

# НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ


Кафедра комп'ютерних наук

**«ЗАТВЕРДЖУЮ»**  
Декан факультету інформаційних технологій  
Олена ГЛАЗУНОВА  
« 12 » вересня 20 23 р.



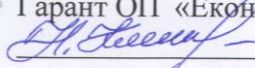
**«СХВАЛЕНО»**  
на засіданні кафедри комп'ютерних наук  
Протокол № 12 від «01» 06 20 23 р.

Завідувач кафедри  
Белла ГОЛУБ



**«РОЗГЛЯНУТО»**

Гарант ОП «Економічна кібернетика»  
Наталія КЛИМЕНКО



## РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ Технології Data Mining

Спеціальність – 051 «Економіка»

Освітня програма – «Економічна кібернетика»

Факультет інформаційних технологій

Розробник: к.е.н. Густера О.М.

Київ – 2023 р.

# 1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

## Технології Data Mining

<b>Галузь знань, спеціальність, освітній ступінь</b>	
Освітній ступінь	Магістр
Спеціальність	051 «Економіка»
Освітня програма	«Економічна кібернетика»
<b>Характеристика навчальної дисципліни</b>	
Вид	Вибіркова
Загальна кількість годин	120
Кількість кредитів ECTS	4
Кількість змістових модулів	2
Форма контролю	іспит
<b>Показники навчальної дисципліни для денної форми навчання</b>	
	денна форма навчання
Курс (рік підготовки)	1
Семестр	2
Лекційні заняття	20 год.
Лабораторні заняття	30 год.
Курсова робота	
Самостійної роботи студента	70 год.
Кількість тижневих аудиторних годин для денної форми навчання	10 год.

## 2. МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Метою викладання дисципліни “ТЕХНОЛОГІЇ DATA MINING” є ознайомити магістрів із найсучаснішою технологією аналізу даних – DataMining. Data Mining – дослідження і знаходження “машиною” (алгоритмами, засобами штучного інтелекту) у сирих даних схованих знань, які раніше не були відомі, нетривіальні, практично корисні, доступні для інтерпретації людиною.

У курсі розглядаються задачі аналізу даних, що став можливим завдяки великим обсягам накопичених даних за роки існування підприємства і використання оперативних баз даних. Із різних оперативних джерел дані поступають в єдине велике сховище даних, де за допомогою спеціальних програмних засобів аналізуються і класифікуються.

Крім того, у курсі розглядаються такі питання:

- методи та стадії технології Data Mining;
- задачі Data Mining;
- методи побудови правил класифікації;
- алгоритми класифікації даних.

Отриманні знання після прослуховування курсу «Технології Data Mining» майбутній спеціаліст зможе застосовувати як при подальшому навчанні, так і після отримання вищої освіти у своїй професійній діяльності. Оволодіння такими знаннями дозволить майбутньому спеціалісту активно включатися у процеси, які пов’язані аналітикою.

Набуття компетентностей:

СК3. Здатність збирати, аналізувати та обробляти статистичні дані, науково-аналітичні матеріали, які необхідні для розв’язання комплексних економічних проблем, робити на їх основі обґрунтовані висновки.

СК4. Здатність використовувати сучасні інформаційні технології, методи та прийоми дослідження економічних та соціальних процесів, адекватні встановленим потребам дослідження.

СК11. Здатність створювати та оцінювати моделі економічних процесів як аналітично так і з використанням універсальних програмних засобів і аналітичних платформ, що застосовуються для аналізу даних.

Програмні результати навчання:

ПР10. Застосовувати сучасні інформаційні технології та спеціалізоване програмне забезпечення у соціально-економічних дослідженнях та в управлінні соціально-економічними системами.

## 3. ПРОГРАМА І СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**Змістовий модуль 1. DataMining – Технології аналізу даних.**

**Тема лекційного заняття 1. DATA MINING – основні положення.** Розвиток технологій баз даних. Накопичення великих обсягів даних.

Чинники виникнення технології DataMining. Відмінності DataMining від інших методів аналізу даних.

**Тема лекційного заняття 2. OLAP і DataMining.**

Технології оперативного аналізу даних. Створення сховища даних.

Передача транзакційних даних різного походження у сховище даних. Служба BI Microsoft SQL Server. Порівняння можливостей двох технологій.

**Тема лекційного заняття 3.** Перспективи технології Data Mining. Виділення типів предметних областей з відповідними їм евристичними.

Створення формальних мов і логічних засобів. Створення методів Data Mining, здатних формувати якісь теорії, що спираються на емпіричні дані. Подолання істотного відставання можливостей інструментальних засобів Data Mining від теоретичних досягнень у цій галузі.

**Тема лекційного заняття 4.** Задачі Data Mining. Класифікація. Регресія. Пошук асоціативних правил. Кластеризація.

**Тема лекційного заняття 5.** Класифікація стадій Data Mining.

Стадія 1. Виявлення закономірностей (вільний пошук).

Стадія 2. Використання виявлених закономірностей для передбачення невідомих значень (прогностичне моделювання).

Стадія 3. Аналіз винятків - стадія призначена для виявлення і пояснення аномалій, знайдених в закономірностях.

**Змістовий модуль 2.** Класифікація даних. Алгоритми.

**Тема лекційного заняття 6.** Класифікація і регресія.

Результати рішення задачі класифікації. Методи рішення. Регресивний аналіз.

**Тема лекційного заняття 7.** Методи побудови правил класифікації. Алгоритм побудови 1-правил. Метод Naïve Bayes.

**Тема лекційного заняття 8.** Спеціальні методи побудови правил класифікації. Древа рішень. Кластеризація. Математичні функції.

**Тема лекційного заняття 9.** Пошук асоціативних правил.

Постановка задачі. Транзакції: набір, таблиця і множина. Приклади використання.

#### 4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістовних модулів і тем	Кількість годин													
	денна форма							Заочна форма						
	ти жн і	ус ь о го	у тому числі					усь ог о	у тому числі					
			л	п	л а б	і н д	с. р.		л	п	л а б	і н д	с. р.	
1	2	3	4	5	6	7	8	9	1 0	1 1	1 2	1 3	1 4	
Змістовий модуль 1. DataMining – Технології аналізу даних														
Тема 1. DATA MINING – основні положення	1	16	2		4		1 0							
Тема 2. OLAP і DataMining	1	32	4		8		2 0							
Тема 3. Класифікація стадій DataMining	2	18	4		4		1 0							
Разом за змістовим модулем 1	2	66	10		1 6		4 0							
Змістовий модуль 2. Класифікація даних. Алгоритм														
Тема 4. Класифікація і регресія	3	16	2		4		1 0							
Тема 5. Методи побудови правил класифікації	4	18	4		4		1 0							
Тема 6. Пошук асоціативних правил	5	20	4		6		1 0							
Разом за змістовим модулем 2	3	54	10		1 4		3 0							
Усього	5	120	20		3 0		7 0							

## 5. ТЕМИ ЛАБОРАТОРНИХ ЗАНЯТЬ

№ з / п	Назва теми	Кількість годин
1	Інтелектуальний аналіз даних за допомогою програмного пакета WEKA	6
2	Інтелектуальний аналіз даних за допомогою програмного пакета WEKA: Класифікація і кластеризація	6
3	Інтелектуальний аналіз даних за допомогою програмного пакета WEKA. Метод найближчих сусідів.	4
4	Знайомство з Statistica. Робота з таблицями даних і побудова графіків	4
5	Описова статистика. Дані в пакеті STATISTICA	4
6	Лінійна регресія. Багатофакторна регресія.	6

## 6. ТЕМИ САМОСТІЙНОЇ РОБОТИ

№ з / п	Назва теми	Кількість годин
1	Побудова сховища даних для задач аналізу в середовищі BI Microsoft SQL Server. Отримання даних у сховище із двох різних джерел.	20
2	Обчислення KPI в середовищі BI Microsoft SQL Server.	10
3	Розробка кваліфікаційних правил за допомогою 1R-алгоритму.	10
4	Визначення алгоритму Naive Bayesу вигляді приклада.	10
5	Знаходження множини частих наборів (L) для прикладу, при $Supp_{min}=0,5$ .	10
6	Використання алгоритму Apriori.	10

## 7. КОНТРОЛЬНІ ПИТАННЯ, КОМПЛЕКТИ ТЕСТІВ ДЛЯ ВИЗНАЧЕННЯ РІВНЯ ЗАСВОЄННЯ ЗНАТЬ СТУДЕНТАМИ

1. Основні засади технології Data Mining.
2. Стадії Data Mining.
3. Класифікація задач Data Mining.
4. Класифікація і регресія. Вирішення цих задач методами Data Mining.
5. Способи представлення результатів класифікації і регресії. Класифікаційні правила.
6. Способи представлення результатів задач Data Mining. Древа рішень і математичні функції.
7. Методи побудови правил класифікації. Алгоритм побудови 1-правил. 8. Пошук асоціативних правил. Алгоритм Apriori.

Комплекти тестів розташовані в осередку ЕНК за посиланням:  
<https://elearn.nubip.edu.ua/course/view.php?id=1141>.

## **8. МЕТОДИ НАВЧАННЯ**

Форми навчання – лекції, лабораторні заняття, самостійна робота, індивідуальна робота – курсовий проект.

При викладанні дисципліни використовуються наступні методи навчання:

М1. Лекція (проблемна, інтерактивна)

М2. Лабораторна робота – для використання набутих знань до розв'язування практичних завдань;

М3. Проблемне навчання – створення проблемної ситуації для зацікавленого і активного сприйняття матеріалу.

М4. Проектне навчання (індивідуальне, малі групи, групове)

## **9. ФОРМИ КОНТРОЛЮ**

При викладанні дисципліни передбачені такі форми контролю:

МК1. Тестування

МК2. Контрольне завдання

МК4. Методи усного

контролю МК5. Екзамен

МК7. Звіт

Для студентів денної форми навчання: усне опитування (МК4) та експрес контроль (МК1) на лабораторних заняттях, захист індивідуальних лабораторних завдань (МК7), аудиторні модульні контрольні роботи (МК2).

## **10. РОЗПОДІЛ БАЛІВ, ЯКІ ОТРИМУЮТЬ СТУДЕНТИ**

Під час контролю знань студентів враховуючи наступні види робіт:

- робота студента на лекційних заняттях;
- захист лабораторної роботи студентом;
- аудиторні модульні контрольні роботи – до 30 балів.

Поточний контроль				Рейт инг з навч ал ьної робо ти R нр	Рейт инг з дода тк ової робо ти R др	Рейт инг штра ф ний R штр	Підсу м кова атеста ція (екза ме н чи залі к)	Зага ль на кіль кіс ть балі в
Зміст о вий моду ль 1	Зміст о вий моду ль 2	Зміст о вий моду ль 3	Зміст о вий моду ль 4					
0- 100	0- 100	0- 100	0- 100	0-70	0-20	0 - 5	0-30	0- 100

1. Відповідно до «Положення про екзамени та заліки у НУБіП України» (наказ про уведення в дію від 26 квітня 2023 р. протокол № 10) , рейтинг студента з навчальної роботи  $R_{нр}$  стосовно вивчення певної дисципліни визначається за формулою

$$R_{нр} = 0,7 \cdot (R^{(1)}_{зм} \cdot K^{(1)}_{зм} + \dots + R^{(n)}_{зм} \cdot K^{(n)}_{зм}) + R_{др} - R_{штр},$$

$K_{дис}$

де  $R^{(1)}_{зм}, \dots, R^{(n)}_{зм}$  – рейтингові оцінки змістових модулів за 100-бальною шкалою;

$n$  – кількість змістових модулів;

$K^{(1)}_{зм}, \dots, K^{(n)}_{зм}$  – кількість кредитів ECTS, передбачених робочим навчальним планом для відповідного змістового модуля;

$K_{дис} = K^{(1)}_{зм} + \dots + K^{(n)}_{зм}$  – кількість кредитів ECTS, передбачених робочим навчальним планом для дисципліни у поточному семестрі;

$R_{др}$  – рейтинг з додаткової роботи;

$R_{штр}$  – рейтинг штрафний.



Наведену формулу можна спростити, якщо прийняти  $K^{(1)}_{ЗМ} = \dots = K^{(n)}_{ЗМ}$ . Тоді вона буде мати вигляд

$$R_{НР} = \frac{0,7 \cdot (R^{(1)}_{ЗМ} + \dots + R^{(n)}_{ЗМ})}{n} + R_{ДР} - R_{ШТР}.$$

**Рейтинг з додаткової роботи**  $R_{ДР}$  додається до  $R_{НР}$  і не може перевищувати 20 балів. Він визначається лектором і надається студентам рішенням кафедри за виконання робіт, які не передбачені навчальним планом, але сприяють підвищенню рівня знань студентів з дисципліни.

**Рейтинг штрафний**  $R_{ШТР}$  не перевищує 5 балів і віднімається від  $R_{НР}$ . Він визначається лектором і вводиться рішенням кафедри для студентів, які матеріал змістового модуля засвоїли невчасно, не дотримувалися графіка роботи, пропускали заняття тощо.

2. Згідно із зазначеним Положенням **підготовка і захист курсового проекту (роботи)** оцінюється за 100 бальною шкалою і далі переводиться в оцінки за національною шкалою.

### Шкала оцінювання

Сума балів за всі види навчальної діяльності	Оцінка за національною шкалою	
	для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	відмінно	зараховано
74-89	добре	
60-73	задовільно	
35-59	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

## 11. МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ

1. Голуб Б. Л., Ящук Д. Ю. Організація сховищ даних: Навчальний посібник до вивчення дисципліни «Організація сховищ даних». – Київ, 2018. – 150 с.
2. Методичні вказівки до виконання лабораторних робіт з дисципліни «Технології Data Mining» для студентів, що навчаються за спеціальністю 122

## 12. РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

### Базові

1. Голуб Б.Л., Ящук Д.Ю. Організація сховищ даних: Навчальний посібник до вивчення дисципліни «Організація сховищ даних». – Київ, 2018. – 150 с.
2. Методичні вказівки до виконання курсового проекту з дисципліни «Організація сховища даних» для студентів, що навчаються за спеціальністю 121 «Інженерія програмного забезпечення», 122 «Комп'ютерні науки» ОС «Магістр» / Голуб Б.Л., Ящук Д.Ю. – Редакційно-видавничий відділ НУБіП України, 2018. – 20 с.
3. Пасічник В. В., Шаховська Н. Б. Сховища даних. Підручник - Магнолія - 2021 - 496с.
4. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейта- майнінг). Київ: КНЕУ, 2007. – 376 с.
5. А.О. Олійник, С.О. Субботін, О.О. Олійник. Інтелектуальний аналіз даних. – Запоріжжя: ЗНТУ, 2012. – 278 с.
6. Д.В. Ланде, І.Ю. Субач, Ю.Є. Бояринова. Основи теорії і практики аналізу інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗЗІ КПІ ім. Ігоря Сікорського», 2018. – 297 с.
7. Черняк О.І. Інтелектуальний аналіз даних: підручник. – К: Знання, 2014. –599с.

### Додаткові

1. Ian H. Witten>DataMining: PracticalMachineLearningToolsandTechniques / Ian H. Witten, EibeFrankandMark A. Hall — 3rd Edition. — Morgan Kaufmann, 2011. — P. 664. и.], 2005. - 159 с.
2. Курс Power BI: <https://powerbi.microsoft.com/ru-ru/learning/>
3. Курс Power BI на платформі Udemy: <https://www.udemy.com/ru/topic/microsoft-power-bi/>
4. Курс Power BI на платформі Coursera:<https://ru.coursera.org/courses?query=power%20bi>
5. Data Warehouse: <https://youtube.com/playlist?list=PLIBwmCV8YuLOXLGkYK1Ifid2lhaVBuh1N>