



ПРИКЛАДНІ МЕТОДИ АНАЛІЗУ ДАНИХ
Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти *Третій (доктор філософії)*

Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>121 Інженерія програмного забезпечення 123 Комп'ютерна інженерія 126 Інформаційні системи та технології</i>
Освітня програма	<i>Інженерія програмного забезпечення Комп'ютерна інженерія Інформаційні системи та технології</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>очна(денна)/заочна</i>
Рік підготовки, семестр	<i>2 курс, весняний семестр</i>
Обсяг дисципліни	<i>5 кредитів / 150 годин. Денна форма: лекцій 18 годин, лаб. роб. 18 годин, СРС: 114 годин</i>
Семестровий контроль/ контрольні заходи	<i>Залік</i>
Розклад занять	<i>Згідно розкладу на весняний семестр поточного навчального року за адресою http://rozklad.kpi.ua</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com Лабораторні: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com</i>
Розміщення курсу	<i>https://classroom.google.com/c/NDEzMDE0NTY2MzQ4?cjc=ijjq23</i>

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна “Прикладні методи аналізу даних” призначена для вивчення основних методів аналізу даних. В рамках даного курсу вивчаються теми, які включають основні статистичні методи навчання: методи навчання з учителем та без учителя. Значну увагу приділено методам, які забезпечують ефективний аналіз великих наборів даних. Практична частина курсу призначена для вивчення прикладних бібліотек та програмних систем, які використовуються при машинному аналізі даних.

Метою вивчення дисципліни «Прикладні методи аналізу даних» є підготовка фахівців, здатних розв'язувати комплексні проблеми в галузі дослідницько-інноваційної діяльності у сфері

аналізу даних для прийняття оптимальних рішень та використання результатів аналізу даних для уточнення наукових висновків та формування прогнозів щодо прийнятих рішень, що передбачає глибоке розуміння підходів до створення традиційних та створення нових поведінкових моделей .

Предметом дисципліни є:

- методи аналізу даних;
- основні статистичні методи навчання;
- методи навчання з учителем, методи навчання без учителя.

Основні результати навчання

Здобувачі наукового ступеня доктора філософії після засвоєння вибіркової навчальної дисципліни мають підсилити такі **компетентності**.

1. Загальні компетентності:

- Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- Здатність розв'язувати комплексні проблеми у сфері інженерії програмного забезпечення.
- Здатність презентувати ідеї, інноваційні розробки і результати досліджень як в науковій так і в професійній спільноті.
- Здатність переосмислювати наявне та створювати нове цілісне знання та/або професійну практику і розв'язувати значущі соціальні, наукові, культурні, етичні та інші проблеми.

2. Фахові компетентності, що узагальнені відповідно до освітніх програм за спеціальностями 121, 123 та 126

Здатність виявляти, ставити та вирішувати проблеми дослідницького характеру в сфері інженерії програмного забезпечення, оцінювати та забезпечувати якість виконуваних досліджень.

- Здатність отримувати нові наукові результати, які створюють нові знання та становлять оригінальний внесок у розвиток інженерії програмного забезпечення та дотичних до неї міждисциплінарних напрямів.
- Здатність відстежувати тенденції розвитку інженерії програмного забезпечення та критично переосмислювати наявні технології.
- Здатність до розроблення нових та вдосконалення існуючих моделей, методів, засобів, процесів у сфері інженерії програмного забезпечення, які забезпечують розвиток або надають нові можливості технологіям розробки та супроводження програмного забезпечення.
- Здатність до застосування сучасних методологій, методів та інструментів інженерії програмного забезпечення в науково-педагогічній та науковій діяльності.
- Здатність проводити експериментальні дослідження з оцінювання ефективності та безпечності програмного забезпечення.

У відповідності до вищезазначеного, підсилені загальні та фахові компетенції дадуть наступні результати навчання:

- Застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи для покращення ефективності програмних систем.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Для успішного оволодіння дисципліною необхідні знання:

- основ математичного аналізу, теорії функцій та математичної статистики;
- основ функціонування операційних систем;
- основ програмування мовою Python.

Відповідно до освітньої програми необхідно попередньо оволодіти знаннями, які включають вміння програмувати алгоритмічними мовами програмування, мати базову математичну підготовку, володіти знаннями з курсу теорії множин та теорії графів. Компетентності, знання та вміння, отримані в рамках вивчення даної дисципліни, можуть бути застосовані для отримання обґрунтованих результатів досліджень та підвищення наукового рівня дисертаційних робіт.

3. Зміст навчальної дисципліни

Розділ 1. Вступ

Тема 1.1. Загальне поняття про аналіз даних та сфери застосування задач прикладного аналізу.

Тема 1.2. Порівняння методів машинного навчання та аналізу даних

Тема 1.3. Класифікація задач машинного навчання. Дедуктивний та індуктивний способи навчання.

Тема 1.4. Навчання з учителем та навчання без учителя. Навчання з підкріпленням.

Тема 1.5. Активне та пасивне навчання. Зв'язок з іншими областями науки.

Розділ 2. Загальне поняття про прикладні задачі аналізу даних

Тема 2.1. Задачі класифікації. Задача оцінювання позичальників.

Тема 2.2. Задачі регресійної оцінки. Задача прогнозування споживчого попиту.

Тема 2.3. Задачі кластеризації. Задача рубрикації текстів.

Тема 2.4. Задача виділення термінів. Задача соціологічного опитування. Задачі пошуку асоціацій.

Розділ 3. Постановка задачі аналізу даних

Тема 3.1. Основні визначення. Об'єкти та ознаки.

Тема 3.2. Типи задач аналізу даних. Поняття про модель алгоритму аналізу.

Тема 3.3. Означення методу навчання. Означення та способи формування функції втрат.

Тема 3.4. Функціонал якості. Мінімізація емпіричного ризику.

Тема 3.5. Асимптотичний приклад мінімізації емпіричного ризику.

Тема 3.6. Емпіричні оцінки спроможності методу аналізу: Hold-out володація.

Тема 3.7. Cross-володація. Leave-one-out cross-володація.

Розділ 4. Метричні методи класифікації

Тема 4.1. Гіпотеза компактності. Формалізація поняття відстані.

Тема 4.2. Постановка задачі для метричних методів класифікації.

Тема 4.3. Принципи перенумерації вибірки.

Тема 4.4. Метричний алгоритм класифікації.

Тема 4.5. Метод найближчих сусідів.

Тема 4.6. Метод k найближчих сусідів (k nearest neighbors, kNN).

Тема 4.7. Проблема вибору оптимального k . Алгоритм k зважених найближчих сусідів.

Розділ 5. Лінійні методи класифікації

Тема 5.1. Постановка задачі для лінійних методів класифікації.

Тема 5.2. Класична модель нейрона МакКаллока і Піттса та опис її роботи.

Тема 5.3. Види типових функцій активації нейрона.

Тема 5.4. Апроксимація та регуляризація емпіричного ризику у перцептронах.

Тема 5.5. Лінійна модель класифікації. Метод градієнтного спуску.

Тема 5.6. Метод стохастичного середнього градієнта. Правило Хебба.

Тема 5.7. Перцептрон Розенблатта. Теорема Новикова про кількість кроків для розділення вибірки.

Тема 5.8. Евристики для поліпшення градієнтних методів навчання. Квадратична регуляризація.

Розділ 6. Метод опорних векторів

Тема 6.1. Основні властивості методу опорних векторів.

Тема 6.2. Вибірка з лінійною роздільністю: оптимальна розділяюча гіперплощина, нормування, ширина роздільної смуги.

Тема 6.3. Лінійно не роздільна вибірка: регуляризація емпіричного ризику, двоїста задача.

Тема 6.4. Ядра і спрямляючі простори: конструктивні способи побудови ядер.

Тема 6.5. Двошарова нейронна мережа для реалізації методу опорних векторів.

Тема 6.6. Переваги та недоліки методу опорних векторів.

Розділ 7. Байєсівська теорія класифікації. Основні положення. Імовірнісна постановка задачі.

Непараметрична класифікація

Тема 7.1. Основні положення теорії імовірностей.

Тема 7.2. Функція розподілу імовірностей дискретної випадкової величини та її властивості.

Тема 7.3. Неперервні випадкові величини, умовна імовірність.

Тема 7.4. Повна імовірність, формула Байєса.

Тема 7.5. Імовірнісна постановка задачі класифікації.

Тема 7.6. Функціонал середнього ризику. Аргументи максимізації та мінімізації.

Тема 7.7. Оптимальний класифікатор Байєса. «Наївний» байєсівський класифікатор.

Тема 7.8. Непараметрична класифікація: одновимірний неперервний випадок, багатовимірний неперервний випадок.

Тема 7.9. Метод парзенівського вікна.

Розділ 8. Байєсівська теорія класифікації. Нормальний дискримінантний аналіз. Розділення суміші розподілів.

Тема 8.1. Багатовимірний нормальний розподіл.

Тема 8.2. Квадратичний дискримінант.

Тема 8.3. Лінійний дискримінант Фішера.

Тема 8.4. EM-алгоритм

Тема 8.5. Суміші багатовимірних нормальних розподілів.

Розділ 9. Логічні алгоритми класифікації

Тема 9.1. Постановка задачі для логічних методів класифікації.

Тема 9.2. Статистичне визначення інформативності.

Тема 9.3. Ентропійне визначення інформативності.

Тема 9.4. Багатокласова інформативність. Зважена інформативність.

Тема 9.5. Методи пошуку інформативних закономірностей: бінаризація кількісних ознак.

Тема 9.6. «Жадібний» алгоритм злиття зон, пошук закономірностей у формі кон'юнкцій, «градієнтний» алгоритм синтезу кон'юнкції.

Тема 9.7. «Жадібний» алгоритм синтезу кон'юнкції, стохастичний локальний пошук.

Тема 9.8. Емерджентний алгоритм синтезу кон'юнкцій, форми закономірностей.

Розділ 10. Списки та дерева ухвалення рішень

Тема 10.1. «Жадібний» алгоритм побудови списку ухвалення рішень, приклади списків ухвалення рішень.

Тема 10.2. Дерева ухвалення рішень: синтез дерев ухвалення рішень, алгоритм побудови дерева ухвалення рішень ID3.

Тема 10.3. Обробка пропусків, оцінювання ймовірностей.

Тема 10.4. Трудомісткість алгоритму ID3, переваги і недоліки алгоритму ID3.

Тема 10.5. Редукція дерев ухвалення рішень.

Тема 10.6. Перетворення дерева рішень у список рішень. Заглядання вперед.

Розділ 11. Зважене голосування правил

Тема 11.1. Принципи голосування: алгоритм простого голосування. Тема 11.2. Алгоритм зваженого голосування, налаштування ваг.

Тема 11.3. Диверсифікація правил, відмови від класифікації.

Тема 11.4. Алгоритм КОРА: принципи побудови алгоритму КОРА, його переваги та недоліки.

Тема 11.5. Алгоритм ТЕМП: принципи роботи алгоритму, переваги та недоліки алгоритму ТЕМП.

Тема 11.6. Алгоритм бустинга: експонентна апроксимація граничної функції втрат.

Тема 11.6. Принципи роботи алгоритму бустинга, переваги та недоліки алгоритму бустинга.

Розділ 12. Штучні нейронні мережі

Тема 12.1. Проблема повноти та задача «XOR». Обчислювальні можливості нейронних мереж.

Тема 12.2. Багат шарові нейронні мережі та метод зворотного поширення помилки.

Тема 12.3. Переваги та недоліки методу зворотного поширення помилки.

Тема 12.4. Евристики для покращення збіжності: вибір початкового наближення.

Тема 12.5. Вибір градієнтного методу оптимізації.

Тема 12.6. Оптимізація структури нейронної мережі: вибір кількості шарів, вибір кількості нейронів у прихованому шарі.

Тема 12.7. Динамічне додавання нейронів, видалення надлишкових зв'язків.

Розділ 13. Кластеризація. Алгоритми кластеризації

Тема 13.1. Цілі кластеризації. Типи кластерних структур.

Тема 13.2. Евристичні графові алгоритми: алгоритм виділення зв'язних компонентів.

Тема 13.3. Алгоритм найкоротшого незамкнутого шляху, алгоритм FOREL.

Тема 13.4. Функціонали якості кластеризації.

Тема 13.5. Статистичні алгоритми: гіпотеза про простір об'єктів і форму кластерів.

Тема 13.6. Метод k -середніх, кластеризація з частковим навчанням.

Тема 13.7. Ієрархічна кластеризація: властивість монотонності, теорема Міллігана.

Тема 13.8. Властивості розтягування і стискання, властивість редукованості, теорема Діде і Моро.

Тема 13.9. Визначення числа кластерів, переваги і недоліки кластеризації.

Розділ 14. Мережі Кохонена

Тема 14.1. Моделі конкурентного навчання: правило жорсткої конкуренції WTA.

Тема 14.2. Правило справедливої конкуренції CWTA, правило м'якої конкуренції WTM.

Тема 14.3. Карти Кохонена, що самоорганізуються, мистецтво інтерпретації карт Кохонена.

Тема 14.4. Недоліки карт Кохонена.

Тема 14.5. Гібридні мережі зустрічного поширення: кусочно-постійна апроксимація, гладка апроксимація.

Тема 14.6. Багатовимірне шкакування: розміщення одного об'єкта методом Ньютона-Рафсона.

Тема 14.7. Субквадратичний алгоритм багатовимірного шкакування, карта подібності, діаграма Шепарда.

Розділ 15. Методи відновлення регресії (оцінки регресії)

Тема 15.1. Метод найменших квадратів.

Тема 15.2. Непараметрична регресія: ядерне згладжування, формула Надарая-Ватсона.

Тема 15.3. Вибір ядра й ширини вікна, проблема викидів (робастна непараметрична регресія)

Тема 15.4. Алгоритм LOWESS (локально зважене згладжування), проблема крайових ефектів.

Тема 15.5. Лінійна регресія: сингулярне розкладання, проблема мультиколінеарності, гребенева регресія.

Тема 15.6. Нелінійні методи відновлення регресії: нелінійна модель регресії, нелінійні одновимірні перетворення ознак.

4. Навчальні матеріали та ресурси

4.1. Базова література

1. Новотарський М.А. Прикладні методи аналізу даних / навчальний посібник для аспірантів // <https://comsys.kpi.ua/upload/PMAD7.pdf>
2. Новотарський М.А. Методичні матеріали до лабораторних робіт з дисципліни «Прикладні методи аналізу даних» / для PhD студентів // <https://t.ly/u68Dc>
3. Новотарський М. А., Нестеренко Б. Б. Штучні нейронні мережі: обчислення. К: Ін-т математики НАН України, 2004. – 408 с.
4. Завантаження та інсталяція Weka: <https://www.onworks.net/uk/software/windows/app-weka>
5. Pentaho Data Mining Community Documentation: <https://pentaho-public.atlassian.net/wiki/spaces/DATAMINING/overview?homepageId=276037757%2F1000>
6. Кузьмич, В. А. Методи та засоби математичного моделювання руху рідин з використанням машинного навчання : дис. ... д-ра філософії : 123 Комп'ютерна інженерія / Кузьмич Валентин Анатолійович. – Київ, 2023. – 213 с. <https://ela.kpi.ua/handle/123456789/64826>
7. Новотарський, М. А. Основи науки про дані [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» / М. А. Новотарський ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 3.95 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2022. – 294 с. <https://ela.kpi.ua/handle/123456789/49912>
8. Новотарський, М. А. Основи програмування алгоритмічною мовою Python [Електронний ресурс] : навч. посіб. для студ. освітньої програми «Комп'ютерні системи та мережі» спеціальності 123 «Комп'ютерна інженерія» / М. А. Новотарський ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 17.93 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2022. – 701 с. <https://ela.kpi.ua/handle/123456789/49913>
9. Nesterenko B. V., Novotarskyi M. A. Formal Means of the Simulation of Parallel Processes and Systems. – К: РН “Академперіодика”, 2016. – 194 р. DOI: <https://doi.org/10.15407/akademperiodyka.309.194>

4.2. Додаткова література

1. Підручник з алгоритмічної мови Python // <https://docs.python.org/uk/3/tutorial/index.html>
2. Підручник з Anaconda // <https://www.ikkaro.com/uk/anaconda/>
3. PyCharm // <https://ubunlog.com/uk/pycharm-ide-python-ppa/>
4. Машинне навчання // <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>
5. TensorFlow tutorials // <https://www.tensorflow.org/tutorials>
6. Novotarskyi M., Stirenko S. , Gordoenko Y. Simulation of 3D transient flow passing through an intestinal anastomosis by Lattice-Boltzmann method // Radio Electronics, Computer Science, Control.– 2018.– №1, P.57-82, DOI: <https://doi.org/10.15588/1607-3274-2018-1-9>
7. Novotarskyi M., Stirenko S. , Gordoenko Y., Kuzmych V. Deep reinforcement learning with sparse distributed memory for “Water World” problem solving // Radio Electronics, Computer Science, Control.– 2021.– № 1, P.136-143, DOI: <https://doi.org/10.15588/1607-3274-2021-1-14>.

Інформаційні ресурси

1. Прикладні методи аналізу даних / дистанційний курс для студентів Phd // <https://classroom.google.com/c/NDEzMDE0NTY2MzQ4?cjc=ijijq23>

5. Методика опанування навчальної дисципліни (освітнього компоненту)

Розподіл навчального часу за видами занять і завдань з дисципліни згідно з робочим навчальним планом. На кредитний модуль виділено 120 годин та 4 кредитів – 18 лекцій (36 годин), 7 лабораторних робіт (18 годин)

Для досягнення мети навчальної дисципліни слід зосередитись в лекційному матеріалі на розгляді теоретичних основ статистичних методів машинного навчання. Такий розгляд має відбуватися починаючи з визначення базових понять, що застосовуються у рамках дисципліни «Прикладні методи аналізу даних». Подальший розгляд відбувається у послідовності від простих методів до складніших методів. Першими розглядаються метричні та лінійні методи класифікації. Курс лекцій завершується розглядом типових задач машинного навчання

Таблиця 1.

Таблиця розподілу навантаження

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Розділ 1. Вступ Тема 1.1. Загальне поняття про аналіз даних та сфери застосування задач прикладного аналізу. Тема 1.2. Порівняння методів машинного навчання та аналізу даних Тема 1.3. Класифікація задач машинного навчання. Дедуктивний та індуктивний способи навчання. Тема 1.4. Навчання з учителем та навчання без учителя. Навчання з підкріпленням. Тема 1.5. Активне та пасивне навчання. Зв'язок з іншими областями науки.	8	1	0	7
Розділ 2. Загальне поняття про прикладні задачі аналізу даних Тема 2.1. Задачі класифікації. Задача оцінювання позичальників. Тема 2.2. Задачі регресійної оцінки. Задача прогнозування споживчого попиту. Тема 2.3. Задачі кластеризації. Задача рубрикації текстів. Тема 2.4. Задача виділення термінів. Задача соціологічного опитування. Задачі пошуку асоціацій.	6	1	0	5

<p>Розділ 3. Постановка задачі аналізу даних</p> <p>Тема 3.1. Основні визначення. Об'єкти та ознаки.</p> <p>Тема 3.2. Типи задач аналізу даних. Поняття про модель алгоритму аналізу.</p> <p>Тема 3.3. Означення методу навчання. Означення та способи формування функції втрат.</p> <p>Тема 3.4. Функціонал якості. Мінімізація емпіричного ризику.</p> <p>Тема 3.5. Асимптотичний приклад мінімізації емпіричного ризику.</p> <p>Тема 3.6. Емпіричні оцінки спроможності методу аналізу: Hold-out волідація.</p> <p>Тема 3.7. Cross-волідація. Leave-one-out cross-волідація.</p>	9	1	0	8
<p>Розділ 4. Метричні методи класифікації</p> <p>Тема 4.1. Гіпотеза компактності. Формалізація поняття відстані.</p> <p>Тема 4.2. Постановка задачі для метричних методів класифікації.</p> <p>Тема 4.3. Принципи перенумерації вибірки.</p> <p>Тема 4.4. Метричний алгоритм класифікації.</p> <p>Тема 4.5. Метод найближчих сусідів.</p>	4	1	0	3
<p>Тема 4.6. Метод k найближчих сусідів (k nearest neighbors, kNN).</p> <p>Тема 4.7. Проблема вибору оптимального k. Алгоритм k зважених найближчих сусідів.</p>				
<p>Розділ 5. Лінійні методи класифікації</p> <p>Тема 5.1. Постановка задачі для лінійних методів класифікації.</p> <p>Тема 5.2. Класична модель нейрона МакКаллока і Піттса та опис її роботи.</p> <p>Тема 5.3. Види типових функцій активації нейрона.</p> <p>Тема 5.4. Апроксимація та регуляризація емпіричного ризику у перцептронах.</p> <p>Тема 5.5. Лінійна модель класифікації. Метод градієнтного спуску.</p> <p>Тема 5.6. Метод стохастичного середнього градієнта. Правило Хебба.</p> <p>Тема 5.7. Перцептрон Розенблатта. Теорема Новикова про кількість кроків для розділення вибірки.</p> <p>Тема 5.8. Евристики для поліпшення градієнтних методів навчання. Квадратична регуляризація.</p>	6	1	0	5

<p>Розділ 6. Метод опорних векторів</p> <p>Тема 6.1. Основні властивості методу опорних векторів.</p> <p>Тема 6.2. Вибірка з лінійною роздільністю: оптимальна розділяюча гіперплощина, нормування, ширина роздільної смуги.</p> <p>Тема 6.3. Лінійно не роздільна вибірка: регуляризація емпіричного ризику, двоїста задача.</p> <p>Тема 6.4. Ядра і спрямляючі простори: конструктивні способи побудови ядер.</p> <p>Тема 6.5. Двошарова нейронна мережа для реалізації методу опорних векторів.</p> <p>Тема 6.6. Переваги та недоліки методу опорних векторів.</p>	9	1	2	6
<p>Розділ 7. Байєсівська теорія класифікації. Основні положення. Імовірнісна постановка задачі. Непараметрична класифікація</p> <p>Тема 7.1. Основні положення теорії імовірностей.</p> <p>Тема 7.2. Функція розподілу імовірностей дискретної випадкової величини та її властивості.</p> <p>Тема 7.3. Неперервні випадкові величини, умовна імовірність.</p> <p>Тема 7.4. Повна імовірність, формула Байєса.</p>	10	1	0	9
<p>Тема 7.5. Імовірнісна постановка задачі класифікації.</p> <p>Тема 7.6. Функціонал середнього ризику. Аргументи максимізації та мінімізації.</p> <p>Тема 7.7. Оптимальний класифікатор Байєса. «Наївний» байєсівський класифікатор.</p> <p>Тема 7.8. Непараметрична класифікація: одновимірний неперервний випадок, багатовимірний неперервний випадок.</p> <p>Тема 7.9. Метод парзенівського вікна.</p>				
<p>Розділ 8. Байєсівська теорія класифікації. Нормальний дискримінантний аналіз. Розділення суміші розподілів.</p> <p>Тема 8.1. Багатовимірний нормальний розподіл.</p> <p>Тема 8.2. Квадратичний дискримінант.</p> <p>Тема 8.3. Лінійний дискримінант Фішера.</p> <p>Тема 8.4. EM-алгоритм</p> <p>Тема 8.5. Суміші багатовимірних нормальних розподілів.</p>	10	1	0	9

<p>Розділ 9. Логічні алгоритми класифікації</p> <p>Тема 9.1. Постановка задачі для логічних методів класифікації.</p> <p>Тема 9.2. Статистичне визначення інформативності.</p> <p>Тема 9.3. Ентропійне визначення інформативності.</p> <p>Тема 9.4. Багатокласова інформативність. Зважена інформативність.</p> <p>Тема 9.5. Методи пошуку інформативних закономірностей: бінаризація кількісних ознак.</p> <p>Тема 9.6. «Жадібний» алгоритм злиття зон, пошук закономірностей у формі кон'юнкцій, «градієнтний» алгоритм синтезу кон'юнкції.</p> <p>Тема 9.7. «Жадібний» алгоритм синтезу кон'юнкції, стохастичний локальний пошук.</p> <p>Тема 9.8. Емерджентний алгоритм синтезу кон'юнкцій, форми закономірностей.</p>	10	1	0	9
<p>Розділ 10. Списки та дерева ухвалення рішень</p> <p>Тема 10.1. «Жадібний» алгоритм побудови списку ухвалення рішень, приклади списків ухвалення рішень.</p> <p>Тема 10.2. Дерева ухвалення рішень: синтез дерев ухвалення рішень, алгоритм побудови дерева ухвалення рішень ID3.</p> <p>Тема 10.3. Обробка пропусків, оцінювання ймовірностей.</p>	6	1	0	5
<p>Тема 10.4. Трудомісткість алгоритму ID3, переваги і недоліки алгоритму ID3.</p> <p>Тема 10.5. Редукція дерев ухвалення рішень.</p> <p>Тема 10.6. Перетворення дерева рішень у список рішень. Заглядання вперед.</p>				
<p>Розділ 11. Зважене голосування правил</p> <p>Тема 11.1. Принципи голосування: алгоритм простого голосування.</p> <p>Тема 11.2. Алгоритм зваженого голосування, налаштування ваг.</p> <p>Тема 11.3. Диверсифікація правил, відмови від класифікації.</p> <p>Тема 11.4. Алгоритм КОРА: принципи побудови алгоритму КОРА, його переваги та недоліки.</p> <p>Тема 11.5. Алгоритм ТЕМП: принципи роботи алгоритму, переваги та недоліки алгоритму ТЕМП.</p> <p>Тема 11.6. Алгоритм бустинга: експонентна апроксимація граничної функції втрат.</p> <p>Тема 11.6. Принципи роботи алгоритму бустинга, переваги та недоліки алгоритму бустинга.</p>	16	2	0	14

<p>Розділ 12. Штучні нейронні мережі</p> <p>Тема 12.1. Проблема повноти та задача «XOR». Обчислювальні можливості нейронних мереж.</p> <p>Тема 12.2. Багат шарові нейронні мережі та метод зворотного поширення помилки.</p> <p>Тема 12.3. Переваги та недоліки методу зворотного поширення помилки.</p>	8	1	2	5
<p>Розділ 13. Кластеризація. Алгоритми кластеризації</p> <p>Тема 13.1. Цілі кластеризації. Типи кластерних структур.</p> <p>Тема 13.2. Евристичні графові алгоритми: алгоритм виділення зв'язних компонентів.</p> <p>Тема 13.3. Алгоритм найкоротшого незамкнутого шляху, алгоритм FOREL.</p> <p>Тема 13.4. Функціонали якості кластеризації.</p> <p>Тема 13.5. Статистичні алгоритми: гіпотеза про простір об'єктів і форму кластерів.</p> <p>Тема 13.6. Метод k-середніх, кластеризація з частковим навчанням.</p> <p>Тема 13.7. Ієрархічна кластеризація: властивість монотонності, теорема Міллігана.</p> <p>Тема 13.8. Властивості розтягування і стискання, властивість редукованості, теорема Діде і Моро.</p> <p>Тема 13.9. Визначення числа кластерів, переваги і недоліки кластеризації.</p>	12	1	4	7
<p>Розділ 14. Мережі Кохонена</p> <p>Тема 14.1. Моделі конкурентного навчання: правило жорсткої конкуренції WTA.</p> <p>Тема 14.2. Правило справедливої конкуренції CWTA, правило м'якої конкуренції WTM.</p> <p>Тема 14.3. Карти Кохонена, що самоорганізуються, мистецтво інтерпретації карт Кохонена.</p> <p>Тема 14.4. Недоліки карт Кохонена.</p> <p>Тема 14.5. Гібридні мережі зустрічного поширення: кусочно-постійна апроксимація, гладка апроксимація.</p> <p>Тема 14.6. Багатовимірне шкакування: розміщення одного об'єкта методом Ньютона-Рафсона.</p> <p>Тема 14.7. Субквадратичний алгоритм багатовимірного шкакування, карта подібності, діаграма Шепарда.</p>	14	1	4	9

Розділ 15. Методи відновлення регресії (оцінки регресії) Тема 15.1. Метод найменших квадратів. Тема 15.2. Непараметрична регресія: ядерне згладжування, формула Надарая-Ватсона. Тема 15.3. Вибір ядра й ширини вікна, проблема викидів (робастна непараметрична регресія) Тема 15.4. Алгоритм LOWESS (локально зважене згладжування), проблема крайових ефектів. Тема 15.5. Лінійна регресія: сингулярне розкладання, проблема мультиколінеарності, гребенева регресія. Тема 15.6. Нелінійні методи відновлення регресії: нелінійна модель регресії, нелінійні одновимірні перетворення ознак.	16	1	6	9
Залік	6	2	0	4
Всього в семестрі:	150	18	18	114

Форма виконання лабораторних робіт

Метою проведення циклу лабораторних робіт є набуття студентами необхідних практичних навичок використання методів та способів представлення та опрацювання даних для отримання результатів, що відповідають меті лабораторної роботи.

Звіт до кожної лабораторної роботи складається з такого комплексу матеріалів:

1. Протокол у форматі *.docx або *.pdf файлу, який містить такі обов'язкові пункти:

Тема лабораторної роботи

Мета лабораторної роботи

Теоретичні відомості

Опис процесу виконання роботи

Отримані результати

Висновок

2. Код у форматі *.i7upb файлу.

3. Файл використаного датасету або посилання на використаний датасет.

Теми занять з лабораторних робіт

1. **Вступне заняття.** Вирішення організаційних питань. Оволодіння програмним забезпеченням, необхідним для виконання лабораторних робіт

2. **Лабораторна робота 1.** Попередня обробка даних.

3. **Лабораторна робота 2.** Кластеризація за методом k-середніх.

4. **Лабораторна робота 3.** Кластеризація (DBSCAN, OPTICS).

5. **Лабораторна робота 4.** Класифікація (Байєсівські методи, дерева).

6. **Лабораторна робота 5.** Класифікація (лінійний дискримінантний аналіз, метод опорних векторів)

7. **Лабораторна робота 6.** Бібліотека TensorFlow для навчання глибоких нейронних мереж.

8. **Лабораторна робота 7.** Основи застосування повнозв'язних мереж для задач регресії.

9. **Завершальне заняття.** Підведення підсумків та прийом заборгованостей.

6. Самостійна робота здобувача вищої освіти денної форми навчання

Види самостійної роботи (114 годин):

1. Підготовка до аудиторних занять, у відповідності з таблицею 1.
2. Підготовка та оброблення проведення розрахунків за первинними даними, отриманими на лабораторних заняттях, виконання лабораторних робіт (рекомендовано 2 години x 7 лабораторних робіт = 14 годин).
3. Підготовка до заліку (4 години).

7. Методика викладання дисципліни на заочній формі навчання

Викладання дисципліни для заочної форми навчання відбувається у обсязі, що співпадає з викладанням для денної форми навчання. Лекції начитуються у оглядовій формі. Обсяг начитаних лекцій регулюється поточним розкладом для даної дисципліни.

Лабораторні роботи для заочної форми навчання ідентичні тим, що видаються студентам заочної форми навчання. Система оцінювання результатів навчання співпадає з системою оцінювання для денної форми навчання

Політика та контроль

8. Політика навчальної дисципліни (освітнього компонента)

Для виконання лабораторних робіт встановлюються дедлайни.

Виконання лабораторних робіт поза встановлених термінів супроводжуються штрафними балами, які вираховуються із оцінки за протокол.

Штрафні бали виставляються за: невчасну здачу лабораторної роботи. Кількість штрафних балів не більше 10. Штрафні бали та жорсткі дедлайни не запроваджуються у період військового стану.

Заохочувальні бали виставляються тільки у випадку додаткового оголошення викладача на протязі семестру за активну участь на лекціях; виконання поточних домашніх завдань, виконання бонусної лабораторної роботи, отримання сертифікату Coursera за темою, яка попередньо узгоджена з викладачем тощо. Кількість заохочуваних балів не більше 10.

Виконанню кожної лабораторної роботи передуює виконання індивідуального завдання і оформлення його у вигляді комплекту матеріалів. Комплект матеріалів до лабораторної роботи, яку студент планує захистити індивідуально на занятті, має бути попередньо завантажений у відповідно сформоване для нього завдання у рамках дистанційного курсу.

Студент, який прийшов на заняття без попередньо завантаженого комплекту матеріалів до захисту лабораторної роботи не допускається. Першим етапом студент захищає результати отримані під час виконання індивідуального завдання до лабораторної роботи шляхом демонстрації роботи коду та відповідей на поточні запитання викладача. На другому етапі – захищає теорію шляхом усного опитування або тестування.

Бали отримані за захист всіх етапів лабораторної роботи входять в оцінку за лабораторну роботу. Студент, який не захистив практичну частину завдання, до захисту теоретичної складової не допускається.

Оцінка, яку студент може отримати за виконання кожної лабораторної наведені в таблиці 3.

Таким чином мінімальна оцінка, яку повинен отримати студент дорівнює 60 балів, максимальна – 100 балів за виконання лабораторних робіт за семестр.

Під час занять з навчальної дисципліни «Прикладні методи аналізу даних» аспіранти повинні дотримуватись певних дисциплінарних правил:

- забороняється запізнюватись на заняття;
- не допускаються сторонні розмови або інший шум, що заважає проведенню занять.

Лабораторні роботи здаються особисто з попередньою перевіркою теоретичних знань, які необхідні для виконання лабораторної роботи. Перевірка практичних результатів включає перевірку коду та виконання тестових завдань.

При проведенні контрольних заходів та при виконанні лабораторних робіт студенти повинні дотримуватись правил академічної доброчесності. При виявленні значного відсотку списування або плагіату викладач може відмовити у прийнятті даної роботи та вимагати доброчесного виконання навчального плану.

9. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Види контролю з навчальної дисципліни «Прикладні методи аналізу даних» включають:

1. Захист лабораторних робіт.
2. Залік.

Таблиця 2

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

Таблиця 3

Оцінювання видів робіт

№ п/п	Вид роботи	Кількість балів
1	Лабораторна робота №1	14
2	Лабораторна робота №2	14
3	Лабораторна робота №3	14
4	Лабораторна робота №4	14
5	Лабораторна робота №5	14
6	Лабораторна робота №6	14
7	Лабораторна робота №7	16

Умови допуску до заліку:

1. Виконані всі лабораторні роботи.
2. Загальний рейтинг за семестр більший за 35 балів

Здобувачі, які виконали всі умови допуску до заліку допускаються до складання заліку на останньому занятті.

Здобувачі, які мають семестровий рейтинг 60 і більше балів можуть отримати залік автоматично без додаткових випробувань.

Здобувачі, які не набрали 60 балів та здобувачі та які бажають підвищити свій семестровий рейтинг, виконують залікову контрольну роботу. Ця оцінка є остаточною. У разі написання залікової контрольної роботи на оцінку, нижчу за отриману «автоматом», попередній рейтинг з дисципліни скасовується і до залікової відомості заносяться бали за залікову контрольну роботу («жорстка» РСО).

Залікова контрольна робота оцінюється у 100 балів максимум і проходить у вигляді тестування у системі ТСЕХАМ.

Критерій оцінювання визначається відсотком правильних відповідей у відповідності з таблицею 2

10. Додаткова інформація з дисципліни (освітнього компонента)

Викладання дисципліни «Прикладні методи аналізу даних» має свою специфіку, яка пов'язана з тим, що сфера застосування методів машинного аналізу даних постійно розширюється. Повсюдна інформатизація приводить до накопичення величезних обсягів даних у наукових дослідженнях, виробництві, на транспорті, в охороні здоров'я. Задачі прогнозування, управління та прийняття рішень часто потребують машинного аналізу даних та машинного навчання, оскільки раніше такі задачі або не ставились взагалі, або вирішувались точними методами, які вимагали великих обчислювальних потужностей через складність відповідних алгоритмів.

Складено: д.т.н, професор, Новотарський Михайло Анатолійович

Ухвалено кафедрою ОТ (протокол № 15 від 29.05.24)

Погоджено Методичною комісією ФІОТ (протокол № 10 від 21.06.24)