

**Національний університет біоресурсів і природокористування  
України**

Кафедра комп'ютерних наук

**«ЗАТВЕРДЖУЮ»**

Декан факультету інформаційних технологій

\_\_\_\_\_ О. Г. Глазунова

«\_\_\_\_\_» \_\_\_\_\_ 20\_\_ р.

**«СХВАЛЕНО»**

на засіданні кафедри комп'ютерних наук

Протокол № \_\_\_\_\_ від «\_» \_\_\_\_\_ 20\_\_ р.

Завідувач кафедри

\_\_\_\_\_ Б. Л. Голуб

**”РОЗГЛЯНУТО ”**

Гарант ОП «Інформаційні управляючі системи та технології»

\_\_\_\_\_ проф., д.т.н., Бондаренко В. Є.

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

**ТЕХНОЛОГІЯ DATA MINING**

Спеціальність 122 «Комп'ютерні науки»

Освітня програма «Інформаційні управляючі системи та технології»

Факультет інформаційних технологій

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

# 1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

## ТЕХНОЛОГІЯ DATA MINING

<b>Галузь знань, спеціальність, освітній ступінь</b>	
Спеціальність	122 «Комп'ютерні науки»
Освітня програма	«Інформаційні управляючі системи та технології»
Освітній ступінь	Магістр
<b>Характеристика навчальної дисципліни</b>	
Вид	Обов'язкова
Загальна кількість годин	120
Кількість кредитів ECTS	4
Кількість змістових модулів	2
Форма контролю	Іспит
<b>Показники навчальної дисципліни для денної та заочної форм навчання</b>	
денна форма навчання	
Рік підготовки	1
Семестр	2
Лекційні заняття	20 год.
Лабораторні заняття	30 год.
Самостійна робота	70 год.
Кількість тижневих годин для денної форми навчання: аудиторних: самостійної роботи студента:	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 і 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 з аналітикою.

### **Набуття компетентностей.**

#### *1. Відповідно ОПП ІУСТ та КЕЕМ*

ЗК2. Здатність застосовувати знання у практичних ситуаціях, насамперед, пов’язаних з природоохоронною галуззю.

ЗК3. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК7. Здатність генерувати нові ідеї (креативність).

СК1. Розуміння теоретичних засад комп’ютерних наук для об’єктивного оцінювання можливостей використання обчислювальної техніки в певних процесах людської діяльності і визначення перспективних інформаційних технологій, у тому числі, геоінформаційних технологій у природоохоронній галуззі.

СК5. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області певного проекту в процесі його реалізації і супроводження.

СК6. Здатність збирати і аналізувати дані (включно з великими), насамперед, пов’язаних з природоохоронною галуззю, для забезпечення якості прийняття рішень на основі моніторингових процесів.

СК8. Здатність застосовувати існуючі і розробляти нові алгоритми розв’язування задач у галузі комп’ютерних наук: алгоритми розв’язання обчислювальних та логічних задач, алгоритми паралельних та розподілених обчислень, алгоритми аналітичної обробки й інтелектуального аналізу великих даних з оцінкою їх ефективності та складності.

СК9. Здатність розробляти програмне забезпечення: розуміти та застосовувати основи логіки для вирішення проблем; вміти конструювати, виконувати та налагоджувати програми за допомогою сучасних інтегрованих

програмних (візуальних) середовищ розробки; розуміти методології програмування, включаючи об'єктно-орієнтоване, структуроване, процедурне та функціональне програмування; порівнювати наявні в даний час мови програмування, методології розробки програмного забезпечення та середовища розробки, а також обирати та використовувати ті, що відповідають певному проекту; вміти оцінювати код для повторного використання або включення до існуючої бібліотеки; вміти оцінювати конфігурацію та вплив на налаштування в умовах роботи з сторонніми програмними пакетами.

СК11. Здатність розробляти та адмініструвати бази даних та знань, володіти сучасними теоріями та моделями даних та знань, методами їх інтерактивної та автоматизованої розробки, технологіями обробки та візуалізації.

## *2. Відповідно ОПП ІІЗ*

ЗК03. Здатність проводити дослідження на відповідному рівні.

ЗК05. Здатність генерувати нові ідеї (креативність). 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.

РН04 Виявляти інформаційні потреби і класифікувати дані для проектування програмного забезпечення.

РН05 Розробляти, аналізувати, обґрунтовувати та систематизувати вимоги до програмного забезпечення.

РН07 Аналізувати, оцінювати і застосовувати на системному рівні сучасні програмні та апаратні платформи для розв'язання складних задач інженерії програмного забезпечення.

### **3 ПРОГРАМА І СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

**Змістовий модуль 1. DataMining – 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 BI Microsoft SQL Server. 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 змістових модулів і тем	Кількість годин													
	денна форма							Заочна форма						
	тижн і	усьо го	у тому числі					усьог о	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	
Змістовий модуль 1. DataMining – технологія аналізу даних														
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 NaiveBayes у вигляді приклада.	4
5	Знаходження множини частих наборів (L) для прикладу, при $Supp_{min}=0,5$ .	4
6	Використання алгоритму Apriori.	6

## **6 САМОСТІЙНА РОБОТА**

1. Области використання DataMining(8 годин).
2. Класи систем DataMining(8 годин).
3. Інструментарій технологіїDataMining(8 годин).
4. Алгоритми навчання (8 годин).
5. Технології інтелектуальних обчислень і апаратне забезпечення (8 годин).
6. Технології інтелектуальних обчислень та український ринок (8 годин).
7. Тестування та інтерпретація отриманих результатів (8 годин).
8. Аналітичні методи DataMining- нейронні мережі (8 годин).
9. Аналітичні методи DataMining: правило індукції(8 годин).

## **6 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. Дерева 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.

## **8 МЕТОДИ НАВЧАННЯ**

М1. Лекція (проблемна, інтерактивна). М2. Лабораторна робота.  
М3. Проблемне навчання. М4. Проектне навчання (індивідуальне). М8. Дослідницький метод.

## **9 ФОРМИ КОНТРОЛЮ**

МК1. Тестування. МК4. Методи усного контролю. МК5. Екзамен. МК7. Звіт.

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

Оцінювання знань студента відбувається за 100-бальною шкалою і переводиться в національні оцінки згідно з табл. 1 «Положення про екзамени та заліки у НУБіП України» (наказ про уведення в дію від 27.12.2019 р. № 1371).

Поточний контроль		Рейтинг з навчальної роботи $R_{НР}$	Рейтинг з додаткової роботи $R_{ДР}$	Рейтинг штрафний $R_{ШТР}$	Підсумкова атестація (екзамен чи залік)	Загальна кількість балів
Змістовий модуль 1	Змістовий модуль 2					
0-100	0-100	0-70	0-20	0-5	0-30	0-100

### Шкала оцінювання

Рейтинг здобувача вищої освіти, бали	Оцінка національна за результати складання екзаменів заліків	
	екзаменів	заліків
90-100	відмінно	зараховано
74-89	добре	
60-73	задовільно	
0-59	незадовільно	не зараховано

Для визначення рейтингу студента із засвоєння дисципліни  $R_{дис}$  (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи  $R_{НР}$  (до 70 балів):  $R_{дис} = R_{НР} + R_{ат}$ .

## 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. 1. 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.

## 12 РЕКОМЕНДОВАНА ЛІТЕРАТУРА

### Базова

1. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям. Паклин Н. Б., Орешков В. И.— СПб.: Изд. Питер, 2009. — 624 с.

2. Дюк В. DataMining: учебный курс / Дюк В., Самойленко А.— СПб.: Изд. Питер, 2001. — 368 с.

3. Журавлёв Ю.И. РАСПОЗНАВАНИЕ. Математические методы. Программная система. Практические применения / Журавлёв Ю.И., Рязанов В.В., Сенько О.В.— М.: Изд. «Фазис», 2006. — 176 с.

4. Зиновьев А. Ю. Визуализация многомерных данных / Зиновьев А. Ю.— Красноярск: Изд. Красноярского государственного технического университета, 2000. — 180 с.

5. Чубукова И. А. DataMining: учебное пособие. / Чубукова И. А.— М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. — 382 с.

### Додаткова

1. Ian H. Witten DataMining: Practical Machine Learning Tools and Techniques /



Ian H. Witten, EibeFrankandMark A. Hall — 3rd Edition. — Morgan Kaufmann, 2011.  
— Р. 664. и.], 2005. - 159 с.

### **13 ІНФОРМАЦІЙНІ РЕСУРСИ**

1. ЕНК по даній дисципліні знаходиться за електронною адресою:  
<http://it.nubip.edu.ua/course/view.php?id=144>