

Data Mining як
мультидисциплінарна
галузь





Data Mining:

- пошук цінної інформації у великій базі даних (data)
- добича гірської руди (mining)



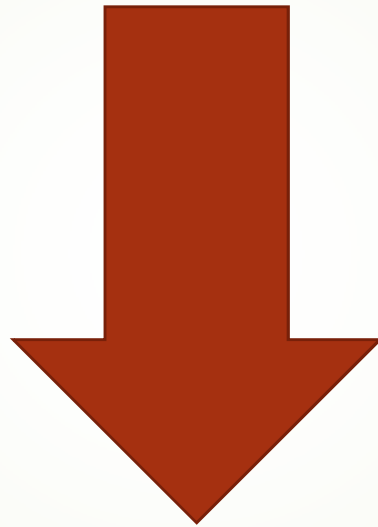
Data Mining:

- добича даних, добування інформації, розкопка даних
- інтелектуальний аналіз даних
- засоби пошуку закономірностей, добування знань, аналіз шаблонів
- “добування зерен знань з гір даних”
- розкопка знань у базах даних, “промивання” даних



Data Mining:

▶ поняття з'явилося у 1978р.



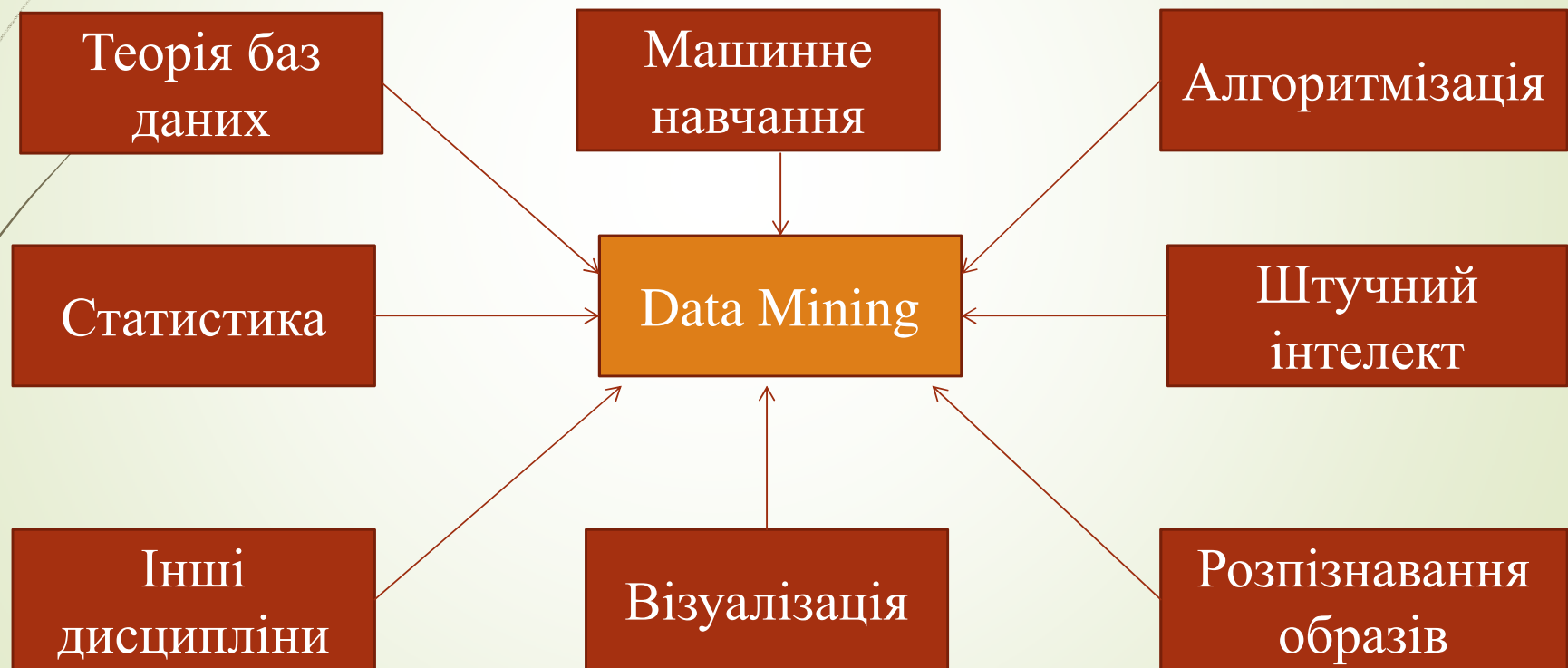
▶ висока популярність у сучасній трактовці – перша половина 1990-х рр.

Data Mining:

The image shows a screenshot of a Google search page for the query "data mining". The search bar contains the text "data mining" and is circled in red. Below the search bar, the navigation menu includes "Все", "Картинки", "Видео", "Новости", "Книги", "Ещё", "Настройки", and "Инструменты". The search results section shows "Результатов: примерно 540 000 000 (0,51 сек.)" circled in red. Below this, the heading "Результаты по запросу 'data mining'" is followed by a "Реклама" section. Five book advertisements are displayed in a row, each with a cover image, title, price, and retailer. A right arrow button is visible on the right side of the advertisement row.

Book Title	Price	Retailer
Big data. Вся технология в ...	144,00 грн.	Yakaboo.ua
Statistica Expert - Data Science	176 175,00 г...	Softlist.com.ua
Introduction To Data Mining 1st	5380,00 грн.	Ubuy Ukraine
Data Mining (Ebook) Foundations and Practice	6570,67 грн.... 272,43 \$ без...	eBooks.com
Анализ больших ...	1110,00 грн.	Читайка

Data Mining як мультидисциплінарна область:



- ***Статистика*** – наука про методи збирання даних, їх обробки та аналізу для виявлення закономірностей, що притаманні явищу, що вивчається
- ***Машинне навчання*** – процес отримання програмою нових знань (приклад - нейронні мережі)
- ***Штучний інтелект*** – науковий напрямок, у рамках якого ставляться та вирішуються завдання апаратного та програмного моделювання видів людської діяльності, які традиційно вважаються інтелектуальними

► *Статистика*

- *більше, ніж Data Mining, базується на теорії*
- *більше зосереджується на перевірці гіпотез*

► *Машинне навчання*

- *більш евристичне*
- *концентрується на покращенні роботи агентів навчання*

► *Data Mining*

- *інтеграція теорії та евристик*
- *сконцентрована на єдиному процесі аналізу даних, включає очищення даних, навчання, інтеграцію та візуалізацію результатів*

Фактори, що посприяли виникненню Data Mining:

- удосконалювання апаратного та програмного забезпечення
- удосконалювання технологій зберігання та запису даних
- накопичення великої кількості ретроспективних даних
- удосконалювання алгоритмів обробки інформації

- **Знання** – сукупність відомостей, що утворює цілісний опис, який відповідає деякому рівню обізнаності про питання, предмет, проблему тощо
- **Data Mining** – процес, мета якого – виявити нові значущі кореляції, образи та тенденції у результаті просіювання великого обсягу даних, що зберігаються, з використанням методик розпізнавання образів та застосування статистичних та математичних методів (*визначення Gartner Group*)

► ***Business Intelligence (BI)*** – програмні засоби, що функціонують у рамках підприємств та забезпечують функції доступу та аналізу інформації, що знаходиться у сховищі даних, а також що забезпечують прийняття правильних та обґрунтованих управлінських рішень.



Класифікація аналітичних систем:

- засоби побудови сховищ даних (*Data Warehousing*)
- системи оперативної аналітичної обробки (*OLAP*)
- інформаційно-аналітичні системи (*Enterprise Information Systems, EIS*)
- засоби інтелектуального аналізу даних (*Data Mining*)
- Інструменти для виконання запитів та побудови звітів (*Query and Reporting Tools*)

Проблеми та обмеження Data Mining:

- Data Mining не може замінити аналітика
- складність розробки та експлуатації додатків Data Mining
- кваліфікація користувача
- добування корисних знань неможливо без хорошого розуміння суті даних
- складність підготовки даних
- великий відсоток помилкових, недостовірних та безглузвих результатів
- висока вартість
- наявність достатньої кількості репрезентативних даних

Перспективи технології Data Mining:

- виділення типів предметних галузей з відповідними евристичними, формалізація яких полегшить розв'язок відповідних завдань Data Mining, що відносяться до цих галузей
- створення формальних мов та логічних засобів, за допомогою яких будуть формалізовані міркування та автоматизація яких стане інструментом вирішення завдань Data Mining у конкретних предметних галузях

Перспективи технології Data Mining:

- створення методів Data Mining, які здатні не тільки здобувати із даних закономірності, але й формувати деякі теорії, що спираються на емпіричні дані
- подолання суттєвого відставання можливостей інструментальних засобів Data Mining від теоретичних досягнень у цій галузі

Особливості галузей застосування

Data Mining:

- вимагають рішень, заснованих на знаннях
- мають навколишнє середовище, що змінюється
- мають доступні, достатні та значущі дані
- забезпечують високі дивіденди від правильних рішень