Затверджено Вченою Радою ОНУ
імені І.І. Мечникова
від “20”грудня 20 16 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

ВБ2 Методи Data Mining
(назва навчальної дисципліни)

Рівень вищої освіти _____ третій (доктор філософії)

Спеціальність _____ 122 – Комп’ютерні науки _____
(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 «Методи Data Mining».

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. "8" 12 2021 р.

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

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

Протокол № 2 від. "10" 12 2021 р.

Голова НМК _____ (О.В. Савастру)
(підпис) (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="13">Загальна кількість:

кредитів – 3

годин – 90

залікових модулів – 1

змістовних модулів – 3

ІНДЗ* – _____
(вид завдання)</td><td>Галузь знань
12 – Інформаційні технології
(шифр і назва)</td><td colspan="2">Нормативна</td></tr>
<tr><td rowspan="4">Спеціальність
122 – Комп’ютерні науки
(шифр і назва)</td><td colspan="2">Рік підготовки:</td></tr>
<tr><td>1</td><td></td></tr>
<tr><td colspan="2">Семестр</td></tr>
<tr><td>1</td><td></td></tr>
<tr><td rowspan="9">Рівень вищої освіти:
доктор філософії</td><td colspan="2">Лекції</td></tr>
<tr><td>16 год.</td><td></td></tr>
<tr><td colspan="2">Практичні, семінарські</td></tr>
<tr><td>14 год.</td><td></td></tr>
<tr><td colspan="2">Лабораторні</td></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="2">Самостійна робота</td></tr>
<tr><td>60 год.</td><td></td></tr>
<tr><td colspan="2">Індивідуальні завдання:</td></tr>
<tr><td colspan="2">Вид контролю: залік</td></tr>
</table>

* – за наявності

1. Мета дисципліни

Метою викладання дисципліни є формування у аспірантів знань щодо 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: набуття компетентностей на основі засвоєння основних теоретичних положень щодо існуючих методів та підходів до інтелектуального аналізу даних; набуття практичних навичок 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**:

а) загальних:

ЗК 01. Здатність до абстрактного мислення, аналізу та синтезу.

б) фахових:

СК 01. Здатність виконувати оригінальні дослідження, досягати наукових результатів, які створюють нові знання у комп'ютерній науці та дотичних до неї міждисциплінарних напрямах і можуть бути опубліковані у провідних наукових виданнях з комп'ютерних наук та суміжних галузей.

СК 02. Здатність застосовувати сучасні інформаційні технології, бази даних та інші електронні ресурси, спеціалізоване програмне забезпечення у науковій та навчальній діяльності.

СК 07. Здатність розуміти, виявляти сутність та розв'язувати проблеми перетворення, аналізу та представлення інформації.

СК 10. Здатність створювати та удосконалювати методи аналізу і обробки інформації, а також розробляти на їх основі відповідні інформаційні технології.

Програмні результати навчання:

ПРН 03. Вміння формулювати і перевіряти гіпотези; використовувати для обґрунтування висновків належні докази, зокрема, результати теоретичного аналізу, експериментальних досліджень і математичного та/або комп'ютерного моделювання, наявні літературні дані.

ПРН 04. Вміння розробляти та досліджувати концептуальні, математичні, інформаційні і комп'ютерні моделі процесів і систем, ефективно використовувати їх для отримання нових знань та/або створення інноваційних продуктів у комп'ютерній науці та дотичних міждисциплінарних напрямах.

ПРН 06. Навички застосовувати сучасні інструменти і технології пошуку, оброблення, аналізу і контекстуалізації інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи.

ПРН 08. Глибоке розуміння загальних принципів та методів комп'ютерних наук, а також методології наукових досліджень, та вміння застосувати їх у власних дослідженнях у сфері комп'ютерних наук та у викладацькій практиці.

ПРН 15. Глибоке розуміння методів аналізу, обробки і перетворення даних, принципів вдосконалення таких методів і створення нових, вміння розробляти на їх основі нові інформаційні технології та відповідні інтелектуальні та спеціалізовані системи.

Очікувані результати навчання. У результаті вивчення навчальної дисципліни аспірант 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.*

2. Зміст навчальної дисципліни

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

Змістовний модуль 2. Методи інтелектуального аналізу даних.
Тема 1. Методи класифікації.
Тема 2. Методи кластерного аналізу.
Тема 3. Методи кореляційного і регресійного аналізу.
Тема 4. Аналіз і прогнозування часових рядів.

Змістовний модуль 3. Нейромережеве моделювання і глибоке навчання.
Тема 1. Нейронні мережі прямого поширення.
Тема 2. Рекурентні і згорткові нейронні мережі.

3. 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	1	4		4
Тема 2.	КО	21	2			
Тема 3.	КО	22	2			8
Змістовний модуль 2. Методи інтелектуального аналізу даних.						
Тема 1.	КО	22	1	1		8

Тема 2.	КО	22	2	5		8
Тема 3.	КО	24	2			8
Тема 4.	КО	24	2			8
Змістовний модуль 3. Нейромережеве моделювання і глибоке навчання.						
Тема 1.	КО	24	2	4		8
Тема 2.	КО	24	2			8
Всього годин		**90**	**16**	**14**		**60**

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

4. Теми семінарських/практичних занять

№ з/п	Назва теми	Кількість годин
1	Підготовка, обробка, аналіз та візуалізація результатів з допомогою MS SQL Analytical Services.	4
2	Кластерний аналіз у середовищі R	4
3	Регресійний аналіз в середовищі R	4
4	Байєсова класифікація в середовищі R	2
	Разом	**14**

5. Теми лабораторних занять

Немає

6. Самостійна робота

Основні види самостійної роботи з Модулю1:

– робота над індивідуальним науково-дослідним завданням (10 год);

– підготовка до практичних занять; опрацювання лекційного матеріалу та рекомендованої літератури; самоперевірка аспирантом власних знань за запитаннями до самоконтроля (розподіл годин за темами модулю наведено в таблиці).

№ з/п	Назва теми	Кількість годин
1	Аналіз існуючих програмних інструментів інтелектуального аналізу даних.	10
2	Використання можливості аналізу табличних даних з алгоритмами машинного навчання в Elastic Search	15
3	Основні методи навчання з підкріпленням: динамічне програмування, методи Монте-Карло, SARSA- та Q-навчання.	15
4	Згорткові нейромережі та їх використання щодо інтелектуального аналізу даних.	20

	Разом	**60**

До самостійної роботи відноситься:
[1] – підготовка до лекцій та лабораторних занять;
[2] – виконання індивідуальних завдань.

7. Індивідуальне навчально-дослідне завдання

Окремого завдання немає. Практична робота виконується на основі дисертаційного дослідження

8. Курсовий проект

Немає

9. Методи навчання

Лекції з використанням мультимедійного презентаційного матеріалу.

10. Методи контролю

Навчальна дисципліна вивчається протягом одного семестру та складається з трьох змістових модулів (ЗМ). По завершенні курсу аспіранти складають залік. Залік виставляється за результатами роботи здобувача вищої освіти впродовж усього семестру.

Здобувач вищої освіти одержує підсумкову оцінку, якщо за результатами поточного контролю він набрав 60 і більше балів.

Якщо за результатами поточного контролю здобувач вищої освіти набрав менше 60 балів, він повинен виконати залікове завдання і з урахуванням його результатів одержати відповідну кількість залікових балів із дисципліни.

Якщо за результатами поточного контролю здобувач вищої освіти набрав 60 і більше балів, проте хоче поліпшити свій підсумковий результат, він також може виконати залікове завдання і з урахуванням його результатів підвищити свою підсумкову оцінку з дисципліни.

Під час викладання матеріалу лекцій лектор може здійснювати контрольні опитування здобувачів. Максимальна оцінка, яку аспірант може отримати за ці опитування, складає 10 балів, які є складовими комплексного підсумкового контролю.

Оцінювання під час поточного та підсумкового контролю здійснюється за 100-бальною шкалою шляхом:

– усне опитування на практичних заняттях;
– оцінювання якості виконання індивідуальної роботи (у вигляді проєкту / реферату).

Підсумкове оцінювання здійснюється шляхом врахування усіх отриманих здобувачем вищої освіти оцінок. Підсумкова оцінка визначається як зважена сума

балів за результатами поточного контролю. За три змістових модулі аспірант може одержати максимум 100 балів.

10.1. Критерії оцінювання на підсумковому модульному контролі:

1. Відповідь повинна бути повною і короткою. Вона не повинна мати в собі матеріал, що не відноситься до сутті питання.
2. Чітко формулювати твердження, вправно застосовувати необхідні формули і знання основних питань програми.
3. Відповіді, що мають помилкові твердження оцінюються виходячи з близькості відповіді до правильної.
4. Пропуски в обґрунтуванні тверджень враховуються і це призводить до зменшення кількості балів.
5. Малі недоліки, неточності при викладенні матеріалу, зменшують кількість балів.
6. Незнання і нерозуміння основної ідеї теоретичного питання або задачі призводить до зняття до 90 % балів.
7. Якщо відповідь на питання відсутня то виставляється нуль балів.

10.2. 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. Багатомірний гауссівський розподіл.
10. Лінійний дискримінантний аналіз.
11. Квадратічний дискримінантний аналіз.
12. Однофакторна лінійна регресійний.
13. Множинна лінійна регресія.
14. Перевірка гіпотез про коефіцієнти регресії.
15. Нелінійна регресія.
16. Логістична регресія

11. Розподіл балів, які отримують аспіранти

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

Загальна сума балів	Оцінка ECTS	Національна шкала	
90 — 100	A – «відмінно»	5 «відмінно»	«залік»
85 — 89	B – «дуже добре»	4 «добре»	
75 — 84	C – «добре»		
70 — 74	D – «задовільно»	3 «задовільно»	
60 — 69	E – «допустимо»		
35 — 59	F – «незадовільно з можливістю повторного складання»	2 «незадовільно»	«незалік»
0 — 34	FX – «незадовільно з обов’язковим повторним курсом»		

12. Методичне забезпечення

Конспект лекцій в електронному форматі; комплекс навчально-методичного забезпечення дисципліни; нормативні документи; презентаційні матеріали.

13. Рекомендована література

13.1. 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. Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques (The Morgan Kaufmann Series in Data Management Systems). - Morgan Kaufmann Publishers Inc., 2011.
3. Steve Lohr. Data-Ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else. – HarperBusiness, 2015.
4. Макленнен Дж. Microsoft SQL Server 2008: Data Mining — интеллектуальный анализ данных. – BHV, 2009.
5. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data

Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p.

13.2. Допоміжна література

1. Макшанов А.В. Технологии интеллектуального анализа данных. — М.: Лань. 2019, - 212 с.
2. Рафалович, В. Data mining, или интеллектуальный анализ данных для занятых. Практический курс / В. Рафалович. - М.: SmartBook, 2018. - 352 с.
3. 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.
4. 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.
5. Саттон Р. С., Барто Э. Дж. Обучение с подкреплением. М.: ДМК Пресс, 2020.
6. С.Рассел, П.Норвиг Искусственный интеллект: современный подход, 2-е изд.. : Пер. С англ. – М.: Издательский дом «Вильямс», 2006. – 1408 с.: ил.
7. Лонца А. Алгоритмы обучения с подкреплением на Python / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2020. – 286 с.: ил.
8. Маккинли, У. Python и анализ данных / У. Маккинли. - М.: ДМК, 2015. - 482 с.
9. Мастицкий, С.Э. Статистический анализ и визуализация данных с помощью R / С.Э. Мастицкий. - М.: ДМК, 2015. - 496 с.
10. Форман, Дж. Много цифр: Анализ больших данных при помощи Excel / Дж. Форман. - М.: Альпина Паблишер, 2019. - 461 с.
11. Чубукова И. А. Data Mining : учебное пособие / И. А. Чубукова. - Москва: ИНТУИТ, БИНОМ. Лаб. знаний, 2008.
12. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP : учебное пособие / А. А. Барсегян [и др.] . - СПб. : БХВ-Петербург, 2008.

14. Електронні інформаційні ресурси

1. Zaki M., Meira W. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014. - Режим доступу: http://www.dataminingbook.info/
2. Data Mining Labs – Режим доступу: http://dmlabs.org/
3. Encog Machine Learning Framework - Режим доступу: https://www.heatonresearch.com/encog/
4. Neural Networks and Deep Learning - Режим доступу: http://neuralnetworksanddeeplearning.com/
5. Кашина О.А., Миссаров М.Д. Электронный курс "Анализ данных в среде R" – Режим доступу: http://zilant.kpfu.ru/course/view.php?id=17341
6. Кашина О.А., Миссаров М.Д. Электронный курс "Статистический анализ данных" – Режим доступу: http://zilant.kpfu.ru/course/view.php?id=17260

7. The R Project for Statistical Computing – Режим доступу: http://www.r-project.org/
8. Курс "Data mining" – Режим доступу: http://www.intuit.ru/studies/courses/6/6/info.
9. Курс "Интеллектуальный анализ данных средствами MS SQL Server 2008" – Режим доступу: http://www.intuit.ru/studies/courses/2312/612/info.
10. Программный комплекс RStudio для интеллектуального анализа данных – Режим доступу: http://www.rstudio.com/
11. Методы интеллектуального анализа данных – Режим доступу: https://www.ibm.com/develo perworks/ru/library/ba-datamining-techniques/