



ПРИКЛАДНІ МЕТОДИ АНАЛІЗУ ДАНИХ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти *Третій (доктор філософії)*

Галузь знань	12 Інформаційні технології
Спеціальність	121 Інженерія програмного забезпечення 123 Комп'ютерна інженерія
Освітня програма	Комп'ютерна інженерія
Статус дисципліни	Вибіркова
Форма навчання	очна(денна)
Рік підготовки, семестр	2 курс, осінній семестр
Обсяг дисципліни	4 кредити
Семестровий контроль/ контрольні заходи	Залік
Розклад занять	Лекцій 18 (36 годин) Лабораторних 9 (18 годин)
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com Лабораторні: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com
Розміщення курсу	https://cutt.ly/LECj7ZQ

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.

Вивчення даної дисципліни дає можливість набутти важливих компетенцій, знань та умінь, оскільки актуальність застосування методів аналізу даних у різних сферах науки і техніки істотно зросла. Також значно розширився діапазон застосувань цих методів при вирішенні різноманітних прикладних задач..

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 навчальної дисципліни мають продемонструвати такі **компетентності**.

1. *Загальні компетентності*: здатність до абстрактного мислення, аналізу і синтезу; здатність до пошуку, оброблення та аналізу інформації з різних джерел щодо методів та технологій добування даних; здатність формування системного наукового світогляду у сфері аналізу даних; здатність набуття універсальних навичок усної та письмової презентації власного наукового дослідження за результатами аналізу даних, застосування сучасних інформаційних технологій у науковій діяльності.

2. *Спеціальні компетентності*: здатність ефективно застосовувати основні методи аналізу даних, основні статистичні методи навчання, кластерні методи машинного навчання та методи навчання без учителя при проведенні наукових досліджень; здатність інтегрувати знання з різних дисциплін, застосовувати системний підхід при проведенні досліджень; здатність аргументувати вибір методу розв'язання наукової задачі, критично оцінювати отримані результати та захищати прийняті рішення щодо вибору методів та способів аналізу даних.

За результатами вивчення навчальної дисципліни «Прикладні методи аналізу даних» мають бути отримані такі **знання**.

1. Мати передові концептуальні та методологічні знання у сфері аналізу даних для прийняття оптимальних рішень, а також дослідницькі навички, достатні для проведення наукових і прикладних досліджень в галузі аналізу даних на рівні останніх світових досягнень з комп'ютерної інженерії, IT-інфраструктур, інформаційних технологій.

2. Знати сучасні методи проведення досліджень у сфері прикладного аналізу даних для прийняття рішень в різних галузях науки шляхом вивчення теоретичних та практичних положень побудови поведінкових моделей функціональних залежностей та використання результатів аналізу даних для уточнення оптимальних параметрів прийняття рішень.

3. Знати і розуміти наукові і математичні положення, що лежать в основі методів аналізу даних для отримання оптимальних параметрів прийняття рішень, методів побудови та дослідження математичних моделей та технологій адаптивних та інтелектуальних обчислень при аналізі даних.

Уміння, які мають бути отримані у рамках вивчення навчальної дисципліни «Прикладні методи аналізу даних».

1. Вміти ефективно здійснювати пошук та критичний аналіз інформації з різних джерел щодо методів та технологій аналізу даних.

2. Вміти розв'язувати задачі синтезу та аналізу об'єктів дослідження при аналізі даних для прийняття оптимальних рішень.

3. Вміти створювати та реалізовувати поведінкові математичні моделі опису функціональних залежностей при аналізі даних;

4. Вміти застосовувати прикладні методи аналізу даних технології адаптивних та інтелектуальних обчислень для передбачення майбутніх станів досліджуваних об'єктів та процесів.

Здобувачі наукового ступеня також мають бути **здатні**.

1. Застосовувати прикладні бібліотеки та програмні системи, які використовуються при машинному аналізі даних.

2. Володіти методами та технологіями програмування з використанням прикладних бібліотек та програмних систем, призначених для машинного аналізу великих наборів даних.

Таке поєднання загальних та спеціальних компетентностей, теоретичних та практичних знань, умінь та здатностей сприяє підвищенню науково-практичного рівня здобувачів наукового ступеня доктора філософії задля здійснення ними ефективних наукових досліджень.

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 з дисциплін: “Програмування”, “Об’єктно-орієнтоване програмування”, “Системне програмування”, “Структури даних та алгоритми”, “Інженерія програмного забезпечення”, “Алгоритми та методи обчислень”, “Дискретна математика”.

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. Багатошарові 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 редуктивності, 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

Базова:

1. Новотарський М.А. Прикладні методи аналізу даних, навч.пос. для спец. 123- комп'ютерна інженерія // <https://cloud.comsys.kpi.ua/s/YpEaokWx3HjcMmZ>
2. Новотарський М.А. Методичні вказівки до практичних занять для спец. 123- комп'ютерна інженерія // <https://cloud.comsys.kpi.ua/s/YpEaokWx3HjcMmZ>
3. Новотарський М.А. Методичні матеріали до лабораторних робіт для спец. 123- комп'ютерна інженерія // <https://cloud.comsys.kpi.ua/s/YpEaokWx3HjcMmZ>
4. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989.
5. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. М.: Наука, 1974.
6. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979.
7. Воронцов К. В. Математические методы обучения по прецедентам. Курс лекций. Москва, ВЦ РАН, 2005. <http://www.cas.ru/voron/teaching.html>
8. Гладков Л. А., Курейчик В. В., Курейчик В. М. Генетические алгоритмы. – М.: Физматлит, 206.– 320 с.
9. Головкин В. А. Нейронные сети: обучение, организация и применение. – М.: ИПР-ЖР, 2001.
10. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. – Новосибирск: ИМ СО РАН, 1999.
11. Закс Ш. Теория статистических выводов.– М.: Мир, 1975.
12. Кулаичев А. П. Методы и средства комплексного анализа данных. – М: ИНФРА-М, 2006.

Додаткова:

11. Лбов Г. С. Методы обработки разнотипных экспериментальных данных. – Новосибирск: Наука, 1981.
12. Мандель И. Д. Кластерный анализ. – М.: Финансы и Статистика, 1988.
13. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. – Киев: Наукова думка, 2004.

14. Хардле В. Прикладная непараметрическая регрессия. – М.: Мир, 1993.
15. Asuncion A., Newman D. UCI machine learning repository: Tech. rep.: University of California, Irvine, School of Information and Computer Sciences, 2007. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
16. Bishop C. M. Pattern Recognition and Machine Learning. – Springer, Series: Information Science and Statistics, 2006. – 740 pp.
17. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning.– Springer, 2001. – 533 pp. <http://http://www-stat.stanford.edu/tibs/ElemStatLearn>.
18. Jordan M. I., Xu L. Convergence results for the EM algorithm to mixtures of experts architectures: Tech. Rep. A.I. Memo No. 1458: MIT, Cambridge, MA, 1993.
19. LeCun Y., Bottou L., Orr G. B., Muller K.-R. Efficient BackProp // Neural Networks: tricks of the trade. – Springer, 1998.
20. Smola A., Schoelkopf B. A tutorial on support vector regression: Tech. Rep. NeuroCOLT2 NC2-TR-1998-030: Royal Holloway College, London, UK, 1998. <http://citeseer.ist.psu.edu/smola98tutorial.html>.
21. Новотарський М. А., Нестеренко Б. Б. Штучні нейронні мережі: обчислення. К: Ін-т математики НАН України, 2004. – 408 с.
22. Завантаження та інсталяція Weka: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>
23. Pentaho Data Mining Community Documentation: <http://wiki.pentaho.com/display/DATAMINING/Pentaho+Data+Mining+Community+Documentation>

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
<p>Розділ 1. Вступ</p> <p>Тема 1.1. Загальне поняття про аналіз даних та сфери застосування задач прикладного аналізу.</p> <p>Тема 1.2. Порівняння методів машинного навчання та аналізу даних</p> <p>Тема 1.3. Класифікація задач машинного навчання. Дедуктивний та індуктивний способи навчання.</p> <p>Тема 1.4. Навчання з учителем та навчання без учителя. Навчання з підкріпленням.</p> <p>Тема 1.5. Активне та пасивне навчання. Зв'язок з іншими областями науки.</p>	6	4	0	2
<p>Розділ 2. Загальне поняття про прикладні задачі аналізу даних</p> <p>Тема 2.1. Задачі класифікації. Задача оцінювання позичальників.</p> <p>Тема 2.2. Задачі регресійної оцінки. Задача прогнозування споживчого попиту.</p> <p>Тема 2.3. Задачі кластеризації. Задача рубрикації текстів.</p>	6	4	0	2

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
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.	6	4	0	2
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.	4	2	0	2
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.	4	2	0	2

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
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.	6	2	2	2
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.	6	2	0	4
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.	6	2	0	4
Розділ 9. Логічні алгоритми класифікації	6	2	0	4

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
<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>				
<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> <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>	6	2	0	4
<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>	6	4	0	2

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Розділ 12. Штучні нейронні мережі Тема 12.1. Проблема повноти та задача «XOR». Обчислювальні можливості нейронних мереж. Тема 12.2. Багатошарові нейронні мережі та метод зворотного поширення помилки. Тема 12.3. Переваги та недоліки методу зворотного поширення помилки.	8	2	2	4
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.	12	4	4	4
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.	14	0	4	10

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
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	0	6	10
Залік	8	2	0	6
Всього в семестрі:	120	38	18	64

6. Самостійна робота аспіранта

Метою проведення циклу лабораторних робіт є набуття студентами необхідних практичних навичок використання методів та способів представлення та опрацювання даних для отримання прихованої інформації, методів дослідження математичних моделей опису шуканої інформації, технології добування даних для розроблення прогнозів щодо досліджуваних об'єктів та процесів.

№ з/п	Назва лабораторної роботи	Кількість ауд. годин
1	<i>Лабораторна робота № 1.</i> Створення зошита в середовищі Jupyter notebook.	2
2	<i>Лабораторна робота № 2.</i> Практичне застосування атрибутів та функцій модуля NumPy.	2
3	<i>Лабораторна робота № 3.</i> Вивчення та практичне застосування бібліотеки Matplotlib.	2
4	<i>Лабораторна робота № 4.</i> Вивчення та практичне застосування бібліотеки Pandas.	2
5	<i>Лабораторна робота № 5.</i> Первинна обробка набору даних "Yelp Dataset JSON" для тематичного дослідження.	2
6	<i>Лабораторна робота № 6.</i> Фільтрування великого набору даних фрагментами	2
7	<i>Лабораторна робота № 7.</i> Збереження структури DataFrame у форматі csv	2

8	Лабораторна робота № 8. Практичний аналіз текстових представлень великих обсягів даних	2
9	Лабораторна робота № 9. Візуалізація одержаних даних та підготовка презентації за результатами досліджень	2

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

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

- забороняється запізнюватись на заняття;
- при вході викладача, на знак привітання, особи, які навчаються в КПІ ім. Ігоря Сікорського повинні встати;
- не допускаються сторонні розмови або інший шум, що заважає проведенню занять;
- виходити з аудиторії під час заняття допускається лише з дозволу викладача.
- не допускається користування мобільними телефонами та іншими технічними засобами без дозволу викладача.

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.

В процесі навчання викладач має право нарахувати до 5 заохочувальних балів за дострокове виконання лабораторної роботи, за проявлений творчий підхід при виконанні індивідуального завдання або за активну участь у обговоренні питань, що пов'язані з тематикою лекції або практичного заняття.

За виконання та здачу лабораторної роботи після зазначеного дедлайну, за значну кількість пропущених занять, або за порушення правил поведінки на заняттях викладач може призначити до 5 штрафних балів.

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

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

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

Поточний контроль: тестування закритими тестами.

Календарний контроль: провадиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Семестровий контроль: залік

Умови допуску до семестрового контролю: семестровий рейтинг більше 40 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	Оцінка
-----------------	--------

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

9. Додаткова інформація з дисципліни (освітнього компонента)

Викладання дисципліни «Прикладні методи аналізу даних» для напрямку «Комп'ютерна інженерія» 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.

Умова зарахування додаткових балів.

В рамках вивчення навчальної дисципліни «Прикладні методи аналізу даних» допускається зарахування балів, одержаних в результаті дистанційних курсів на платформі "Coursera", за умови попереднього погодження програми даного курсу з викладачем та за умови отримання офіційного сертифікату.

Робочу програму навчальної дисципліни (силабус):

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

Ухвалено кафедрою ОТ (протокол № 18 від 25.05.21)

Погоджено Методичною комісією ФІОТ (протокол № 10 від 14.06.21)