**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 вказівки

до виконання практичних занять

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 вказівки до виконання практичних занять для здобувачів ступеня вищої освіти «Магістр» зі спеціальності 071 «Облік і оподаткування». – Мелітополь, ТДАТУ, 2022 р. – 67 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

This project has been funded with support from the European Commission. This document reflects the views only of the author, and the

**ЗМІСТ**

ВСТУП

………………………………………………………………………...4 Практичне заняття № 1. ІТ-фактор впливу в сучасних управлінських завданнях…………………………………………………………………………. 6

Практичне заняття № 2. Моделювання та аналіз даних………………………8 Практичне заняття № 3. Архітектура та методи зберігання даних…………15 Практичне заняття № 4. Видобування, перетворення та навантаження (ETL).19 Практичне заняття № 5. Надання інформації (звітування, інформаційні панелі)………………………………………………………………………30 Практичне заняття № 6. Аналітичний життєвий цикл та методи: кластеризація, класифікація, машинне навчання…………………………37

Практичне заняття № 7. Моделювання великих даних……………………….40 Практичне заняття № 8. Архітектура та розгортання…………………………49 ПИТАННЯ ДО ІСПИТУ …………………………………………………..55 СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ ……………………………59

ГЛОСАРІЙ……………………………………………………………………….60

**ВСТУП**

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.

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.

Завдання вивчення дисципліни “ Аналіз великих даних та бізнес аналітика ”:

Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв’язанні задач проектування і використання ІСТ.

Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп’ютерних систем та середовищ з метою їх запровадження у професійній діяльності.

Видобувати знання шляхом інтеграції та аналізу великих даних, отриманих з різноманітних та різнорідних джерела інформації. Вміти обґрунтовувати вибір абстрактних типів даних та структур даних при проектуванні програмного забезпечення ІСТ.

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. Мета проведення практичних занять – формування практичних знань і умінь, необхідних для аналізу і роботи з великими даними.. Методичні вказівки до виконання практичних занять містять основні положення кожного змістового модуля, список рекомендованої літератури з кожної теми, а також тести за змістовими модулями дисципліни.

**Практичне заняття № 1.**

**ІТ-фактор впливу в сучасних управлінських завданнях**

План

1. Зміст поняття «іноваційні технології».

2. Класифікація ІТ-факторів.

3. Клімат ІТ-менеджменту

4. Ефективність процесу ІТ-управління та використання ІТ

5. Критерії оптимізації організаційних структур на основі ІТ

Таблиця 1.1

| **Ключові поняття:** | |
| --- | --- |
| - Іфнормаційні технології  - ІТ фактори  - ІТ процес | - Модель структурного рівняння - ІТ клімат  - Бізнес-цілі |

**Питання для обговорення**

1. Які напрями можно виділити в управління ІТ-послугами? 2. В чому полягає Клімат ІТ-менеджменту?

3. Сутність управлінських ІТ-знань

4. Програмний пакет weka для аналітики великих даних

5. У яких 4х сферах реалізуються 17 бізнес-цілей згідно збалансованої системи показників (Balanced Scorecard)?

**Питання для самоперевірки та контролю**

1. Дайте визначення ІТ.

2. В чому полягає вирівнювання ІТ-цілей і ІТ-процесів

3. Які стратегії впровадження ІТ в систему управління підприємства ви знаєте?

4. Назвіть напрямки удосконалення систем управління на основі ІТ.

**Тестові завдання**

1) Що розуміють під органом управління в організації?

А) сукупність ланок керованої системи

Б) ланку або сукупність ланок керуючої системи

В) окремі ланки управління організацією

Г) осіб, які здійснюють управлінський вплив на трудовий колектив 2) Розставити інформаційні процеси в логічнійй послідовності: А) Передавання

Б) Пошук

В) Отримання

Г) Кодування

Д) Відображення

Е) Використання

Ж) Зберігання

З) Опрацювання

3) Інформаційна система – це:

А) система, яка включає в себе об’єкт, який одночасно розглядається як єдине ціле, і як різнорідні елементи, об’єднані для досягнення поставлених цілей; Б) система, яка спрямована на збереження і маніпулювання інформацією у проблемній області;

В) система, яка включає в себе керований об’єкт, керуючий об’єкт і виконавчий орган;

Г) система, яка включає в себе весь обсяг знань, що відчужені від творців та зафіксовані на матеріальних носіях і призначених для загального використання.

4) В управління ІТ-послугами можна виділити 3 напрямки, оберіть зайве: А) Співставлення ІТ-послуг з поточними і майбутніми потребами бізнесу Б) Підвищення якості ІТ-послуг;

В) Портфель конкурентоспроможної продукції та послуг; Г)Оптимізація довгострокових витрат на надання ІТ-послуг. 5) На що вказує зрілість ІТ-процесів?

А) на зрілість ІТ-функції;

Б) на фінансову віддачу;

В) не має вірної відповіді.

**Практична ситуація 1.1**

Керівництво промислового підприємства "Успіх" отримує інформаційні ресурси з наступних джерел:

інформацію про конкурентів – з інформаційних баз, що доступні у мережі Internet;

інформацію про обсяги продажів – з власних фінансових документів; інформацію про потенційних клієнтів – з результатів маркетингових досліджень, що проводяться науково-дослідним інститутом; інформацію про стан кредитного ринку – зі спеціалізованих економічних журналів; інформацію про зміни в законодавстві – з інформаційно-пошукової системи "Ліга:Закон";

інформацію про обсяги постачання матеріалів – з документів відділу матеріально-технічного постачання.

Завдання до практичної ситуації

1. Визначте, до якого виду інформації (в залежності від джерела її виникнення) відноситься кожен з перерахованих інформаційних ресурсів. 2. Наведіть характерні ознаки цих видів інформації.

ВІКОРИСТАННЯ ПРОГРАМНОГО ПАКЕТА WEKA ДЛЯ АНАЛІТИКИ ВЕЛИКИХ ДАНИХ

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.

Практичне заняття № 2.

**Моделювання та аналіз даних**

План

1. Аналіз даних

2. Аналітичний підхід до моделювання

3. Інформаційний підхід до моделювання

4. Етапи аналізу даних

Таблиця 2.1

| **Ключові поняття:** | |
| --- | --- |
| - Аналіз даних  - Аналітичний підхід  - Інформаційний підхід | - Модель  - Моделювання |

**Питання для обговорення**

1. Етапи аналізу даних

2. Класифікація моделювання

3. Рух від моделі до результату

**Питання для самоперевірки та контролю**

1. Дайте визначення Аналіз даних – це?

2. Що називають процесом Моделювання ?

3. В чому полягає аналітичний підхід до моделювання?

4. Розкрийте сутність інформаційного підходу.

5. Назвіть етапи аналізу даних.

Приклад аналітичний підхід до моделювання

Розглянемо фізичну систему. Тіло масою m, на яке впливає сила F, скочується по похилій площині з прискоренням a. Досліджуючи такі системи, Ньютон отримав математичне співвідношення F = ma. Це математична модель фізичної системи.

При побудові цієї моделі були прийняті наступні гіпотези. 1. Поверхня ідеальна (тобто коефіцієнт тертя дорівнює нулю). 2. Тіло знаходиться у вакуумі (тобто опір повітря дорівнює нулю). 3. Маса тіла незмінна.

4. Тіло рухається з однаковим постійним прискоренням в будь-якій точці.

При моделюванні багатьох фізичних явищ ми використовуємо закон Ньютона і робимо висновки.

П'ять найкращих інструментів для аналізу даних:

Excel Упродовж довгого часу, у якому є великий інструмент аналітики, що отримав широке розповсюдження, був Excel. По-прежнему варто научитись користуватися Excel. Завдяки цьому він володіє індивідуальною функціональністю та потужністю, а багато компаній використовують його виключно. Панди Більш досвідчені компанії все ще можуть працювати з фреймворками, які можуть це робити, що Excel просто не може. Тот, з яким у мене більше всього досвіду, - це Панди.

Pandas - це платформа даних на основі Python, здатна виконати обширне перетворення, візуалізацію та аналіз даних. Це швидко заселяє отраслевим стандартом. R R - це повноцінний мову програмування, популярний в академічних кругах, але він також широко використовується в отраслях. Він побудований з нуля для статистичного аналізу, тому це відмінний мову для вивчення.

KNIME - це набір інструментів з відкритим вихідним кодом, який підтримує створення робочих процесів перетворення для кожної частини конвеєра аналізу, включаючи створення моделей машинного навчання. Більшість аналітиків не роблять це регулярно, але не мають такої здатності ніколи не помічає.

SAS - це середа та мова, які значною мірою спрощують прийом, обробку та аналіз даних. Він досить старий і має багато спеціалізованих модулів для завдань, маркетингу в соціальних мережах.

**Практична робота. Табличний процесор Microsoft Excel. Аналіз даних: підбір параметра, таблиці підстановки, консолідація даних** Завдання:

Використовуючи функцію Подбор параметра та дані наведеної таблиці, знайдіть, при якому значенні терміну вкладу сума виплат становитиме певне значення.

Створити таблицю підстановки, яка відображає вплив відсоткової ставки та терміну вкладу на суму виплат за заданими в таблиці даними. Послідовно виконайте всі завдання.

Функція "Подбор параметра"

Підбір параметра - спосіб пошуку певного значення комірки шляхом зміни значення в іншій комірці. При підборі параметра значення в комірці змінюється до тих пір, поки формула, залежна від цієї комірки, не поверне необхідний результат. Microsoft Excel змінює значення в одній конкретній комірці доти, поки формула, залежна від цієї комірки, не повертає потрібний результат.

Завдання 1. Використовуючи функцію Подбор параметра та дані наведеної таблиці, знайдіть, при якому значенні терміну вкладу сума виплат становитиме 12 000грн.



Алгоритм розв’язання

1. Виділіть комірку, в якій міститься значення суми до виплат, яке обчислене за формулою.

Для заданої таблиці це буде коміркаС16.

2. Меню Сервис à Подбор параметра.

а) Перевірте в діалоговому вікні Подбор параметра, чи задана адреса комірки С16, яка містить значення суми до виплат, в текстовому полі Установить в ячейке.

б) В діалоговому вікні Подбор параметра встановити текстовий курсор в текстове віконце Значение і введіть з клавіатури потрібне значення, у даному прикладі 12000.

в) В діалоговому вікні Подбор параметра встановити текстовий курсор в текстове віконце Изменяя значение ячейки і клацніть на комірці, в якій міститься значення терміну вкладу, у даному прикладі С13.

г) В діалоговому вікні Подбор параметра клацніть на кенопці Ок. 3. Проаналізуйте таблицю. Яке значення терміну вкладу з’явилось в комірці С13?

Знайдене значення терміну вкладу і буде тим шуканим значенням терміну вкладу, при якому значення суми виплат досягне заданого розміру ( у даному разі 12000 грн.).

Завдання 2. Використовуючи функцію Подбор параметра та дані наведеної таблиці, знайдіть, при якому значенні відсоткової ставки сума виплат становитиме 12 000грн.



Зауваження. Розв’язання виконайте самостійно. Використайте наведений для прикладу 1 алгоритм, замінюючи в ньому значення терміну вкладу на значення відсоткової ставки.

Функція "Таблица подстановки"

Ця функція використовується для дослідження впливу одного чи двох параметрів на значення певної величини, яка задана формулою, яка містить ці параметри. Якщо параметр один, то список його значень і список значень досліджуваної величини утворюють таблицю, яка називається таблицею підстановки з однією змінною, що відображає функціональну залежність S=f(P).

Завдання 3. Створити таблицю підстановки, яка відображає вплив відсоткової ставки на суму виплат та коефіцієнт нарощування за заданими в таблиці даними.



Зауваження. Розв’язання виконайте самостійно. Використайте наведений для прикладу 1 алгоритм, замінюючи в ньому значення терміну вкладу на значення відсоткової ставки.

Функція "Таблица подстановки"

Ця функція використовується для дослідження впливу одного чи двох параметрів на значення певної величини, яка задана формулою, яка містить ці параметри. Якщо параметр один, то список його значень і список значень досліджуваної величини утворюють таблицю, яка називається таблицею

підстановки з однією змінною, що відображає функціональну залежність S=f(P).

Завдання 3. Створити таблицю підстановки, яка відображає вплив відсоткової ставки на суму виплат та коефіцієнт нарощування за заданими в таблиці даними.



Алгоритм розв’язання

1. Складіть список значень відсoткової ставки у коміркахD22:D36.

2. Введіть формулу розрахунку суми виплат в коміркуЕ22. =$C$12\*(1+$C$15/100)^ $C$14

3. Введіть формулу розрахунку коефіцієнта нарощування в комірку F22:

=(1+$C$15/100)^ $C$14 .

4. Виділіть діапазон комірокD22:F36 .

5. Меню Данные, клацніть на пункті Таблица подстановки. У діалоговому вікні Таблица подстановки встановіть текстовий курсор у віконціПодставлять значения по строкам в: і клацніть на комірці C15, потім – на кнопці ОК.

6. Проаналізуйте отримані результати - вміст

комірок діапазонуD22:F34.

Завдання 4. Створити таблицю підстановки, яка відображає вплив відсоткової ставки та терміну вкладу на суму виплат за заданими в таблиці даними.



Якщо параметрів два, то список їх значень і список значень досліджуваної величини S=f(P,Т).

Алгоритм

1. Складіть список значень відсoткової ставки в діапазоні комірок Е67:Е82. Об’єднайте комірки і створіть зміст об’єднаних комірок як на рисунку нижче.

2. Введіть формулу розрахунку суми виплат в комірку Е66

=$B$49\*(1+$B$51/100)^ $B$50 .

3. Введіть значення терміну вкладу в діапазон комірок F66:K66. 4. Виділіть діапазон комірокE66:K82.

5. Меню Данные, клацніть на пункті Таблица подстановки. а) У діалоговому вікні Таблица подстановки встановіть

текстовий курсор у віконці Подставлять значения по столбцам в: і клацніть на комірці В50.

б) текстовий курсор у віконці Подставлять значения по строкам в: і клацніть на комірці В51.

6. Проаналізуйте отримані результати - вміст

комірок діапазонуD22:F36.

Приклад інформаційний підхід до моделювання

У банківському ризик-менеджменті широко відома модель Дюрана для розрахунку рейтингу кредитоспроможності позичальника, яка набула поширення в 40-50-і рр. XX ст. На основі власного досвіду Дюран розробив бальну модель для оцінки позичальника за сукупністю його майнових і соціальних параметрів (вік, стать, професія і т. д.). Подолавши певний поріг,

позичальник вважався кредитоспроможним. Ця модель являє собою аналітичну залежність y = f (X), де y - рейтинг, X - набір ознак позичальника. Якщо перед сучасним російським банком постане завдання розрахувати рейтинг позичальника, банк може скористатися моделлю Дюрана. Однак чи буде адекватною для сучасної російської дійсності модель, розроблена в середині минулого століття на Заході? Природно, не буде, так як вона не враховує закономірності між характеристиками російських позичальників (вік, освіта, дохід і т. д.) і дефолтних по кредитах.

Якщо ж банк візьме власні дані по кредитних історій і на їх основі побудує модель, розраховується рейтинг клієнта, то, цілком ймовірно, вона виявиться працездатною. У першому випадку, коли ми брали модель Дюрана, ми використовували аналітичний підхід. У другому - інформаційний; для побудови моделі нам знадобилися дані - кредитні історії позичальників банку.

Практичне заняття № 3.

**Архітектура та методи зберігання даних**

**План**

1. Архітектура бази даних: поняття, визначення

2 Інструменти бізнес-аналітики

3. Сучасний бізнес-аналіз.

Таблиця 3.1

| **Ключові поняття:** | |
| --- | --- |
| - Архітектура  - Бізнес-аналітика  - Business Intelligence | - Tableau Public |

**Питання для обговорення**

Визначення Архітектура бази даних

Рівні архітектури БД

Business Intelligence

**Питання для самоперевірки та контролю**

1. Види БД.

2. Які інструменти бізнес-аналітики ви знаєте?

3. Як працює Business Intelligence?

4. Назвіть переваги програмного забезпечення Business Intelligence 5. В чому полягає сучасний бізнес-аналіз?

Бізнес-аналітика (BI)-це процес прийняття рішення на основі даних, заснований на збірних даних. Це часто використовується менеджерами та керівниками, щоб зробити ефективні ідеї. В результаті до BI завжди відноситься взаємозалежне "Business Analytics" або "Data Analytics". BI включає управління виробництвом, аналітику, прогнозне моделювання, аналіз даних та тексти та т. Д. Наприклад, у бібліотеці тисячі книг. Нам потрібно розподілити їх за жанрами та організувати у різних проходах, щоб люди без праці знаходили книгу. Корпоративна інформація та дані, подібні книзі в бібліотеці. Важливо керувати цією інформацією в аккуратній та доступній структурах. Інструменти BI використовуються для організації вашого інформаційного сховища, відображення візуалізованих даних, побудови аналітичних даних та т. Ін. Д. Тому просто у використанні інструменту бізнес аналітики крайні можливості для бізнес-операцій.

Приклад. Публікаця в Tableau Public

За допомогою функцій перекладання ви можете редагувати та створювати свої аналітичні діаграми без глибокого розуміння структур даних. Це дозволяє людям візуалізувати ідеї та знищувати їх з вашими командами. Оцінка Ринка: 98

Безкоштовна версія: Tableau Public

Розвиток: Облако, SaaS, Інтернет; Установлено - Mac / Windows; Мобільний - Android Native / iOS Native; Особливості BI: спеціальний аналіз; Спеціальний запит; Специальные отчеты; Сравнительный анализ; Бюджетирование і прогнозування; Приборная доска; Аналіз данных; Візуалізація даних; Ключові показники ефективності; OLAP; Показатели ефективності; Прогнозна аналітика; Аналіз рентабельності; Стратегическое планирование; Індикатори тренда / проблеми.

Тут я розкажу вам про те, як опублікувати вашу візуалізацію з Tableau Desktop у Tableau Public.

Після цього, як ви створили свою інформаційну панель, виконали наступні дії: - Сервер, Tableau Public, Сохранить в Tableau Public як

\* Можливо, спочатку вам запропонувати войти у свій профіль у Tableau Таблиця робить аналіз даних інтуїтивно понятним та простим у використанні. Виберіть назву для своєї візуалізації та натисніть «Зберегти» (Сохранить). Далі ви можете переглянути повідомлення у всьому цьому: Це повідомлення означає, що ваше з'єднання з набором даних Sample Superstore тепер встановлено. Tableau Public не може розміщувати живі з'єднання, тому вам потрібно змінити ваше з'єднання у вибір (наприклад, скриншот екрану ваших даних).

Перейдіть до одного з ваших робочих листів та виконайте наступні дії: -Ще раз натисніть праву кнопку «Sample-Superstore» (Вибірка Супермагазин) у верхній частині екрана та виберіть «Extract Data» (Видалення даних).

Натисніть кнопку «Витяг» (Видалення) та перегляньте інформаційну панель, щоб повторити публікацію. Коли закончите, натисніть «Сохранить», ви потрапите на сторінку Tableau Public.



Тут ви бачите, що моя інформаційна панель вже опублікована. Вона неактивна, тому що я автоматично встановлює свої візуалізації у «приховані» (скриті). Ви можете зробити так, або немає, перейти на сторінку настрою. Я робив це тому, що якщо візуалізація не буде скритою, вона з'явиться у лінте діяльності ваших підписників, навіть якщо ви видалите її привітання. Якщо ви налаштували автоматичне скриття, щоб побачити, як його відредагувати, перегляньте його в браузері, а потім, коли ви будете готові, ви можете показати його, натиснувши на значок у маленькій глазі.

Практичне заняття № 4.

**Видобування, перетворення та навантаження (ETL)**

План

1 Процес ETL у сховищі даних

2 Дизайн процесу ETL та підтримка інструментів

3 CDC та віртуалізація даних

Таблиця 4.1

| Ключові поняття: | |
| --- | --- |
| Витяг  завантаження  перетворення  Інтегратор даних | ETL  Сховище даних  Віртуалізація даних |

**Питання для обговорення**

Дорожня карта процесу ETL: видобування, перетворення та навантаження.

Програмні засоби ETL.

Інтеграція даних із Microsoft Dataverse.

Cloud Data Center (CDC).

**Питання для самоперевірки та контролю**

1) В чому полягає ETL?

2) Розкрийте сутність дизайна процесу ETL

3) Інструменти ETL

4) Що таке віртуальний ЦОД Cloud Data Center (CDC)?

Програмні засоби ETL

Самі популярні комерційні програмні засоби, що реалізують функції ETL-підсистем у корпоративному сховищі даних, є:

IBM WebSphere DataStage

Informatica PowerCenter

Інтегратор даних

Oracle Інтегратор даних

SAP BusinessObjects

Сервер інтеграції даних SAS и другие.

До категорій умовно безкоштовних можна віднести:

Oracle Warehouse Builder

Talend Open

Studio Скриптелла и другие.

Що таке віртуальний ЦОД Cloud Data Center (CDC)?

Віртуальний центр обробки даних (скорочений в ЦОД) - це пул ресурсів хмарної інфраструктури, розроблений спеціально для нужд корпоративного бізнесу. У таких рішеннях сьогодні зацікавлені прогресивні компанії.

ВЦОД фізично розміщується у традиційних центральних процесах обробки даних та надається провайдером в оренду. Проще говорячи, всі використовувані компактні пристрої об'єднуються в єдиній системі, яка може бути доступною в будь -який час. При цьому разом використовуються основні ресурси - процесор, пам’ять, дискове просторове та пропускна здатність. Всё, як і в фізичному КОДа - з серверами, кластерами зберігання та множиною сетевих компонентів, - тільки у віртуалізованому просторі.

Управління вЦОДом здійснюється з консолі з будь -якої точки світу, тому адміністратор може працювати удалено. Спектр вирішених завдань сопоставлений або навіть більш широкий, що використовується традиційним ЦОДам. Для цього автономність від провайдера та рівень захисту від зовнішніх скачкових навантажень в інших ресурсах у хмарі - вище, якщо віртуальний КОД розміщений у ізольованій мережі vLAN.

Усі інші організації використовують архітектуру вЦОД, щоб спростувати процес виділення ІТ-ресурсів, оптимізувати використання

емкости, знизити ризики та максимізувати виробництво. Ця технологія дозволяє ІТ-спеціалістам перетворюватись у «стратегічні партнери» для бізнесу, сприяти змінам у відповідь на зміну риночного ландшафту, і не залишати статей обов'язковими капітальними витратами.

**Інтеграція даних із Microsoft Dataverse**

Інтегратор даних (для адміністраторів) – це служба інтеграції від точки до точки, яка використовується для інтеграції даних в Dataverse. Підтримує інтегрування даних між програмами Finance and Operations та Dataverse. Вона також підтримує інтегрування даних у програми Finance and Operations та Dynamics 365 Sales. Ця служба доступна для широкого загалу з липня 2017 року.

Платформа інтегратора даних

Інтегратор даних (для адміністраторів) складається з платформи інтеграції даних, готових шаблонів, наданих нашими робочими групами програми (наприклад, програми Finance and Operations та Dynamics 365 Sales) і користувацьких шаблонів, створених нашими клієнтами та партнерами. Ми створили прикладно-агностичну платформу, яка може масштабувати різні джерела. Її суть полягає в тому, що ви створюєте підключення (для інтеграції кінцевих точок), вибираєте один зі спеціальних шаблонів із наперед визначеними зіставленнями (які можна додатково налаштувати), а також створюєте та виконуєте проект інтеграції даних.

Шаблони інтеграції служать проектом із наперед визначеними сутностями та зіставленнями полів, щоб активувати потік даних від джерела до призначення. Крім того, вони надають можливість трансформувати дані, перш ніж імпортувати їх. У багатьох випадках схеми між вихідними та цільовими програмами можуть дуже відрізнятися, а шаблон із наперед визначеними сутностями і зіставленнями полів слугує прекрасною відправною точкою для проекту інтеграції (рис. 4.1).



Рис. 4.1. Платформа інтегратора даних

Процедура налаштування проекту інтеграції даних

Потрібно виконати три основні кроки:

Створення підключення (надання облікових даних для джерел даних). Створення набору підключень (визначення середовищ для підключень, створених на попередньому кроці).

Створення проекту інтеграції даних на основі шаблону (створення або використання попередньо визначених зіставлень принаймні для однієї сутності).

Створивши проект інтеграції ви отримуєте можливість запустити проект вручну, а також налаштувати оновлення за розкладом на майбутнє. Далі в цій статті пояснюються ці три кроки.

Процедура створення підключення

Перш ніж створити проект інтеграції даних, необхідно підготувати підключення для кожної системи, з якою ви плануєте працювати на порталі Microsoft Power Apps. Розглядайте ці підключення як точки інтеграції. Створення підключення

Перейти до Power Apps.

У розділі "Дані" виберіть Підключення, а потім виберіть Нове підключення. Ви можете вибрати підключення зі списку підключень або знайти своє підключення.

Вибравши своє підключення, натисніть кнопку Створити. Відобразиться запит на введення облікових даних.

Коли ви надасте облікові дані, підключення буде додано до списку підключень.

Примітка

Переконайтеся, що обліковий запис, який було задано для кожного підключення, має доступ до сутностей для відповідних програм. Крім того, бізнес-партнери для кожного підключення можуть бути в різних клієнтах. Процедура створення набору підключень

Набори підключень – це набори з двох підключень, середовищ для підключень, відомостей про зіставлення організації та ключів інтеграції, які можна повторно використовувати з проектами. Ви можете почати використовувати набір підключень для розробки, а потім переключитися на інший для виробництва. Одна із ключових складових інформації, яка зберігається з набором підключення, — це зіставлення організаційних

одиниць, наприклад, зіставлення між юридичною особою (або компанією) Finance and Operations і організацією Dynamics 365 Sales або підрозділами. У наборі підключень можна зберігати кілька зіставлень організації. Створення набору підключень

Перейдіть до центру адміністрування Power Apps.

Перейдіть на вкладку Інтеграція даних в області переходів ліворуч. Перейдіть на вкладку Набори підключень, а потім виберіть Новий набір підключень.

Укажіть ім’я для набору підключень.

Виберіть раніше створені підключення, а потім виберіть відповідне середовище.

Повторіть ці кроки, вибравши наступне підключення (вважайте їх джерелом і призначенням без визначеного порядку).

Укажіть зіставлення організації із підрозділом (якщо системи Finance and Operations і Sales інтегровані).

Примітка

Для кожного набору підключень можна задати кілька зіставлень. Заповнивши всі поля, виберіть Створити.

Новий набір підключень, який ви щойно створили, відобразиться на сторінці списку "Набори підключень".

Набір підключень готовий до використання в різних проектах інтеграції. Процедура створення проекту інтеграції даних

Проекти дають змогу використовувати потік даних між системами. Проект містить зіставлення для однієї або кількох сутностей. Зіставлення вказують, які поля зіставляються з іншими полями.

Створення проекту інтеграції даних

Перейдіть до центру адміністрування Power Apps.

Перейдіть на вкладку Інтеграція даних в області переходів ліворуч. На вкладці Проекти виберіть Новий проект у верхньому правому куті.

Укажіть ім’я для проекту інтеграції.

Виберіть один із доступних шаблонів (або створіть власний шаблон). У цьому випадку ми переносимо сутність «Продукти» з Finance and Operations до Sales. .

Натисніть кнопку Далі та виберіть створений раніше набір підключень (або створіть новий набір підключень).

Переконайтеся, що вибрано правильний набір підключень, підтвердивши імена підключення та середовища.

Виберіть Далі, а потім виберіть юридичну особу для зіставлень організаційних одиниць.

Перегляньте та прийміть повідомлення про конфіденційність і надайте згоду на наступному екрані.

Приступіть до створення проекту, а потім запустіть проект, який у свою черзі виконає проект.

На цьому екрані можна помітити кілька вкладок – Планування і Журнал виконання, а також деякі кнопки – Додати завдання, Оновити сутності і Розширений запит – які буде описано далі.

Журнал виконання

Журнал виконання показує історію всіх виконань проекту з іменем проекту, позначкою часу виконання проекту та станом виконання разом з кількістю upsert і/або помилок.

Приклад журналу виконання проекту.

Приклад успішного виконання, зі станом "Завершено" з # upsert. (Додавання/оновлення – це логіка, що полягає або в оновленні запису, якщо він вже існує, або у вставленні нового.)

Невдалі виконання можна деталізувати, щоб побачити причину. Нижче наведено приклад невдачі з помилками перевірки проекту. У цьому разі помилка перевірки проекту пов’язана з відсутніми вихідними полями в зіставленнях сутності.

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.

Якщо виконання проекту перебуває в стані «ПОМИЛКА», то він повторить виконання під час наступного запланованого запуску.

Якщо виконання проекту перебуває в стані «ПОПЕРЕДЖЕННЯ», то вам потрібно буде вирішити проблеми в джерелі. Він повторить виконання під час наступного запланованого запуску.

У будь-якому разі також можна вибрати «повторне виконання» вручну. Процедура налаштування оновлення за розкладом

Ми підтримуємо два типи виконань/записів сьогодні:

Записи вручну (виконання та оновлення проекту вручну)

Записи за розкладом (автоматичне оновлення)

Після створення проекту інтеграції ви отримуєте можливість запустити його вручну або налаштувати записи за розкладом, що дає змогу налаштувати автоматичне оновлення для своїх проектів.

Налаштування записів за розкладом

Перейдіть до центру адміністрування Power Apps.

Ви можете запланувати проекти двома різними способами. Виберіть цей проект і перейдіть на вкладку Планування або запустіть планувальник на сторінці списку проектів, натиснувши три крапки поруч із назвою проекту.

Виберіть Повторювати кожн. та, заповнивши всі поля, виберіть Зберегти розклад.

Ви можете встановити частоту на 1 хвилину або повторювати запис певну кількість годин, днів, тижнів або місяців. Зверніть увагу на те, що наступне оновлення не почнеться, доки не буде завершено виконання попереднього

завдання проекту.

Також зверніть увагу, що в розділі "Сповіщення" можна вибрати сповіщення електронною поштою, які оповіщатимуть вас про виконання завдань, що

завершилися з попередженнями та/або не були завершені через помилки. Ви також можете вказати кількох одержувачів, зокрема групи, через кому. 

Практичне заняття № 5

**Надання інформації (звітування, інформаційні панелі)**

План

Таблиця 5.1

| Ключові поняття: | |
| --- | --- |
| - Звітування  - Інформаційна панель | - Google Analytics  - Інформація |

**Питання для обговорення**

Проектування та впровадження інформаційних панелей.

Типи звітів, які повільно змінюються.

Google Analytics.

Створення й налаштування інформаційних панелей.

**Питання для самоперевірки та контролю**

Інформаційна панель – це?

Типи звітів, які повільно змінюються

Які типи інформаційних панелей вам відомі?

Створення й налаштування інформаційних панелей

Налаштовуйте свої інформаційні панелі й віджети так, щоб оглядати обліковий запис у зручний для себе спосіб.

Кожне представлення даних в Analytics за умовчанням включає "Мою інформаційну панель". Такі інформаційні панелі містять набори всіх найважливіших параметрів, проте головна їх перевага полягає в тому, що панелі можна створювати й налаштовувати відповідно до своїх потреб.

Інформаційна панель містить один або кілька віджетів (не більше 12), за допомогою яких можна здійснювати огляд найважливіших параметрів і показників. У цій статті розповідається, як створити й налаштувати нову інформаційну панель (або змінити наявні).

Створення інформаційної панелі

Щоб створити інформаційну панель:

Увійдіть у Google Analytics.

Відкрийте потрібне представлення даних.

Відкрийте Звіти.

Натисніть НАЛАШТУВАННЯ > Інформаційні панелі.

Натисніть +Нова інформаційна панель.

У діалоговому вікні Створення інформаційної панелі виберіть Чисте полотно (без віджетів) або Інформаційна панель для початківців (набір віджетів за умовчанням). Можна також імпортувати конфігурацію інформаційної панелі з Галереї рішень.

Меню створення інформаційної панелі

Дайте інформаційній панелі описову назву й натисніть Створити інформаційну панель.

Додавання віджетів на інформаційну панель

Інформаційна панель може містити перелічені нижче типи віджетів. Показник – просте числове представлення певного вибраного показника.

Часова шкала – графік змінення вибраного показника за певний період часу. Можна порівнювати основні та другорядні показники. Географічна карта – карта вибраного регіону з нанесеними на неї показниками. Наведіть курсор миші на карту, щоб побачити фактичні значення показників.

Таблиця – не більше двох показників, які відповідають вибраному параметру, представлені у формі таблиці.

Кругова діаграма – секторна діаграма для вибраного показника, згрупована за параметром. Наведіть курсор миші на сектор, щоб побачити фактичні значення показників.

Гістограма – стовпчикова діаграма для вибраного показника, згрупована за щонайбільше двома параметрами. Наведіть курсор миші на стовпець, щоб побачити фактичні значення показників.

Редактор віджетів

Стандартні віджети й віджети реального часу

Деякі з наявних віджетів можуть показувати дані в реальному часі. Показники в цих віджетах оновлюються автоматично (на відміну від стандартних, які оновлюються лише під час відкривання або перезавантаження інформаційної панелі).

Залежно від типу, віджети реального часу можуть показувати лише дані про активних користувачів і переглянуті сторінки. Нижче наведено доступні типи віджетів реального часу.

Лічильник – показує кількість активних користувачів на сайті. За потреби можна групувати цих користувачів за вибраним параметром. Часова шкала – графік переглядів сторінок на вашому сайті за останні 30–60 хвилин.

Географічна карта – карта територій перебування активних користувачів.

Таблиця – зведені в таблицю дані про активних користувачів за щонайбільше трьома вибраними параметрами.

Щоб додати віджет на інформаційну панель:

Створіть нову інформаційну панель, вибравши Чисте полотно, або натисніть + Додати віджет у наявній інформаційній панелі, щоб відкрити редактор віджетів.

Виберіть тип віджета.

Налаштуйте параметри, показники й інші опції віджета (їх доступність залежить від типу вибраного віджета). Щоб знайти потрібний параметр або показник, прокрутіть список чи скористайтеся вікном пошуку.

Показ даних у віджеті можна обмежити, натиснувши посилання Додати фільтр. За допомогою фільтрів у визначений параметр можна включати або виключати дані, які відповідають критеріям фільтрування. У визначення фільтра можна додавати кілька рядків, проте всі вони мають виконуватися, щоб фільтрування відбулося.

Принцип дії фільтрів інформаційної панелі та звітів відрізняється від фільтрів представлення даних. Фільтри представлення даних назавжди змінюють відображувані дані, тоді як фільтри звітів та інформаційної панелі лише обмежують показ даних у звіті або на інформаційній панелі. Фільтри інформаційної панелі працюють лише на тій панелі, на якій їх створено.

Ви можете пов’язати віджет зі звітом або URL-адресою. У такому випадку назва віджета стане посиланням, за яким можна перейти до

відповідного звіту чи на веб-сторінку. Щоб зробити посилання на звіт, почніть вводити його назву. Analytics автоматично доповнить її, пов’язавши віджет із наявним звітом Можна також скопіювати та вставити URL-адресу звіту в це поле.

Введіть Назву віджета або виберіть запропоновану.

Натисніть Зберегти.

Редагування віджета

Щоб Редагувати наявний віджет, наведіть курсор миші на його назву й натисніть значок олівця.

Щоб видалити віджет, наведіть курсор на його назву й натисніть значок Закрити (X).

Редагування віджета Примітка. Видалення віджета неможливо відмінити.

Клонування віджета

Створити точну копію віджета можна за допомогою посилання Копіювати віджет. Це зручно, якщо потрібно використати один віджет як основу для іншого.

нті.

Клонування віджета

Додавання сегментів на інформаційну панель

Сегменти на інформаційну панель можна додавати так само, як у звітах Analytics. Це дає змогу порівнювати та протиставляти показники, отримані з різних сеансів або груп користувачів.

Щоб змінити наявний сегмент, натисніть його мітку вгорі інформаційної панелі, а щоб додати новий – порожню мітку +Додати сегмент. Докладніше про сегменти.

Налаштування макета інформаційної панелі

Натисніть посилання Налаштувати інформаційну панель, щоб змінити порядок розташування й зовнішній вигляд віджетів на сторінці

Як створювати та редагувати інформаційні панелі

На інформаційних панелях можна переглядати зведену статистику ефективності всіх оголошень облікового запису.

Маючи доступ до найважливіших даних, ви можете швидко виявляти проблеми та можливості для розвитку бізнесу.

Дані на інформаційних панелях і формат їх показу легко редагуються. Можна також ділитися даними, завантажувати їх і додавати примітки.

Примітка. Інформаційна панель може містити щонайбільше 10 карток. Коли їх кількість буде максимальною, ви вже не зможете додавати інші примітки, таблиці, діаграми або картки показників.

Вказівки

Як створити інформаційну панель

Для цього:

Увійдіть в обліковий запис Google Ads.

Угорі сторінки натисніть значок звітів .

Виберіть Інформаційні панелі.

Натисніть кнопку з плюсом .

Як додати картку показників

Інформаційні панелі складаються з карток, які містять вибрані вами показники ефективності. Вставивши картку, ви можете змінити її розмір, потягнувши за краї, або перетягнути її в інше місце на інформаційній панелі. Щоб вставити картку показників:

Увійдіть в обліковий запис Google Ads.

Натисніть значок звітів і виберіть Інформаційні панелі.

Натисніть кнопку з плюсом , щоб створити нову інформаційну панель. Або виберіть наявну інформаційну панель зі списку й натисніть Редагувати на наступній сторінці.

Натисніть значок картки показників над інформаційною панеллю. Виберіть потрібні показники й часовий діапазон і натисніть Далі. Введіть потрібний фільтр.

Ви можете налаштовувати дані карток показників, ставлячи прапорці Міні-діаграма, % змін і Абсолютна зміна.

Завершивши, натисніть Далі.

Введіть заголовок і опис (необов'язково) для таблиці або діаграми. Натисніть Додати.

Як додати таблицю або діаграму

У редакторі звітів можна створити нову таблицю чи діаграму або додати наявну зі збережених.

Як створити таблицю або діаграму

Над інформаційною панеллю натисніть значок таблиці або діаграми .

Виберіть Створити нову таблицю або Створити нову діаграму. Натисніть Далі.

Відкриється редактор звітів, де ви зможете створити таблицю чи діаграму.

Завершивши, натисніть Додати на інформаційну панель.

Дайте назву звіту й натисніть Зберегти.

Введіть опис (необов'язково) і натисніть Додати.

Як додати збережену таблицю або діаграму

Над інформаційною панеллю натисніть значок таблиці або діаграми .

Натисніть Вибрати збережену таблицю з редактора звітів або Вибрати збережену діаграму з редактора звітів.

У спадному меню виберіть потрібну таблицю або діаграму. Натисніть Далі.

Введіть заголовок і опис (необов'язково) для таблиці або діаграми. Натисніть Додати.

Як додати примітку

Додавши картки показників, можна також створити примітку для інформаційної панелі. Це зручно, якщо ви плануєте надати до інформаційної панелі доступ іншим користувачам або завантажити її.

Щоб додати примітку, натисніть значок над інформаційною панеллю. Введіть текст примітки й натисніть Додати.

Як надати доступ до інформаційної панелі

Інформаційні панелі можуть переглядати всі, хто користується вашим обліковим записом Google Ads, окрім людей із доступом лише до електронної пошти. Ви можете надсилати звіти іншим користувачам електронною поштою. Щоб надати доступ до інформаційної панелі:

Завершіть редагувати інформаційну панель і натисніть Зберегти.

Над інформаційною панеллю натисніть значок електронного листа .

Введіть електронні адреси користувачів, яким потрібно надати доступ до інформаційної панелі.

Додайте примітку (необов’язково).

Виберіть частоту надсилання даних інформаційної панелі.

Натисніть Поділитися.

Як завантажити дані

Щоб завантажити відомості з інформаційної панелі, натисніть значок . Дані панелі завантажаться у форматі .pdf.

Як переглянути історичні дані

Дані на інформаційній панелі відображаються в межах часових проміжків, які ви вказали для кожної картки показників. Можна також змінити загальну дату панелі, щоб переглянути дані за певний день.

Наприклад, на вашій інформаційній панелі є 2 картки показників: одна з даними про CTR усіх пошукових кампаній за останні 2 тижні, а інша з даними про покази медійних кампаній за останні 4 тижні. Якщо змінити загальну дату панелі на "3 місяці тому", ви побачите дані CTR за 2 тижні та статистику показів за 4 тижні до вказаної дати.

Щоб змінити дату інформаційної панелі, натисніть поле дати й виберіть потрібну.

Практичне заняття № 6

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

План

1. Що таке аналітичний життєвий цикл.

2. Сучасна архітектура даних. Машинне навчання.

3. Сутність класифікації. Знайомство з методами кластеризації даних Таблиця 6.1

| Ключові поняття: | |
| --- | --- |
| - Життєвий цикл | - Кластеризація |

| - Архітектура даних  - Машинне навчання | - Класифікація |
| --- | --- |

**Питання для обговорення**

1. Системи бізнес -аналітики.

2. Життєвий цикл аналітики.

3. Модель управління даними для BI.

4. Методи кластеризації даних.

5. Кластерний аналіз.

**Питання для самоперевірки та контролю**

1. В яких сферах бізнесу використовуються системи бізнес -аналітики? 2. Життєвий цикл аналітики – це?

3. Назвіть етапи життєвого циклу аналітики даних

4. В чому різниця між сучасною архітектурою даних та традиційною? 5. Назвіть проблеми управління даними

6. В чому полягає модель управління даними для BI наступного покоління?

7. Яким чином відбувається покращення BI за допомогою AI? 8. Назвіть методами кластеризації даних

ПРИКЛАД ЖИТТЯ ЦИКЛІДАЦІЇ АНАЛІТИКИ ДАНИХ

Розглянемо приклад мережі роздрібних магазинів, яка хоче оптимізувати ціни на свою продукцію для збільшення свого доходу. Мережа магазинів налічує тисячі товарів у сотнях торгових точок, що робить це дуже складним сценарієм. Визначивши мету мережі магазинів, ви знайдете потрібні дані, підготуєте їх та пройдете процес життєвого циклу Data Analytics. Ви спостерігаєте різні типи клієнтів, наприклад, звичайних клієнтів та клієнтів, таких як підрядники, які купують оптом. На вашу думку, поводження з різними типами клієнтів може дати вам рішення. Однак у вас недостатньо інформації про це, і вам потрібно обговорити це з командою клієнтів. У цьому випадку вам потрібно отримати визначення, знайти дані та провести перевірку гіпотез, щоб перевірити, чи різні типи клієнтів впливають на результати моделі та отримати правильний результат. Після того, як ви переконаєтесь у

результатах моделі, ви зможете розгорнути її, інтегрувати її у бізнес, і ви налаштуєтесь розгортати ціни, які, на вашу думку, є найбільш оптимальними у торгових точках магазину.

ЯК ЗРОБИТИ КЛАСТЕРНОГО АНАЛІЗУ В EXCEL

Для приклада шість об'єктів спостереження. Кожен має два характерних параметра (рис 6.1)



Рис. 6.1. Приклад

У якості розстановки між об'єктами беремо евклідове стан. Формула розрахунку (рис. 6.2):

Рис. 6.2. Формула

Розраховані дані розміщуємо в матриці відстаней.

Найближчими один до одного об'єктами є об'єкти 4 і 5. Отже, їх можна об'єднати в одну групу - при формуванні нової матриці залишаємо найменше значення (рис 6.3.).



Рис. 6.3. Об’єднання

З нової матриці видно, що можна об'єднати в один кластер об'єкти [4, 5] і 6 (як найбільш близькі один до одного за значеннями). Ми залишаємо найменше значення і формуємо нову матрицю (рис. 6.4):



6.1. Матриця з найменшим значенням

Об'єкти 1 і 2 можна об'єднати в один кластер (як найбільш близькі з наявних). Вибираємо найменше значення і формуємо нову матрицю відстаней. В результаті отримуємо три кластери (рис. 6.5.):



Рис. 6.5. Кластери.

Найближчі об'єкти - 1, 2 і 3. Об'єднаємо їх.



Рис. 6.6. Найближчий сусід

Ми провели кластерний аналіз за методом «найближчого сусіда» (рис. 6.6). В результаті отримано два кластери, відстань між якими - 7,07. Величезне значення має кластерний аналіз в економічному аналізі. Інструмент дозволяє виокремлювати з величезної сукупності періоди, де значення відповідних параметрів максимально близькі і де динаміка найбільш схожа. Для дослідження, наприклад, товарної і господарської кон'юнктури цей метод відмінно підходить.

Практичне заняття № 7

**Моделювання великих даних**

План

1. Big data: які дані вважаються великими

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.



2. Big Data в маркетингу

3 Важливість моделювання даних у світі великих даних

4. Поради створення ефективних моделей великих даних

Таблиця 7.1

| Ключові поняття: | |
| --- | --- |
| - Big Data  - Моделювання  - Постачальники інфраструктури | - Датамайнери  - Системні інтегратори  - Споживачі |

**Питання для обговорення**

1) Засоби роботи з Великими даними.

2) Основи використання Big Data в маркетингу.

3) Ефективні моделі великих даних.

4) Прийняття управлінських рішень з Big Data.

5) Google Analytics

**Питання для самоперевірки та контролю**

**1.** Дайте визначення Big data – це?

**2.** Назвіть основні показники Big Data.

**3.** Які вигоди використання ІТ в маркетингу?

**4.** Моделювання даних – це?

**5.** В чому важливість моделювання даних у світі великих даних?

Приклади Big Data

Нью-Йоркська фондова біржа щодня генерує 1 терабайт даних про торги за минулу сесію.

Соціальні медіа: статистика показує, що в бази даних Facebook щодня завантажується 500 терабайт нових даних, генеруються в основному через завантажень фото та відео на сервери соціальної мережі, обміну повідомленнями, коментарями під постами і так далі.

Реактивний двигун генерує 10 терабайт даних кожні 30 хвилин під час польоту. Так як щодня відбуваються тисячі перельотів, то обсяг даних досягає петабайт.

Приклад роботи в Google Analytics (русифікована версія)

Початок.

Перейдіть на головну сторінку сервісу і створіть новий обліковий запис Google або увійдіть в існуючий (рис. 7.1, 7.2).

Рис. 7.1. Головн сторінка сервісу Google

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.



Рис. 7.1. Головн сторінка сервісу Google, вхід

Ви можете отримувати статистику веб-сайту або мобільний додатки. У будь-якому випадку, вам потрібно вказати: Назва аккаунта, щоб не заплутатися, якщо їх багато. На одному акаунті можна відстежувати відразу кілька сайтів. В рамках одного облікового запису Google можна створити не більше 100 акаунтів (рис. 7.3).

Рис. 7.3. Новий акаунт

Назва веб-сайту (додатки), якщо їх у вас багато, для спрощення навігації. URL для веб-сайтів. Галузь. Подумайте, яка категорія вам ближче. Надалі ви зможете порівнювати свої показники з середніми по галузі. Більше категорія ні на що не впливає.

Часовий пояс, щоб звіти приходили в нормальний час, а не о 3 годині ночі. Тонкі налаштування UTM для зв'язку з Google Ads та іншою рекламою.

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.



Рис. 7.4. Налаштування ресурсів

Налаштування спільного використання даних - ви можете дозволити або заборонити Google використовувати вашу статистику в своїх цілях. Сервіс збирає інформацію анонімно, так що не переживайте: ніхто не дізнається вашу конверсію і відвідуваність.



Рис. 7.5. Налаштування спільного використання даних

Коли все буде готово, натисніть на «Отримати ідентифікатор відстеження».

Щоб Google Analytics могла збирати дані про ваш сайт, потрібно встановити лічильник. Це скрипт, який завантажується на кожній сторінці, збирає статистику і відправляє її в Google. Є кілька способів підключити лічильник: Вставити скрипт в код сайту. Якщо хочете, щоб аналітика працювала на всіх сторінках, помістіть лічильник в Header або Footer. Найкраще помістіть його в Header між тегами <head> і </ head>. Скопіюйте код з розділу «Відстеження веб-сайтів».

Через Диспетчер тегів Google ви підключаєте скрипт контейнера до сайту. У цей контейнер можна помістити відразу кілька скриптів. В результаті вам не потрібно буде кожен раз йти в код і щось там міняти. Відредагували контейнер, і зміни автоматично вступили в силу (рис 7.6). Рис. 7.6. Робоча область

Через плагін для WordPress. Вам достатньо вказати ID свого кабінету в Google Analytics (рис. 7.7).

Commission cannot be held responsible for any use which may be made of the information contained there in.

Рис. 7.7. WordPress

Тепер зайдіть в налаштування плагіна і скопіюйте ідентифікатор відстеження (вказано на сайті Google Analytics) в поле Google Analytics ID. Не забудьте зберегти зміни. Плагін від MonsterInsights дозволяє підключатися через акаунт Google. Вам потрібно лише авторизуватися. Він сам знайде ресурси, які ви відстежуєте.

Через ваш конструктора сайту. Багатьом з них, наприклад, Tilda, досить вказати тільки ідентифікатор лічильника Google Analytics для його автоматичної прив'язки до сайту. Шукайте дані можливості в налаштуваннях аналітики для ресурсу. Виберіть один із способів і підключіть Google Analytics до свого сайту. Я раджу використовувати «Диспетчер тегів». Напевно ви користуєтеся «Яндекс.Метрика» або ще чимось (картами кліків, наприклад). Якщо ви випадково закрили сторінку і не можете знайти код відстеження та свій ID, відкрийте розділ «Адміністратор» → «Ресурс» → «Відстеження» → «Код відстеження».

Статистика може відображатися не відразу. Не хвилюйтесь. Моя, наприклад, довантажити через кілька годин.

Огляд інтерфейсу.

В меню Google Analytics є 4 розділи:

Головна сторінка;

звіти;

мої звіти;

адміністратор.

Звіти.

Основний розділ, з яким ви будете працювати найчастіше. Тут можна подивитися будь-яку статистику, налаштувати цілі для підрахунку конверсії і створити зведення. Зліва знаходиться меню для навігації по різним звітів.

Клацніть по розділу, щоб розкрити його і побачити список підрозділів (у яких теж можуть бути підрозділи).

Основний розділ, з яким ви будете працювати найчастіше. Тут можна подивитися будь-яку статистику, налаштувати цілі для підрахунку конверсії і створити зведення. Зліва знаходиться меню для навігації по різним звітів. Клацніть по розділу, щоб розкрити його і побачити список підрозділів (у яких теж можуть бути підрозділи).

Рис. 7.8. Список звітів, які можливо продивитись

Або введіть назву звіту в пошуку.

Вгорі сторінки знаходиться меню експорту. Ви можете зробити звіт доступним, експортувати його у вигляді таблиці або PDF, додати в зведення (багато різних звітів на одній сторінці) (рис 7.9.).

Навчання і підказки є майже на кожній сторінці з звітами. Підказки слушні, зрозумілі і написані російською мовою. Раджу їх читати.

Більшу частину сторінки займають самі звіти (рис. 7.10)

Рис. 7.10 Звіт активні користувачі

Більш детальна інструкція з роботи в Google Аналітика за посиланням:

Практичне заняття № 8

**Архітектура та розгортання**

План:

1. Стиль архітектури великих даних

2. Еволюція моделей розгортання в епоху великих даних

3. Моделі розгортання в хмарі великих даних

Таблиця 8.1

| Ключові поняття: | |
| --- | --- |
| - Архітектура  - Розгортання | - Хмари  - Еволюція моделей |

**Питання для обговорення**

Що таке лямбда-архітектура.

Архітектура W інфраструктура платформи Big Data.

Apache Spark

**Питання для самоперевірки та контролю**

Для чого призначена архітектура великих даних?

Розкрийте стиль архітектури великих даних.

Еволюція моделей розгортання в епоху великих даних

Назвіть моделі розгортання в хмарі великих даних

Які ключові моделі хмар важливі для обговорення?

Приклад. Архітектура W інфраструктура платформи Big Data. Візьмемо Функціональні вимоги до платформи Big Data і для наочності помістимо їх в матрицю впливу архітектури та інфраструктури (рис. 8.1).

Рис. 8.1. Матриця впливу на виконання вимог

Дивлячись на матрицю, здається, що все в порядку: багато вимог залежать від обраної архітектури платформи Big Data, багато вимог залежать від інфраструктури, на якій платформа буде розгорнута. Нагадаю, одне з головних заявлених переваг платформи Hadoop є «commodity hardware», а, судячи з матриці, так і не скажеш.

Подивимося, що нам дає архітектура Hadoop на прикладі референсной архітектури розміщення кластера в ЦОД від Cloudera (рис. 8.2). У кластері присутні вузли різного призначення:

Name Node - це «Master» вузол кластера, найбільш критичний вузол і, за сумісництвом, єдина точка відмови, зберігає всю мета-інформацію про файлову систему кластера і підтримує її працездатність. Немає Name Node - кластер перетворюється в купу непотрібного заліза. На ньому, напевно, можна буде Майні криптовалюта, але для цього є більш ефективні рішення. Зменшити ризики, пов'язані з його втратою, допоможе Standby Name Node. Природно, розташовуватися вони повинні в різних стійках і бути підключені до різних комутаторів.

Рис. 8.2 Архітектура розгортання вузлів кластера Hadoop в ЦОД. Management Node - це вузли, що дозволяють працювати базовим сервісів

кластера також в відмовостійкості режимі. Таких вузлів може бути досить багато, відмова одного з них не вплине сильно на роботу кластера. Data / HBase Node - це вузли, на яких вже безпосередньо зберігаються і обробляються дані. Це найчисленніший тип вузлів кластера. Вони рівномірно «розмазуються» по стійках і генерують високий трафік (особливо в процесах реплікації).

Edge / DS Node - строго кажучи, не члени кластера. Ці вузли потрібні для завантаження даних в кластер / вивантаження з кластера, а також для обробки даних у випадках, коли паралельна обробка на кластері неможлива. У ряді випадків ці вузли містять GPU для виконання завдань навчання моделей Data Science, на них розгортається ETL / ELT інструментарій, сервера додатків, моніторингу, перевірки якості і т.д. На цих вузлах працюють і користувачі.

Kafka Node - вузли розподіленого брокера повідомлень Kafka, використовуються в задачах потокової обробки з архітектурою Pub / Sub. Вузли також рівномірно «розмазуються» по стійках.

Крім того, в кожній стійці є дубльовані ToR-комутатори, а кожен вузол кластера по 2x10Gb інтерфейсів з'єднаний з обома комутаторами. У свою чергу, ToR комутатори з'єднуються з комутаторами рівня Spine і найбільш

інтенсивний трафік, як видно зі схеми, переміщується у напрямку «Схід Захід», тобто не покидає межі ЦОД-а. На рівні вище знаходяться комутатори ядра мережі.

Така архітектура допомагає нам виконати вимоги:

• Висока доступність

• Отказоустойчивость

• Надійність

• Горизонтальний масштабованість

• Висока продуктивність.

Apache Spark - це Big Data фреймворк з відкритим вихідним кодом для розподіленої пакетної і потокової обробки неструктурованих і слабоструктурованих даних, що входить в екосистему проектів Hadoop.

Класичний MapReduce, Apache компонент Hadoop для обробки даних, проводить обчислення в два етапи:

1. Map, коли головний вузол кластера (master) розподіляє завдання по робочих вузлів (node).

2. Reduce, коли дані згортаються і передаються назад на головний вузол, формуючи остаточний результат обчислень.

Поки всі процеси етапу Map не закінчаться, процеси Reduce не почнуться. При цьому всі операції проходять по циклу читання-запис з жорсткого диска. Це обумовлює затримки в обробці інформації. Таким чином, технологія MapReduce добре підходить для задач розподілених обчислень в пакетному режимі, але через затримки (latency) не може використовуватися для потокової обробки в режимі реального часу. Для вирішення цієї проблеми було створено Apache Spark і інші Big Data фреймворки розподіленої потокової обробки (Storm, Samza, Flink).

На відміну від класичного обробника ядра Apache Hadoop c дворівневої концепцією MapReduce на базі дискового сховища, Spark використовує спеціалізовані примітиви для рекуррентной обробки в оперативній пам'яті. Завдяки цьому багато обчислювальні завдання реалізуються в Спарк значно швидше. Наприклад, можливість багаторазового доступу до завантажених в пам'ять призначених для користувача даних дозволяє ефективно працювати з алгоритмами машинного навчання (Machine Learning) (рис. 8.3).



Порівняння Apache Hadoop і Spark

Як влаштований apache spark: архітектура і принцип роботи Спарк складається з наступних компонентів:

Ядро (Core);

SQL - інструмент для аналітичної обробки даних за допомогою SQL запитів;

Streaming - надбудова для обробки потокових даних, про яку докладно ми розповідали тут і тут;

MLlib - набір бібліотек машинного навчання;

GraphX - модуль розподіленої обробки графів.

Spark може працювати як в середовищі кластера Hadoop під керуванням YARN, так і без компонентів ядра хадуп, наприклад, на базі системи управління кластером Mesos. Спарк підтримує кілька популярних розподілених систем зберігання даних (HDFS, OpenStack Swift, Cassandra, Amazon S3) і мов програмування (Java, Scala, Python, R), надаючи для них API

інтерфейси (рис. 8.4).



Рис. 8.4. Компоненти Спарк

Справедливості заради варто відзначити, що Spark Streaming, на відміну від, наприклад, Apache Storm, Flink або Samza, і не виконує жодних потоки Big Data цілком. Замість цього реалізується мікропакетний підхід (micro-batch), коли потік даних розбивається на невеликі пакети тимчасових інтервалів. Абстракція Spark для потоку називається DStream (discretized stream, Дискретизований потік) і являє собою мікро-пакет, що містить кілька відмовостійких розподілених датасета, RDD (resilient distributed dataset).

Саме RDD є основним обчислювальним примітивом Спарк, над яким можна робити паралельні обчислення і перетворення за допомогою вбудованих і довільних функцій, в тому числі за допомогою тимчасових вікон (window-based operations). Детальніше про тимчасові вікна ми розповідали тут на прикладі Apache Kafka Streams (рис. 8.5).

Принцип роботи спарк де і як використовується apache spark . Завдяки наявності різнопрофільних інструментів для аналітичної

обробки даних «на льоту» (SQL, Streaming, MLLib, GraphX), Спарк активно використовується в системах інтернету речей (Internet of Things, IoT) на стороні IoT-платформ, а також в різних бізнес-додатках, в т.ч. на базі методів Machine Learning. Наприклад, Спарк застосовується для прогнозування відтоку клієнтів (Churn Predict) і оцінки фінансових ризиків. Однак, якщо тимчасова затримка обробки даних (latency) - це критичний чинник, Apache Spark не підійде і варто розглянути альтернативу у вигляді клієнтської бібліотеки Kafka Streams або фреймворків Storm, Flink, Samza.

По набору компонентів і функціональними можливостями Spark можна порівняти з іншим Big Data інструментом розподіленої потокової обробки - Apache Flink.

**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

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

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.



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

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

**ГЛОСАРІЙ**

BI - це технології, інструменти, системи та програми для складання, аналізу,

комбінації та виставки бізнес -звіту з активним способом виконання бізнес - рішень.

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.

Hevo - Рекомендований інструмент ETL детальний моніторинг, щоб ви завжди були в курсі своїх даних.

SEO – це складна суміш на сторінках, поза сторінками та технічних факторів, які безпосередньо впливають на рейтинг вашого веб-сайту в пошуковій мережі – і, як наслідок, на вашу здатність залучати більше органічного трафіку. Xplenty - це хмарне рішення ETL, що забезпечує прості візуалізовані конвеєри даних для автоматизованих потоків даних у широкому діапазоні джерел та напрямків.

Аналіз даних - це дослідження, пов'язані з обрахуванням багатовимірної системи даних, що має безліч параметрів

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.

Електронна пошта – це один з найважливіших маркетингових каналів, що забезпечує середню рентабельність інвестицій 42 долари за кожен витрачений 1 долар.

Інформаційна панель (інфопанель, дашборд) — тип графічного інтерфейсу користувача, що забезпечує наочну презентацію основних показників продуктивності (ОПП), значимих для конкретної цілі чи підприємчого процесу

ІТ − це комплекс взаємозалежних, наукових, технологічних, інженерних дисциплін, що вивчають методи ефективної організації праці людей, зайнятих обробкою й зберіганням інформації; обчислювальна техніка й методи організації й взаємодії з людьми й виробничим устаткуванням, їхні практичні додатки, а також зв'язані з усім цим соціальні, економічні й культурні проблеми.

Машинне навчання - клас методів штучного інтелекту, характерною рисою яких є не пряме рішення задачі, а навчання за рахунок застосування рішень безлічі подібних завдань.

Метод класифікації – це метод аналізу даних, який дозволяє оцінити ймовірність приналежності екземплярів даних до деякого класу залежно від значень їх атрибутів.

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 Повне оновлення - Повністю стирання вмісту однієї або декількох таблиць і перезавантаження з новими даними.

Системні інтегратори - компанії, які впроваджують системи аналізу великих даних на стороні клієнта.

Скайвія - це хмарна платформа даних для інтеграції, резервного копіювання, управління та доступу до даних без кодування, розроблена Devart. Споживачі - компанії, які купують програмно-апаратні комплекси і замовляють алгоритми у консультантів. Це компанії з галузей фінансів, телекомунікацій, рітейлу.

Сховище даних - це система, яка підтримує процес бізнес-аналітики. Він перетворює дані в значущу інформацію для аналізу бізнесу. Тому це цінний ресурс для керівництва організації при прийнятті рішень.

Хмарні обчислення - це метод надання набору спільних обчислювальних ресурсів, що включає програми, обчислення, зберігання, мережу, платформи розробки та розгортання, а також бізнес -процеси.

Штучний інтелект - властивість інтелектуальних систем виконувати творчі функції, які традиційно вважаються прерогативою людини; наука і технологія створення інтелектуальних машин, особливо інтелектуальних комп'ютерних програм.