Затверджено Вченою Радою ОНУ
імені І.І. Мечникова
від "20" грудня 2016 р. № 4

Одеський національний університет імені І.І. Мечникова
Кафедра математичного забезпечення комп'ютерних систем

"ЗАТВЕРДЖУЮ"
Проректор з науково-педагогічної роботи

"____" _____ 20___ р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ОК19 «Інтелектуальний аналіз даних і методи machine learning»
(назва навчальної дисципліни)

Рівень вищої освіти бакалавр

Спеціальність 126 – Інформаційні системи та технології
(код і назва спеціальності (тей)

Факультет математики, фізики та інформаційних технологій
(назва інституту, факультету)

2021 – 2022

Робоча програма складена на основі навчальної програми з дисципліни «Інтелектуальний аналіз даних і методи machine learning».

Розробник:
Петрушина Т.І., к.ф.-м.н., доцент кафедри МЗКС

Робоча програма затверджена на засіданні кафедри математичного забезпечення комп'ютерних систем

Протокол № 5 від. "08" 12 2021 р.

Завідувач кафедри _____ (підпис) (Є.В. Малахов) (прізвище та ініціали)

Схвалено навчально-методичною комісією (НМК) з ІТ факультету МФІТ

Протокол № 2 від. "10" 12 2021 р.

Голова НМК _____ (підпис) (О.В. Савастру) (прізвище та ініціали)

Переглянуто та затверджено на засіданні кафедри _____

Протокол № ___ від. "____"_____20__ р.

Завідувач кафедри _____ (підпис) (_____) (прізвище та ініціали)

Переглянуто та затверджено на засіданні кафедри _____

Протокол № ___ від. "____"_____20__ р.

Завідувач кафедри _____ (підпис) (_____) (прізвище та ініціали)

1. Опис навчальної дисципліни

<table>
<tr><td rowspan="2">Найменування показників</td><td rowspan="2">Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень</td><td colspan="2">Характеристика навчальної дисципліни</td></tr>
<tr><td>денна форма навчання</td><td>заочна форма навчання</td></tr>
<tr><td rowspan="15">Загальна кількість:

кредитів – 7

годин – 210

залікових модулів – 1

змістовних модулів – 3

ІНДЗ* – _____
(вид завдання)</td><td>Галузь знань
<u>12 – Інформаційні технології</u>
(шифр і назва)</td><td colspan="2">Нормативна</td></tr>
<tr><td rowspan="5">Спеціальність
<u>126 – Інформаційні системи та технології</u>
(шифр і назва)</td><td colspan="2">Рік підготовки:</td></tr>
<tr><td>4</td><td></td></tr>
<tr><td colspan="2">Семестр</td></tr>
<tr><td>1 (7)</td><td></td></tr>
<tr><td colspan="2">Лекції</td></tr>
<tr><td rowspan="9">Рівень вищої освіти:
<u>бакалавр</u></td><td>44 год.</td><td></td></tr>
<tr><td colspan="2">Практичні, семінарські</td></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="2">Лабораторні</td></tr>
<tr><td>44 год.</td><td></td></tr>
<tr><td colspan="2">Самостійна робота</td></tr>
<tr><td>122 год.</td><td></td></tr>
<tr><td colspan="2">Індивідуальні завдання:</td></tr>
<tr><td colspan="2">Вид контролю: екзамен</td></tr>
</table>

* – за наявності

2. Мета та завдання навчальної дисципліни

Метою вивчення дисципліни є формування у студентів уявлення про набори технологій та методів інтелектуального аналізу даних (Data Mining), призначених для обробки більших об'ємів вхідних даних за рахунок автоматизації процесу виведення нових, коректних та потенційно корисних знань, які можуть бути використані для подальшого прийняття рішень.

Завдання:

– вивчення методів та підходів інтелектуального аналізу даних,

– вивчення можливостей практичної реалізації моделей аналізу та обмеженості їх застосування,

– ознайомлення з способами та методами вибору найбільш відповідних методів для вирішення практичних завдань.

Процес вивчення дисципліни спрямований на формування елементів наступних **компетентностей**:

1) загальних:

ЗК.01 Здатність до дослідницької роботи

ЗК.03 Здатність аналізувати та синтезувати науково-технічну, природничо-наукову та загальнонаукову інформацію

ЗК.07 Базові знання в області фундаментальної та прикладної математики та уміння їх застосовувати в науково-дослідній і професійній діяльності

ЗК.09 Знання психологічних закономірностей інформаційної взаємодії людини з технічними пристроями і програмним забезпеченням комп'ютерних систем, знання основ мікроергономіки та ергономічних стандартів

2) спеціальних (фахових):

ФК.01 Здатність до математичного та логічного мислення, знання основних понять, положень і методів фундаментальної математики та вміння їх використовувати під час розв'язання конкретних виробничих завдань

ФК.03 Знання закономірностей випадкових явищ і вміння застосовувати ймовірносно-статистичні методи для вирішення професійних завдань

ФК.06 Знання основних підходів, методів і технологій штучного інтелекту, уміння розробляти та застосовувати моделі відображення знань, стратегії логічного виведення, технологій інженерії знань, технологій і інструментальних засобів побудови інтелектуальних систем

ФК.11 Знання сучасних технологій та інструментальних засобів розробки програмних систем, уміння їх застосовувати на всіх етапах життєвого циклу

ФК.19 Знання концепцій сховищ даних, їх оперативної аналітичної обробки та інтелектуального аналізу; уміння виявляти в даних раніше не відомих знань, необхідних для прийняття рішень в різних сферах професійної діяльності

ФК.28 Ґрунтовна математична підготовка, а також підготовка з теоретичних, методичних і алгоритмічних основ інформаційних технологій для використання математичного апарату під час вирішення прикладних і наукових завдань в області інформаційних систем і технологій

Програмні результати навчання:
ПРН 06. Уміння використовувати основні поняття, ідеї та застосовувати методи фундаментальної та прикладної математики під час розв'язання конкретних задач в сфері інформаційних технологій.
ПРН 09. Уміння розробляти та застосовувати моделі представлення знань, стратегії логічного виведення, технологій інженерії знань, технологій і інструментальних засобів побудови інтелектуальних систем і систем штучного інтелекту
ПРН 20. Уміння моделювати системи та процеси, стани та поведінки складних об'єктів інформатизації в процесі розроблення інформаційних систем і технологій
ПРН 21. Уміння застосовувати технології роботи зі сховищами даних, здійснювати їх аналітичну обробку та інтелектуальний аналіз для забезпечення надійної роботи інформаційних систем
Очікувані результати навчання. У результаті вивчення навчальної дисципліни студент повинен
***знати**: основні поняття та підходи інтелектуального аналізу даних; основні стадії проведення інтелектуального аналізу даних; методи інтелектуального аналізу даних; способи підвищення ефективності прийняття рішення за рахунок застосування методів інтелектуального аналізу даних;*
***вміти**: вибирати засоби аналізу, найбільш ефективні для конкретних даних з урахуванням їх природи, погрешності, просторового та тимчасового вирішення, а також завдань дослідження; правильно розуміти та інтерпретувати отримані результати дослідження; вміти критично оцінювати можливості та обмежування використованих методів.*
На вивчення навчальної дисципліни відводиться 210 годин, що становить 7 кредити ЄКТС.

3. Зміст навчальної дисципліни

Змістовний модуль 1. Основи інтелектуального аналізу даних.
Тема 1. Основні поняття технологій збору і систематизації даних.
Література: [1, 3].
Тема 2. Передобробка даних, візуалізація, первинний статистичний аналіз.
Література: [1, 2, 4, 23].
Тема 3. Основні завдання та класифікація методів їх вирішення.
Література: [1, 3, 4].

Змістовний модуль 2. Методи інтелектуального аналізу даних.
Тема 1. Методи класифікації.
Література: [1,2, 4, 7].
Тема 2. Методи кластерного аналізу.
Література: [3, 6, 7].
Тема 3. Методи кореляційного і регресійного аналізу.

Література: [5, 6, 8].

Тема 4. Аналіз і прогнозування часових рядів.

Література: [1, 7, 9].

Змістовний модуль 3. Нейромережеве моделювання і глибоке навчання.

Тема 1. Нейронні мережі прямого поширення.
Література: [1,5, 6,8].

Тема 2. Рекурентні і згорткові нейронні мережі.
Література: [5, 9, 13].

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин					
	Денна форма					
	Форма контролю	Усього	у тому числі			
			л	п	лаб	ср
1	2	3	4	5	6	7
Змістовний модуль 1. Основи інтелектуального аналізу даних.						
Тема 1.	КО	9	2			6
Тема 2.	КО	21	4		6	8
Тема 3.	КО	22	6		6	10
Змістовний модуль 2. Методи інтелектуального аналізу даних.						
Тема 1.	КО	22	4		4	14
Тема 2.	КО	22	4		4	14
Тема 3.	КО	24	4		4	14
Тема 4.	КО	24	6		6	14
Змістовний модуль 3. Нейромережеве моделювання і глибоке навчання.						
Тема 1.	КО	24	4		4	14
Тема 2.	КО	24	4		4	14
Всього годин		**210**	**44**		**44**	**122**

Форма контролю: **КО** –контрольне опитування (поточне)

5. Теми семінарських занять

Немає

6. Теми практичних занять

Немає

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Підготовка, обробка, аналіз та візуалізація результатів з допомогою MS SQL Analytical Services.	16
2	Кластерний аналіз у середовищі R	10
3	Регресійний аналіз в середовищі R	10
4	Байєсова класифікація в середовищі R	8
	Разом	**44**

8. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	Аналіз існуючих програмних інструментів інтелектуального аналізу даних.	30
2	Використання еволюційних алгоритмів для інтелектуального аналізу даних.	30
3	Основні методи навчання з підкріпленням: динамічне програмування, методи Монте-Карло, SARSA- та Q-навчання.	30
4	Згорткові нейромережі та їх використання щодо інтелектуального аналізу даних.	32
	Разом	**122**

До самостійної роботи відноситься:
[1] – підготовка до лекцій та лабораторних занять;

9. Індивідуальне навчально-дослідне завдання

Немає

10. Методи навчання

Лекції з використанням мультимедійного презентаційного матеріалу.

11. Методи контролю

Підсумковим контролем з дисципліни є екзамен.

Критерії оцінювання на підсумковому контролі:
Екзаменаційний квиток по дисципліні складається з двох частин: теоретичної та практичної. Мінімальна кількість балів, зараховуються як позитивний результат, дорівнює 60 (за 100-бальною шкалою).
Бали розподіляються наступним чином: 60 балів - теоретична частина та 40 балів - практична.
Теоретична частина містить 2 питання, практична - 1 питання.
За бездоганну відповідь на кожне теоретичне питання студент отримує - 30 балів. При цьому відповідь вважається бездоганною, якщо студент повністю розкрив сутність питання, послідовно і логічно його доповів, навів приклади, проілюстрував відповідь необхідним і достатнім кількістю записів, графіків, формул, схем; зробив посилання на відповідні літературні джерела.
За бездоганне виконання завдання практичної частини студент отримує - 40 балів. Завдання практичної частини іспиту вважається виконаним бездоганно, якщо при його вирішенні отримано правильну відповідь, послідовно і логічно викладено рішення, отримано всі сформульовані в завданні результати.

11.1. **Питання до підсумкового контролю**

1. Класифікація завдань аналізу даних.
2. Відстані для числових даних.
3. Кореляційні подібності та відстані.
4. Відстані по строках і перестановках.
5. Евристичні алгоритми кластерного аналізу.
6. Ієрархічна кластеризація.
7. Числові характеристики кластерної перегородки.
8. Байєсівський підхід до проблеми класифікації.
9. Багатовимірний гаусівський розподіл.
10. Лінійний дискримінантний аналіз.
11. Квадратичний дискримінантний аналіз.
12. Однофакторна лінійна регресійний.
13. Множинна лінійна регресія.
14. Перевірка гіпотез про коефіцієнти регресії.
15. Нелінійна регресія.
16. Логістична регресія

12. **Розподіл балів, які отримують студенти**

Поточне тестування та самостійна робота									Сума
Змістовий модуль 1			Змістовий модуль 2				Змістовий модуль 3		
Т1	Т2	Т3	Т1	Т2	Т3	Т4	Т1	Т2	
8	10	12	10	10	10	10	15	15	100

Т1, Т2 ... – теми змістових модулів

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
85-89	B	добре	
75-84	C		
70-74	D	задовільно	
60-69	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання

0-34	F	незадовільно з обов’язковим повторним вивченням дисципліни	не зараховано з обов’язковим повторним вивченням дисципліни

13. Методичне забезпечення

Конспект лекцій в електронному форматі.

14. Рекомендована література

14.1. Основна література

1. Chris Neil. Machine Learning For Beginners: A Comprehensive Guide To Understand Machine Learning. How It Works And How Is Correlated To Artificial Intelligence And Deep Learning. - Alicex Ltd, 2020, - 188 p.
2. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p.
3. Witten, I. H. Data mining: practical machine learning tools and techniques. / Ian H. Witten, Frank Eibe, Mark A. Hall. – 3rd ed. – Morgan Kaufmann Publishers, 2011. – 630 p.

14.2. Допоміжна література

4. Саттон Р. С., Барто Э. Дж. Обучение с подкреплением. М.: ДМК Пресс, 2020.
5. Чубукова И. А. Data Mining : учебное пособие / И. А. Чубукова. - Москва: ИНТУИТ, БИНОМ. Лаб. знаний, 2008.
6. Рафалович В. Data mining, или интеллектуальный анализ данных для занятых. Практический курс / В. Рафалович. - М.: SmartBook, 2018. - 352 с.
7. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. - 3-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2009.
8. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP: учебное пособие / А. А. Барсегян [и др.] . - СПб. : БХВ-Петербург, 2008.
9. А.С.Бирюков, В.В.Рязанов, А.С.Шмаков. Решение задач кластерного анализа коллективами алгоритмов. Журнал вычислительной математики и математической физики, Т.48, 2008, N 1, стр. 176-192.
10. С. Хайкин - Нейронные сети: полный курс, 2-е издание. : Пер. с англ. - М. : Издательский дом "Вильямс", 2006. - 1104 с. : ил. - Парал. тит. англ.
11. С.Рассел, П.Норвиг Искусственный интеллект: современный подход, 2-е изд.. : Пер. С англ. – М.: Издательский дом «Вильямс», 2006. – 1408 с.: ил.
12. Han J. Data Mining: Concepts and Techniques (Second Edition) / J. Han, M. Kamber – Morgan Kaufmann Publishers, 2006. – 800 p.
13. Н.В.Ковшов, В.Л.Моисеев, В.В.Рязанов. Алгоритмы поиска логических закономерностей в задачах распознавания. Журнал вычислительной математики и математической физики, Т.48, 2008, N 2, стр. 329-344.

14. Емельянов В. В., Курейчик В. В., Курейчик В. М. Теория и практика эволюционного моделирования. - М.: ФИЗМАТЛИТ, 2003. - 432 с.
15. Штовба С.Д. Муравьиные алгоритмы. Exponenta Pro, 2003г.
16. Лонца А. Алгоритмы обучения с подкреплением на Python / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2020. – 286 с.: ил.

15. Електронні інформаційні ресурси

1. Data Mining Labs – Режим доступу: http://dmlabs.org/
2. Encog Machine Learning Framework - Режим доступу: https://www.heatonresearch.com/encog/
3. Neural Networks and Deep Learning - Режим доступу: http://neuralnetworksanddeeplearning.com/
4. Кашина О.А., Миссаров М.Д. Электронный курс "Анализ данных в среде R" – Режим доступу: http://zilant.kpfu.ru/course/view.php?id=17341
5. Кашина О.А., Миссаров М.Д. Электронный курс "Статистический анализ данных" – Режим доступу: http://zilant.kpfu.ru/course/view.php?id=17260
6. The R Project for Statistical Computing – Режим доступу: http://www.r-project.org/
7. Курс "Data mining" – Режим доступу: http://www.intuit.ru/studies/courses/6/6/info.
8. Курс "Интеллектуальный анализ данных средствами MS SQL Server 2008" – Режим доступу: http://www.intuit.ru/studies/courses/2312/612/info.
9. Программный комплекс RStudio для интеллектуального анализа данных – Режим доступу: http://www.rstudio.com/
10. Методы интеллектуального анализа данных – Режим доступу: https://www.ibm.com/develo perworks/ru/library/ba-datamining-techniques/