

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра моделювання систем і технологій

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної роботи



Пантелеймонов А.В.

30 серпня 2019р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Інтелектуальний аналіз даних

рівень вищої освіти	перший (бакалаврський)
галузь знань	12 Інформаційні технології
напрямок підготовки	122 Комп'ютерні науки та інформаційні технології
освітня програма	Комп'ютерні науки
спеціалізація	
вид дисципліни	Обов'язкова
факультет	Комп'ютерних наук

2019 / 2020 навчальний рік

Програму рекомендовано до затвердження вченою радою факультету комп'ютерних наук

Протокол від « 27 » червня 2019 року № 2

РОЗРОБНИК ПРОГРАМИ: старший викладач кафедри моделювання систем і технологій Діденко Євген Володимирович.

Програму схвалено на засіданні кафедри моделювання систем і технологій

Протокол від « 30 » травня 2019 року № 15

Завідувач кафедри моделювання систем і технологій



М.В.Ткачук

Програму погоджено методичною радою факультету комп'ютерних наук

Протокол від « 20 » червня 2019 року № 9

Голова методичної комісії



А.Г.Бердніков



ВСТУП

Програма навчальної дисципліни «Інтелектуальний аналіз даних» складена відповідно до освітньо-професійної програми підготовки першого (бакалаврського) рівня спеціальності 122 Комп'ютерні науки та інформаційні технології.

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 уміння 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 – 3

1.4. Загальна кількість годин – 90

1.5. Характеристика навчальної дисципліни	
<u>Нормативна / за вибором</u>	
Денна форма навчання	Денна форма навчання
Рік підготовки	
4-й	4-й
Семестр	
7-й	-й
Лекції	
16 год.	год.
Практичні, семінарські заняття	
год.	год.
Лабораторні заняття	
16 год.	год.
Самостійна робота	
58 год.	год.
В т. ч. Індивідуальні завдання	
20 год.	

1.6. Заплановані результати навчання

Згідно з вимогами освітньо-професійної програми студенти повинні досягти таких результатів 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 анотування текстів. Засоби аналізу текстової інформації.

Задача анотування текстів. Засоби аналізу текстової інформації.

Розділ 3. Кластеризація та веб-майнінг**Тема 7. 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. Области застосування кластерного аналізу.

Тема 8. 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					
	усього	у тому числі				
		л	п	лаб.	інд.	с. р.
1	2	3	4	5	6	7
Розділ 1 Традиційні та сучасні технології інтелектуальної обробки даних.						
Тема 1. Введення в інтелектуальний аналіз даних.	7	2				3
Тема 2. Методи інтелектуального аналізу даних	9	2		2		3
Тема 3. Етапи проведення інтелектуального аналізу даних.	9	2		4		3
Підготовка до контрольної роботи	5					4
Разом за розділом 1	30	6		6		13
Розділ 2. Аналіз текстової інформації – Text Mining.						
Тема 4. Задача та етапи аналізу текстів Text Mining.	12	2				5
Тема 5. Методи класифікації текстових документі	12	2		4		5
Тема 6. Задача анотування текстів. Засоби аналізу текстової інформації.	11	2				5
Разом за розділом 2	35	6		4		15
Розділ 3. Кластеризація та веб-майнінг						
Тема 7. Кластерний аналіз	14	2		4		5
Тема 8. Веб-майнінг	11	2		2		5
Розрахунково-графічна робота 1	10				10	
Розрахунково-графічна робота 2	10				10	
Разом за розділом 3	45	4		6	20	10
Усього годин	90	16		16	20	38

4. Теми семінарських (практичних, лабораторних) занять

№ теми	Назва теми	Кількість годин
Розділ 1 Традиційні та сучасні технології інтелектуальної обробки даних.		
1	«Ознайомлення з інтелектуальним аналізом даних». Основні методи та приклади. <i>«Лабораторна робота №1». Обробка даних експериментів та розрахунок основних статистичних характеристик.</i>	2
2	<i>«Лабораторна робота №2». Розрахунок статистичних характеристик та кластеризація дата-сету.</i>	4
Розділ 2. Аналіз текстової інформації – Text Mining.		
3	<i>«Лабораторна робота №3». Використання наївного байєвського класифікатора для класифікації текстової інформації.</i> Робота з текстовими даними в Python (стеммінг).	4
Розділ 3. Кластеризація та веб-майнінг		
4	<i>«Лабораторна робота №4». Кластеризація методом k-середніх</i> Робота з графічним відображенням в у Python.	4
5	<i>«Лабораторна робота №5». Веб-майнінг.</i> Створення парсеру веб-сторінок та розрахунок Page-rank.	2
Усього годин		16

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	Основні завдання інтелектуального аналізу даних. Опрацювання навчального матеріалу.	3
2	Методи інтелектуального аналізу даних. Опрацювання навчального матеріалу.	3
3	Етапи проведення інтелектуального аналізу даних. Опрацювання навчального матеріалу.	3
4	Підготовка до контрольної роботи.	4
Разом за розділ 1		13
Розділ 2. Аналіз текстової інформації – Text Mining.		
5	Задача та етапи аналізу текстів Text Mining. Опрацювання навчального матеріалу.	5
6	Загальний опис процесу отримання понять з тексту. Методи класифікації текстових документів. Опрацювання навчального матеріалу.	5
7	Представлення текстових документів. Задача анотування текстів. Засоби аналізу текстової інформації. Опрацювання навчального матеріалу.	5
Разом за розділ 2		15
Розділ 3. Кластеризація та веб-майнінг		
8	Кластерний аналіз. Опрацювання навчального матеріалу. Опрацювання навчального матеріалу.	5
9	Веб-майнінг та розрахунок page-rank для маленької мережі веб-сторінок. Опрацювання навчального матеріалу.	5
Разом за розділ 3		10
Усього годин		38

6. Індивідуальні завдання

Контрольна робота
Розрахунково- графічні роботи-2

7. Методи контролю

Протягом навчального семестру проводиться поточний контроль знань, який складається з виконання: 5 лабораторних робіт, контрольної роботи (модульна контрольна робота) та виконання 2 розрахунково-графічних робіт.

Максимальна оцінка за лабораторну роботу – 10 або 5 балів.

Максимальна оцінка за контрольну роботу – 10 балів.

Максимальна оцінка за розрахунково-графічну роботу – 10 балів.

Кількість балів, які студент може отримати протягом навчального семестру складає 60 балів.

За підсумками навчального семестру студент який виконав усі контрольні точки: лабораторні роботи, розрахунково-графічні роботи, контрольну роботу та отримав за підсумками семестру не менше 50 балів – отримує допуск до складання заліку, в іншому випадку студент не допускається до складання заліку.

Залік складається з двох теоретичних та одного практичного питання.

Максимальна оцінка за теоретичне питання – 10 балів.

Максимальна оцінка за практичне питання – 20 балів.

Максимальна оцінка за залік – 40 балів.

8. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання							Залік	Сума		
Розділ 1		Розділ 2		Розділ 3		Контрольна робота			Розрахунково-графічні роботи -1	Розрахунково-графічні роботи-2
1	2	3	4	5	10	10	10	60	40	100
10	5	5	5	5						

Критерії поточної оцінки знань студентів (контрольна робота, крок оцінювання 2 бали)

Кількість балів	Критерії оцінювання
10	Студент демонструє стійкі системні знання та продуктивно їх використовує; вміє вільно використовувати нові інформаційні технології для поповнення власних знань та розв'язування задач; має стійкі навички управління інформаційною системою у нестандартних ситуаціях.
8	Студент демонструє міцні знання, самостійно визначає проміжні цілі власної навчальної діяльності, оцінює нові факти, явища; вміє самостійно знаходити додаткові відомості та використовує їх для реалізації поставлених перед ним навчальних цілей, судження його (її) логічні і достатньо обґрунтовані; має певні навички управління інформаційною системою.
6	Студент демонструє вивчений матеріал у стандартних ситуаціях; пояснює основні процеси, що відбуваються під час роботи інформаційної системи та наводить власні приклади на підтвердження деяких тверджень; вміє виконувати навчальні завдання.
4	Студент знайомий з основними поняттями навчального матеріалу; може самостійно відтворити значну частину навчального матеріалу і робити певні узагальнення; вміє виконати просте навчальне завдання.

Кількість балів	Критерії оцінювання
2	Студент демонструє, значну (більше половини) частину навчального матеріалу може відтворити репродуктивно; може викладача виконати просте навчальне завдання; має елементарні, нестійкі навички необхідні для виконання завдань.
0	Студент демонструє фрагментарні знання при незначному загальному їх обсязі (менше половини навчального матеріалу).

**Критерії поточної оцінки знань студентів
(лабораторна робота за 10 бальною шкалою, крок оцінювання 2 бали)**

Кількість балів	Критерії оцінювання
10	Оцінюється робота студента, який вільно володіє теоретичним матеріалом необхідним для виконання лабораторної роботи, оперує поняттями та категоріями, вміє встановити зв'язок між теоретичною базою та практикою, залучає до відповіді самостійно опрацьовану літературу. Відповідає на додаткові запитання, наводить аргументи на підтвердження власних думок, здійснює аналіз та робить висновки.
8	Оцінюється робота студента, який достатньо повно володіє теоретичним матеріалом необхідним для виконання лабораторної роботи, орієнтується в ньому, вміє встановити зв'язок між теоретичною базою та практикою. Відповідь з незначною кількістю помилок, відповідає на додаткові запитання, але не має аргументованої думки, висновки не повні.
6	Оцінюється робота студента, який виконав проект який містить значну кількість недоліків і помилок, неповне висвітлення змісту питань володіє матеріалом і частково відповідає на додаткові питання, недостатньо відповідає на питання, не може зробити аргументовані висновки. Презентація частково відповідає наступним вимогам: послідовність, інформативність матеріалу, цілісність, відеоряд.
4	Проект виконано не правильно, студень показує істотне незрозуміння проблеми, в роботі не виконано всі завдання; або студент не здатен захистити результати поданої лабораторної роботи
2	Проект відзначається фрагментарністю, студень показує істотне незрозуміння проблеми, в роботі не виконано всі завдання; однак студент володіє умінням здійснювати первинну обробку навчальної інформації без подальшого її аналізу
0	Оцінюється робота студента, який не виконав лабораторної роботи

**Критерії поточної оцінки знань студентів
(лабораторна робота за 5 бальною шкалою, крок оцінювання 1 балів)**

Кількість балів	Критерії оцінювання
5	Оцінюється робота студента, який вільно володіє теоретичним матеріалом необхідним для виконання лабораторної роботи, оперує поняттями та категоріями, вміє встановити зв'язок між теоретичною базою та практикою, залучає до відповіді самостійно опрацьовану літературу. Відповідає на додаткові запитання, наводить аргументи на підтвердження власних думок, здійснює аналіз та робить висновки.
4	Оцінюється робота студента, який достатньо повно володіє теоретичним матеріалом необхідним для виконання лабораторної роботи, орієнтується в ньому, вміє встановити зв'язок між теоретичною базою та практикою. Відповідь з незначною кількістю помилок, відповідає на додаткові запитання, але не має аргументованої думки, висновки не повні.
3	Оцінюється робота студента, який виконав проект який містить значну кількість недоліків і помилок, неповне висвітлення змісту питань володіє матеріалом і частково відповідає на додаткові питання, недостатньо відповідає на питання, не може зробити аргументовані висновки. Презентація частково відповідає наступним вимогам:

Кількість балів	Критерії оцінювання
	послідовність, інформативність матеріалу, цілісність, відеоряд.
2	Проект виконано не правильно, студень показує істотне незрозуміння проблеми, в роботі не виконано всі завдання; або студент не здатен захистити результати поданої лабораторної роботи
1	Проект відзначається фрагментарністю, студень показує істотне незрозуміння проблеми, в роботі не виконано всі завдання; однак студент володіє умінням здійснювати первинну обробку навчальної інформації без подальшого її аналізу
0	Оцінюється робота студента, який не виконав лабораторної роботи

**Критерії поточної оцінки знань студентів
(індивідуальне завдання, крок оцінювання 2 бали)**

Критеріями оцінювання індивідуальних завдань студента є:

1. самостійність виконання завдання;
2. правильність, точність, оптимальність реалізації поставленого завдання;
3. дизайнерське рішення (єдине стильове рішення, композиція, врахування психологічних особливостей сприйняття інформації людиною);
4. завершеність завдання.
5. вміння захищати результати проведеного дослідження.

Кількість балів	Критерії оцінювання
10	Оцінюється робота студента, який вільно володіє матеріалом розробленого проекту, творче його осмислив, оперує поняттями та категоріями, вміє встановити зв'язок між теоретичною базою та практикою, залучає до відповіді самостійно опрацьовану літературу. Відповідає на додаткові запитання, наводить аргументи на підтвердження власних думок, здійснює аналіз та робить висновки. Презентація відповідає наступним вимогам: послідовність, інформативність матеріалу, цілісність, відеоряд.
8	Оцінюється робота студента, який достатньо повно володіє матеріалом розробленого проекту, орієнтується в ньому, оперує необхідним колом понять та категорій щодо проекту, вміє встановити зв'язок між теоретичною базою та практикою. Відповідь з незначною кількістю помилок, відповідає на додаткові запитання, але не має аргументованої думки, висновки не повні. Презентація частково відповідає наступним вимогам: послідовність, інформативність матеріалу, цілісність, відеоряд.
6	Оцінюється робота студента, який виконав проект який містить значну кількість недоліків і помилок, неповне висвітлення змісту питань володіє матеріалом і частково відповідає на додаткові питання, недостатньо відповідає на питання, не може зробити аргументовані висновки. Презентація частково відповідає наступним вимогам: послідовність, інформативність матеріалу, цілісність, відеоряд.
4	Проект виконано не правильно, студень показує істотне незрозуміння проблеми, в роботі не виконано всі завдання; або студент не здатен захистити результати поданого проекту
2	Проект відзначається фрагментарністю, студень показує істотне незрозуміння проблеми, в роботі не виконано всі завдання; однак студент володіє умінням здійснювати первинну обробку навчальної інформації без подальшого її аналізу
0	Оцінюється робота студента, який не виконав індивідуальне завдання

**Критерії підсумкового оцінювання знань студентів
(залік, теоретичне питання, крок оцінювання 2 бали)**

Оцінка в балах	Пояснення
10	Теоретичний зміст курсу освоєний цілком, необхідні практичні навички роботи з освоєним матеріалом сформовані, всі навчальні завдання, які передбачені програмою навчання виконані в повному обсязі, відмінна робота без помилок або з однією незначною помилкою.
8	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.
6	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
2-4	Теоретичний зміст курсу частково освоєно (менше половини навчального матеріалу), 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.
0	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

**Критерії підсумкового оцінювання знань студентів
(залік, практичне питання, крок оцінювання 5 балів)**

Оцінка в балах	Пояснення
20-16	Студент правильно розв'язав практичне завдання (задачу), вміє пояснити методику розв'язання та зміст застосовуваного методу знань і алгоритму. Вміє аргументувати свої думки.
15-11	Студент показує знання методики розв'язання практичного завдання (задачі) та змісту застосовуваного методу знань і алгоритму. Проте допущені окремі незначні помилки у розрахунку.
10-6	Студент показує знання методики розв'язання практичного завдання (задачі) та змісту застосовуваного методу знань і алгоритму. Проте допущені помилки у розрахунку, не дають можливості зробити правильні висновки.
5-0	Студент неправильно розв'язав практичне завдання (задачу).), показав незадовільне знання

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

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/>