



## «Сучасні методи аналізу даних»

**Рівень вищої освіти:** доктор філософії

**Кваліфікація:** доктор філософії, комп'ютерні науки

**Заняття:** 4 семестр

лекції: 72 год.

Практичні заняття: 48 год.

**Кількість кредитів:** 4

**Мова викладання:** українська

### Викладач(і):

**Мороз Борис Іванович**



Професор, доктор технічних наук (лекційні та практичні)

<https://pzks.nmu.org.ua/ua/teachers/teachers.php>

**E-mail:** [Moroz.B.I@nmu.one](mailto:Moroz.B.I@nmu.one)

### 1. Анотація до курсу

Сучасні методи аналізу даних - це дисципліна для вивчення процесу аналізу даних, яка заснована на пошуку в даних схованих закономірностей (шаблонів інформації). Сучасні методи аналізу даних є мультидисциплінарною областю, яка виникла і розвивається на базі таких наук як математична і статистика, розпізнавання образів, штучний інтелект. Вивчення дисципліни «Сучасні методи аналізу даних» переслідує наступні основні цілі: 1) освоєння теоретичних основ прикладного інтелектуального статистичного аналізу даних; 2) формування практичних навичок для вирішення прикладних завдань з використанням програмного забезпечення інтелектуального аналізу даних. Теоретичний матеріал курсу «Сучасні методи аналізу даних» ґрунтується на припущенні, що дані є різнотипними: безперервними, ранжируемими і номінальними. В рамках лекційного курсу розглядаються завдання, алгоритми і методи інтелектуального даних, що включають розвідувальний аналіз даних, кластеризації, класифікацію (методи машинного навчання), прогнозування, аналіз асоціацій і послідовностей

### 2. Мета та завдання курсу

**Мета дисципліни** – розвинути систему знань, умінь і навичок аспірантів в області використання комп'ютерних технологій в науці і освіті, складові основу формування компетентності аспіранта по застосуванню сучасних методів аналізу даних у професійній діяльності. Вивчення загальних принципів та методів застосування сучасних методів аналізу даних в науці..

#### **Завдання курсу:**

У результаті вивчення курсу аспіранти повинні вивчити: основні завдання і принципи інтелектуального аналізу; кореляційно-регресійний аналіз даних; дисперсійний аналіз; регресійний аналіз даних; методи машинного навчання: метод опорних векторів, дерева

класифікації, факторний, дисперсійний, компонентний, дискримінантний методи аналізу; аналіз послідовностей; методи інтелектуального аналізу для великих обсягів даних; структуру та функціональні можливості програмного забезпечення прикладного інтелектуального аналізу даних.

### 3. Результати навчання:

Отримання знань та навичок у використанні сучасних методів аналізу даних для наукових досліджень. У результаті навчання аспіранти навчатимуться: вибирати відповідні методи аналізу даних, для аналізу даних в рамках наукових досліджень; робити попередню обробку даних; визначати методи, які найбільш підходять для аналізу даних; вибирати засоби інформаційних комп'ютерних технологій для обробки даних.

### 4. Структура курсу.

<b>Види та тематика навчальних занять</b>	<b>Обсяг складових, години</b>
<b>ЛЕКЦІЇ</b>	<b>72</b>
<b>Тема 1. Кореляційно-регресійний аналіз даних</b> Завдання оцінювання параметрів. Кореляційний аналіз кількісних ознак. Кореляційний аналіз порядкових ознак. Кореляційний аналіз номінальних ознак	6
<b>Тема 2. Кореляційно-регресійний аналіз даних</b> Кореляційний аналіз змішаних ознак. Множинна кореляція. Приклади здійснення кореляційного аналізу	6
<b>Тема 3. Дисперсійний аналіз.</b> ранговий однофакторний аналіз Краскела–Уолліса, Критерій Джонкхієра, М-критерій Бартлетта, Оцінка дисперсії лінійного контрасту, G-критерій Кокрена (Кочрена), метод множинних порівнянь (Шеффе)	6
<b>Тема 4. Дисперсійний аналіз.</b> Двофакторний аналіз, F-розподіл Фішера, Критерій Пейджа, Q-критерій Кокрена, Приклад виконання дисперсійного аналізу	6
<b>Тема 5. Перевірка статистичних гіпотез.</b> Параметричні тести Z-критерій, t-критерій, Одновибірковий критерій Стюдента, Проблема Беренса – Фішера, F-критерій Фішера.	6
<b>Тема 6. Перевірка статистичних гіпотез.</b> Непараметричні тести, Критерій омега-квадрат, Критерій Смірнова, W-критерій Уїлкоксона, U-критерій Манна – Уїтні T-критерій Уїлкоксона, Критерій $\chi^2$ (хі-квадрат), Критерій серій Вальда – Волфовиця, Критерій знаків.	6
<b>Тема 7. Визначення моделей розподілу емпіричних даних</b> Критерій $\omega^2$ (Крамера – фон Мізеса), Критерій Смірнова, Приклад ідентифікації функції розподілу однорідної вибірки, Приклад ідентифікації функції розподілу неоднорідної вибірки	6
<b>Тема 8. Факторний аналіз</b> Методи факторного аналізу, Критерій Кайзера, Критерій кам'янистого осипу, Метод Якобі, Метод головних факторів	6
<b>Тема 9. Завдання та методи класифікації даних</b> Параметричні методи класифікації без навчання, Кластерний аналіз, кореляції Пірсона, Еквідова відстань. Ієрархічні (агломеративні та дивізимні), Метод середнього зв'язку Кінга	6

<b>Види та тематика навчальних занять</b>	<b>Обсяг складових, години</b>
<b>Тема 10. Завдання та методи класифікації даних</b> Класифікація з навчанням, Дискримінантний аналіз, Лінійний дискримінантний аналіз Фішера, Приклади здійснення класифікації даних	6
<b>Тема 11. Елементи кластерного аналізу</b> Визначення кластерного аналізу, Методи латентного семантичного індексування (LSA), Метод k-means, Метод суфіксних дерев, Самоорганізаційні карти	6
<b>Тема 12. Розробка моделей регресивного аналізу з використанням методології Oracle Data Miner та Oracle R Enterprise</b> Розуміння зв'язків, Введення в регресійний аналіз, OLS (ordinary least squares) основи регресії, OLS регресії які використовують Oracle Advanced Analytics, GLM та огляд регресії хребта GLM регресія використовуючи Oracle SQL та PLSQL APIs GLM регресія використовуючи Oracle R Enterprise GLM регресія використовуючи Oracle SQL Developer GUI	6
<b>ПРАКТИЧНІ ЗАНЯТТЯ</b>	<b>48</b>
<p style="text-align: center;"><b>Практична робота №1</b></p> <b>Тема: Методи класифікації даних</b> <u>Мета роботи:</u> Ознайомитися та отримати навички побудови моделей класифікації за допомогою Data Mining GUI бібліотеки Weka. <u>Завдання:</u> Побудувати модель класифікації за допомогою методу класифікації.	8
<p style="text-align: center;"><b>Практична робота №1</b></p> <b>Тема: Методи класифікації даних</b> <u>Мета роботи:</u> Ознайомитися та отримати навички побудови моделей класифікації за допомогою Data Mining GUI бібліотеки Weka. <u>Завдання:</u> Побудувати модель класифікації за допомогою методу класифікації.	8
<p style="text-align: center;"><b>Практична робота №2</b></p> <b>Тема: Класифікація даних за допомогою класифікатора Байєса</b> <u>Мета роботи:</u> Здобути навички програмної розробки методу класифікації.. <u>Завдання:</u> Розробити програму для класифікації даних.	8
<p style="text-align: center;"><b>Практична робота №3</b></p> <b>Тема: Кластеризації даних</b> <u>Мета роботи:</u> Ознайомитися та отримати навички кластеризації даних за допомогою Data Mining GUI бібліотеки Weka. <u>Завдання:</u> Виконати кластеризацію тестових даних за допомогою методу кластеризації.	8

<b>Види та тематика навчальних занять</b>	<b>Обсяг складових, години</b>
<p align="center"><b>Практична робота №4</b></p> <p><b>Тема: Розробка програмного додатку кластеризації даних</b>  <u>Мета роботи:</u> Здобути навички програмної розробки методу кластеризації даних.  <u>Завдання:</u> Розробити програму для ієрархічної кластеризації даних.</p>	8
<p align="center"><b>Практична робота №5</b></p> <p><b>Тема: Використання Weka API для регресійного аналізу</b>  <u>Мета роботи:</u> Ознайомитися та отримати навички практичного застосування методів бібліотеки Weka для регресійної класифікації даних.  <u>Завдання:</u> Розробити програму для побудови регресійної моделі даних.</p>	8
<p align="center"><b>Практична робота №6</b></p> <p><b>Тема: Методи побудови асоціативних правил</b>  <u>Мета роботи:</u> Ознайомитися та набути навички побудови асоціативних правил за допомогою Weka.  <u>Завдання:</u> Побудувати асоціативні правила для тестових екземплярів даних за допомогою зазначеного методу.</p>	8
<b>Разом</b>	<b>120</b>

## 5 . Рекомендовані джерела інформації

1. Марченко О. О., Россада Т. В. Актуальні проблеми Data Mining: навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. Київ: Київський національний університет імені Тараса Шевченка, 2017. 150 с.
2. Моделювання та оптимізація систем: підручник / [Дубовой В. М. , Кветний Р. Н. , Михальов О. І., Усов А. В. ] – Вінниця : ПП ТД «Едельвейс», 2017 – 804 с. URL: <https://nmetau.edu.ua/ua/mdiv/i2001/p24552>
3. Вовк В.М. Оптимізаційні методи і моделі : Навч. посібник / В.М. Вовк, Л.М. Зомчак. – Львів: ЛНУ імені Івана Франка, 2014. – 360 с.
4. Методи аналізу даних: навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с
5. Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеев-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 1,72 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 73 с.
6. Brendan Tierney Predictive Analytics Using Oracle Data Miner: Develop & Use Data Mining Models in Oracle Data Miner, SQL & PL/SQL: - Oracle Press, 2014. – 429 - ISBN: 978-0-07-182175-9

### Додаткова:

1. Грабовецький, Б. Є. Методи експертних оцінок: теорія, методологія, напрямки використання : монографія / Б. Є. Грабовецький. - Вінниця :ВНТУ, 2010.-171с. Режим доступу:  
<http://hrabovecky.vk.vntu.edu.ua/file/a0a40b7bd74c5d39fe693b7b2c99f38f.pdf>
2. 50 top free data mining software. URL: <https://www.predictiveanalyticstoday.com/top-free-data-mining-software/#>

### 6. Технічне обладнання та/або програмне забезпечення.

Комп'ютер Pentium і3 ОЗУ 8 МБ та краще.  
Дистанційна платформа MOODL.  
Програмний додаток (безкоштовний) Weka

## 7. Система оцінювання та вимоги

**7.1. Навчальні досягнення аспірантів за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:**

Сума балів за навчальні досягнення аспіранта	Оцінка за національною шкалою
90 – 100	відмінно
75-89	добре
60-74	задовільно
0-59	незадовільно

**7.2.** Аспіранти можуть отримати підсумкову оцінку з дисципліни на підставі поточного оцінювання знань за умови, якщо набрана кількість балів з поточного тестування та самостійної роботи складатиме не менше 60 балів.

Максимальне оцінювання:

Теоретична частина	Лабораторна частина		Бонус	Разом
	При своєчасному складанні	При несвоєчасному складанні		
65	30	20	5	<b>100</b>

Підсумковий контроль відбувається у формі письмової роботи.

Білет містить 23 запитання, з яких 20 – тестів, 3 задачі.

Практичні роботи приймаються за контрольними запитаннями до кожної з роботи.

### 7.3. Критерії оцінювання підсумкової роботи:

18 тестових завдань з чотирма варіантами відповідей, 1 правильна відповідь оцінюється у 3 бали.

Задача – 1 правильна відповідь оцінюється в 5 балів, причому

**5 балів** – відповідність еталону, з одиницями виміру;

**4 бали** – відповідність еталону, без одиниць виміру або помилками в розрахунках.

**3 бали** – незначні помилки у формулах, без одиниць виміру.

**2 бали** – присутні суттєві помилки у рішенні

**1 бал** – наведені формули повністю не відповідають еталону.

**0 балів** – рішення не наведене.

### 7.4. Критерії оцінювання лабораторної роботи:

**5 балів** – Достатня зрозумілість відповіді

**4 бали** – Добра зрозумілість відповіді

**3 бали** – Задовільна зрозумілість відповіді

**0 балів** – Незадовільна зрозумілість відповіді

## 8. Політика курсу

**8.1. Політика щодо академічної доброчесності.** Академічна доброчесність аспірантів є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату

(відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). У НТУ «Дніпровська політехніка» політика щодо академічної доброчесності регламентується положенням "Положення про систему запобігання та виявлення плагіату у Національному технічному університеті "Дніпровська політехніка" ..[http://www.nmu.org.ua/ua/content/activity/us\\_documents/System\\_of\\_prevention\\_and\\_detection\\_of\\_plagiarism.pdf](http://www.nmu.org.ua/ua/content/activity/us_documents/System_of_prevention_and_detection_of_plagiarism.pdf).

У разі порушення аспірантом академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

### **8.2. Комунікативна політика.**

Аспіранти повинні мати активовану університетську пошту.

Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

### **8.3. Політика щодо перескладання.**

Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

### **8.4. Відвідування занять.**

Для аспірантів денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, відрядження, які необхідно підтверджувати документами у разі тривалої (два тижні) відсутності. Про відсутність на занятті та причини відсутності аспірант має повідомити викладача або особисто, або через старосту. Якщо аспірант захворів, ми рекомендуємо залишатися вдома і навчатися за допомогою дистанційної платформи. Аспірантам, чий стан здоров'я є незадовільним і може вплинути на здоров'я інших аспірантів, буде пропонуватися залишити заняття (така відсутність вважатиметься пропуском з причини хвороби). Лабораторні заняття не проводяться повторно, ці оцінки неможливо отримати під час консультації. **За об'єктивних причин (наприклад, міжнародна мобільність) навчання може відбуватись в он-лайн формі за погодженням з керівником курсу.**

**8.5. Бонуси.** Аспіранти, які регулярно відвідували лекції (мають не більше двох пропусків без поважних причин) та мають написаний конспект лекцій отримують додатково 5 бали до результатів оцінювання за поточним контролем.

**8.6. Участь в анкетуванні.** Наприкінці вивчення курсу та перед початком сесії аспірантам буде запропоновано анонімно заповнити електронні анкети (MicrosoftFormsOffice 365), які буде розіслано на ваші університетські поштові скриньки. Заповнення анкет є важливою складовою вашої навчальної активності, що дозволить оцінити дієвість застосованих методів викладання та врахувати ваші пропозиції стосовно покращення змісту навчальної дисципліни.