

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

"ЗАТВЕРДЖУЮ"

Заступник керівника

(проректор з науково-педагогічної роботи)



М.В. Афанасьєв

### Аналітика великих даних

робоча програма навчальної дисципліни

Галузь знань	12 "Інформаційні технології"
Спеціальність	126 "Інформаційні системи та технології"
Освітній рівень	другий (магістерський)
Освітня програма	Інформаційні системи та технології

Вид дисципліни  
Мова викладання,  
навчання та оцінювання

вибіркова  
українська

Завідувач кафедри економічної кібернетики

Гур'янова Л.С.

Харків  
ХНЕУ ім. С. Кузнеця  
2019

*Лист...*  
*04.12*

Затверджено на засіданні кафедри економічної кібернетики  
Протокол № 1 від 27.08.2019 р.

Розробник:

Чаговець Любов Олексіївна, к.е.н., доц. кафедри економічної кібернетики

**Лист оновлення та перезатвердження  
робочої програми навчальної дисципліни**

Навчальний рік	Дата засідання кафедри – розробника РПНД	Номер протоколу	Підпис завідувача кафедри

## 1. Вступ

**Анотація навчальної дисципліни.** Дисципліна "Аналітика великих даних" є однією з вибіркових дисциплін професійного циклу для підготовки магістрів за спеціальністю "Інформаційні системи та технології".

**Предметом** дисципліни є сукупність методів і моделей порівняльного аналізу великих даних. **Об'єктом** вивчення дисципліни "Аналітика великих даних" є інформаційні екосистеми різного рівня ієрархії. Наукову основу дисципліни складають теоретичні методи і моделі, математичний апарат, сучасні концепції, які визначають різні підходи до порівняльного аналізу великих даних. У процесі навчання студенти отримують необхідні знання під час лекційних занять і виконуючи лабораторні роботи. Також велике значення в процесі вивчення та закріплення знань має самостійна робота студентів і виконання індивідуальних завдань.

Програма навчальної дисципліни складається з таких змістових модулів:

1. Методологічні основи аналітики великих даних.
2. Методи предиктивної аналітики великих даних.

**Мета навчальної дисципліни:** вивчення теоретичних основ і можливостей практичного застосування методів аналізу великих даних для дослідження процесів та систем різного призначення.

Курс	<b>1М</b>	
Семестр	<b>1</b>	
Кількість кредитів ECTS	<b>5</b>	
Аудиторні навчальні заняття	<b>лекції</b>	<b>12</b>
	<b>лабораторні</b>	<b>28</b>
Самостійна робота		<b>110</b>
Форма підсумкового контролю	<b>іспит</b>	

### Структурно-логічна схема вивчення навчальної дисципліни:

<b>Попередні дисципліни</b>	<b>Наступні дисципліни</b>
Вища математика	Комплексний тренінг
Системи БІ	Дипломна робота
Бізнес-моделювання	Стратегія ІС
Бізнес-аналіз	
Системний аналіз та проектування інформаційних систем	

## 2. Компетентності та результати навчання за дисципліною:

Компетентності	Результати навчання
<p>ЗК6. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.</p> <p>ЗК7. Здатність бути критичним і самокритичним.</p> <p>ЗК9. Вміння виявляти, ставити та вирішувати проблеми.</p>	<p>ПР2. Здатність робити презентації за професійною тематикою різного обсягу та складності рідною та іноземною мовами як для фахівців, так і для нефахівців.</p> <p>ПР3. Здатність ефективно працювати в групі, в тому числі і на лідерських позиціях з метою вирішення різноманітних дослідницьких та практичних завдань.</p> <p>ПР4. Здатність навчати інших та самонавчатися за різними аспектами професійної діяльності з метою підвищення рівня професійних та загальних компетентностей.</p>
<p>СК1. Здатність до управління на основі планів, гібридних і гнучких підходів до розробки.</p>	<p>ПР1. Здатність формулювати та вирішувати дослідницьке завдання, для його вирішення збирати, оброблювати та систематизувати інформацію та формулювати висновки.</p>
<p>СК4. Здатність до вибору відповідних технологій управління даними на основі вимог предметної області.</p>	<p>ПР5. Здатність до автономної роботи для вирішення конкретних професійних та дослідницьких завдань.</p>
<p>СК5. Здатність до вибору і використання відповідних методів аналітики.</p>	<p>ПР7. Здатність застосовувати технології моделювання при розв'язанні типових задач, проектуванні та використанні ICT.</p>
<p>СК6. Володіння навичками інтеграції і підготовки даних з різних джерел для цілей аналітики.</p>	<p>ПР8. Здатність проводити аналіз предметної області, технологій та інструментів за допомогою різноманітних методів та підходів.</p>
<p>СК7. Здатність приймати участь в стратегічному плануванні ІС.</p>	<p>ПР9. Здатність використовувати новітні технології ICT з метою їх запровадження у професійну діяльність, аналізувати та відстежувати якість їх впровадження.</p>
<p>СК13. Здатність до розуміння того як застосовувати творчі підходи до вирішення проблем, які пов'язані з технологіями.</p>	<p>ПР10. Здатність до генерування нових ідей в області ICT та оцінки їх перспективності і ризику впровадження.</p>
<p>СК16. Здатність до моніторингу нових технологій для розуміння їх потенціалу для підтримки предметної області.</p>	<p>ПР12. Здатність демонструвати знання сучасного рівня та новітніх технологій ICT, а також впроваджувати їх у професійну діяльність.</p>

### **3. Програма навчальної дисципліни**

#### **Змістовий модуль 1.**

##### **Методологічні основи аналітики великих даних**

###### **Тема 1. Базові поняття аналізу великих даних**

- 1.3. Поняття великих даних. Особливості обробки та аналізу великих даних.
- 1.4. Архітектура великих даних. Фреймворки та екосистема обробки великих даних.
- 1.5. Модель розподілених обчислень MapReduce.
- 1.6. Етапи моделювання та дослідження великих даних. Функціональні зв'язки аналітики великих даних.

###### **Тема 2. Розвідувальний аналіз даних.**

- 2.1. Розподіл даних та вибірок. Сутність і причини виникнення помилок в даних.
- 2.2. Методи дослідження похибок в даних.
- 2.3. Методи бутстрепа.

###### **Тема 3. Методи машинного навчання. Виявлення логічних закономірностей у великих даних**

- 3.1. Методи статистичного машинного навчання.
- 3.2. Деревоподібні моделі пошуку логічних закономірностей. Алгоритм CART.
- 3.3. Ансамблеве навчання та випадкові ліси.
- 3.4. Моделі Байєса для моделювання великих даних.
- 3.5. Мережні графи та моделі спільноти. Візуалізація мережного графу.
- 3.6. Методи нейромережного моделювання та глибинного навчання.

#### **Змістовий модуль 2**

##### **Методи предиктивної аналітики великих даних**

###### **Тема 4. Основи предиктивного моделювання великих даних.**

- 4.1. Особливості предиктивного моделювання.
- 4.2. Поняття і класифікація предиктивних моделей
- 4.3. Критерії якості предиктивної моделі.

###### **Тема 5. Множинні регресії та прогнозування.**

- 5.1. Сутність та процедури побудови багатофакторної регресії.
- 5.2. Методи оцінки корельованості предикторів.
- 5.3. Особливості обробки даних та побудова моделей з нестандартними помилками.
- 5.4. Екстраполяція на основі регресії.
- 5.3. Нелінійна регресія. Вибір і порівняння регресійних моделей

## **Тема 6. Методи аналізу одновимірних часових рядів.**

- 6.1. Поняття та види часових рядів.
- 6.2. Моделі аналізу тренду часового ряду та кривих зростання.
- 6.3. Моделі декомпозиції динамічних рядів.

### **Теми лабораторних занять**

*Лабораторна робота 1.* Оцінка параметрів розподілу випадкових величин.

*Лабораторна робота 2.* Розвідувальний аналіз даних.

*Лабораторна робота 3.* Методи і моделі супервайзерного навчання.

*Лабораторна робота 4.* Побудова та використання методів дерев класифікацій.

*Лабораторна робота 5.* Моделі наївного байесовського класифікатора.

*Лабораторна робота 6.* Побудова та аналіз простої предиктивної моделі.

*Лабораторна робота 7.* Побудова та аналіз множинної лінійної моделі.

*Лабораторна робота 8.* Побудова та аналіз моделей динаміки.

## **4. Порядок оцінювання результатів навчання**

Система оцінювання сформованих компетентностей у студентів враховує види занять, які згідно з програмою навчальної дисципліни передбачають лекційні, лабораторні заняття, а також виконання самостійної роботи. Оцінювання сформованих компетентностей у студентів здійснюється за накопичувальною 100-бальною системою. Відповідно до Тимчасового положення "Про порядок оцінювання результатів навчання студентів за накопичувальною бально-рейтинговою системою" ХНЕУ ім. С. Кузнеця, контрольні заходи містять:

поточний контроль, що здійснюється протягом семестру під час проведення лекційних, лабораторних занять і оцінюється сумою набраних балів (максимальна сума – 60 балів; мінімальна сума, що дозволяє студенту скласти іспит, – 35 балів);

Проміжний модульний контроль рівня знань передбачає виявлення опанування студентом матеріалів лекційного модуля та вміння застосовувати його для вирішення практичної ситуації і проводиться у вигляді захисту індивідуальних завдань та тестування. Тестове завдання може містити як запитання, що стосуються суто теоретичного матеріалу, так і запитання, спрямовані на вирішення невеличкого практичного завдання. Тести для проміжного контролю вибираються із загального переліку тестів за відповідними модулями.

Підсумковий/семестровий контроль проводиться у формі семестрового екзамену відповідно до графіку навчального процесу.

Порядок проведення поточного оцінювання знань студентів. Оцінювання знань студента під лабораторних занять та виконання індивідуальних завдань проводиться за такими критеріями:

розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються; ступінь засвоєння фактичного матеріалу навчальної дисципліни; ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що

розглядаються; вміння поєднувати теорію з практикою під час розгляду практичних задач та завдань, винесених на самостійний розгляд; логіка, структура, стиль викладу матеріалу за індивідуальними завданнями і під час їх захисту, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки; арифметична правильність виконання індивідуального завдання; здатність проводити критичну та незалежну оцінку певних проблемних питань; вміння пояснювати альтернативні погляди та наявність власної точки зору, позиції на певне проблемне питання; застосування аналітичних підходів; якість і чіткість викладення міркувань; логіка, структуризація та обґрунтованість висновків; самостійність виконання роботи; грамотність подачі матеріалу, використання методів порівняння, узагальнення понять та явищ. Загальними критеріями, за якими здійснюється оцінювання самостійної роботи студентів, є: глибина і міцність знань, рівень мислення, вміння систематизувати знання за окремими темами, вміння робити обґрунтовані висновки, володіння категорійним апаратом, прийомами виконання практичних завдань, вміння знаходити необхідну інформацію, здійснювати її систематизацію та обробку, самореалізація на лабораторних заняттях.

**Підсумковий контроль** знань та компетентностей студентів з навчальної дисципліни здійснюється на підставі проведення семестрового екзамену, завданням якого є перевірка розуміння студентом програмного матеріалу в цілому, логіки та взаємозв'язків між окремими розділами, здатності творчого використання накопичених знань, вміння формулювати своє ставлення до певної проблеми. Екзаменаційний білет охоплює програму дисципліни і передбачає визначення рівня знань та ступеня опанування студентами компетентностей. Виконання кожного екзаменаційного завдання оцінюється таким чином (табл. 1)

Таблиця 1

#### Розподіл балів

Завдання	Кількість балів
Завдання 1	25
Завдання 2	7
Завдання 3	8
Разом	40

**Перше завдання** спрямоване на оцінку когнітивних знань студента за дисципліною, що дозволяє визначити рівень володіння навчальним матеріалом, для цього використовуються наступні критерії (табл. 2):

Таблиця 2

#### Критерії оцінювання завдання 1

Кількість вірних відповідей	Кількість балів	Кількість вірних відповідей	Кількість балів
1	2	3	4
1	1	14	14

Закінчення табл. 2

1	2	3	4
2	2	15	15
3	3	16	16
4	4	17	17
5	5	18	18
6	6	19	19
7	7	20	20
8	8	21	21
9	9	22	22
10	10	23	23
11	11	24	24
12	12	25	25
13	13		

**Друге завдання** спрямоване на оцінку компетентностей, пов'язаних з умінням застосовувати здобуті знання для вирішення широкого класу задач управління даними з предметної області із застосуванням методів аналітики великих даних. Оцінювання завдання здійснюється у відповідності з критеріями (табл. 2)

Таблиця 3

### Критерії оцінювання завдання 2

Бал	Вимоги до відповіді на завдання
1	2
7	На поставлені завдання дано правильні відповіді. Розв'язання виконано в логічній послідовності. Хід розв'язання задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання. Наведено логічні висновки про результати розрахунків.
6	На всі поставлені завдання дано відповіді, деякі з них мають помилки арифметичного характеру. Розв'язання виконано в логічній послідовності. Хід розв'язання задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання.
5	На всі поставлені завдання дано відповіді з помилками арифметичного та логічного характеру, що впливають на загальний відповідь завдань. Хід розв'язано задачі не супроводжується поясненнями, які використовують понятійний апарат методу розв'язання.
4	Всі поставлені завдання розв'язано з помилками арифметичного та логічного характеру. Наведено формальну постановку і загальний алгоритм рішення, однак, розв'язок містить істотні логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні. Хід рішення задачі не супроводжується поясненнями.



1	2
3	Наведена формальна постановка і загальний алгоритм розв'язання, однак, розв'язання містить істотні логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні.
2	Наведена лише формальна постановка рішення, рішення містить істотні логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні, пояснення відсутні.
1	Студент зміг тільки запропонувати загальний алгоритм розв'язання задачі.
0	Завдання не розв'язано

**Третє завдання** спрямовано на виявлення творчого мислення студента, його здатності до формування рекомендацій на підставі інтеграції отриманих знань і застосування їх для вибору та обґрунтування комплексу методів і моделей, математичного апарату, сучасних концепції, які визначають різні підходи до порівняльного аналізу великих даних. Оцінювання завдання здійснюється у відповідності з критеріями (табл. 4)

Таблиця 4

**Критерії оцінювання завдання 3**

Бал	Вимоги до відповіді на завдання
1	2
8	На поставлені завдання дано правильні відповіді. Розв'язання виконано в логічній послідовності. Хід розв'язання задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання. Наведено логічні висновки про результати розрахунків. Наведено рекомендації за виконаними результатами моделювання.
7	На поставлені завдання дано відповіді з помилками арифметичного характеру, які несуттєво впливають на загальні висновки. Розв'язання виконано в логічній послідовності. Хід розв'язання задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання. Висновки за результатами розрахунків присутні.
6	На всі поставлені завдання дано відповіді з істотними помилками арифметичного та логічного характеру, що впливають на загальну відповідь за завданням. Хід рішення задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання.
5	Наведено рішення завдання, формальна постановка і загальний алгоритм рішення, однак, рішення містить істотні арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні. Хід рішення не супроводжується поясненнями.

1	2
4	Наведено рішення завдання, формальна постановка і загальний алгоритм рішення, однак, рішення містить істотні логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки невірні. Хід рішення задачі не супроводжується поясненнями.
3	Наведена формальна постановка і загальний алгоритм рішення, однак, рішення містить істотні логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки невірні. Пояснення відсутні
2	Наведена лише формальна постановка рішення, рішення містить істотні логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не наведено.
1	Студент зміг запропонувати алгоритм розв'язання задачі.
0	Завдання не розв'язано

Результат семестрового екзамену оцінюється в балах (максимальна кількість – 40 балів, мінімальна кількість, що зараховується, – 25 балів) і проставляється у відповідній графі екзаменаційної "Відомості обліку успішності".

Студента слід **вважати атестованим**, якщо сума балів, одержаних за результатами підсумкової/семестрової перевірки успішності, дорівнює або перевищує 60. Мінімумально можлива кількість балів за поточний і модульний контроль упродовж семестру – 35 та мінімумально можлива кількість балів, набраних на екзамені, – 25.

Підсумкова оцінка з навчальної дисципліни розраховується з урахуванням балів, отриманих під час екзамену, та балів, отриманих під час поточного контролю за накопичувальною системою (табл. 5).

Таблиця 5

## Розподіл балів за тижнями

Теми змістовного модулю		Лекції	Лабораторні заняття	Індивідуальні завдання за темами	Поточні контрольні роботи	Разом
1		2	3	4	5	6
ЗМ 1	Тема 1	1 тиждень	1	2		3
		2 тиждень				
	Тема 2	3 тиждень	1	2	4	7
		4 тиждень				
		5 тиждень	1	2	4	7
		6 тиждень				
	Тема 3	7 тиждень	1	2	4	7
		8 тиждень				

		1	2	3	4	5	6
ЗМ 2	Тема 4	9 тиждень	1	2	4	4	11
		10 тиждень					
	Тема 5	11 тиждень	1	2	8		11
		12 тиждень					
	Тема 6	13 тиждень		2	8	4	14
		14 тиждень					
		15 тиждень					
		16 тиждень					
		17 тиждень					
	Разом за змістовими модулями		6	14	32	8	60
Іспит						40	
Разом						100	

Сумарний результат у балах за семестр складає: "60 і більше балів – зараховано", "59 і менше балів – не зараховано" та заноситься у залікову "Відомість обліку успішності" навчальної дисципліни.

### Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проєкту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
82 – 89	B	добре	
74 – 81	C		
64 – 73	D	задовільно	
60 – 63	E		
35 – 59	FX	незадовільно	не зараховано
1 – 34	F		

## 5. Рекомендована література

### 5.1. Основна

1. Бізнес-аналітика багатовимірних процесів : навчальний посібник / Т. С. Клебанова, Л. С. Гур'янова, Л. О. Чаговець та ін. – Харків. : Вид. ХНЕУ ім. С. Кузнеця, 2018. – 272 с.

2. Клебанова Т. С. Нечітка логіка та нейронні мережі в управлінні підприємством : монографія / Т. С. Клебанова, Л. О. Чаговець, О. В. Панасенко. – Х. : ВД «ІН-ЖЕК», 2011. – 240 с.

3. Сенько А. Работа с BigData в облаках. Обработка и хранение данных с примерами из Microsoft Azure/ А. Сенько. – СПб. : Питер, 2019. – 448 с.

4. Форман Дж. Много цифр: Анализ больших данных при помощи Excel / Дж. Форман; 2-е изд. – Москва: Альпина Паблишер, 2019. – 461 с.

5. Фрэнкс Б. Революция в аналитике. Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Б. Фрэнкс. – М. : Альпина Паблишер. – 2018. – 316 с.

## 5.2. Додаткова

6. Боровиков В. П. Популярное введение в современный анализ данных и машинное обучение на STATISTICA / В. П. Боровиков, – М.: Горячая линия – Телеком, 2018. – 354 с.

7. Bruce P. Practical Statistics for Data Scientists: 50 Essential Concepts / P. Bruce, E. Bruce. – Boston, Sebastopol, Tokyo, Beijing, Farnham: O'Reilly, 2017. – 304 p.

8. Chagovets L. Fuzzy Logic and Neural Networks Application in Estimation of Economic Security / L.Chagovets, N. Chernova, O. Panasenko Oksana, I. Medvicka // Conference Proceedings of the 2nd International Scientific Conference “Economic and Social-Focused Issues of Modern World” (October 16 – 17, 2019, Bratislava, Slovak Republic). – Pp. 20-29.

9. Davenport T. H. Big Data at Work: Dispelling the Myths, Uncovering the Opportunities / T. H. Davenport, Harvard Business Review Press, 2014 – 228 p.

10. Бутаков Н.А. Обработка больших данных с Apache Spark / Н. А. Бутаков, М. В. Петров, Д.В. Насонов. – СПб: Университет ИТМО, 2019. – 50 с.

11. Инструментальные средства моделирования систем в информационной экономике / Под ред. докт. экон. наук, проф. В.С. Пономаренко, докт. экон. наук, проф. Т.С. Клебановой. – Харьков: ВШЭМ – ХНЭУ им. С. Кузнеца, 2019. – 472 с.

12. Информационная экономика: этапы развития, методы управления, модели / Под ред. докт. экон. наук, проф. В.С. Пономаренко, докт. экон. наук, проф. Т.С. Клебановой. – Харьков: ВШЭМ – ХНЭУ им. С. Кузнеца, 2018. – 668 с.

13. Mayer-Schönberger V. Big Data: A Revolution That Will Transform How We Live, Work, and Think / V. Mayer-Schönberger, K. Cukier, E. Dolan, Mariner Books; Reprint edition, 2014. – 272 p.

14. Моделирование процессов управления в информационной экономике / Под ред. докт. экон. наук, проф. В.С. Пономаренко, докт. экон. наук, проф. Т.С. Клебановой. – Бердянск, Издатель Ткачук А.В., 2017. – 386 с.

15. Орельен Ж. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем / Ж. Орельен. – СПб. : ООО «Диалектика», 2019. – 688 с.

16. Радченко И. А. Технологии и инфраструктура Big Data / И. А. Радченко, И. Н. Николаев. – СПб. : Университет ИТМО, 2018. – 52 с.

17. Силен Д. Основы Data Science и Big Data. Python и наука о данных / Д. Си-

лен Д., А. Мейсман, М. Али. – СПб.: Питер, 2017. — 336 с.

18. Smolan R. The Human Face of Big Data / R. Smolan, J. Erwitte. – Against All Odds Productions; First Edition edition, 2012. – 224 p.

### **5.3. Інформаційні ресурси**

19. Електронний курс «Аналітика великих даних». – [Електронний ресурс]. – Режим доступу: <https://pns.hneu.edu.ua/course/view.php?id=5417>.

20. Цифрова трансформація бізнесу. Big Data (Великі дані). – [Електронний ресурс]. – Режим доступу : <https://www.it.ua/knowledge-base/technology-innovation/big-data-bolshie-dannye>.