуто на засіданні кафедри інформаційних технологій та кібербезпеки  
(протокол від 04.10.2019 № 7) )

### **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.

## ПОЯСНЮВАЛЬНА ЗАПИСКА

Програма вивчення вибіркової навчальної дисципліни «Моделі, методи та засоби аналітичної обробки великих масивів даних» складена відповідно до освітньо-професійної програми підготовки магістра за спеціальністю 125 "Кібербезпека".

**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 ПЄОМ з їх використанням.

**Міждисциплінарні зв'язки:** викладання дисципліни «Моделі, методи та засоби аналітичної обробки великих масивів даних» базується на знаннях дисциплін «Вища математика» та «Алгоритмізація та програмування».

Програма навчальної дисципліни складається з таких тем:

1. Тема № 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері.
2. ТЕМА № 2. Задачі і етапи опрацювання великих даних.
3. ТЕМА № 3. Елементи теорії множин і теорії графів.
4. ТЕМА № 4. Статистичне дослідження великих даних.
5. ТЕМА № 5. Алгоритми ієрархічної кластеризації.
6. ТЕМА № 6. Методи і алгоритми чіткої кластеризації.
7. ТЕМА № 7. Інструментальні засоби обробки Великих Даних у правоохоронній діяльності.

### **1. Мета та завдання навчальної дисципліни**

1.1. Метою викладання навчальної дисципліни «Моделі, методи та засоби аналітичної обробки великих масивів даних» є 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.

1.2. Основними завданнями вивчення дисципліни „Моделі, методи та засоби аналітичної обробки великих масивів даних” є 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.

1.3. Згідно з вимогами освітньо-професійної програми студенти повинні:  
**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.

1.4. Форма підсумкового контролю - екзамен.

На вивчення навчальної дисципліни відводиться 180 годин/6 кредитів ECTS.

1.5. Програмні компетентності:

<b>Програмні компетентності, які формуються при вивченні навчальної дисципліни:</b>		
<b>Інтегральна компетентність</b>	Здатність самостійно досліджувати і розроблювати комплексні системи забезпечення кібербезпеки, викладати і здійснювати аналітичну діяльність в області кібербезпеки.	
<b>Загальні компетентності (КЗ)</b>	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).
<b>Фахові компетентності (КФ)</b>	КФ.2.	Здатність формувати, 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 формувати (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
<b>Програмні результати навчання</b>	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.

## **2. Короткий опис змісту 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.

### **3. Рекомендована література (основна, допоміжна), інформаційні ресурси в Інтернеті**

#### **Основна література.**

1. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
2. Aggarwal C.C., Reddy C.K. Data Clustering. Algorithms and Applications.- New York: CRC Press, Taylor & Francis Group, 2014. – 648p.
3. Конспект лекцій.

#### **Допоміжна література.**

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

#### **Інформаційні ресурси в Інтернеті**

1. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.
2. 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)
3. Кластерний аналіз [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/modules/stcluan.html>

### **4. Засоби оцінювання здобувачів вищої освіти**

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.