



Інтелектуальний аналіз даних

Шифр та назва спеціальності

122 – Комп'ютерні науки

Інститут

ННІ комп'ютерних наук та інформаційних технологій

Освітня програма

Управління проектами у сфері інформаційних технологій

Кафедра

Управління проектами в інформаційних технологіях

Рівень освіти

Магістр

Тип дисципліни

Спеціальна (фахова), Обов'язкова

Семестр

1

Мова викладання

Українська,

Викладачі, розробники



Харченко Алла Олександрівна

Alla.Kharchenko@khpi.edu.ua

К.е.н., доцент, доцент кафедри стратегічного управління НТУ «ХПІ»

Авторка понад 70 наукових і навчально-методичних публікацій. 2 статті у виданнях, індексованих у Scopus. (h-index = 3, Google Scholar:

<https://scholar.google.com/citations?hl=en&user=VIRAtqIAAAAJ>;

ORCID: <https://orcid.org/0000-0002-8536-2857>;

Scopus:

<https://www2.scopus.com/authid/detail.uri?authorId=57220059467>).

Курси: «Основи економіки та економетрики», «Дослідження операцій», «Інтелектуальний аналіз даних», «Основи бізнес планування», «Основи бізнес аналізу».

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна спрямована на формування у студентів системи знань, умінь і навичок, застосування механізмів інтелектуального аналізу для вирішення прикладних задач обробки великих даних, вирішення складних теоретичних і практичних задач, які відповідають завданням сучасної обробки даних – інтелектуального аналізу даних (Data Mining), виявлення закономірностей, необхідних для прийняття рішень.

Мета та цілі дисципліни

Метою навчальної дисципліни є формування системи наукових знань і практичних умінь та навичок щодо використання інструментарію аналізу даних та способів вирішення типових задач аналізу даних за допомогою технологій машинного навчання.

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

СК1. Усвідомлення теоретичних засад комп'ютерних наук.

СК2. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.

СК3. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК4. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень.

СК. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

Результати навчання

П РН1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.

РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

РН 8. Розробляти математичні моделі та методи аналізу даних (включно з великими).

РН 9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

РН 11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

Обсяг дисципліни

Загальний обсяг дисципліни 90 год. (3 кредитів ECTS): лекції – 32 год., лабораторні роботи – 16 год., самостійна робота – 42 год.

Передумови вивчення дисципліни (пререквізити)

Для успішного проходження курсу необхідно мати знання та практичні навички з наступних дисциплін: Алгоритмізація і програмування; Дискретна математика; Теорія ймовірності, ймовірнісні процеси та математична статистика; Алгоритми та структури даних; Організація баз даних та знань

Особливості дисципліни, методи та технології навчання

Інтерактивні лекції з презентаціями, лабораторні заняття, командна робота, метод зворотного зв'язку з боку студентів, проблемне навчання

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Особливості технології Data Mining та її відмінності від інших методів аналізу даних

Підтеми / перелік питань

Тема 2. Предпроцесний аналіз даних

Поняття та типи даних. Метадані. Підготовка даних. Огляд даних. Попередній аналіз даних. Візуалізація даних.

Тема 3. Класифікації даних

Постановка задач класифікації даних. Види класифікації. Процес класифікації. Точність класифікації: оцінка рівня помилок. Методи, що застосовуються для розв'язання задач класифікації. Класифікація методом опорних векторів. Методи та моделі дискримінантного аналізу. Класифікація за допомогою методу найближчого сусіда. Байєсівська (наївна)

класифікація. Класифікація за допомогою дерев рішень. Класифікація за допомогою штучних нейронних мереж.

Тема 4. Штучна нейронна мережа

Штучна нейронна мережа принцип роботи, основні складові. Схема штучного нейрону. Типи та архітектура нейронної мережі. Підходи до навчання нейронної мережі. Прості нейронні мережі (Feedforward Neural Networks). Згорткові нейронні мережі (Convolutional Neural Networks, CNN). Рекурентні нейронні мережі (Recurrent Neural Networks, RNN). Довгі короткострокові пам'яті (Long Short-Term Memory, LSTM). Генеративно-змагальні мережі (Generative Adversarial Networks, GAN). Автокодері (Autoencoders). Трансформери (Transformers). Нейронні мережі з підкріпленням (Reinforcement Learning Neural Networks).

Тема 5. Кластерний аналіз даних

Задачі кластерного аналізу. Основні етапи кластерного аналізу. Типи кластерних структур. Міри відстані. Методи кластерного аналізу. Ієрархічна кластеризація. Види ітеративних (ітераційних) методів кластеризації. Метод к-середніх. Нечіткі методи класифікації. Критерії якості класифікації методами кластерного аналізу.

Тема 6. Зниження розмірності

Основні поняття і завдання зниження розмірності. Підходи до відбору інформативних ознак. Методи зниження (редукції) ознак. Поняття системи діагностичних ознак. Метод «центра ваги». Методи факторного аналізу. Компонентний аналіз.

Тема 7. Генетичний алгоритм

Генетичний кроссовер. Основні поняття генетичних алгоритмів. Функція пристосованості (фітнес-функція). Класичний генетичний алгоритм. Створення початкової популяції. Розмноження. Селекція та формування нового покоління.

Тема 8. Аналіз часовий рядів

Складові часового ряду. Сезонні ARMA та ARIMA-моделі. Фрактальні та вейвлет методи роботи з часовими рядами.

Тема 9. Асоціативні правила

Задача пошуку асоціативних правил. Послідовне відображення шаблонів даних. Методи побудови класифікаційних правил. Алгоритм Apriori та його різновиди. 1R алгоритм. Метод Naive Bayes.

Тема 10. Методи прийняття рішень на основі нечітких множин та нечіткої логіки

Основи теорії нечітких множин. Методи нечіткої логіки. Нечіткі експертні системи. Етапи побудови нечіткого логічного виводу. Нечіткі алгоритми Мамдані, Цукамото, Такагі-Сугено та ін. Нейро-нечіткі технології.

Тема 11. Адаптивні методи кластеризації

Загальна схема процедури адаптивної кластеризації. Визначення якості кластеризації. Показники чіткості. Нечіткі алгоритми кластеризації

Тема 12. Способи візуального аналізу даних

Візуалізація інструментів Data Mining. Візуалізація Data Mining моделей. Методи візуалізації. Представлення даних в одному, двох і трьох вимірах

Тема 13. Сховища даних

Концепція сховища даних. Організація СД. Структура систем підтримки прийняття рішень з фізичним СД. Структура систем підтримки прийняття рішень з віртуальним СД. Вітрини даних. ETL процес. Очистка даних. Сховища даних та аналіз.

Тема 14. OLAP системи

Багатовимірна модель даних. Визначення OLAP систем. Особливості OLAP систем. Тест FASMI. Архітектура OLAP систем (MOLAP, ROLAP, HOLAP)

Теми практичних занять

Не передбачено навчальним планом дисципліни

Теми лабораторних робіт

Тема 1. Первинна обробка статистичних даних. Описова статистика

Тема 2. Використання дискримінантної функції в задачах класифікації даних

Тема 3. Розв'язання задачі класифікації шляхом побудови класифікаційних правил

Тема 4. Кластеризація даних методом k-means

Тема 5. Редукція даних, як важливий етап їх підготовки

Тема 6. Генерація асоціативних правил

Тема 7. Аналіз часових рядів

Тема 8. Оптимізація функції за допомогою генетичних алгоритмів

Самостійна робота

Вивчення тем курсу за допомогою рекомендованої літератури. Підготовка звітів з виконання лабораторних робіт. Підготовка до захисту лабораторних робіт. Підготовка до перевірки знань у вигляді тестування. Індивідуальне розрахункове завдання.

Література та навчальні матеріали

1. Ситник В. Ф. Інтелектуальний аналіз даних (дейтамайнінг): Навч. Посібник/ В. Ф. Ситник, М.Т. Краснюк - К: КНЕУ, 2007. - 376 с.
2. Малкіна В. М., Зінов'єва О.Г. Інтелектуальний аналіз даних: Лабораторний практикум. Частина I – Мелітополь: Люкс, 2021. – 150 с.
3. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2018. 152 с.
4. Черняк О.І. Інтелектуальний аналіз даних: Підручник / О.І. Черняк, П.В. Захарченко ; Київ. нац. ун-т ім. Т. Шевченка. — К. : Знання, 2014. — 599 с.
5. Пасічник В.В., Шаховська Н.Б. Сховища даних: Навчальний посібник. – Львів: «Магнолія 2006», 2008. - 496 с.
6. Обчислювальні аспекти аналізу даних на основі карт Кохонена. / Годич О. В та ін. Національний університет "Львівська політехніка", 2011. 6372 с.
7. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. Київ : ІСЗЗІ КПІ ім. Ігоря Сікорського, 2018. 300 с.
8. Фісун М.Т., Кравець І.О., Казмірчук П.П., Ніколенко С.Г. Інтелектуальний аналіз даних: практикум: навчальний посібник. Львів : Новий Світ-2000, 2018. 160 с.
9. McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. O'Reilly Media. September 20, 2022. - 621 p.
10. Foster Provost Data Science for Business: What you need to know about data mining and data-analytic thinking (Paperback). – 2013, 414 p.
11. Torgo L. Data Mining with R: Learning with Case Studies 2nd ed. — Taylor & Francis;Chapman and Hall/CRC, 2017. — 426 p. — (Chapman & Hall/CRC Data Mining and Knowledge Discovery). — ISBN: 9781482234893, 1482234890
12. Mishra P. R Data Mining Blueprints New York: Packt Publishing, 2016. — 254 p.
13. Saiz A.Z., Gonzalez C.Q., Gil L.H., Ruiz D.M. An Introduction to Data Analysis in R: Hands-on Coding, Data Mining, Visualization and Statistics from Scratch Springer, 2020. — 291 p. — ISBN 978-3-030-48997-7 (eBook).
14. Kretowski, M. Evolutionary Decision Trees in Large-Scale Data Mining /M. Kretowski , - Cham: Springer International Publishing, 2019, 180 p.
15. Ranga Suri N.N.R., Murty N., Athithan M.G. Outlier Detection: Techniques and Applications. A Data Mining Perspective / N.N.R. Ranga Suri, N. Murty, M.G. Athithan.,- Cham: Springer International Publishing, 2019, 214 p.

Інформаційні ресурси

1. Stanford University Data Mining Lecture Notes– 2020. – Режим доступу до ресурсу: <http://infolab.stanford.edu/~ullman/mining/2003.html>.
2. University of Florida Introduction to Data Mining [Електронний ресурс]. – 2020. – Режим доступу до ресурсу: <https://www.cise.ufl.edu/class/cis4930fa15idm/notes.html>.
3. Massachusetts Institute of Technology Data Mining » Lecture Notes [Електронний ресурс]. – 2020. – Режим доступу до ресурсу: <https://ocw.mit.edu/courses/sloan-school-ofmanagement/15-062-data-mining-spring-2003/lecture-notes/>.
4. Georgia State university Data Mining Tentative Lecture Notes [Електронний ресурс]. – 2020. – Режим доступу до ресурсу: <https://grid.cs.gsu.edu/~cscyqz/courses/dm/dmlectures.html/>.

5. University of Toronto Machine Learning and Data Mining Lecture Notes [Електронний ресурс]. - 2020. - Режим доступу до ресурсу: <https://www.dgp.toronto.edu/~hertzman/411notes.pdf>.
6. Mining of Massive Datasets [Електронний ресурс]. - 2020. - Режим доступу до ресурсу: <http://www.mmds.org/>.

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді іспиту (30%) та поточного оцінювання (70%).

Дисципліна передбачає 8 лабораторних робіт, За виконання та захист лабораторної роботи студент отримує оцінку згідно 100 шкалі (60 балів мінімальний бал, 100 балів — максимальний бал)

Поточне оцінювання складається з середньо зваженого значення, яке включає оцінку за тест (10%), оцінку за розрахункове завдання (30%) та підсумкову оцінку за всі лабораторні роботи (60%).

Тест складається з 35 запитань, залежно від складності питання, відповідь оцінюються в 2-3 бали. Завдання тесту оцінюються у 100 шкалі (60 балів мінімальний бал, 100 балів — максимальний бал) Посилання на тест: <https://forms.office.com/e/91KUt6Gqsm>

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

27.08.2024



Завідувач кафедри
Марина ГРИНЧЕНКО

27.08.2024



Гарант ОП
Ігор КОНОНЕНКО