**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**ТАВРІЙСЬКИЙ ДЕРЖАВНИЙ АГРОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ ДМИТРА МОТОРНОГО**

**Факультет економіки та бізнесу**

**Кафедра фінансів, обліку і оподаткування**

**АНАЛІЗ ВЕЛИКИХ ДАНИХ ТА БІЗНЕС-АНАЛІТИКА**

Методичні рекомендації

з самостійного вивчення дисципліни

«**Аналіз великих даних та бізнес-аналітика**»

для для здобувачів ступеня вищої освіти «Магістр»

зі спеціальності 071 «Облік і оподаткування»

. **Мелітополь, 2022**

УДК 336.7(075)

«Аналіз великих даних та бізнес-аналітика» Методичні рекомендації з самостійного вивчення дисципліни для здобувачів ступеня вищої освіти «Магістр» зі спеціальності 071 «Облік і оподаткування». – Мелітополь, ТДАТУ, 2022 р. – 36 с.

**Розробники**:

**Кюрчев В.М.,** д.т.н., професор кафедри експлуатації та технічного сервісу машин, ТДАТУ

**Гончар В.В. ,** д.е.н., професор, завідувач кафедри маркетингу та бізнес адміністрування ПДТУ

**Сокіл О.Г.,** д.е.н., професор кафедри фінансів, обліку та оподаткування ТДАТУ

**Рецензенти**:

**Тропманн-Фрік Марина** , д-р екон. наук, професор, Університет прикладних наук Гамбургу (Німеччина)

**Хаджинова О. В.**, д-р екон. наук, професор, директор ННІЕМ ПДТУ

Розглянуто та рекомендовано до друку на засіданні кафедри фінансів, обліку і оподаткування

Протокол № *11* від « *07*» червня 2022 року

Завідувач кафедри фінансів, обліку і оподаткування

д.е.н, професор \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Олег СОКІЛ

Схвалено методичною комісією факультету економіки та бізнесу для здобувачів освітнього рівня «Магістр» за спеціальності 075 «Маркетинг»

Протокол № *7* від «*20*»  *червня* 2022 року

Голова, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_ Анна КОСТЯКОВА

Схвалено Навчально-методичною радою Таврійського державного агротехнологічного університету імені Дмитра Моторного для здобувачів освітнього рівня «Магістр» за спеціальності 075 «Маркетинг»

Протокол № *1* від «*26*»  *серпня* 2022 року

Голова, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_ Олександр ЛОМЕЙКО

© Гончар В.В., 2022 рік

**ЗМІСТ**

ВСТУП 4 1 МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДО САМОСТІЙНОЇ ПІДГОТОВКИ 6 2 МЕТОДИЧНІ ВКАЗІВКИ ДО САМОСТІЙНОГО ВИВЧЕННЯ ТЕОРЕТИЧНОГО МАТЕРІАЛУ 8 3 ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ ЗА ТЕМАМИ КУРСУ «МАРКЕТИНГ ІННОВАЦІЙ» 13 4 ПИТАННЯ ДО ІСПИТУ З ДИСЦИПЛІНИ «МАРКЕТИНГ ІННОВАЦІЙ» 30 СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ 32

**ВСТУП**

Методичні рекомендації з вивчення курсу «Аналіз великих даних та бізнес-аналітика» розроблено для організації самостійної роботи студентів з метою підвищення якості освіти, підготовки кадрів та визначення порядку організації, контролю та оцінки самостійних робіт студентів.

Навчальна дисципліна призначена для підготовки нового покоління професіоналів в області менеджменту/бізнес-аналітики, які володіють методами обробки великих даних, сучасними аналітичними інструментами, методами та моделями прийняття управлінських рішень, які інформаційно та інноваційно зорієнтовані на створення нових цінностей для клієнтів.

За допомогою найновіших цифрових технологій (хмарні обчислення, мобільні технології, соціальні технології, машинне навчання, інтернет речей) менеджери/бізнес-аналітики здатні ідентифікувати, збирати, аналізувати великі масиви даних, інтерпретувати та трансформувати їх для глибшого розуміння бізнесу в реальному часі з метою більш швидкого прийняття кращих рішень, підвищення ефективності роботи компанії, оптимізації її конкурентоспроможності, пом’якшення ризиків.

Для опанування матеріалу навчальної дисципліни «Аналіз великих даних та бізнес-аналітика» окрім лекційних, практичних занять, значну увагу необхідно приділяти самостійній роботі.

Основні види самостійної роботи студента:

1. Вивчення додаткової літератури

2. Робота з законодавчими, нормативними та інструктивними матеріалами.

3. Підготовка до практичних занять

4. Підготовка до проміжного і модульного контролю

Предметом вивчення навчальної дисципліни є інструменти та аналітичні методи для використання даних для прийняття рішень, збору та впорядкування даних у масштабі та отримають розуміння того, як аналіз даних може допомогти інформувати зміни в організаціях.

Метою викладання навчальної дисципліни “ Аналіз великих даних та бізнес-аналітика” є підготувати фахівців зі знаннями у галузі великих даних; надання фахівцям навичок у галузі діяльності з удосконалення організації праці, виробництва та управління даними; вивчити принципи, методи та форми організації управління великими даними.

Згідно з вимогами освітньо-професійної програми студенти повинні:

знати:

- різні типи структур даних, формати файлів, джерела даних та мови, які спеціалісти використовують у своїх повсякденних завданнях; - різні типи сховищ даних, такі як Бази даних, Сховища даних, Мапи даних, Озера даних та Конвеєри даних;

- процес вилучення, перетворення та завантаження (ETL), який використовується для вилучення, перетворення та завантаження даних у сховища даних;

- обробка великих даних;

- різні компоненти сучасної екосистеми даних та ролі, яку в цій екосистемі відіграють аналітики даних, науковці даних та бізнес-аналітики; - навички, необхідні для того, щоб бути аналітиком даних, і як виглядає типовий проект аналізу великих даних.

вміти:

- розробляти алгоритми великих даних;

- розуміти основні концепції програмування, включаючи абстракцію даних, їх зберігання та структури;

- розуміти обчислювальне мислення, яке включає декомпозицію, розпізнавання шаблонів та абстрагування, керовані даними проблема та дизайн алгоритму для великих даних Інтерпретації подання та аналізу даних;

- розуміти ключові математичні поняття, включаючи зменшення розмірів та парадокси моделей;

- використовувати аналітичні інструменти, такі як Java, про які студенти дізнаються;

- аналізувати різні типи аналізу даних та ключові етапи процесу аналізу даних;

- розуміти різні компоненти сучасної екосистеми даних та роль аналітиків даних, науковців даних та бізнес-аналітиків у цій екосистемі. Засвоєння ґрунтовних знань і набуття практичних вмінь, відповідних основним професійним компетенціям, які необхідно сформувати у майбутніх фахівців з маркетингу та бізнес-адміністрування в процесі вивчення дисципліни «Аналіз великих даних та бізнес-аналітика», є можливими тільки за умови поєднання теорії та практики в навчальному процесі.

**1 МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДО САМОСТІЙНОЇ ПІДГОТОВКИ**

Інформаційною базою для виконання самостійних робіт стають підручники та навчальні посібники, методичні рекомендації, збірки даних, науково-популярні видання, відповідні ресурси мережі Інтернет, раніш виконані роботи з даної теми та інше.

При вивченні курсу «Аналіз великих даних та бізнес-аналітика» студенти повинні виконати усі вимоги плану та програми, самостійно вивчити частину матеріалу, виконати по завданню викладача семестрове домашнє завдання, використовуючи рекомендовану літературу та методичні рекомендації до самостійної та поза аудиторної роботи студентів за даною дисципліною.

Самостійна робота студента є невід'ємною частиною учбового плану означеної для засвоєння дисципліни у робочому учбовому плані, яка забезпечується методичними та інформаційними ресурсами.

На самостійну роботові студента денної форми навчання відводиться 118 годин, яка враховує підготовку до практичних (семінарських) занять та виконання самостійного домашнього завдання. Щодо заочної форми навчання на самостійну роботу припадає 170 годин відповідно, які включають в собі самостійну підготовку теоретичної частини курсу базуючись на вивчені опорного конспекту лекцій. Для закріплення вивченого необхідно виконати самостійне домашнє завдання.

Оцінювання самостійної роботи студентів виконується на основі поточного контролю виконання індивідуальних завдань та підсумкового контролю (іспит). Підготовка студентів виконується на основі рекомендацій із вивчення теоретичного матеріалу (наведеного переліку контрольних питань з кожного модуля, тем для виконання контрольних робіт); рекомендацій до підготовки до практичних занять; екзаменаційних питань.

Уміння студентів самостійно працювати над вивченням конкретного предмета – важливий чинник підвищення якості підготовки спеціалістів. Самостійна робота виконується відповідно до рекомендованої літератури щодо самостійної роботи з курсу з метою надання студенту можливості отримати визначений обсяг знань, вивчаючи спеціальну технічну літературу, поглибити знання у напрямку вирішення науково -економічних проблем. А також для підготовки до поточного контролю, модульних контрольних робіт та написання та захисту рефератів.

Перелік тем для самостійного вивчення

| №  з/п | Підготовка до практичних занять та виконання самостійного домашнього завдання за темами: | Кількість  годин |
| --- | --- | --- |
| 1 | ІТ -фактор впливу на сучасні управлінські завдання | 13 |
| 2 | Моделювання та аналіз даних | 15 |
| 3 | Архітектура та методи сховища даних | 15 |

| 4 | Видобування, перетворення та завантаження (ETL) | 15 |
| --- | --- | --- |
| 5 | Надання інформації (звітність, інформаційні панелі) | 15 |
| 6 | Аналітичний життєвий цикл і методи:  кластеризація, класифікація, машинне навчання | 15 |
| 7 | Моделювання великих даних | 15 |
| 8 | Архітектура та впровадження | 13 |
|  | Разом | 118 |

Розрахунок часу для самостійної роботи студента за видами

| №  з/п | Вид роботи | Кількість  годин |
| --- | --- | --- |
| 1 | Опрацювання програмного матеріалу, що викладається на лекціях | 20 |
| 2 | Підготовка до лабораторних робіт | - |
| 3 | Підготовка до практичних (семінарських) занять | 20 |
| 4 | Виконання індивідуальних завдань (реферати, творчі, розрахунково -графічні роботи, презентації тощо) | 20 |
| 5 | Підготовка до заходів контролю (модульні контрольні роботи) | 22 |
| 6 | Дизайн курсу | - |
| 7 | Підготовка самостійного домашнього завдання | 36 |
|  | Разом | 118 |

Самостійна робота виконується відповідно до методичних рекомендацій щодо самостійної роботи студента.

Для виконання самостійного домашнього завдання необхідно спочатку ознайомитись з коротким викладенням курсу «Аналіз великих даних та бізнес аналітика», що наведено на стор. 9 – 13, та звернутись до конспекту лекцій з дисципліни «Аналіз великих даних та бізнес-аналітика» (для студентів спеціальності 075 «Маркетинг» , 075 «Менеджмент» всіх форм навчання.) На сторінках 13 – 42 викладено перелік завдань для самостійного домашнього завдання. При виконані завдань для самостійної роботи необхідно спочатку вивчити теоретичні питання користуючись конспектом лекцій, підручниками та навчальними посібниками з курсу.

Якщо для студента при вивченні даного курсу виникають труднощі при використанні літератури, потрібно звернути увагу на список рекомендованої літератури. Слід не обмежуватися тільки однією навчальною літературою. Глибоке знання даної дисципліни передбачає вивчення першоджерел, монографій і періодичних видань. Періодичні видання зосереджують свою

This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in.

Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі.



увагу в основному на аналізі сучасних тенденцій розвитку економіки. Для більш чіткого усвідомлення змісту економічних категорій, які використовуються в аналізі великих даних, радимо скористатися економічними словниками та економічними енциклопедіями, а також збірниками наукових праць які розміщено на сайті Інституційного репозитарію ПДТУ (Electronic Institutional Repository Pryazovskyi State Technical University) http://eir.pstu.edu.

Для самостійної підготовки студенту потрібно вивчити теоретичний матеріал, що викладено у опорному конспекті лекцій який надається викладачем. Або керуючись робочою програмою та методичними вказівками до практичних занять і рекомендованою літературою самостійно оволодіти знаннями з дисципліни «Аналіз великих даних та бізнес-аналітика».

Для студентів денної форми навчання самостійне домашнє завдання враховується в підсумку балів що отримує студент наприкінці семестру, і є обов’язковим до виконання.

Для студентів заочної або прискореної форми навчання виконання самостійного домашнього завдання є допуском до здачі іспиту.

**2 МЕТОДИЧНІ ВКАЗІВКИ ДО САМОСТІЙНОГО ВИВЧЕННЯ ТЕОРЕТИЧНОГО МАТЕРІАЛУ**

Студент повинен знати: теоретичну основу дисципліни; давати визначення понять по нижченаведеним темам курсу; знати відповіді на контрольні питання по кожному модулю; виконати модульні контрольні роботи по наведеним темам.

**Змістовий модуль 1. Сучасні аспекти цифрової трансформації бізнесу**

**Тема 1. ІТ-фактор впливу в сучасних управлінських завданнях** Самостійну працю над темою належить розпочати з визначення змісту курсу «Аналіз великих даних та бізнес-аналітика» як самостійної дисципліни, а насамперед – з визначення самого поняття «аналіз даних». Аналіз даних - це дослідження, пов'язані з обрахуванням багатовимірної системи даних, що має безліч параметрів.

Використання ІТ визначається як міра, в якій організація використовує ІТ для підтримки оперативних та стратегічних завдань. Клімат ІТ менеджменту, управлінські ІТ-знання, ефективність процесу ІТ-управління та використання ІТ. Модель структурного рівняння розроблена для оцінки відносних ефектів та взаємозв’язків між цими конструкціями. Управлінські ІТ- знання є домінуючим фактором у поясненні високого рівня використання ІТ, а також те, що як управлінські ІТ-знання, так і ефективність процесу ІТ управління впливають на ІТ-клімат.

У зв'язку з цим студентам рекомендується приділити достатню увагу вивченню основних положень кожної теорії.

Закінчуючи вивчення теми, потрібно розглянути і ефективність процесу ІТ-управління.

**Тема 2. Моделювання та аналіз даних**

Ця тема має винятково важливе значення для засвоєння дисципліни. В ній доводиться, що моделювання даних допомагає зрозуміти інформаційні вимоги.

Модель даних - це опис того, як дані повинні використовуватися для задоволення вимог, заданих кінцевим користувачем. Моделювання даних відрізняється залежно від типу бізнесу, оскільки бізнес-процеси або кожен сектор є різним, і його потрібно визначити на етапі моделювання. Початковим кроком є аналіз ситуації, збір даних. Процес моделювання даних починається зі збору вимог. При розробці належної моделі даних важливо спілкуватися із зацікавленими сторонами щодо вимог. Моделювання даних - це акт дослідження структур, орієнтованих на дані. Це можна використовувати для різних цілей. Однією з важливих функцій моделювання даних є те, що вона допомагає зрозуміти інформаційні вимоги. Особливо це полегшує життя як розробникам, так і кінцевим користувачам. Як зазначалося вище, моделювання даних допомагає кінцевим користувачам визначити свої вимоги, а розробники можуть розробити систему, яка відповідає зазначеним вимогам.

Студентам при самостійному вивченні теми слід розглянути Аналітичний і інформаційний підходи до моделювання. Моделювання - універсальний метод отримання, опису та використання знань. Застосовується в будь-якій професійній діяльності. Аналітичний підхід до моделювання базується на тому, що дослідник при вивченні системи відштовхується від моделі. При інформаційному підході відправною точкою є дані, що характеризують досліджуваний об'єкт, і модель «підлаштовується» під дійсність.

Завершувати самостійне вивчення теми рекомендується розглядом питання етапи аналізу даних. Аналіз даних - це всього лише послідовність кроків, кожен з яких відіграє ключову роль для наступних.

**Тема 3. Архітектура та методи зберігання даних**

Самостійну працю над цією темою належить розпочати з вивчення поняття – архітектура зберігання даних. Архітектура бази даних - комплекс структурних компонентів БД, а також засобів, що забезпечують їх взаємодію як один з

одним, так і з кінцевим користувачем, системним персоналом. Студенти повинні засвоїти, що дзайн сховища даних використовує певну тематику. Він надає інформацію про тему, а не про діяльність підприємства. Ці теми можуть бути пов’язані з продажами, рекламою, маркетингом тощо.

Замість того, щоб зосередитись на ділових операціях або операціях, зберігання даних робиться акцент на бізнес-аналітиці (BI), тобто; відображення та аналіз даних для прийняття рішень. Він також пропонує пряму та стислу інтерпретацію певної теми, виключаючи дані, які можуть бути не корисними для тих, хто приймає рішення.

Велику увагу належить приділити вивченню зарубіжного досвіду сучасного бізнес-аналізу.

**Тема 4. Видобування, перетворення та навантаження (ETL)** Тему присвячено розгляду Extract, Transform, Load (ETL) або Витяг, Перетворення та Завантаження. ETL - один з основних процесів в управлінні сховищами даних, який включає в себе: вилучення даних із зовнішніх джерел; їх трансформація і очищення, щоб вони відповідали потребам бізнес-моделі; і завантаження їх в сховище даних.

Зараз трудомістка пакетна операція, ETL що рекомендується частіше для створення менших цільових сховищ даних, які вимагають менш частого оновлення, тоді як використовуються інші методи інтеграції даних, включаючи ELT (витяг, завантаження, перетворення), CDC та віртуалізація даних. інтегрувати все більші обсяги потоків даних, що постійно змінюються, або потоків даних у режимі реального часу.

Завершувати вивчення теми рекомендується розглядом дизайном процесу ETL та підтримкою інструментів.

**Змістовий модуль 2. Діджиталізація бізнес-моделей**

**Тема 5. Надання інформації (звітування, інформаційні панелі)** Цю тему рекомендується розпочати з вивчення понять Звітування, Інформаційна панель, Google Analytics, Інформація.

Конфігурація як код - це практика зберігання конфігурації вашої системи як набору контрольованих версій, зручних для читання файлів конфігурації, а не в базі даних. Ці файли конфігурації можна використовувати повторно в усіх середовищах, щоб уникнути дублювання ресурсів.

У міру зростання кількості інформаційних панелей та джерел даних в організації, керування змінами вручну може стати схильним до помилок. Заохочення повторного використання стає важливим, щоб уникнути переробки декількох команд однакових інформаційних панелей.

Під час самостійного вивчення теми студенти повинні визначити основні завдання Проектування та впровадження інформаційних панелей Інформаційна панель (інфопанель, дашборд) — тип графічного інтерфейсу користувача, що забезпечує наочну презентацію основних показників продуктивності (ОПП), значимих для конкретної цілі чи підприємчого процесу. Інфопанелі є динамічними звітами в режимі реального. Студенти повинні засвоїти типи звітів, які повільно змінюються.

**Тема 6. Аналітичний життєвий цикл та методи: кластеризація, класифікація, машинне навчання**

Вивчення теми треба розпочати з визначення аналітичний життєвий цикл. У сучасному цифровому світі дані мають величезне значення. Протягом свого життя він проходить різні етапи, під час створення, тестування, обробки, споживання та повторного використання. Життєвий цикл Data Analytics намічає ці етапи для професіоналів, які працюють над проектами аналізу даних. Ці фази розташовані у круговій структурі, яка формує життєвий цикл Data Analytics. Кожен крок має своє значення та характеристики

Доступ до даних та їх збір можуть бути досить простими, якщо дані перебувають у сховищі даних, однак якщо дані розкидані у різних файлах, базах даних та операційних системах, що може бути нудним процесом. Якщо дані знаходяться в операційній системі, їх доведеться витягти в окреме сховище, щоб інтелектуальний аналіз даних не заважав нормальній роботі. Дані також потрібно оцінювати на предмет повноти, надмірності, відсутніх значень, дублювання, а іноді може знадобитися зменшення складності, наприклад, округлення даних та зменшення кількості десяткових значень. Наступним кроком є підготовка даних, і цей етап може бути як мистецтвом,

так і наукою. Це справді найбільш трудомістка частина процесу видобування даних, і це часто може зайняти від 60 до 80% загальних зусиль. Однак, як тільки буде створений сценарій коду підготовки даних, побудова наступної моделі на даних одного типу може бути набагато швидшою, оскільки все, що потрібно зробити, це запустити цей скрипт на необроблених даних та завантаженні даних, і підготовка виконується автоматично (зазвичай з від декількох хвилин до декількох годин або на ніч).

Продовженням вивчення даної теми буде питання Сучасна архітектура даних. Машинне навчання. Сутність класифікації. Знайомство з методами кластеризації даних.

**Тема 7. Моделювання великих даних**

Тему рекомендується розпочати з вивчення основних методів моделювання. Модель - об'єкт або опис об'єкта, системи для заміщення (при певних умовах, припущеннях, гіпотезах) однієї системи (тобто оригіналу) іншою системою для кращого вивчення оригіналу або відтворення будь-яких його властивостей.

Моделювання - універсальний метод отримання, опису та використання знань.

Моделювання даних - це складна наука, яка передбачає організацію корпоративних даних так, щоб вона відповідала потребам бізнес -процесів Моделювання великих даних залежить від багатьох факторів, включаючи структуру даних, які операції можуть виконуватися з даними та які обмеження накладені на моделі. Big Data створює нові вимоги, засновані на складності збору даних, їх зберігання, аналізу та візуалізації даних. Продовжити засвоєння теми необхідно з вивчення Big Data в маркетингу та засвоїти важливість моделювання даних у світі великих даних.

**Тема 8. Архітектура та розгортання**

Самостійне вивчення питань цієї теми дасть змогу майбутнім бакалаврам ознайомитися з новими можливостями для управління та управління внутрішніми та зовнішніми інформаційними потоками, щоб перетворити їх у стратегічні ресурси, здатні визначати стратегії позиціонування, які відповідають потребам споживачів, які стають дедалі

більш поінформованими та вимогливими. Як технічні, накладають кардинальну зміну бізнес-моделей та людських ресурсів з точки зору інформаційної спрямованості та вдосконалення корпоративного інформаційного активу як підтримку процесу прийняття стратегічних рішень.

Певна увага повинна бути приділена студентами опрацюванню питання стилю архітектури великих даних та еволюції моделей розгортання в епоху великих даних.

Менеджерам часто важко приймати рішення щодо інноваційних технологій, головним чином через відсутність технічних знань. Буде надано інструмент для підтримки цих оцінок з метою зменшення ризику, пов’язаного з інвестицією у великі дані. Цей інструмент пов'язує звичайні бар'єри, що перешкоджають успішному проекту, з технічним вибором, який слід зробити під час впровадження архітектури великих даних.

**3 ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ ЗА ТЕМАМИ КУРСУ** «**Аналіз великих даних та бізнес-аналітика**» **Змістовий модуль 1. Сучасні аспекти цифрової трансформації бізнесу**

Лабораторна робота № 1.

ВІКОРИСТАННЯ ПРОГРАМНОГО ПАКЕТА WEKA ДЛЯ АНАЛІТИКИ ВЕЛИКИХ ДАНИХ

Навчальні питання:

1. Основи роботи з програмним пакетом WEKA.

2. Створення набору даних для завантаження в WEKA.

3. Створення регресійної моделі в WEKA.

Завдання

1. Ознайомитись з програмним пакетом з відкритим вихідним кодом для аналізу даних Waikato Environment for Knowledge Analysis (WEKA). 2. Провести регресивний аналіз для інтелектуального аналізу даних. 3. Розглянути основні поняття: метод найменших квадратів, нормальний розподіл, коефіцієнт детермінації R-квадрат.

4. Перетворити набір даних у формат, зрозумілий для програмного пакету WEKA.

5. Завантажити дані до програмного пакету WEKA.

6. Створити регресівну модель в програмному пакеті WEKA. 7. Створити звіт з лабораторної роботи, в якому стисло описати виконані дії та навести необхідні ілюстрації.

Методичні рекомендації

Waikato Environment for Knowledge Analysis (WEKA), є вільно поширюваним програмним пакетом з відкритим вихідним кодом для аналізу даних. WEKA забезпечує графічний користувальницький інтерфейс для роботи з файлами даних і генерації візуальних результатів (у вигляді таблиць і графіків). Крім того, можливо інтегрувати WEKA, як і будь-яку іншу бібліотеку, у свої 6 власні додатки, наприклад, для автоматизації аналізу даних на стороні сервера, використовуючи стандартний API.

WEKA поширюється по ліцензії GNU General Public License (GPL). WEKA має чотирі графічних інтерфейсу для роботи з даними. 

Рис. 1.1. Вікно вибору інтерфейсу в WEKA

Метод регресійного аналізу є найпростішим, але одним з найменш ефективнх методів інтелектуального аналізу даних.

Метод найменших квадратів (МНК, OLS, Ordinary Least Squares) – математичний метод, який застосовується для вирішення різних задач, заснований на мінімізації суми квадратів деяких функцій від шуканих змінних. Він може використовуватися зокрема для апроксимації точкових значень деякою функцією. МНК є одним з базових методів регресійного аналізу для оцінки невідомих параметрів регресійних моделей за вибірковими даними.

Нормальний закон розподілу (normal law of distribution) (закон Гаусса) відіграє виключно важливу роль в теорії ймовірностей і займає серед інших законів розподілу особливий стан. Це закон, який найчастіше зустрічається на практиці.

Коефіцієнт детермінації (позначається як R2 – R-квадрат) – статистичний показник, що використовується в статистичних моделях як міра залежності варіації залежної змінної від варіації незалежних змінних. Вказує наскільки отримані спостереження підтверджують модель.

Для завантаження даних в WEKA, їх слід перетворити у формат, зрозумілий для цього програмного пакету. Найбільш підходящим форматом для завантаження даних в WEKA є формат Attribute-Relation File Format (ARFF), який спочатку визначає тип завантажуваних даних, а потім вказує власне дані.

Контрольні запитання

1. Що таке WEKA?

2. Поясніть метод регресійного аналізу.

3. Поясніть метод найменших квадратів.

4. Що таке закон нормального розподілу?

**Лабораторна робота №2**

Створення баз даних. Робота з таблицями.

Мета роботи: оволодіти практичними навиками створення баз даних, роботи з таблицями.

Програма виконання роботи

1. Завантажити Access. 2. Створити базу даних “Студенти”. 3. Створити за допомогою конструктора таблиці бази даних згідно завдання на практичну роботу.

4. Ввести дані в таблиці в режимі редагування.

5. Переглянути введені дані таблиць.

6. Зберегти базу даних.

Завдання на практичну роботу

Сформувати базу даних для обробітку інформації по студентам університету.

При заповненні таблиць бази даних дані взяти довільні. Кількість об’єктів у головній таблиці бази даних повинна бути не менше 10, а в додаткових – не менш 5.

Структура таблиць бази даних “Студенти”:

∙ Студенти (Призвище, Ім’я, По батькові, Рік народження, Код факультету, Код спеціальності, Група, Курс, Рік вступу, Код області, Адреса, Місце проживання) – головна таблиця;

∙ Список факультетів (Код факультету, Назва факультету); ∙ Список спеціальностей (Код спеціальності, Назва спеціальності); ∙ Склад факультетів (Код факультету, Код спеціальності);

∙ Список груп (Код групи, Код факультету, Назва групи, );

∙ Список областей (Код області, Назва). Теоретичні відомості База даних (БД) − це систематизоване сховище інформації.

Телефонний довідник − прекрасний приклад базі даних.

Теоретичні відомості:

Бази даних створюються в тих випадках, коли необхідно

відслідковувати, аналізувати і зберігати інформацію за деякий час. Система управління базою даних (СУБД) − це програма, основними функціями якою є створення

структури БД, запис, пошук, сортування, обробка (аналіз) та друк інформації, яка зберігається в базі даних. В комп’ютерній базі даних інформація подається у вигляді таблиці, дуже схожій на електронну таблицю. Назва стовпців − це імена полів, а самі стовпці − поля. Рядки даних звуться записами.

Microsoft Access − одна із простіших систем управління реляційними базами даних, яка входить до Microsoft Office.

Працюючи з Access, можна вирішити наступні задачі:

∙ розбити дані на логічно зв’язані частини;

∙ ввести, редагувати і знаходити необхідні дані;

∙ знаходити безліч даних по заданим умовам;

∙ створювати форми і звіти;

∙ автоматизувати виконання стандартних задач;

∙ графічно встановлювати зв’язок між даними;

∙ вставляти малюнки у форми і звіти;

∙ створювати власні програми з меню, діалоговими вікнами та командними кнопками, які готові до роботи з базою даних. Проектування бази даних.

Перед тим як приступити до створення таблиць, форм і звітів необхідно витрати деякий час на проектування базі даних. Сконцентруйте вашу увагу на даних, виконавцях та задачах.

Деякі поради, яких треба притримуватися при розробки бази даних:

♣ Почніть з аналізу існуючій бази даних (створеної вручну або за допомогою комп’ютера). Перегляньте формі і звіти, які використовуються. Визначте джерело даних (якщо це комп’ютер, продумайте можливість імпортування даних). Переговоріть з людьми, які використовують інформацію з бази даних, обговоріть проблеми і побажання, які виникають у зв’язку з цим. Перегляньте задачі, які вирішуються за допомогою бази даних, та продумайте ті, які потрібно вирішувати в майбутньому.

♣ Ідентифікував дані та задачі, які необхідно вирішувати, розділити їх на групи. Ці групи пізніше стануть таблицями.

♣ Визначити типи даних, які будуть зберігатися в кожній таблиці. Ці групи даних таблиці називаються полями.

♣ Перегляньте загальні елементи таблиць (наприклад, назва клієнта може бути загальним елементом таблиці клієнтів і таблиці накладних). Ці загальні елементи мають назву ключових полів.

♣ Обміркуйте оформлення форм і звітів.

♣ Визначте умови вибору для запитів.

♣ Обміркуйте, як автоматизувати стандартні задачі баз даних, таких як виконання запитів і друк звітів.

♣ Обміркуйте проблеми безпеки даних, таких як стратегія створення копій, поділення даних і обмеження доступу при роботі в мережі. ♣ по даній предметній області Таким чином, результатом процесу проектування бази даних є розподілення даних по таблицям, структура таблиць, встановлення зв’язку таблиць (визначення полів для зв’язку). Початок роботи з Access

Для завантаження Access необхідно виконати команду: “Пуск”◊ “Программы”◊ “Microsoft Access”.

В вікні, що з’явиться, є можливість створення нової бази даних та відкриття раніше створеної бази даних.

Створення нової бази даних

Після позначення пункту “Нова настільна база даних” необхідно вказати ім’я бази і місце її знаходження (диск, папку). Після натиснення кнопки “Створити”буде створено файл з розширенням .accdb. На екрані з’явиться вікно бази даних з переліком об’єктів БД (таблиці, запити, форми, звіти, сторінки, макроси, модулі).

Вікно бази даних має корінці для кожного типа об’єктів бази даних. Якщо вибираємо тип об’єкта , наприклад таблиці, то на екрані з’являються список імен таблиць бази даних. Для того щоб відкрити необхідний об’єкт,

двічі клацнути по його імені або виділити ім’я та вибрати команду «Відкрити» з контекстного меню. Для створення нових об’єктів бази даних або модифікації існуючих об’єктів бази даних використовуйте меню “Створити”.

Контрольні запитання

1. Яке призначення етапу проектування бази даних?

2. Визначити характеристики типів даних?

3. Яке призначення полів таблиць і їх характеристики?

4. Охарактеризувати об’єкти бази даних?

5. Визначити можливості роботи з таблицями бази даних?

**Лабораторна робота № 3**

ВІЗУАЛІЗАЦІЯ ДАНИХ ЗА ДОПОМОГОЮ СИСТЕМИ STATISTICA Навчальні питання:

1. Основні поняття та принципи роботи з таблицями в системі Statistica. 2. Типи графічних даних в системі Statistica.

3. Використання функції для аналізу великиї даних в системі Statistica. Завдання

1. Ознайомитися з принципом роботи системи Statistica.

2. Створити новий аркуш у системі Statistica.

3. Ознайомитися з принципами роботи з рядками та стовпцями на аркуші.

4. Заповнюти таблицю довільними даними та виконати обчислення стандартних функцій.

5. Додати новий аркуш у системі Statistica.

6. Заповнюти таблицю з трьома зміними. Перейменувати їх на х, у і z. 7. Побудувати графік по зміними з допомогою пункту меню GraphicsHistograms.

8. Змінити властивості графіка – колір, товщину ліній тощо. 9. Використовуючи вікно з налаштування властивостей графіка змінити тип графіка.

10. Аналогічно побудувати іншій тип графіка – Graphs-2D Graphs Scatterplots.

11. Додати на графіки назву та діапазон значень осей.

12. Аналогічно побудувати 3D графік за трьома змінними (Surface Plots).

13. Змінити властивості графіка – кут нахилу та повороту. 14. Створити нову таблицю з чотирьома змінними: номер залікової книжки, прізвище, курс, успішність (середній бал). Заповнюти таблицю довільними значеннями.

15. Відсортувати значення в заліжності від типу даних змінних: текст – в алфавітному порядку, числові дані – за зростанням.

16. Побудувати графік залежності середнього балу від курсу навчання. 17. Змінити властивості графіка, функцію кривої з лінійної на довільну нелінійну.

18. Проаналізувати графік та описати залежності.

19. Побудувати гістограму для змінної успішність.

20. Змінити функцію кривої Normal на будь-яку іншу, використовуючи вікно властивостей графіка.

21. Створити звіт з лабораторної роботи, в якому стисло описати виконані дії та навести необхідні ілюстрації.

Методичні рекомендації

Система «STATISTICA», розроблена компанією Statsoft, є однією з найбільш популярних статистичних програм для пошуку закономірностей, прогнозування, класифікації, візуалізації даних. Може застосовуватися в економіці, промисловості, медицині, наукових дослідженнях і інших сферах людської діяльності. Клієнтами Statsoft є найбільші компанії зі світовим ім’ям.

STATISTICA – це інтегрована система аналізу та управління даними. STATISTICA - це інструмент розробки користувальницьких додатків в бізнесі, економіці, фінансах, промисловості, медицині, страхуванні та інших областях. STATISTICA легка в освоєнні і використанні.

У системі існує можливість застосовувати класичні й новітні методи проведення аналізу даних: кластерний, факторний, кореляційний, дисперсійний аналіз, лінійну й нелінійну регресії, нейронні мережі й ін. Візуалізація вихідних, проміжних, вихідних даних може бути здійснена вибором з великої кількості різних графіків, піктографіків і діаграм.



Рис. 3.1. Вікно системи STATISTICA

STATISTICA складається з набору модулів, в кожному з яких зібрані тематично зв’язкові групи процедур. При перемиканні модулів можна або залишати відкритим тільки одне вікно програми STATISTICA, або всі викликані раніше модулі, оскільки кожен з них може виконуватися в окремому вікні.

Контрольні запитання

1. Що таке система «STATISTICA» і для чого вона потрібна? 2. Охарактеризуйте особливості графічного формату \*.stg.

3. Як експортувати графік в інший застосунок?

4. Як імпортувати на графік системи STATISTICA графічний об’єкт з іншої програми?

5. Як помістити текст на графік STATISTICA (звіти, таблиці тощо)?

**Лабораторна робота № 4.**

НАЛАШТУВАННЯ РЕПЛІКАЦІЇ НА СКБД MYSQL

Навчальні питання:

1. Основне призначення репликації даних.

2. Види репликацій даних.

3. Використання репликацій в MySQL.

Завдання

1. Ознайомитися з поняттями реплікації даних.

2. Розглянути типи реплікацій: Master-slave; Master-slave для деякількох slaveсерверов; Master-master.

3. Визначити, коли виконується затримання репликації та як отримати актуальні дані при аналізі.

4. Визначити проблеми репликації в MySQL та як їх вирішувати. 5. Переваги асинхронної репликації.

6. Встановити окрему операційну систему Ubuntu Server 16.04.3 LTS на віртуальну машину VirtualBox.

7. Налаштувати статичний ІР-адрес та доменне ім’я вузлу. 8. Налаштувати доступ за ssh та відключити використання IPv6. 9. Встановити на віртуальну машину MySQL-сервер.

10. Налаштувати master-slavе реплікації:

− налаштувати провідний сервер;

− створити базу даних, в якої створити таблицу з наступними даними: create database clusterdb; create table clusterdb.simples (id int not null primary key); insert into clusterdb.simples values (999),(1),(2),(3); − переглянути зміст таблиці: select \* from

clusterdb.simples

11. Налаштувати права на реплікацію:

− створити профіль користувача: mysql -u root -p - -prompt='master> '

− створити та призначити права користувачу для репліки: grant replication slave on \*.\* to 'slave\_user'@'%' identified by 'slavepass';

− обновити права доступу: flush privileges; 12. Заблокувати всі таблиці в створенної базі даних:

use clusterdb;

flush tables with read lock;

13. Перевірити статус провідного серверу: show master status; 14. Вийти з коснолі MySQL.

15. Створити дамп бази даних: mysqldump -u root -p clusterdb > /home/user/clusterdb.sql

16. Зайти в консольта розблокувати таблиці.

17. Створити базу на провідному сервері.

18. Налаштувати провідній сервер.

19. Активувати репликації на провідному сервері.

20. Перевірити роботу репликації.

21. Налаштувати права на реплікацію.

22. Активувати репликації на первому мастер-сервері.

23. Перевірити роботу репликації.

24. Створити звіт з лабораторної роботи, в якому стисло описати виконані дії та навести необхідні ілюстрації.

Методичні рекомендації

Реплікація (англ. Replication) – механізм синхронізації вмісту декількох копій об'єкта (наприклад, вмісту бази даних). Під репликацией також розуміють процес копіювання даних з одного джерела на інший (або на безліч інших).

Реплікація є однією з технік масштабування даних в розподілених системах. Реплікація діляться на синхронні і асинхронні.

У разі синхронної реплікації, якщо одна репліка (копія) оновлюється, всі інші репліки того ж фрагмента даних також повинні бути оновлені в одній і тій же транзакції. Це означає, що всі репліки залишаються повинні суперечити одна одній. Недоліком даного методу є високі накладні витрати на синхронізацію реплік.

У разі асинхронної реплікації оновлення однієї репліки поширюється на інші через деякий час, а не в тій же транзакції. Таким чином, при асинхронної реплікації вводиться затримка, або час очікування, протягом якого окремі репліки можуть бути фактично неідентичних. У той же час накладні витрати на реплікацію істотно зменшуються.

Master-slave реплікація працює з одним основним сервером бази даних, який називається ведучим. На ньому відбуваються всі зміни в даних (будь-які запити MySQL). Ведений сервер постійно копіює всі зміни з майстра. З додатка на сервер відправляються запити читання даних (запити SELECT). Отже, провідний сервер відповідає за зміни даних, а ведений – за читання.

Master-slave реплікація на кілька slave серверів – перевага цього типу реплікації в тому, що можна використовувати більш одного відомого сервера. Зазвичай слід використовувати не більше 20 відомих серверів при роботі з одним ведучим. MySQL реалізує асинхронну реплікацію. Це означає, що дані на відомому сервері можуть з'явитися з невеликою затримкою. Контрольні запитання

1. Дати визначення реплікації даних.

2. В чому відмінність синхронної та асинхронної реплікації? 3. В чому полягає master-slave реплікація?

4. В чому полягає master-master реплікація?

**Змістовий модуль 2. Діджиталізація бізнес-моделей**

**Лабораторна робота № 5.**

ВИКОРИСТАННЯ ПРОГРАМНОГО ПАКЕТА WEKA:

КЛАСИФІКАЦІЯ І КЛАСТЕРИЗАЦІЯ

Навчальні питання:

1. Класифікація в WEKA.

2. Кластеризація в WEKA.

3. Метод найближчих сусідів.

Завдання

1. Ознайомитись з методом класифікації – алгоритмом аналізу даних для визначення покрокового способу прийняття рішень.

2. Створити набір даних для аналізу за різними атрибутами. 3. Створити конкретну моділь аналізу засобами пакету WEKA. 4. Ознайомитися з методом кластерного аналізу та розділити дані на окреми групи за певними ознаками.

5. Підготувати дані для кластерного аналізу засобами пакету WEKA. 6. Створити модель кластерізації.

7. Проаналізувати отримані результати.

8. Створити звіт з лабораторної роботи, в якому стисло описати виконані дії та навести необхідні ілюстрації.

Методичні рекомендації

Метод класифікації (також відомий як метод класифікаційних дерев або дерев прийняття рішень) – це алгоритм аналізу даних, який визначає покроковий спосіб прийняття рішення в залежності від значень конкретних параметрів. Дерево цього методу має наступний вигляд: кожен вузол являє собою точку прийняття рішення на підставі вхідних параметрів. Залежно від конкретного значення параметра переход здійснюється до наступного вузла, від нього – до наступного вузла, і так далі, поки не дійде до листка, який і дає остаточне рішення.



Рис. 5.1. Класифікація даних в WEKA

Кластеризація дозволяє розбити дані на групи, кожна з яких має певні ознаки. Метод кластерного аналізу використовується в тих випадках, коли необхідно виділити деякі правила, взаємозв’язки або тенденції у великих наборах даних. Залежно від потреб, ви можете виділити кілька різних груп даних. Одна з явних переваг кластеризації в порівнянні з класифікацією полягає в тому, що для розбиття множини на групи може використовуватися будь-який атрибут (метод класифікації використовує тільки певну підмножину атрибутів).



Рис. 5.2. Кластерізація даних в WEKA

Контрольні запитання

1. Які існують методи інтелектуального аналізу даних?

2. Під якими іншими назвами відомий метод класифікації? 3. У чому полягає основна перевага класифікаційних дерев?

4. В чому полугає сутність підходу створення моделі на основі навчальної послідовності?

5. З якою метою в методі класифікаційного аналізу набір даних ділиться на дві частини?

6. Описати принцип відсікання гілок.

7. Охарактеризувати сенс проблеми помилкового розпізнавання. 8. В результаті аналізу для тестового і навчального набору даних одержана приблизно однакова точність. Як це вплине на нові дані які будуть використовуватися в цій моделі в майбутньому?

**Лабораторна робота № 6.**

ВИКОРИСТАННЯ ПРОГРАМНОГО ПАКЕТА WEKA: МЕТОД НАЙБЛИЖЧИХ СУСІДІВ

Навчальні питання:

1. Математичний апарат методу найбліжчих сусідів.

2. Реалізація методу найближчих сусідів в WEKA.

Завдання

1. Ознайомитись з методом найближчих сусідів для аналізу великих даних. 2. Створити математичну модель методу найближчих сусідів. 3. Підготувати файл з даними для аналізу методом найближчих сусідів за допомгою пакету WEKA.

4. Створити модель методу найближчих сусідів для набору даних. 5. Проаналізувати результат обчислення моделі.

6. Змінити кількість найбліжчих сусідів в моделі та виконати обчислення. 7. Визначити недоліки методу найближчих сусідів.

8. Створити звіт з лабораторної роботи, в якому стисло описати виконані дії та навести необхідні ілюстрації.

Методичні рекомендації

Метод найближчих сусідів, так само відомий як метод колаборативної фільтрації або метод навчання на прикладах, є досить корисним способом інтелектуального аналізу даних, що дозволяє використовувати накопичені приклади з відомим результатом для прогнозування очікуваного результату для нових зразків даних.



Рис. 6.1. Метод найближчих сусідів в WEKA

Алгоритм методу найближчих сусідів в чомусь схожий з алгоритмом, використовуваним в методі кластеризації. Метод визначає відстань між невідомою точкою і всіма відомими точками даних. Визначення відстані - цілком тривіальна процедура, що легко виконується в рамках електронних таблиць, так що досить потужний комп’ютер впорається з цим завданням практично миттєво. Найпростіший і найбільш поширений спосіб визначення відстані – це нормалізована евклідова відстань.

Контрольні запитання

1. Які інші назви використовуються для методу найближчих сусідів? 2. В чом полягає основна відмінність регресійного аналізу від методу найближчих сусідів?

3. Описати математичний алгоритм методу найближчих сусідів. 4. Яким чином можна розширити алгоритм методу найближчих сусідів?

**Лабораторна робота № 7**

Створення й редагування бази даних (БД).

**Мета:** опанувати способами створення й редагування бази даних.

**Теоретична частина:**

***База даних*** (БД) - це сукупність даних, які мають властивості структурованості й взаємозв'язку, а також незалежності від прикладних програм. У БД зберігається інформація про об'єкти.

Головна перевага автоматизованого ведення БД – швидкий пошук необхідних відомостей і подання їх у зручній формі. Це здійснюють прикладні

програми – ***системи керування базами даних (СУБД).*** Більшість СУБД підтримує бази даних реляционного типу. У реляционной моделі БД дані зберігаються в таблицях. Кожна таблиця повинна описувати один об'єкт, описувати його повністю й не містити повторюваних записів. Таблиці зв'язуються по ключових полях.

Найбільше поширення в нас у країні одержав пакет СУБД - ***Access*** фірми Microsoft. У процесі роботи будуть використатися кнопки груп об'єктів Access: - ***таблиці*** – основні об'єкти БД. У них зберігаються дані; - ***запити*** – покликані для пошуку даних, які відповідають існуючим умовам;

- ***форми*** – дозволяють одержувати дані в необхідному виді. Вони використаються також для пошуку даних;

- ***звіти*** – з їхньою допомогою дані подаються на принтер у зручному й наочному виді;

- ***макроси*** – це макрокоманди. Якщо якісь оператори з базою використаються часто, то має сенс згрупувати кілька команд в один макрос і позначити його виділеною комбінацією клавіш;

- ***модулі*** – це програмні процедури, написані мовою Visual Basic. **Завдання для самостійної роботи:**

**2 рівень:**

1. Cоздайте таблиці з додатка № 1 (таблиці «Видавництво» й «Книга Автор»)

2. Задайте зв'язку між таблицями.

**3 рівень:**

1. Створіть форми для уведення даних для таблиць «**Видавництво**», «**Книги**», «**Книга-Автор**».

2. Створіть у формі **«Книги»** поля зі списками в наступних полях: 1.1. **«Код автора»** й **«Код книги»** у таблиці **« Книги - Автор».**

1.2. **«Код видавництва», «Тема», «Тип**

**обкладинки», «Формат»** у таблиці **«Книги».**

**4 рівень:**

1. Розробіть БД для магазина **«Відеопрокат».** Опишіть всі об'єкти, продумайте зв'язку, виділите первинний і зовнішній ключі. Розпишіть типи даних для кожного поля.

**Контрольні запитання:**

**1 рівень:**

1. Що таке БАЗА ДАНИХ?

2. Які СУБД ви знаєте?

**2 рівень:**

1. Опишіть основні категорії об'єктів БД.

2. Що таке об'єкт?

3. Опишіть способи створення таблиць.

**3 рівень:**

1. Укажіть потужності співвідношень?

2. Співвідношення якої потужності припустимі в реляционной БД?

**4 рівень:**

1. Які моделі БД ви знаєте?

2. Укажіть на розходження між різними моделями?

3. Чому реляционная БД одержала найбільш широке застосування на сьогоднішній день?

Додаток 1







**Лабораторна робота №8**

Пошук інформації в базі даних.

**Мета:** навчитися створювати запити й знаходити потрібну інформацію в базі даних.

**Теоретична частина:**

Базою даних (БД) є сукупність даних, які певним чином структуровані й взаємозалежні між собою, незалежні від прикладних програм. У БД зберігається інформація про об'єкти. Для пошуку необхідної інформації можна скористатися ***фільтром***. Для того щоб вибрати потрібний запис, потрібно відкрити таблицю, що містить необхідні вам запису. Для цього варто встановити курсор на слово, по якому ви хочете проводити пошук, і нажати кнопку **Фільтр** по виділеному слову.

При необхідності можна скористатися засобом «**Пошук**». У діалогове вікно необхідно ввести значення поля й запустити пошук.

Запити дозволяють відібрати дані, що втримуються в різних таблицях бази, а також виконати відбір відповідно до заданих умов. Створення запиту можливо за допомогою **Майстра** або в режимі **Конструктора**, що дозволяє задавати різні умови відбору й використати функції. Умови пошуку – логічне вираження. Просте логічне вираження є операцією відносин (>, <, =, <>, >=, <=). Складне логічне вираження містить логічні операції **AND, OR, NOT**.

**Завдання для самостійної роботи:**

**2 рівень:**

1. Виведіть на екран дані про книгу «Собаче серце» і його авторі М. Булгакові.

1. Знайдіть дані про письменника А.С. Пушкіні.

**3 рівень:**

1. Знайдіть книги із циклу **Фантастика** за ціною 45 грн. 2. Виведіть кількість книг кожного автора на складі. 2. Створіть можливість переглядати книги певного автора.

**4 рівень:**

1. Відобразите інформацію про книги М.И. Шолохова, які були видані до 1941 року.

2. Визначите термін служби книги.

3. Визначите книги **«Ювіляри»,** яким здійснилося 10, 25, 50 або 100 років.

**Контрольні запитання:**

**1 рівень:**

1. Що таке фільтр? База даних?

2. Як здійснюється пошук у БД?

**2 рівень:**

1. До скількох полів можна застосовувати фільтр?

2. Які способи створення запитів ви знаєте?

**3 рівень:**

1. Які типи запитів ви знаєте?

2. Чи можна використати дані одного запиту в іншому? Якщо так, то навіщо це робити?

**4 рівень:**

1. Для чого застосовується умова відбору в запитах?

2. Які данні ми можемо використати в умовах відбору?

**Питання до іспиту**

1. Розробка додатків для технологій аналітичної обробки (ADAPT) 2. ADAPT: Розміри

3. Архітектура "системи великих даних"

4. Складові сховища даних: куб даних

5. ADAPT: Паралельні ієрархії

6. Підходи до впровадження - можливі рішення

7. Порівняльні дані - Сховище основних даних

8. Компоненти сховища даних: база даних

9. ETL - Видобуток

10. Складові сховища даних: Марки даних

11. Складові сховища даних: менеджер

12. Сховище даних/ Ринки даних - Варіанти архітектури

13. Інтеграція даних: перспектива бізнесу

14. ADAPT: Роз'єми

15. Джерела даних

16. Джерела даних - вимоги до якості

17. Багатовимірна модель даних - оператори

18. Сховище даних - Характеристика

19. ADAPT: Факти та заходи

20. Сховище даних проти озера даних

21. ETL - Мета та вимоги

22. ETL - Інструменти

23. Видобуток: початкові навантаження проти дельта -навантажень 24. Озеро даних: Визначення

25. Видобування: види доставки даних

26. Гомогенізація: прості перетворення

27. Впровадження інформаційних панелей: Вимоги

28. Багатовимірна модель даних - основні елементи

29. ADAPT: Атрибути та учасники

30. Споживачі інформації: інформаційні панелі

31. ETL - Виклики

32. Оператори OLAP: свердління і свердління поперек

33. Видобуток: Стратегія моніторингу

34. Забезпечення інформацією в BI

35. Забезпечення інформацією в BI

36. Оператори OLAP: поворот / обертання

37. Інтеграція неструктурованих даних

This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in. Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі.



38. Очікувані переваги через використання BI (інформаційні панелі) 39. Логічна модель даних для зберігання даних

40. АДАПТ: Ієрархії

41. Моделювання в сховищі даних

42. Багатовимірна модель даних - Гіперкуб

43. Багатовимірні відносини між сутностями (ME/R-модель) 44. Архітектура сховища даних із розширенням ODS

45. Трансформація: Виклики

46. OLAP-оператори: кістки

47. Вимоги до інформаційних панелей: „SMART“

48. Онлайн -аналітична обробка (OLAP): FASMI

49. Фази зберігання даних

50. Візуальний дизайн панелей приладів - Діаграми

51. Вимоги до систем BI 52. ROLAP - Розміри / Ієрархія класифікації 53. Схема зіставлення: нормалізована проти денормалізована 54. Оператори OLAP: згортання та деталізація

55. Семантичне моделювання в сховищі даних

56. Трансформація: Завдання в процесі ETL

57. Типи інформаційних панелей: Стратегічні інформаційні панелі 58. Візуальний дизайн панелей приладів - Ваги

59. OLAP-оператори: Зріз

60. Візуальний дизайн/ кодування

**СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ**

1. Заяць В. М. Роль інформаційних технологій у формуванні стратегічного мислення менеджера / В. М. Заєць // Актуальні проблеми економіки. – 2009. – №6 (96). – С. 280-288.

2. Європейська Бізнес Асоціація: www.eba.com.ua. - Нові підходи в управлінні ІТ або як бізнес-цілі, пов’язані з ІТ-процесами.

3. MastersInDataScience.org is owned and operated by 2U, Inc.. Here Are 10 Key Benefits of Business Intelligence Software. [Електроний носій]. - https://www.mastersindatascience.org/learning/benefits-of-business intelligence/

4. Дерево рішень: приклад. Алгоритми побудови дерева рішень. Електроний носій]. - http://yrok.pp.ua/serednya-osvta/12020-derevo-rshen priklad-algoritmi-pobudovi-dereva-rshen.html

5. Інтеграція даних із Microsoft Dataverse. Електроний носій]. - https://docs.microsoft.com/uk-ua/power-platform/admin/data-integrator 6. The Inexorable Rise of Self Service Data Integration. [Електроний носій]. - https://blogs.gartner.com/andrew\_white/2015/05/22/the-inexorable-rise-of self-service-data-integration/

7. 15 найкращих інструментів ETL у 2021 році (повний оновлений список). [Електроний носій]. - https://uk.myservername.com/15-best-etl-tools-2021 8. Проектування інформаційних систем: навчальний посібник / В.С. Авраменко, А.С. Авраменко. – Черкаси: Черкаський національний університет ім. Б. Хмельницького, 2017. – 434 с.: іл.

9. Будсвіт Україна » Маркетинг » 3 типи маркетингових інформаційних панелей та способи їх використання. [Електроний носій]. - https://budsvit.net.ua/3-typy-marketyngovyh-informaczijnyh-panelej-ta sposoby-yih-vykorystannya

10.Data Analytics Lifecycle: An Easy Overview For 2021 [Електроний носій]. - https://www.jigsawacademy.com/blogs/hr-analytics/data-analytics lifecycle/

11.Ахмед А. А. Гад-Елраб. Сучасний бізнес-аналіз: аналіз великих даних та штучний інтелект для створення цінності, керованої даними. [Електроний носій]. - https://www.intechopen.com/chapters/76332

12.Інтелектуальний аналіз даних: Комп’ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп’ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеєв-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 1,72 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 73 с.: Іл

13.By Craig S. Mullins. The Importance of Data Modeling in a Big Data World. [Електроний носій]. - https://www.dbta.com/Editorial/Think-About-It/The Importance-of-Data-Modeling-in-a-Big-Data-World-145915.aspx

14.Big data architecture style. - [Електроний носій]. - https://docs.microsoft.com/en-us/azure/architecture/guide/architecture styles/big-data

15.By Judith S. Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman. The Evolution of Deployment Models in the Big Data Era. [Електроний носій]. -https://www.dummies.com/programming/big-data/engineering/big-data cloud-deployment-models/

16.By Judith S. Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman. Big Data Cloud Deployment Models. [Електроний носій]. - https://www.dummies.com/programming/big-data/engineering/big-data cloud-deployment-models/

17.Hariri, R.H., Fredericks, E.M. & Bowers, K.M. Uncertainty in big data analytics: survey, opportunities, and challenges. J Big Data 6, 44 (2019). https://doi.org/10.1186/s40537-019-0206-3

18.R. Lovas, A. Farkas, A. C. Marosi et al., “Orchestrated Platform for Cyber Physical Systems,” Complexity, vol. 2018, Article ID 8281079, 16 pages, 2018.

19.R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, “Intelligent Manufacturing in the context of industry 4.0: a review,” Engineering Journal, vol. 3, no. 5, pp. 616–630, 2017.

20.R. M. Müller, H.-J. Lenz. 2013. Business Intelligence

21.Y. Su, X. Meng, Q. Kang, and X. Han, “Dynamic Virtual Network Reconfiguration Method for Hybrid Multiple Failures Based on Weighted Relative Entropy,” Entropy, vol. 20, no. 9, p. 711, 2018.

22.Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis und Florian Mansmann. „Visual Analytics“. 2010

23. Дерево рішень: приклад. Алгоритми побудови дерева рішень. Електроний носій]. - http://yrok.pp.ua/serednya-osvta/12020-derevo-rshen priklad-algoritmi-pobudovi-dereva-rshen.html

24.Пример использования кластерного анализа STATISTICA в автостраховании. Електроний носій]. - http://statsoft.ru/solutions/ExamplesBase/branches/detail.php?ELEMENT\_I D=1573

25.Google Analytics для начинающих: самое полное руководство. Електроний носій]. - https://texterra.ru/blog/google-analytics-dlya nachinayushchikh-samoe-polnoe-rukovodstvo-v-runete.html