

ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ АНАЛІЗУ ДАНИХ

Силабус курсу

1. Загальні відомості

Лектор: д.е.н., проф. Катерина Юріївна Кононова

(CV, <http://www.cyber.kharkov.ua/page-kononova.html>)

Асистент: аспірант Антон Олегович Дек

Консультавання: понеділок 10⁰⁰ – 12⁰⁰

Кафедра економічної кібернетики та прикладної економіки,

ауд. 3-16, 3-17, 3-18,

м. Харків, вул. Миросицька, 1.

тел. (057) 707 51 94,

e-mail: кафедра економічної кібернетики та прикладної економіки

cyber@karazin.ua,

Кононова К.Ю. kateryna.kononova@karazin.ua,

Дек А.О. dek@karazin.ua

2. Характеристика та обсяг навчальної дисципліни

Кількість кредитів – 3 кредитів, загальна кількість годин – 90.

За вибором	
Денна форма навчання	Заочна форма навчання
Рік підготовки	
2-й	2-й
Семестр	
3-й	2-й
Лекції	
8 год.	8 год.
Практичні заняття	
37 год.	10 год.
Самостійна робота, у тому числі	
45 год.	72 год.

3. Опис курсу

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

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

Методи навчання: лекції, лабораторні заняття із розробкою програмного коду для розв'язання задач з використанням технологій машинного навчання, виконання індивідуальних завдань.

Результати навчання

Здобувачі вищої освіти будуть *знати*: методи підготовки даних, моделі лінійної регресії, дерев рішень і випадкового лісу, логістичної регресії, опорних векторів, K найближчих сусідів та баєсової класифікації, ієрархічної кластеризація та на основі k-means.

вміти: збирати, обробляти та аналізувати статистичні дані, науково-аналітичні матеріали, необхідні для вирішення комплексних економічних завдань та застосовувати сучасні інформаційні технології у соціально-економічних дослідженнях.

Результати навчання сприятимуть формуванню вмінь аргументувати нові та складні ідеї щодо аналізу соціально-економічних даних, а також здатності виймати корисну інформацію з баз даних.

4. Тематичний план навчальної дисципліни

Тема 1. Основні поняття інтелектуального аналізу даних

Завдання аналізу даних. CRISP-DM методологія. Підготовка даних

Тема 2. Регресія

Лінійна регресія. Древа рішень і випадковий ліс

Тема 3. Класифікація

Логістична регресія. Метод опорних векторів. Метод K найближчих сусідів. Баєсова класифікація. Древа рішень і випадковий ліс

Тема 4. Кластеризація

Ієрархічна кластеризація. Кластеризація на основі k-means

5. Теми практичних занять

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

6. Завдання для самостійної роботи

Самостійна робота здобувачів вищої освіти передбачає: вивчення тем вказаних в таблиці нижче, опрацювання рекомендованої літератури (п. 11), виконання індивідуального завдання (п. 7).

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1.	Збір та підготовка даних за обраною тематикою	6	8
2.	Аналіз даних з використанням лінійної регресії	7	8
3.	Аналіз даних з використанням дерев рішень	3	4
4.	Аналіз даних з використанням логістичної регресії	4	4

5.	Аналіз даних з використанням методів опорних векторів, K найближчих сусідів та Баєсової класифікації	4	5
6.	Аналіз даних з використанням дерев рішень	2	3
7.	Аналіз даних з використанням ієрархічної кластеризації	2	2
8.	Аналіз даних з використанням k-means	2	3
	Разом	60	78

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. Методи контролю та критерії оцінювання

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 (одну задачу, 40 балів).

Схема нарахування балів

Поточний контроль, самостійна робота					Залікова робота	Сума
Розділ 1- Розділ 2	Розділ 3	Розділ 4	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 уміннями, 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 уміннями

	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 уміє застосовувати знання на практиці. Не вирішив завдання взагалі або допустив грубих помилок. Звіт не сформовано. Код скрипту містить значні помилки

9. Політика курсу щодо дотримання принципів академічної доброчесності. Суворе дотримання принципів академічної доброчесності згідно до [Положення](https://www.univer.kharkov.ua/docs/antiplagiat_nakaz_polozhennya.pdf) про систему запобігання та виявлення академічного плагіату у наукових та навчальних працях працівників і здобувачів вищої освіти Харківського національного університету імені В. Н. Каразіна (введено в дію наказом ректора № 0501-1/173 від 14.05.2015 р., https://www.univer.kharkov.ua/docs/antiplagiat_nakaz_polozhennya.pdf).

10. Оскарження результатів підсумкового контролю здійснюється згідно до [Положення](https://drive.google.com/drive/folders/0BwRBoOD_AIQSZ0xDtkt2eDNrX0U) про організацію освітнього процесу в Харківському національному університеті імені В. Н. Каразіна (введено в дію наказом ректора № 0211-1/342 від 10.07.2018 р., https://drive.google.com/drive/folders/0BwRBoOD_AIQSZ0xDtkt2eDNrX0U).

11. 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.

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/>