

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

Кафедра інформаційних систем і технологій

“ЗАТВЕРДЖУЮ”

Декан факультету інформаційних
технологій

_____ Глазунова О.Г.

“ ____ ” _____ 2020 р.

РОЗГЛЯНУТО І СХВАЛЕНО

на засіданні кафедри інформаційних
систем і технологій

Протокол № 9 від “22” квітня 2020 р.

Завідувач кафедри

_____ Швиденко М.З.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»

галузь знань	<u>05 «Соціальні та поведінкові науки»</u>
спеціальність	051 «Економіка»
освітня програма	«Економічна кібернетика»
факультет	Інформаційних технологій
розробники:	<u>доц. к.е.н., Харченко В.В.</u>

Київ – 2020 р.

1. Опис навчальної дисципліни «Інтелектуальний аналіз даних»

Галузь знань, напрям підготовки, спеціальність, освітній ступінь		
Освітній ступінь	Магістр	
Галузь знань	05 «Соціальні та поведінкові науки»	
Спеціальність	051 «Економіка»	
Освітня програма	«Економічна кібернетика»	
Характеристика навчальної дисципліни		
Вид	Нормативна	
Загальна кількість годин	120 год.	
Кількість кредитів ECTS	4 ECTS	
Кількість змістових модулів	2	
Курсовий проект (робота) (за наявності)	-	
Форма контролю	Іспит	
Показники навчальної дисципліни для денної та заочної форм навчання		
	денна форма навчання	заочна форма навчання
Рік підготовки (курс)	1	
Семестр	1	
Лекційні заняття	30 год.	
Практичні, семінарські заняття		
Лабораторні заняття	30 год.	
Самостійна робота	60 год.	
Індивідуальні завдання		
Кількість тижневих аудиторних годин для денної форми навчання	12 год.	

2. Мета та завдання навчальної дисципліни

Дисципліна «Інтелектуальний аналіз даних» належить до спеціальних дисциплін, які забезпечують формування знань та навиків магістрів економічних спеціальностей по впровадженню комп'ютерних технологій у бізнесі, основ бізнес аналітики та її використанню у прийнятті управлінських рішень.

Мета: набуття теоретичних і практичних знань з основ інтелектуального аналізу даних.

Завдання курсу:

- оволодіння основними поняттями інтелектуального аналізу даних;
- ознайомлення з новітніми інформаційними технологіями аналітичної обробки первинної інформації;
- набуття практичних навичок по використанню інтелектуального аналізу даних у економіці.

Для вивчення дисципліни необхідні знання з основ інформатики, системного програмного забезпечення, офісного програмного забезпечення, систем управління базами даних, комп'ютерних мереж.

Засвоєння матеріалу забезпечується на лекціях, лабораторно-практичних заняттях та самостійній роботі у комп'ютерних класах, обладнаних локальними мережами, Інтернет і новітнім програмним забезпеченням. При викладанні дисципліни використовуються активні методи навчання, системних підхід, модульрейтингова система контролю навчання студентів.

У результаті вивчення навчальної дисципліни студент повинен:

знати: основні поняття інтелектуального аналізу даних; основні математичні моделі, що лежать в основі інтелектуального аналізу даних.

вміти: використовувати програмне забезпечення для використання процедур інтелектуального аналізу при обробці та аналізу первинної економічної інформації.

Навчальна дисципліна забезпечує формування ряду фахових 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.

У результаті вивчення навчальної дисципліни студент набуде певні програмні результати, а саме:

ПР 8. Збирати, обробляти та аналізувати статистичні дані, науково-аналітичні матеріали, необхідні для вирішення комплексних економічних завдань.

3. Програма та структура навчальної дисципліни для повного терміну денної форми навчання.

ЗМІСТОВИЙ МОДУЛЬ I. Введення в інтелектуальний аналіз даних

Тема 1. Введення в інтелектуальний аналіз даних.

Визначення інтелектуального аналізу даних (Data Mining), історія та передумови виникнення. Data Mining і витяг знань з даних (Knowledge Discovery in Databases). Місце і роль Data Mining в процесі прийняття рішень. Основні завдання Data Mining, види моделей. Data Mining і статистичний аналіз. Data Mining і сховища даних. Data Mining і OLAP. Data Mining як самостійна галузь індустрії програмного забезпечення.

Тема 2. Методи інтелектуального аналізу даних.

Класи задач (типи закономірностей), вирішуються за допомогою Data Mining: класифікація, кластеризація, регресія (прогнозування), асоціація, послідовні шаблони. Методи отримання і аналізу даних: описові - дисперсійний та регресійний аналіз; аналіз часових рядів, кластерний аналіз, еволюційні - дерева рішень, генетичні алгоритми; штучні нейронні мережі, метод «найближчих сусідів».

Тема 3. Інтелектуальний аналіз даних в СУБД MicrosoftSQLServer.

Засоби інтелектуального аналізу даних в Analysis Services: набір стандартних алгоритмів інтелектуального аналізу даних; конструктор інтелектуального аналізу даних, призначений для створення і перегляду моделей інтелектуального аналізу даних, керування ними та побудови прогнозів; розширювану мову інтелектуального аналізу даних (Data MiningExtensions to SQL, DMX).

Тема 4. Етапи проведення інтелектуального аналізу даних.

Етапи побудови моделі методами Data Mining: Постановка завдання, ідентифікація бізнес-проблеми, типові «вузькі місця» бізнесу. Підготовка даних, визначення джерел даних для аналізу, вибір, очищення та попередня обробка даних. Побудова моделей із зазначенням алгоритмів інтелектуального аналізу

даних і їх параметри. Перевірка моделі: оцінка якості роботи створеної моделі перед початком її використання в "виробничому середовищі".

ЗМІСТОВИЙ МОДУЛЬ II. Практичні аспекти вирішення задач інтелектуального аналізу даних

Тема 5. Алгоритм дерева рішень.

Методи дерева рішень. Опис графа набору даних (вузли, гілки, листя). Основні алгоритми побудови Дерева рішень у випадку двох можливих дискретних станів результуючої змінної. «Ліси» рішень. Зростаючі «дерева» (Boosted trees) Випадкові «ліси» (Random forests).

Тема 6. Лінійні та нелінійні регресійні моделі

Статистичні методи обробки даних. Оцінювання параметрів розподілу; перевірка статистичних гіпотез; дисперсійний та регресійний аналіз; аналіз часових рядів. Класична модель ARIMA (АРПСС). Експоненційне згладжування з сезонними компонентами. Спектральне розкладання Фур'є. Поліномінальний і регресійний аналіз лагів.

Тема 7. Кластерний аналіз

Основні завдання кластерного аналізу: розробка типології або класифікації, дослідження концептуальних схем групування об'єктів, висунення гіпотез на основі дослідження даних, перевірка гіпотез. Основні етапи кластерного аналізу. Алгоритми кластеризації: м'яка і тверда кластеризація. Області застосування кластерного аналізу.

Тема 8. Нейронні мережі. Практичні аспекти вирішення задач інтелектуального аналізу даних.

Визначення та характеристика нейронних мереж. Область їх застосування. Класифікація нейромережових архітектур. Моделі нейронних мереж. Класифікація методів навчання. Особливості сучасних нейронних мереж.

Специфічні галузеві рішення Data Mining на практиці. Бізнес-завдання: в умовах жорсткої конкуренції. Застосування Data Mining при дослідженні задач: Роздрібна торгівля, Банківська справа. Телекомунікації. Страхування. Інші застосування в бізнесі

Структура навчальної дисципліни «Інтелектуальний аналіз даних»

Назви змістових модулів і тем	Кількість годин												
	Денна форма							Заочна форма					
	тижні	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
Змістовий модуль 1. Введення в інтелектуальний аналіз даних													
Тема 1. Введення в інтелектуальний аналіз даних	1	11	2		2		7						
Тема 2. Методи інтелектуального аналізу даних	1-2	16	4		4		8						
Тема 3. Інтелектуальний аналіз даних в СУБД MicrosoftSQLServer	2-3	15	4		4		7						
Тема 4. Етапи проведення інтелектуального аналізу даних	2-3	16	4		4		8						
Разом за змістовим модулем 1		58	14		14		30						
Змістовий модуль 2. Практичні аспекти вирішення задач інтелектуального аналізу даних													
Тема 5. Алгоритм дерева рішень	3-4	15	4		4		7						
Тема 6. Лінійні та нелінійні регресійні моделі	3-4	16	4		4		8						
Тема 7. Кластерний аналіз	4-5	15	4		4		7						
Тема 8. Нейронні мережі. Практичні аспекти вирішення задач інтелектуального аналізу даних	5	16	4		4		8						
Разом за змістовим модулем 2		62	16		16		30						
Усього годин		120	30		30		60						

4. Теми семінарських занять Не передбачені навчальним планом.

5. Теми практичних занять Не передбачені навчальним планом.

6. Теми лабораторних робіт

№ з/п	Назва теми	Кількість годин
1	Використання інструментів "AnalyzeKeyInfluencers" і "DetectCategories" надбудови інтелектуального аналізу даних для MicrosoftOffice.	4
2	Використання інструментів "FillFromExample" і "Forecast"	4
3	Використання інструментів "HighlightExceptions" і "ScenarioAnalysis"	4
4	Використання інструментів "Prediction Calculator" і "ShoppingbasketAnalysis"	4
5	Використання інструментів Data Mining Client для Excel 2007 для підготовки даних	2
6	Використання інструментів Data Mining Client для Excel 2007 для створення моделі інтелектуального аналізу даних	4
7	Аналіз точності прогнозу і використання моделі інтелектуального аналізу	4
8	Побудова моделі кластеризації, трасування і перехресна перевірка	4
Всього		30

7. Контрольні питання, комплекти тестів для визначення рівня засвоєння знань студентами.

Перелік контрольних питань

1. Data Mining як самостійна галузь індустрії програмного забезпечення.
2. Місце і роль Data Mining в процесі прийняття рішень.
3. Методи отримання і аналізу даних.
4. Поняття дерева рішень.
5. Що таке генетичні алгоритми?
6. Програмна мова інтелектуального аналізу даних.
7. Методи оцінки якості роботи створеної моделі інтелектуального аналізу.
8. Особливості дерева рішень у випадку двох можливих дискретних станів результуючої змінної.

9. Експоненційне згладжування з сезонними компонентами.
10. Спектральне розкладання Фур'є
11. Специфічні галузеві рішення Data Mining на практиці.
12. Застосування засобів інтелектуального аналізу в АПК
13. Суть системи діалогової обробки запитів
14. Сфери застосування сховищ даних.
15. В чому різниця OLAP і OLTP.
16. Класи продуктів Business Intelligence.
17. Інструментальні засоби бізнесу-інтелекту.
18. Алгоритм Байєса.
19. Часові ряди в економіці.
20. Поняття кластеризації.
21. Алгоритми кластеризації.
22. Особливості налаштування SQL сервера для інтелектуального аналізу.
23. Параметри точності прогнозу.
24. Методи трасування і перехресної перевірки.
25. Історія нейронних мереж
26. Типи регресійних моделей. Особливості.
27. Алгоритм дерева рішень C4.5.
28. Застосування Data mining в економіці.
29. Місце аналітика в процесі прийняття управлінських рішень.
30. Інтелектуальний аналіз даних на основі відкритого програмного забезпечення.

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ			
ОС Магістр Освітня програма “Економічна кібернетика”	Кафедра інформаційних систем і технологій 2020-2021 навч. рік	ЕКЗАМЕНАЦІЙНИЙ БІЛЕТ № 1 з дисципліни “Інтелектуальний аналіз даних”	Затверджую Зав. кафедри Швиденко Михайло Зіновійович _____ р.
Екзаменаційні запитання			
1. Спрощений алгоритм Байєса.			
2. Опишіть процес проведення інтелектуального аналізу даних.			

10 тестових завдань

1. Data Mining –

1	грубий аналіз окремих атрибутів даних (тип, довжина, спектр значень, дискретні значення і їх частота, унікальність, наявність null-значень)
2	виконує угруповання, узагальнення, пошук асоціацій, послідовностей, тобто допомагає знайти специфічні моделі у великих наборах даних
3	виявлення і видалення помилок і невідповідностей в даних з метою поліпшення їх якостей
4	вивантаження даних засобами oltp-систем в проміжні структури

2. Інструментальні засоби бізнесу-інтелекту

1	програмне забезпечення, що дає можливість користувачам спостерігати й використати більші обсяги складних даних
2	програмне забезпечення, що дає користувачеві можливість спостерігати дані в різних вимірах,
3	це інтерактивна, комп'ютерна система, яка полегшує колективне вирішення неструктурованих
4	програмне забезпечення, що дозволяє формувати запити до даних по змісту або зразку

3. Коректність використання алгоритму Байєса можлива за умов (відмітити зайве)

1	вхідні атрибути повинні бути взаємно незалежними;
2	атрибути можуть бути тільки дискретними або дискретизованими(в процесі дискретизації безліч значень безперервного числового атрибута розбивається на інтервали і далі йде робота з номером інтервалу);
3	алгоритм вимагає меншої кількості обчислень, ніж інші алгоритми інтелектуального аналізу, що подаються Microsoft SQLServer 2008, тому він часто використовується для первинного дослідження даних. З тієї ж причини, даний алгоритм кращий для аналізу великих наборів даних з великою кількістю вхідних атрибутів.
4	всі вірні

4. Алгоритм що дозволяє виявити поєднання елементів даних що часто зустрічаються і використовувати виявлені закономірності для побудови прогнозу

(у бланку відповідей впишіть вірну відповідь одним словом)

5. Алгоритм класифікації, заснований на обчисленні умовної ймовірності значень прогнозованих атрибутів —

1	кластеризація
2	класифікація
3	Байєса
4	лінійна регресія

6. Ймовірність задана до дослідження

1	априорна
2	незалежна
3	несумісна
4	апостеріорна

7. Набір числових значень, зібраних в послідовні моменти часу –	
1	транзакції
2	тимчасові ряди
3	часові ряди
4	неперервні ряди
8. Поставити у відповідність категорії прогнозів	
А. істинний позитивний прогноз	1. клієнт, для якого прогноз показав істину, насправді зацікавлений в покупці велосипеда
Б. істинний негативний прогноз	2. клієнт, для якого прогноз показав незацікавленість у купівлі, насправді не збирається купувати велосипед
В. помилковий позитивний прогноз	3. клієнт хоче зробити покупку, хоча насправді це не так
Г. помилковий негативний прогноз	4. клієнт не хоче зробити покупку, хоча насправді він в ній зацікавлений
9. На основі даних, що обробляються, формуються шаблони, які при використанні дозволяють віднести розглянутий приклад до однієї з можливих груп (вказати модель інтелектуального аналізу)	
1	Класифікація (Classify)
2	Кластер (Cluster)
3	Пошук взаємозв'язків (Associate)
4	Оцінка (Estimate)
10. Виберіть правильне визначення	
1	Організаційне забезпечення — сукупність організаційних, методичних та технологічних документів, що регламентують процес людино-машинного оброблення інформації в АІС
2	Інформаційне забезпечення — сукупність програм, які реалізують мету та задачі АІС і забезпечують функціонування комплексу технічних засобів системи
3	Лінгвістичне забезпечення — сукупність документів, що регламентують діяльність персоналу в АІС, взаємодію з технічними засобами і між собою в процесі розв'язування задач управління
4	Ергономічне забезпечення — сукупність методів і засобів, призначених для створення оптимальних умов високоефективної та безпомилкової діяльності людини в АІС і найшвидшого її освоєння.

8. Методи навчання.

Пояснювально-ілюстративний, репродуктивний, проблемне викладання, частково-пошуковий, дослідницький.

9. Форми контролю.

Модульний контроль, поточний контроль, підсумковий контроль.

Розподіл балів, які отримують студенти. Оцінювання студента відбувається згідно положенням «Про екзамени та заліки у НУБіП України» від 27.02.2019 р. протокол № 7 з табл. 1.

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

Оцінка національна	Рейтинг здобувача вищої освіти, бали
Відмінно	90 – 100
Добре	74 – 89
Задовільно	60 – 73
Незадовільно	0 – 59

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

10. Методичне забезпечення

1. Інтелектуальний аналіз даних: Підручник / Черняк О.І., Захарченко П.В./ К.: Знання, 2014 р. - 599 с.

11. Рекомендована література

Основні джерела:

1. Джеми Макленнен, Чжаохуэй Танг, Богдан Криват. Microsoft SQL Server 2008 :Datamining - интеллектуальный анализ данных СПб. : БХВ-Петербург, 2009
2. Кондрат Карлберг. Бизнес-анализ с помощью Excel.: Пер.с англ. – К.: Диалектика, 1997. – 448с.: ил.
3. Чубукова И. А. Data Mining: учебное пособие. — М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. — 382 с

Додаткові друковані джерела:

1. А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д.Тесс, С.И. Елизаров Анализ данных и процессов СПб.: БХВ-Петербург, 2009
2. Г.А.Титаренко. Автоматизированные информационные технологии в экономике. М.: Компьютер ЮНИТИ, 1998. – с.

3. Інформаційні системи і технології в економіці: Посібник для студентів вищих навчальних закладів/ За редакцією В.С. Пономаренка. – К.: Видавничий центр "Академія", 2002. – 544с.
4. Круглов В.В., Дли М.И. Интеллектуальные информационные системы: компьютерная поддержка систем нечеткой логики и нечеткого вывода. – М.: Физматлит, 2002.
5. Матвеев Л.А. Компьютерная поддержка решений: Учебник – СПб: "Специальная литература", 1998. – 472 с.
6. Матвеев Л.А. Компьютерная поддержка решений: Учебник – СПб: «Специальная Литература», 1998. – 472 с.
7. Плакунов М.К. Планирование на малых и средних предприятиях средствами EXCEL. – СПб.: Питер, 2004. – 156с.
8. Решение экономических задач на компьютере / Каплан А.В., Каплан В.Е., Мащенко М.В., Овечкина Е.В. – М.: ДМК Пресс; СПб.: Питер, 2004.- 600с.
9. Системи підтримки прийняття рішень: Навчальний посібник/ О.І.Пушкар, В.М.Гірковатий, О.С.Євсєєв, Л.В.Потрашкова; За ред. д-ра екон. наук, проф. Пушкаря О.І.-Х.:ВД "ІНЖЕК", 2006.-304 с.

13. Інформаційні ресурси

1. Електронні навчальні курси створені в системі дистанційного навчання MOODLE , адреса – <http://elearn.nubip.edu.ua>)
2. Службы AnalysisServices — видеоролики по интеллектуальному анализу данных (SQL Server 2008)
3. Службы SQLServerAnalysisServices — интеллектуальный анализ данных
4. Технический справочник по алгоритму взаимосвязей (Майкрософт)
5. Технический справочник по алгоритму кластеризации (Майкрософт)
6. Технический справочник по алгоритму логистической регрессии (Майкрософт)