

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ІВАНО-ФРАНКІВСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ
УНІВЕРСИТЕТ НАФТИ І ГАЗУ
Інститут інформаційних технологій
Кафедра комп'ютерних систем і мереж



ПЛАНІ СЕМІНАРСЬКИХ ЗАНЯТЬ
з дисципліни

ПЛАНУВАННЯ І ОПРАЦЮВАННЯ РЕЗУЛЬТАТІВ
ЕКСПЕРИМЕНТІВ
(назва навчальної дисципліни)

для аспірантів спеціальності

F7(123) - Комп'ютерна інженерія
(шифр і назва)

Плани семінарських занять містить рекомендації щодо їх проведення та перелік питань, що розглядаються в межах обраних тем з дисципліни: **"Планування і опрацювання результатів експериментів"**. Для студентів, що навчаються за освітньо-науковою програмою на здобуття ступеня **доктор філософії** за спеціальністю **"F7(123) - Комп'ютерна інженерія"** / Мойсеєнко О.В. - Івано-Франківськ: Видавництво ІФНТУНГ, 2025. – 20 с.

Укладач:

зав. кафедри комп'ютерних систем
і мереж, к.т.н., доцент

_____ Олена МОЙСЕЄНКО

Рецензент:

професор кафедри інженерії програмного
забезпечення, д.т.н., професор

_____ Василь ШЕКЕТА

Відповідальний за випуск:

зав. кафедри комп'ютерних систем
і мереж, д.т.н., професор

_____ Степан МЕЛЬНИЧУК

ЗМІСТ

1. Рекомендації щодо проведення семінарських занять	- 3
2. Тематичний план семінарських занять	- 3
3. План семінарського заняття №1	- 4
4. План семінарського заняття №2	- 6
5. План семінарського заняття №3	- 8
6. План семінарського заняття №4	- 10
7. План семінарського заняття №5	- 12
8. План семінарського заняття №6	- 14
9. План семінарського заняття №7	- 16
10. План семінарського заняття №8	- 18
Література	- 20

РЕКОМЕНДАЦІЇ ЩОДО ПРОВЕДЕННЯ СЕМІНАРСЬКИХ ЗАНЯТЬ

Основною метою семінарських занять є поглиблення та закріплення знань, отриманих як на заняттях так і під час самостійного вивчення відповідних тем курсу.

Опитування студентів на семінарі проходить як за бажанням студента, так і за ініціативою викладача. Крім основних виступів студентів на семінарі, викладач оцінює також істотні доповнення до викладеної проблеми.

На семінарських заняттях можуть обговорюватися повідомлення, доповіді та реферати. У кінці заняття викладач підводить підсумки роботи, оцінює виступи і доповнення кожного студента, акцентуючи увагу на найбільш вдалим відповідях, недоліках у висвітленні теми. Крім цього, студенти отримують завдання щодо підготовки до наступного семінарського заняття.

Готуючись до семінарського заняття, студент повинен опрацювати рекомендовані до даної теми джерела та літературу, продумати відповіді на кожне питання, скласти план і стислі тези свого виступу. Важливе значення у підготовці до семінару мають консультації викладачів та наукового керівника, час і місце проведення яких доводиться до відома студентів.

Готуватися до семінарського заняття необхідно заздалегідь. Всебічна та ґрунтовна підготовка – важлива передумова створення творчої атмосфери при обговоренні передбачених планом семінару питань, змістовних виступів і доповнень.

На семінарі студенти повинні доповнювати свої записи новим матеріалом з виступів своїх товаришів та викладача.

ТЕМАТИЧНИЙ ПЛАН СЕМІНАРСЬКИХ ЗАНЯТЬ

	Теми семінарських занять	Обсяг годин
М 1	Експерименти та опрацювання даних	36
1	Аспекти аналізу первинних даних	6
2	Розподіли даних і вибірок	6
3	Статистична перевірка гіпотез	6
4	Застосування кореляційно-регресійного аналізу для різних категорій даних	6
5	Особливості застосування дисперсійного аналізу	2
6	Особливості застосування факторного аналізу	2
7	Класифікація та кластеризація	4
8	Планування експериментів	4

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №1

Тема № 1. Аспекти аналізу первинних даних.

Навчальний час: 6 год.

Мета і завдання семінару: Ознайомити учасників з методами та інструментами для якісного аналізу первинних даних, а також навчити їх виявляти закономірності, обробляти та візуалізувати дані для ефективного прийняття рішень.

Питання для перевірки базових знань за темою семінару:

1. Первинні дані, їх основні типи, основні джерела первинних даних
2. Найпоширеніші методи збору первинних даних
3. У чому полягає різниця між очищенням даних та їх трансформацією?
4. Що таке структуровані дані та як вони відрізняються від непрямокутних структур?
5. Поясніть поняття "прямокутні дані"
6. Які характеристики мають кадри даних (DataFrame) у бібліотеці Pandas?
7. У чому полягає роль індексів у структурі даних?
8. Розрахунок середнього значення та коли його доцільно використовувати?
9. Визначення медіани і чому вона вважається робастною оцінкою?
10. Ідентифікація викидів у наборі даних
11. Розрахунок стандартного відхилення і що воно характеризує
12. Які переваги мають оцінки варіабельності на основі відсотків?
13. У яких випадках варто застосовувати міжквартильний розмах (IQR)?
14. Як інтерпретувати коробчасту діаграму (boxplot)?
15. Різниця між гістограмою та оцінкою густини ймовірності (KDE)?
16. Побудова частотної таблиці та які висновки можна зробити на її основі?
17. Визначення моди та коли її варто використовувати для категоріальних даних?
18. Розрахунок математичного очікування та яка його роль у статистиці?
19. У яких випадках мода є більш інформативною за середнє значення?

План семінару:

1. Ознайомлення з теоретичними основами первинного аналізу даних:
– поняття первинних даних: структура, типи, джерела.
– методи збору та попередньої обробки даних.
2. Розвідувальний аналіз даних: елементи структурованих даних; прямокутні дані; кадри даних та індекси; непрямокутні структури даних.
3. Оцінки центрального розташування: середнє; медіана та робастні оцінки; викиди (промахи).
4. Оцінки варіабельності: стандартне відхилення та пов'язані з ним оцінки; оцінки на основі відсотків.
5. Обстеження розподілу даних: відсотки і коробчасті діаграми; частотна таблиця та гістограми; оцінки густини ймовірності.
6. Обстеження двійкових та категоріальних даних: мода, математичне очікування.

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань, виконання розрахунків параметрів для запропонованих даних тощо.

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №2

Тема № 2. Розподіли даних і вибірок.

Навчальний час: 6 год.

Мета і завдання семінару: сформувати розуміння основних концепцій розподілів даних та вибірок, навчити аналізувати їх характеристики, а також застосовувати ці знання для прийняття обґрунтованих рішень у задачах обробки результатів експерименту. Ознайомити з базовими поняттями розподілу даних, навчити аналізувати вибірки та оцінювати їх характеристики візуалізовувати та інтерпретувати розподіли.

Питання для перевірки базових знань за темою семінару:

1. Зміщення у вибірці та як його уникнути?
2. Різниця між вибіркоvim середнім і популяційним середнім і коли співпадають?
3. Систематична помилка відбору та її вплив на результати досліджень
4. Як можна виявити та мінімізувати систематичну помилку відбору?
5. У чому суть та важливість центральної граничної теореми (ЦГТ)?
6. Обчислення стандартної помилки і що вона характеризує?
7. Принцип методу бутс трап, різниця між повторним відбором та бутстрапуванням.
8. Для яких типів даних бутстрап є найбільш ефективним методом оцінювання?
9. Поняття і розрахунок довірчого інтервалу і яку інформацію він надає?
10. Які фактори впливають на ширину довірчого інтервалу?
11. Особливість стандартного нормального розподілу?
12. Що таке "довгохвостий розподіл" і його важливість у обробці результатів експериментів
13. Як побудувати квантиль-квантильний графік і що він демонструє?
14. Особливості t-розподілу, коли застосовують і як відрізняють від нормального розподілу?
15. Як кількість ступенів свободи впливає на форму t-розподілу?
16. Коли використовують біноміальний розподіл і які параметри визначають його форму?
17. Що описує і коли застосовується розподіл Пуассона?
18. Як параметр λ (лямбда) впливає на форму розподілу Пуассона?
19. Основні властивості і коли використовуються експонентний розподіл?
20. Чим експонентний розподіл відрізняється від розподілу Пуассона?
21. Як визначити інтенсивність відмов у системах із випадковими відмовами?
22. Які параметри впливають на оцінку інтенсивності відмов?
23. Особливості, параметри, що визначають форму та масштаб розподілу Вейбулла?

План семінару:

1. Випадковий відбір та зміщена вибірка: зміщення; довільний вибір; вибіркoве / популяційне середнє.
2. Систематична помилка відбору: регресія до середнього; додаткові матеріали для читання.
3. Вибірковий розподіл статистик: центральна гранична теорема; стандартна помилка.
4. Бутстрап: основні поняття; повторний відбір проти бутстрапування;
5. Довірчі інтервали: поняття; розрахунок для математичного сподівання, дисперсій, часток.
6. Нормальний розподіл: стандартний нормальний розподіл та квантиль-квантильні графіки; довгохвости розподілу; t -Розподіл Стюдента.
7. Біноміальний розподіл; розподіл Пуассона та інші з ним пов'язані розподіли.
8. Експонентний розподіл.
9. Оцінка інтенсивності відмов; розподіл Вейбулла

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань тощо.

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №3

Тема № 3. Статистичні експерименти і перевірка значущості.

Навчальний час: 6 год.

Мета і завдання семінару: ознайомити із принципами планування та проведення статистичних експериментів, а також навчити коректно застосовувати методи перевірки статистичної значущості для обґрунтованих висновків у дослідженнях та практичних задачах. Навчити формулювати гіпотези, обирати статистичні критерії. Ознайомити з практичними аспектами перевірки значущості. Навчити підбирати методи відповідно до типу даних та їх розподілу.

Питання для перевірки базових знань за темою семінару:

1. Що таке статистичний експеримент і які його основні етапи?
2. Роль випадковості та контролю змінних у забезпеченні коректних результатів експерименту?
3. У чому різниця між нульовою та альтернативною гіпотезами?
4. Поясніть відмінність між помилкою I та II роду.
5. Що таке рівень значущості (α) і як він впливає на результати перевірки гіпотез?
6. Що таке потужність критерію і чому вона важлива?
7. Поясніть поняття значення критерію та критичного значення критерію.
8. Різниця між лівосторонньою, правосторонньою та двосторонньою критичними областями?
9. Що таке ступені свободи і як вони впливають на результати статистичних тестів?
10. Які основні етапи планування статистичного експерименту?
11. Які критерії необхідно врахувати під час формулювання нульової та альтернативної гіпотез?
12. У яких випадках доцільно застосовувати t-тест?
13. Інтерпретація тесту χ^2 у задачах перевірки незалежності чи відповідності розподілу?
14. Коли застосовується U-тест Манна-Вітні і чим він відрізняється від параметричних методів?
15. Що таке p-значення, і як його слід інтерпретувати?
16. Яку роль відіграє "ефект розміру" (effect size) у статистичному аналізі?
17. У чому полягає суть перестановного тесту?
18. Чим відрізняються вичерпний та бутстрапівський перестановні тести?
19. Як перестановочні тести застосовуються в обробці результатів експерименту?
20. Які методи корекції використовуються для контролю помилок у множинному тестуванні?

План семінару:

1. Поняття статистичного експерименту та його етапи: роль випадковості та контролю змінних у забезпеченні коректних результатів; нульова та альтернативна гіпотези; види помилок (I та II роду); визначення рівня значущості (α) та потужності критерію; поняття значення критерію, критичне значення критерію; ліво-, право- та двосторонні критичні області; ступені свободи.
2. Планування експерименту, формулювання гіпотез
3. Методи перевірки статистичних гіпотез: t-тест (для порівняння середніх значень); χ^2 -тест (для перевірки незалежності або відповідності розподілу); U-тест Манна-Вітні та інші непараметричні методи.
4. Аналіз залежних та незалежних вибірок: інтерпретація p-значення; аналіз ефекту розміру (effect size) для оцінки практичної значущості.
5. Повторний відбір: перестановний тест; вичерпний та бутстрапівський перестановні тести; перестановочні тести: сухий залишок для науки про дані.
6. Множинне тестування.
7. Потужність та розмір вибірки.
8. Проведення практичних кейсів. Демонстрація розрахунків у Python.

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань тощо.

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №4

Тема № 4. Застосування кореляційно-регресійного аналізу для різних категорій даних.

Навчальний час: 6 год.

Мета і завдання семінару: Сформувані у слухачів розуміння принципів кореляційно-регресійного аналізу та його особливостей при роботі з різними категоріями даних, навчити правильно обирати та застосовувати відповідні методи для виявлення залежностей та прогнозування у практичних задачах.

Питання для перевірки базових знань за темою семінару:

1. Діаграма розсіювання і як вона допомагає виявити кореляцію між змінними?
2. Кореляційне поле і як воно відображає зв'язок між величинами?
3. Як розрізнити позитивну, негативну та відсутню кореляцію на діаграмі розсіювання?
4. Лінія регресії і яку інформацію вона надає?
5. Обчислення коефіцієнту лінійної кореляції та що означає, як оцінити його достовірність?
7. Кореляційне відношення і коли його доцільно застосовувати?
8. Як визначити достовірність нелінійного кореляційного зв'язку?
9. Множинна кореляція і як вона використовується у практичному аналізі даних?
10. Фактори, що впливають на інтерпретацію множинної кореляції?
11. Які основні етапи побудови математичної моделі на основі регресійного аналізу?
12. Оцінювання достовірності коефіцієнтів у регресійній моделі
13. Які критерії оцінки адекватності побудованої регресійної моделі існують?
14. Які показники дозволяють порівняти якості різних регресійних моделей?
15. Які небезпеки пов'язані з екстраполяцією в регресійному аналізі?
16. Різниця між довірчим та передбачуваним інтервалом у прогнозуванні?
17. Що таке фіктивні змінні та як вони застосовуються в регресійних моделях?
18. Як врахувати багаторівневі та порядкові факторні змінні в регресійному аналізі?
19. Мультиколінеарність і як вона впливає на точність регресійної моделі?

План семінару:

1. Кореляція: діаграми розсіювання; кореляційне поле; кореляційна залежність між величинами.
2. Рівняння регресії: лінія регресії; коефіцієнт лінійної кореляції; достовірність коефіцієнта кореляції.
3. Нелінійний кореляційний зв'язок: кореляційні відношення; достовірність зв'язку.
4. Поняття множинної кореляції.
5. Основи регресійного аналізу: побудова математичних моделей на основі регресійного аналізу; оцінка достовірності коефіцієнтів регресійних моделей; оцінка адекватності побудованої регресійної моделі; порівняння якостей регресійних моделей.
6. Прогнозування: небезпеки екстраполяції; довірчий і передбачуваний інтервали.
7. Факторні змінні у регресії: подання фіктивні змінних; багаторівневі факторні змінні; порядкові факторні змінні.
8. Інтерпретація рівняння регресії: корельовані предиктори; мультиколінеарність; спотворювані змінні; взаємодія та головні ефекти.
9. Перевірка припущень: діагностика регресії: викиди; впливові значення; гетероскедастичність, ненормальність і корельовані помилки; графіки частинних залишків та нелінійність.
10. Нелінійна регресія: параболічна регресія; сплайнова регресія; узагальнені адитивні моделі.

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань тощо.

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №5

Тема № 5. Особливості застосування дисперсійного аналізу

Навчальний час: 2 год.

Мета і завдання семінару: Сформувати розуміння теоретичних основ та практичних аспектів застосування дисперсійного аналізу як статистичного методу, розвинути навички проведення дисперсійного аналізу та інтерпретації отриманих результатів. Завдання заняття: Розкрити теоретичні основи дисперсійного аналізу, його сутність та місце серед інших статистичних методів. Ознайомити з різними видами дисперсійного аналізу (однофакторний, двофакторний, багатофакторний). Розвинути практичні навички проведення розрахунків при дисперсійному аналізі та вміння інтерпретувати результати дисперсійного аналізу та робити обґрунтовані висновки. Розглянути аспекти застосування дисперсійного аналізу в ІТ.

Питання для перевірки базових знань за темою семінару:

1. Поняття дисперсійного аналізу і його призначення в статистиці?
2. Які основні відмінності дисперсійного аналізу від інших статистичних методів?
3. Чим відрізняється однофакторний дисперсійний аналіз від двофакторного?
4. Коли доцільно застосовувати багатофакторний дисперсійний аналіз?
5. Які переваги та недоліки має кожен із видів дисперсійного аналізу?
6. Що таке дисперсійний аналіз з повторними вимірюваннями?
7. Які основні передумови застосування дисперсійного аналізу?
8. Що таке гомогенність (однорідність) дисперсій і як її перевіряють?
9. Які наслідки можуть бути від порушення передумов дисперсійного аналізу?
10. Як формулюються нульова та альтернативна гіпотези в дисперсійному аналізі?
11. Що означає F-критерій у дисперсійному аналізі?
12. Інтерпретація р-значення при перевірці гіпотез?
13. Які типи помилок можливі при перевірці статистичних гіпотез?
14. Розрахунок загальної, міжгрупової та внутрішньогрупової варіації?
15. Які складові включає таблиця дисперсійного аналізу (ANOVA-таблиця)?
16. Обчислення F-статистика та яке її критичне значення?
17. Ступені свободи і як вони визначаються в дисперсійному аналізі?
18. Які показники використовуються для оцінки сили зв'язку між фактором і відгуком?
19. Альтернативні методи опрацювання при порушенні передумов дисперсійного аналізу?
20. Як враховувати взаємодію факторів при багатофакторному дисперсійному аналізі?

План семінару:

1. Актуальність дисперсійного аналізу в наукових дослідженнях з ІТ.
2. Огляд основних понять: дисперсія, фактори, рівні факторів.
3. Дисперсійний аналіз: основні типи однофакторний (ANOVA) / багатофакторний (MANOVA); аналіз взаємодій між факторами; гіпотези в дисперсійному аналізі; критерії значущості: F-критерій Фішера, р-значення; припущення дисперсійного аналізу (нормальність, гомогенність дисперсій, незалежність спостережень).
4. Застосування дисперсійного аналізу в ІТ: ANOVA у дослідженні продуктивності алгоритмів; аналіз ефективності різних підходів у машинному навчанні; дисперсійний аналіз у дослідженні користувацького досвіду та UX.

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань тощо.

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №6

Тема № 6. Особливості застосування факторного аналізу

Навчальний час: 2 год.

Мета і завдання семінару: Сформувати у здобувачів розуміння теоретичних основ та практичних аспектів застосування факторного аналізу як багатовимірного статистичного методу, розвинути навички проведення факторного аналізу та інтерпретації його результатів у ІТ. Завдання семінару: розкрити сутність факторного аналізу та основні завдання, які він вирішує. Ознайомити з різними моделями факторного аналізу та особливостями їх застосування. Розвинути навички проведення процедури факторного аналізу (вибір методу виділення факторів, визначення оптимальної кількості факторів, вибір методу обертання) та вміння інтерпретувати виділені фактори і оцінювати якість факторного рішення.

Питання для перевірки базових знань за темою семінару:

1. Факторний аналіз, його основні цілі та відмінність від інших статистичних методів?
2. Різниця між латентними змінними та спостережуваними змінними у факторному аналізі.
3. Різниця між розвідувальним та підтверджувальним факторним аналізом?
4. Умови застосування розвідувального та підтверджувального факторного аналізу?
5. Поняття факторних навантажень і їх інтерпретації.
6. Що таке спільності (communalities) і що вони показують?
7. Обчислення частки поясненої дисперсії?
8. Що таке власні значення (eigenvalues) і як вони використовуються у факторному аналізі?
9. Визначення достатності розміру вибірки для факторного аналізу?
10. Методи перевірки на придатність даних до факторного аналізу ви знаєте?
11. Що показує тест Бартлетта та міра вибіркової адекватності Кайзера-Мейєра-Олкіна (КМО)?
12. Етапи проведення факторного аналізу, виділення факторів, оптимальну кількість факторів?
13. Методи обертання факторів і їх призначення.
14. Оцінка якості проведеного факторного аналізу в науковій публікації?
15. Типові помилки при проведенні та інтерпретації факторного аналізу?
16. Коректність інтерпретації виділених факторів та оцінка якості факторного рішення.
17. Критерії, що використовують для оцінки значущості факторних навантажень.
18. Як визначити, чи адекватно факторна модель описує дані?

План семінару:

1. Теоретичні основи факторного аналізу: фактори, змінні, факторні навантаження; латентні змінні; кореляційні та коваріаційні матриці.
2. Основні типи факторного аналізу: експлораторний (EFA – Exploratory Factor Analysis); конфірмаційний (CFA – Confirmatory Factor Analysis).
3. Математичні засоби: модель факторного аналізу; метод головних компонент (Principal Component Analysis, PCA); метод головних факторів (Principal Factor Method); обертання факторів (Varimax, Promax).
3. Алгоритм проведення факторного аналізу: вимоги до даних (лінійність, відсутність мультиколінеарності); перевірка припущень (Kaiser-Meyer-Olkin test, Bartlett's test); вибір кількості факторів (критерій власних значень (Eigenvalues > 1), графік осипу (Scree Plot), перехресна перевірка за допомогою різних методів).
4. Інтерпретація факторних навантажень, візуалізація та оцінка факторної структури.
6. Аналіз результатів та інтерпретація отриманих факторів.
7. Основні переваги та обмеження факторного аналізу.

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань тощо.

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №7

Тема № 7. Класифікація та кластеризація

Навчальний час: 4 год.

Мета і завдання семінару: Сформувати у здобувачів фундаментальне розуміння методів класифікації та кластеризації, їх теоретичних основ, алгоритмів реалізації та практичного застосування в аналізі даних, а також розвинути навички вибору, застосування та інтерпретації результатів відповідних методів для вирішення дослідницьких та прикладних задач. Завдання семінару: Розкрити сутність понять "класифікація" та "кластеризація", їх спільні риси та принципові відмінності. Ознайомити студентів з основними видами методів класифікації (методи машинного навчання з учителем) та особливостями їх застосування. Розвинути навички підготовки даних та вміння інтерпретувати результати їх класифікації та кластеризації.

Питання для перевірки базових знань за темою семінару:

1. Відмінність між класифікацією та кластеризацією.
2. Які типи задач можна вирішувати за допомогою методів класифікації?
3. Методи k-найближчих сусідів (k-NN) та опорних векторів (SVM)
4. Принцип роботи дерев рішень, наївний класифікатор Байєса.
5. Основні алгоритми кластеризації.
6. Принцип роботи алгоритму k-середніх (k-means).
7. Ієрархічна кластеризація та її види, визначення оптимальної кількості кластерів.
8. Метрики оцінювання якості класифікації.
9. Що показує матриця помилок (confusion matrix) і як її інтерпретувати?
10. У чому полягає суть крос-валідації і навіщо вона потрібна?
11. Вплив наявності викидів на результати класифікації та кластеризації.
12. Яке значення має нормалізація даних для різних алгоритмів?
13. Як обирати ознаки для класифікації та кластеризації?
14. Переваги та недоліки різних методів класифікації при роботі з великими обсягами даних?
15. В яких випадках доцільно застосовувати ієрархічну кластеризацію, а в яких - k-means?
16. Як можна використовувати результати кластеризації для побудови систем рекомендацій?
17. Що таке перенавчання (overfitting) і як його уникнути при класифікації?
18. Які проблеми можуть виникнути при кластеризації даних з великою кількістю вимірювань?
19. Як вирішується проблема незбалансованих класів при класифікації?
20. Які методи можна використати для зменшення розмірності даних перед кластеризацією?
21. Як визначити значимість ознак для класифікації та кластеризації?

План семінару:

1. Загальні поняття: основні відмінності між класифікацією та кластеризацією; задачі класифікації та кластеризації в аналізі даних та машинному навчанні.
2. Основні типи класифікаційних алгоритмів: бінарна та багатокласова класифікація; одноетикетна та багатоетикетна (multi-label) класифікація.
3. Популярні алгоритми класифікації: логістична регресія; дерева рішень; метод опорних векторів (SVM); байєсівські класифікатори; нейронні мережі.
4. Метрики оцінки класифікаційних моделей: точність (Accuracy); повнота (Recall); точність передбачень (Precision); F1-міра; матриця помилок (Confusion Matrix).
5. Основні підходи до кластеризації: центроїдна (K-Means); ієрархічна кластеризація.
6. Оптимальна кількість кластерів: метод "Ліктя"(Elbow Method); оцінка силуету (Silhouette Score).
7. Інтерпретація результатів кластеризації, приклади комбінування класифікації і кластеризації.

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань тощо.

ПЛАН СЕМІНАРСЬКОГО ЗАНЯТТЯ №8

Тема № 8. Планування експериментів

Навчальний час: 4 год.

Мета і завдання семінару: Сформувати у здобувачів системне розуміння теоретичних основ та практичних методів планування експерименту, зокрема повного і дробового факторного експерименту, розвинути навички застосування факторних експериментів для ефективного дослідження складних систем та процесів, аналізу експериментальних даних та інтерпретації отриманих результатів. Завдання семінару: Розкрити сутність та принципи планування експерименту як наукового методу дослідження. Сформувати розуміння основних понять теорії планування експерименту: фактори, рівні факторів, відгук системи, план експерименту. Вивчити методологію повного факторного експерименту (ПФЕ), його математичний апарат та сферу застосування. Сформувати вміння критично оцінювати обмеження методів факторного експерименту та обирати оптимальний план експерименту для конкретних дослідницьких задач.

Питання для перевірки базових знань за темою семінару:

1. Планування експерименту, основні етапи і основна мета.
2. Фактори (контрольовані / неконтрольовані) та рівні факторів у контексті експерименту.
3. Які основні переваги використання планування експерименту в дослідженнях?
4. Що таке повний факторний експеримент (ПФЕ)?
5. Які основні переваги та недоліки повного факторного експерименту?
6. Як визначається кількість експериментальних комбінацій у ПФЕ?
7. Яке значення взаємодії факторів у повному факторному експерименті?
8. Як проводиться аналіз результатів у повному факторному експерименті?
9. Дробовий факторний експеримент (ДФЕ)
10. Що таке дробовий факторний експеримент (ДФЕ) і коли його доцільно використовувати?
11. Як скорочується кількість експериментальних комбінацій у ДФЕ?
12. Що таке роздільність (resolution) дробового факторного експерименту і його ролі?
13. Які ризики виникають при використанні дробового факторного експерименту?
14. Які критерії вибору між повним та дробовим факторним експериментом?
15. Як факторний аналіз допомагає оптимізувати технологічні процеси?

План семінару:

1. Сутність та принципи експериментальних досліджень: класифікація експериментів; переваги планування експерименту перед традиційними методами.
2. Опорні поняття: фактори та їх рівні; відгук системи; функція відгуку та поверхня відгуку; математична модель об'єкта дослідження.
3. Етапи планування та проведення експерименту: постановка задачі; вибір факторів та їх рівнів; побудова плану експерименту; проведення експерименту; аналіз результатів.
4. Повний факторний експеримент (ПФЕ): характеристики ПФЕ; типи планів 2^k ; побудова матриці планування ПФЕ. Властивості матриці планування; взаємодія факторів; математична модель ПФЕ; поліноміальні моделі. Розрахунок коефіцієнтів регресії.
5. Статистичний аналіз результатів ПФЕ: оцінка дисперсії відтворюваності; перевірка значущості коефіцієнтів; перевірка адекватності моделі.
6. Дробовий факторний експеримент (ДФЕ): теоретичні основи ДФЕ; регулярні дробові репліки; генеруючі співвідношення; визначаючий контраст.
7. Плани ДФЕ типу $2^{(k-p)}$: побудова матриці планування ДФЕ; принципи вибору генеруючих співвідношень; обробка та аналіз результатів ДФЕ; особливості розрахунку коефіцієнтів моделі.

Форми контролю знань: обговорення питань, доповідей, рефератів; презентація виконаних завдань тощо.

ЛІТЕРАТУРА

4.1 Основна література

1. Practical Statistics for Data Scientists. by Peter Bruce, Andrew Bruce. Released May 2017. Publisher(s): O'Reilly Media, Inc. ISBN: 9781491952962
2. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. - Запоріжжя: КПУ, 2011. - 268 с
3. Єріна А.М. Статистичне моделювання та прогнозування / А.М. Єріна. - Київ : КНТЕУ, 2001. - 196 с.
4. Мойсеєнко, О.В. Прикладна статистика та прогнозування [Текст] : лабораторний практикум / О. В. Мойсеєнко. – Івано-Франківськ : ІФНТУНГ, 2022.

4.2 Додаткова література

5. Основи теорії планування експерименту: Розділ дисципліни "Методика та організація наукових досліджень" [Електронний ресурс] : навч. посіб. для студ. освітньої програми "Комп'ютерні системи та мережі" за спеціальністю 123 "Комп'ютерна інженерія" / А.М.Волокита, В.Л.Селіванов О. А; КПП ім. Ігоря Сікорського. — Режим доступу: <https://ela.kpi.ua/server/api/core/bitstreams/7c625af3-4aba-449d-a474-66fc54071f3a/content>.
- 6 John D. Kelleher, Brendan Tierney. Data Science (Наука про дані. Базовий курс) – Режим доступу: <https://books.google.com.ua/books?id=wdPWDwAAQBAJ&printsec=frontcover&hl=uk#v=onepage&q&f=false>.
7. Нілд Томас Математика для Data Science. Керуємо даними за допомогою лінійної алгебри, теорії ймовірностей і статистики.. – Режим доступу: <https://www.twirpx.com/file/4259355/>.

4.3 Інформаційні ресурси в Інтернеті

8. Нормативна база освітнього процесу в ІФНТУНГ. Електронний ресурс: <https://nung.edu.ua/department/navchalnyy-viddil/04-normatyvno-pravova-baza-universytetu>
9. http://103.203.175.90:81/fdScript/RootOfEBooks/E%20Book%20collection%20-%202023%20-%20E/CSE%20ITAIDSML/Practical_Statistics_for_Data_Scientist_by_Peter_Bruce,_Andrew_Bruce.pdf
10. Руденко В.М. Математична статистика. Навч. посіб. [Електронний ресурс] / В.М. Руденко. – К.: Центр учбової літератури, 2012. – 304 с. Режим доступу: http://shron1.chtyvo.org.ua/Rudenko_Volodymyr/Matematychna_statystyka.pdf.
11. Щоголев С.А. Основи теорії ймовірностей та математичної статистики: навчально-методичний посібник [Електронний ресурс] / С.А. Щоголев. – Одеса : «Одеський національний університет імені І.І. Мечникова», 2015. – 206 с. – Режим доступу: http://fs.onu.edu.ua/clients/client11/web11/metod/fiz/schegolev_OTV.pdf
12. Математика для Data Science: матеріали для самопідготовки. Електронний ресурс: <https://campus.epam.ua/ua/blog/555>