Силабус освітнього компонента
Програма навчальної дисципліни

"Інтелектуальний аналіз даних"

Шифр та назва спеціальності
122 – Комп'ютерні науки

Інститут
ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма
Інженерія програмного забезпечення

Кафедра
Програмна інженерія та інтелектуальні технології управління (321)

Рівень освіти
Бакалавр

Тип дисципліни
Спеціальна (фахова), Обов'язкова

Семестр
8

Мова викладання
Українська, англійська

Викладачі, розробники

[Єршова Світлана Іванівна
Svetlana.Ershova@khpi.edu.ua

Ст. викладачка кафедри ПІІТУ.
Кількість наукових та навчальних публікацій – понад 20.
(Google Scholar: https://scholar.google.com.tw/;
ORCID: https://orcid.org/0000-0003-3893-117X)
Scopus (Web of Science): https://ceur-ws.org/Vol-2753/paper25.pdf https://www.scopus.com/record/display.uri?eid=2-s2.0-85134395161&origin=resultslist&sort=plf-f&retries=1, https://ceur-ws.org/Vol-3403/paper37.pdf, https://ceur-ws.org/Vol-3403/paper37.pdf.
Детальніше про викладача на сайті кафедри

Загальна інформація

Анотація

В програмі дисципліни головна увага приділяється розгляду технологій, основних понять, принципів, сучасним досягненням та перспективам розвитку інтелектуального аналізу даних. Розглядаються процеси видобутку даних, попередньої обробки даних, основні поняття теорії асоціативних правил, методи кластерного аналіз; основні питання, пов'язані з побудовою моделі векторного простору і проведенням багатовимірного аналізу даних на основі технологій Data Mining, Text Mining, Web Mining.
Набуття практичних навичок інтелектуального аналізу даних здійснюється на базі Python , WEKA.

Мета та цілі дисципліни

Метою та завданням навчальної дисципліни «Інтелектуальний аналіз даних» є формування у фахівців з комп'ютерних наук комплексу знань, необхідних для розуміння проблем, які виникають під час інтелектуального аналізу даних, освоєнню моделей та методів розв’язування завдань інтелектуального аналізу даних.
Навчальний курс займає провідне місце у формуванні майбутнього спеціаліста з комп'ютерних наук. Отримані у процесі вивчення дисципліни “Інтелектуальний аналіз даних" знання та навички є невід’ємними складовими формування професійної компетентності та важливим аспектом професійної підготовки студентів.

Формат занять

Дисципліна «Інтелектуальний аналіз даних» є навчальною дисципліною з циклу спеціальних (фахової) підготовки за спеціальністю 122 – Комп'ютерні науки.
Лекції, лабораторні роботи, самостійна робота, розрахункове завдання, консультації. Підсумковий контроль – залік.

Компетентності

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.
ЗК2. Здатність застосовувати знання у практичних ситуаціях.
ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.
ЗК6. Здатність вчитися й оволодівати сучасними знаннями.
ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
СК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.
СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.
СК16. Здатність реалізовувати високопродуктивні обчислення на основі хмарних сервісів і технологій, паралельних і розподілених обчислень при розробці й експлуатації розподілених систем паралельної обробки інформації.
СК17. Здатність застосовувати теоретичні та практичні основи сучасної теорії управління складними організаційно-технічними та соціально-економічними системами для побудови інтелектуальних систем управління, у процесі проєктування інтелектуальних систем використовувати сучасні технології обробки інформації та методи обчислювального інтелекту.

Результати навчання

ПР3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.
ПР4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.
ПР12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій Data Mining, Text Mining, Web Mining.
ПР16. Виконувати паралельні та розподілені обчислення, застосовувати чисельні методи та алгоритми для паралельних структур, мови паралельного програмування при розробці та експлуатації паралельного та розподіленого програмного забезпечення.
ПР17. Застосовувати для побудови інтелектуальних систем управління теоретичні та практичні основи сучасної теорії управління, проєктувати інтелектуальні системи з використанням сучасних технологій обробки інформації та методів обчислювального інтелекту.

Обсяг дисципліни

Дисципліна викладається у 8 семестрі в обсязі 120 год. (4 кредити ECTS), зокрема: лекції – 20 год., лабораторні – 20 год., розрахункове завдання, самостійна робота – 80 год.

Передумови вивчення дисципліни (пререквізити)

Курс «Інтелектуальний аналіз даних» базується на дисциплінах «Дискретна математика», «Теорія ймовірності та математична статистика», «Бази даних», «Кібербезпека»

Особливості дисципліни, методи та технології навчання

Методи викладання та навчання:
інтерактивні лекції з презентаціями, дискусії, лабораторні заняття, командна робота, кейс-метод, метод зворотного зв'язку з боку студентів, проблемне навчання.

Форми оцінювання:
письмові індивідуальні завдання до лабораторних робіт (CAS), оцінювання знань на лабораторних заняттях (CAS), експрес-опитування (CAS), онлайн-тести (CAS), підсумковий/семестровий контроль у формі семестрового заліку, відповідно до графіку навчального процесу (FAS).

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. «Основні поняття інтелектуального аналізу даних».
Поняття «інформація, дані, знання».
Сутність інтелектуального аналізу даних. Цілі інтелектуального аналізу даних. Сфери практичного використання інтелектуального аналізу даних .
Тема 2. «Процеси видобутку даних".
Загальні відомості про процес видобутку даних. Процес CRISP-DM. Процес SEMMA. Застосування процесів інтелектуального аналізу даних.
Тема 3. «Попередня обробка даних».
Попередня обробка даних. Обробка неінформативних змінних, відсутніх даних, очищення даних.
Тема 4. «Зменшення кількості атрибутів».
Схеми вибірки для фіксованих наборів даних. Зменшення кількості атрибутів.
Тема 5. «Міркування на основі пам'яті».
Методи міркування на основі пам'яті. Збіг, зважений збіг.
Тема 6. «Асоціації».
Основні поняття теорії асоціативних правил. Правила асоціації у видобутку знань.
Послідовна асоціація. Алгоритми родини «Апріорі».
Тема 7 «Кластеризація».
Сутність кластерізації. Підготовка даних. Алгоритм проведення кластерного аналізу. Статистики, що пов'язані з кластерним аналізом. Тестування моделі. Методи кластерного аналізу.
Тема 8 «Text Mining».
Видобуток тексту. Представлення текстових документів. Побудова моделі векторного простору.
Тема 9 «Web Mining».
Веб -майнінг. Веб -контент, використання Інтернету, видобуток веб -структур.
Тема 10 «Обробка веб -даних».
Інтелектуальні методи обробки веб -даних.

Теми практичних занять

Практичні заняття в рамках дисципліни не передбачені.

Теми лабораторних робіт

Тема 1. Попередня обробка даних за допомогою Weka.
Тема 2. Рішення задачі кластеризації за допомогою Weka.
Тема 3. Обробка даних JSON за допомогою Python.
Тема 4. Обробка XML даних за допомогою Python.
Тема 5. Реалізація веб -сканера за допомогою Python.

Самостійна робота

Студентам рекомендуються додаткові матеріали (відео, статті) для самостійного вивчення та опрацювання.

Література та навчальні матеріали

Основна література

1. Берри Пол Head First Python. Легкий для сприйняття довідник/ Пол Берри,- Видавництво Фабула, 2021, 624 с.
2. Іванов С.М., Максишко Н.К., Бречко Д.О. Інтелектуальний аналіз даних: конспект лекцій/ С.М. Іванов, Н.К Максишко Н.К., Д.О. Бречко,- Запоріжжя: ЗНУ, 2020, 156 с.
3. Kretowski, M. Evolutionary Decision Trees in Large-Scale Data Mining /M. Kretowski ,- Cham: Springer International Publishing, 2019, 180 p.
4. 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.
5. Samanta D., Banerjee A. Computationally Intensive Statistics for Intelligent/ D. Samanta, A.Banerjee ,- Springer, 2021, 218 p.

Додаткова література

1. Akerkar Ranjendra Big Data in Emergency Management: Exploitation Techniques for Social and Mobile Data/ Ranjendra Akerkar,- Springer, 2020, 201 p.
2. Anandan, R. A Closer Look at Big Data Analytics/ R. Anandan ,- Nova Science Publishers. Inc. , 2021, 366 p.
3. Botros Silvia, Tinley Jeremy High Performance MySQL: Proven Strategies for Operating at Scale 4th Edition/ Silvia Botros, Jeremy Tinley,- O'Reilly Media, 2021 , 388 p.
4. Sakarkar Gaurav, Patil Gaurav, Dutta Preteek Machine Learning Algorithms Using Python Programming/ Gaurav Sakarkar, Gaurav Patil, Preteek Dutta ,- Nova Science Publishers. Inc., 2021, 218 p.
5. Srinivas M., Sucharitha G., Matta A., Chatterjee P. /M. Srinivas, G. Sucharitha, A. Matta, P. Chatterjee Machine Learning Algorithms and Applications: Theory and Applications,-Wiley-Scrivener Publishing, 2021, 368 p.

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді заліку (40%) та поточного оцінювання (60%):
- 5 лабораторних робіт (по 5%);
- 1 контрольна робота (5%);
- 1 розрахункове завдання (10%);
- самостійна робота (20%).

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 11.04.2023

Завідувач кафедри
Ігор ГАМАЮН

Гарант ОП
Юлія ЛІТВІНОВА