МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЗАПОРІЗЬКА ПОЛІТЕХНІКА»

Кафедра системного аналізу та обчислювальної математики
(найменування кафедри)

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Методи Data Mining
(назва навчальної дисципліни)

Освітня програма: Інтелектуальні технології та прийняття рішень в складних системах
(назва освітньої програми)

Спеціальність: 124 – Системний аналіз
(найменування спеціальності)

Галузь знань: 12 – Інформаційні технології
(найменування галузі знань)

Ступінь вищої освіти: бакалаврський
(назва ступеня вищої освіти)

Затверджено на засіданні кафедри
СА та ОМ
(найменування кафедри)

Протокол № 18 від 16.08.2021 р.

м. Запоріжжя 2021

1. Загальна інформація	
Назва дисципліни	*Машинне навчання*
Рівень вищої освіти	*Перший (бакалаврський) рівень*
Викладач	*Широкорад Дмитро Вікторович, к.ф.-м.н., ст. викладач*
Контактна інформація викладача	*Телефон кафедри 769-8247, Телефон викладача +380631182567, hoveringphoenix@gmail.com*
Час і місце проведення навчальної дисципліни	*Предметна аудиторія кафедри, комп'ютерні класи згідно розкладу занять*
Обсяг дисципліни	*розподіл годин: лекції-30, лабораторні заняття-14, вид контролю- залік,*
Консультації	*Згідно з графіком консультацій*

2. Пререквізити і постреквізити навчальної дисципліни

Пререквізити: основи математичного аналізу, лінійна алгебра, програмування

3. Характеристика навчальної дисципліни

Загальні компетентності:

- К01. Здатність до абстрактного мислення, аналізу та синтезу;
- К02. Здатність застосовувати знання у практичних ситуаціях
- К04. Знання та розуміння предметної області та розуміння професійної діяльності;
- К07. Здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- К10. Здатність працювати автономно;

Фахові компетентності:

- К18. Здатність формалізувати проблеми, описані природною мовою, у тому числі за допомогою математичних методів, застосовувати загальні підходи до математичного моделювання конкретних процесів;
- К20. Здатність визначати основні чинники, які впливають на розвиток фізичних, економічних, соціальних процесів, виокремлювати в них стохастичні та невизначені показники, формулювати їх у вигляді випадкових або нечітких величин, векторів, процесів та досліджувати залежності між ними.
- К22. Здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, оптимізації, обробки інформації, інтелектуального аналізу даних.

Результати навчання:

- ПР08. Володіти сучасними методами розробки програм і програмних комплексів та прийняття оптимальних рішень щодо складу програмного забезпечення, алгоритмів процедур і операцій.
- ПР09. Вміти створювати ефективні алгоритми для обчислювальних задач системного аналізу та систем підтримки прийняття рішень.
- ПР12. Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу.

Результати навчання:

В результаті вивчення дисципліни «*Методи Data Mining»* студент повинен

знати:

- основні підходи до візуалізації даних;
- методи обробки текстових даних для виявлення необхідної інформації;
- алгоритми організації інтернет-пошуку;
- методи кластеризації текстових документів та пошуку шаблонів у різноманітних даних;

вміти:

- будувати наочні візуалізації;
- видобувати інформацію з текстових даних;
- проводити кластеризацію документів;
- шукати патерни в даних різної природи.

4. Мета вивчення навчальної дисципліни

Метою вивчення курсу «Методи Data Mining» є вивчення інструментів аналізу великих даних з метою пошуку корисних фактів.

5. Завдання вивчення дисципліни

Завданням вивчення навчальної дисципліни є набуття компетенцій та навичок з методів видобутку даних як для структурованих даних, які відповідають чітко визначеній схемі, так і для неструктурованих даних, які існують у вигляді тексту природною мовою.

6. Зміст навчальної дисципліни

У навчальній дисципліні «Методи Data Mining» розглядаються виявлення шаблонів, кластеризація, пошук тексту, видобування та аналітика тексту, візуалізацію даних.

7. План вивчення навчальної дисципліни

№ тижня	Назва теми	Форми організації навчання	Кількість годин
1.	**Візуалізація даних.** Взаємодія комп'ютера та людини. Візуалізація числових та нечислових даних. Панель візуалізації.	Лекції, лабораторні заняття, самостійна робота	4
2	**Обробка текстових даних.** Техніки обробки природньої мови. Модель векторного простору. TF перетворення. Імплементація TR систем. Швидкий пошук.	Лекції, лабораторні заняття, самостійна робота	4
3	**Мовна модель.** Статистична мовна модель. Функція пошуку ймовірності запиту. Методи згладжування.	Лекції, лабораторні заняття, самостійна робота	4
4	**Інтернет-пошук.** Індексація у веб. Аналіз посилань. Ранжування результатів пошуку. Рекомендаційні системи.	Лекції, лабораторні заняття, самостійна робота	4
5.	**Майнінг та аналіз тексту.** Аналіз змісту природної мови. Майнінг асоціацій слів. Майнінг парадигматичних відношень Синтагматичні відношення.	Лекції, лабораторні заняття, самостійна робота	4
6.	**Імовірнісне тематичне моделювання.** Юніграм мовні моделі. Оцінкі моделі суміші. EM-алгоритм. Імовірнісний латентно-семантичний аналіз. Латентне розміщення Діріхле.	Лекції, лабораторні заняття, самостійна робота	4
7.	**Кластеризація документів.** Генеративні імовірнісні моделі. Підходи на основі подібності. Категоризація тексту.	Лекції, лабораторні заняття, самостійна робота	4
8	**Пошук шаблонів.** Шаблони та асоціативні правила. Закриті шаблони та макс-шаблони. Алгоритм Apriori. Алгоритм Frequent Pattern-Growth (FPG).	Лекції, лабораторні заняття, самостійна робота	4
9	**Майнінг складних шаблонів.** Багаторівневі асоціації. Багатовимірні асоціації. Кількісні асоціації. Майнінг негативних кореляцій. Стиснуті шаблони**.**	Лекції, лабораторні заняття, самостійна робота	4
10	**Послідовні шаблони.** Пошук послідовних шаблонів на основі апріорних алгоритмів. SPADE. PrefixSpan. Clospan.	Лекції, лабораторні заняття, самостійна робота	4

11	**Майнінг просторових асоціацій.** Просторові шаблони колокацій. Шаблони для кількох траєкторій. Шаблони семантичного руху. Шаблони періодичного руху.	Лекції, лабораторні заняття, самостійна робота	4

8. Самостійна робота

Необхідним елементом успішного засвоєння навчального матеріалу дисципліни є самостійна робота студентів з вітчизняною та закордонною спеціальною літературою. Самостійна робота є основним засобом оволодіння навчальним матеріалом у час, вільний від обов'язкових навчальних занять.

Основні види самостійної роботи, запропоновані студентам:

- *вивчення лекційного матеріалу;*
- *робота з рекомендованою літературою;*
- *вивчення окремих тем або питань, що передбачені для самостійного опрацювання;*
- *вивчення основних термінів та понять з галузі регресійного аналізу;*
- *підготовка до лабораторних занять;*
- *підготовка до проміжного та підсумкового контролю;*
- *контрольна перевірка кожним студентом особистих знань за питаннями для самостійного поглибленого вивчення та самоконтролю;*
- *робота над індивідуальним завданням.*

9. Система та критерії оцінювання курсу

Проміжний контроль знань студентів включає такі заходи:

- *виконання та захист лабораторних робіт, що передбачені учбовим планом;*
- *аудиторні контрольні роботи (2 за семестр), що проводяться перед кожним підсумковим контролем.*

Підсумковий контроль знань студентів включає: залік.

Оцінювання знань при проміжному контролю можливо таким чином:

- *за відмінний захист лабораторних робіт – 50 балів;*
- *при відмінному виконані всіх завдань підсумкової контрольної роботи (контрольної роботи для заочної форми) – 50 балів.*

10. Політика курсу

Академічна доброчесність. Студенти не видають за свої результати роботи інших людей. При використанні чужих ідей і тверджень у власних роботах обов'язково посилаються на використані джерела інформації. Під час оцінювання результатів навчання не користуються недозволеними засобами, самостійно виконують навчальні завдання, завдання поточного та підсумкового контролю.

Відвідування занять. Очікується, що всі студенти відвідають усі лекції та практичні заняття курсу. За згоди декана та викладача дозволяється перейти на індивідуальний графік занять. У будь-якому випадку студенти зобов'язані дотримуватися термінів виконання усіх видів робіт, передбачених робочою програмою курсу.

Література

1. Рогушина Ю.В., Гладун А.Я. Data Mining: пошук знань в даних. – К. : АДЕФ-Україна, Універсаріум, 2016. – 452 с.

2. Провост Ф., Фоусет Т. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані. – К.: Наш формат, 2019. – 400 с.

3. Марченко О. О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.

4. Олійник А.О. Інтелектуальний аналіз даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О.О.Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.

5. Литвин В.В. Методи та засоби інженерії даних та знань / В.В.Литвин – Львів: Магнолія-2006, 2012. – 241 с.

6. Гладун А. Я. Семантичні технології: принципи та практики /А. Я. Гладун, Ю. В. Рогушина. – К.: ТОВ «ВД «АДЕФ-Україна», 2016. – 308 с.

7. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг) – Київ: КНЕУ, 2007. — 376 с.
8. Чубукова И.А. Data Mining – Курс лекцій інтернет-університету INTUIT, 2006. – 328 с.

Публікації з тематики курсу

1. Shyrokorad, D. V. Evolution of the Ni-Al Janus-like clusters under the impacts of low-energy Ar and Ar13 projectiles / D. V. Shyrokorad, G. V. Kornich, S. G. Buga // Materials Today Communications. – V. 23. (2020) 101107-12 (Изд. Elsevier) – doi: 10.1016/j.mtcomm.2020.101107.
2. Shyrokorad D.V. Formation of the core-shell structures from bimetallic Janus-like nanoclusters under low-energy Ar and Ar13 impacts: A molecular dynamics study / D. V. Shyrokorad, G. V. Kornich, S. G. Buga // Computational Materials Science. – 2019. – V. 159. – P. 110-119.
3. Dumin O., Plakhtii V., Pochanin G., Shyrokorad D.. Object classification using artificial neural network processing of data obtained by impulse GPR with 1 Tx+ 4Rx antenna system // 2020 IEEE Ukrainian Microwave Week (UkrMW). – Kharkiv, (Ukraine, September 21-25). – 2020. – P. 1140-1144.