



Силабус курсу

Інтелектуальний аналіз даних

Ступінь вищої освіти-бакалавр

Освітньо-професійна програма: «Інженерія програмного забезпечення»

Спеціальність – 121 Інженерія програмного забезпечення

Рік навчання: 1, Семестр: 2

Кредитів: 5 Мова викладання: українська

Керівник курсу

к.е.н., доцент Гончар Людмила Іванівна

l.honchar(@)wunu.edu.ua

ІІІ

Контактна інформація

Опис дисципліни

Курс дозволить сформувати у студентів знання про основи інтелектуальної технології Data Mining (Data Science), Big Data, моделей Data Mining, статистичних і кібернетичних методів Data Mining, інтелектуальної технології Text Mining, а також умінь щодо використання доступного програмного забезпечення інтелектуального аналізу даних, методів реалізації генетичних алгоритмів та формування динамічних бізнес-процесів у мережі Інтернет тощо.

Систематизовано теоретичні засади базових відомостей із основ інтелектуального аналізу даних (ІАД), світового досвіду та сучасного стану програмного забезпечення Data Mining (Data Science) в Україні та світі, вмінні застосувати Data Science, Big Data, Data Mining в бізнесі, технологію машинного навчання.

Структура курсу

Години (лек./пр.)	Тема	Результати навчання	Завдання
2/1	Тема 1. Розвиток та призначення DataMining	Ознайомлення з поняттям ІАД. Моделі DataMining. Статистичні та кібернетичні методи DataMining. Інтеграція DataMining і OLAP. Відкриття. Пророче моделювання. Аналіз аномалій.	Тести Лабораторна робота Питання
3/1	Тема 2. Дерево методів дейтамайнінгу	Ознайомлення із поняттям класифікації в ІАД. Кластерізація. Асоціація. Упорядкування. Прогнозування.	Тести Лабораторна робота Питання
2/1	Тема 3. Збережені дані. Дистильовані дані.	Ознайомитись із методом “nearestneighbor”. Міркування за аналогією. Кластерний аналіз. Крос-табуляція. Задача пошуку асоціативних правил.	Тести Лабораторна робота Питання
4/1	Тема 4. Характеристика процесів та активностей дейтамайнінгу	Вивчити характеристики процесів і активностей дейтамайнінгу Умовна логіка. Тренди і варіації. Прогнозування. Виявлення ідевітації.. Процеси дейтамайнінгу. Типи процесів дейтамайнінгу. Онтологічний підхід до використання знань на ІТ-підприємстві.	Тести Лабораторна робота Питання

4/1	Тема 5. . Генетичні алгоритми	Ознайомитись із поняттям генетичних алгоритмів. Генетичні успадкування - концептуальна засада генетичних алгоритмів. алгоритми. Загальна схема генетичних алгоритмів. Генетичні успадкування. Доступне програмне забезпечення генетичних алгоритмів.	Тести Лабораторна робота Питання
3/1	Тема 6. Програмні агенти в СППР	Вивчити призначення і основні характеристики програмних агентів. DSS Agent. Агенто-орієнтовані системи на IT - підприємстві.	Тести Лабораторна робота Питання
2/2	Тема 7. Poly Analyst. Mine Set – візуальний інструмент аналітика. Knowledge STUDIO. Інтелектуальна технологія Text Mining	Ознайомитись із програмним продуктом Poly Analyst. Mine Set – візуальний інструмент аналітика. Knowledge STUDIO. Інтелектуальна технологія Text Mining	Тести Лабораторна робота Питання
2/1	Тема 8. Засоби штучного інтелекту в СППР. Нейронні мережі	Ознайомитись з архітектурою нейромереж. Визначення та еволюція нейронних мереж. Застосування нейронних мереж. Математична модель штучного нейронна.. Навчання та використання нейромереж. Готове програмне забезпечення нейромереж (нейропакети).	Тести Лабораторна робота Питання
2/1	Тема 9. Нейронні мережі з TensorFlow	Використання функції активації. Оптимізатори. Згортковий шар. Обробка природної мови за допомогою RNN. Створення ігрового генератора.	Тести Лабораторна робота Питання
2/2	Тема 10. Машинне навчання	Ознайомитись із машинним навчанням. Машинне навчання з Python. Основи машинного навчання. Тренування та тестування даних. Класифікація. Побудова моделі. Кластеризація. Приховані моделі Маркова.	Тести Лабораторна робота Питання
4/2	Тема 11. Основні технології Big Data. Сучасні інструменти Data Science	Ознайомитись із поняттям Big Data. Характеристика Big Data. Приклади задач Big Data. Сучасні технології Big Data. Парадигма MapReduce. Платформа для розподілених обчислень Hadoop. Базові відомості про Data Science. Методи використання навчальної інформації. Методи багатомірного розвідувального аналізу (багатовимірний не графічний та багатовимірний графічний). Метрики оцінки якості моделей машинного навчання. Методи прогнозування.	Тести Лабораторна робота Питання

Літературні джерела

1. Andrew Moore. Statistical Data Mining Tutorials [<http://www.autonlab.org/tutorials/>]

2. Великі дані. [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Великі_дані
3. Big Data і блокчейн. [Електронний ресурс]. – Режим доступу: <https://forklog.com/big-data-i-blokchejn-proryv-v-oblasti-analiza-dannyh/>
4. 515K Hotel Reviews Data in Europe [Режим електронного доступу] https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe?select=Hotel_Reviews.csv
5. The Python Package Index (PyPI) is a repository of software for the Python programming language [Режим електронного доступу] <https://pypi.org/>
6. NumPy and Pandas Tutorial – Data Analysis with Python <https://cloudxlab.com/blog/numpy-pandas-introduction/>
7. PolyAnalyst 6.5 Technical capabilities and system requirements, Megaputer intelligence, 2019.
8. "Ukrainian Data Mining Group" - <http://www.udmg.com.ua/>
9. "Блог Data Mining в Україні" - <https://datamining.com.ua/>
10. "Data Mining: моделі та методи" - <http://study.uaitlab.com/datamining/>
11. "Data Science UA" - <https://datascienceua.com/>
12. Pierre Baldi and Søren Brunak. Chapter 6. Neural networks: applications. In Bioinformatics: The Machine Learning Approach. MIT press, 2020.
13. Machine Learning [Електронний ресурс] Автор brilliant.org <https://brilliant.org/wiki/machine-learning/#techniques-in-machine-learning>
14. <https://www.python.org/about/apps/>
15. <https://www.kaggle.com/docs/notebooks>
16. <https://numpy.org/doc/stable/user/whatisnumpy.html>
17. Keras Models – Types and Examples [Електронний ресурс] Автор data-
18. flair.training <https://data-flair.training/blogs/keras-models/>
19. Fruits 360 [Електронний ресурс] 2020 Автор Mihai Oltean <https://www.kaggle.com/moltean/fruits>
20. Наївний басів класифікатор - Вікіпедія. *Вікіпедія*.
URL: https://uk.wikipedia.org/wiki/%D0%9D%D0%B0%D1%97%D0%B2%D0%BD%D0%B8%D0%B9_%D0%B1%D0%B0%D1%94%D1%81%D1%96%D0%B2_%D0%BA%D0%BB%D0%B0%D1%81%D0%B8%D1%84%D1%96%D0%BA%D0%B0%D1%82%D0%BE%D1%80 (дата звернення: 10.09.2024).
21. 10 найкращих інструментів аналізу даних
URL: https://uk.myservername.com/10-best-data-analysis-tools#1_Microsoft_Power_BI (дата звернення: 30.08.2024).
22. RapidMiner studio - rapidminer documentation. *Home - RapidMiner Documentation*.
URL: <https://docs.rapidminer.com/latest/studio/> (дата звернення: 1.09.2024).
23. Power BI Tutorial - Microsoft
URL: <https://learn.microsoft.com/uk-ua/power-bi/create-reports/desktop-excel-stunning-report> (дата звернення: 15.08.2024).

24. <http://dspace.nbu.gov.ua/bitstream/handle/123456789/150922/06-Rogushina.pdf?sequence=1> - Засоби та методи аналізу неструктурованих даних
25. <https://core.ac.uk/download/pdf/149247424.pdf> - контент аналіз
26. <https://actualproblems.dp.ua> › article › інтелектуальний аналіз тексту
27. <https://uk.wikipedia.org/wiki> інтелектуальний аналіз тексту

Політика оцінювання

Модуль 1		Модуль 2	Модуль 3	Модуль 4
20%	20%	5%	15%	40%
Поточне оцінювання	Модульний контроль	Тренінги	Самостійна робота	Екзамен
Оцінка за даний модуль визначається як середнє арифметичне з оцінок, отриманих на практичних заняттях	Підсумкова письмова робота за темами №1-11	Оцінка за виконання одного вибраного завдання тренінгу	Оцінка за виконаний і представлений реферат на вибрану тему	1. Теоретичні питання: 2 питання по 30 балів - тах 60 балів. 2. Практичне завдання - тах 40 балів

Шкала оцінювання

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90-100	відмінно	A (відмінно)
85-89	добре	B (дуже добре)
75-84		C (добре)
65-74	задовільно	D (задовільно)
60-64		E (достатньо)
35-59	незадовільно	FX (незадовільно з можливістю повторного складання)
1-34		F (незадовільно з обов'язковим повторним курсом)