

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра економічної кібернетики та прикладної економіки

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
роботи

_____ А.В. Пантелеймонов _____

“ _____ ” _____ 20__ р.

Робоча програма навчальної дисципліни

Інтелектуальні системи аналізу даних

рівень вищої освіти _____ другий (магістерський) _____

галузь знань _____ 05 «Соціальні та поведінкові науки» _____

спеціальність _____ 051 Економіка _____

освітня програма _____ Прикладна економіка _____

спеціалізація _____ _____

вид дисципліни _____ за вибором _____

факультет _____ економічний _____

2018 / 2019 навчальний рік

Програму рекомендовано до затвердження Вченою радою економічного факультету

“ 22 ” червня 2018 року, протокол №8

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

Програму схвалено на засіданні кафедри економічної кібернетики та прикладної економіки

Протокол від “ 11 ” червня 2018 року, протокол №11

Завідувач кафедри економічної кібернетики та прикладної економіки

_____ Т.В. Меркулова
(підпис) (прізвище та ініціали)

Програму погоджено методичною комісією економічного факультету

Протокол від “ 21 ” червня 2018 року, протокол №9

Голова методичної комісії економічного факультету

_____ В.А. Євтушенко
(підпис) (прізвище та ініціали)

ВСТУП

Програма навчальної дисципліни «Інтелектуальні системи аналізу даних» складена відповідно до освітньо-професійної програми підготовки

_____ другий (магістерський) _____
(назва рівня вищої освіти)
спеціальності _____ 051 Економіка _____
освітніх програм _____ Прикладна 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 Data Mining.

1.3. Кількість кредитів – 3 кредити.

1.4. Загальна кількість годин – 90 годин.

1.5. Характеристика навчальної дисципліни	
Нормативна / за вибором	
За вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
2-й	1-й
Семестр	
3-й	2-й
Лекції	
18 год.	8 год.
Лабораторні заняття	
18 год.	4 год.
Самостійна робота, у тому числі	
54 год.	78 год.

1.6. Заплановані результати навчання:

Компетентності:

ЗК1. Здатність критично мислити та генерувати нові ідеї.

ФК3. Здатність використовувати сучасні інформаційні технології та економіко-математичні методи і моделі для дослідження економічних та соціальних процесів.

ФК5. Здатність формулювати професійні задачі в сфері економіки, вибирати належні напрями і відповідні методи для їх розв'язання, беручи до уваги наявні ресурси.

ФК8. Здатність застосовувати науковий підхід до формування та обґрунтування ефективних стратегій в економічній діяльності.

Результати:

1. Проводити дослідження, генерувати нові ідеї, здійснювати інноваційну діяльність.
2. Збирати, обробляти та аналізувати статистичні дані, науково-аналітичні матеріали, необхідні для вирішення комплексних економічних завдань.
3. Застосовувати сучасні інформаційні технології у соціально-економічних дослідженнях.
4. Застосовувати наукові підходи до формування та обґрунтування ефективних стратегій в економічній діяльності.

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. Алгоритми аналізу даних: прогнозування

Тема 4. Прогнозування на основі лінійної регресії

Тема 5. Авторегресійна модель

Тема 6. Прогнозування з використанням нейронних мереж

Розділ 3. Алгоритми аналізу даних: класифікація

Тема 7. Класифікація на основі логістичної регресії

Тема 8. Класифікація з використанням дерев рішень

Розділ 4. Алгоритми аналізу даних: кластеризація

Тема 9. Ієрархічна кластеризація

Тема 10. Кластеризація на основі k-means

Тема 11. Кластеризація з використанням карт Кохонена

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	16	2		2		12	14	2				12
Розділ 2. Алгоритми аналізу даних: прогнозування												
Разом за розділом 2	26	6		6		14	27	2		2		23
Розділ 3. Алгоритми аналізу даних: класифікація												
Разом за розділом 3	22	4		4		14	24	2		2		20
Розділ 4. Алгоритми аналізу даних: кластеризація												
Разом за розділом 4	26	6		6		14	25	2				23
Усього годин	90	18		18		54	90	8		4		78

4. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1.	Підготовка даних	2	
2.	Лінійна регресія	2	1
3.	Авторегресійна модель	2	1
4.	Нейронні мережі	2	1
5.	Логістична регресія	2	
6.	Дерева рішень	2	1
7.	Ієрархічна кластеризація	2	
8.	Кластеризація k-means	2	
9.	Карти Кохонена	2	
	Разом	18	4

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	12	12
	Разом	12	12
	Розділ 2		
2.	Аналіз даних з використанням лінійної регресії	4	7
3.	Аналіз даних з використанням авторегресійної моделі	5	8
4.	Аналіз даних з використанням нейронних мереж	5	8
	Разом	14	23
	Розділ 3		
5.	Аналіз даних з використанням логістичної регресії	7	10
6.	Аналіз даних з використанням дерев рішень	7	10
	Разом	14	20
	Розділ 4		
7.	Аналіз даних з використанням ієрархічної кластеризації	4	7
8.	Аналіз даних з використанням k-means	5	8
9.	Аналіз даних з використанням карт Кохонена	5	8
	Разом	14	23
	Разом	54	78

6. Методи контролю

Методи контролю передбачають поточний контроль, виконання самостійних та контрольних робіт. Додаткові бали студент отримує за роботу на занятті. Підсумковий контроль проводиться у формі екзамену, що проводиться у письмовій формі і включає практичне завдання (одну задачу, 40 балів).

8. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання				Екзаменаційна робота (письмова)	Сума
Самостійні роботи	Контрольні роботи	Додаткові бали	Разом		
20	30	10	60	40	100

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	

9. Рекомендована література

Основна література

1. 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
2. Дюк В., Самойленко А. Data Mining: учебный курс. – СПб.: Изд. Питер, 2001
3. Киселев М., Соломатин Е.. Средства добычи знаний в бизнесе и финансах. – Открытые системы, № 4, 1997
4. Криват Б., Макленнен Д., Танг Ч. Microsoft SQL Server 2008: Data Mining – интеллектуальный анализ данных, ВHV, 2009
5. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям. – СПб.: Изд. Питер, 2009
6. Черняк А. И. Интеллектуальный анализ данных: учебник / А.И. Черняк, П.В. Захарченко; Киевский национальный университет им. Т.Шевченко. – К.: Знание, 2014. – 599 с.
7. Чубукова И. А. Data Mining: учебное пособие. – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006

Допоміжна література

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

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

10. <http://mymagictools.blogspot.com/2015/07/r.html?view=classic>
11. <http://www.algorithmist.ru/2011/05/clustering-with-example-in-r.html>
12. http://re9ulus.github.io/2015/12/07/trees_in_r/
13. <https://ranalytics.github.io/data-mining/105-Cohonen-Maps.html>

14. https://clarkdatalabs.github.io/soms/SOM_NBA
15. <https://www.shanelynn.ie/self-organising-maps-for-customer-segmentation-using-r/>
16. <https://ranalytics.github.io/data-mining/105-Cohonen-Maps.html>
17. <https://habrahabr.ru/company/ods/blog/>
18. <https://cran.r-project.org/web/packages/>