

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ**

СИСТЕМА УПРАВЛІННЯ ЯКІСТЮ

Система забезпечення якості освітньої діяльності та якості вищої освіти
сертифікована на відповідність ДСТУ ISO 9001:2015 / ISO 9001:2015

Кафедра цифрової економіки та системного аналізу

ЗАТВЕРДЖЕНО

вченою радою

(пост. п. __ від «__» _____ 2021 р.)

Ректор

_____ А. А. Мазаракі

**ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ /
DATA ANALYSIS TECHNOLOGIES**

**РОБОЧА ПРОГРАМА /
COURSE OUTLINE**

освітній ступінь	бакалавр / bachelor
галузь знань	12 Інформаційні технології / Information Technologies
спеціальність	124 Системний аналіз / System analysis
спеціалізація	Системний аналіз / System analysis Інформаційні технології та бізнес-аналітика (Data Science) / Information technologies and business intelligence (Data Science)

Київ 2021

**Розповсюдження і тиражування без офіційного дозволу КНТЕУ
заборонено**

Автор: А. А. Роскладка, доктор економічних наук, професор

Робочу програму розглянуто і затверджено на засіданні кафедри
цифрової економіки та системного аналізу 10.02.2021р., протокол № 7

Рецензенти: В. В. Кулаженко, кандидат економічних наук,
М. Г. Шарафутдінов, директор з розвитку компанії «Center
Research & Development», бізнес-аналітик

**ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ/
DATA ANALYSIS TECHNOLOGIES**

**РОБОЧА ПРОГРАМА /
COURSE OUTLINE**

освітній ступінь	бакалавр / bachelor
галузь знань	12 Інформаційні технології / Information Technologies
спеціальність	124 Системний аналіз / System analysis
спеціалізація	Системний аналіз / System analysis Інформаційні технології та бізнес-аналітика (Data Science) / Information technologies and business intelligence (Data Science)

1. СТРУКТУРА ДИСЦИПЛІНИ ТА РОЗПОДІЛ ГОДИН ЗА ТЕМАМИ (ТЕМАТИЧНИЙ ПЛАН)

Для вивчення дисципліни відводиться 360 год. / 12 кредити ЄКТС (лекцій – 58 год., лабораторних занять – 116 год., самостійної роботи – 186 год.), підсумковий контроль – курсова робота, екзамен

Назва теми	Кількість годин				Форма контролю
	Усього годин / кредитів	з них:			
		лекцій	лабораторні заняття / МК	самостійна робота аспірантів	
Тема 1. Передобробка даних	52	10	12	30	Т, ІДЗ, Пр, П
Тема 2. Асоціація даних	18	2	6	10	Т, ІДЗ, Пр, П
Тема 3. Кластеризація даних	34	2	16	16	Т, ІДЗ, Пр, П, МК
Тема 4. Класифікація та регресія даних	48	6	22	20	Т, ІДЗ, Пр, П, МК
Тема 5. Технології інтелектуальної обробки даних	18	8	0	10	Т
Тема 6. Інструментальні засоби аналізу даних	38	6	12	20	Т, ІДЗ, Пр, П
Тема 7. Створення моделі даних	76	10	26	40	Т, ІДЗ, Пр, П, МК
Тема 8. Побудова аналітичних звітів	76	14	22	40	Т, ІДЗ, Пр, П, МК
Підсумковий контроль – курсова робота, екзамен					
Разом	360/12	58	116	186	x

Примітка: Т – тестування; ІДЗ – виконання індивідуальних домашніх завдань; П – перевірка індивідуальних завдань; Пр. – презентація індивідуального завдання; МК – модульний контроль.

2. ТЕМАТИКА ТА ЗМІСТ ЛЕКЦІЙНИХ, ПРАКТИЧНИХ (СЕМІНАРСЬКИХ) ЗАНЯТЬ І САМОСТІЙНОЇ РОБОТИ ЗДОБУВАЧІВ

Результати навчання	Навчальна діяльність ¹	Робочий час здобувача, год
1	2	3
<p>Знати: основні поняття <i>Data Science, Artificial Intelligence, Business Intelligence</i>, етапи розв'язування задач аналізу даних, технології <i>Data Mining</i>.</p> <p>Вміти: оцінювати якість даних, здійснювати консолідацію, очищення та збагачення даних.</p>	<p style="text-align: center;">Тема 1. Передобробка даних.</p> <p style="text-align: center;">Лекція 1. Вступ у Data Science. Поняття про дані План лекції</p> <ol style="list-style-type: none"> 1. Поняття про дані. 2. Типи та види даних. Форми представлення даних. 3. Вимірювання і шкали в аналізі даних. Реляційні та багатовимірні дані. Метадані. 4. Формати зберігання даних. 5. Типи наборів даних. Транзакційні дані. <p style="text-align: center;">Лекція 2. Основні задачі аналізу даних. Knowledge Discovery in Databases і Data Mining. План лекції</p> <ol style="list-style-type: none"> 1. Характеристика фахівця з аналізу даних. <i>Soft skills</i> та <i>hard skills</i> аналітика даних. 2. Технологія <i>Knowledge Discovery in Databases</i>. Формування вибірки даних. 3. Технологія <i>Data Mining</i>. Задачі <i>Data Mining</i>: класифікація, регресія, кластеризація, асоціація, послідовність даних. 4. Актуальні бізнес-задачі. 5. Етапи аналізу даних. <p style="text-align: center;">Лекція 3. Задачі консолідації даних. Сховища даних та OLAP-системи. План лекції</p> <ol style="list-style-type: none"> 1. Поняття консолідації. Джерела даних. 2. Основні задачі консолідації даних. Схема процесу консолідації. 3. Архітектури сховищ даних. Спеціалізовані сховища даних. 4. ETL-процес. 5. OLAP-системи. Схеми побудови сховищ даних. <p style="text-align: center;">Лекція 4. Перевірка якості даних. Методи та алгоритми очищення даних. План лекції</p> <ol style="list-style-type: none"> 1. Типовий набір інструментів оцінки якості даних в аналітичній платформі. 2. Очищення від шумів і згладжування рядів даних. 3. Фільтрація даних. Відновлення пропущених значень. 4. Редагування аномальних значень. Методи корегування аномальних значень. 5. Проблема репрезентативності. Обробка дублікатів і протиріч. 6. Семплінг. Усунення незначущих факторів. 	<p>2</p> <p>2</p> <p>2</p> <p>2</p>

¹+20% інтерактивних методів навчання виділено курсивом

	<p style="text-align: center;">Лекція 5. Трансформація даних.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Основні методи трансформації даних. 2. Перетворення часових рядів. Квантування даних. 3. Сортування даних. Злиття даних. Об'єднання даних. 4. Нормалізація даних. 5. Групування та розгрупування даних. <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-4. <i>Додатковий:</i> 6, 8-10, 15, 26, 27. <i>Інтернет-ресурси:</i> 28-30</p>	2
	<p style="text-align: center;">Лабораторне заняття 1</p> <p style="text-align: center;">Основи аналізу даних у системах Deductor і Loginom</p> <p>Мета: ознайомитися з основами роботи в системах <i>Deductor Studio</i> і <i>Loginom</i>, ознайомитися з базовими конструкціями проєктів у <i>Deductor Studio</i> та <i>Loginom</i>, навчитися оперувати основними вузлами сценарію.</p> <p>Завдання: Відкрити сайт розробників аналітичних платформ <i>Deductor</i> та <i>Loginom</i>. Завантажити безкоштовну версію аналітичної платформи <i>Deductor Studio Academic</i>. Завантажити безкоштовну версію аналітичної платформи <i>Loginom Community Edition</i> https://loginom.com/download. Порівняти функціонал аналітичних платформ <i>Deductor</i> та <i>Loginom</i>. Завантажити або переглянути в режимі <i>on-line</i> демонстраційні матеріали із сайту розробника, що стосуються аналітичних платформ <i>Deductor</i> та <i>Loginom</i>.</p>	2
	<p style="text-align: center;">Лабораторне заняття 2</p> <p style="text-align: center;">Імпорт з різних джерел та первинна обробка даних у системі Deductor</p> <p>Мета: набути практичних навичок імпорту даних у систему <i>Deductor Studio</i> з використанням текстових файлів та файлів із роздільниками формату <i>csv</i>.</p> <p>Завдання: Імпортувати в <i>Deductor</i> текстовий файл згідно із завданням, коректно налаштувавши параметри імпорту. Перейменувати мітки вузлів, встановлені за замовченням. Додати до вузла вузол «Налаштування набору даних». Експортувати набір даних у текстовий файл з настройками, які пропонуються за замовченням.</p> <p>Створити файл <i>Surname_import2</i> у середовищі <i>MS Excel</i>, що містить табличні дані. Зберегти файл у форматі <i>*.csv</i>. Імпортувати файл в аналітичну платформу <i>Deductor</i> з використанням майстра імпорту.</p> <p>Використовуючи майстер обробки, створити декілька вузлів (Кореляційний аналіз, Фільтрація, Сортування, Групування тощо). Дані за введеними вузлами представити у вигляді таблиць, діаграм, гістограм. Змінити порядок гілок у сценарії. Виконати експорт окремих результатів у текстові файли.</p>	2
	<p style="text-align: center;">Лабораторне заняття 3</p> <p style="text-align: center;">Імпорт з різних джерел та первинна обробка даних у системі Loginom</p>	2

Мета: набути практичних навичок імпорту даних у аналітичній платформі *Logiplot* з використанням текстових файлів, файлів із роздільниками формату *csv* та файлів *Excel*.

Завдання: Виконати завдання імпорту на налаштування текстового файлу в аналітичній платформі *Logiplot*. Зберегти пакет під іменем *Surname_txt.lgp*

Виконати завдання імпорту та налаштування файлу із роздільниками в аналітичній платформі *Logiplot*. Зберегти пакет під іменем *Surname_csv.lgp*

Імпортувати доданий до завдання файл *Excel*. Налаштувати типи полів та інші параметри коректного імпорту. Побудувати візуалізатори (таблицю, статистику, графіки, діаграми тощо) для імпортованого файлу. Зберегти пакет під іменем *Surname_xls.lgp*

Додати до пакету в *Logiplot* декілька обробників для імпортованих наборів даних із попередніх завдань.

Лабораторні заняття 4

Оцінка якості даних

Мета: ознайомитися з основними методами передобробки даних: фільтрація, оцінка якості даних, заповнення пропусків, редагування викидів з використанням вбудованих інструментів *Logiplot*.

Завдання: У файлі *Food_data.xlsx* є дані про вартість умовного середнього набору продуктів харчування в м. Києві. Для прогнозування даного показника на перспективу необхідно заповнити пропуски, які візуально спостерігаються в таблиці.

Перетворити файл *Excel* у текстовий файл формату **.txt*. Імпортувати файл у систему *Logiplot* з вірним форматкуванням типів полів (рядковий, цілий). За допомогою налаштування візуалізації обрати спосіб відображення даних у вигляді таблиці та діаграми. Для ліквідації проблем із даними використати майстер обробки «*Imputation*» (Заповнення пропусків), встановивши при налаштуванні опцію обробки як упорядкованого набору. Переконайтеся, що після обробки дані приймуть вигляд, придатний для подальшого аналізу.

Перейменувати гілку з імпортом файлу *Food_data.txt*, щоб її назва відображала зміст виконаної переобробки даних.

Лабораторне заняття 5

Очищення даних

Мета: ознайомитися з основними методами очищення даних: парціальна передобробка, спектральна передобробка, факторний аналіз, кореляційний аналіз, усунення дублікатів і протиріч з використанням вбудованих інструментів *Logiplot*.

Завдання: Виконати спектральну обробку з даними файлу «*Trade.txt*». В якості візуалізації обрати діаграму з неперервними лініями зеленого кольору. Виконати факторний аналіз даних з файлу «*Anketa1.txt*». Виконати імпорт файлу та встановити вірні типи полів (рядковий, цілий, логічний). Залишити три фактори з чотирьох для подальшої обробки. Визначити інтервал порогу значущості, при якому з множини чотирьох факторів залишається тільки три. В таблиці залишити для відображення

2

2

тільки значення трьох факторів впливу, імені, прізвища та ймовірного результату щодо повернення кредиту.
 Виконати кореляційний аналіз даних з файлу «region.txt». Виконати імпорт файлу та встановити вірні типи полів (рядковий, цілий, дійсний). Встановити ручний вибір незначущих факторів і обрати фактори «Середня за місяць нарахована заробітна плата», «Основні фонди в економіці», «Об'єм промислової продукції», «Інвестиції в основний капітал». Поріг значущості встановлюємо на рівні 0,3. Виявити дублікати на прикладі даних файлу «Anketa.txt».

Лабораторне заняття 6 Трансформація даних

2

Мета: ознайомитися з основними методами трансформації даних: розбиття дати, квантування, налаштування набору даних, фільтрація, заміна значень, злиття з використанням вбудованих інструментів *Logiplot*.

Завдання: Здійснити перетворення дати. У файлі *Credit.csv* є інформація за кредитами, отриманими у банку. Завантажити його в аналітичну платформу з вірним форматуванням типів полів (рядковий, цілий, дата/час). Отримати дані за сумами взятих кредитів у розрізі тижнів. У якості візуалізатора обрати «Куб». Отримати крос-діаграму суми взятих кредитів за тижнями в розрізі цілей кредитування

Здійснити квантування даних. У якості початкових даних обрати дані, отримані при розбитті дати файлу «Credit.csv». У якості візуалізатора обрати куб (факт – «Сума кредиту», виміри – «Вік» та «Дата кредитування (Рік+Тиждень)»).

Виконати налаштування набору даних. У якості початкових даних обрати дані, отримані при квантуванні віку кредиторів з файлу «Credit.csv». Визначити на яку загальну суму і на який термін було надано кредити за тижнями року; визначити на яку загальну суму і на який термін було відмовлено у наданні кредитів за тижнями року.

Виконати фільтрацію даних. Початкові дані взяти із файлу *Credit.csv*. Здійснити вибірку позичальників, які отримали позику на суму менше 10000 грн. В якості візуалізації результату обрати статистику і переконатися у правильному застосуванні фільтру за знайденими мінімальним та максимальним значенням суми кредиту.

Провести заміну значень. Створити звіт про суми кредитів на різні цілі у розрізі статі позичальників. В якості візуалізації результату обрати куб (факт – «Сума кредиту», виміри – «Стать» та «Мета кредитування»)

Виконати злиття даних. Початкові дані взяти із файлів *TradeSales.txt* (обсяги продажів) та *TradeRest.txt* (обсяги залишків). Додати до даних про продажі дані про відповідні залишки. Для злиття даних вибрати майстер обробки «Join» (Злиття) і в ньому – вузол із залишками, з яким необхідно з'єднати дані за типом «Внутрішнє з'єднання».

Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Передобробка даних», ознайомлення з основною та додатковою літературою, інтернет-ресурсами,

30

	<p>виконання індивідуальних завдань з використанням комп'ютера. <i>Індивідуальна робота:</i> Розробити шаблон пакету в системі <i>Loginot</i>. В розробленому пакеті передбачити процедури імпорту даних із текстового файлу, файлу формату <i>csv</i>, файлу <i>Excel</i>. У шаблоні передбачити приєднання різних обробників подій. Використати всі можливі візуалізатори передобробки даних. Використати розроблений шаблон для виконання лабораторних робіт за темою «Передобробка даних» Список рекомендованих джерел: <i>Основний:</i> 1-4. <i>Додатковий:</i> 6, 8-10, 15, 26, 27. <i>Інтернет-ресурси:</i> 28-30</p>	
<p>Знати: основи афінитивного аналізу, принципи генерації популярних наборів даних та асоціативних правил.</p> <p>Вміти: використовувати алгоритм <i>apriori</i> для побудови та аналізу асоціативних правил.</p>	<p style="text-align: center;">Тема 2. Асоціація даних.</p> <p style="text-align: center;">Лекція 6. Асоціація даних. Асоціативні правила.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Афінитивний аналіз. Поняття типової транзакції. Предметний набір. 2. Основні поняття <i>Rules Mining</i>. Асоціативні правила. Умова та наслідок асоціативного правила. 3. Підтримка та достовірність правил. Значущість асоціативних правил. Міри корисності правил. Ліфт, левередж та покращення асоціативних правил. 4. Алгоритм <i>a priori</i>. Генерація асоціативних правил. 5. Секвенціальний аналіз. <p style="text-align: center;">Список рекомендованих джерел: <i>Основний:</i> 1-3. <i>Додатковий:</i> 7, 18, 20. <i>Інтернет-ресурси:</i> 28.</p>	2
	<p style="text-align: center;">Лабораторне заняття 7</p> <p style="text-align: center;">Асоціативні правила у стимулюванні оптових покупців</p> <p>Мета: навчитися генерувати асоціативні правила та на основі алгоритму <i>a priori</i>.</p> <p>Завдання: Дано базу транзакцій, утворену за результатами аналізу діяльності міжнародної компанії <i>Dapop</i> з продажу продукції власного виробництва. З метою збільшення прибутку було вирішено створити вигідні для оптових споживачів пакетні пропозиції з придбання декількох видів продукції одночасно, а потім пропонувати споживачам придбання інших видів продукції, яка найчастіше закуповується разом. З цією метою керівництво компанії поставило задачу перед аналітичним відділом визначити зв'язок між видами продукції, які найчастіше закуповують оптові споживачі. Представлено аналіз закупівель 20 оптових споживачів. Використовуючи алгоритм <i>a priori</i>, визначити популярні набори видів продукції, які купляють мінімум 5 споживачів. Побудувати асоціативні правила з рівнем достовірності не менше 70%. Впорядкувати правила за мірою добутку підтримки і достовірності і визначити 5 найбільш сильних правил. Дати письмову інтерпретацію знайденим правилам (визначити який вид продукції слід додатково запропонувати покупцям).</p>	2

	<p style="text-align: center;">Лабораторні заняття 8-9</p> <p>Асоціативні правила в стимулюванні роздрібних продажів Мета: Навчитися виділяти правила з максимальною достовірністю з метою розробки більш досконалої маркетингової стратегії для оптимізації закупок товарів. Оволодіти навичками роботи з обробником <i>Association Rules</i> в <i>Loginot</i> для реалізації алгоритму <i>a priori</i>. Завдання: Роздрібна мережа по продажу товарів побутової хімії поставила задачу аналізу купівельних кошиків для оптимізації розміщення товарів на вітринах і проведення крос-продажів. Відділ маркетингу надав 5000 чеків, у яких відображені покупки, зроблені клієнтами магазинів. Передбачити, які товари покупці можуть вибрати в залежності від того, що вже є в їхніх кошиках; виявити найбільш популярні товарні набори, які складаються з більше ніж одного предмету; запропонувати рекламні акції типу: «Кожному, хто купив A і B – товар C в подарунок». Використовуючи алгоритм <i>a priori</i>, побудувати асоціативні правила і зробити їх інтерпретацію. Сформувати власну базу даних для подібної задачі, провести генерацію популярних наборів і визначити систему асоціативних правил. Серед початкових даних повинно бути: не менше 20 різних видів товару; кожен чек повинен містити від 2 до 4 різних видів товару; кількість чеків не менше 50.</p>	4
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Асоціація даних», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульного контролю. <i>Індивідуальна робота:</i> Розробити шаблон пакету в системі <i>Loginot</i>. В розробленому пакеті передбачити процедури імпорту даних із файлу <i>Excel</i> з ідентифікатором транзакції та найменуванням товару або послуги. У шаблоні передбачити використання обробника «<i>Association Rules</i>». У якості візуалізаторів використати аналіз «<i>What/If</i>», популярні набори та асоціативні правила. Використати розроблений шаблон для виконання лабораторних робіт за темою «Асоціація даних» Список рекомендованих джерел: <i>Основний:</i> 1-3. <i>Додатковий:</i> 7, 18, 20. <i>Інтернет-ресурси:</i> 28.</p>	10
<p>Знати: поняття кластеризації та сегментації даних</p> <p>Вміти: застосовувати на практиці алгоритми <i>k-means, g-means</i>, проводити кластеризацію</p>	<p style="text-align: center;">Тема 3. Кластеризація даних</p> <p style="text-align: center;">Лекція 7. Алгоритми кластеризації. Кластерний аналіз.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Формальна постановка задачі кластеризації. 2. Задачі кластеризації: вивчення даних, полегшення аналізу, стиснення даних, прогнозування, виявлення аномалій. 3. Приклади кластеризації в різних областях знань. 4. Базові алгоритми кластеризації. 5. Ієрархічні та ітеративні методи кластерного аналізу. 6. Карти та мережі Кохонена. 	2

<p>даних, використовувати аналітичні платформи для задач кластеризації та сегментації даних</p>	<p>Список рекомендованих джерел: <i>Основний:</i> 1- 3. <i>Додатковий:</i> 5, 7, 20. <i>Інтернет- ресурси:</i> 28.</p>	
	<p>Лабораторні заняття 10-11 Алгоритм k-means. Загальні принципи кластеризації. Мета: набуття практичних навичок роботи з алгоритмом кластеризації <i>k-means</i>. Завдання: Дано 10 об'єктів (A, B, C, D, E, F, G, H, I, K), які представлено координатами точок на площині. Використовуючи алгоритм <i>k-means</i>, провести кластерний аналіз даних, розбивши початкову множину точок на 2 кластери. За початкові центри кластерів прийняти точки E і J. Використовуючи алгоритм <i>k-means</i>, провести кластерний аналіз даних, розбивши початкову множину точок на 3 кластери. За початкові центри кластерів прийняти точки A, F і J. Порівняти отримані результати і обрати оптимальну кількість кластерів. На кожному етапі кластерного аналізу провести візуалізацію результатів розбиття на кластери. Для визначення належності об'єкта до певного кластеру використати евклідову відстань та відстань Манхеттена.</p>	4
	<p>Лабораторні заняття 12-13 Задача кластеризації регіонів України Мета: отримання практичних навичок використання алгоритмів кластеризації в аналітичній платформі <i>Logiplot</i>. Завдання: Існує база даних про обсяг послуг на душу населення за регіонами України. Відібрати п'ять показників: обсяг платних послуг, обсяг побутових послуг, обсяг транспортних послуг, обсяг послуг зв'язку та обсяг комунальних послуг. Необхідно розподілити регіони України на функціональні групи за спостережуваним обсягом наданих послуг та виявити приховані закономірності. Існує база даних ріелтєрської фірми, що містить 1720 записів оцінок вартості проданого житла в одному з міст України. Для створення правил попередньої оцінки вартості житла, окрім продажної вартості, фіксувалися також 11 різних параметрів квартири. Необхідно провести кластерний аналіз квартир з метою виділення чотирьох типів квартир, що існують на ринку, а також кластеризацію змінних, виділивши групи змінних, що характеризують одну властивість.</p>	4
	<p>Лабораторне заняття 14 Кластерний аналіз з використанням нейромережі Мета: проведення кластеризації даних за допомогою нейронної мережі. Завдання: Використати алгоритм кластеризації <i>g-means</i> для класифікації даних на прикладі Ірисів Фішера. База даних міститься у файлі <i>Iris.txt</i> та складається з даних про 150 примірників ірису, по 50 примірників з трьох видів - Ірис щетинистий (<i>Iris setosa</i>), Ірис віргінський (<i>Iris virginica</i>) і Ірис різнокольоровий (<i>Iris versicolor</i>). Для кожного екземпляра</p>	2

вимірювалися чотири характеристики (в сантиметрах): довжина чашолистка (англ. *Sepal length*); ширина чашолистка (англ. *Sepal width*); довжина пелюстки (англ. *Petal length*); ширина пелюстки (англ. *Petal width*). На підставі цього набору даних побудувати правило класифікації, що визначає вид рослини за даними вимірів. Провести класифікацію ірисів за допомогою нейронної мережі з використанням інструменту «*Neural Network (Classification)*» в системі *Loginot*.

Лабораторні заняття 15-16

Сегментація даних на основі карт Кохонена

Мета: набуття навичок роботи з обробником і візуалізатором *Self-Organizing Networks* в процесі розв'язування бізнес-задачі сегментації в *Loginot*.

Завдання: Керівництво філії регіональної телекомунікаційної компанії, яка надає послуги мобільного зв'язку, поставило задачу сегментації абонентської бази. Її цілі полягають в наступному:

- побудова профілів абонентів шляхом виявлення їх схожої поведінки в плані частоти, тривалості та часу дзвінків, а також щомісячних витрат;

- оцінка найбільш і найменш прибуткових сегментів.

Ця інформація може в подальшому використовуватися для:

- розробки маркетингових акцій, направлених на визначення групи клієнтів;

- розробки нових тарифних планів;

- оптимізації витрат на адресу SMS-розсилку про нові послуги і тарифи;

- попередження відтоку клієнтів у інші компанії.

Вихідні дані представлені у вигляді файлу *mobile.txt*.

Розв'язати бізнес-задачу сегментації абонентів за допомогою підходу, який базується на алгоритмі Кохонена і складається з двох кроків: кластеризація об'єктів алгоритмом Кохонена; побудова та інтерпретація карти Кохонена.

4

Лабораторне заняття 17

Модульна контрольна робота 1

«Передобробка, асоціація та кластеризація даних»

Мета: перевірити рівень знань, отриманих студентами при вивчення передобробки, асоціації та кластеризації даних.

Завдання: Існують дані для аналізу споживчого кошику роздрібною мережі, що займається продажем продовольчих товарів. Дані представлені за 10 магазинами мережі. Набір даних становить 2500 чеків.

1. Провести аналіз якості даних, за потреби вилучити аномальні значення, заповнити пропущені дані, усунути дублікати значень які логічно суперечать умові задачі.

2. Провести кластеризацію магазинів, виділивши два кластера.

3. Здійснити аналіз споживчого кошику за допомогою пошуку асоціативних правил для магазинів, що увійшли у найменший за обсягом кластер.

2

	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Кластеризація даних», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульного контролю.</p> <p><i>Індивідуальна робота:</i> Розробити шаблон пакету в системі <i>LogiNot</i> для проведення кластеризації даних. В розробленому пакеті передбачити процедури імпорту даних із файлу <i>Excel</i> з характеристиками об'єктів, що підлягають кластеризації. У шаблоні передбачити використання обробників «<i>Clustering</i>», «<i>Transaction Clustering</i>», «<i>EM Clustering</i>».</p> <p>Використати розроблений шаблон для виконання лабораторних робіт за темою «Кластеризація даних»</p> <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1- 3. <i>Додатковий:</i> 5, 7, 20. <i>Інтернет- ресурси:</i> 28.</p>	16
<p>Знати: основні методи класифікації даних, регресії та прогнозування даних, логістичної регресії</p> <p>Вміти: проводити класифікацію даних з використанням дерева рішень, здійснювати регресійний аналіз та прогнозування даних з використанням лінійної, нелінійної, динамічної та логістичної регресії.</p>	<p style="text-align: center;">Тема 4. Класифікація та регресія даних</p> <p style="text-align: center;">Лекція 8. Класифікація даних. Дерева рішень.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Огляд методів класифікації. Процес класифікації. 2. Точність класифікації. Оцінювання класифікаційних методів. 3. Структура дерева рішень. Алгоритми побудови дерев рішень. 4. Міри ефективності дерев рішень. Критерії вибору найкращих атрибутів розгалуження. Критерії оптимізації дерев рішень. 5. Алгоритми машинної класифікації. <p style="text-align: center;">Лекція 9. Регресійний аналіз даних. Логістична регресія.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Кореляційно-регресійний аналіз. Статистичні методи аналізу. 2. Лінійна регресія. Регресія з категоріальними вхідними змінними. 3. Логістична регресія. Оцінки максимальної правдоподібності. 4. Тест Чоу. ROC-аналіз. 5. Множинна логістична регресія. Простий байесівській класифікатор. <p style="text-align: center;">Лекція 10. Технології прогнозування даних. Аналіз часових рядів.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Методи прогнозування даних. 2. Часовий ряд та його компоненти. 3. Моделі прогнозування часових рядів. 4. Прогнозування в торгівлі та логістиці. <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1, 3. <i>Додатковий:</i> 6, 7, 11, 12, 14. <i>Інтернет- ресурси:</i> 28.</p>	2 2 2

Лабораторні заняття 18-19

4

Регресійний аналіз даних. Лінійна і квадратична регресія

Мета: обробка даних та побудова прогнозу з використанням обробника *Linear Regression*.

Завдання:

1. *Лінійна регресія.* Є набір даних щодо вартості квартир в залежності від кількості кімнат, загальної площі, житлової площі, площі кухні, типу будинку тощо (усього 11 характеристик житла). Необхідно побудувати регресійну модель, яка дозволить оцінити вартість одно- та двокімнатних квартир.

Імпортуйте в *LogiNot* набір даних з файлу *Flats.xlsx*. За допомогою вузла *Row Filter* здійсніть відбір одно- та двокімнатних квартир. На основі відфільтрованих даних побудуйте лінійну модель з використанням обробника *Linear Regression*. При налаштуванні параметрів лінійної регресії:

- поділ на навчальну та тестову вибірку не здійснювати;
- обмеження на діапазон вихідних значень не задавати;
- в якості методу відбору змінних використати прямий відбір (*forward*). Імовірності включення та видалення змінної залишити за замовченням;
- в якості результату аналізу обрати *Regression Report* (Звіт за регресією).

На основі проведеного аналізу: оцініть якість регресійної моделі за допомогою коефіцієнта детермінації; побудуйте регресійну модель множинної лінійної регресії виду

$$y = a_0 + a_1x_1 + a_2x_2 \dots + a_nx_n$$

2. *Квадратична регресія.* Є дані щодо заробітної плати робітників, їх вік та стаж. Необхідно визначити залежність заробітної плати від віку робітників.

Імпортуйте в *LogiNot* набір даних з файлу *Salary.xlsx*. Налаштуйте типи полів згідно із даними у файлі імпорту. При виборі засобів відображення результату оберіть таблицю та діаграму залежності заробітної плати від стажу робітника.

За допомогою вузла *Sort* здійсніть упорядкування записів вибірки за зростанням заробітної плати. У якості засобів відображення результату оберіть таблицю та діаграму залежності заробітної плати від віку робітника

Зобразити залежність заробітної плати від віку робітників за допомогою моделі квадратичної регресії:

$$y = a_0 + a_1x + a_2x^2$$

Звести модель до лінійного вигляду шляхом введення додаткового стовпця «Вік у квадраті». Для цього використати метод обробки *Calculator*. У якості засобів відображення результату оберіть таблицю.

Побудувати модель регресії з використанням майстра обробки *Linear Regression*. При налаштуванні параметрів лінійної регресії:

- поділ на навчальну та тестову вибірку не здійснювати;
- обмеження на діапазон вихідних значень не задавати;
- в якості метода відбору змінних використати повне включення (*Enter*). Імовірності включення та видалення змінної залишити за замовченням;
- в якості результату аналізу обрати *Regression Report* (Звіт

за регресією) та Diagram (Діаграма).
На основі проведеного аналізу оцініть якість регресійної моделі за допомогою коефіцієнта детермінації.

Лабораторне заняття №20

Прогнозування даних. Динамічна регресія

Мета: Побудова прогнозу на основі динамічної регресії з використанням обробника *Linear Regression* в системі *LogiNot*.

Завдання: Є набір даних за місяцями поточного року про загальний обсяг роздрібного товарообігу магазину, який нещодавно відкрито у торговельному центрі.

Необхідно побудувати прогнозну модель часового ряду і здійснити прогноз товарообігу до липня наступного року.

Імпортуйте в *LogiNot* набір даних з файлу *Turnover.xlsx*. Налаштуйте типи полів згідно із даними у файлі імпорту (цілий тип, дійсний тип, дата). Оберіть спосіб відображення даних імпорту у вигляді таблиці та діаграми. Побудуйте лінійну модель на основі імпортованих даних з використанням обробника *Linear Regression*. При налаштуванні параметрів лінійної регресії:

- поділ на навчальну та тестову вибірку не здійснювати (навчальна вибірка становить 100%);
- обмеження на діапазон вихідних значень не задавати;
- в якості метода відбору змінних використати повне включення (*Enter*). Імовірності включення та видалення змінної залишити за замовченням;
- в якості результату аналізу обрати *Regression Report* (Звіт за регресією) та *Diagram* (Діаграма);
- при налаштуванні стовпців діаграми мітками стовпця вказати *Роздрібний товарообіг та розрахований за рівнянням регресії Роздрібний товарообіг_OUT*. У якості підписів даних обрати *Місяць*

На основі проведеного аналізу:

- оцініть якість регресійної моделі за допомогою коефіцієнта детермінації;
- оцініть статистичну значущість моделі (модель є статично значущою, якщо параметр значущості у таблиці дисперсійного аналізу близький до нуля);
- побудуйте динамічну регресійну модель лінійної регресії виду

$$y = a_0 + a_1 t$$

Здійсніть прогнозування за допомогою динамічної регресійної моделі з використанням обробника подій *ARIMAX*. При налаштуванні параметрів прогнозування:

- додати до прогнозної моделі крок прогнозу і початкові дані
- прогнозування здійснити на 2 місяці наперед;
- в якості засобу відображення обрати *Forecasting Diagram*.

Вивести на графік діаграми значення початкових даних товарообігу та прогнозні значення на майбутній період. Перевірити узгодженість знайдених прогнозних значень із отриманим рівнянням лінійної регресії, підставивши в рівняння прогнозні значення періодів.

2

	<p style="text-align: center;">Лабораторне заняття №21 Декомпозиція часових рядів</p> <p>Мета: Побудова прогнозу часового ряду в <i>Logiplot</i> з урахуванням тренду та сезонності.</p> <p>Завдання: Є набір даних за тижнями чотирьох років про споживчі ціни на свіжі помідори в цілому по Україні. Необхідно провести декомпозицію часового ряду на складові: тренд, сезонність та залишок.</p> <p>Імпортуйте в <i>Logiplot</i> набір даних з файлу <i>Tomatoes.xlsx</i>. Налаштуйте типи полів згідно із даними у файлі імпорту (цілий тип, дійсний тип, рядковий тип). Оберіть спосіб відображення даних імпорту у вигляді таблиці та діаграми.</p> <p>Обчисліть період сезонності на основі перетворених даних. В якості методу обробки використовуйте вузол <i>Autocorrelation</i>.</p> <p>При налаштуванні параметрів автокореляції:</p> <ul style="list-style-type: none"> - кількість відліків встановити рівним 164 (за кількістю періодів спостереження); - в якості результату аналізу обрати таблицю та діаграму; - при налаштуванні стовпців діаграми міткою стовпця вказати Споживча ціна на свіжі помідори. У якості підписів даних обрати Лаг. <p>Здійснити декомпозицію часового ряду на сезонну і трендову складові. При налаштуванні параметрів декомпозиції:</p> <ul style="list-style-type: none"> - період сезонності встановити рівним 12 місяців; - тип тренду – лінійний; - в якості засобу відображення обрати діаграму декомпозиції часового ряду та налаштування тренду і сезонних індексів; - при налаштуванні діаграми декомпозиції поставити мітки біля всіх полів, окрім початкового ряду без тренду. <p>На основі проведеного аналізу тренду та сезонних індексів</p> <ul style="list-style-type: none"> - записати рівняння тренду у вигляді <ul style="list-style-type: none"> - $y = a \cdot t + b$ - за знайденим рівнянням оцінити величину (в середньому, в грн) щотижневого зниження споживчої ціни на свіжі помідори. 	2
	<p style="text-align: center;">Лабораторне заняття №22 Прогнозування з використанням нейронних мереж</p> <p>Мета: засвоїти методи й засоби прогнозування в системі <i>Logiplot</i> при інтелектуальному аналізі даних за допомогою нейронних мереж.</p> <p>Завдання: Існує файл <i>multi.txt</i>, що містить таблицю з трьома полями: <i>Аргумент1</i>, <i>Аргумент2</i>, <i>Добуток</i>. Деякі з добутків у даному файлі відсутні. Необхідно побудувати модель прогнозу добутку, подаючи на вхід моделі два множника і отримуючи на виході їх добуток.</p> <p>Для побудови моделі використати обробник <i>Neural Network (Regression)</i>. Поля <i>Аргумент1</i> та <i>Аргумент2</i> задати як вхідні, поле <i>Добуток</i> – як вихідне. Здійснити розбиття початкової множини на навчаючу та тестову вибірки. Спосіб розбиття – <i>Random</i>, обсяг навчаючої вибірки – 100%. Задати функцію активації,</p>	2

кількість прихованих шарів та кількість нейронів у шарі. У якості алгоритму навчання обрати *Resilient Propagation (RPROP)*. Приклад вважати розпізнаним, якщо помилка менша за 0.005. Умовою зупинки навчання вказати 100000 epoch. Візуалізаторами роботи нейромережі слід обрати діаграму, діаграму розсіювання, граф нейромережі та аналіз «What/If».

4

Лабораторні заняття №23-24
Логістична регресія. Оцінка кредитоспроможності позичальників.

Мета: побудова моделі бінарної логістичної регресії та визначення рівня її якості.

Завдання: Є набір даних про кредитні історії 700 клієнтів банку і 150 клієнтів, які мають намір звернутися до банку за кредитом. Інформація про клієнтів банку містить такі дані: вік, освіта, стаж комерційної діяльності, кількість повних років постійного місця проживання, річний дохід (у тисячах доларах США), борги (% від річного доходу), борги за кредитною картою (тисяч доларів США), інші борги (тисяч доларів США). Для клієнтів банку зазначається, чи є він боржником банку.

Необхідно, використовуючи 70% випадкову вибірку клієнтів, які вже брали кредит у банку, побудувати регресійну логістичну модель, яка дозволить передбачити неповернення кредиту клієнтом. Використовуючи 30% клієнтів, що залишилися, необхідно з'ясувати ступінь придатності побудованої моделі для передбачення випадків неповернення кредиту клієнтами банку. Визначити, які змінні можуть бути виключені з моделі без втрати її якості.

Імпортувати в Logiport набір даних з файлу *Bank Credit.xlsx*. Налаштувати типи полів згідно із даними у файлі імпорту (цілий тип, дійсний тип). Обрати спосіб відображення даних імпорту у вигляді таблиці. Перевірте якість даних. У звіті наявні пропуски, викиди і екстремальні значення. Методи редагування викидів та екстремальних значень залишити без змін. На основі вузла оцінки якості даних заповнити пропуски в базі даних, використавши майстер обробки *Imputation*. На основі бази даних, що не містить пропусків, провести редагування викидів, використавши майстер обробки *Eliminate Outliers*.

Для прогнозування ймовірності відгуку на подію сформуванню цільову бінарну змінну «Відгук» на основі змінної «Чи є клієнт боржником банку?» В якості методу обробки використати вузол *Calculator*. Розбити множину клієнтів на дві вибірки: навчальну і тестову. У якості методу обробки використати вузол *Partitioning*. При налаштуванні розбиття:

- визначити обсяг навчальної множини у 70%;
- задати метод вибірки даних – *Stratified*, який забезпечує однакову частку відгуків в обох вибірках;
- для здійснення стратифікації вказати поле *Відгук*;
- метод відображення даних – *Partitioning*.

Побудувати модель прогнозування ймовірності перспектив повернення боргу кредитором за допомогою логістичної регресії,

обравши в якості метода обробки вузол *Logistic Regression*.
 При налаштуванні параметрів логістичної регресії:

- обсяг навчальної множини встановити на рівні 70%;
- спосіб поділу початкової множини *By Column*;
- метод відбору змінних в регресійній моделі – прямий відбір (*Forward*);
- істинні частоти «подій» та «не-подій» визначити із навчальної множини;
- параметри перетворення шансів у лінійну шкалу: референсна точка – 600, відношення шансів у референсній точці – 50, крок подвоєння відношення шансів – 20, обрати тільки додатні бали з округленням до цілого значення;
- в якості способів відображення даних обрати коефіцієнти регресії, звіт за регресією, якість класифікації та «*What/If*».

Провести аналіз отриманих візуалізаторів логістичної регресії.
 Провести аналіз візуалізатора коефіцієнтів регресії. Провести аналіз результатів за візуалізатором якості класифікації, побудувавши *ROC*-криву та *Lift*-криву, а також розрахувати основні метрики бінарних класифікаторів: *AUC*, *KS*, *Gini*.
 Провести аналіз візуалізатора «*What/If*».

2

Лабораторне заняття №25

Скорингова карта для оцінки кредитоспроможності

Мета: побудова скорингової карти для прогнозування кількості відмов у видачі кредитів для клієнтів банку.

Завдання: У комерційному банку є продукт «Нецільовий споживчий кредит»: позики надаються на будь-які цілі з прийняттям рішення за один день. В даний час рішення про видачу кредиту приймається на основі скорингової карти, побудованої експертним способом, з відсотком відмови, що дорівнює 55%, при цьому обсяг простроченої заборгованості великий. Накопичена статистична інформація про позичальників і якості обслуговування ними боргу за кілька місяців. Керівництво банку, розуміючи, що високий рівень відмов перешкоджає розширенню роздрібного бізнесу в області споживчого кредитування, поставило перед відділом роздрібних ризиків завдання розробити нову скорингову карту, яка дозволила б значно скоротити число відмов у видачі та знизити суму простроченої заборгованості.
 Імпортувати файл *loans.txt*, який містить 2709 кредитів з відомими результатами платежів протягом кількох місяців після отримання кредиту. Використати правило: якщо клієнт банку має хоча б одне порушення термінів повернення кредиту більш як на 60 днів, то він відноситься до класу ненадійних.
 Провести квантування клієнтів на три категорії за віком. Побудувати модель логістичної регресії та визначити її коефіцієнти. Побудувати таблиці спряженості для тестової та навчальної вибірки. На основі коефіцієнтів логістичної регресії побудувати скорингову карту. Провести прогнозування кількості відмов у кредитуванні на основі скорингової моделі.

	<p style="text-align: center;">Лабораторне заняття №26 Дерева рішень</p> <p>Мета: набуття практичних навичок роботи з деревами рішень як важливим інструментом класифікації даних.</p> <p>Завдання: <i>Завдання 1. Визначення кредитного рейтингу. Є набір даних про кредитні рейтинги 2464 клієнтів банку (результуюча змінна), які користувалися банківськими послугами в минулому, а також такі дані про клієнтів, як вік (кількість років), рівень доходу (низький, середній високий), кількість наявних кредитних карт (менше 5; 5 і більше), рівень освіти (вища, середня) і кількість кредитів на покупку автомобіля (0; 1; 2 і більше).</i></p> <p><i>Необхідно за допомогою дерева рішень з'ясувати, за якими правилами є можливість визначити клієнта з негативним кредитним рейтингом (високим ризиком неповернення кредиту).</i></p> <p><i>Завдання 2. Аналіз даних ріелтєрської фірми. У файлі Realtor.xlsx є база даних ріелтєрської фірми, що містить 1721 запис оцінок вартості проданого житла у Києві. Для створення правил попередньої оцінки вартості житла, крім продажної вартості, фіксувалося ще 11 різних параметрів квартири.</i></p> <p><i>Потрібно за допомогою дерева рішень створити правила, за якими призначається низька (менше 50 тис. дол.), Середня (від 50 тис. до 100 тис. дол.) і висока (більше 100 тис. дол.) продажна вартість квартири.</i></p> <p style="text-align: center;">Лабораторне заняття №27 Факторний аналіз даних</p> <p>Мета: набуття практичних навичок використання методу головних компонент обробника подій <i>Factor Analysis</i> системи <i>Logitot</i>.</p> <p>Завдання: <i>Завдання 1. Зниження вимірності даних. Представлено дані розподілу областей України за професійно-кваліфікаційною структурою керівників та спеціалістів сільськогосподарських підприємств. Для характеристики професійно-кваліфікаційної структури обрано дев'ять показників.</i></p> <p><i>Необхідно:</i></p> <ol style="list-style-type: none"> 1) побудувати матрицю фактичних навантажень, використовуючи метод головних компонент; 2) виділити головні фактори, що визначають рівень забезпеченості кадрами; 3) знайти чисельні значення головних компонент для всіх об'єктів спостережень. <p><i>Завдання 2. Зниження вимірності даних для регресійної моделі. Існують дані щодо врожайності картоплі у 36 фермерських господарствах і факторів, які на неї впливають.</i></p> <p><i>Необхідно, використовуючи метод головних компонент, скоротити кількість факторів, що впливають на врожайність картоплі, і побудувати регресійну модель залежності врожайності від визначених факторів.</i></p> <p style="text-align: center;">Лабораторне заняття 28 Модульна контрольна робота 2 «Класифікація та регресія даних»</p> <p>Мета: перевірити рівень знань, отриманих студентами при вивченні теми «Класифікація та регресія даних».</p>	<p style="text-align: center;">2</p> <p style="text-align: center;">2</p> <p style="text-align: center;">2</p>
--	--	--

	<p>Завдання: Початковими даними є витрати на рекламу туристичної агенції, помісячні обсяги продажів туристичного продукту та факти відміни авіарейсів для подорожуючих за останні три роки. Використовуючи апарат лінійної регресії побудувати регресійну модель залежності обсягів продажу від витрат на рекламу та здійснити прогноз діяльності туристичного підприємства на наступні півроку.</p> <p>Використовуючи модель динамічної регресії, розрахувати прогноз урахуванням фактору сезонності.</p> <p>Використовуючи модель логістичної регресії, побудувати прогноз відміни авіарейсу для клієнта туристичної компанії.</p>	
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Класифікація та регресія даних», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульного контролю.</p> <p><i>Індивідуальна робота:</i></p> <p>Розробити шаблон пакету в системі <i>Loginom</i> для проведення класифікації та регресії даних і побудови прогнозу моделі. В розробленому пакеті передбачити процедури імпорту даних із файлу <i>Excel</i> з даними для прогнозу. У шаблоні передбачити використання обробників «<i>Linear Regression</i>», «<i>Logistic Regression</i>», «<i>ARIMAX</i>».</p> <p>Використати розроблений шаблон для виконання лабораторних робіт за темою «Класифікація та регресія даних»</p> <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1, 3. <i>Додатковий:</i> 6, 7, 11, 12, 14. <i>Інтернет-ресурси:</i> 28.</p>	20
<p>Знати: принципи візуального аналізу даних (<i>Visual Mining</i>), аналізу текстової інформації (<i>Text Mining</i>), аналізу даних в режимі реального часу (<i>Real-Time Mining</i>), аналізу даних з мережі Інтернет (<i>Web Mining</i>).</p>	<p style="text-align: center;">Тема 5. Технології інтелектуальної обробки даних</p> <p style="text-align: center;">Лекція 11. Візуальний аналіз даних - <i>Visual Mining</i>.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Виконання візуального аналізу даних. 2. Характеристики засобів візуалізації даних. Методи візуалізації. 3. Методи геометричних перетворень. 4. Методи, орієнтовані на пікселі. Ієрархічні образи. <p style="text-align: center;">Лекція 12. Аналіз текстових даних - <i>Text Mining</i>.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Методи класифікації текстових документів. 2. Попередня обробка тексту. Видалення стоп-слів. Стеммінг. <i>N</i>-грами. Зведення реєстра. 3. Класифікація текстів. Методи кластеризації текстових документів: ієрархічні, бінарні. 4. Задача анотування текстів. Пошук асоціацій. Первинний витяг ключових понять. Навігація текстом. Засоби аналізу текстової інформації. <p style="text-align: center;">Лекція 13. Аналіз даних у реальному часі - <i>Real Time Data Mining</i>.</p>	<p>2</p> <p>2</p> <p>2</p>

	<p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Ідея <i>Data Mining</i> у реальному часі. Адаптація системи до загальної концепції. Адаптивне добування даних. 2. Статичний <i>Data Mining</i> і <i>Data Mining</i> в реальному часі 3. Рекомендаційні машини. Класифікація рекомендаційних машин. 4. Аналіз ринкового кошика і секвенційний аналіз. 5. Інструменти <i>Data Mining</i> у реальному часі. <p style="text-align: center;">Лекція 14. Аналіз даних мережі Інтернет - <i>Web Mining</i>.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Складнощі аналізу даних з мережі Інтернет. 2. Етапи <i>Web Mining</i>. Категорії <i>Web Mining</i>. Аналіз використання веб-ресурсів. 3. Методи <i>Web Mining</i> з точки зору вирішуваних завдань і реалізованих підходів. 4. Пошук інформації. Аналіз структури сегмента мережі. Використання знань з веб-ресурсів. 5. Персоналізація інформації. Пошук шаблонів в поведінці користувачів. <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-3. <i>Додатковий:</i> 5-7, 9, 11, 12, 14-18, 26. <i>Інтернет-ресурси:</i> 28.</p>	2
	<p>Самостійна робота: опрацювання матеріалу лекцій з теми «Технології інтелектуальної обробки даних», ознайомлення з основною та додатковою літературою, інтернет-ресурсами.</p> <p><i>Індивідуальна робота:</i> На основі опрацювання спеціалізованої літератури, фахових періодичних видань, Internet-джерел підготувати за допомогою програмного забезпечення (<i>Microsoft PowerPoint, Prezi, Keynote</i> тощо) презентацію на одну із запропонованих тем:</p> <ol style="list-style-type: none"> 1. Деревовидні карти та канонічні дерева у процесі <i>Visual Mining</i>. 2. Рекомендаційні системи зі штучним інтелектом. 3. Веб-системи аналітики даних у режимі реального часу. 4. Інтелектуальний аналіз текстових даних. 5. Інструментальні засоби добування інформації з Інтернет-джерел. <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-3. <i>Додатковий:</i> 5-7, 9, 11, 12, 14-18, 26. <i>Інтернет-ресурси:</i> 28.</p>	10
<p>Знати: принципи роботи аналітичних платформ <i>Deductor Studio, Loginom, RapidMiner,</i></p>	<p style="text-align: center;">Тема 6. Інструментальні засоби аналізу даних</p> <p style="text-align: center;">Лекція 15. Аналіз даних в аналітичних платформах</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Програмне забезпечення в області аналізу даних. 2. Аналітичні платформи <i>Deductor Studio</i> та <i>Loginom</i>. 3. Системи аналітики даних <i>RapidMiner, Tableau, Weka, Orange, NodeXL, Qlik</i>. 	2

<p><i>Tableau, Weka, Orange, NodeXL, Qlik, Microsoft Power BI.</i></p> <p>Вміги: використовувати середовище <i>Desktop</i> для імпорту даних з різних джерел</p>	<p>Лекція 16. Технології аналізу даних у продуктах Microsoft Corporation. Архітектура Power BI План лекції</p> <ol style="list-style-type: none"> 1. Технології аналізу даних у продуктах Microsoft Corporation. 2. Технологія моделювання даних у Microsoft Power Pivot. 3. Інтерактивний інструмент Microsoft Power View для дослідження, графічного відображення та представлення даних. 4. Надбудова Microsoft Power Query в задачах бізнес-аналітики. 5. Хмарні технології Microsoft для аналізу та візуалізації даних. 6. Організація бізнес-аналітики рівня Business Intelligence (BI). <p>Лекція 17. Імпорт даних до Power BI Desktop План лекції</p> <ol style="list-style-type: none"> 1. Налаштування середовища Power BI. 2. Інтерфейс Power BI Desktop. 3. Завантаження даних Power BI з різних інформаційних джерел. 4. Імпорт даних із реляційних баз даних, текстового файлу, вхідного каналу даних та сервісів аналізу. <p>Список рекомендованих джерел: <i>Основний:</i> 3, 4. <i>Додатковий:</i> 5-7, 9, 11, 12, 14-19. <i>Інтернет-ресурси:</i> 28-30.</p>	<p>2</p> <p>2</p>
	<p>Лабораторне заняття 29 Вступ до Power BI. Створення бази даних в системі Power BI Desktop</p> <p>Мета: набуття базових навичок роботи в аналітичній платформі Power BI.</p> <p>Завдання: У таблиці представлено дані Державної служби статистики України http://www.ukrstat.gov.ua щодо цін за оренду однокімнатної квартири у м. Києві. Здійснити реєстрацію у програмі Microsoft Power BI Desktop. Безпосередньо в програмі ввести дані з таблиці, використовуючи дані за останні 24 місяці. Налаштувати необхідний формат даних таблиці (ціле число, дата, текстове поле, грошові одиниці). Провести візуалізацію даних, використовуючи такі інструменти: - графік; - стовпчикова діаграма; - кругова діаграма.</p> <p>Змінити кольори елементів візуалізацій, шрифт і стиль надписів, встановлених за замовченням. Продемонструвати інтерактивність візуалізацій, по черзі виділивши їх окремі компоненти. Побудувати агреговані таблиці, здійснивши ієрархію дат до рівня кварталу та року. Таблиці розташувати на окремих аркушах в Power BI. Використавши панель аналітики, за даними таблиці: - встановити на графіку лінії максимуму, мінімуму та середнього значення; - побудувати прогноз ціни оренди однокімнатної квартири на наступні 6 місяців.</p> <p>Лабораторні заняття 30-31 Імпорт даних із текстових, csv-файлів та файлів Excel</p>	<p>2</p>

	<p>Мета: отримання практичних навиків імпорту даних до аналітичної платформи <i>Microsoft Power BI Desktop</i> із зовнішніх джерел у вигляді текстових файлів з роздільниками та файлів <i>Excel</i>.</p> <p>Завдання: Створіть новий файл <i>Power BI Desktop</i> і завантажте таблицю синглів із файлу <i>million-sellers.csv</i>, що продаються мільйонами примірників, на веб-сайті <i>The Guardian</i>. Створіть стовпчикову діаграму мільйонних продажів відомих виконавців. Переконайтеся, що насиченість кольору встановлена в кольоровій діаграмі відповідно до обсягу продажів і міток даних, які застосовуються в кожному рядку стовпців. Створіть кругову діаграму, яка показує обсяг продажів за рік. Змініть налаштування деталізації, щоб після натискання виконавця на графіку стовпців першої діаграми, відбувалася відповідна фільтрація даних кругової діаграми. Завантажте файл <i>Menu.xlsx</i>, який містить розбивку калорій для різних типів піци <i>Pizza Express</i>. Створіть діаграму для порівняння середніх калорій для 3 різних типів піци. Додайте легенду до кругової діаграми (щоб вона з'явилася внизу) і налаштуйте мітки даних, щоб показати значення категорії та деталей для кожного фрагмента. Додайте кільцеву діаграму, щоб показати розподіл калорій за кожним видом піци. Діаграма повинна відображати середнє значення кількості калорій для кожної піци. Налаштуйте візуальну взаємодію так, щоб при натисканні на тип піци на першій діаграмі отримувати фільтровані дані на другій.</p> <p style="text-align: center;">Лабораторні заняття 32-33</p> <p style="text-align: center;">Імпорт даних із баз даних та веб-ресурсів</p> <p>Мета: отримання практичних навиків даних із веб-ресурсів.</p> <p>Завдання: Завантажте дані про переможців Чемпіонату Європи з футболу УЄФА, які можна знайти в таблиці результатів на сторінці Вікіпедії за адресою http://en.wikipedia.org/wiki/UEFA_European_Football_Championship Не використовуйте україномовну або російськомовну версії цієї сторінки, оскільки вони відрізняються від англійської версії, яка є найбільш повною. Зробити попередню очистку даних перед завантаженням їх у платформу. Відкрийте редактор <i>Power Query</i> та видаліть всі непотрібні стовпці, залишивши тільки <i>Year (Рік)</i> і <i>Final Winners (Переможці фіналу)</i>. З використанням меню <i>Replace</i> виконайте заміну значень, видаливши слово «<i>Details</i>» з усіх рядків стовпця <i>Year</i>. За допомогою пункту <i>Rename</i> залишити у стовпці <i>Final Winner</i> тільки переможців фіналів, яким можна присвоїти ім'я <i>Country (Країна)</i>. Створити візуалізацію звіту з використанням інструменту <i>Map</i>. Змінити величину маркера країни на карті, щоб розмір маркера відображав кількість перемог для кожної країни. Розфарбувати точку даних <i>Германія</i> в різні кольори, щоб розділити переможців з ФРН та об'єднаної Німеччини. Змінити колір та стиль заголовку карти. Побудуйте кругову діаграму, що відображає внесок кожної країни у загальну кількість перемог. Проведіть сортування секторів діаграми так, щоб країни найбільшою кількістю перемог відображалися на круговій діаграмі поруч.</p>	4
	<p>Мета: отримання практичних навиків даних із веб-ресурсів.</p> <p>Завдання: Завантажте дані про переможців Чемпіонату Європи з футболу УЄФА, які можна знайти в таблиці результатів на сторінці Вікіпедії за адресою http://en.wikipedia.org/wiki/UEFA_European_Football_Championship Не використовуйте україномовну або російськомовну версії цієї сторінки, оскільки вони відрізняються від англійської версії, яка є найбільш повною. Зробити попередню очистку даних перед завантаженням їх у платформу. Відкрийте редактор <i>Power Query</i> та видаліть всі непотрібні стовпці, залишивши тільки <i>Year (Рік)</i> і <i>Final Winners (Переможці фіналу)</i>. З використанням меню <i>Replace</i> виконайте заміну значень, видаливши слово «<i>Details</i>» з усіх рядків стовпця <i>Year</i>. За допомогою пункту <i>Rename</i> залишити у стовпці <i>Final Winner</i> тільки переможців фіналів, яким можна присвоїти ім'я <i>Country (Країна)</i>. Створити візуалізацію звіту з використанням інструменту <i>Map</i>. Змінити величину маркера країни на карті, щоб розмір маркера відображав кількість перемог для кожної країни. Розфарбувати точку даних <i>Германія</i> в різні кольори, щоб розділити переможців з ФРН та об'єднаної Німеччини. Змінити колір та стиль заголовку карти. Побудуйте кругову діаграму, що відображає внесок кожної країни у загальну кількість перемог. Проведіть сортування секторів діаграми так, щоб країни найбільшою кількістю перемог відображалися на круговій діаграмі поруч.</p>	4

	<p style="text-align: center;">Лабораторне заняття №34 Дослідження Power Pivot</p> <p>Мета: набуття навичок роботи з <i>Power Pivot</i> в аналітичній платформі <i>Power BI</i>.</p> <p>Завдання: Задано матрицю із відображенням продажів за роками та країнами. На вкладці «Глобальне завантаження даних» переконайтесь, що параметр «Автоматична дата / час для нових файлів» не увімкнено. Побудуйте візуалізатор Table. Для цього знайдіть поле <i>EnglishProductCategoryName</i> та перетягніть його до категорії рядків добре під поле <i>CalendarYear</i>.</p> <p>Виберіть візуальний елемент <i>Фільтр</i> та додайте поле <i>EnglishProductCategoryName</i>. Отримайте різні результати, застосувавши різні категорії з використанням фільтрів.</p> <p>Дослідіть міру, яка називається <i>Total Sales Amount</i> у таблиці <i>FactInternetSales</i>. Клацніть міру та зауважте, що у рядку формул над таблицею знаходиться код мовою <i>DAX</i>, який використовується для обчислення міри.</p>	2
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Інструментальні засоби аналізу даних», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульного контролю.</p> <p><i>Індивідуальна робота:</i></p> <p>Знайдіть веб-ресурси з довідковою інформацією про імпорт даних інших категорій до аналітичної платформи <i>Microsoft Power BI</i>. Здійсніть практичну реалізацію імпорту з двох із перелічених джерел даних, які не були розглянуті на лабораторних роботах з даної теми: <i>XML</i>, <i>Microsoft Access Databases</i>, <i>JSON</i>, <i>SQL</i>, <i>Oracle Databases</i>, <i>ODBC</i>, <i>Google Analytics</i>, <i>Azure SQL Database</i>.</p> <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 3, 4. <i>Додатковий:</i> 5-7, 9, 11, 12, 14-19. <i>Інтернет-ресурси:</i> 28-30.</p>	20
<p>Знати: алгоритми трансформації даних у <i>Power Query</i>, функції мови <i>DAX</i> для роботи з даними</p> <p>Вміти: створювати схеми даних типу «зірка» та «сніжинка», встановлювати та модифікувати зв'язки в моделі даних, використовувати мову <i>DAX</i> для створення мір та</p>	<p style="text-align: center;">Тема 7. Створення моделі даних</p> <p style="text-align: center;">Лекція 18. Трансформація даних у редакторі Power Query План лекції</p> <ol style="list-style-type: none"> 1. Запити у <i>Power BI</i>. 2. Зміна даних у <i>Power Query</i>. 3. Трансформація, очищення та фільтрування даних у <i>Power BI</i>. 4. Об'єднання даних. Додавання даних. Розщеплення даних. 5. Приведення даних до необхідної форми. 6. Групування та агрегування даних. <p style="text-align: center;">Лекція 19. Створення моделі даних План лекції</p> <ol style="list-style-type: none"> 1. Створення зв'язків таблиці. 2. Схеми зірки та сніжинки у <i>Power BI Desktop</i>. 3. Денормалізація даних у моделі. 4. Модифікація зв'язків моделі. 5. Створення зручної моделі. 	2

<p>обчислювальних стовпців у моделі даних</p>	<p style="text-align: center;">Лекція 20. Основи мови DAX План лекції</p> <ol style="list-style-type: none"> 1. Мова запитів DAX. 2. Оператори DAX. 3. Робота з текстовими функціями. 4. Використання інформаційних та логічних функцій. 5. Отримання даних із суміжних таблиць. Використання математичних, тригонометричних та статистичних функцій у DAX. <p style="text-align: center;">Лекція 21. Створення мір та обчислювальних стовпців у DAX План лекції</p> <ol style="list-style-type: none"> 1. Створення обчислювальних стовпців і мір у Power BI. 2. Зміна контексту запиту. 3. Каскадні розрахунки стовпців. 4. Використання функцій фільтра у створених мірах. 5. Створення напівадитивних мір. <p style="text-align: center;">Лекція 22. Аналіз часових даних у DAX План лекції</p> <ol style="list-style-type: none"> 1. Аналіз часових даних у моделі даних Power BI. 2. Створення та застосування таблиці дат. 3. Оцінки на основі часового періоду. 4. Зміна контексту дати. 5. Використання функцій дати та часу DAX. <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-3. <i>Додатковий:</i> 11, 12, 14. <i>Інтернет-ресурси:</i> 28.</p>	<p style="text-align: center;">2</p> <p style="text-align: center;">2</p> <p style="text-align: center;">2</p>
	<p style="text-align: center;">Лабораторні заняття 35-36 Робота з наборами даних</p> <p>Мета: отримання навиків роботи з редактором запитів <i>Power Query</i>.</p> <p>Завдання: Завантажте файл <i>CarSales.xlsx</i> з обсягами продажу автомобілів. Увійдіть у редактор <i>Power Query</i> для редагування існуючих записів серед завантажених даних. Здійсніть декілька перетворень даних (додавання та видалення стовпців та рядків, перейменування стовпців, фільтрування даних).</p> <p>Проведіть модифікацію даних, залишивши тільки перші 50 записів початкового набору даних. У наборі, що залишився, проведіть очистку даних, видаливши порожні рядки та дублікати записів. Відсортуйте дані спочатку за стовпцем <i>Make</i>, а потім за полем <i>CountryName</i>.</p> <p>Проведіть розширену фільтрацію даних для стовпця <i>SalePrice</i>. Вказавши в одному фільтрі декілька числових значень ціни.</p> <p>Створіть додатковий стовпець <i>Count</i> шляхом застосування операції <i>Count Rows</i> для даних стовпця <i>Make</i>. Виконайте самостійно декілька агрегацій даних з різних стовпців, використовуючи операції</p> <ul style="list-style-type: none"> - <i>Count Distinct Rows</i> - <i>Sum</i> - <i>Average</i> - <i>Median</i> 	<p style="text-align: center;">4</p>

- *Min*
- *Max*
- *All Rows*.

Збережіть зміни у редакторі запитів.

Лабораторні заняття 37-38

Форматування даних у *Power Query*

Мета: набуття навичок роботи з *Power Query* в аналітичній платформі *Power BI*; створення запитів на імпорт даних; фільтрування та перетворення даних; додавання та формування даних; групування та агрегування даних.

Завдання: *Запустити Power Query Editor. На вкладці Home вибрати спадне меню New Source та оберіть опцію Text/CSV. Оберіть файл FlightPerformance.csv. На панелі Query Settings перейменуйте запит у FlightDelays. У список Applied Steps додайте пункт «Promoted headers» і зазначити перший рядок у якості заголовків. Перевірити типи кожної колонки, щоб побачити, чи Query Editor оновив FlightDate до типу даних Date та колонки числового типу (наприклад, «FlightNum») до типу даних Number. На вкладці Home вибрати Choose Columns, очистити усі виділення, а потім виберіть лише колонки Carrier, Origin, OriginCityName та DepDelay. Скористатися стовпцем DepDelay та застосувати фільтр, щоб запит витягував лише рядки із затримкою вильоту рейсу понад 15 хвилин. Вибрати запит FlightDelays у списку Queries та параметр Append Queries у розділі Combine section на вкладці Home. Додайте запит FlightDelays2 до запиту FlightDelays. Перейменувати дві результуючі колонки OriginCity та OriginState. Замінити West Palm Beach / Palm Beach в колонці OriginCity на просто Palm Beach, вибравши Replace values. На панелі Query Settings перейменувати запит на DelaySummary. Залишити колонки Carrier, Origin та DepDelay, а решту – видалити. Знайти групування середньої затримки та максимальної затримки за походженням та перевізником. На вкладці Transform використати округлення, щоб округлити AveDelay до найближчої хвилини.*

Додати запит під назвою Carriers, який отримує код та опис з файлу Carriers.csv. Переконайтеся, що перший рядок був запропонований як заголовок даних.

Вибрати запит DelaySummary та опцію Merge Queries на вкладці Home. Використовуючи запит Carriers, об'єднати стовпець Description, що відповідає стовпцю Carrier у запиті DelaySummary, та стовпець Code у запиті Carriers. Переконайтеся, що таблиці Carriers, DelaySummary та FlightDelays були додані до моделі Power BI.

Лабораторні заняття 39-40

Створення моделі даних

Мета: набуття навичок роботи з *Power Pivot* в аналітичній платформі *Power BI*; створення відношень між таблицями; денормалізація даних; встановлення ієрархії даних.

Завдання: *Завантажити файл з базу даних Adventureworks.accdb. Переключитися до перегляду даних у Power BI Desktop; перейменувати та (або) приховати стовпці, зазначені*

4

4

в таблиці за допомогою команд *Rename or Hide From Report*. Перейти до перегляду моделі (*Model view*), яка повинна містити таблиці з датою, клієнтами та продажами в Інтернеті. Переконайтеся, що таблиця *Customer* та таблиця *Internet Sales* мають взаємозв'язок, визначений між ними.

В моделі даних створити відношення між *DateKey* з таблиці *Date* та *OrderDateKey* в таблиці *Internet Sales*. Аналогічно створити відношення між *DateKey* та *ShipDateKey*. Використати вікно *Edit relationship* і спробувати зробити останнє відношення активним (повинна бути помилка, оскільки може бути лише одне активне відношення між двома таблицями в моделі).

За допомогою дизайнера *Power Query* вибрати стовпці *ProductKey*, *ProductAlternateKey*, *ProductSubCategoryKey*, *WeightUnitMeasureCode*, *SizeUnitMeasureCode*, *EnglishProductName*, *ListPrice*, *Size*, *SizeRange*, *Weight* та *Color* у запиті *DimProduct*. Змінити ім'я запиту на *Product*.

Об'єднати *DimProductSubcategory* із запитом *Product* за допомогою лівого зовнішнього з'єднання. Повторити попередні кроки для об'єднання *EnglishProductCategoryName* із запитом *DimProductCategory* у запит *Product*.

Видалити стовпці *ProductSubcategoryKey* та *ProductCategoryKey* із запиту *Product*.

Завантажити дані про продукт у модель *Power BI*, приховати стовпець *ProductKey* у вікні *Report*. Створити зв'язок між таблицями *Internet Sales* та *Product* за допомогою *ProductKey*.

Створити ієрархію, вибравши стовпець *Category* в таблиці *Product*. Створити ієрархію календаря з назвою *Calendar* у таблиці *Date* використовуючи рік, квартал та місяць.

Лабораторні заняття 41-42

Створення обчислювальних стовпців

Мета: створення обчислювальних стовпців з використанням текстових функцій *DAX*, функцій дати *DAX*; використання у виразі даних із пов'язаних таблиць; реалізація умовної логіки в обчислювальних виразах.

Завдання: Відкрити файл, який містить модель даних, що складається з даних про продажі, даних про товари та даних магазину. Перейти до подання даних і вибрати таблицю *Sales*. Створити новий стовпець (*New column*), ввівши в рядок формули над таблицею введіть наступний вираз для обчислення маржі:

$$\text{Margin} = [\text{SalesAmount}] - [\text{TotalCost}]$$

Повторити цю процедуру, щоб створити стовпець *Margin Percent* з наступною формулою:

$$\text{Margin Percent} = \text{DIVIDE}([\text{Margin}], [\text{SalesAmount}], \text{BLANK}())$$

Для таблиці *Calendar* використати функцію *WeekDay* для створення стовпця *DayOfWeek*.

Використати функцію *Format*, щоб отримати стовпець *Weekday* (назви днів тижня).

Створити стовпець «Year», «Month #» та «Month». Скориставшись стовпцями *Year* і *Month*, створити ієрархію *Calendar* у таблиці *Date*.

4

У таблицю *Products* додати обчислювальний стовпець *Weight Label* за формулою:

= if (ISBLANK ([Weight]), BLANK (), [Weight] & "" & [WeightUnitMeasureID])

Створити стовпець *Product Category* за допомогою функції *RELATED*:

= RELATED('ProductCategory'[ProductCategory])

Скориставшись функцією *RELATED*, створити стовпець *ProductSubcategory*.

Приховати таблиці «*ProductCategory*» та «*ProductSubcategory*».

Перейти до таблиці *Store* і створити стовпець *Years Open* за допомогою формули:

=TRUNC(YEARFRAC([OpenDate],
If(ISBLANK([CloseDate]),TODAY(),[CloseDate])),0)

Створити стовпець *Lifetime Sales* за формулою:

=SUMX(RELATEDTABLE(Sales),Sales[SalesAmount]).

Лабораторні заняття 43-44

Створення мір у DAX

Мета: створення агрегованих мір; зміна контексту даних у мірах; використання неактивних співвідношень у мірах; створення комплексної міри; створення *KPI*.

Завдання: Відкрийте файл зі звітом, який містить модель даних, що складається з даних про продажі, товари та зберігання. Під таблицею *Sales* створити міру *Sum Sales*, обравши команду *New measure*:

Sum Sales = Sum(Sales[SalesAmount])

Створити міри *Max Sales Quantity*, *Min Sales Quantity* та *Ave Sales Quantity* під таблицею *Sales* та застосувати візуальний елемент *Matrix* для перевірки контексту запиту.

Додати міри *Sum Sales* та *Max Sales Quantity* у область *Values*. Під таблицею *Product* додати *ProductCategory* до області *Rows*.

Додати візуальний фрагмент *Slicer* у звіт, який дозволяє фільтрувати поле *Continent* у таблиці *Geography*.

Порівняти співвідношення продажів із загальними продажами для всіх товарів. Відкрити конструктор моделі *Power Pivot* в режимі перегляду даних, обрати таблицю *Sales* і додати наступну міру до таблиці *Sales*:

All Product Sales=CALCULATE([Sum Sales],ALL('Product'))

Міра *All Product Sales* використовує функцію *CALCULATE* для зміни будь-якого фільтра товару, застосованого до контексту запиту. Відформатувати міру як валюту.

Переключитися на матрицю. Замінити міру *Max Sales Quantity* на *All Product Sales*.

Додайте наступну міру до таблиці *Sales* (відформатуйте міру у відсотках):

Product Sales Ratio=Divide([Sum Sales],[All Product Sales],Blank())

На вкладці *Model view* створити взаємозв'язок між таблицею *Date* і таблицею *Store*.

Перейти на вкладку *Data view* і додати наступну міру до таблиці *Store*. Оскільки використовується неактивний зв'язок, то потрібно скористатися функцією *USERELATIONSHIP*:

Store Count=CALCULATE(DISTINCTCOUNT([StoreKey]),

4

USERRELATIONSHIP(Store[OpenDate], 'Date'[Datekey]))
 Визначити найкращий день продажу для категорії товару, створивши міру *Sale Quantity* в таблиці *Sales*:

$$\text{Sale Quantity} = \text{SUM}([\text{SalesQuantity}])$$
 Визначити дату дня з найвищим показником продажу. Для цього спершу створити функцію фільтрації, яка повертає дати, коли кількість продажу (*Sale Quantity*) дорівнює кількості найкращих днів продажу (*Top Sale Day Quantity*) за період:

$$\text{Filter}(\text{VALUES}('Date'[Datekey]),$$

$$[\text{Sale Quantity}] = \text{CALCULATE}([\text{Top Sale Day Quantity}],$$

$$\text{VALUES}('Date'[Datekey])))$$
 Потім цей фільтр вставити у функцію *CALCULATE*, яка повертає найсвіжішу дату. Відформатувати міру як коротку дату (*More Formats* у спадному меню *Format*):

$$\text{Top Sale Day} = \text{CALCULATE}(\text{MAX}('Date'[Datekey]),$$

$$\text{Filter}(\text{VALUES}('Date'[Datekey]), [\text{Sale Quantity}] =$$

$$\text{CALCULATE}([\text{Top Sale Day Quantity}], \text{VALUES}('Date'[Datekey])))$$

Лабораторні заняття 45-46

Операції з функціями часу

Мета: створення таблиці дат; використання функцій операції з часом для аналізу даних; створення матриці зростання «місяць за місяцем»; створення звіту про рівень запасів.

Завдання: Імпортувати файл, що містить дані про запаси та продажі з тестової бази даних *Contoso*.

Обрати вкладку *Model view* і переконатися у відсутності таблиці *Date*. Створити таблицю *Date*, обравши на вкладці *Modeling* опцію «*New Table*». Ввести наступний вираз *DAX*, щоб створити набір дат на основі мінімальних та максимальних дат продажів:

Calendar =

$$\text{CALENDAR}(\text{Min}('Sales'[DateKey]), \text{MAX}(\text{Sales}[DateKey]))$$

Використовуючи *DAX*, додати стовпці *Year*, *Month* та *Month No* у таблицю *Calendar*.

Відсортувати стовпець *Month* за стовпцем *Month No* та відформатувати стовпець *Date*, щоб відобразити дату без часу.

Створити відношення між таблицями *Sales* та *Calendar*

Обрати таблицю *Sales* у вікні *Data view*. Додати наступні міри та відформатувати їх як валюту:

$$\text{Sum of Sales} = \text{SUM}(\text{Sales}[\text{SalesAmount}])$$

$$\text{YTD Sales} = \text{TOTALYTD}([\text{Sum of Sales}], 'Calendar'[Date])$$

$$\text{MTD Sales} = \text{TOTALMTD}([\text{Sum of Sales}], 'Calendar'[Date])$$

Створити тривалу тримісячну міру продажу із наступним виразом:

$$\text{Rolling 3 Month Sales} = \text{CALCULATE}([\text{Sum of Sales}],$$

$$\text{DATESINPERIOD}('Calendar'[Date], \text{LASTDATE}('Calendar'[Date]),$$

$$-3, \text{MONTH}))$$

Порівняти зростання продажів від певного місяця до наступного. Створити показник продажів за попередній місяць у таблиці *Sales*:

$$\text{Prev Month Sales} = \text{CALCULATE}([\text{Sum of Sales}],$$

$$\text{PARALLELPERIOD}('Calendar'[Date], -1, \text{MONTH}))$$

4

	<p>Використати попередні та поточні продажі, щоб створити міру зростання продажів. Відформатувати міру у відсотках.</p> $\text{Monthly Sales Growth} = \text{Divide}$ $([\text{Sum of Sales}] - [\text{Prev Month Sales}], [\text{Prev Month Sales}], \text{BLANK}())$ <p>Створити зв'язок між таблицями Inventory та Calendar. Додати в таблицю Inventory наступну міру підрахунку запасів:</p> $\text{Inventory Count} = \text{SUM}(\text{'Inventory'}[\text{UnitsInStock}])$ <p>Використовуючи непорожній фільтр та функцію CALCULATE, обчислити останню непорожню кількість для продукту. Додати наступну міру до таблиці Inventory:</p> $\text{Product Quantity} = \text{CALCULATE}([\text{Inventory Count}], \text{LASTNONBLANK}(\text{'Calendar'}[\text{Date}], [\text{Inventory Count}]))$ <p>Визначити товарні запаси у всіх вибраних магазинах, використовуючи наступну міру:</p> $\text{Inventory Level} = \text{SUMX}(\text{Values}(\text{Store}[\text{StoreKey}]), [\text{Product Quantity}])$ <p style="text-align: center;">Лабораторне заняття 47 Модульна контрольна робота 3 «Передобробка даних в системі Power BI Desktop»</p> <p>Мета: перевірити рівень знань, отриманих студентами при вивчення теми «Передобробка даних в системі Power BI Desktop».</p> <p>Завдання: Завантажити набір даних у вигляді таблиць згідно із індивідуальним завданням. Налаштувати необхідний формат даних таблиць (ціле число, дійсне число, відсотки, дата, текстове поле, грошові одиниці).</p> <p>Створити дві нові таблиці, в яких здійснити сортування даних за заданим полем за зростанням і спаданням.</p> <p>Створити обчислювальний стовпець та міру з використанням мови DAX згідно із вказаними параметрами.</p> <p>Побудувати модель даних і встановити всі можливі активні зв'язки між таблицями.</p> <p>Провести візуалізацію даних, використовуючи такі інструменти:</p> <ul style="list-style-type: none"> - графік; - стовпчикова діаграма; - кругова діаграма. <p>Побудувати прогноз за вказаними показником та горизонтом прогнозування.</p>	2
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Передобробка даних в системі Power BI Desktop», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульної контрольної роботи.</p> <p><i>Індивідуальна робота:</i></p> <p>Знайти серед веб-ресурсів порталу відкритих даних https://data.gov.ua набір даних за однією з груп: будівництво, економіка та бізнес, освіта і культура, сільське господарство, транспорт, держава, земля, охорона здоров'я, соціальний захист, фінанси, екологія, молодь і спорт, податки, стандарти, юстиція.</p> <p>Завантажити набір даних до платформи Power BI за прямим доступом до відповідної веб-сторінки. Провести трансформацію</p>	40

	<p>даних, залишивши тільки стовпці, які будуть використані для дослідження. Побудувати модель завантажених даних. Додати до існуючих таблиць в моделі принаймні один додатковий обчислювальний стовпець і одну міру. Побудувати звіт з використанням базових елементів візуалізації даних.</p> <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-3. <i>Додатковий:</i> 11, 12, 14. <i>Інтернет-ресурси:</i> 28.</p>	
<p>Знати: основні інструменти візуалізації аналітичних звітів, засоби захисту й публікації звітів, оновлення даних у побудованих звітах.</p> <p>Вміти: створювати та формувати інтерактивні візуальні компоненти <i>Power BI Desktop</i>, публікувати аналітичні звіти на платформі <i>Power BI Service</i>, створювати вбудований код звіту на веб-ресурсах</p>	<p style="text-align: center;">Тема 8. Побудова аналітичних звітів</p> <p style="text-align: center;">Лекція 23. Візуалізація звітів у Power BI Desktop План лекції</p> <ol style="list-style-type: none"> 1. Панелі інструментів <i>Power BI Desktop</i>. 2. Робота з таблицями. 3. Створення матриці. 4. Одномірні та багатомірні картки. 5. Створення різних типів діаграм. <p style="text-align: center;">Лекція 24. Форматування візуальних елементів звітів План лекції</p> <ol style="list-style-type: none"> 1. Створення інтерактивних візуалізацій. Керування інтерактивністю. 2. Форматування діаграм. 3. Розширене форматування. 4. Візуалізація аналітики. <p style="text-align: center;">Лекція 25. Фільтрація даних та використання слайсерів План лекції</p> <ol style="list-style-type: none"> 1. Фільтри візуального рівня. 2. Розширені текстові фільтри. 3. Специфічні фільтри рівня візуалізації. 4. Використання слайсерів. 5. Визначення підходу до інтерактивного відбору даних. <p style="text-align: center;">Лекція 26. Створення власних візуалізацій План лекції</p> <ol style="list-style-type: none"> 1. Розширена інтерактивність та власні візуалізації. 2. Спеціальні візуалізації <i>Power BI</i>. 3. Спеціальні візуалізації від R та Python. 4. Створення візуалізацій на основі карт. 5. Створення власної карти фігури. <p style="text-align: center;">Лекція 27. Публікація та поширення аналітичних звітів План лекції</p> <ol style="list-style-type: none"> 1. Робоча область користувача. 2. Додавання плиток на інформаційну панель. 3. Спільне використання звітів. 4. Публікація та обмін звітами. 5. Налаштування оновлення даних в опублікованих звітах. <p style="text-align: center;">Лекція 28. Power BI Service. Використання служби PowerBI.com</p>	<p style="text-align: center;">2</p> <p style="text-align: center;">2</p> <p style="text-align: center;">2</p> <p style="text-align: center;">2</p> <p style="text-align: center;">2</p> <p style="text-align: center;">2</p>

	<p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Технічні вимоги до служби Power BI Service. 2. Безпека та дозволи робочої області. 3. Встановлення та використання шлюзів даних. 4. Створення звітів у веб-службі Power BI.com <p style="text-align: center;">Лекція 29. Створення комплексних рішень у Power BI</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Використання спеціальних візуальних засобів. 2. Впровадження геопросторового аналізу. 3. Впровадження безпеки на основі рядків. 4. Створення шаблонів та пакетів Power BI Desktop. 5. Приклади побудови комплексних рішень аналітики даних. <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 3. <i>Додатковий:</i> 11, 12, 14. <i>Інтернет-ресурси:</i> 28.</p>	2
	<p style="text-align: center;">Лабораторні заняття 48-50</p> <p style="text-align: center;">Створення базових візуалізацій</p> <p>Мета: створення стандартних візуальних зображень; створення візуалізацій на основі карти; вивчення візуальних взаємодій.</p> <p>Завдання: <i>Завантажити до аналітичної платформи дані колекції міста Сан-Франциско. Використовуючи інструмент Table visualization, створити таблицю з полями Type of request, Ave Days Open CY та Ave Days Open PY. На панелі Filters відфільтрувати таблицю так, щоб відображалися лише ті випадки, коли Status закритий. Відформатувати таблицю рядками, що чергуються (смужками) і змінити кегль шрифту на 10. Створити 100-відсоткову гістограму з панелі інструментів Visualizations, додавши поле Number of Cases до області Values, Type – до області Axis, а Year – до області Legend. Змінити кольори даних на золотий та оранжевий та увімкнути мітки даних. Змініть мітки даних на чорні. Створити візуальний елемент Treemap у панелі інструментів. Додати Number of Cases до Values та Neighborhood до Group, Ave Days Open до Tooltips. На новій сторінці додати лінійну діаграму, що показує дату закінчення тижня (Week Ending Date) на осі та кількість випадків (Number of Cases) як значення. Провести візуалізацію з використанням елемента Slicer із панелі інструментів Visualizations. Додати Week Ending Date в область Fields. Використати Slicer для регулювання діапазону дат на графіку.</i></p> <p style="text-align: center;">Лабораторні заняття 51-52</p> <p style="text-align: center;">Створення та форматування інтерактивних візуалізацій</p> <p>Мета: отримання практичних навиків створення та форматування візуалізаторів Матриця, Картка, Багаторядкова картка, різні типи Діаграм.</p> <p>Завдання: <i>Завантажити файл із даними CarSalesDataForReports.pbix. Створити матрицю, використовуючи поля CountryName (з таблиці Країни), Make (із</i></p>	4

таблиці Виробництво) та Колір (з таблиці Кольори). Додати поля SalePrice і Cost Plus (з таблиці InvoiceLines) та Пробіг (з таблиці Виробництво). Додати проміжні підсумки до матриці. Створити ієрархію стовпців матриці, додавши рівень VehicleType. За допомогою пункту See Data відобразити у новому вікні неформатовану версію даних для матриці. Відсортувати дані матриці за величиною загальної валової маржі у порядку спадання значень.

Розгорнути поля таблиці InvoiceLines у списку Поля та вибрати поле SalePrice поле. Відобразити на картці загальну суму всіх продажів. Змінити розмір картки так, щоб фігурка вписувалася в рамки картки. Змінити одиниці, кількість десяткових знаків, колір та розмір тексту даних, що відображається на картці. Видалити мітку, яка відображається автоматично, а також змінити її розмір тексту. Додати заголовок, який ви можете формувати самостійно. Змінити колір фону картки.

Створити багаторядкову карту з полями: назва країни, ціна автомобіля, потужність, наявність запчастин. Змінити розмір багаторядкової картки для декількох країн, щоб відобразити всі фігури в одному рядку для кожної країни.

Провести візуалізацію даних звіту CarSalesDataForReports.pbix, самостійно підібравши найбільш вдалі показники для відображення за допомогою кожного з таких типів діаграм:

- секторна діаграма;
- гістограма;
- стовпчикова діаграма;
- лінійна діаграма;
- діаграма з областями;
- точкова діаграма;
- бульбашкова діаграма;
- діаграма послідовності;
- водоспадна діаграма;
- кільцева діаграма;
- стрічкова діаграма.

Лабораторне заняття 53

Створення власних візуальних елементів

Мета: розширення можливостей Microsoft Power BI за допомогою створення користувацьких візуальних зображень.

Завдання: Відкрити додаткові візуальні елементи, що присутні у галереї візуальних елементів за адресою <http://app.powerbi.com/visuals>. Імпортувати обраний візуальний елемент з використанням команди Get It Now.

Ознайомитися з веб-ресурсом OKViz (<http://okviz.com>). Інтегрувати до власного звіту візуальну синоптичну панель, яка дозволяє створювати області на зображеннях, призначати кольори і відображати інформацію, керовану даними.

Відкрити галерею візуальних програм Power BI - веб-сайт спільноти (<http://visuals.powerbi.com>), який дозволяє розробникам завантажувати нові Power BI-образи, а користувачам відкривати та завантажувати ці візуальні зображення. Ознайомитися зі спеціальними засобами для розробників Power BI – інструментами спеціальної програми для візуальних розробників

2

(<https://github.com/Microsoft/PowerBI-Visuals/>), які інтегруються з Power BI, щоб допомогти розробникам у налагодженні та тестуванні візуального коду.

Проаналізувати специфікацію дизайну, що міститься у стандарті IVisual інтерфейсу на сайті <https://github.com/Microsoft/PowerBI-visuals/wiki>.

Використати базові конструкції мови TypeScript – програми вільного та відкритого джерела (<http://www.typescriptlang.org>), що розроблена та підтримувана Microsoft для кодування програм на стороні клієнта та сервера (Node.js).

Лабораторні заняття 54-55

Фільтрація даних у звіті

Мета: навчитися використовувати різні типи фільтрів для обмеження даних, що використовуються у звіті, сторінці чи візуальному зображенні.

Завдання: Завантажити файл із даними CarSalesDataForReports.pbix. Здійснити фільтрацію даних за трьома рівнями фільтрів:

- фільтри рівня звітів;
- фільтри рівня сторінки;
- фільтри рівня візуалізації.

Створити кластерну діаграму за допомогою полів Make (з таблиці Stock) та SalePrice (з таблиці InvoiceLines).

Створити візуальний фільтр, встановивши прапорець зліва від кожного з наступних елементів:

- a) Jaguar;
- b) Rolls-Royce;
- c) Triumph.

Провести фільтрацію різних типів даних в таблиці: числових даних, дати і часу, логічних даних.

Використати розширений текстовий фільтр для поля Make.

Створити кластерну діаграму SalePrice від ClientName. Встановити конкретні фільтра для різних рівнів візуалізації і обрати для кожного рівня власний колір.

Створити декілька фільтрів рівня сторінки. Застосувати фільтрацію на рівні звіту, який ви створили з SalePrice від ClientName. Відфільтрувати за Color на візуальному рівні та FullYear на рівні сторінки. Розгорнути таблицю Countries. перетягнути поле CountryName в список полів рівня звітів. Відобразяться розширені параметри фільтрації для цього поля.

Переконатися, що для цього фільтра вибраний тип базового фільтрування. Вибрати United Kingdom зі списку доступних країн. Усі візуалізації в поточному звіті (на кожній сторінці) будуть відфільтровані для відображення лише даних за цей рік.

Лабораторне заняття 56

Елементи керування аналітичним звітом

Мета: покращення інформаційних панелей шляхом додавання елементів керування звітом.

4

2

Завдання: Завантажити один зі звітів, створених на минулих лабораторних роботах. Налаштувати панель перегляду, використовуючи параметри *Phone Layout, Desktop Layout, Page View Options, Show/Hide Gridlines, Snap Objects to Grid*. Змінити спосіб реагування візуальних елементів звіту з використанням параметру *Edit Interactions*. Застосувати різне перенесення та вирівнювання об'єктів звіту за допомогою параметрів *Bring Forward, Send Backward, Align Objects, Distribute Objects*. Випробувати різні формати сторінок (16:9, 4:3, Кортана, Власні розміри). Обрати окремі візуальні зображення і застосувати до них різні параметри вирівнювання: *Align Left, Align Right, Align Center, Align Top, Align Middle, Align Bottom*. Використати різні параметри розподілення візуальних зображень: *Distribute Horizontally, Distribute Vertically*. Додати до звіту:

- текстове поле із слоганом компанії;
- зображення із логотипом компанії;
- гіперпосилання на сайт компанії.

Змінити колір фону сторінки або використати фонове зображення. Налаштувати навігацію сторінками звіту, використавши кнопки та дії над ними з указанням конкретних сторінок переходу.

Лабораторне заняття 57 **Публікація аналітичних звітів**

2

Мета: публікація звіту на порталі *Power BI Service*, створення інформаційних панелей на порталі та встановлення графіка оновлення даних.

Завдання: Завантажити файл, що містить дані про продажі від торгових компаній *Nordwind*. Увійти до служби *Power BI Service*. Обрати параметр *Publish* і вибрати в якості пункту призначення *My Workspace*. У вікні «*Pin to dashboard*» вибрати нову інформаційну панель і назвати її «*Northwind Sales*». Додати кількість замовлень за категоріями/товарами на інформаційну панель «*Northwind Sales*». Розгорнути панель навігації та вибрати інформаційну панель «*Northwind Sales*». Задати декілька питань в категорії «*Ask a question*». Закріпити цю таблицю на приладовій панелі. Вибрати «*Edit details*». Змінити заголовок і поставити відмітку в «*Display last refresh time*». Додати плитку з текстовим полем. Перебудувати і змінити розміри плиток. Налаштувати графік оновлення з використанням параметру «*SCHEDULE REFRESH*» в меню. У вікні, що з'явилося, розгорнути вузол облікових даних джерела даних та натиснути посилання «*Edit credentials*». Переконавшись, що для методу аутентифікації встановлено значення «*Anonymous*», а рівень конфіденційності встановлено на «*Public*» перед входом. Створіть графік оновлення, який працює щодня о 7:00 год.

Лабораторне заняття 58

Модульна контрольна робота 4

«Візуалізація аналітичних звітів»

2

Мета: перевірити рівень знань, отриманих студентами при вивчення теми «Візуалізація аналітичних звітів».

Завдання: Ви працюєте в компанії з обладнання для велосипедів, і вас просили проаналізувати дані про продажі. Використовуючи Power BI Desktop, вам потрібно створити аналітичний звіт, який дозволяє менеджеру з продажів відстежувати ефективність роботи торгової команди.

Імпортуйте до аналітичної платформи дані з бази даних SalesRepAnalysis.accdb. Оберіть таблиці Employee, SalesTerritory та ResellerSales.

Фільтруйте таблицю «Співробітник», щоб включати лише продавців. Додайте стовпець імені, який поєднує стовпці Ім'я та Прізвище. У таблиці продажів використовуйте OrderDateKey для створення стовпця OrderDate. Змінити тип даних OrderDateKey з цілого числа на текст. Згрупуйте SalesAmount по EmployeeKey, SalesTerritoryKey, Order Year і Quarter. Створіть стовпець TimePeriod, який об'єднує стовпці Рік та Квартал. Для створення відповідних зв'язків між таблицями створіть Таблицю періодів, що містить календарний рік та квартал

Додайте наступну міру до таблиці продажів:

Загальний обсяг продажів = сума (продажі [сума продажу])

Додайте наступні міри до таблиці квот:

Загальна квота = сума (квота [квота])

Варіант = Продажі [Загальний обсяг продажів] -Кота [Загальна квота]

Відсоткове відхилення = DIVIDE (квота [варіація], квота [загальна квота], BLANK ())

Створіть візуалізації для аналізу продуктивності продавця на сторінці дизайнера звітів, додавши такі наочні зображення:

- зріз, використовуючи рік із таблиці Період;
- складена діаграма стовпців, що показує загальний обсяг продажів за країною та кварталом;
- картка з декількома рядами із загальним обсягом продажів та загальною квотою;
- картка з декількома рядами, що відображає варіацію та відсоткове відхилення;
- кластеризована стовпчаста діаграма, що показує загальний обсяг продажів та загальний обсяг продажів за Продавця;
- текстове поле для заголовка.

Додайте до звіту ще одну сторінку та додайте до неї такі візуальні зображення:

- зріз поля «Рік» з таблиці Період;
- зріз поля Salesperson з таблиці Rep;
- дві таблиці з продавцем, загальним обсягом продажів, загальною квотою та Відсотки поля варіації;
- лінійна діаграма, що показує загальний обсяг продажів та загальну квоту за TimePeriod.

Фільтруйте таблиці так, щоб одна відображала п'ятірку найкращих продавців за відсотковою відмінністю, а інша

	<p>показувала п'ять найгірших продавців за відсотковою відмінністю.</p> <p>Відредагуйте візуальні взаємодії, так що частина року впливала на таблиці, але не на лінійну діаграму. Зріз продавця повинен впливати на лінійну діаграму, але не на таблиці.</p>	
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Візуалізація аналітичних звітів», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульної контрольної роботи.</p> <p><i>Індивідуальна робота:</i></p> <p>Обрати на веб-ресурсі порталу відкритих даних https://tradingeconomics.com набір економічних показників України: https://tradingeconomics.com/ukraine/indicators.</p> <p>Завантажити набір даних до платформи <i>Power BI</i> за прямим доступом до відповідної веб-сторінки. Побудувати звіт з використанням різних типів стовпчикових та лінійних діаграм, секторних, кільцевих та бульбашкових діаграм, карток та багаторядкових карток, карт. Здійснити візуальну фільтрацію даних за трьома обраними показниками. Опублікувати аналітичний звіт на платформі <i>Power BI Service</i>.</p> <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 3.</p> <p><i>Додатковий:</i> 11, 12, 14.</p> <p><i>Інтернет-ресурси:</i> 28.</p>	40
Разом		360

5. СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

Основний

1. Cuesta H., Kumar S. Practical Data Analysis. Birmingham : Packt Publishing Ltd, 2016. 316 p.
2. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data /EMC Education Services. Indianapolis : John Wiley & Sons, Inc, 2015. 432 p.
3. Microsoft Power BI Cookbook: Creating Business Intelligence Solutions of Analytical Data Models, Reports, and Dashboards. Birmingham : Packt Publishing Ltd, 2017. 802 p.
4. Гладун А.Я. Data mining: пошук знань в даних / А. Я. Гладун, Ю. В. Рогущина. –Київ: АДЕФ-Україна, 2016. – 451 с.

Додатковий

5. *Бізнес-аналітика: Конспект лекцій. Для студентів усіх форм навчання спеціальностей 073 «Менеджмент» та 8.03060101 «Менеджмент організацій та адміністрування» / Н. М. Шведа. – Тернопіль, ТНТУ, 2016. – 70 с. **
6. Матвійчук А. В. Штучний інтелект в економіці: нейронні мережі, нечітка логіка: монографія. Київ : КНЕУ, 2011. 439 с.
7. Олійник А. О., Субботін С. О., Олійник О. О. Інтелектуальний аналіз даних: навч. посібн. Запоріжжя : ЗНТУ, 2012. 278 с.
8. Adamson C. Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance. Wiley Publishing Inc., 2006. 318 p.
9. Albright S. C., Winston W., Zappe C. Data Analysis and Decision Making. Boston : Cengage Learning, 2016. 948 p.
10. Cao L., Yu P. S., Zhang C., Zhang H. Data Mining for Business Applications. Springer Science; Business Media, 2008. 402 p.
11. Coodley M. O. Introduction to Microsoft Power BI: bring your data to life! CreateSpace Independent Publishing Platform, 2016. 128 p.
12. Etaati L. Advance Analytics with Power BI and R. Auckland : Radacad Systems Limited, 2017. 179 p.
13. Fabrice G., Hamilton N. J. Quality Measures in Data Mining. Berlin; Heidelberg: Springer-Verlag, 2007. 361 p.
14. Ferrari A., Russo M. Introducing Microsoft Power BI. Redmond : Microsoft Press, 2016. 407 p.
15. Han J., Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2006. 800 p.
16. Linoff G. S. Data Analysis Using SQL and Excel. Indianapolis: Wiley, 2015. 792 p.
17. Linoff G. S., Berry M. J. A. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Indianapolis: Wiley, 2011. 888 p.
18. Milton M. Head First Data Analysis: A learner's guide to big numbers, statistics, and good decisions. Sebastopol: O'Reilly Media, 2009. 435 p.

19. *Roskladka A., Ivanova O., Kulazhenko V. Data Scientist: a glance into the future // Зовнішня торгівля: економіка, фінанси, право. 2019. № 3. С. 109-120*
20. Upton G. *Categorical data analysis by example. New Jersey: John Wiley & Sons Inc, 2017. 198 p.*
21. Witten I. H., Eibe F., Hall M. A. *Data mining: practical machine learning tools and techniques. Morgan Kaufmann Publishers, 2011. 630 p.*
22. Лысенко Ю.Г., Н. Н. Иванов Н. Н., Минц А. Ю. *Нейронные сети и генетические алгоритмы : учебн. пособ. Донецк : ООО «Юго-Восток, Лтд», 2003. 265 с.*
23. Олійник А. О., Субботін С. О., Олійник О. О. *Ітеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і нейромережних моделей : монографія. Запоріжжя : ЗНТУ, 2009. 375 с.*
24. Рідкокаша А.А., Голдер К. К. *Основи систем штучного інтелекту : навч. посібн. Черкаси : ВІДЛУННЯ-ПЛЮС, 2002. 240 с.*
25. Руденко О. Г., Бодяньський Є. В. *Штучні нейронні мережі. Харків : Компанія СМІТ, 2006. 404 с.*
26. Ситник В. Ф., Краснюк М. Т. *Інтелектуальний аналіз даних (дейтамайнінг) : навч. посібн. Київ : КНЕУ, 2007. 376 с.*
27. Субботін С. О. *Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень : навч. посібн. Запоріжжя : ЗНТУ, 2008. 341 с.*

Інтернет-ресурси

28. Microsoft Power BI Guided Learning URL: <https://docs.microsoft.com/uk-ua/power-bi/guided-learning>.
29. Офіційний сайт платформи Loginom. URL: <https://loginom.com/>
30. RapidMiner Academy: learn data science and RapidMiner from leading industry experts. URL: <https://academy.rapidminer.com/>.

* Курсивом виділені джерела, що є у бібліотеці КНТЕУ