

Черкаський державний технологічний університет  
Факультет інформаційних технологій і систем

«ЗАТВЕРДЖУЮ»

Голова вченої ради ФІТІС



І.Б. Трегубенко

(підпис)

Протокол № 5

« 17 » лютого

20 20 р.

## СИЛАБУС

навчальної дисципліни

«Інтелектуальний аналіз даних»

Шифр за ОПП – ВППБ9

Освітній рівень – бакалаврський

Галузь знань – 12 «Інформаційні технології»

Спеціальність – 126 «Інформаційні системи та технології»

Освітня програма – «Web-технології, Web-дизайн»

2019-2020 навчальний рік

Силабус навчальної дисципліни «Інтелектуальний аналіз даних» підготовки здобувачів освітнього рівня «бакалавр» за спеціальністю 126 «Інформаційні системи та технології», освітня програма «Web-технології, Web-дизайн» – 13 стор.

Силабус складений на основі програми навчальної дисципліни «Інтелектуальний аналіз даних», шифр (за ОПП) – ВППБ9.

Розробник силабусу:

Єгорова О.В., к.т.н., доцент кафедри інформаційних технологій проектування

(прізвище та ініціали, наук. ст., вчене зв., посада НПП кафедри, що розробив силабус)

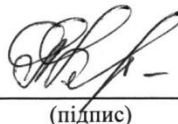
Силабус затверджений на засіданні кафедри інформаційних технологій проектування  
(найменування кафедри)

Протокол № 8 від « 10 » січня 20 20 р.

Обговорено та рекомендовано до затвердження методичною комісією факультету інформаційних технологій і систем

« 14 » лютого 20 20 р., протокол № 4

Голова методичної комісії  
факультету інформаційних технологій і систем



(підпис)

(А.Р. Карапетян)

## 1 ІНФОРМАЦІЯ ПРО ВИКЛАДАЧА

Прізвище, ім'я, по батькові	Єгорова Ольга В'ячеславівна
Науковий ступінь	кандидат технічних наук
Наукове звання	
Посада	доцент кафедри інформаційних технологій проектування
Місце роботи	Черкаський державний технологічний університет, Кафедра інформаційних технологій проектування
Адреса кафедри	к.603, бул. Шевченка, 460, м. Черкаси, 18006
Контактний телефон	+38 099 648 75 66
Профайл викладача	<a href="https://chdtu.edu.ua/fitis/kitp/staff/item/7613-yehorova-olha-v-iacheslavivna">https://chdtu.edu.ua/fitis/kitp/staff/item/7613-yehorova-olha-v-iacheslavivna</a>
e-mail	<a href="mailto:yegorovaov@gmail.com">yegorovaov@gmail.com</a>
Профайл дисципліни	<a href="http://fitis.moodle.chdtu.edu.ua/course/view.php?id=536">http://fitis.moodle.chdtu.edu.ua/course/view.php?id=536</a>
Розклад консультацій	

## 2 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Галузь знань, спеціальність, освітня програма, освітній рівень	Загальні характеристики		Навчальне навантаження з дисципліни	
			денна форма навчання	заочна форма навчання
<u>Галузь знань</u> 12 «Інформаційні технології»	вибіркова		Курс підготовки	
			4	-
<u>Спеціальність</u> 126 «Інформаційні системи та технології»	Загальна кількість кредитів ЄКТС	4	Семестр підготовки	
	Загальна кількість годин	120	7	-
<u>Освітня програма</u> «Web-технології, Web-дизайн»	Кількість аудиторних годин	48	Лекції	
			16 год	-
	Кількість годин самостійної роботи	72	Практичні, семінарські	
			-	-
<u>Освітній рівень</u> бакалаврський	Мова навчання – українська		Лабораторні	
			32 год.	-
			Самостійна робота	
			72 год.	-
			Форма підсумкового контролю	
			залік	-

### 3 МЕТА І ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

<b>Мета викладання дисципліни</b>	Теоретична та практична підготовка здобувачів вищої освіти у напрямку розв'язання задач обробки великих масивів інформації, проектування інформаційного забезпечення інформаційних систем та розробки сценаріїв можливих дій в умовах невизначеності з використанням інтелектуальних методів обчислень.
<b>Завдання вивчення дисципліни</b>	Забезпечити розуміння і засвоєння здобувачами вищої освіти технологій інтелектуального аналізу даних, підготовка до їх вибору, реалізації і використання при вирішенні прикладних задач, ознайомлення зі станом та перспективами розвитку інтелектуальних методів обчислень як одного з напрямів штучного інтелекту.

### 4 РЕЗУЛЬТАТИ НАВЧАННЯ

№ з/п	Результати навчання
1	ПР 6. Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.
2	ПР 7. Обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу інформаційних систем та технологій.
3	Демонструвати знання методів та технологій Data Mining.
4	Демонструвати здатність розробки алгоритмів інтелектуального аналізу даних.
5	Демонструвати здатність розробки інтелектуальних підсистем в складі web-орієнтованих систем.

### 5 ПРЕРЕКВІЗИТИ

«Вища математика», «Теорія ймовірності та математична статистика», «Алгоритмізація та програмування», «Бази даних та знань», «Архітектура інформаційних систем і технологій».

### 6 ПОСТРЕКВІЗИТИ

«Безпека інформаційних і комунікаційних систем», «Професійний практикум», «Технології та системи підтримки прийняття рішень», «Прикладні інтелектуальні системи обробки даних», «Електронна комерція та електронний бізнес».

## 7 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

<b>Змістовий модуль № 1</b> <i>Методи інтелектуального аналізу даних</i>
<b>Тема 1. Основні поняття Data Mining.</b>
Поняття даних. Особливості обробки даних. Типи даних. Задачі Data Mining. Етапи Data Mining. Моделі Data Mining. Методи Data Mining. Стандарти Data Mining. Інструменти Data Mining.
<b>Тема 2. Препроцесінг інформації.</b>
Нормалізація і стандартизація вихідних значень. Аналітико-евристичні алгоритми визначення інформативних ознак. Алгоритм «вибілювання» входів. Нейромережеве визначення значущих факторів. Методика «box-counting».
<b>Тема 3. Кореляційний і регресійний аналіз.</b>
Поняття зв'язку між величинами. Кореляційний аналіз. Загальні положення регресійного аналізу. Метод найменших квадратів. Парна лінійна регресія. Множинна лінійна регресія. Тестування і усунення мультиколінеарності. Тестування і усунення гетероскедастичності. Автокореляція: причини і наслідки. Множинна нелінійна регресія. Алгоритм Брандона.
<b>Тема 4. Методи класифікації.</b>
Постановка задачі класифікації даних. Правила класифікації. Методи побудови правил класифікації. Точність класифікації: оцінка рівня помилок. Алгоритм 1-rule. Наївний баєсівський класифікатор. Поняття дерева рішень. Методика «розділай та володарюй». Алгоритм ID3. Алгоритм C4.5. Алгоритм CART. Метод опорних векторів. Метод «найближчого сусіда». Застосування штучних нейронних мереж для класифікації.
<b>Тема 5. Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій.</b>
Основні поняття дискримінантного аналізу. Канонічні дискримінантні функції. Класифікаційні функції. Врахування апіорних ймовірностей. Класифікація за допомогою канонічних дискримінантних функцій. Класифікаційна матриця. Обґрунтування за допомогою розбиття вибірки.
<b>Тема 6. Кластерний аналіз. Ієрархічна та секційна кластеризації.</b>
Постановка задачі кластеризації. Метрики визначення відстані між об'єктами. Ієрархічні агломеративні методи кластеризації. Ієрархічні дивізимні методи кластеризації. Неієрархічні методи кластеризації: процедура Мак-Кіна, метод k-середніх, сітчасті методи кластеризації, алгоритм fuzzy c-means, алгоритм Гюстафсона-Кеселля.
<b>Тема 7. Пошук асоціативних правил.</b>
Постановка задачі пошуку асоціативних правил. Секвенційний аналіз. Різновиди задач пошуку асоціативних правил. Метод Apriori. Побудова FP-дерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.
<b>Тема 8. Статистична обробка часових рядів і прогнозування.</b>
Аналіз часових рядів з допомогою Python. Експоненціальне згладжування. Модель Хольта-Вінтерса. Крос-валідація на часових рядах. Лінійні та

нелінійні моделі на часових рядах. Вилучення ознак. Алгоритм XGBoost.
<b>Змістовий модуль № 2</b> <i>Застосування інтелектуальних технологій</i>
<b>Тема 1. Візуалізація даних.</b>
Постановка задачі візуалізації даних. Етапи візуального аналізу даних. Методи геометричних перетворень. Методи, орієнтовані на пікселі. Ієрархічні образи. OLAP-аналіз. Засоби візуалізації даних, призначені для оцінки якості моделей: матриця класифікації, діаграма розсіювання, ретропрогноз, графіки контролю процесу навчання.
<b>Тема 2. Аналіз текстової інформації Text Mining.</b>
Постановка задачі аналізу текстової інформації. Етапи аналізу текстів. Добування ключових понять із тексту. Класифікація текстових документів. Методи кластеризації текстових документів. Анотування текстів. Засоби аналізу текстової інформації.
<b>Тема 3. Добування знань із Web – Web Mining.</b>
Основні поняття Web Mining. Методи добування web-контенту. Добування web-структур. Дослідження використання web-ресурсів.
<b>Тема 4. Засоби аналізу процесів Process Mining.</b>
Автоматизація виконання бізнесів-процесів. Технологія аналізу процесів. Методи Process Mining. Бібліотека алгоритмів Process Mining – ProM.
<b>Тема 5. Розподілений аналіз даних.</b>
Системи мобільних агентів. Використання мобільних агентів для аналізу даних. Побудов систем аналізу розподілених даних.

## 8 ТЕМАТИЧНИЙ ПЛАН НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

№ теми	Назва модулів і тем	Форми організації навчання, кількість годин						Література, інформаційні ресурси
		Денна форма			Заочна форма			
		Лекції	Лабораторні роботи	Самостійна робота	Лекції	Лабораторні роботи	Самостійна робота	
Змістовий модуль № 1. Методи інтелектуального аналізу даних								
1	Основні поняття Data Mining	1	0	3				1,2
2	Препроцесінг інформації	1	4	3				2,3
3	Кореляційний і регресійний аналіз	1	8	3				1,3
4	Методи класифікації	1	4	3				2,4
5	Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій	1	0	4				4,5

6	Кластерний аналіз. Ієрархічна та секційна кластеризації.	1	4	4				1,3
7	Пошук асоціативних правил	1	4	4				2,3
8	Статистична обробка часових рядів і прогнозування	1	0	4				3,4
<b>Змістовий модуль № 2. Застосування інтелектуальних технологій</b>								
9	Візуалізація даних	1	0	8				1,5
10	Аналіз текстової інформації Text Mining	1	4	9				1,5
11	Добування знань із Web – Web Mining	2	4	9				1,5
12	Засоби аналізу процесів Process Mining	2	0	9				1,5
13	Розподілений аналіз даних	2	0	9				1,5
	<b>Разом</b>	16	32	72				

## 9 ЛАБОРАТОРНІ РОБОТИ

№ з/п	Назва лабораторної роботи	Кількість годин	
		Денна	Заочна
1	Препроцесінг інформації	4	
2	Однофакторний кореляційно-регресійний аналіз	4	
3	Багатофакторний кореляційно-регресійний аналіз	4	
4	Методи кластеризації	4	
5	Методи побудови дерев рішень	4	
6	Методи побудови асоціативних правил	4	
7	Text Mining	8	

### **МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ**

Методичні рекомендації до лабораторних робіт

[https://drive.google.com/open?id=1-uKq5h\\_amO6w4yLyaek\\_VeeCQfwjWCdy](https://drive.google.com/open?id=1-uKq5h_amO6w4yLyaek_VeeCQfwjWCdy)

## 10 САМОСТІЙНА РОБОТА

**10.1 Рекомендації до самостійної роботи здобувачів вищої освіти денної форми навчання.**

### **МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ**

Методичні рекомендації до самостійної роботи

<https://drive.google.com/open?id=1NYJRgDYwThElsj4T6tumb7YLNjN3johY>

# 11 СИСТЕМА ОЦІНЮВАННЯ НАВЧАЛЬНИХ ДОСЯГНЕНЬ

## 11.1 МЕТОДИ КОНТРОЛЮ

В організації навчального процесу застосовуються контрольні заходи у формі вхідного, поточного, модульного, рейтингового і підсумкового контролю.

*Вхідний контроль* проводиться перед вивченням нового курсу з метою визначення рівня підготовки здобувачів вищої освіти з дисциплін, які забезпечують цей курс. За результатами вхідного контролю розробляються заходи з надання індивідуальної допомоги здобувачам вищої освіти, коригування навчального процесу з відповідного курсу.

*Поточний контроль* здійснюється під час проведення лекцій та лабораторних занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретних видів навчальної діяльності.

*Модульний контроль* успішності здобувачів вищої освіти здійснюється для перевірки рівня засвоєння навчального матеріалу в кінці кожного навчального модуля.

*Рейтинговий контроль* є інструментом комплексного оцінювання якості навчальної роботи здобувача вищої освіти з усіх кредитних модулів на певному етапі навчання. Рейтинговий контроль успішності здобувачів вищої освіти проводиться на 8-9 навчальних тижнях.

*Семестровий контроль* з дисципліни проводиться відповідно до навчального плану у вигляді заліку в терміни, встановлені графіком навчального процесу, та в обсязі навчального матеріалу, визначеному робочою програмою дисципліни.

*Залік* – це вид підсумкового контролю, за якого засвоєння здобувачем вищої освіти навчального матеріалу з дисципліни оцінюється на підставі результатів поточного, проміжного контролів (тестування, поточного опитування, виконання індивідуальних завдань та певних видів робіт на лабораторних заняттях) протягом семестру і модульного контролю.

## 11.2 ПИТАННЯ ДО ЗАЛІКУ

1. Поняття даних. Особливості обробки даних
2. Типи даних
3. Задачі Data Mining
4. Етапи Data Mining. Моделі Data Mining
5. Методи Data Mining
6. Стандарти Data Mining
7. Інструменти Data Mining
8. Нормалізація і стандартизація вихідних значень
9. Аналітико-евристичні алгоритми визначення інформативних ознак
10. Алгоритм «вибілювання» входів
11. Нейромережеве визначення значущих факторів. Методика «box-counting»
12. Поняття зв'язку між величинами. Кореляційний аналіз
13. Загальні положення регресійного аналізу



14. Метод найменших квадратів
15. Парна лінійна регресія
16. Множинна лінійна регресія
17. Тестування і усунення мультиколінеарності
18. Тестування і усунення гетероскедастичності
19. Автокореляція: причини і наслідки
20. Множинна нелінійна регресія
21. Алгоритм Брандона
22. Постановка задачі класифікації даних
23. Правила класифікації. Методи побудови правил класифікації
24. Точність класифікації: оцінка рівня помилок
25. Алгоритм 1-rule
26. Наївний баєсівський класифікатор
27. Поняття дерева рішень. Методика «розділяй та володарюй»
28. Алгоритм ID3
29. Алгоритм C4.5
30. Алгоритм CART
31. Метод опорних векторів
32. Метод «найближчого сусіда»
33. Застосування штучних нейронних мереж для класифікації
34. Основні поняття дискримінантного аналізу
35. Канонічні дискримінантні функції
36. Класифікаційні функції
37. Класифікація за допомогою канонічних дискримінантних функцій
38. Класифікаційна матриці
39. Постановка задачі кластеризації
40. Метрики визначення відстані між об'єктами
41. Ієрархічні агломеративні методи кластеризації
42. Ієрархічні дивізімні методи кластеризації
43. Процедура Мак-Кина
44. Метод k-середніх
45. Сітчасті методи кластеризації
46. Алгоритм fuzzy c-means
47. Алгоритм Гюстафсона-Кеселля
48. Постановка задачі пошуку асоціативних правил
49. Секвенційний аналіз
50. Різновиди задач пошуку асоціативних правил
51. Метод Apriori. Побудова FP-дерев пошуку шаблонів даних
52. Min-max асоціації у базах даних
53. Побудова hash-дерев
54. Аналіз часових рядів з допомогою Python
55. Експоненціальне згладжування
56. Модель Хольта-Вінтерса
57. Крос-валідація на часових рядах

58. Лінійні та нелінійні моделі на часових рядах. Вилучення ознак. Алгоритм XGBoost
59. Постановка задачі візуалізації даних
60. Етапи візуального аналізу даних
61. Методи геометричних перетворень
62. Методи, орієнтовані на пікселі
63. Ієрархічні образи. OLAP-аналіз
64. Засоби візуалізації даних, призначені для оцінки якості моделей: матриця класифікації, діаграма розсіювання, ретропрогноз, графіки контролю процесу навчання
65. Постановка задачі аналізу текстової інформації
66. Етапи аналізу текстів
67. Добування ключових понять із тексту
68. Класифікація текстових документів
69. Методи кластеризації текстових документів
70. Анотування текстів
71. Засоби аналізу текстової інформації
72. Основні поняття Web Mining
73. Методи добування web-контенту
74. Добування web-структур
75. Дослідження використання web-ресурсів
76. Автоматизація виконання бізнесів-процесів
77. Технологія аналізу процесів
78. Методи Process Mining
79. Бібліотека алгоритмів Process Mining – ProM
80. Системи мобільних агентів
81. Використання мобільних агентів для аналізу даних
82. Побудов систем аналізу розподілених даних

### 11.3 КРИТЕРІЇ ОЦІНЮВАННЯ

#### *ДЕННА ФОРМА*

Для здобувачів денної форми навчання	
Вид навчальної роботи	Кількість балів максимум
<b><u>Постійна частина</u></b>	
ЗМІСТОВИЙ МОДУЛЬ № 1 «Методи інтелектуального аналізу даних» – 60 годин	
Захист лабораторної роботи № 1	5
Захист лабораторної роботи № 2	5
Захист лабораторної роботи № 3	5
Захист лабораторної роботи № 4	5
Захист лабораторної роботи № 5	5
Захист лабораторної роботи № 6	5

Модульна контрольна робота № 1	20
<i>Всього за змістовим модулем № 1</i>	50
<b>ЗМІСТОВИЙ МОДУЛЬ № 2</b> «Застосування інтелектуальних технологій» – 60 годин	
Захист лабораторної роботи № 7	15
Захист лабораторної роботи № 8	15
Модульна контрольна робота № 2	20
<i>Всього за змістовим модулем № 2</i>	50
<b><u>Додаткова частина</u></b>	
Підготовка та захист реферату за індивідуальною темою	20
Участь у Днях студентської науки	20
Участь у науковій конференції чи семінарі за темою дисципліни	20
Оформлення наочного стенда за індивідуальною темою	20
<b><u>Штрафна частина</u></b>	
Пропуск одного заняття без поважної причини	-5
Несвоєчасний захист звіту з лабораторної роботи	-5
<b>ПІДСУМКОВА СЕМЕСТРОВА ОЦІНКА</b>	<b>100</b>

### ЗАОЧНА ФОРМА

Вид навчальної роботи	Кількість балів максимум
<i>Контрольна робота з дисципліни (відповідно до отриманого завдання)</i>	60
Залік	40
Разом	100

## 12 РЕКОМЕНДОВАНА ЛІТЕРАТУРА

### Основна

1. Гладун А. Я., Рагушина Ю. В. Data Mining: пошук знань в даних. К.: ТОВ «ВД «АДЕФ Україна», 2016. 452 с.
2. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань : навчальний посібник. Львів: «Магнолія 2006», 2015. 276 с.
3. Замятин А. В. Интеллектуальный анализ данных : учеб. пособие. Томск: Издательский Дом Томского государственного университета, 2016. 120 с.
4. Sarkar D., Bali R., Sharma T. Practical Machine Learning with Python. Apress, 2018. 545 p. URL: [www.apress.com/978-1-4842-3206-4](http://www.apress.com/978-1-4842-3206-4).
5. Zaki M. J., Meira W. Jr. Data mining and analysis : Fundamental Concepts and Algorithms. New York : Cambridge University Press, 2014. 604 p.

### Допоміжна

6. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для . Запоріжжя: КПУ, 2011. 268 с.

7. Анализ данных и процессов : учеб. пособие / А. А. Барсегян, М. С. Куприянов, Холод И.И. и др.; 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с.
8. Снитюк В. Є. Прогнозування. Моделі. Методи. Алгоритми : навчальний посібник. К.: Маклаут, 2008. 364 с.
9. Паклин Н. Б., Орешков В. И. Бизнес-аналитика : от данных к знаниям. СПб.: Питер, 2009. 624 с.
10. Литвин В.В., Пасічник В.В., Яцишин Ю.В. Інтелектуальні системи : підручник. Львів: Новий світ – 2000, 2009. 406с.
11. Зайченко Ю. П. Нечеткие модели и методы в интеллектуальных системах : учебное пособие. К.: Слово, 2008. 344с.
12. Математичне забезпечення інформаційно-керуючих систем: підручник / А. М. Гуржій, З. В. Дудар, В. М. Левикін, Б. В. Шамша. Х. : Компанія Сміт, 2006. 448 с.

### **13 ІНФОРМАЦІЙНІ РЕСУРСИ**

1. Wiki-портал <http://www.machinelearning.ru>
2. Электронный журнал «Нечеткие системы и мягкие вычисления». <http://fuzzy.tversu.ru/>.
3. Evolutionary Complexity Research Group (EPlex): <http://eplex.cs.ucf.edu/>
4. Классификация <http://rcs.chemometrics.ru/Tutorials/classification.htm#Ch3.1>

### **14 ПЕРЕЛІК НОРМАТИВНИХ ДОКУМЕНТІВ**

1. ДСТУ 3008:2015. Звіти у сфері науки і техніки. Структура та правила оформлювання / Нац. Стандарт України. Вид. офіц. [На заміну ДСТУ 3008-95; чинний від 2017-07-01]. Київ : ДП «УкрНДНЦ», 2016. 31 с. (Інформація та документація).
2. ДСТУ 8302:2015. Бібліографічне посилання. Загальні положення та правила складання / Нац. Стандарт України. Вид. офіц. [Уведено вперше; чинний від 2016-07-01]. Київ : ДП «УкрНДНЦ», 2016. 17 с. (Інформація та документація). – З внесеними правками.

### **15 ПОЛІТИКА ДИСЦИПЛІНИ**

Для успішного вивчення дисципліни та проходження контрольних заходів здобувачі вищої освіти зобов'язані:

- не запізнюватися на заняття;
- не пропускати заняття (у разі хвороби надати довідку або її ксерокопію);
- своєчасно і самостійно виконувати всі передбачені програмою завдання до лабораторних робіт;
- брати очну участь у контрольних заходах;
- оволодіти навчальним матеріалом для самостійного вивчення з дисципліни у вільний від обов'язкових занять час;

- підтримувати зворотній зв'язок з викладачем на всіх етапах вивчення дисципліни;
- дотримуватися академічної доброчесності.