

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра моделювання систем та технологій

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної роботи



Пантелеймонов А. В.

” 2018 р.

Робоча програма навчальної дисципліни

Інтелектуальний аналіз даних

(назва навчальної дисципліни)

рівень вищої освіти	перший (бакалаврський рівень)
галузь знань	0501 Інформатика та обчислювальна техніка (шифр і назва)
Напрямок підготовки	6.050101 комп'ютерні науки (шифр і назва)
вид дисципліни	обов'язкова
факультет	комп'ютерних наук

2018 / 2019 навчальний рік

Програму рекомендовано до затвердження вченою радою факультету (інституту, центру)

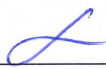
“29” серпня 2018 року, протокол № 9

РОЗРОБНИК ПРОГРАМИ:

старший викладач **Діденко Євген Володимирович**

“11”червня 2018 року протокол №14

Завідувач кафедри моделювання систем і технологій

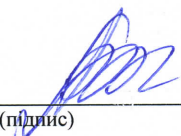

(підпис)

Краснобаєв В.А.
(прізвище та ініціали)

Програму погоджено методичною комісією факультету комп'ютерних наук

“ 27 ”червня 2018 року протокол № 7

Голова методичної комісії


(підпис)

Васильєва Л. В.
(прізвище та ініціали)



ВСТУП

Програма навчальної дисципліни “Інтелектуальний аналіз даних” складена відповідно до освітньо-професійної програми підготовки

бакалавр

(назва рівня вищої освіти, освітньо-кваліфікаційного рівня)

напряму Комп'ютерні науки

спеціалізації

1. Опис навчальної дисципліни

1.1. Мета викладання навчальної дисципліни

Засвоєння студентами основних теоретичних відомостей та практичних вмінь з курсу. Підготувати студента до ефективного використання як класичних, так і сучасних методів інтелектуального аналізу даних та обробки інформації з використанням обчислювальних методів та нейромереж для проектування елементів автоматизованих комп'ютерних систем та їх використання у подальшій професійній діяльності; допомогти набутти навички практичної роботи з програмними засобами для інтелектуальної обробки даних.

1.2. Основні завдання вивчення дисципліни

Оволодіння студентами науковими основами, сучасною методологією та особливостями застосування інтелектуальної обробки даних; засвоєння майбутніми фахівцями теоретичних основ інформаційних систем, орієнтованих на застосування стандартів Data Mining; набуття умінь програмувати окремі елементи систем Data Mining різного призначення і різної проблемної орієнтації на всіх стадіях життєвого циклу інформаційної системи; отримання практичних навичок використання і адаптації деяких найбільш відомих систем та бібліотек Data Mining.

1.3. Кількість кредитів 4

1.4. Загальна кількість годин 120

1.5. Характеристика навчальної дисципліни	
Нормативна / за вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
4-й	4-й
Семестр	
7-й	7-й
Лекції	
16 год.	год.
Практичні, семінарські заняття	
8 год.	год.
Лабораторні заняття	
8 год.	год.
Самостійна робота	
88 год.	год.
Індивідуальні завдання	
год.	

1.6. Заплановані результати навчання

знати:

- основні поняття та визначення інтелектуального аналізу даних;
- моделі та методи побудови моделей та аналізу залежностей у даних;
- сучасні програмні засоби для проектування і розробки систем інтелектуального аналізу даних;
- критерії порівняння моделей і методів інтелектуального аналізу даних.

вміти:

- обґрунтовувати й аналізувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні практичних задач;
- використовувати сучасні програмні засоби для проектування та дослідження систем інтелектуального аналізу даних;
- створювати програми для інтелектуального аналізу даних при розв'язку конкретних практичних задач;
- аналізувати результати побудови та використання систем інтелектуального аналізу даних при вирішенні прикладних задач.

2. Тематичний план навчальної дисципліни

Розділ 1. Традиційні та сучасні технології інтелектуальної обробки даних.

Тема 1. Введення в інтелектуальний аналіз даних.

Визначення інтелектуального аналізу даних (Data Mining), історія та передумови виникнення. Data Mining і витяг знань з даних (Knowledge Discovery in Databases). Місце і роль Data Mining в процесі прийняття рішень. Основні завдання Data Mining, види моделей. Data Mining і статистичний аналіз. Data Mining і сховища даних.

Тема 2. Методи інтелектуального аналізу даних

Класи задач (типи закономірностей), вирішуються за допомогою Data Mining: класифікація, кластеризація, регресія (прогнозування), асоціація, послідовні шаблони. Методи отримання і аналізу даних: описові - дисперсійний та регресійний аналіз; аналіз часових рядів, кластерний аналіз, еволюційні - дерева рішень, генетичні алгоритми; штучні нейронні мережі, метод «найближчих сусідів».

Тема 3. Етапи проведення інтелектуального аналізу даних.

Етапи побудови моделі методами Data Mining: постановка завдання, підготовка даних, визначення джерел даних для аналізу, вибір, очищення та попередня обробка даних. Побудова моделей із зазначенням алгоритмів інтелектуального аналізу даних і їх параметри.

Розділ 2. Аналіз текстової інформації – Text Mining.

Тема 4. Задача та етапи аналізу текстів Text Mining.

Задача та етапи аналізу текстів Text Mining. Загальний опис процесу отримання понять з тексту. Методи класифікації текстових документів.

Тема 5. Методи класифікації текстових документів

Методи кластеризації текстових документів: ієрархічні, бінарні. Представлення текстових документів.

Тема 6. Задача анутовування текстів. Засоби аналізу текстової інформації.

Задача анутовування текстів. Засоби аналізу текстової інформації.

Розділ 3. Кластеризація та веб-майнінг

Тема 9. Кластерний аналіз

Основні завдання кластерного аналізу: розробка типології або класифікації, дослідження концептуальних схем групування об'єктів, висунення гіпотез на основі дослідження даних, перевірка гіпотез. Основні етапи кластерного аналізу. Алгоритми кластеризації: м'яка і тверда кластеризація. Області застосування кластерного аналізу.

Тема 10. Веб-майнінг

Застосування методів і алгоритмів Data Mining для виявлення і пошуку залежностей і знань в мережі Інтернет. Розрахунок показника Page-rank

3. Структура навчальної дисципліни

Назви розділів і тем	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 Традиційні та сучасні технології інтелектуальної обробки даних.												
Тема 1. Введення в інтелектуальний аналіз даних.	12	2				10						
Тема 2. Методи інтелектуального аналізу даних	14	2		2		10						
Тема 3. Етапи проведення інтелектуального аналізу даних.	15	2	2	2		9						
Разом за розділом 1	41	6	2	4		29	105					
Розділ 2. Аналіз текстової інформації – Text Mining.												
Тема 6. Задача та етапи аналізу текстів Text Mining.	12	2				10						
Тема 7. Методи класифікації текстових документів	16	2	2	2		10						
Тема 8. Задача	17	2				15						

анотування текстів. Засоби аналізу текстової інформації.												
Разом за розділом 2	45	6	2	2		35						
Розділ 3. Кластеризація та веб-майнінг												
Тема 9. Кластерний аналіз	16	2	2	2		12						
Тема 10. Веб- майнінг	18	2	2			12						
Разом за розділом 3	34	4	4	2		24						
Усього годин	120	16	8	8	0	88						

4. Теми практичних та лабораторних занять

№ теми	Назва теми	Кількість годин
Розділ. 1 Традиційні та сучасні технології інтелектуальної обробки даних.		
Практичні заняття		
3	«Ознайомлення з інтелектуальним аналізом даних ». Основні методи та приклади.	2
Лабораторні заняття		
2	«Лабораторна робота №1». Обробка даних експериментів та розрахунок основних статистичних характеристик	2
3	«Лабораторна робота №2». Розрахунок статистичних характеристик та кластеризація дата-сету	2
Розділ 2. Аналіз текстової інформації – Text Mining.		
Лабораторні заняття		
7	«Лабораторна робота №3». Використання найвнього баїсовського класифікатору для класифікації текстової інформації	2
Практичні заняття		
7	Робота з текстовими даними в Python (стеммінг)	2
Розділ 3. Кластеризація та веб-майнінг		
Практичні заняття		

9	Робота з графічним відображенням в у Python.	2
10	Створення парсеру веб-сторінок та розрахунок Page-rank	2
	Лабораторні заняття	
9	«Лабораторна робота №4». Кластеризація методом k-середніх	2

5. Завдання для самостійної роботи

№ з/п		Види, зміст самостійної роботи. Назва теми.	Кількість годин
Розділ 1. Традиційні та сучасні технології інтелектуальної обробки даних.			
1	Опрацювання навчального матеріалу	Основні завдання інтелектуального аналізу даних.	10
2	Опрацювання навчального матеріалу	Методи інтелектуального аналізу даних	10
3	Опрацювання навчального матеріалу	Етапи проведення інтелектуального аналізу даних.	9
		Разом	29
Розділ 2. Аналіз текстової інформації – Text Mining.			
6	Опрацювання навчального матеріалу	Задача та етапи аналізу текстів Text Mining.	10
7	Опрацювання навчального матеріалу	Загальний опис процесу отримання понять з тексту. Методи класифікації текстових документів	10
8	Опрацювання навчального матеріалу	Представлення текстових документів. Задача анотування текстів. Засоби аналізу текстової інформації.	15
		Разом	35
Розділ 3. Кластеризація та веб-майнінг			

9	Опрацювання навчального матеріалу	Кластерний аналіз	12
10	Опрацювання навчального матеріалу	Веб-майнінг та розрахунок page-rank для маленької мережі веб-сторінок	12
		Разом	24

6. Індивідуальні завдання

7. Методи контролю

Контроль засвоєння навчального матеріалу здійснюється шляхом:

- Проведення контрольної роботи
- Проведення заліку з використанням інформаційних технологій (проведення електронного тестування).

8. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання											Залік	Сума
Розділ 1			Розділ 2			Розділ 3			Контрольна робота	Разом		
1	2	3	6	7	8	9	10	11	10	60	40	100
20			20			10						

Шкала оцінювання

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

Оцінка в балах	Оцінка за національною шкалою	
Оцінка	Пояснення	
90 – 100	“зараховано”	Теоретичний зміст курсу освоєний цілком, необхідні практичні навички роботи з освоєним матеріалом сформовані, всі навчальні завдання, які передбачені програмою навчання виконані повному обсязі, відмінна робота без помилок або з однією незначною помилкою.
70 – 89	“зараховано”	Теоретичний зміст курсу освоєний цілком, практичні навички роботи з освоєним матеріалом в основному сформовані, всі навчальні завдання, які передбачені програмою навчання виконані, якість виконання жодного з них не оцінено мінімальним числом балів, деякі види завдань виконані з помилками, робота з декількома незначними помилками, або з однією – двома значними помилками.
50-69	“зараховано”	Теоретичний зміст курсу освоєний не повністю, але прогалини не носять істотного характеру, необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, більшість передбачених програмою навчання навчальних завдань виконано, деякі з виконаних завдань, містять помилки, робота з трьома значними помилками.
1-49	„не зараховано”	Теоретичний зміст курсу не освоєно, необхідні практичні навички роботи не сформовані, всі виконані навчальні завдання містять грубі помилки, додаткова самостійна робота над матеріалом курсу не приведе до значимого підвищення якості виконання навчальних завдань, робота, що потребує повної переробки

9. Рекомендована література

Основна література

1. Ковальчук А.М. Основи проектування та розробки інформаційних систем: Збірка навчальних матеріалів./ Ковальчук А.М., Левицький В.Г., Самолюк І.І., Янчук В.М.- Ж.: ЖДТУ, 2009. - 54с. .
2. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод – Спб: БХВ-Петербург, 2004. - 365 с

3. Елманова Н. Введение в OLAP-технологии Microsoft / Наталия Елманова, Алексей Федоров – М.: Диалог-МИФИ, 2002. - 272 с
4. Паклин Н. Бизнес-аналитика. От данных к знаниям / Н. Паклин, В. Орешков – СПб: Питер, 2013. - 704 с.
5. Чубукова И. А. Data Mining/ Чубукова И. А. - М: Бином. Лаборатория знаний, 2008. - 384 с.
6. Ian H. Witten. Data Mining: Practical Machine Learning Tools and Techniques / Ian H. Witten, Eibe Frank and Mark A. Hall — [3rd Edition] — Morgan Kaufmann, 2011. — P. 664. — ISBN 9780123748560

Допоміжна література

1. Jiawei Han, Micheline Kamber, Jian Pei Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems),Morgan Kaufmann,2011.
2. Michael J. A. Berry, Gordon S. Data Mining Techniques: For Marketing, Sales, and Customer Support
3. Ian H. Witten, Eibe Frank, Morgan Kaufmann Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems)

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

1. Язык програмування з документацією, Python 3.x., [Електронний ресурс]. URL: www.python.org
2. Набор бібліотек для Python, Anaconda 3.x., [Електронний ресурс]. URL: www.continuum.io/ (includes Numpy, Matplotlib, sklearn, Ipython Notebook))
3. Набір інструментів Open AI Gym, [Електронний ресурс]. URL: <https://gym.openai.com>
4. Система командного спілкування, Slack, [Електронний ресурс]. URL: <https://slack.com/>
5. Репозиторії, Gitlab / Github / BitBucket, [Електронний ресурс]. URL: <https://gitlab.com/> , <https://github.com/> , <https://bitbucket.org/>