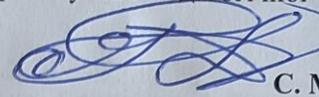


**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
Економічний факультет**

«ЗАТВЕРДЖУЮ»

Проректор з науково-педагогічної роботи та
розвитку

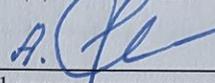


С. М. Кваша

«18» червня 2021 р.

РОЗГЛЯНУТО І СХВАЛЕНО:

на засіданні Вченої ради економічного факультету
Протокол № 9 від «15» червня 2021 р.
Декан економічного факультету



А. Д. Діброва

на засіданні кафедри статистики та економічного
аналізу

Протокол № 15 від «20» травня 2021 р.



Завідувач кафедри

І. Д. Лазаришина

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«ДЕЙТАМАЙНІНГ»

Рівень вищої освіти Третій (освітньо-науковий)

Галузь знань 07 «Управління та адміністрування»

Спеціальність 076 «Підприємництво, торгівля та біржова діяльність»

Освітньо-наукова програма «Підприємництво, торгівля та біржова діяльність»

Гарант ОНП

М. М. Ільчук

Розробники:

д. е. н., проф., Савчук В. К.,
к. е. н., доц. Воляк Л. Р.

2021

**1. Опис навчальної дисципліни
«Дейтамайнінг»**

Галузь знань, спеціальність, освітня програма, освітній ступінь		
Галузь знань	07 «Управління та адміністрування»_	
Спеціальність	076 Підприємництво, торгівля та біржова діяльність (шифр і назва)	
Освітня програма	Підприємництво, торгівля та біржова діяльність	
Освітній ступінь	Доктор філософії	
Характеристика навчальної дисципліни		
Вид	Вибіркова	
Загальна кількість годин	150	
Кількість кредитів ECTS	5	
Курсовий проект (робота) (якщо є в робочому навчальному плані)	- (назва)	
Форма контролю	Іспит	
Показники навчальної дисципліни для денної та заочної форм навчання		
	денна / вечірня форма навчання	заочна форма навчання
Рік підготовки	2	2
Семестр	3	3
Лекційні заняття	20 год.	20 год.
Практичні заняття	30 год.	30 год.
Самостійна робота	100 год.	100 год.
Індивідуальні завдання	-	
Кількість тижневих годин для денної форми навчання: аудиторних	5 год.	

2. Мета, завдання та компетентності навчальної дисципліни

Метою навчальної дисципліни «Дейтамайнінг» є вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining, Knowledge Discovery in Data), аналітичного дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються Data Mining; розгляд практичних прикладів застосування Data Mining; підготовка аспірантів до самостійної роботи з вирішення задач засобами Data Mining і розробки інтелектуальних систем.

Data Mining – мультидисциплінарна область, яка виникла і розвивається на базі таких наук як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та ін.

Завданнями вивчення дисципліни «Дейтамайнінг» є:

- здійснення пошуку, аналіз та критичне осмислення інформації, отриманої з різних джерел;
- генерування і продукування власних науково-теоретичних та практично-спрямованих ідей;
- робота з різними джерелами, аналіз, систематизація отриманої інформації;
- опанування базовими принципами побудови моделей даних;
- ознайомлення з концепцією Knowledge Discovery in Data (виявлення знань в даних) і Data Mining («видобування» знань);
- навчитися ефективно використовувати методи здобуття знань з великих масивів даних;
- отримати практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач та навчитися інтерпретувати отримані результати;
- проводити емпіричні дослідження та використовувати економіко-математичні методи для встановлення тенденцій розвитку об'єктів дослідження;
- застосовувати інформаційні технології у науковій діяльності, сучасні методи наукових досліджень із використанням новітніх прикладних пакетів і програмних продуктів для наукового обґрунтування та підтвердження / спростування висунутих гіпотез;
- ідентифікувати наукові та практичні проблеми, здійснювати апробацію результатів наукових досліджень, висновків і практичних рекомендацій.

Згідно з вимогами освітньо-наукової програми аспіранти повинні:

знати :

- систему та показники розвитку економічних систем, методи збирання, обробки та аналізу даних;
- базові методологічні засади аналізу даних;
- основні поняття, задачі та стадії інтелектуального аналізу даних;

- підходи до збереження, представлення та обробки інформації в сучасних інформаційних системах;
- методи побудови моделей та аналізу взаємозв'язків у великих масивах даних;
- сучасні програмні засоби для проектування і розробки систем інтелектуального аналізу даних;
- концепції сховищ даних, їх оперативної аналітичної обробки для практичного використання;
- прикладні пакети аналізу для обробки даних.

вміти :

- генерувати нові ідеї на підставі креативного мислення;
- здійснювати пошук, оброблення та аналіз інформації з різних джерел;
- оцінювати та забезпечувати якість дослідницької та наукової діяльності;
- працювати з різними інформаційними ресурсами, отримувати, обробляти, аналізувати, систематизувати інформацію;
- обґрунтовувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні поставленої практичної задачі;
- проводити необхідну попередню обробку даних, визначати тип задачі аналізу, вирішувати її адекватно з обраним методом та оптимально визначеними параметрами, оцінювати результати, робити змістовні висновки та інтерпретацію;
- виявляти та усвідомлювати економічні зв'язки і процеси, які мають загальнодержавний характер для типологічно однорідних умов (економічних систем, видів діяльності);
- проводити емпіричні дослідження та використовувати економіко-математичні методи аналізу для встановлення тенденцій розвитку об'єктів дослідження;
- застосовувати інформаційні технології, сучасні методи наукових досліджень із використанням новітніх прикладних пакетів і програмних продуктів для обґрунтування та підтвердження / спростування гіпотез;
- застосовувати технології роботи зі сховищами даних, здійснювати їх аналітичну обробку та інтелектуальний аналіз для забезпечення надійної роботи інформаційних систем;
- проектувати інформаційне забезпечення (логічну та фізичну структури баз даних) інформаційних систем;
- здійснювати апробацію і впровадження результатів наукових досліджень, висновків, методичних та практичних рекомендацій.

Результатом вивчення дисципліни є набуття аспірантами таких компетентностей:

загальні (ЗК)

- ЗК 1. Здатність генерувати нові ідеї (креативність).
- ЗК 3. Здатність працювати автономно.
- ЗК 4. Здатність приймати обґрунтовані рішення.

ЗК 5. Здатність розробляти та управляти проектами

Спеціальні (фахові) компетенції (СК)

СК 2. Здатність ідентифікувати, формулювати та вирішувати актуальні наукові та прикладні проблеми у сфері підприємництва, торгівлі та біржової діяльності.

СК 3. Здатність планувати, організовувати, проводити спеціальні наукові дослідження та впроваджувати їх результати.

СК 4. Здатність продукувати нові ідеї, розв'язувати комплексні проблеми та приймати науково обґрунтовані рішення в професійній діяльності.

СК6. Здатність усно і письмово презентувати та обговорювати результати наукових досліджень та/або інноваційних розробок українською та англійською мовами, глибоке розуміння англомовних наукових текстів за напрямом досліджень.

2. Програма та структура навчальної дисципліни

3.

Тема 1. Концепція дейтамайнінгу.

Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Поняття Business Intelligence. Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором і корекцією моделей. Методи первісної обробки даних. Методи дослідження структури даних: візуалізація та автоматичне групування даних.

Тема 2. Алгоритми Data Mining: класифікація і регресія.

Постановка задачі класифікації та представлення результатів. Методи побудови правил класифікації. Методи побудови математичних функцій. Класифікація об'єктів у випадку невідомих розподілень даних. Методи оцінювання помилок класифікації. Методи вирішення задач регресії.

Тема 3. Інтелектуальний аналіз часових рядів.

Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів.

Тема 4. Алгоритми Data Mining: кластеризація.

Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстаннях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.

Тема 5. Технології нейронних мереж та генетичні алгоритми.

Визначення та еволюція нейронних мереж. Математична модель штучного нейрона. Програмне забезпечення нейромереж. Апаратне забезпечення нейромереж. Діапазон застосування нейромереж, їх переваги та недоліки.

Тема 6. Дерева рішень як метод дейтамайнінгу.

Сутність дерев рішень, особливості їх використання та побудови. Дерева рішень для дослідження альтернатив. Сутність класифікації альтернативних варіантів та методи її проведення. Загальний опис дерев класифікації. Способи побудови класифікаційних дерев рішень.

Тема 7. Сховища даних та оперативний аналіз даних (OLAP).

Визначення сховища даних, порівняння з базами даних, використання. Архітектура сховища даних. ETL-процеси (добування, перетворення й завантаження даних). Вітрини даних, куби даних, багатомірна модель даних. Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.

Структура навчальної дисципліни
для денної та заочної форм навчання

Назви тем	Кількість годин					
	усього	у тому числі				
		лекц.	сем.	прак	лаб	с.р.
1	2	3	4	5	6	7
Тема 1. Концепція дейтамайнінгу	20	2	-	2	-	14
Тема 2. Алгоритми Data Mining: класифікація і регресія	20	2	-	2	-	14
Тема 3. Інтелектуальний аналіз часових рядів	28	4	-	4	-	18
Тема 4. Алгоритми Data Mining: кластеризація	28	4	-	4	-	18
Тема 5. Технології нейронних мереж та генетичні алгоритми	20	4	-	6	-	12
Тема 6. Дерева рішень як метод дейтамайнінгу	17	2	-	6	-	12
Тема 7. Сховища даних та оперативний аналіз даних (OLAP)	17	2		6		12
Разом годин	150	20	0	30	0	100

4. Семінарські заняття

Навчальним планом не передбачено

5. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1.	Тема 1. Концепція дейтамайнінгу. Ведення а програму R	2
2.	Тема 2. Алгоритми Data Mining: класифікація і регресія	2
3.	Тема 3. Інтелектуальний аналіз часових рядів	4
4.	Тема 4. Алгоритми Data Mining: кластеризація	4
5.	Тема 5. Технології нейронних мереж та генетичні алгоритми	6
6.	Тема 6. Дерева рішень як метод дейтамайнінгу	6
7.	Тема 7. Сховища даних та оперативний аналіз даних (OLAP)	6
	Разом	30

6. Теми лабораторних занять

Навчальним планом не передбачено

7. Самостійна робота

№ з/п	Назва теми	Кількість годин
1.	Тема 1. Концепція дейтамайнінгу	14
2.	Тема 2. Алгоритми Data Mining: класифікація і регресія	14
3.	Тема 3. Інтелектуальний аналіз часових рядів	18
4.	Тема 4. Алгоритми Data Mining: кластеризація	18
5.	Тема 5. Технології нейронних мереж та генетичні алгоритми	12
6.	Тема 6. Дерева рішень як метод дейтамайнінгу	12
7.	Тема 7. <u>Сховища даних та оперативний аналіз даних (OLAP)</u>	12
	Разом	100

8. Контрольні запитання, комплекти тестів для визначення рівня засвоєння знань здобувачів

1. Сутність поняття Data Mining
2. Область застосування і задачі Data Mining
3. Методи і моделі Data Mining
4. Поняття Business Intelligence
5. Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання.
6. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором і корекцією моделей.
7. Методи первісної обробки даних.
8. Візуалізація та автоматичне групування даних.
9. Постановка задачі класифікації та представлення результатів
10. Методи побудови правил класифікації.
11. Методи побудови математичних функцій
12. Класифікація об'єктів у випадку невідомих розподілень даних.
13. Методи оцінювання помилок класифікації
14. Методи вирішення задач регресії.
15. Поняття нечітких часових рядів
16. Методи моделювання часових рядів
17. Методи аналізу поведінки часових рядів.
18. Методи прогнозування поведінки часових рядів
19. Постановка задачі кластеризації та представлення результатів
20. Види кластерів
21. Міри близькості, засновані на відстаннях
22. Базові алгоритми кластеризації
23. Адаптивні методи кластеризації
24. Визначення та еволюція нейронних мереж
25. Математична модель штучного нейрона.
26. Програмне забезпечення нейромереж

27. Апаратне забезпечення нейромереж
28. Діапазон застосування нейромереж, їх переваги та недоліки
29. Сутність дерев рішень, особливості їх використання та побудови
30. Древа рішень для дослідження альтернатив.
31. Сутність класифікації альтернативних варіантів та методи її проведення.
32. Загальний опис дерев класифікації
33. Способи побудови класифікаційних дерев рішень.
34. Сутність поняття сховища даних
35. Використання даних сховищ
36. Архітектура сховища даних
37. ETL-процеси (добування, перетворення й завантаження даних).
38. Вітрини даних, кубики даних, багатомірна модель даних
39. Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP..
40. Основні властивості пакету аналізу R.
41. Статистична обробка даних в R.

Тести для визначення рівня засвоєння знань аспірантами

Питання 1. Розкрийте суть концепції дейтамайнінгу	
Питання 2. Розкрийте суть інтелектуального аналізу часових рядів	
Питання 3. До яких рядів динаміки належать показники, що характеризують розміри явища за певні проміжки часу?	
1	Моментні
2	Дискретні
3	Інтервальні
4	Ряди середніх
Питання 4. Для застосування кореляційного аналізу потрібні наявні передумови	
1	Достатня варіація досліджуваної ознаки
2	Числовий вираз ознаки
3	Достатня однорідність досліджуваної сукупності
4	Неоднорідність досліджуваної сукупності
Питання 5. Древа рішень відносяться до групи ...	
1	кібернетичних методів
2	логічних методів
3	статистичних методів
4	методів крос-табуляції
Питання 6. Одним з алгоритмів навчання нейронної мережі, які найчастіше використовуються є алгоритм...	
1	прямого розповсюдження помилки
2	зворотного розповсюдження помилки
3	диференційний

4	лінійний
Питання 7. Data Mining - це процес виявлення в сирих даних	
1	неочевидних закономірностей
2	раніше сформульованих гіпотез
3	великої кількості закономірностей
4	об'єктивних закономірностей
Питання 8 До статистичних методів Data Mining відносять:	
1	Аналіз зав'язків
2	Дерева рішень;
3	Аналіз часових рядів
4	Кластерний аналіз
5	Штучні нейронні мережі
Питання 9. Вектор у середовищі R можна створити за допомогою функції:	
1	#
2	c
3	sum
4	length
5	<-
Питання 10. Для виведення даних описової статистики (сума, дисперсія тощо) у середовищі R використовується функція:	
1	Var()
2	Sd()
3	Summary()
4	View()
Питання 11. Для побудови лінійної регресійної моделі у середовищі R використовується функція:	
1	lm()
2	str()
3	plot()
4	fit()
5	predict()
Питання 12. Щоб змінити робочу директорію у середовищі R використовується функція:	
1	getwd ()
2	setwd ()
3	ls()
4	read.table ()

9. Методи навчання

Методами навчання є способи спільної діяльності й спілкування викладача і здобувачів освітнього ступеня доктор філософії, що забезпечують вироблення позитивної мотивації навчання, оволодіння системою професійних знань, умінь і навичок, формування наукового світогляду, розвиток пізнавальних сил, культури розумової праці майбутніх фахівців.

Під час навчального процесу використовуються наступні методи навчання:

Залежно від джерела знань: словесні (пояснення, бесіда, дискусія, діалог); наочні (демонстрація, ілюстрація); практичні (рішення задач, ділові ігри).

За характером пізнавальної діяльності: пояснювально-наочний проблемний виклад; частково-пошуковий та дослідницький методи.

За місцем в навчальній діяльності:

- методи організації й здійснення навчальної діяльності, що поєднують словесні, наочні і практичні методи; репродуктивні й проблемно-пошукові; методи навчальної роботи під керівництвом викладача й методи самостійної роботи здобувачів освітнього ступеня доктор філософії;

- методи контролю й самоконтролю за навчальною діяльністю: методи усного, письмового контролю; індивідуального й фронтального, тематичного і систематичного контролю.

У процесі викладання навчальної дисципліни для активізації навчально-пізнавальної діяльності здобувачів освітнього ступеня передбачено застосування таких навчальних технологій:

- *робота в малих групах* дає змогу структурувати практичні заняття за формою і змістом, створює можливості для участі кожного аспіранта в роботі за темою заняття, забезпечує формування особистісних якостей та досвіду соціального спілкування;

- *дискусії* передбачають обмін думками і поглядами учасників з приводу даної теми, а також розвивають мислення, допомагають формувати погляди і переконання, вміння формулювати думки й висловлювати їх, вчать оцінювати пропозиції інших колег, критично підходити до власних поглядів;

- *мозкові атаки* – метод розв'язання невідкладних завдань, сутність якого полягає в тому, щоб висловити як найбільшу кількість ідей за обмежений проміжок часу, обговорити і здійснити їх селекцію;

- *кейс-метод* – метод аналізу конкретних ситуацій/проблем, який дає змогу наблизити процес навчання до реальної практичної діяльності фахівців і передбачає розгляд виробничих, управлінських та інших бізнесових ситуацій/проблем, складних конфліктних випадків, інцидентів у процесі вивчення конкретного навчального матеріалу;

- *презентації* – виступи перед аудиторією, що використовуються для подання певних досягнень, результатів вивчення матеріалу, звіту про виконання самостійної роботи.

10. Форми контролю

Відповідно до «Положення про екзамени та заліки у здобувачів вищої освіти ступеня доктор філософії в Національному університеті біоресурсів і природокористування України» затвердженого вченою радою НУБіП України 29 травня 2020 року, протокол № 10, видами контролю знань здобувачів вищої освіти ступеня доктор філософії є поточний контроль, проміжна та підсумкова атестації.

Поточний контроль здійснюється під час проведення практичних занять і має на меті перевірку рівня підготовленості здобувачів вищої освіти до виконання конкретної роботи.

Проміжна атестація проводиться після вивчення програмного матеріалу і має визначити рівень знань здобувачів вищої освіти з програмного матеріалу, отриманих під час усіх видів занять і самостійної роботи.

Форми та методи проведення проміжної атестації, засвоєння програмного матеріалу розробляються лектором дисципліни і затверджуються відповідною кафедрою у вигляді тестування, письмової контрольної роботи, колоквіуму, результату експерименту, що можна оцінити чисельно, розрахункової чи розрахунково-графічної роботи тощо.

Засвоєння аспірантом програмного матеріалу вважається успішним, якщо рейтингова оцінка його становить не менше, ніж 60 балів за 100-бальною шкалою.

Атестація проводиться у формі екзамену і підтверджує рівень засвоєння здобувачем вищої освіти теоретичного та практичного матеріалу з навчальної дисципліни.

Аспіранти зобов'язані складати екзамен відповідно до вимог навчального плану у термін, передбачений графіком освітнього процесу. Зміст екзамену визначається робочою програмою дисципліни.

11. Розподіл балів, які отримують здобувачі

Оцінювання знань здобувачів вищої освіти відбувається за 100-бальною шкалою і переводиться в національні оцінки згідно з табл. 1 «Положення про екзамени та заліки у здобувачів вищої освіти ступеня доктор філософії в Національному університеті біоресурсів і природокористування України», затвердженого Вченою радою НУБіП України від 29.05.2020 р., протокол №10.

Для визначення рейтингу здобувача вищої освіти із засвоєння дисципліни $R_{\text{дис}}$ (до 100 балів) одержаний рейтинг з атестації (до 30 балів)

додається до рейтингу з навчальної роботи $R_{НР}$ (до 70 балів): $R_{ДИС} = R_{НР} + R_{АТ}$

Таблиця 1. Співвідношення між рейтингом здобувача вищої освіти і національними оцінками

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

12. Методичне забезпечення

1. Воляк Л.Р. Електронний навчальний курс “Дейтамайнінг» <https://elearn.nubip.edu.ua/enrol/index.php?id=2900>
2. Навчально-наукова лабораторія біоеконометрики та дейтамайнінгу.

13. Рекомендовна література

Базова

1. Акіменко В.В., Загородній Ю.В. Проектування СППР на основі нечіткої логіки: Навч.- методич. посіб. Київ: КНУ, 2007. 94 с.
2. Дэви С. Основы Data Science и Big Data. Python и наука о данных. Санкт-Петербург: Питер, 2017. 336 с.
3. Ситник В.Ф., Краснюк М.Т. Ітелектуальний аналіз даних (дейтамайнінг): навч. посіб. Київ: КНЕУ, 2007. 376 с.
- Хайкин С. Нейронные сети: полный курс, 2-е изд., пер. с англ. Москва: Вильямс, 2006. 1104 с.
4. Уайт Т. Nadoor: Подробное руководство: Санкт-Петербург: Питер, 2013. 672 с.
5. Черняк О.І., Захарченко П.В. Ітелектуальний аналіз даних: підручник. Київ: Знання, 2014. 599 с.

Допоміжна

1. Грінгарт Семюель. Інтернет речей. пер. з англ. Л. Герасимчука. Харків: КК, 2018. 175 с.

2. Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим : пер. с англ. Инны Гайдюк. Москва : Манн, Иванов и Фербер, 2014. 240 с.

3. Наглядная статистика. Используем R. /А.Б.Шипунов и др. Москва: ДМК Пресс, 2012. 298 с.

4. Чак Л. Надоор в действии. Москва: ДМК Пресс, 2012. 424 с.

5. Чубукова И. А. Data Mining: учебн. пособ. Москва: Интернет университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. 382 с.

Інформаційні ресурси

1. Електронна бібліотека НУБіП України. URL: <https://nubip.edu.ua/structure/library>

1. Великі дані. URL: <https://uk.wikipedia.org/wiki/>

2. Big Data for Development: From Information- to Knowledge Societies. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145

3. Big Data от А до Я. URL: <https://habrahabr.ru/company/dca/blog/267361/>

4. Big Data и блокчейн. URL: <https://forklog.com/big-data-i-blokchejn-proryv-v-oblasti-analiza-dannyh>

