

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра економічної кібернетики та прикладної економіки

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
роботи

Микола АЗАРСНКОВ

2020 р.



Робоча програма навчальної дисципліни

Інтелектуальні системи аналізу даних

рівень вищої освіти	<u>третій (доктора філософії)</u>
галузь знань	<u>05 Соціальні та поведінкові науки</u>
спеціальність	<u>051 Економіка</u>
освітня програма	<u>«Економіка»</u>
вид дисципліни	<u>обов'язкова</u>
факультет	<u>економічний</u>

Програму рекомендовано до затвердження Вченою радою економічного факультету

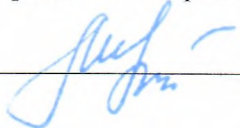
« 26 » червня 2020 року, протокол № 5

РОЗРОБНИКИ ПРОГРАМИ: Катерина Юріївна Кононова д.е.н., професор, професор кафедри економічної кібернетики та прикладної економіки

Програму схвалено на засіданні кафедри економічної кібернетики та прикладної економіки

Протокол від « 04 » червня 2020 року № 10

Завідувач кафедри економічної кібернетики та прикладної економіки


_____ Тамара МЕРКУЛОВА

Програму погоджено з гарантом освітньо-наукової програми «Економіка»

Гарант освітньо-наукової програми
(керівник проектної групи) «Економіка»


_____ Володимир СОБОЛСВ

Програму погоджено науково-методичною комісією економічного факультету

Протокол від « 23 » червня 2020 року № 10

Голова науково-методичної комісії


_____ Вікторія СВТУШЕНКО

ВСТУП

Програма навчальної дисципліни «Інтелектуальні системи аналізу даних» складена відповідно до освітньо-наукової програми третього (доктора філософії) рівня вищої освіти спеціальності 051 «Економіка»

1. Опис навчальної дисципліни

1.1. Мета викладання навчальної дисципліни: формування системи теоретичних знань та практичних навичок з аналізу даних.

1.2. Основні завдання вивчення дисципліни: оволодіння принципами, методами та інструментарієм аналізу даних та способами рішення типових задач аналізу бізнес-даних за допомогою технології машинного навчання.

1.3. Кількість кредитів – 3 кредити.

1.4. Загальна кількість годин – 90 годин.

1.5. Характеристика навчальної дисципліни	
Обов'язкова	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
2-й	2-й
Семестр	
3-й	3-й
Лекції	
8 год.	8 год.
Практичні заняття	
37 год.	6 год.
Самостійна робота, у тому числі	
45 год.	76 год.
Індивідуальні завдання	

1.6. Заплановані результати навчання:

Компетентності:

ЗК02. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

СК03. Здатність використовувати сучасні інформаційні та комунікаційні технології, бази даних та інші електронні ресурси, спеціалізоване програмне забезпечення у науковій та освітній діяльності.

СК06. Здатність обґрунтовувати економічні рішення на основі розуміння закономірностей розвитку соціально-економічних систем і процесів із застосуванням математичних методів та моделей на мікро-, мезо-, та макрорівнях.

Програмні результати:

РН01. Мати теоретичні знання з економіки, соціально-економічних систем і на межі предметних галузей, а також дослідницькі навички, достатні для проведення фундаментальних і прикладних досліджень на рівні останніх світових досягнень з відповідного напрямку, отримання нових знань та/або здійснення інновацій.

РН03. Розробляти та досліджувати фундаментальні та прикладні моделі соціально-економічних процесів і систем, ефективно використовувати їх для отримання нових знань та/або створення інноваційних продуктів у економіці та дотичних міждисциплінарних напрямках.

PH04. Застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу великих масивів даних та/або складної структури, програмне забезпечення та інформаційні системи.

PH05. Розробляти та реалізовувати наукові та/або інноваційні проекти, які дають можливість переосмислити наявне та створити нове цілісне знання та/або професійну практику і розв'язувати значущі проблеми фундаментальної економічної науки з врахуванням соціальних, економічних, екологічних та правових аспектів, лідерства, автономності та відповідальності.

2. Тематичний план навчальної дисципліни

Розділ 1. Основні поняття інтелектуального аналізу даних

Тема 1. Завдання аналізу даних

Тема 2. CRISP-DM методологія

Тема 3. Підготовка даних

Розділ 2. Регресія

Тема 4. Лінійна регресія

Тема 5. Деревя рішень і випадковий ліс

Розділ 3. Класифікація

Тема 6. Логістична регресія

Тема 7. Метод опорних векторів

Тема 8. Метод К найближчих сусідів

Тема 9. Баєсова класифікація

Тема 10. Деревя рішень і випадковий ліс

Розділ 4. Кластеризація

Тема 11. Ієрархічна кластеризація

Тема 12. Кластеризація на основі k-means

3. Структура навчальної дисципліни

Назви розділів	Кількість годин											
	денна форма						заочна форма					
	усього	у тому числі					усього	у тому числі				
л		п	лаб.	інд.	с. р.	л		п	лаб.	інд.	с. р.	
1	2	3	4	5	6	7	8	9	10	11	12	13
Розділ 1. Основні поняття інтелектуального аналізу даних												
Разом за розділом 1	22	2	10			10	16	2	1			16
Розділ 2. Регресія												
Разом за розділом 2	24	2	10			12	24	2	2			20
Розділ 3. Класифікація												
Разом за розділом 3	24	2	10			12	25	2	2			20
Розділ 4. Кластеризація												
Разом за розділом 4	20	2	7			11	25	2	1			20
<i>Усього годин</i>	90	8	37			45	90	8	6			76

4. Теми практичних занять

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1.	Підготовка даних	10	1
2.	Лінійна регресія	5	1
3.	Дерева рішень і випадковий ліс	5	1
4.	Логістична регресія	4	1
5.	Метод опорних векторів, K найближчих сусідів та Баєсова класифікація	3	0,5
6.	Дерева рішень і випадковий ліс	3	0,5
7.	Ієрархічна кластеризація	4	0,5
8.	Кластеризація на основі k-means	3	0,5
	Разом	37	6

5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи	Кількість годин	
		денна	заочна
	Розділ 1		
1.	Збір та підготовка даних за обраною тематикою	10	16
	Разом	10	16
	Розділ 2		
2.	Аналіз даних з використанням лінійної регресії	6	10
3.	Аналіз даних з використанням дерев рішень	6	10
	Разом	12	20
	Розділ 3		
4.	Аналіз даних з використанням логістичної регресії	3	8
5.	Аналіз даних з використанням методів опорних векторів, K найближчих сусідів та Баєсової класифікації	3	8
6.	Аналіз даних з використанням дерев рішень	3	4
	Разом	12	20
	Розділ 4		
7.	Аналіз даних з використанням ієрархічної кластеризації	5	10
8.	Аналіз даних з використанням k-means	6	10
	Разом	11	20
	Разом	45	72

6. Методи контролю

Методи контролю передбачають поточний контроль, виконання самостійних та контрольної роботи. Додаткові бали здобувач вищої освіти отримує за роботу на занятті. Підсумковий контроль проводиться у формі заліку, що проводиться у письмовій формі і включає практичне завдання (40 балів).

7. Індивідуальні завдання (самостійна робота)

Аналіз даних за тематикою дисертаційного дослідження.

Завдання передбачає, що кожен здобувач вищої освіти самостійно збирає дані за тематикою дисертаційного дослідження; проводить їх попередній аналіз та виконує підготовку даних. Згідно до структури та обсягу даних, дослідження проводиться за такими напрямками:

- побудова та аналіз моделей регресії;
- побудова та аналіз моделей класифікації;

- побудова та аналіз моделей кластеризації.

За результатами виконання завдання здобувач вищої освіти готує звіт-презентацію та захищає її.

8. Схема нарахування балів

Поточний контроль, самостійна робота					Залікова робота	Сума
Самостійна робота			Контрольна робота, передбачена навчальним планом	Разом		
T1-T3, T4-T5	T6-10	T11-12				
10	10	10	30	60	40	100

T1 – T14 – теми розділів.

Критерії оцінювання (денна та заочна форма)

1) Контрольна робота, передбачена навчальним планом (30 балів)

Контрольна робота складаються з практичних завдань (всі теми), які виконуються та здаються в електронному вигляді.

Оцінка	Критерії оцінювання
25-30	Завдання виконані в повному обсязі та без помилок. Здобувач вищої освіти демонструє глибоке володіння теоретичним матеріалом, показує уміння користуватися навичками та різними прийомами рішення практичних завдань.
24-19	Зробив всі практичні завдання з окремими незначними помилками. Проявив знання та розуміння основних положень з навчальної дисципліни.
18-13	При виконанні завдань допущені суттєві помилки. Здобувач вищої освіти показує знання основних положень навчального матеріалу на рівні запам'ятовування, але не достатнього розуміння.
12-0	Завдання не виконані зовсім або при їх виконанні допущені грубі помилки.

2) Поточний контроль та самостійна робота

- Розділи 1 та 2 (10 балів)
- Розділ 3 (10 балів)
- Розділи 4 (10 балів)

Оцінка	Критерії оцінювання
10-8	Завдання виконано в повному обсязі та чітко з дотриманням усіх вимог. Своєчасно представлено на перевірку.
7-6	Завдання виконано повністю, допущено незначні помилки. АБО Завдання виконано в повному обсязі та чітко з дотриманням усіх вимог. Представлено на перевірку із запізненням.
6-4	Завдання виконано не повністю, допущено значні помилки. Своєчасно представлено на перевірку. АБО Завдання виконано повністю, допущено незначні помилки. Представлено на перевірку із запізненням.
3-0	Завдання виконано не повністю, допущено значні помилки. Представлено на перевірку зі значним запізненням.

3) Залікова робота (40 балів)

Робота складається з 2-х практичних завдань.

Звіт по кожному завданню включає:

- Скрипт з завантаженими та обробленими без помилок екзаменаційними даними
- Документ звіту - файл у форматі MS Word «Прізвище.docx»

Шкала оцінювання практичних завдань

Кількість балів	Критерії оцінювання
18-20	Здобувач вищої освіти правильно обирає метод розв'язання задачі, володіє різнобічними вміннями, навичками та прийомами рішення завдань. Завдання виконане без помилок. Звіт містить розгорнуті авторські коментарі. Код скрипту чітко структурований
15-17	Здобувач вищої освіти правильно застосовує теоретичні знання та положення при рішенні практичної задачі, володіє необхідними вміннями та навичками роботи з програмами. Виконав завдання з окремими незначними помилками. Звіт містить стислі авторські коментарі. Код скрипту чітко структурований
12-14	Здобувач вищої освіти правильно застосовує теоретичні знання та положення при рішенні практичної задачі, володіє необхідними вміннями та навичками роботи з програмами. Виконав завдання з окремими незначними помилками. Звіт не містить авторських коментарів. Код скрипту структурований не чітко
9-11	Здобувач вищої освіти при розв'язанні практичної задачі допустив значну помилку. Не досить вільно володіє вміннями та навичками роботи з програмами. Звіт не містить авторських коментарів. Код скрипту структурований не чітко
6-8	Здобувач вищої освіти при розв'язанні практичної задачі допустив значну помилку. Звіт не сформовано. Код скрипту містить незначні помилки
3-5	Здобувач вищої освіти при розв'язанні практичної задачі допустив значну помилку. Звіт не сформовано. Код скрипту містить значні помилки
0-2	Здобувач вищої освіти не вміє застосовувати знання на практиці. Не вирішив завдання взагалі або допустив грубих помилок. Звіт не сформовано. Код скрипту містить значні помилки

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка	
	для чотирирівневої шкали оцінювання	для дворівневої шкали оцінювання
90 – 100	відмінно	зараховано
70 – 89	добре	
50 – 69	задовільно	
1 – 49	незадовільно	не зараховано

9. Рекомендована література

Основна література

1. Кононова К. Машинне навчання: методи та моделі : [підручник] // ХНУ імені В.Н. Каразіна, 2019. – 307 с.
2. Кононова К. Інтелектуальні системи аналізу даних: [нав.-мет. посібник] // ХНУ імені В.Н. Каразіна, 2019. – 100 с.

3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP: БХВ-Петербург, 2007
4. Дюк В., Самойленко А. Data Mining: учебный курс. – СПб.: Изд. Питер, 2001
5. Киселев М., Соломатин Е.. Средства добычи знаний в бизнесе и финансах. – Открытые системы, № 4, 1997
6. Криват Б., Макленнен Д., Танг Ч. Microsoft SQL Server 2008: Data Mining – интеллектуальный анализ данных, ВHV, 2009
7. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям. – СПб.: Изд. Питер, 2009
8. Черняк А. И. Интеллектуальный анализ данных: учебник / А.И. Черняк, П.В. Захарченко; Киевский национальный университет им. Т.Шевченко. – К.: Знание, 2014. – 599 с.
9. Чубукова И. А. Data Mining: учебное пособие. – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006

Допоміжна література

10. Asteriou, D., Hall, S.G. (2011). ARIMA Models and the Box–Jenkins Methodology. Applied Econometrics (Second ed.). Palgrave MacMillan.
11. David J. C. MacKay (2003). Information Theory, Inference and Learning Algorithms. Cambridge University Press.
12. Hand D., Manila H., and Smyth P.(2001) Principles of Data Mining, MIT Press.
13. Hastie T., Tibshirani R., Friedman J. (2001). The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, New York.
14. Witten I.H., Frank E. (2005). Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco.
15. Kohonen, T., Honkela, T. (2007). Kohonen Network. Scholarpedia.
16. Mills, T.C. (1990). Time Series Techniques for Economists. Cambridge University Press.
17. Tan, P., Steinbach, M., Kumar, V. (2005). Introduction to Data Mining, Addison Wesley.
18. Rokach, L., Oded, M. (2005) Clustering methods. Data mining and knowledge discovery handbook. Springer US.
19. Ville B. (2001). Microsoft Data Mining. Digital Press.

10. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

20. <http://mymagictools.blogspot.com/2015/07/r.html?view=classic>
21. <http://www.algorithmist.ru/2011/05/clustering-with-example-in-r.html>
22. http://re9ulus.github.io/2015/12/07/trees_in_r/
23. <https://ranalytics.github.io/data-mining/105-Cohonen-Maps.html>
24. https://clarkdatalabs.github.io/soms/SOM_NBA
25. <https://www.shanelynn.ie/self-organising-maps-for-customer-segmentation-using-r/>
26. <https://ranalytics.github.io/data-mining/105-Cohonen-Maps.html>
27. <https://habrahabr.ru/company/ods/blog/>
28. <https://cran.r-project.org/web/packages/>