

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ,
МОЛОДІ ТА СПОРТУ УКРАЇНИ**

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

**Робоча програма
навчальної дисципліни
"ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ"
для студентів спеціальності
8.000011 "Прикладна економіка"
всіх форм навчання**

Харків. Вид. ХНЕУ, 2012

Затверджено на засіданні кафедри економічної кібернетики.
Протокол № 2 від 08.09.2011 р.

Укладачі: Мілов О. В.
Мілевський С. В.

P58 Робоча програма навчальної дисципліни "Інтелектуальний аналіз даних" для студентів спеціальності 8.000011 "Прикладна економіка" всіх форм навчання / укл. Мілов О. В., Мілевський С. В. – Х. : Вид. ХНЕУ, 2012. – 32 с. (Укр. мов.)

Подано тематичний план навчальної дисципліни та її зміст за модулями й темами, наведено плани лекцій, лабораторних занять, роботи щодо закріплення знань та систему оцінювання студентів.

Рекомендовано для студентів спеціальності 8.000011 "Прикладна економіка".

Вступ

Дисципліна "Інтелектуальний аналіз даних" знайомить студентів з технологією Data Mining, докладно розглядаються методи, інструментальні засоби й застосування Data Mining. Опис кожного методу супроводжується конкретним прикладом його використання.

Обговорюються відмінності Data Mining від класичних статистичних методів аналізу й OLAP-систем, розглядаються типи закономірностей, що виявляють Data Mining (асоціація, класифікація, послідовність, кластеризація, прогнозування). Описується сфера застосування Data Mining. Вводиться поняття Web Mining. Докладно розглядаються методи Data Mining: нейронні мережі, дерева рішень, методи обмеженого перебору, генетичні алгоритми, еволюційне програмування, кластерні моделі, комбіновані методи. Знайомство з кожним методом проілюстровано рішенням практичного завдання за допомогою інструментального засобу, що використовує технологію Data Mining. Викладаються основні концепції сховищ даних і місця Data Mining у їхній архітектурі. Вводяться поняття OLTP, OLAP, ROLAP, MOLAP. Обговорюється процес аналізу даних за допомогою технології Data Mining. Докладно розглядаються етапи цього процесу. Аналізується ринок аналітичного програмного забезпечення, описуються продукти від провідних виробників Data Mining, обговорюються їхні можливості.

Методи, що викладаються, були реалізовані на ЕОМ і пройшли практичну апробацію в різних сферах економіки і управління. Теоретичний матеріал підкріплений практичними прикладами, що дозволяють краще засвоїти матеріал, що викладається.

Розглянуті в дисципліні системи доцільно використовувати для рішення задач соціально-економічного прогнозування і планування розвитку промислових галузей, підприємств і в інших службах, що утворюють інфраструктуру міст, областей і регіонів.

Навчальна дисципліна "Інтелектуальний аналіз даних" базується на знаннях, одержаних при вивченні дисциплін математичного циклу (інформатики, основ теорії баз даних, технології обробки інформації).

Наукову основу дисципліни складають теоретичні моделі, математичний апарат, сучасні концепції та парадигми, які визначають підходи до вивчення характеристик систем інтелектуального аналізу даних.

Програмою дисципліни передбачено проведення лекційних та ла-

бораторних занять. Лабораторні заняття базуються на матеріалах лекцій, включають розгляд конкретних об'єктів і систем та побудову математичних моделей інтелектуального аналізу даних.

Самостійна робота студентів передбачає підготовку до лабораторних занять, проведення індивідуальних консультацій.

Матеріал дисципліни "Інтелектуальний аналіз даних" використовується у викладенні дисциплін "Багатомірний статистичний аналіз", "Прогнозування в соціально-економічних системах", а також при виконанні дипломних робіт магістра.

Структуру навчальної дисципліни наведено у табл. 1.

Таблиця 1

Структура навчальної дисципліни

Навчальна дисципліна: підготовка магістрів	Галузь знань, спеціальність, освітньо- кваліфікаційний рівень	Характеристика навчальної дисципліни
Кількість кредитів відповідних ECTS – 3; у тому числі: змістовних модулів – 1; самостійна робота; ІНДЗ	Шифр та галузь знань: 1801 "Специфічні категорії"	Виборкова. Рік підготовки: 5. Семестр: 9
Кількість годин: усього – 108; за змістовними модулями: модуль 1 – 108 годин	Шифр та назва спеціальності: 8.000011 "Прикладна економіка"	Лекції: кількість годин – 15. Лабораторні: кількість годин – 15. Самостійна робота: кількість годин – 78, (у тому числі індивідуальна робота 18 годин)
Кількість тижнів викладення навчальної дисципліни – 15. Кількість годин за тиждень – 2	Освітньо-кваліфікаційний рівень: магістр	Вид контролю: ПМК

1. Кваліфікаційні вимоги до студентів

У процесі навчання студенти отримують необхідні знання під час лекційних занять і виконуючи лабораторні роботи. Також велике значення в процесі вивчення та закріплення знань має самостійна робота сту-

дентів та виконання індивідуальних завдань. Усі види занять розроблено відповідно до вимог кредитно-модульної системи організації навчального процесу.

В результаті вивчення дисципліни студент повинен:

мати уявлення:

- про роль та цілі застосування комп'ютеризованого аналізу й інтерпретації даних у пізнанні навколишнього світу та вирішенні практичних завдань;
- про прикладні методи аналізу даних і видобутку знань, основних напрямках та тенденцій їхнього розвитку;
- про можливості й перспективи застосування методів інтелектуального аналізу даних у дослідженні об'єктів і процесів професійної діяльності в області управління та організації;
- про організацію й застосування сучасних інформаційних технологій і програмних систем для обробки експериментальних і статистичних даних;

знати:

- предмет та об'єкти вивчення, структуру даного курсу, сутність використаних у ньому основних методів досліджень, його місце в навчальному плані для підготовки дипломованих фахівців з даного напрямку;
- базові визначення й поняття, основні етапи аналізу та інтерпретації даних, загальну структуру і функціонування інтегрованих інформаційно-аналітичних систем (систем підтримки прийняття рішень);
- основи організації й технології обробки експериментальних і статистичних даних на ЕОМ;
- змістовну й математичну постановку основних завдань аналізу даних, способи їхнього розв'язання;
- сутність і властивості основних алгоритмів інтелектуального аналізу;
- приклади рішення практичних завдань аналізу даних;

вміти:

- зводити словесні постановки завдань до типових математичних й відносити їх до відповідних розділів математики та зв'язувати з відомими засобами обробки експериментальних даних;
- створювати і користуватися типовими математичними моделями для дослідження випадкових явищ та процесів за результатами спостережень при розв'язанні завдань: порівняння, виявлення й відновлення закономірностей, класифікації, прогнозування;
- ставити завдання, вибирати засоби й методи обробки експериментальних даних для ефективного застосування;

- представляти результати аналізу в зручному для сприйняття вигляді, інтерпретувати їх відповідно до поставленого завдання;
- здійснювати самооцінку й самоконтроль при розв'язанні завдань обробки статистичних даних.

Компетенції, які формуються навчальною дисципліною "Інтелектуальний аналіз даних", наведено в табл. 2.

Таблиця 2

Компетенції, які формуються навчальною дисципліною

Назва компетенції	Зміст компетенції	Вміння студента щодо даної компетенції
1. Аналітична	1.1. Здатність до проведення аналізу поставлених проблем	1.1.1. Визначити чинники, що впливають на результати аналізу. 1.1.2. Прогнозувати найімовірніший перебіг процесів на перспективу та аналізувати й оцінювати можливі наслідки зміни явищ
2. Організаційно-методологічна	2.1. Здатність упроваджувати сучасні методи інтелектуального аналізу даних	2.1.1. Обґрунтовано вибирати відповідний метод аналізу. 2.1.2. Володіти методами інтелектуального аналізу даних
3. Контрольна	3.1. Здатність перевіряти якість та практичність отриманих результатів аналізу	3.1.1. Готувати інформацію для аналізу, здійснювати розрахунки щодо аналізу та перевіряти їх на відповідність плановим завданням. 3.1.2. Використовувати відповідні критерії для аналізу рівня достовірності отриманих оцінок

2. Тематичний план навчальної дисципліни

На початку вивчення дисципліни кожен студент має бути ознайомлений як з програмою дисципліни і формами організації навчання, так і зі структурою, змістом та обсягом кожного з її навчальних модулів, а також з усіма видами контролю та методикою оцінювання навчальної роботи.

Навчальний процес згідно з програмою навчальної дисципліни "Інтелектуальний аналіз даних" здійснюється у таких формах: лекційні та лабораторні заняття; виконання індивідуального науково-дослідного завдання; самостійна робота студентів; контрольні заходи.

Вивчення студентом навчальної дисципліни відбувається шляхом послідовного і ґрунтовного опрацювання навчального модулю. Навчальний модуль – це відносно окремий самостійний блок дисципліни, який логічно об'єднує кілька навчальних елементів дисципліни за змістом та взаємозв'язками.

Тематичний план даної дисципліни складається з одного модулю (табл. 3).

Таблиця 3

Структура залікового кредиту навчальної дисципліни

Тема	Кількість годин на:			
	лекції	лабораторні заняття	самостійну роботу	індивідуальну роботу
Тема 1. Інтелектуальний аналіз даних – Data Mining	3	3	7	7
Тема 2. Візуальний аналіз даних – Visual Mining	2	2	6	6
Тема 3. Аналіз текстової інформації – Text Mining	2	2	5	5
Тема 4. Добування знань з Web – Web Mining	2	2	5	5
Тема 5. Засоби аналізу процесів – Process Mining	2	2	5	5
Тема 6. Пошук асоціативних правил – Rules Mining	2	2	5	5
Тема 7. Розподілений аналіз даних	2	2	6	6
Усього годин	15	15	39	39

3. Зміст навчальної дисципліни за темами

Тема 1. Інтелектуальний аналіз даних – Data Mining

Докладно розглядаються поняття Data Mining. Описано виникнення, перспективи, проблеми Data mining. Наведений погляд на технологію Data Mining як на частину ринку інформаційних технологій. Докладно розглядається поняття даних. Пояснюється значення понять "об'єкт" та

"атрибут", "вибірка", "залежна й незалежна змінна". Докладно обговорюються типи шкал. Приводяться різні типи наборів даних. Коротко розглянуті поняття бази даних і СУБД. Описуються стадії Data Mining та дії, які виконуються в рамках цих стадій. Розглянуто відомі класифікації методів Data Mining. Наведено порівняльну характеристику деяких методів, заснована на їхніх властивостях. Характеризується основна суть завдань Data Mining та їхня класифікація. Докладно розглянуті поняття "інформація", "знання", а також зіставлення й порівняння цих понять.

Тема 2. Візуальний аналіз даних – Visual Mining

Розглядаються питання візуального аналізу даних. Наведені характеристики засобів візуалізації даних, методів візуалізації та методів геометричних перетворень. Порівнюються методи, орієнтовані на пікселі, а також методи аналізу ієрархічних образів та відображення іконок. Розглядаються методи та засоби візуального подання інформації, зокрема, способи подання інформації в одне-, двох-, тривимірному вимірах, а також способи відображення інформації в більш ніж трьох вимірах. Описано принципи якісної візуалізації. Викладено основні тенденції в області візуалізації.

Тема 3. Аналіз текстової інформації – Text Mining

Формулюються задачі аналізу текстів (етапи аналізу текстів, попередня обробка тексту, задачі Text Mining). Розглядаються етапи аналізу текстів, такі як витяг ключових понять із тексту (загальний опис процесу витягу понять із тексту, стадія локального аналізу, стадія інтеграції й висновку понять), класифікація текстових документів (опис задач класифікації текстів, методи класифікації текстових документів), методи кластеризації текстових документів (наведення текстових документів, ієрархічні методи кластеризації текстів, бінарні методи кластеризації текстів), анотування текстів (виконання анотування текстів, методи витягу фрагментів для анотації). Порівнюються різноманітні засоби аналізу текстової інформації (засоби Oracle - Oracle Text, засоби від IBM - Intelligent Miner for Text, засоби SAS Institute - Text Miner, засоби Мега-комп'ютер Интеллидженс - TextAnalyst).

Тема 4. Добування знань з Web – Web Mining

Розглянуті проблеми аналізу інформації з Web, етапи Web Mining, Web Mining та інші Інтернет-технології, а також категорії Web Mining. Опи-

сані методи витягу Web-контента (витяг Web-контента в процесі інформаційного пошуку, витяг Web-контента для формування баз даних), а також методи витягу Web-структур (представлення Web-структур, оцінка важливості Web-структур, пошук Web-документів з урахуванням гіперпосилань, кластеризація Web-структур). Наведені результати досліджень використання Web-ресурсів (дослідницька інформація, етап препроцесінгу, етап витягу шаблонів, етап аналізу шаблонів та їхнє застосування).

Тема 5. Засоби аналізу процесів – Process Mining

Розглянуті засоби автоматизації виконання бізнес-процесів (бізнес-процеси, формалізація бізнес-процесів, Workflow-системи, сервісно-орієнтована архітектура, проектування бізнес-процесів). Виконаний аналіз процесів (технологія Process Mining, аналіз протоколів, стандарт запису протоколів MXML, задачі Process Mining, проблеми аналізу протоколів). Порівнюються методи Process Mining (перші імовірнісні методи Process Mining, метод побудови диз'юнктивної Workflow-схеми, α -алгоритм, методи на основі генетичних алгоритмів). Описана бібліотека алгоритмів Process Mining-Pro (архітектура Pro, Pro Import Framework).

Тема 6. Пошук асоціативних правил – Rules Mining

Виконана постановка задачі. Розглянуті форми подання результатів (правила класифікації, дерева класифікації, математичні функції), методи побудови правил класифікації (алгоритм побудови правил, метод Naive Bayes), а також методи побудови дерев класифікації (методика "розділяй і пануй", алгоритм покриття), методи побудови математичних функцій (загальний вид, лінійні методи, метод найменших квадратів, нелінійні методи, Support Vector Machines (SVM), регуляризаційні мережі (Regularization Networks), дискретизації й рідкі сітки). Розглянута постановка задачі пошуку асоціативних правил (формальна постановка задачі, секвенціальний аналіз, різновиди задач пошуку асоціативних правил), алгоритми (алгоритм Apriori, різновид алгоритму Apriori).

Тема 7. Розподілений аналіз даних

Розглядаються системи мобільних агентів (основні поняття, стандарти багатоагентних систем, системи мобільних агентів, система мобільних агентів JADE). Продемонстровано використання мобільних агентів для аналізу даних (проблеми розподіленого аналізу даних, агенти-аналітики, варіанти аналізу розподілених даних). Побудована система

аналізу розподілених даних (загальний підхід до реалізації системи, агент для збору інформації про базу даних, агент для збору статистичної інформації, агент для вирішення одного завдання інтелектуального аналізу даних, агент для вирішення інтегрованого завдання інтелектуального аналізу даних).

4. Плани лекцій

Тема 1. Інтелектуальний аналіз даних

1. Видобуток даних – Data Mining.
2. Задачі Data Mining.
3. Класифікація задач Data Mining.
4. Задача класифікації й регресії (задача пошуку асоціативних правил, задача кластеризації).
5. Практичне застосування Data Mining (інтернет-технології, торгівля, телекомунікації, промислове виробництво, медицина, банківська справа, страховий бізнес, інші області застосування).
6. Моделі Data Mining (прогнозні моделі, описові моделі).
7. Методи Data Mining (базові методи, нечітка логіка, генетичні алгоритми, нейронні мережі).
8. Процес виявлення знань (основні етапи аналізу, підготовка вихідних даних).
9. Управління знаннями (Knowledge Management).
10. Засоби Data Mining.

Література: основна [1; 12; 13; 18]; додаткова [1; 5; 8].

Тема 2. Візуальний аналіз даних – Visual Mining

1. Виконання візуального аналізу даних.
2. Характеристики засобів візуалізації даних.
3. Методи візуалізації.
4. Методи геометричних перетворень.
5. Відображення іконок.
6. Методи, орієнтовані на пікселі.
7. Ієрархічні образи

Література: основна [1; 2; 4]; додаткова [7; 11; 15].

Тема 3. Аналіз текстової інформації – Text Mining

1. Задача аналізу текстів (етапи аналізу текстів, попередня обробка тексту, задачі Text Mining).

2. Добування ключових понять із тексту (загальний опис процесу добування понять із тексту, стадія локального аналізу, стадія інтеграції й висновку понять).

3. Класифікація текстових документів (опис задачі класифікації текстів, методи класифікації текстових документів).

4. Методи кластеризації текстових документів (подання текстових документів, ієрархічні методи кластеризації текстів, бінарні методи кластеризації текстів).

5. Задача анотування текстів (виконання анотування текстів, методи добування фрагментів для анотації).

6. Засоби аналізу текстової інформації (засоб Oracle – Oracle Text, засоб від IBM – Intelligent Miner for Text, засоб SAS Institute – Text Miner, засоб Мега-комп'ютер Інтелідженс – Text Analyst).

Література: основна [1; 2; 4]; додаткова [6; 7; 12].

Розділ 4. Добування знань з Web – Web Mining

1. Web Mining (проблеми аналізу інформації з Web, етапи Web Mining, Web Mining й інші інтернет-технології, категорії Web Mining).

2. Методи добування Web-контенту (добування Web-контенту в процесі інформаційного пошуку, добування Web-контенту для формування баз даних).

3. Добування Web-структур (подання Web-структур, оцінка важливості Web-структур, пошук Web-документів з урахуванням гіперпосилань, кластеризація Web-структур).

4. Дослідження використання Web-ресурсів (дослідницька інформація, етап препроцесінгу, етап добування шаблонів, етап аналізу шаблонів й їхнє застосування).

Література: основна [1; 2; 3; 5]; додаткова [7; 8; 11; 12].

Тема 5. Засоби аналізу процесів – Process Mining

1. Автоматизація виконання бізнесів-процесів (бізнеси-процеси, формалізація бізнесів-процесів, Workflow-системи, сервісно-орієнтована архітектура, проектування бізнесів-процесів).

2. Аналіз процесів (технологія Process Mining, аналіз протоколів, стандарт запису протоколів MXML, задачі Process Mining, проблеми аналізу протоколів).

3. Методи Process Mining (перші імовірнісні методи Process Mining, метод побудови диз'юнктивної Workflow-схеми, α -алгоритм, методи на основі генетичних алгоритмів).

4. Бібліотека алгоритмів Process Mining— Pro (архітектура Pro, Pro Import Framework).

Література: основна [1; 3; 5]; додаткова [11; 15].

Тема 6. Пошук асоціативних правил – Rules Mining

1. Постановка задачі.

2. Подання результатів (правила класифікації, дерева класифікації, математичні функції).

3. Методи побудови правил класифікації (алгоритм побудови правил, метод Naive Bayes).

4. Методи побудови дерев класифікації (методика "розділяй і пануй", алгоритм покриття).

5. Методи побудови математичних функцій (загальний вид, лінійні методи, метод найменших квадратів, нелінійні методи, Support Vector Machines (SVM), регуляризаційні мережі (Regularization Networks), дискретизації й рідкі сітки).

6. Постановка задачі пошуку асоціативних правил (формальна постановка задачі, секвенціальний аналіз, різновиди задачі пошуку асоціативних правил).

7. Подання результатів.

8. Алгоритми (алгоритм Apriori, різновиди алгоритму Apriori).

Література: основна [1; 2; 3; 4]; додаткова [10; 13; 14; 15].

Тема 7. Розподілений аналіз даних

1. Системи мобільних агентів (основні поняття, стандарти багато-агентних систем, системи мобільних агентів, система мобільних агентів JADE).

2. Використання мобільних агентів для аналізу даних (проблеми розподіленого аналізу даних, агенти-аналітики, варіанти аналізу розподілених даних).

3. Система аналізу розподілених даних (загальний підхід до реалізації системи, агент для збору інформації про базу даних, агент для збору статистичної інформації, агент для вирішення однієї задачі інтелектуального аналізу даних, агент для вирішення інтегрованої задачі інтелектуального аналізу даних).

Література: основна [1, 3, 4, 5]; додаткова [9, 11, 14].

5. Плани лабораторних занять

Лабораторне заняття – форма навчального заняття, спрямована на формування вмінь та навичок роботи з пакетами прикладних програм з прогнозування характеристик соціально-економічних явищ шляхом індивідуального виконання студентом відповідно до сформульованих завдань. Проведення лабораторного заняття ґрунтується на попередньо підготовленому методичному матеріалі — наборі завдань різної складності для розв'язання їх студентами на занятті.

На кожному лабораторному занятті (табл. 4), до виконання лабораторної роботи, студент має відповісти на контрольні питання, які відбивають його готовність до виконання лабораторної роботи, зокрема оволодіння необхідними теоретичними знаннями та усвідомлення мети роботи. По закінченні виконання лабораторної викладач оцінює ступінь оволодіння відповідними навичками та ступінь досягнення мети даної роботи.

Для здачі лабораторної роботи студенту необхідно оформити індивідуальний звіт, у якому повинні бути: постановка задачі, роздруковані основні результати моделі, аналіз розрахунків і чіткі висновки та економічна інтерпретація результатів. Підсумкові оцінки за виконання кожної лабораторної роботи вносяться у відповідний журнал. Отримані студентом оцінки за окремі лабораторні роботи враховуються при виставленні підсумкової оцінки з даної навчальної дисципліни.

Таблиця 4

Перелік тем лабораторних робіт

Назва теми	Тема лабораторних робіт	Кількість годин	Література
1	2	3	4
Тема 1. Інтелектуальний аналіз даних – Data Mining	Інтелектуальний аналіз даних за допомогою аналітичної платформи Deductor	3	Основна [1; 2]; додаткова [12]
Тема 2. Візуальний аналіз даних - Visual Mining	Засоби візуального аналізу даних в пакеті Statistica	2	Основна [1; 2; 4]; додаткова [10; 11]
Тема 3. Аналіз текстової інформації - Text Mining	Вивчення пакету інтелектуального аналізатора текстової інформації – Text Analyst	2	Основна [1; 2]; додаткова [12; 14]

1	2	3	4
Розділ 4. Добування знань з Web - Web Mining	Создание нейросети в пакете NeuroPro	2	Основна [1; 2; 4]; додаткова [13; 15]
Тема 5. Засоби аналізу процесів - Process Mining	Аналіз бізнесів-процесів за допомогою бібліотеки алгоритмів Process Mining – Pro (архітектура Pro, Pro Import Framework)	2	Основна [1; 2; 4]; додаткова [15; 16; 17]
Тема 6. Пошук асоціативних правил - Rules Mining	Применение деревьев классификации для сегментации данных в пакете See 5.	2	Основна [1; 2; 3; 4; 5]; додаткова [12; 13]
Тема 7. Розподілений аналіз даних	Розподілений аналіз даних за допомогою пакету Agent Sheet	2	Основна [1; 4]; додаткова [12; 14]
Разом годин за темами		15	

6. Самостійна робота студентів

Необхідним елементом успішного засвоєння матеріалу навчальної дисципліни є самостійна робота студентів з вітчизняною та закордонною спеціальною літературою, спеціальними засобами імітаційного моделювання, періодичними виданнями тощо.

Основні види самостійної роботи, які запропоновані студентам:

1. Вивчення лекційного матеріалу.
2. Опрацювання та вивчення рекомендованої літератури.
3. Вивчення основних термінів та понять за темами дисципліни.
4. Підготовка до лабораторних занять, колоквиумів, тестування.
5. Контрольна перевірка кожним слухачем особистих знань за запитаннями для самоконтролю.

Перелік питань для самостійного опрацювання надано в табл. 5.

Таблиця 5

Перелік питань для самостійного опрацювання

Назва теми	Питання для самостійного опрацювання	Рекомендована література
1	2	3
Тема 1. Інтелектуальний аналіз даних – Data Mining	1. Основні цілі й завдання ІАД. 2. Засоби й методи ІАД. 3. Призначення, основні етапи розвитку й види систем ІАД	Основна [1; 2]; додаткова [10; 16]

1	2	3
Тема 2. Візуальний аналіз даних - Visual Mining	1.Способи візуалізації інформації в системах ІАД. 2. Системи для візуалізації багатомірних даних	Основна [1; 2]; додаткова [12; 13; 14]
Тема 3. Аналіз текстової інформації - Text Mining	1.Обробка текстів на природній мові. 2.Природні та штучні мови	Основна [1; 2]; додаткова [12; 15; 16]
Розділ 4. Добування знань з Web - Web Mining	1.Організація інформації в Internet. 2.Моделі знань у мережах	Основна [1; 2; 3]; додаткова [11; 12]
Тема 5. Засоби аналізу процесів - Process Mining	1. Моделювання бізнес-процесів. 2. Мережі Петрі як засіб моделювання	Основна [1; 3]; додаткова [6; 7; 10; 16]
Тема 6. Пошук асоціативних правил - Rules Mining	1.Асоціативні правила як модель знань. 2.Використання асоціативних правил у експертних системах	Основна [1; 2; 3; 4]; додаткова [7; 8; 9; 11; 12]
Тема 7. Розподілений аналіз даних	1.Інтелектуальні агенти в економіці. 2.Моделювання поведінки інтелектуальних агентів. 3.Багато-агентні системи	Основна [1; 4]; додаткова [8; 9; 11; 14]

Для оцінки ступеня опрацювання матеріалу, який запропоновано для самостійної роботи застосовується колоквіум – усна бесіда з викладачем за питаннями для самостійної роботи. Колоквіум проводиться один раз у семестр і включає матеріал першого змістовного модуля.

7. Контрольні запитання для самодіагностики

1. Для якої шкали застосовують тільки такі операції як: дорівнює, не дорівнює, більше, менше?
2. На які дві групи підрозділяються методи Data Mining за принципом роботи з вихідними навчальними даними?
3. Яка зі стадій може вважатися додатковою або невід'ємною частиною однієї з основних стадій Data mining?
4. Чи правильне таке формулювання: "Асоціація є часткою випадкової послідовності з часовим лагом, що дорівнює нулю"?
5. Що перебуває в основі так званої інформаційної піраміди?
6. На які категорії підрозділяються задачі Data Mining, залежно від моделей, що використовуються?

7. Для вирішення яких задач застосовуються ієрархічні алгоритми?
8. У чому подібність задач класифікації й прогнозування?
9. Який з напрямів включає виявлення закономірностей у діях користувача web-вузла або їхньої групи?
10. Які фактори впливають на результат класифікації в наївно-байєсовським підході?
11. Які переваги характерні для використання байєсовських мереж?
12. Які задачі вирішуються за допомогою методу опорних векторів?
13. Що являє собою група синапсів нейрону?
14. Визначте головну функцію штучного нейрона.
15. У чому полягає навчання мереж, що самоорганізуються?
16. Визначте додаток до традиційних методів візуалізації?
17. Назвіть характеристики однієї з основних тенденцій в області візуалізації.
18. Що означає інтегрованість сховища даних?
19. Назвіть основні концепції сховища даних.
20. Якщо набір даних упорядкований та у ньому присутня сезонна або циклічна компонента, то яку мінімальну кількість даних необхідно мати для можливості аналізу?
21. Що повинна робити якісна програма очищення даних?
22. Які характеристики повинна мати якісна програма очищення даних?
23. На якому етапі перетинається робота фахівця предметної області й фахівця з видобутку даних?
24. Назвіть характеристики, властиві SAS Enterprise Miner.
25. Рішення яких задач передбачають алгоритми аналізу даних в PolyAnalyst?
26. Опишіть можливості пакета Deductor з заповнення пропусків.
27. Які алгоритми реалізовані в пакеті Deductor?
28. Чи існує необхідність тимчасового або постійного копіювання даних для аналізу в системі KXEN?
29. Який компонент KXEN дозволяє виявити природні групи (кластери) у наборі даних?
30. Назвіть слабкі сторони використання готового програмного забезпечення Data Mining.

8. Індивідуальне навчально-дослідне завдання

Індивідуальне навчально-дослідне завдання (далі – ІНДЗ) передбачає: систематизацію, закріплення, розширення теоретичних і практичних знань з дисципліни та їх застосування при дослідженні динаміки розвитку соціально-економічних систем різного рівня ієрархії; розвиток навичок самостійної роботи й оволодіння методикою дослідження та експерименту, пов'язаних з темою ІНДЗ.

Індивідуальне навчально-дослідне завдання виконується самостійно при консультуванні викладачем протягом вивчення дисципліни у відповідності до графіка навчального процесу. Студент має надати ІНДЗ для перевірки наприкінці семестру, але не пізніше терміну проведення підсумкового модульного контролю. Оцінка за виконання ІНДЗ враховується при виставленні загальної оцінки з дисципліни.

Тематика ІНДЗ має носити проблемний характер. Студент має право самостійно обрати тему та зміст роботи з обов'язковим її узгодженням з викладачем. У протилежному випадку тема має бути запропонована викладачем (варіанти тем наведено нижче).

Індивідуальне завдання складається з: титульної сторінки, змісту, вступу, основної частини, завершення, списку використаної літератури, додатків (за необхідністю).

Вступ має розкривати актуальність обраної студентом теми, її проблематику, мету написання роботи.

Основна частина роботи (може включати декілька підрозділів) має включати характеристику сучасного стану проблеми прогнозування, огляд різних підходів щодо вирішення проблеми.

Завершення має включати обґрунтовані висновки студента щодо досягнення мети роботи, можливо рекомендації щодо застосування методів прогнозування.

Обсяг ІНДЗ повинен становити у друкованому варіанті 15 – 20 сторінок. Орієнтована кількість сторінок у розділах: вступ – 1 с.; основна частина – 12 – 18 с.; завершення – 1 – 2 с.

Список використаної літератури необхідно скласти у певному порядку: спочатку наводяться законодавчі та нормативні акти, статистичні довідники, потім загальна та спеціальна література за алфавітом.

8.1. Тематика ІНДЗ

1. Застосування систем ІАД для інформаційно-аналітичного забезпечення прийняття рішень (на основі обраного прикладу).
2. Основні форми подання й способи пояснення (інтерпретації) результатів ІАД.
3. Основні етапи аналізу й інтерпретації даних.
4. Стратегії уведення, подання й організації вихідних даних в інформаційних системах з ІАД.
5. Вимірювальні шкали, подання змінних, уведення й редагування даних.
6. Трансформація даних і файлів у методах ІАД (на обраному методі, прикладі рішення завдання).
7. Основні способи завантаження даних у системи ІАД з різних джерел, у тому числі на основі ODBC.
8. Об'єднання й агрегування (файлів) даних у інформаційно-аналітичних системах із застосуванням засобів ІАД.
9. Базові розвідницькі методи аналізу даних.
10. Табличні способи подання результатів ІАД.
11. Застосування методів і засобів картографії для аналізу даних й інтерпретації його результатів.
12. Побудова й інтерпретації карт візуалізації залежностей (на обраному прикладі, у тому числі - аналізу категоріальних даних).
13. Побудова, застосування та інтерпретації карт категоріальних даних.
14. Підготовка звітів та експорт результатів аналізу із систем ІАД.
15. Класичні методи багатомірного статистичного аналізу.
16. Роль і місце методів класичного статистичного аналізу для розв'язання основних завдань ІАД.
17. Застосування методів класифікації й кластеризації для сегментації даних.
18. Використання дерев рішень у завданнях класифікації (на обраному прикладі).
19. Компоненти й моделі тимчасових рядів (тренда, сезонності, циклу, випадкового компонента).
20. Введення в нейронні мережі: основні принципи організації, побудови й застосування.

21. Рішення завдань класифікації нейронними мережами (на обраному прикладі).
22. Застосування нейронних мереж до завдань регресійного аналізу (на обраному прикладі).
23. Застосування нейронних мереж до завдань прогнозування тимчасових рядів (на обраному прикладі).
24. Деревя рішень, принципи їхньої побудови, практичного застосування й інтерпретації результатів (характеристика дерев).
25. Аналіз за допомогою дерева рішень: звичайні завдання аналізу за допомогою дерева рішень, області додатка аналізу за допомогою дерева рішень.
26. Застосування дерева рішень для класифікації (на обраному прикладі).
27. Застосування дерева рішень для прогнозування (на обраному прикладі).
28. Порівняльний аналіз засобів (систем) аналізу й прогнозування на основі дерев рішень (на прикладі не менше трьох систем).
29. Автоматизація виконання повторюваних завдань у системах ІАД.
30. Об'єднання й агрегування (файлів) даних в інформаційно-аналітичних системах із застосуванням засобів ІАД.
31. Створення звітів й експорт результатів із систем ІАД у різні додатки (на прикладі обраної системи).
32. Вплив розмірності й обсягів вихідних даних на вибір і застосування засобів ІАД.
33. Методи аналізу різнотипних експериментальних даних.
34. Застосування генетичних алгоритмів для аналізу даних (на прикладі пакета SNN).
35. Організація й реалізація оптимального вибору структури й параметрів моделей в ІАД (на обраному прикладі).

9. Індивідуально-консультативна робота

Індивідуально-консультативна робота здійснюється за графіком індивідуально-консультативної роботи у формі індивідуальних занять, консультацій, перевірки виконання індивідуальних завдань, перевірки та захисту завдань, що винесені на поточний контроль тощо.

Формами організації індивідуально-консультативної роботи є:

- а) за засвоєнням теоретичного матеріалу: консультації індивідуальні (запитання – відповідь), консультації групові (розгляд типових прикладів – ситуацій);

б) за засвоєнням практичного матеріалу: консультації індивідуальні і групові;

в) для комплексної оцінки засвоєння програмного матеріалу: індивідуальне здавання виконаних робіт.

10. Методики активізації процесу навчання

При викладанні навчальної дисципліни "Інтелектуальний аналіз даних" для активізації навчально-пізнавальної діяльності студентів передбачено застосування таких навчальних технологій, як: проблемні лекції, колоквиум, презентації, самооцінка, банки візуального супроводження (табл. 6).

Таблиця 6

Використання навчальних технологій для активізації процесу навчання

Методики активізації процесу навчання	Практичне застосування навчальних технологій
1	2
Проблемні лекції направлено на розвиток логічного мислення студентів, коло питань теми обмежується двома-трьома ключовими моментами, використовуються досвід закордонних навчальних закладів з роздачею студентам під час лекцій друкованого матеріалу та виділенням головних висновків з питань, що розглядаються. При читанні лекцій студентам даються питання для самостійного розмірковування, на які лектор відповідає сам, не чекаючи на відповіді студентів	Проблемна лекція з питання "Парадигма інтелектуального аналізу даних – Data Mining" (тема 1)
	Проблемна лекція з питання "Візуальний аналіз та обробка графічних даних – Visual Mining" (тема 2)
	Проблемна лекція з питання "Структура систем аналізу текстової інформації – Text Mining" (тема 3)
	Проблемна лекція з питання "Технології добування знань з Web - Web Mining"(тема 4)
	Проблемна лекція з питання "Моделювання бізнес-процесів та їх аналіз – Process Mining"(тема 5)
	Проблемна лекція з питання "Правила як модель знань та їх використання в експертних системах" (тема 6)
	Проблемна лекція з питання "Від інтелектуального агента до базатоагентної системи" (тема 7)
Презентації – виступи перед аудиторією, що використовуються для представлення певних досягнень.	Демонстрація та обговорення результатів виконання лабораторних робіт та теоретичної підготовки до їх виконання
	Демонстрація та обговорення результатів самостійної підготовки з окремих питань
	Презентація індивідуального навчально-дослідного завдання

1	2
Самооцінка сприяє підвищенню зацікавленості студентів у виступах колег, виробленню об'єктивності в оцінці їх діяльності, підвищує відповідальність студентів за підготовку до виступу, дозволяє провести аналіз помилок у структурі або змісті виступу, виявити приховані конфлікти у групі	Виступи студентів з доповідями (презентаціями) з питань самостійної роботи, виконання ІНДЗ або підготовки до лабораторних робіт оцінюються іншими студентами групи, за критеріями: володіння матеріалом (60 %), вміння пояснити слухачам (30 %), емоційність (10 %). Для проведення самооцінки використовується спеціальний бланк з зазначенням прізвища студента, критеріїв оцінки та вагових коефіцієнтів, який побудовано таким чином, щоб забезпечити анонімність (аби уникнути конфліктів у групі) та об'єктивність оцінок. Викладач також виставляє оцінку за наведеними критеріями. Загальна оцінка визначається як середня.
Банки візуального супроводження за допомогою наочності сприяють активізації творчого сприйняття змісту дисципліни	<p>Опорні конспекти, які тезово відображають ключові моменти заняття, основні формули, визначення та схеми (лекції з тем 1 – 7)</p> <p>Презентація методу побудови дерев класифікації (лекція за темою 4)</p> <p>Комплект форм для застосування самооцінки</p> <p>Методичні вказівки до виконання лабораторних робіт, тексти лекцій</p> <p>Система інтерактивного тестування для поточної самоперевірки знань CLARION</p>

11. Система поточного та підсумкового контролю знань студентів

Система оцінювання знань, вмінь та навичок студентів враховує види занять, які згідно з програмою навчальної дисципліни "Інтелектуальний аналіз даних" передбачають лекційні та лабораторні заняття, а також самостійну роботу і виконання індивідуальних завдань. Контрольні заходи включають поточний і підсумковий контроль.

Перевірка та оцінювання знань студентів проводиться в наступних формах:

1. Оцінювання знань студента під час лабораторних занять.
2. Виконання індивідуального навчально-дослідного завдання.
3. Проведення проміжного тестування.
4. Проведення поточного модульного контролю.

Оцінювання знань **студента під час лабораторних занять** має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Об'єктами поточного контролю є:

а) систематичність, активність та результативність роботи протягом семестру над вивченням програмного матеріалу дисципліни; виконання лабораторних робіт; відвідування занять;

б) виконання завдань для самостійного опрацювання;

в) рівень виконання модульних завдань.

Оцінювання виконання проводиться за 12-бальною шкалою за такими критеріями:

а) ступінь засвоєння та розуміння фактичного матеріалу навчальної дисципліни;

б) ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;

в) вміння поєднувати теорію з практикою при розв'язанні задач моделювання реальних економічних систем, при виконанні завдань, винесених для самостійного опрацювання, та завдань, винесених на розгляд аудиторії;

г) логіка, структура, стиль викладу матеріалу в письмових роботах і під час виступів в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки.

Оцінка "відмінно" ставиться за умови відповідності виконання студентом завдань та його усної відповіді (захисту роботи) всім чотирьом зазначеним критеріям. Відсутність тієї або іншої складової знижує оцінку на відповідну кількість балів.

При оцінюванні виконання лабораторних робіт увага також приділяється якості, самостійності та своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу).

Індивідуальне навчально-дослідне завдання оцінюється за критеріями: самостійності виконання; повноти й глибини розглядання досліджуваної економічної системи; логічності та рівня деталізації побудованої моделі; наявності ілюстрацій (таблиці, рисунки, схеми тощо); наявності конкретних пропозицій; якості оформлення; своєчасності подання; вміння обґрунтовувати якість побудованої моделі, захищати наведені висновки.

Проміжний тестовий контроль проводиться по закінченні вивчення кожної теми дисципліни. Під час проведення поточного тестування визначається рівень знань студентів з теоретичних питань навчальної дисципліни.

Тестові завдання охоплюють теоретичний матеріал теми, який вивчаються в межах навчальної дисципліни "Інтелектуальний аналіз даних" та згруповані за двома модулями, кожен з яких складається з тестових завдань різного рівня складності.

Тестові завдання розрізняються за принципом побудови відповіді.

Тестові завдання закритого типу:

1. *Альтернативні* тестові завдання передбачають наявність двох варіантів відповіді типу "так" – "ні", "правильно" – "неправильно", їх використовують для перевірки правильності вибору або прийняття рішення в згорнутій формі. Для правильної відповіді на них слід констатувати лише хибність чи істинність наведеного твердження. У випадку, якщо твердження, що підлягає оцінці на істинність, містить хоча б один виняток, його вважають хибним. Типовим початком питання є "Чи правильно, що...".

2. Тестові завдання з *множинними відповідями* "правильно" – "неправильно" передбачає наведення кількох варіантів відповідей (до 5), з яких правильною вважається лише одна. За ними тестується глибина знань, розуміння різних аспектів явищ, процесів тощо.

3. Тестові завдання з *множинними відповідями* "правильно" передбачає наведення кількох варіантів відповідей (до 5), з яких правильними може виявитися не одна. За ними тестується глибина знань, розуміння різних аспектів явищ, властивостей, процесів тощо. При формулюванні питання зазначають: "Оберіть усі можливі ...", "Оберіть усі ..., що відповідають умовам ...".

4. Тестові завдання *на поєднання відповідних частин* відносяться до найскладніших. Їх сутність полягає у формуванні двох (чи більше) колонок інформації, одна з яких містить певні терміни, елементи дисципліни, а друга – відповідні їм визначення, властивості і т. ін, але в іншому порядку. Студентам пропонується поставити визначення (властивості, графіки і т. ін.) у тому ж порядку, що й терміни (елементи). В якості відповіді наводиться послідовність позначок відповідних визначень.

5. Тестові завдання *на відтворення правильної послідовності*, які потребують переструктурування даних або елементів будь-якої комбінації, використовуються в разі тестування вмінь та знань правильної послідовності дій, алгоритмів, технологічних прийомів тощо при застосуванні певних методів (методик).

6. Тестові завдання *відкритого типу* передбачають вільні відповіді тестованих, вони є завданнями без запропонованих варіантів відповідей і

використовуються для виявлення знань термінів, визначень, понять і т. ін. Їх сутність полягає у наведенні у запитанні твердження, деякі слова якого замінені пропусками. У процесі відповіді замість пропуску студенту потрібно вписати відповідний професійний термін, що перетворить дане твердження на істинне.

Для поточного тестування з окремих тем використовуються альтернативні тестові запитання, які сприймаються студентами на слух. Для відповіді використовуються спеціальні бланки відповідей.

Тестове завдання містить від 15 до 30 запитань щодо перевірки знань основних категорій навчальної дисципліни залежно від теми.

Оцінка рівня відповідей студентів на тестові завдання розраховується за формулою:

$$\text{оцінка} = \text{округл} \left[\frac{\text{кількість вірних відповідей} * 12}{\text{кількість запитань}} \right], \quad (1)$$

де округл[] – функція округлення за загальними правилами.

Поточно-модульний контроль здійснюється та оцінюється за двома складовими: практичний (лабораторний) модульний контроль і лекційний модульний контроль. Оцінка за практичну складову модульного контролю виставляється за результатами оцінювання знань та вмінь студента під час лабораторних занять, виконання індивідуального навчально-дослідного завдання та проміжного тестового контролю згідно з графіком навчального процесу.

Лекційний модульний контроль здійснюється в письмовій формі за відповідними тестами, зміст яких містить питання всіх тем модуля.

Таким чином, після вивчення всіх тем студенти денної форми навчання виконують завдання поточного модульного контролю.

Зразок завдання модульного контролю

1. Закономірності, знайдені в процесі використання технології Data повинні мати такі властивості:

- а) бути очевидними;
- б) бути неочевидними;
- в) бути практично корисними;
- г) бути об'єктивними;
- д) чим більше знайдено закономірностей, тим краще.

2. Оберіть характеристику, найбільш підходящу для Data Mining
- а) підходить для розуміння ретроспективних даних;
 - б) опирається на ретроспективні дані для одержання відповідей на питання про майбутнє;
 - в) підходить для узагальнення ретроспективних даних.
3. Дані являють собою:
- а) факти й графіки;
 - б) текст;
 - в) картинки, звуки, аналогові або цифрові відео-сегменти;
 - г) всі разом.
4. Для якої шкали застосовні тільки такі операції як "дорівнює", "не дорівнює", "більше", "менше":
- а) номінальна шкала;
 - б) порядкова шкала;
 - в) інтервальна шкала?
5. На які дві групи підрозділяються методи Data Mining за принципом роботи з вихідними навчальними даними:
- а) безпосереднє використання даних або збереження даних;
 - б) виявлення й використання формалізованих закономірностей;
 - в) статистичні методи;
 - г) кібернетичні методи?
6. Яка з перерахованих нижче стадій може вважатися додатковою або частиною однієї з основних стадій Data mining:
- а) виявлення закономірностей (вільний пошук);
 - б) використання виявлених закономірностей для прогнозування невідомих значень (прогнозтичне моделювання);
 - в) валідація.
7. Нечітка логіка й дерева рішень:
- а) відносяться до статистичних методів Data mining;
 - б) відносяться до кібернетичних методів Data mining;
 - в) не є методами Data mining.
8. Чи правильне таке формулювання: "Асоціація є часткою випадкової послідовності з часовим лагом, що дорівнює нулю":
- а) формулювання правильне;
 - б) послідовність є часткою випадкової асоціації;
 - в) ні послідовність, ні асоціація не є окремими випадками один одного?

9. В основі так званої інформаційної піраміди перебуває категорія:

- а) дані;
- б) знання;
- в) інформація.

10. Задачі Data Mining, залежно від використовуваних моделей під-розділяються на:

- а) навчання з учителем;
- б) навчання без учителя;
- в) дескриптивні;
- г) прогнозуючі.

11. Виділіть два основних напрямки Web Mining:

- а) Web Content Mining;
- б) Web Usage Mining;
- в) Web Text Mining.

12. Яке з перерахованих нижче напрямків має на меті виявлення закономірностей у діях користувача web-вузла або їхньої групи:

- а) Web Content Mining;
- б) Web Usage Mining;
- в) Web Text Mining?

13. Кінцевий вузол дерева рішень називають також:

- а) вузлом перевірки;
- б) вузлом рішення;
- в) листом.

14. Використання байесовських мереж має наступні переваги:

- а) дозволяє уникнути проблеми переучування;
- б) визначає залежності між всіма змінними;
- в) на результат класифікації впливають тільки індивідуальні значення вхідних змінних.

15. За допомогою методу опорних векторів вирішуються задачі:

- а) бінарної класифікації;
- б) не тільки бінарної класифікації;
- в) прогнозування.

16. Традиційні методи візуалізації можуть знаходити наступне застосування:

- а) представляти користувачеві інформацію в наочному вигляді;
- б) компактно описувати закономірності, що властиві вихідному набору даних;

- в) знижувати розмірність або стискати інформацію;
- г) спрощувати розрахунки в моделі;
- д) відновлення пробілів у наборі даних.

17. Знаходження шумів і викидів у даних:

- а) можливо за допомогою засобів візуалізації;
- б) неможливо за допомогою засобів візуалізації;
- в) не є функцією візуалізації.

18. Назвіть характеристики однієї з основних тенденцій в області візуалізації:

- а) збільшення розмірів структур даних, що представляють за допомогою візуалізації;
- б) ускладнення структур даних, що представляють за допомогою візуалізації;
- в) зменшення розмірів структур даних, що представляють за допомогою візуалізації.

19. Якісна програма очищення даних повинна:

- а) виправляти неправильні дані;
- б) створювати невеликий за обсягом звіт про підозрілі записи;
- в) вимагати мінімальних витрат на встановлення, обслуговування й ручні перевірки;
- г) виправляти абсолютно всі підозрілі дані.

20. Якісна програма очищення даних повинна мати такі характеристики:

- а) виправляти неправильні дані;
- б) створювати невеликий за обсягом звіт про підозрілі записи;
- в) вимагати мінімальних витрат на встановлення, обслуговування й ручні перевірки;
- г) може частково зачіпати правильні дані.

21. Така функція інструментів очищення даних як поліпшення означає:

- а) додавання до даних додаткових фактів про записи, які в них не надавалися;
- б) видалення дублікатів даних;
- в) видалення шумів і викидів у системі даних.

22. На якому етапі перетинається робота фахівця предметної області й фахівця з видобутку даних:

- а) аналіз бізнес-процесів;
- б) аналіз даних;
- в) підготовка даних;
- г) всі відповіді неправильні?

23. Аналіз предметної області й інтерпретація результатів, отриманих у результаті Data Mining – це точки дотику таких фахівців як:

- а) фахівця предметної області;
- б) фахівця з видобутку даних;
- в) адміністратора баз даних;
- г) всіх разом.

24. Аналіз вимог до даних і збір даних – це точки дотику таких фахівців як:

- а) фахівця предметної області;
- б) фахівця з видобутку даних;
- в) адміністратора баз даних;
- г) всіх разом.

25. Існують наступні варіанти рішень з впровадження інструментів Data Mining:

- а) розробка Data Mining-продукту на замовлення сторонньою компанією;
- б) розробка Data Mining-продукту самотужки;
- в) комбінація цих варіантів, у т.ч. використання різних бібліотек, компонентів та інструментальних наборів для розроблювачів створення убудованих додатків Data Mining.

26. Інструменти Data Mining можуть вирішувати:

- а) тільки одну задачу Data Mining;
- б) кілька задач Data Mining;
- в) всі задачі Data Mining;
- г) це залежить від конкретного інструмента.

27. Рішення яких задач передбачає алгоритми аналізу даних в PolyAnalyst:

- а) моделювання;
- б) прогнозування;
- в) кластеризація;
- г) класифікація;
- д) текстовий аналіз;
- е) всі відповіді правильні.

28. Deductor Studio:

- а) може функціонувати без сховища даних;
- б) може одержувати інформацію з будь-яких інших джерел;
- в) не може функціонувати без сховища даних.

29. Опишіть можливості пакета Deductor по заповненню пропусків:

- а) немає можливості заповнення пропусків;

б) є можливість заповнення пропусків методом апроксимації;
 в) є можливість заповнення пропусків за допомогою алгоритму, що підставляє найбільш імовірне значення замість пропущених даних.

30. Які з представлених алгоритмів реалізовані в пакеті Deductor:

- а) нейронні мережі;
- б) автокореляція;
- в) дерева рішень;
- г) карти, що самоорганізуються;
- д) асоціативні правила;
- е) всі відповіді правильні?

Оцінка з модульного контролю розраховується за формулою (1).

Підсумкова оцінка з дисципліни складається як середня з двох складових, що враховує оцінки кожного виду контролю (дві оцінки за результатами поточного модульного контролю за роботу протягом семестру з коефіцієнтом 0,65 та оцінка за виконання та захист індивідуального науково-дослідного завдання з коефіцієнтом 0,35).

Підсумкова оцінка з дисципліни згідно з Методикою переведення показників успішності знань студентів Університету в систему оцінювання за шкалою ECTS конвертується в підсумкову оцінку за шкалою ECTS (табл. 7).

Таблиця 7

**Переведення показників успішності знань студентів
у систему оцінювання за шкалою ECTS**

Відсоток студентів, які зазвичай успішно досягають відповідної оцінки	Оцінка за шкалою ECTS		Оцінка за бальною шкалою, що використовується в ХНЕУ	Оцінка за національною шкалою
10	відмінне виконання	A	12 – 11	відмінно
25	вище середнього рівня	B	10	
30	взагалі робота правильна, але з певною кількістю помилок	C	9 – 7	добре
25	непогано, але зі значною кількістю недоліків	D	6	задовільно
10	виконання задовольняє мінімальні критерії	E	5 – 4	
–	потрібне повторне перескладання	PX	3	незадовільно
–	повторне вивчення дисципліни	P	2 – 1	

12. Рекомендована література

12.1. Основна

1. Барсегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко та ін. – 2-е изд., перераб. и доп. – СПб. : БХВ-Петербург, 2004. – 336 с.
2. Дюк В. "Data Mining" : учебный курс / В. Дюк, А. Самойленко. – СПб. : Питер, 2001. – 368 с.
3. Задания и методические рекомендации к выполнению лабораторных работ по теме "Нейронные сети" курса "Прогнозирование социально-экономических процессов" для студентов специальности 8.050102 "Экономическая кибернетика" дневной форм обучения / сост. Н. А. Дубровина, А. В. Милов, С. В. Прокопович. – Х. : Изд. ХНЭУ, 2005. – 60 с.
4. Корнеев В. В. Базы данных. Интеллектуальная обработка информации / В. В. Корнеев, А. Ф. Гареев, С. В. Васютин и др. – М. : Издатель Могачева С. В. ; Издательство Нолидж, 2001. – 496 с.
5. Рассел С. Искусственный интеллект: современный подход / Рассел С., Норвинг П. – М. : Издательский дом "Вильямс", 2006. – 1408 с.

12.2. Додаткова

6. Вагин В. Н. Дедукция и обобщение в системах принятия решений / Вагин, В. Н. – М. : Наука, 1988. – 383 с.
7. Дюк В. "Обработка данных на ПК в примерах" / Дюк В. – СПб. : Питер, 1997. – 240 с.
8. Загоруйко Н. Г. Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. – Новосибирск : Изд. Ин-та математики СО РАН, 1999. – 270 с.
9. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. Пер. с англ. – М. : Издательский дом "Вильямс", 2001. – 292 с.
10. Круглов В. В. Искусственные нейронные сети: Теория и практика / В. В. Круглов, В. В. Борисов. – М. : Горячая линия – Телеком, 2001. – 383 с.
11. Лбов Г. С. Анализ данных и знаний : учебн. пособ. / Г. С. Лбов. – Новосибирск : Изд. НГТУ, 2001. – 90 с.
12. Романов А. Н. Советующие информационные системы в экономике : учебн. пособ. для вузов / А. Н. Романов, Б. Е. Одинцов. – М. : ЮНИТИ-ДАНА, 2000. – 488 с.
13. Тюрин Ю. Н. Статистический анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – М. : ИНФРА-М, 1998. – 528 с.

14. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – 2-е изд.; пер. с англ. – М. : Вильямс. 2006. – 1204 с.

15. Ханк Д. Э. Бизнес-прогнозирование / Д. Э. Ханк, Д. У. Уичерн, А. Дж. Райтс. – 7-е изд., пер. с англ. – М. : Вильямс, 2003. – 644 с.

12.3. Ресурси мережі Internet

16. Електронна бібліотека. – Режим доступу : <http://all-ebooks.com>.

17. Свободна енциклопедія [Електронний ресурс]. – Режим доступу : http://ru.wikipedia.org/wiki/Data_mining.

18. Дюк В. А. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях [Электронный ресурс] / В. А. Дюк, А. В. Флегонтов, И. К. Фомина. – Режим доступа : ftp://lib.herzen.spb.ru/text/dyuk_138_77_84.pdf.

19. Шапот М. Интеллектуальный анализ данных в системах поддержки принятия решений [Электронный ресурс] / М. Шапот. – Режим доступа : <http://www.osp.ru/os/1998/01/179360/>.

Зміст

Вступ	3
1. Кваліфікаційні вимоги до студентів	4
2. Тематичний план навчальної дисципліни	6
3. Зміст навчальної дисципліни за темами	7
4. Плани лекцій	10
5. Плани лабораторних занять	13
6. Самостійна робота студентів	14
7. Контрольні запитання для самодіагностики	15
8. Індивідуальне навчально-дослідне завдання	17
8.1. Тематика ІНДЗ	18
9. Індивідуально-консультативна робота	19
10. Методика активізації процесу навчання	20
11. Система поточного та підсумкового контролю знань студентів	21
12. Рекомендована література	30

НАВЧАЛЬНЕ ВИДАННЯ

**Робоча програма
навчальної дисципліни
"ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ"**

для студентів спеціальності
8.000011 "Прикладна економіка"
всіх форм навчання

Укладачі: **Мілов** Олександр Володимирович
Мілевський Станіслав Валерійович

Відповідальний за випуск **Клебанова Т. С.**

Редактор **Шаповал Г. В.**

Коректор **Бриль В. О.**

План 2012 р. Поз. № 317.

Підп. до друку. Формат 60 × 90 1/16. Папір MultiCopy. Друк Riso.

Ум.-друк. арк. 2,0. Обл.-вид. арк. 2,5. Тираж прим. Зам. №

Видавець і виготівник – видавництво ХНЕУ, 61166, м. Харків, пр. Леніна, 9а

*Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
Дк № 481 від 13.06.2001 р.*

**Робоча програма
навчальної дисципліни
"ІНТЕЛЕКТУАЛЬНИЙ
АНАЛІЗ ДАНИХ"
для студентів спеціальності
8.000011 "Прикладна економіка"
всіх форм навчання**