


# НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Кафедра комп'ютерних наук

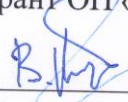
**ЗАТВЕРДЖУЮ**  
Декан факультету інформаційних технологій  
Олена ГЛАЗУНОВА  
« 12 » вересня 20 23 р.



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



**«РОЗГЛЯНУТО»**  
Гарант ОП «Програмне забезпечення інформаційних систем»  
Віктор КИРИЧЕНКО



## РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ Технологія Data Mining

**Спеціальність** – 121 «Інженерія програмного забезпечення»

**Освітня програма** – «Програмне забезпечення інформаційних систем»

**Факультет** інформаційних технологій

**Розробник:** зав. кафедри комп'ютерних наук , доцент, к.т.н. Голуб Б. Л.

Київ – 2023 р.

# 1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

## Технологія Data Mining

<b>Галузь знань, спеціальність, освітній ступінь</b>	
Освітній ступінь	Магістр
Спеціальність	121 «Інженерія програмного забезпечення»
Освітня програма	«Програмне забезпечення інформаційних систем»
<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 “ТЕХНОЛОГІЯ 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 «Технологія 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 загальних компетентностей:

ЗК03. Здатність проводити дослідження на відповідному рівні.

ЗК05. Здатність генерувати нові ідеї (креативність).

Набуття фахових компетентностей:

СК02. Здатність розробляти і реалізовувати наукові та/або прикладні проекти у сфері інженерії програмного забезпечення.

СК07. Здатність критично осмислювати проблеми у галузі інформаційних технологій та на межі галузей знань, інтегрувати відповідні знання та розв'язувати складні задачі у широких або мультидисциплінарних контекстах.

СК08. Здатність розробляти і координувати процеси, етапи та ітерації життєвого циклу програмного забезпечення на основі застосування сучасних моделей, методів та технологій розроблення програмного забезпечення.

## 3. ПРОГРАМА І СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**Змістовий модуль 1. Data Mining – технологія 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.**

Технологія оперативного аналізу даних. Створення сховища даних.

Передача транзакційних даних різного походження у сховище даних. Служба

ВІMicrosoftSQLServer. Порівняння можливостей двох технологій.

**Тема лекційного заняття 3. Перспективи технології DataMining.**  
 Виділення типів предметних областей з відповідними їм евристичними.  
 Створення формальних мов і логічних засобів. Створення методів DataMining, здатних формувати якісь теорії, що спираються на емпіричні дані. Подолання істотного відставання можливостей інструментальних засобів DataMining від теоретичних досягнень у цій галузі.

**Тема лекційного заняття 4. Задачі DataMining.**

Класифікація. Регресія. Пошук асоціативних правил. Кластеризація.

**Тема лекційного заняття 5. Класифікація стадій DataMining.**

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.**

Дерева 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						
	тижн і	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	14	
<b>Змістовий модуль 1. DataMining – технологія аналізу даних</b>														
Тема 1. DATA MINING – основні положення	1	16	2		4		10							
Тема 2. OLAP і DataMining	1	32	4		8		20							
Тема 3. Класифікація стадій DataMining	2	18	4		4		10							
Разом за змістовим модулем 1	2	66	10		16		40							

Змістовий модуль 2. Класифікація даних. Алгоритм												
Тема 4. Класифікація і регресія	3	16	2		4		10					
Тема 5. Методи побудови правил класифікації	4	18	4		4		10					
Тема 6. Пошук асоціативних правил	5	20	4		6		10					
Разом за змістовим модулем 2	3	54	10		14		30					
Усього	5	120	20		30		70					

## 5. ТЕМИ ЛАБОРАТОРНИХ ЗАНЯТЬ

№ 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.	6
2	Обчислення KPI в середовищі BI Microsoft SQL Server.	6
3	Розробка кваліфікаційних правил за допомогою 1R-алгоритму.	4
4	Визначення алгоритму Naive Bayes у вигляді приклада.	4
5	Знаходження множини частих наборів (L) для прикладу, при $Supp_{min}=0,5$ .	4
6	Використання алгоритму Apriori.	6

## 6. ТЕМИ САМОСТІЙНОЇ РОБОТИ

№ з/п	Назва теми	Кількість годин
Модуль 1		
1	Огляд не реляційних баз даних, область використання. Порівняльний аналіз з реляційними базами даних.	10
2	Мова R. Область застосування. Приклади коду.	10
3	Поняття науки про дані (Data Science). Основні напрямки та методи.	10
4	Машинне навчання (Machine Learning). Основні напрямки та методи.	10

№ з/п	Назва теми	Кількість годин
5	Глибоке навчання (Deep Learning). Основні напрямки та методи.	10
Модуль 2		
6	Проходження курсу «Structuring Machine Learning Projects»	20
<b>Всього годин:</b>		<b>70</b>

## 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=919>.

## 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.

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  $R_{\text{НР}}$  стосовно вивчення певної дисципліни визначається за формулою

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

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

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

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

$R_{\text{ДР}}$  – рейтинг з додаткової роботи;

$R_{\text{ШТР}}$  – рейтинг штрафний.

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

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

**Рейтинг з додаткової роботи**  $R_{\text{ДР}}$  додається до  $R_{\text{НР}}$  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**  $R_{\text{ШТР}}$  не перевищує 5 балів і віднімається від  $R_{\text{НР}}$ . 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	
0-34	незадовільно	не зараховано

### 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 «Технологія Data Mining» для студентів, що навчаються за спеціальністю 122 «Комп'ютерні науки» ОС «Магістр» / Голуб Б.Л. – Редакційно-видавничий відділ НУБіП України, 2021.– 60 с.

### 12. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

#### Базова

1. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. — К.: КНЕУ, 2007. — 376 с.

2. Wang, Lipo; Fu, Xiuju (2005). Data mining with computational intelligence. URL: <http://surl.li/hyljq>.

3. Голуб Б. Л., Ящук Д. Ю. Організація сховищ даних: Навчальний посібник до вивчення дисципліни «Організація сховищ даних». – Київ, 2018. – 150 с.

#### Додаткова

1. 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. и.], 2005. - 159 с.

2. ДСТУ 2874-94. Системи оброблення інформації. Бази даних. Терміни та визначення. К. : Держстандарт України, 1995. 29 с.

3. ДСТУ 2938-94. Системи оброблення інформації. Основні поняття. Терміни та визначення. К. : Держстандарт України, 1995. 32 с.

### 13. ІНФОРМАЦІЙНІ РЕСУРСИ

1. ЕНК по даній дисципліні знаходиться за електронною адресою: <https://elearn.nubip.edu.ua/course/view.php?id=919>