

วิชาเหมืองข้อมูล Data Mining (4124305)

บทที่ 1

2/2568

แนะนำเหมืองข้อมูลเบื้องต้น (Introduction to Data Mining)



Asst. Prof. Dr. Nattapong Songneam

Department of Computer Science,
Faculty of Science and Technology,
Phranakhon Rajabhat University

แก้ไขล่าสุด : 10.11.2568

ดาต้า ไมนิง (Data mining)

Review

รอบ ๆ ตัวเรามีข้อมูลมากมายที่เกิดขึ้นจากกิจกรรมต่าง ๆ ในชีวิตประจำวันของเรา ซึ่งเราเองก็เป็นส่วนหนึ่งในการสร้างข้อมูลเหล่านั้น ตัวอย่างเช่น:

- ข้อมูลการซื้อขายสินค้าในร้านสะดวกซื้อ เช่น การเข้า 7/11 ที่บันทึกข้อมูลสินค้าที่ซื้อจำนวนและเวลา
- ข้อมูลการใช้บัตรเครดิต เช่น การเติมน้ำมันหรือซื้อบัตรชมภาพยนตร์ ซึ่งบ่งชี้ถึงพฤติกรรมการใช้จ่ายของลูกค้า
- ข้อมูลการติดต่อสอบถามผ่าน Call Center ซึ่งช่วยเก็บข้อมูลเกี่ยวกับคำถามหรือข้อสงสัยของลูกค้าที่ติดต่อเข้ามา
- ข้อมูลการใช้โทรศัพท์และแอปพลิเคชัน เช่น การติดต่อสื่อสารผ่านแอปพลิเคชันไลน์ หรือการใช้งานโทรศัพท์มือถือในการสื่อสาร
- ข้อมูลการเลือกอาหาร ที่เลือกจากร้านอาหารผ่านแอปพลิเคชันต่าง ๆ เช่น Grab, Line Man, Food Panda, Robinhood, Shoppy Food, หรือ Moo Sang
- ข้อมูลรูปภาพบนเครือข่ายสังคมออนไลน์ เช่น รูปภาพและโพสต์ที่แชร์บน Facebook, Instagram หรือ Twitter
- ข้อมูลเหล่านี้ไม่เพียงแต่สะท้อนพฤติกรรมของเรา แต่ยังเป็นแหล่งข้อมูลที่มีค่าที่สามารถนำมาวิเคราะห์เพื่อเข้าใจแนวโน้มและพฤติกรรมผู้บริโภคได้ดียิ่งขึ้น

ดาต้า ไมนิ่ง (Data mining)

Review

Data Mining (การขุดค้นข้อมูล) เริ่มต้นจากการพัฒนาเทคโนโลยีที่สามารถจัดการและวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) โดยใช้เทคนิคทางสถิติ, การคำนวณเชิงคณิตศาสตร์, และเครื่องมือการเรียนรู้ของเครื่อง (Machine Learning) เพื่อค้นหความสัมพันธ์หรือรูปแบบที่ซ่อนอยู่ในข้อมูลจากแหล่งข้อมูลที่หลากหลาย

ในช่วงปี 1990s เป็นช่วงที่ Data Mining เริ่มมีการพัฒนามากขึ้น โดยการใช้เครื่องมือเชิงสถิติและการคำนวณคณิตศาสตร์ในการขุดค้นข้อมูลขนาดใหญ่ เช่น ฐานข้อมูลทางธุรกิจหรือการเก็บข้อมูลจากเว็บไซต์ ต่อมาในปี 2000 มีการนำ Machine Learning และ Artificial Intelligence (AI) มาใช้ในการพัฒนาการขุดค้นข้อมูลให้มีความแม่นยำและเป็นระบบมากขึ้น

ดาต้า ไมนิง (Data mining)

Review



ดาต้า ไมนิง (Data mining)

Review

Review / Background

Database System

SQL

Big Data

A.I.

ฐานข้อมูลคืออะไร

คือการนำข้อมูลมารวมกันเป็นกลุ่ม หรือรวมไว้ที่เดียวกัน แล้ว

เราจะจัดการอะไรกับฐานข้อมูล

ข้อมูล คืออะไร คือข้อเท็จ ที่สร้างขึ้นมา ซึ่งไม่ใช่ข้อคิดเห็น หรือความคิดของคน

สารสนเทศ (Information) คือข้อมูลที่ ผ่านการประมวลผล ทำให้อยู่ในรูปที่นำไปใช้ประโยชน์ได้

ประเภทของฐานข้อมูล

ฐานข้อมูลแบบเชิงสัมพันธ์ (RDBMS)

ฐานข้อมูลแบบเครือข่าย (Network Database)

ฐานข้อมูลแบบกระจาย (Distributed Database)

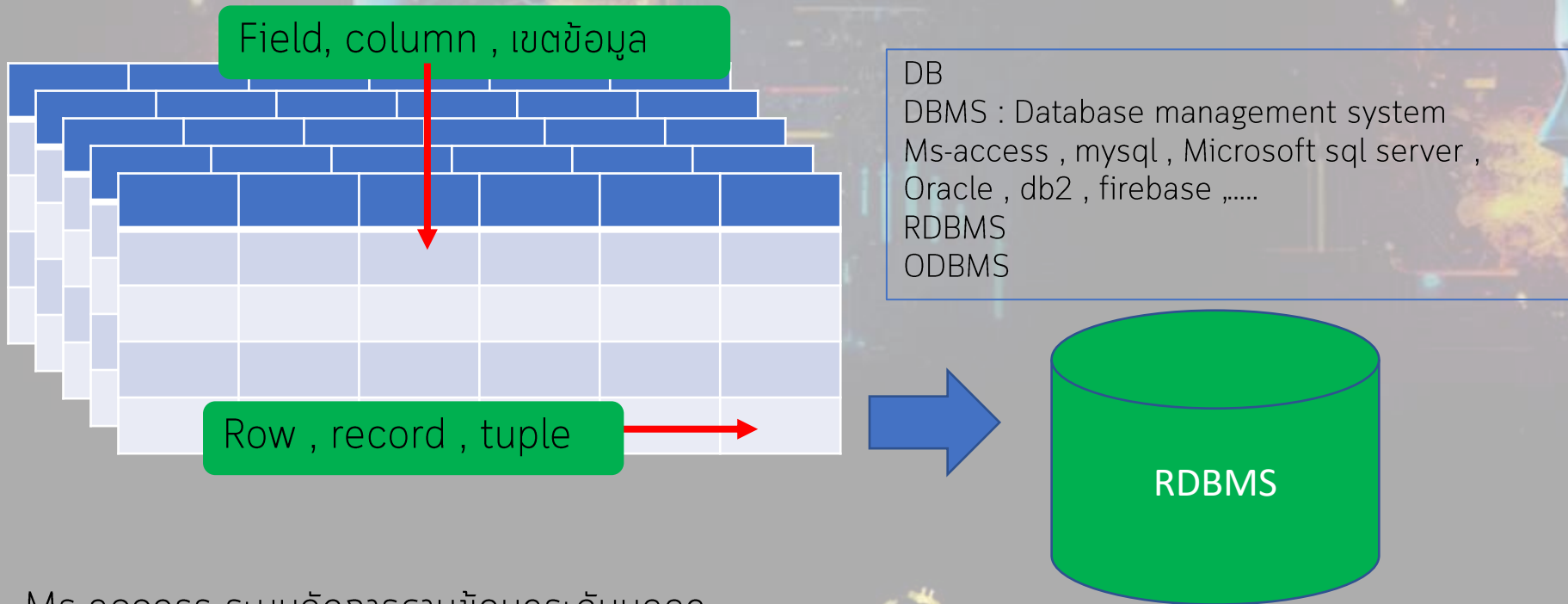
ฐานข้อมูลแบบเชิงสัมพันธ์ (RDBMS)

- MySQL
- MS-SQL
- ACCESS
- Infomix
- DB2
- Oracle

ฐานข้อมูลเชิงสัมพันธ์ (Relational Database)

Review

อยู่ในรูปแบบของตาราง หลาย ๆ ตารางเข้าเชื่อมโยงกัน (ตารางคือข้อมูลในลักษณะ 2 มิติ ประกอบไปด้วย row , column)



DB
DBMS : Database management system
Ms-access , mysql , Microsoft sql server ,
Oracle , db2 , firebase ,....
RDBMS
ODBMS

Ms-access ระบบจัดการฐานข้อมูลระดับบุคคล

ฐานข้อมูลแบบเชิงสัมพันธ์ (RDBMS)

ID	Name	Price	Unit	Type	
001	แปรงสีฟัน	60	ชิ้น	p01	

TbProduct

TypeID	Type Name	Detail			
p01	ของใช้	เป็นของใช้ใน ชีวิต			

TbProductType

CustID	Name	Address	ประวัติการทำงาน		
c001	สมชาย	กรุงเทพ	ปตท		
c001	สมชาย	เชียงใหม่	เซลล์		
c001	สมชาย	สงขลา	ThaiBev		

TbProduct

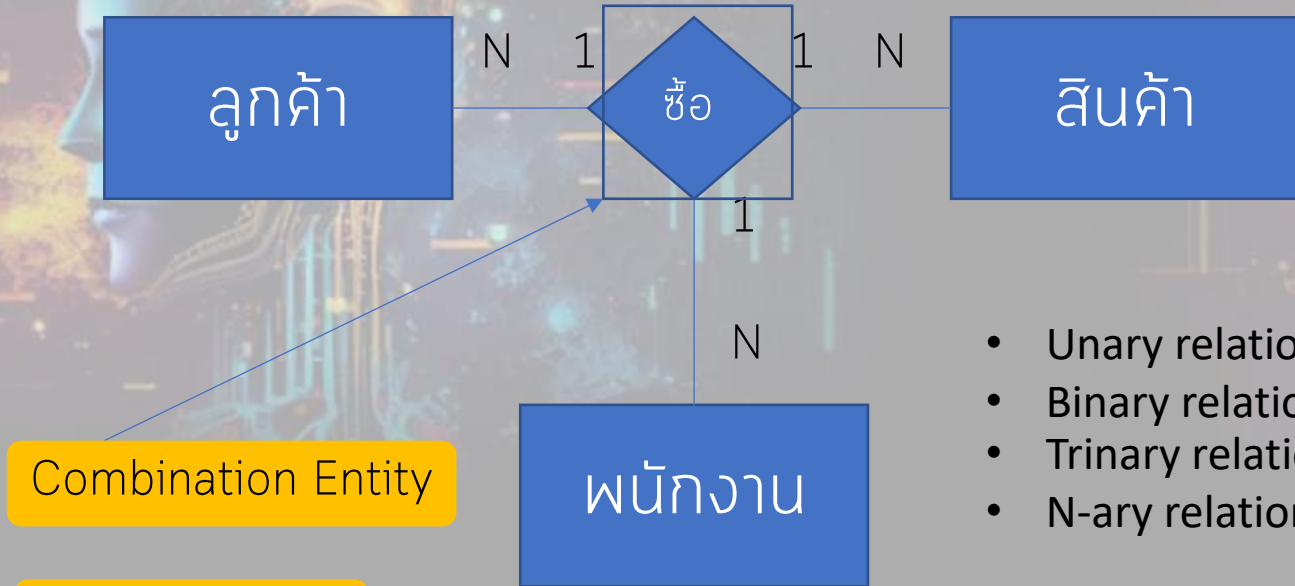
Multi value Attribute

Tbwork

Decomposition

composition

ERD



Combination Entity

รายละเอียดการซื้อ

- Unary relationship
- Binary relationship
- Trinary relationship
- N-ary relationship

ER-Mapping

การออกแบบระบบร้านสะดวกซื้อ (ร้าน 7-11 , โลตัส food go , cj mall ,)



ภาษาในการสืบค้นข้อมูล (SQL)

Review

ชื่อ SQL: Structure Query Language

1. DDL : Data Definition Language ภาษาสำหรับนิยามโครงสร้างของข้อมูล CREATE database , Table ..
2. DML : Data Manipulation Language ภาษาที่ใช้จัดการกับข้อมูล เช่น Select , Delete , Insert , Update ..
3. DCL : Data Control Language : ภาษาควบคุมข้อมูล เช่น การจัดการสิทธิ์ผู้ใช้งานระบบ

ภาษาในการสืบค้นข้อมูล (SQL)

Review

DDL : Data Definition Language ภาษาสำหรับนิยามโครงสร้างของข้อมูล CREATE database ,Table

SQL

```
CREATE TABLE customers (  
    customer_id INT PRIMARY KEY,  
    first_name VARCHAR(255) NOT NULL,  
    last_name VARCHAR(255) NOT NULL,  
    email VARCHAR(255) UNIQUE NOT NULL  
);
```

โปรดใช้โค้ดด้วยความระมัดระวัง [ดูข้อมูลเพิ่มเติม](#)

ภาษาในการสืบค้นข้อมูล (SQL)

Review

DDL : Data Definition Language ภาษาสำหรับนิยามโครงสร้างของข้อมูล **CREATE database ,Table**

เป็นคำสั่งไขโครงสร้างของตารางโดยการเพิ่มคอลัมน์

SQL

```
ALTER TABLE customers ADD COLUMN phone_number VARCHAR(20);
```

โปรดใช้โค้ดด้วยความระมัดระวัง [ดูข้อมูลเพิ่มเติม](#)



ภาษาในการสืบค้นข้อมูล (SQL)

Review

DDL : Data Definition Language ภาษาสำหรับนิยามโครงสร้างของข้อมูล **CREATE database ,Table**

- สร้างดัชนี:

SQL

```
CREATE INDEX idx_customers_last_name ON customers (last_name);
```

โปรดใช้โค้ดด้วยความระมัดระวัง [ดูข้อมูลเพิ่มเติม](#)



- สร้างผู้ใช้:

SQL

```
CREATE USER 'new_user' IDENTIFIED BY 'new_password';
```

โปรดใช้โค้ดด้วยความระมัดระวัง [ดูข้อมูลเพิ่มเติม](#)



ภาษาในการสืบค้นข้อมูล (SQL)

Review

DML : Data Manipulation Language ภาษาที่ใช้จัดการกับข้อมูล เช่น **Select , Delete , Insert , Update ,**

แทรกข้อมูล:

SQL

```
INSERT INTO customers (customer_id, first_name, last_name, email)
VALUES (101, 'John', 'Doe', 'johndoe@example.com');
```

โปรดใช้โค้ดด้วยความระมัดระวัง [ดูข้อมูลเพิ่มเติม](#)



ภาษาในการสืบค้นข้อมูล (SQL)

Review

DML : Data Manipulation Language ภาษาที่ใช้จัดการกับข้อมูล เช่น Select , Delete , Insert , Update ,

อัปเดตข้อมูล:

SQL

```
UPDATE customers SET last_name = 'Smith' WHERE customer_id = 101;
```

โปรดใช้โค้ดด้วยความระมัดระวัง ดูข้อมูลเพิ่มเติม



ภาษาในการสืบค้นข้อมูล (SQL)

Review

DML : Data Manipulation Language ภาษาที่ใช้จัดการกับข้อมูล เช่น Select , Delete , Insert , Update ,

ลบข้อมูล:

SQL

```
DELETE FROM customers WHERE customer_id = 101;
```

โปรดใช้ด้วยความระมัดระวัง [ดูข้อมูลเพิ่มเติม](#)



ภาษาในการสืบค้นข้อมูล (SQL)

Review

ดึงข้อมูล:

SQL

```
SELECT * FROM customers WHERE first_name = 'John';
```

โปรดใช้โค้ดด้วยความระมัดระวัง [ดูข้อมูลเพิ่มเติม](#)

เพศ
F/M
หญิง
ชาย

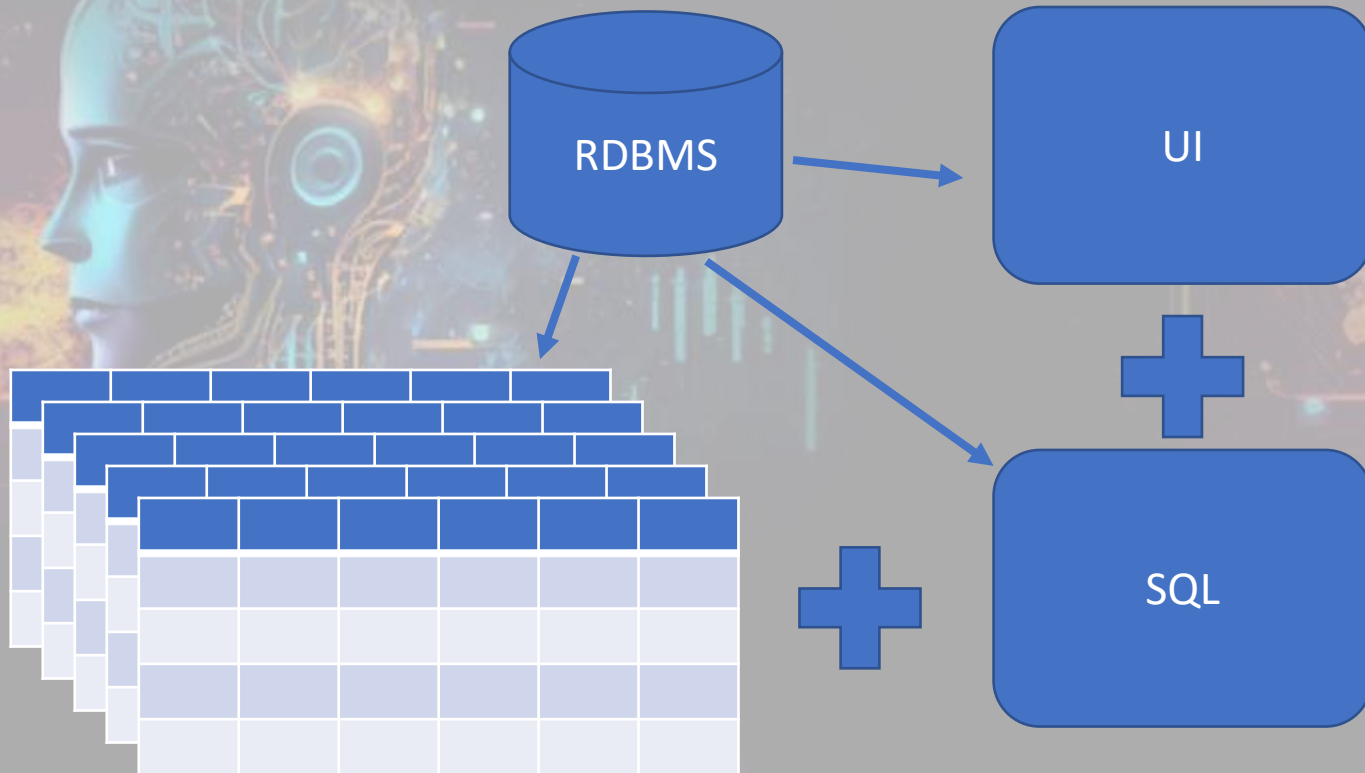
เพศ
F/M
Female
Male

RDBMS



Review

องค์ประกอบของ RDBMS



Database + Data Mining

Review

เก็บข้อมูล ค้นหาได้ สร้าง หรือ ลบได้

ปัจจุบัน การทำงานของฐานข้อมูล ไม่เพียงพอ ถ้าเราจะทำนาย อะไรบ้างโดยใช้ฐานข้อมูล

เกิดจาก ข้อมูลขนาดใหญ่ ที่เรียก Big Data นำข้อมูลปริมาณมาก ๆ จากหลายแหล่ง เข้าด้วยกัน รวมเป็น คลังข้อมูล (Data Warehouse)

Database + Data Mining

Review

เกิดจาก ข้อมูลขนาดใหญ่ ที่เรียก Big Data นำข้อมูลปริมาณมาก ๆ จากหลายแหล่ง เข้าด้วยกัน รวมเป็นคลังข้อมูล (Data Warehouse)



Big Data คืออะไร

Review

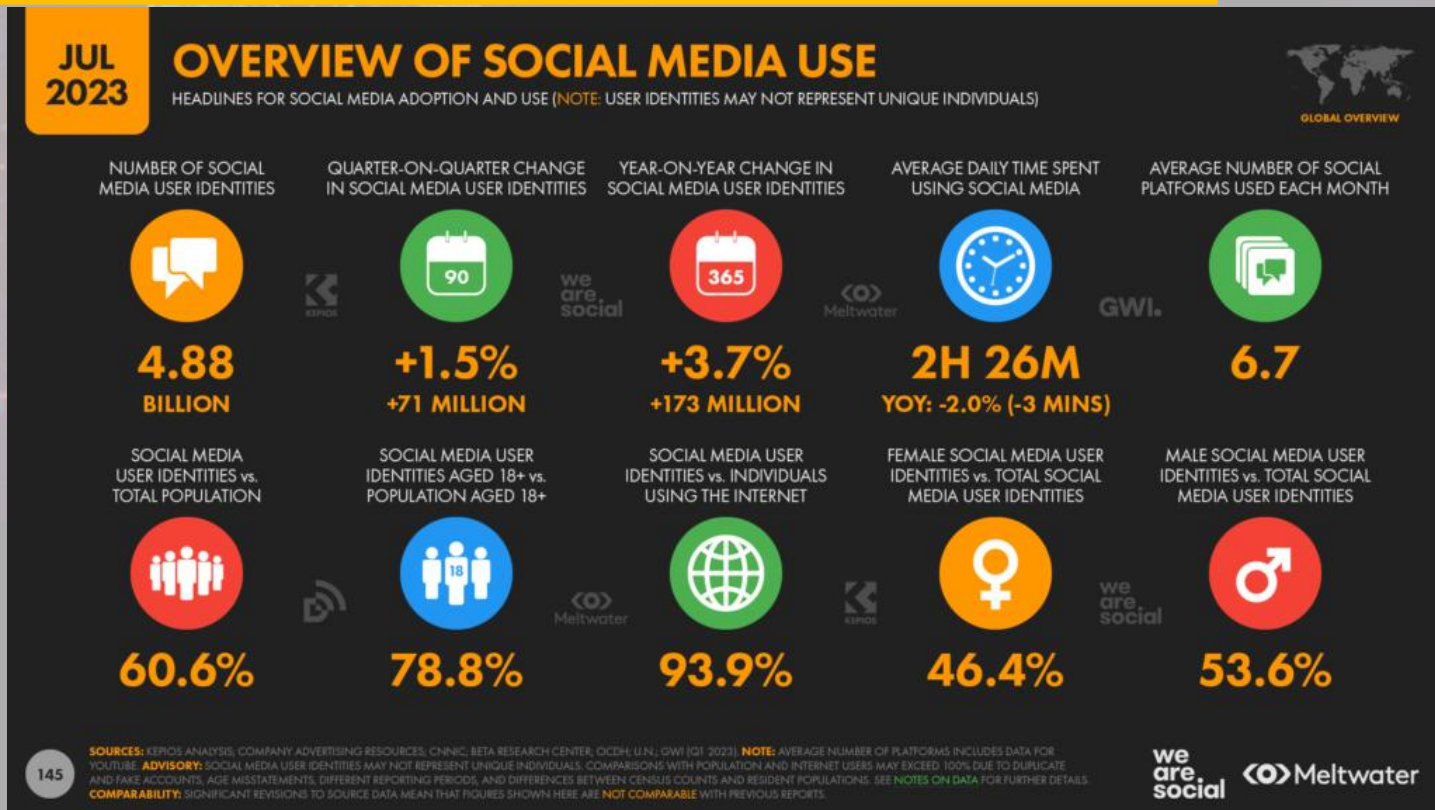
Big Data คือ ข้อมูลขนาดใหญ่/ปริมาณมาก หรือ ข้อมูลจำนวนมากมหาศาล ทุกเรื่อง ทุกแง่มุม ทุกรูปแบบ ซึ่งอาจเป็นข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) เช่น ข้อมูลที่เก็บอยู่ในตารางข้อมูลต่างๆ หรืออาจเป็นข้อมูลกึ่งมีโครงสร้าง (Semi-Structured Data) เช่น ล็อกไฟล์ (Log files) การเข้าระบบ การเข้าเว็บไซต์ หรือแม้กระทั่งข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อมูลการโต้ตอบปฏิสัมพันธ์ผ่านสังคมเครือข่าย (Social Network) เช่น Facebook, twitter, LINE หรือ ไฟล์จำพวกมีเดีย เป็นต้น ซึ่งอาจจะเป็นข้อมูลภายในองค์กรและภายนอกที่มาจาก การติดต่อระหว่างองค์กร หรือจากทุกช่องทางการติดต่อกับลูกค้า แต่ทั้งหมดนี้ก็ยังคงเป็นเพียงข้อมูลดิบที่รอการนำมาประมวลและวิเคราะห์เพื่อนำผลที่ได้มาสร้างมูลค่าทางธุรกิจ ข้อมูลเหล่านี้อาจจะไม่ได้อยู่ในรูปแบบที่องค์กรสามารถนำไปใช้ได้ทันที แต่อาจมีข้อมูลที่เป็นประโยชน์ต่อองค์กรบางอย่างแฝงอยู่

Big Data มีคุณลักษณะสำคัญอยู่ 4 อย่าง คือ

- **ปริมาตร (Volume)** หมายถึง ข้อมูลนั้นมันต้องมีขนาดใหญ่มาก ซึ่งไม่สามารถประมวลผลปริมาณของข้อมูลด้วยระบบฐานข้อมูลได้ จำเป็นต้องใช้คลังข้อมูล (Data Warehouse) และซอฟต์แวร์ฮาดูป (Hadoop) ทำงานประสานกันในการบริหารจัดการข้อมูล
- **ความเร็ว (Velocity)** หมายถึง ข้อมูลดังกล่าวต้องมีอัตราการเพิ่มขึ้นอย่างรวดเร็ว เช่น ข้อมูลจากภาพถ่ายโทรศัพท์ที่ถูกอัปโหลดขึ้น ข้อมูลการพิมพ์สนทนา ข้อมูลวิดีโอ รวมไปถึงข้อมูลการสั่งซื้อสินค้า ผุดงาย ๆ คือ ข้อมูลที่มีการเพิ่มขึ้นตลอดเวลาแบบไม่มีหยุดยั้งนั้นแหละ
- **ความหลากหลาย (Variety)** หมายถึง รูปแบบข้อมูลต้องมีความหลากหลาย อาจจะเป็นข้อมูลที่มีโครงสร้าง ไม่มีโครงสร้าง และกึ่งมีโครงสร้าง ซึ่งผมไม่ขอลงลึกนะเพราะมันซับซ้อนมาก แต่เอาเป็นว่ารูปแบบข้อมูลของ Big Data มันมีทุกอย่าง ไม่ได้จำกัดแค่พวกข้อความ อีเมล รูปภาพ ฯลฯ เท่านั้น
- **Veracity** ไม่สามารถนำมาใช้เป็นข้อมูลที่สมบูรณ์ เพื่อการประกอบการพิจารณาได้

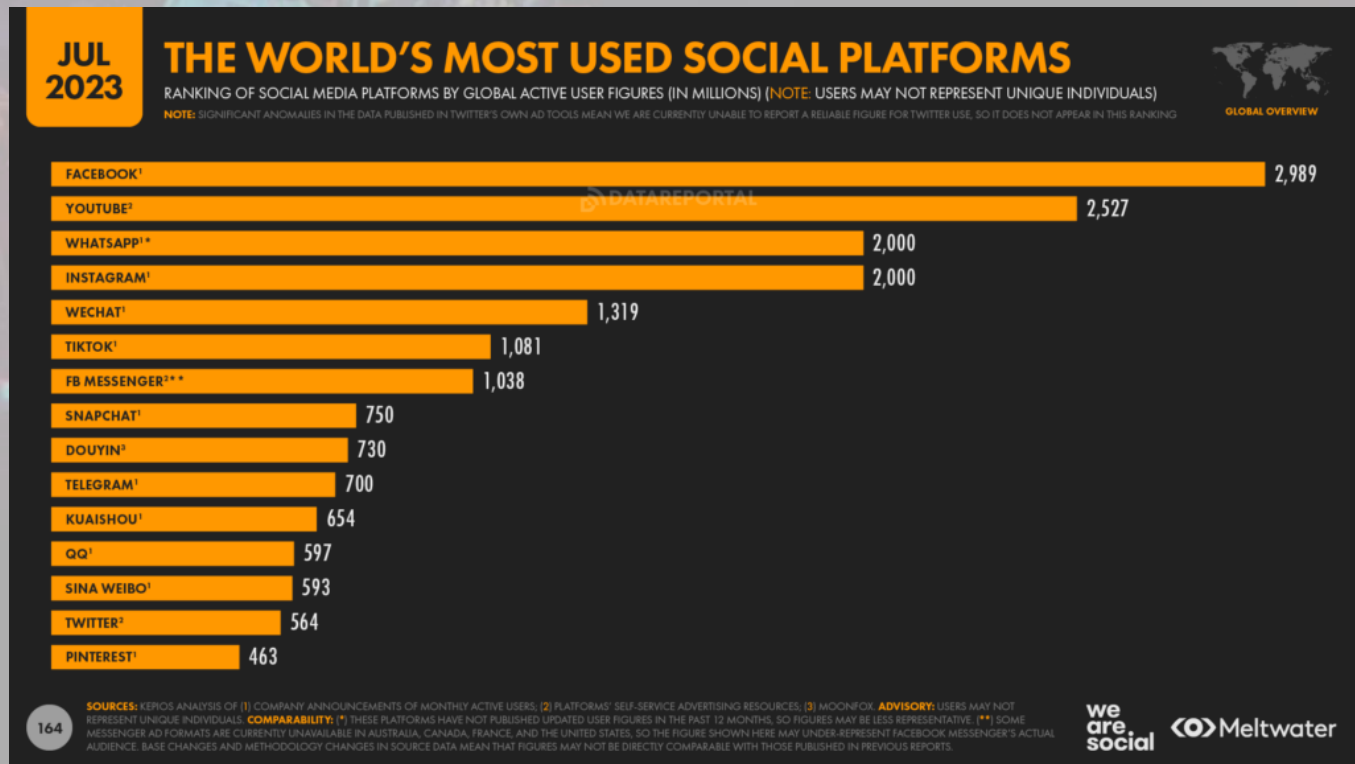
SOCIAL MEDIA

Review



SOCIAL MEDIA

Review



ทำไมต้อง Big Data?

Review

ปัจจุบัน เราทุกคนใช้งานโซเชียลมีเดีย เช่น YouTube, Facebook, Twitter, Google, Netflix, Walmart, Starbucks สิ่งหนึ่งที่ทำให้โซเชียลมีเดียเหล่านี้ประสบความสำเร็จ คือ Big Data เป็นการนำข้อมูลจำนวนมากมาวิเคราะห์ เพื่อหาโอกาสทางธุรกิจ ใช้ประกอบการตัดสินใจในเรื่องสำคัญๆ ทั้งการพัฒนาด้านการขายและการตลาด การปรับปรุงสินค้า บริการให้ตอบโจทยความต้องการของผู้บริโภคยุคใหม่ ที่เปลี่ยนแปลงอย่างรวดเร็ว รวมถึงภาคการผลิตที่นำข้อมูล Big Data ไปใช้ในการวิเคราะห์เพื่อเพิ่ม Productivity ในกระบวนการผลิตและการดำเนินงาน



การใช้ประโยชน์จาก Big Data

Review

ในปัจจุบันนี้ การนำ Big Data มาใช้ในภาครัฐ เพื่อแก้ปัญหาความเดือดร้อนและลดความเหลื่อมล้ำ โดยนำข้อมูลในระบบราชการจากหลายหน่วยงาน เช่น ข้อมูลสาธารณสุข ทะเบียนราษฎร์ ที่ตั้งของธุรกิจ โรงพยาบาล สถานบำบัด สถานการณ์จ้างงานฯ มาวิเคราะห์และการเชื่อมโยงกัน เกิดเป็นข้อมูลขนาดใหญ่ Big Data ของภาครัฐ ผ่านกระบวนการวิเคราะห์เชื่อมโยงเพื่อตอบการให้บริการของภาครัฐ ตัวอย่างเช่น รัฐบาลต้องการช่วยเหลือผู้มีรายได้น้อย แต่แทนที่จะช่วยเหลือโดยให้เงินอุดหนุนที่เท่าๆ กันแบบปูพรมทั่วประเทศ ก็นำ Big Data ซึ่งเป็นข้อมูลจากแหล่งต่าง ๆ มาใช้ชี้จำเพาะว่าบุคคลใดที่ถือว่ามียาได้น้อย พร้อมทั้งกำหนดระดับและลักษณะความช่วยเหลือที่แตกต่างกัน เช่น ผู้มีรายได้น้อยที่สูงอายุ เป็นผู้พิการ อยู่กับบ้าน ให้ลูกหลานดูแล รัฐบาลช่วยโดยสนับสนุนขเทียมให้อุปถัมภ์เข้ารับการกายภาพบำบัด พร้อมทั้งเลือกอาชีพที่เหมาะสมกับกายภาพของผู้สูงอายุ

การฝึกอาชีพเพื่อเพิ่มรายได้ ให้กับผู้มีรายได้น้อย พร้อมทั้งจับคู่กับแหล่งงานที่อยู่ใกล้เคียงกับที่พักอาศัย อีกทั้งยังติดตามและเสนอโอกาสฝึกอาชีพใหม่ๆ เพิ่มเติม เพื่อให้มีรายได้ที่สูงขึ้นและพัฒนาคุณภาพชีวิตให้ดีขึ้น ซึ่งถ้าวิเคราะห์ดูจะเห็นว่า ข้อมูลจำนวนมากเกิดการบูรณาการและวิเคราะห์ เพื่อใช้สำหรับการตัดสินใจในการให้บริการของภาครัฐได้ตรงกลุ่มเป้าหมาย โดยในปัจจุบันนี้ จะเห็นได้จากการใช้บัตรประชาชนเพียงบัตรเดียวก็สามารถเข้าถึงบริการภาครัฐได้มากขึ้น

การใช้ประโยชน์จาก Big Data

Review

Big Data สำหรับภาคเอกชนที่นำมาใช้ประโยชน์ เช่น เว็บไซต์อี-คอมเมิร์ซ ที่จัดเก็บข้อมูลพฤติกรรมการซื้อสินค้าของลูกค้าอย่างต่อเนื่อง และมีระบบที่ทำหน้าที่คัดเลือกสินค้าอื่น ๆ ที่คาดว่าจะต้องการเพิ่มเติม แล้วนำเสนอขึ้นมาให้โดยอัตโนมัติบนหน้าเว็บไซต์อี-คอมเมิร์ซของลูกค้ารายนั้น ๆ ทั้งนี้ ลูกค้าแต่ละคน ไม่จำเป็นต้องนำเสนอสินค้าเดียวกัน จากการสังเกตพฤติกรรมการซื้อสินค้า พบว่าภาคเอกชนจะมีการเก็บข้อมูล ชื่อ ที่อยู่ เพศ เชื้อชาติ อายุ ประวัติการซื้อสินค้า ชนิดสินค้า เวลาที่ซื้อ มูลค่าสินค้า นำมาวิเคราะห์จับคู่กับสินค้าอื่นที่มีศักยภาพ ทั้งนี้ เจือปนหรือสูตรการจับคู่อาจแตกต่างกันไป ตามกลุ่มลูกค้าหรือประชากรในแต่ละประเทศ หรือตามกลุ่มสังคมหรือวัฒนธรรม

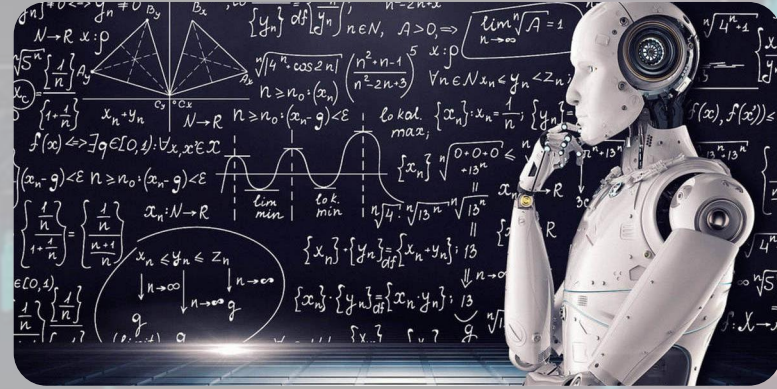
นอกจากนั้น ภาคเอกชนได้นำข้อมูล Big Data มาใช้ประโยชน์ เพื่อยกระดับธุรกิจ ด้วยการพัฒนาเทคโนโลยีแชทบอท (Chatbot) ที่สามารถรับมือกับความต้องการข้อมูลของลูกค้าที่ติดต่อเข้ามาจำนวนมากผ่าน Messaging Application ได้อย่างมีประสิทธิภาพ รวดเร็ว จับใจ พร้อมให้บริการตลอด 24 ชั่วโมง และนี่คือจุดเปลี่ยนสำคัญของการให้บริการที่จะเข้ามาใช้งานแทนคน (Agent)

ความหมายของปัญญาประดิษฐ์ (Artificial Intelligence: AI)

Review

หมายถึง “การทำให้คอมพิวเตอร์สามารถคิดหาเหตุผลได้ เรียนรู้ได้ ทำงานได้เหมือนสมองมนุษย์” (ทักษิณา สอนานนท์, 2539, น.13)

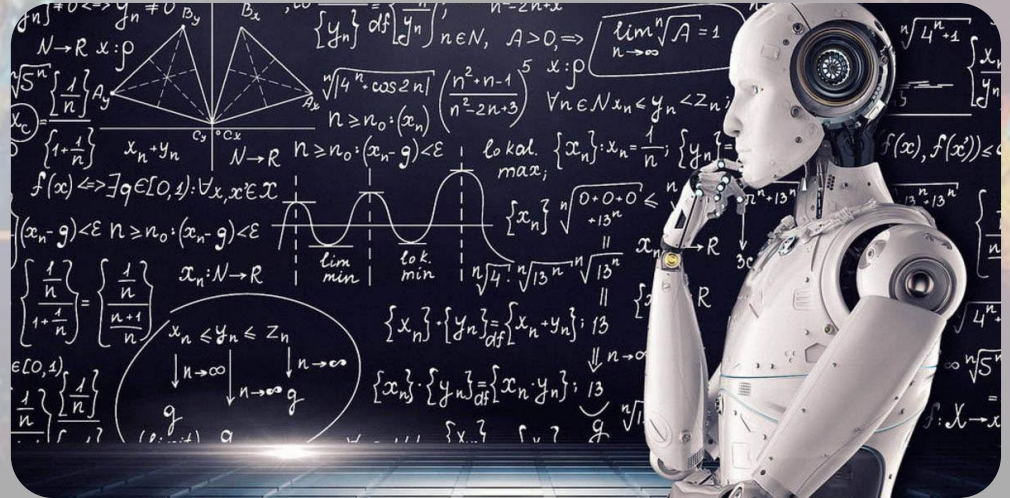
หมายถึง ศาสตร์แขนงหนึ่งของวิทยาการคอมพิวเตอร์ ที่ต้องการประดิษฐ์เครื่องจักร เช่น คอมพิวเตอร์ หรือ หุ่นยนต์ให้สามารถคิดและมีพฤติกรรมเลียนแบบมนุษย์ในกระบวนการตัดสินใจ แก้ปัญหาได้ ซึ่งอาจจะต้องมีการวินิจฉัย หาเหตุผล จากความรู้ที่จัดเก็บไว้ และนำความรู้นั้นมาเชื่อมโยง เพื่อหาข้อสรุป หรือผลลัพธ์ของปัญหานั้นได้ในที่สุด



ความหมายของปัญญาประดิษฐ์ (Artificial Intelligence:AI)

Review

ปัญญาประดิษฐ์ (Artificial Intelligence:AI) คือสาขาหนึ่งของวิทยาการคอมพิวเตอร์ที่มุ่งเน้นการพัฒนาระบบคอมพิวเตอร์ที่สามารถแสดงพฤติกรรมที่ฉลาดได้ ปัญญาประดิษฐ์มีขอบเขตกว้างขวางมากครอบคลุมตั้งแต่ระบบที่ทำงานในลักษณะที่คล้ายคลึงกับมนุษย์ เช่น การคิด การแก้ปัญหา และการวางแผน ไปจนถึงระบบที่ทำงานในลักษณะที่แตกต่างจากมนุษย์ เช่น ระบบที่เรียนรู้จากข้อมูลได้เอง ระบบที่ตัดสินใจได้เอง และระบบที่สามารถสร้างสิ่งใหม่ๆ ขึ้นมาได้



ความหมายของปัญญาประดิษฐ์ (Artificial Intelligence:AI)

Review

ปัญญาประดิษฐ์แบ่งออกเป็นสาขาย่อย ๆ มากมาย แต่ละสาขามีจุดมุ่งหมายและวิธีดำเนินการที่แตกต่างกัน ตัวอย่างสาขาย่อยของปัญญาประดิษฐ์ ได้แก่*

1. การเรียนรู้ของเครื่อง (Machine Learning): พัฒนาระบบคอมพิวเตอร์ที่สามารถเรียนรู้จากข้อมูลได้เอง
2. การประมวลผลภาษาธรรมชาติ (Natural Language Processing) : พัฒนาระบบคอมพิวเตอร์ที่สามารถเข้าใจและประมวลผลภาษามนุษย์ได้
3. การมองเห็นด้วยคอมพิวเตอร์ (Computer Vision) : พัฒนาระบบคอมพิวเตอร์ที่สามารถมองเห็นและเข้าใจภาพได้
4. หุ่นยนต์ (Robotics): พัฒนาหุ่นยนต์ที่สามารถทำงานหรือดำเนินการต่างๆ ได้ด้วยตนเอง

ลักษณะและความสามารถของปัญญาประดิษฐ์ (AI)

Review

1. สามารถเรียนรู้ และทำความเข้าใจประสบการณ์ที่ผ่านมา
2. สามารถทำความเข้าใจกับข้อความที่มีความคลุมเครือหรือมีความขัดแย้งกันได้
3. มีความสามารถในการตอบสนองที่รวดเร็วและถูกต้องต่อสถานการณ์ใหม่ๆ
4. สามารถใช้เหตุผลเพื่อแก้ปัญหา และให้คำแนะนำได้อย่างมีประสิทธิภาพ
5. สามารถจัดการกับสถานการณ์ที่มีความยุ่งยาก ซับซ้อนได้
6. สามารถทำความเข้าใจและวินิจฉัย เพื่อสรุปความอย่างมีประสิทธิภาพ
7. สามารถประยุกต์องค์ความรู้ให้เหมาะสมกับสภาพแวดล้อม
8. สามารถพิจารณาถึงความสำคัญขององค์ประกอบที่แตกต่างกันในแต่ละสถานการณ์ได้

ข้อจำกัดของปัญญาประดิษฐ์ (AI)

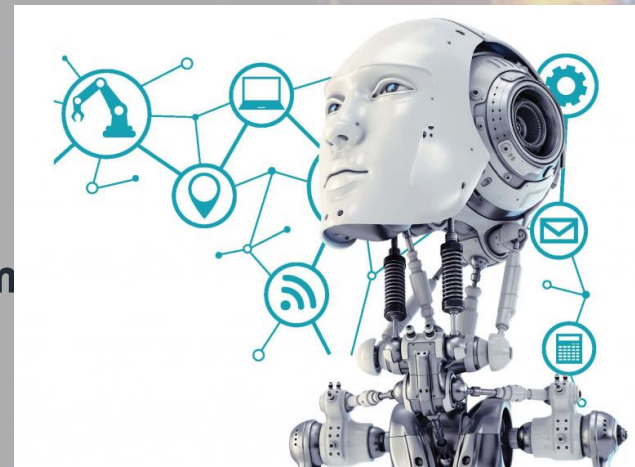
Review

ความสามารถที่ใช้เปรียบเทียบ	ปัญญาธรรมชาติ	ปัญญาประดิษฐ์ (AI)
ความละเอียดและยุติธรรม	ล่าเหยียงได้ และบางครั้งเกิดความไม่รอบคอบได้	ละเอียด รอบคอบ และยุติธรรมตามที่ได้ถูกโปรแกรมไว้
การนำประสบการณ์ออกมาใช้	นำออกมาใช้ได้เลย	ต้องมีการแปลงสารสนเทศก่อน
ความคิดริเริ่มสร้างสรรค์	มีความคิดริเริ่มสร้างสรรค์	ไม่มี
การให้เหตุผล	ต้องอาศัยประสบการณ์สูง	ถ้าเป็นปัญหาเฉพาะด้าน จะสามารถให้เหตุผลได้ดี

การประยุกต์ใช้ปัญญาประดิษฐ์ในศาสตร์แขนงอื่น

Review

1. หุ่นยนต์ (Robotic)
2. ระบบผู้เชี่ยวชาญ (Expert System: ES)
3. การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) และเทคโนโลยีเสียง (Voice/ Speech Technology)
4. คอมพิวเตอร์โครงข่ายประสาท (Neural Computing)
5. ตรรกะคลุมเครือ (Fuzzy Logic)
6. ตัวแทนปัญญา (Intelligent Agent)
7. ระบบช่วยสอนอัจฉริยะ (Intelligent Tutoring System)
8. ระบบความเป็นจริงเสมือน (Virtual Reality System)



ความแตกต่างระหว่าง Data Science, Data Analytics, และ Data Engineering:

ด้าน	Data Science	Data Analytics	Data Engineering
วัตถุประสงค์	ใช้เทคนิคทางสถิติ, ดนิตศาสตร์ และ Machine Learning เพื่อสร้างโมเดลที่สามารถทำนายหรือหาความสัมพันธ์จากข้อมูล	วิเคราะห์ข้อมูลที่มีอยู่เพื่อดึงข้อมูลเชิงลึกและช่วยในการตัดสินใจ	ออกแบบและพัฒนาโครงสร้างพื้นฐานที่จำเป็นสำหรับการจัดการข้อมูล
กิจกรรมหลัก	การสร้างโมเดล, การเรียนรู้ของเครื่อง (ML), การทำนาย, การจำแนกประเภทข้อมูล	การทำ EDA (Exploratory Data Analysis), การวิเคราะห์เชิงสถิติ, การสร้างรายงาน	การสร้างและดูแลฐานข้อมูล, การทำ ETL (Extract, Transform, Load), การจัดการข้อมูล
ทักษะที่ใช้	สถิติ, ดนิตศาสตร์, Machine Learning, การเขียนโปรแกรม (Python, R), การสร้างโมเดล	สถิติ, การวิเคราะห์ข้อมูล, การสร้างกราฟ/แผนภูมิ, การใช้เครื่องมือ BI (เช่น Power BI, Tableau)	การเขียนโปรแกรม (Python, Java, SQL), การออกแบบฐานข้อมูล, การจัดการ Big Data, Cloud Computing
เครื่องมือที่ใช้	Python, R, TensorFlow, Scikit-learn, Jupyter Notebook	Excel, Power BI, Tableau, Google Data Studio, Python (Pandas, Matplotlib)	SQL, Apache Hadoop, Apache Spark, Airflow, Kafka, AWS, GCP
ประเภทของข้อมูล	ข้อมูลที่มีโครงสร้างและไม่มีโครงสร้าง (structured and unstructured data)	ข้อมูลที่มีโครงสร้าง (structured data)	ข้อมูลที่มีโครงสร้าง (structured data), Big Data
ตัวอย่างการใช้งาน	การทำนายแนวโน้มการขายในอนาคต, การจำแนกประเภทลูกค้า	การวิเคราะห์ข้อมูลธุรกิจเพื่อรายงานหรือการตัดสินใจ, การวิเคราะห์ยอดขาย	การสร้างระบบฐานข้อมูล, การทำ ETL, การสร้าง data pipelines สำหรับการใช้งานต่าง ๆ

ความแตกต่างระหว่าง Data Science, Data Analytics, และ Data Engineering:



1. Data Science (วิทยาศาสตร์ข้อมูล)

Data Science คือ สาขาที่เน้นการใช้เทคนิคทางคณิตศาสตร์, สถิติ, และ Machine Learning เพื่อค้นหาความสัมพันธ์ในข้อมูล, สร้างโมเดลการทำนาย, และทำการวิเคราะห์เชิงลึก เพื่อให้ได้ข้อมูลที่สามารถใช้ในการตัดสินใจหรือคาดการณ์ในอนาคต ตัวอย่างเช่น การทำนายพฤติกรรมของลูกค้า การจำแนกประเภทข้อมูล หรือการวิเคราะห์เทรนด์ในตลาดต่าง ๆ

• กิจกรรมหลัก: การสร้างโมเดล, การทำนายข้อมูล, การเรียนรู้ของเครื่อง (Machine Learning)

• เครื่องมือที่ใช้: Python, R, TensorFlow, Scikit-learn

ทำความรู้จักทักษะแห่งอนาคต Data Analytics

รองรับการก้าวเข้าสู่ Smart City

Data Analytics

คือการวิเคราะห์ข้อมูล
เพื่อให้ได้ insight
สำหรับช่วยในการตัดสินใจ
ที่แม่นยำมากขึ้น

เริ่มง่าย ๆ ด้วยการ
จัดเก็บข้อมูล
อย่างเป็นระบบ
และตรงตามความเป็นจริง

เครื่องมือยอดนิยมทำ
Data Visualization
Excel
Power BI



สิ่งสำคัญคือ
Critical thinking
การคิดวิเคราะห์
อย่างมีวิจารณญาณ
จะช่วยให้วิเคราะห์ข้อมูล
ได้อย่างมีประสิทธิภาพ
และมีประโยชน์ต่อองค์กร
อย่างแท้จริง



2. Data Analytics (การวิเคราะห์ข้อมูล)

Data Analytics คือ สาขาที่มุ่งเน้นการใช้เทคนิคในการวิเคราะห์ข้อมูลที่มีอยู่ เช่น การสำรวจข้อมูล (Exploratory Data Analysis: EDA), การทำสถิติ, และการสร้างกราฟเพื่อให้เข้าใจข้อมูลและช่วยในการตัดสินใจทางธุรกิจ ตัวอย่างเช่น การวิเคราะห์ยอดขาย การสร้างรายงานจากข้อมูลที่มีอยู่ เพื่อให้ผู้บริหารสามารถตัดสินใจได้ดีขึ้น

- กิจกรรมหลัก: การสำรวจข้อมูล, การวิเคราะห์สถิติ, การสร้างรายงาน
- เครื่องมือที่ใช้: Excel, Tableau, Power BI, Python (Pandas, Matplotlib)

ทำความรู้จักทักษะแห่งอนาคต Data Analytics

รองรับการก้าวเข้าสู่ Smart City

Data Analytics

คือการวิเคราะห์ข้อมูล
เพื่อให้ได้ insight
สำหรับช่วยในการตัดสินใจ
ที่แม่นยำมากขึ้น

เริ่มง่าย ๆ ด้วยการ
จัดเก็บข้อมูล
อย่างเป็นระบบ
และตรงตามความเป็นจริง

เครื่องมือยอดนิยมทำ
Data Visualization
Excel
Power BI

สิ่งสำคัญคือ
Critical thinking
การคิดวิเคราะห์
อย่างมีวิจารณญาณ
จะช่วยให้วิเคราะห์ข้อมูล
ได้อย่างมีประสิทธิภาพ
และมีประโยชน์ต่อองค์กร
อย่างแท้จริง

3. Data Engineering (วิศวกรรมข้อมูล)

Data Engineering คือ สาขาที่เกี่ยวข้องกับการสร้างและดูแลโครงสร้างพื้นฐานข้อมูล เช่น การออกแบบและพัฒนาระบบฐานข้อมูล, การทำกระบวนการ ETL (Extract, Transform, Load) ซึ่งเกี่ยวข้องกับการดึงข้อมูลจากแหล่งต่าง ๆ, การแปลงข้อมูล และการโหลดข้อมูลเข้าสู่ระบบที่สามารถใช้งานได้ การจัดการ Big Data และการสร้าง Data Pipelines สำหรับการประมวลผลข้อมูลอย่างมีประสิทธิภาพ

- กิจกรรมหลัก: การออกแบบและดูแลฐานข้อมูล, การทำ ETL, การจัดการ Big Data
- เครื่องมือที่ใช้: SQL, Apache Hadoop, Apache Spark, Kafka, AWS, Google Cloud

บทที่ 1

เหมืองข้อมูลเบื้องต้น

1.1 แนะนำเหมืองข้อมูลเบื้องต้น

1.1.1 เหมืองข้อมูล คืออะไร

1.1.2 คุณลักษณะของเหมืองข้อมูล

1.1.3 วัฒนาการของการทำเหมืองข้อมูล

1.1.4 ประโยชน์ของเหมืองข้อมูล

1.1.5 ตัวอย่างการนำเหมืองข้อมูลไปใช้งาน

1.2 ความสำคัญของเหมืองข้อมูล

1.2.1 ทำไมจึงต้องมี เหมืองข้อมูล

1.2.2 วัตถุประสงค์ในการใช้เหมืองข้อมูล

1.2.3 เป้าหมายหลักของการทำเหมืองข้อมูล

1.3 ขั้นตอนการทำเหมืองข้อมูล

1.3.1 ขั้นตอนการทำเหมืองข้อมูล

1.3.2 การได้มาซึ่งข้อมูล/แหล่งข้อมูล

1.3.3 กระบวนการในการทำ data mining

1.4 เทคนิคในการทำเหมืองข้อมูล

1.4.1 ส่วนประกอบการทำเหมืองข้อมูล

1.4.2 ลักษณะเฉพาะของข้อมูลที่มีการใช้ทำเหมืองข้อมูล

1.4.3 ประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล

1.4.4 ประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล : Data Warehouse

1.5 เทคนิคในการทำเหมืองข้อมูล



บทที่ 1

1.1 แนะนำเหมืองข้อมูลเบื้องต้น

1.1.1

เหมืองข้อมูล คืออะไร

บทที่ 1

เหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมาก(big data ข้อมูลขนาดใหญ่)เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

เหมืองข้อมูล เป็นสาขาหนึ่งของวิทยาศาสตร์ ในอดีตวิทยาศาสตร์ จะพิสูจน์ความจริง อาศัยการทดลองและสังเกต แต่ปัจจุบันในยุคของข้อมูลมหาศาล การค้นหาคำจริง จึงเกิดการค้นหาคำจริงจากข้อมูล

การทำเหมืองข้อมูล เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้ในการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง
- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย
- การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
- การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

- การค้นหาข้อมูลโดยอาศัยเทคโนโลยีการทำเหมืองข้อมูล ภายใต้ฐานข้อมูลขนาดใหญ่หรือคลังข้อมูล ซึ่งข้อมูลอาจถูกสะสมมานานหลายปี
- ผู้ใช้งานระบบสารสนเทศไม่จำเป็นต้องมีทักษะในการเขียนโปรแกรม เนื่องจากมีเครื่องมือช่วยค้นหาข้อมูลจากคลังข้อมูลได้อย่างรวดเร็ว (Rapid Miner Studio9.4 , Weka)
- ผู้ใช้ต้องกำหนดขอบเขตการค้นหาข้อมูลให้ชัดเจนเพื่อความรวดเร็ว
- อาจมีการประมวลผลข้อมูลแบบขนาน เพื่อเพิ่มประสิทธิภาพในการค้นหาและวิเคราะห์ข้อมูล
- เครื่องมือสำหรับทำเหมืองข้อมูลสามารถใช้งานร่วมกับโปรแกรม Spreadsheet และเครื่องมือพัฒนาต่าง ๆ ได้เป็นอย่างดี

1.1.3

วิวัฒนาการของการทำเหมืองข้อมูล

บทที่ 1

- ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่น่าเชื่อถือ และป้องกันการสูญหายได้เป็นอย่างดี
- ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อประโยชน์ในการนำไปวิเคราะห์ และการตัดสินใจอย่างมีคุณภาพ
- ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปในฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกด้านขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ
- ปี 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล โดยการสร้างแบบจำลองและความสัมพันธ์ทางสถิติ

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

ประโยชน์ของเหมืองข้อมูล

- ช่วยชี้แนวทางการตัดสินใจและคาดการณ์ผลลัพธ์ที่จะได้จากการตัดสินใจ
- เพิ่มความเร็วในการวิเคราะห์ฐานข้อมูลขนาดใหญ่
- ค้นหาส่วนประกอบที่ซ่อนอยู่ภายในเอกสาร รวมถึงความสัมพันธ์ของส่วนประกอบต่าง ๆ ด้วย
- เชื่อมโยงหน่วยงานต่าง ๆ ภายในองค์กร
- การจัดกลุ่มข้อมูล เช่น จัดกลุ่มลูกค้าทั้งหมดของบริษัทประกันภัยที่ประสบอุบัติเหตุลักษณะเดียวกันเพื่อดำเนินการต่าง ๆ ตามนโยบายของบริษัท

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

1.1.5

ตัวอย่างการนำเหมืองข้อมูลไปใช้งาน

การตลาด

- การทำนายผลการตอบสนองกับการเปิดตัวสินค้าใหม่
- การทำนายยอดขายเมื่อมีการลดราคาสินค้า
- การทำนายกลุ่มลูกค้าที่น่าจะใช้สินค้าของเรา

การเงินการธนาคาร

- การคาดการณ์ถึงโอกาสในการชำระหนี้ของลูกค้าว่าสูงเท่าไร?
- ค้นหาลูกค้าขาดคุณภาพ เพื่อหลีกเลี่ยงความเสี่ยงในการปล่อยกู้
- ค้นหาลูกค้าชั้นดี เพื่อเสนอการปล่อยกู้
- ทำนายแนวโน้มของพฤติกรรมการใช้บัตรเครดิต

สถานีโทรทัศน์หรือวิทยุ

- ค้นหารายการที่ดีและเหมาะสมต่อช่วงเวลาที่สุด เพื่อวางแผนผังรายการในแต่ละเดือน

ฮาร์ดแวร์และซอฟต์แวร์คอมพิวเตอร์

- ค้นหาช่วงเวลาที่เหมาะกับการผลิตฮาร์ดคอมพิวเตอร์ตัวใหม่ เพื่อป้อนสู่ตลาด
- การทำนายอายุการใช้งานของ Disk Drive หรือ อุปกรณ์ต่าง ๆ

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>



บทที่ 1

1.2 ความสำคัญของเหมืองข้อมูล

1.2.1

ทำไมจึงต้องมี เหมืองข้อมูล

บทที่ 1

ข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลหากเก็บไว้เฉย ๆ ก็จะไม่เกิดประโยชน์ดังนั้นจึงต้องมีการสกัดสารสนเทศ หรือการคัดเลือกข้อมูลออกมาใช้งานส่วนที่เราต้องการ

ในอดีตเราได้ใช้คนเป็นผู้สืบค้นข้อมูลต่าง ๆ ในฐานข้อมูลซึ่งผู้สืบค้นจะทำการสร้างเงื่อนไขขึ้นมาตาม ภูมิปัญญาของผู้สืบค้น (SQL: Select * FROM WHERE)

ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูลเดียวอาจไม่ให้ความรู้เพียงพอและลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูงและมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นที่จะต้องรวบรวมฐานข้อมูล หลาย ๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่า “ คลังข้อมูล ” (Data Warehouse) ดังนั้นเราจึงจำเป็นต้องใช้ Data Mining ในการดึงข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะนำข้อมูลนั้นมาใช้งานให้เกิดประโยชน์สูงสุด

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

วัตถุประสงค์ในการใช้เหมืองข้อมูล

1. เพื่อการค้นพบองค์ความรู้ใหม่ในฐานข้อมูล (Knowledge discovery in databases)
2. เพื่อการสกัดองค์ความรู้ที่ซ่อนเร้นอยู่ (Knowledge extraction)
3. เพื่อจัดการกับข้อมูลในอดีต (Data archeology)
4. เพื่อสำรวจข้อมูล (Data exploration)
5. เพื่อค้นหา Pattern ของข้อมูลที่ซ่อนอยู่ (Data pattern processing)
6. เพื่อใช้ขุดเจาะข้อมูล (Data dredging)
7. เพื่อเก็บเกี่ยวผลประโยชน์ให้ได้มาซึ่งสารสนเทศที่มีประโยชน์

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

เป้าหมายหลักของการทำเหมืองข้อมูล

คุณลักษณะและเป้าหมายหลักของ Data Mining คือ ใช้สกัดหรือค้นหา Pattern ของข้อมูลที่ฝังลึกและซ่อนเร้นอยู่ภายในฐานข้อมูลขนาดใหญ่ โดยใช้สถาปัตยกรรม Client-Server (Client/server architecture) ใช้เครื่องมือสมัยใหม่ที่สามารถแสดงผลแบบกราฟฟิก (Visualization) ผู้ใช้สามารถดูข้อมูลแบบเจาะลึก (data drills) และสามารถใช้เครื่องมือในการสอบถามข้อมูลได้อย่างง่ายดาย โดยไม่ต้องอาศัยความชำนาญของ programmer บ่อยครั้งเราอาจค้นพบผลลัพธ์ที่เราไม่คาดหวังมาก่อน เครื่องมือจะทำให้เราใช้งานได้ง่าย ซึ่งเครื่องมือนอกจากจะแสดงผลกราฟิกได้แล้วยังรวม Spreadsheets (Excel , Google sheet) เอาไว้ด้วย

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

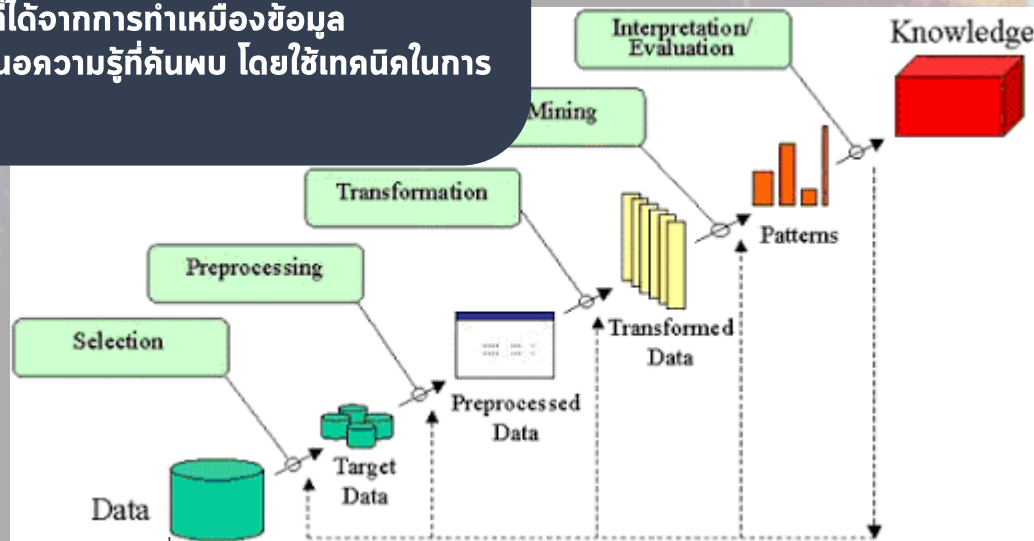
บทที่ 1
1.3 ขั้นตอนการทำเหมืองข้อมูล

1.3.1

ขั้นตอนการทำเหมืองข้อมูล

บทที่ 1

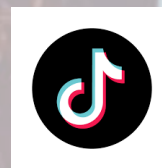
- 1) Data Cleaning เป็นขั้นตอนสำหรับการตัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- 2) Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- 3) Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- 4) Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- 5) Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- 6) Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- 7) Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ



Google form → การทำแบบสำรวจ , แบบสอบถาม

Google sheet → Internet , website , database

Social media → line , facebook , tiktok...



1000 รายการ
(การเรียนการสอน)

10000 รายการ
(งานวิจัย)

100000 รายการ
(งานวิจัย)

ประวัติของนักศึกษา ย้อยหลัง 10 ปี

Data Mining จะเป็นการสังเคราะห์ข้อมูลอย่างละเอียดจากฐานข้อมูลขนาดใหญ่(Big Data) หรืออาจวิเคราะห์มาจากรายการ Transaction โดยเรียนรู้ข้อมูลจากอดีต หรือปัจจุบันผลลัพธ์ที่ได้จากการสังเคราะห์ของ Data Mining อาจจะเป็นข้อมูลแบบ Unknow , Valid, หรือ Actionable ซึ่งความหมายของข้อมูลทั้ง 3 ประเภทนี้ มีดังนี้

1. ข้อมูลแบบ Unknow
2. ข้อมูลแบบ Valid
3. ข้อมูลแบบ Actionable

1. ข้อมูลแบบ Unknow เป็นข้อมูลที่ผู้ใช้งานไม่เคยรู้มาก่อน ไม่ชัดเจน ไม่สามารถตั้งสมมติฐานล่วงหน้าว่าจะเป็นแบบใด เช่น

Ex : ห้างสรรพสินค้าแห่งหนึ่งค้นพบพฤติกรรมของผู้บริโภค ที่พ่อบ้านมักซื้อเบียร์และผ้าอ้อมในวันศุกร์ตอนเย็น ดังนั้นเป็นสัญญาณให้เจ้าของกิจการควรจะเตรียมสินค้าไว้เพื่อจำหน่าย ในขณะที่ห้างคู่แข่งอาจจะไม่รู้ข้อมูลเหล่านี้

Ex: เจ้าของร้านขายรถยนต์พบว่ารถยนต์ขนาดใหญ่ ราคาแพงมักจะถูกซื้อโดยคนที่สูงอายุ ซึ่งเจ้าของร้านไม่เคยรู้มาก่อน แต่ข้อมูลดังกล่าวไม่เป็นลักษณะของ Unknow เพราะสมมติฐานดังกล่าวมีอยู่ คือ คนที่มีอายุมักมีฐานะดีขึ้น เมื่อเทียบกับคนในวัยที่อายุน้อย

2. ข้อมูลแบบ Valid เมื่อผู้ใช้เริ่มใช้เทคนิคของ Data Mining จะค้นพบสิ่งที่น่าสนใจตลอดเวลา แต่จะต้องพิจารณาด้วยว่าสิ่งนั้นถูกต้อง (Valid) หรือไม่ เช่น มักจะพบว่ามีความสัมพันธ์ของการซื้อสินค้า 2 อย่างเสมอ เมื่อจำนวนความหลากหลายของสินค้ามากขึ้น แต่ไม่ได้หมายความว่าต้องให้ห้างสรรพสินค้า เก็บสินค้าในคลังมากขึ้น เพราะข้อมูลที่ได้ อาจเกิดความคลาดเคลื่อน เพราะฉะนั้นจะต้องทำการตรวจสอบความถูกต้อง (Validation and Checking) ของข้อมูลและวิเคราะห์ความถูกต้องอีกครั้ง

ข้อมูลแบบ Valid มีความสำคัญต่อการทำเหมืองข้อมูล เนื่องจากข้อมูลแบบ Valid จะช่วยให้กระบวนการทำเหมืองข้อมูลเป็นไปอย่างมีประสิทธิภาพและถูกต้องมากขึ้น ตัวอย่างเช่น หากข้อมูลตัวเลขในชุดข้อมูลไม่ถูกต้อง จะทำให้การวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลต่างๆ ผิดพลาดได้ หากข้อมูลตัวอักษรในชุดข้อมูลไม่ถูกต้อง จะทำให้การค้นหาคำสำคัญในชุดข้อมูลผิดพลาดได้ เป็นต้น

การได้มาซึ่งข้อมูล/แหล่งข้อมูล

ในการเตรียมข้อมูลสำหรับการทำเหมืองข้อมูล จึงจำเป็นต้องตรวจสอบความถูกต้องของข้อมูลก่อนเสมอ โดยใช้เครื่องมือหรือวิธีการตรวจสอบความถูกต้องของข้อมูลที่เหมาะสมกับประเภทของข้อมูล

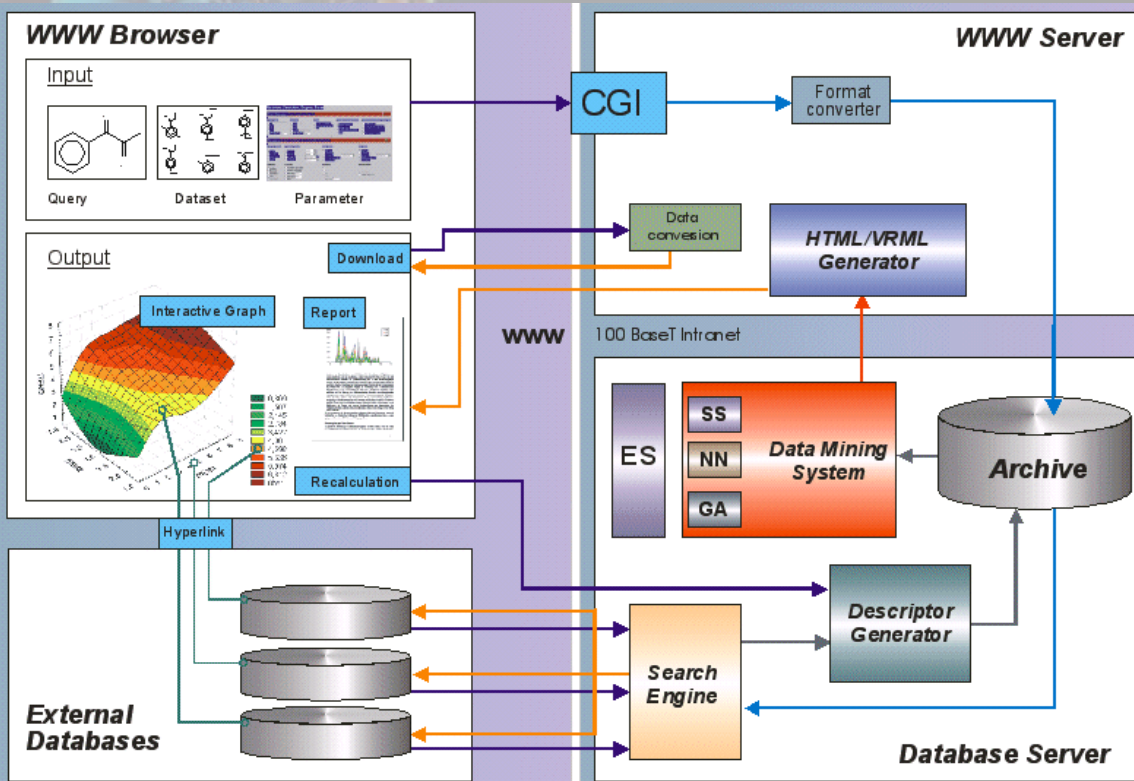
- ตัวอย่างข้อมูลแบบ Valid คือข้อมูลที่ถูกต้อง ครบถ้วน และไม่ขัดแย้งกับข้อมูลอื่นๆ ในชุดข้อมูลเดียวกัน ตัวอย่างข้อมูลแบบ Valid อาจมีรูปแบบดังนี้
- ข้อมูลตัวเลขต้องเป็นตัวเลขที่ถูกต้อง เช่น ตัวเลขเต็ม ทศนิยม หรือเลขศูนย์
- ข้อมูลตัวอักษรต้องเป็นตัวอักษรที่ถูกต้อง เช่น ตัวอักษรภาษาอังกฤษ ตัวอักษรไทย ตัวเลข หรือสัญลักษณ์พิเศษ
- ข้อมูลวันที่ต้องเป็นวันที่ที่ถูกต้อง เช่น รูปแบบ dd/mm/yyyy หรือ yyyy-mm-dd
- ข้อมูลเวลาต้องเป็นเวลาที่ถูกต้อง เช่น รูปแบบ hh:mm:ss หรือ hh:mm
- ข้อมูลที่อยู่ต้องเป็นที่อยู่ที่มีอยู่จริง เช่น ถนน ตำบล อำเภอ จังหวัด ประเทศ
- ข้อมูลอื่นๆ ต้องเป็นข้อมูลที่ถูกต้องตามเกณฑ์ที่กำหนด

3. ข้อมูลแบบ Actionable : ข้อมูลจะต้องถูกแปลงออกมาและนำมาตัดสินใจ เพื่อสร้างความได้เปรียบในเชิงธุรกิจ บางครั้งข้อมูลที่เรากันพบเป็นสิ่งที่คู่แข่งได้ทำไปแล้ว (เราช้าไป) หรืออาจผิดกฎหมาย ซึ่งจะต้องมีวิจารณญาณในการใช้ด้วย บางทีข้อมูลดังกล่าวอาจไม่มีประโยชน์อะไร

1.3.2

การได้มาซึ่งข้อมูล/แหล่งข้อมูล

บทที่ 1



1.3.3

กระบวนการในการทำ data mining

บทที่ 1



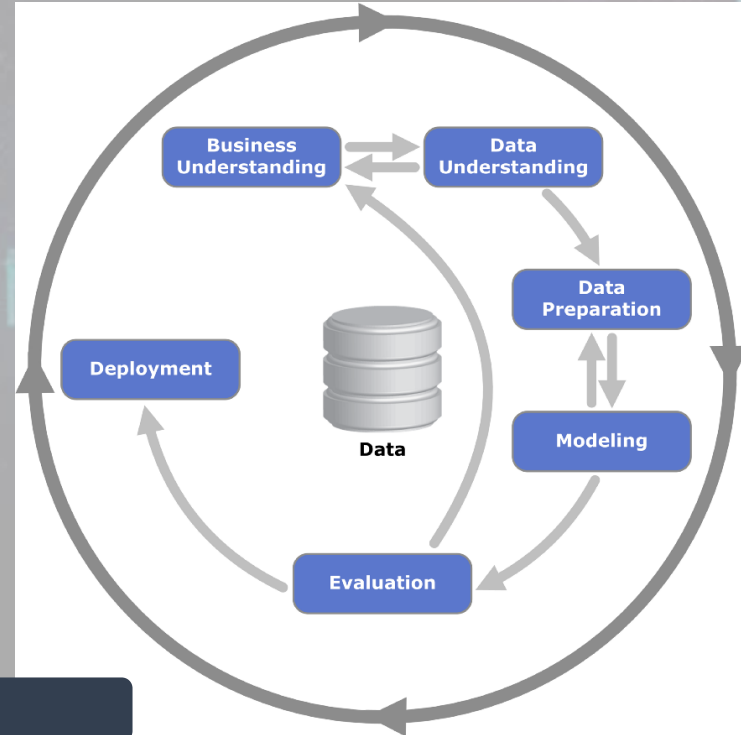
<https://nutdnuy.medium.com/เหมืองข้อมูล-data-mining-c21c64ae6e3b>

1.3.3

กระบวนการในการทำ data mining

บทที่ 1

ทิศทาง Data mining จะเริ่มขยายไปนอก database หรือที่เรียกว่า data science



<https://nutdnuy.medium.com/เหมืองข้อมูล-data-mining-c21c64ae6e3b>

บทที่ 1

1.4 ส่วนประกอบการทำเหมืองข้อมูล

ส่วนประกอบการทำเหมืองข้อมูล

- 1) แหล่งเก็บข้อมูล Database, Data Warehouse, World Wide Web และ Other Info Repositories (XML) เป็นแหล่งข้อมูลสำหรับการทำเหมืองข้อมูล
- 2) Database Management System หรือ Data Warehouse Server ทำหน้าที่นำเข้าข้อมูลตามคำขอของผู้ใช้
- 3) Knowledge Base ได้แก่ ความรู้เฉพาะด้านในงานที่จะเป็นประโยชน์ต่อการสืบค้น หรือประเมินความน่าสนใจของรูปแบบผลลัพธ์ที่ได้
- 4) Data Mining Engine เป็นส่วนประกอบหลักประกอบด้วยโมดูลที่รับผิดชอบงานทำเหมืองข้อมูลประเภทต่าง ๆ ได้แก่ การหากฎความสัมพันธ์ การจำแนกประเภท การจัดกลุ่ม ฯลฯ
- 5) Pattern Evaluation Module ทำงานร่วมกับ Data Mining Engine โดยใช้มาตรวัดความน่าสนใจในการกลั่นกรองรูปแบบผลลัพธ์ที่ได้ เพื่อให้การค้นหามุ่งเน้นเฉพาะรูปแบบที่น่าสนใจ
- 6) User Interface ส่วนติดต่อประสานระหว่างผู้ใช้กับระบบการทำเหมืองข้อมูล ช่วยให้ผู้ใช้สามารถระบุงานทำเหมืองข้อมูลที่ต้องการทำ ดูข้อมูลหรือโครงสร้างการจัดเก็บข้อมูล ประเมินผลลัพธ์ที่ได้

ลักษณะเฉพาะของข้อมูลที่มีการใช้ทำเหมืองข้อมูล

- 1) **ข้อมูลขนาดใหญ่(Big data)** เกินกว่าจะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า หรือโดยใช้ Database Management System (DBMS) ในการจัดการฐานข้อมูล
- 2) **ข้อมูลที่มาจากหลายแหล่ง** โดยอาจรวบรวมมาจากหลายระบบปฏิบัติการหรือหลาย DBMS เช่น Oracle , DB2 , MS SQL , MS Access เป็นต้น มากจาก internet
- 3) **ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดเวลาที่ทำการ Mining** หากข้อมูลที่มีอยู่นั้นเป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาจะต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้นไว้และนำฐานข้อมูลที่บันทึกไว้มาทำ Mining แต่เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลา จึงทำให้ผลลัพธ์ที่ได้จากการทำ Mining สมเหตุสมผลในช่วงเวลาหนึ่งเท่านั้น ดังนั้นเพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องเหมาะสมอยู่ตลอดเวลาจึงต้องทำ Mining ใหม่ทุกครั้งในช่วงเวลาที่เหมาะสม
- 4) **ข้อมูลที่มีโครงสร้างซับซ้อน** เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย ข้อมูลเหล่านี้สามารถนำมาทำ Mining ได้เช่นกันแต่ต้องใช้เทคนิคการทำ Data Mining ขั้นสูง

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

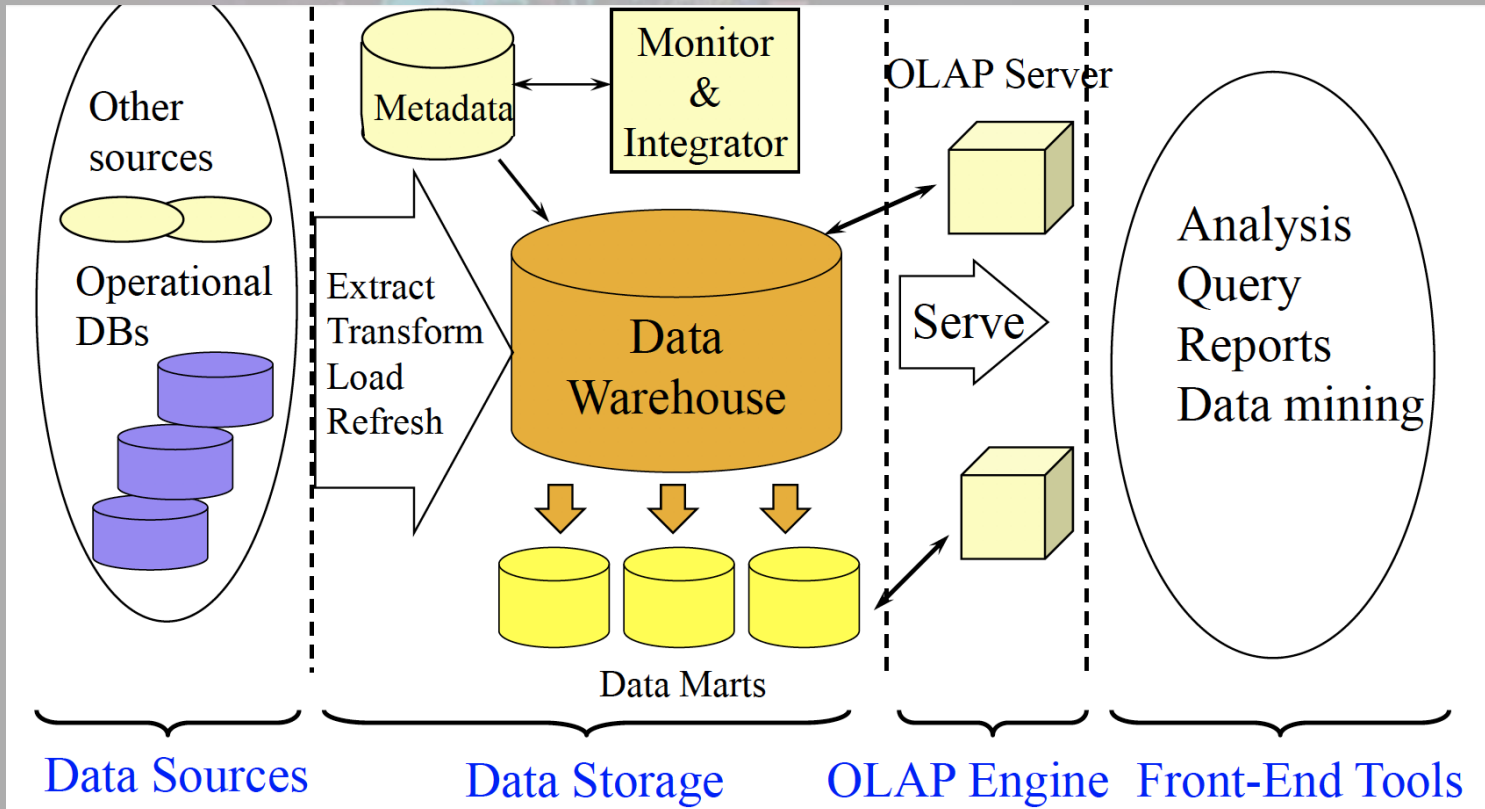
ประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล

- 1) **Relational Database** เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย Entity Relationship Model
- 2) **Data Warehouses** เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่ ๆ เดียวกัน
- 3) **Transactional Database** ประกอบด้วยข้อมูลที่แต่ละทรานแซกชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปชื้อลูกค้าและรายการสินค้าที่ลูกค้ารายชื้อ
- 4) **Advanced Database** เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่นๆ เช่น ข้อมูลแบบ Object-Oriented ข้อมูลที่เป็น Text File ข้อมูลมัลติมีเดีย ข้อมูลในรูปของ Web

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

1.4.4

ประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล : Data Warehouse



บทที่ 1

1.5 เทคนิคในการทำเหมืองข้อมูล

1.5.1

เทคนิคในการทำเหมืองข้อมูล

บทที่ 1

1. การวิเคราะห์ข้อมูลเชิงสถิติในงานเหมืองข้อมูล

บทที่ 4

การวิเคราะห์ข้อมูลเชิงสถิติในงานเหมืองข้อมูล

2. กฎความสัมพันธ์ (Association rule)

บทที่ 5

กฎความสัมพันธ์ (Association rule)

3. การจำแนกข้อมูล (Data classification)

บทที่ 7-9

Decision Tree, SVM , KNN , NN , Naive Bays

4. การแบ่งกลุ่มข้อมูล (Data clustering)

บทที่ 10

K-Mean, Hierarchical Clustering

5. การสร้างมโนภาพ (Visualization)

บทที่ 10

K-Mean, Hierarchical Clustering

1.5.1

เทคนิคในการทำเหมืองข้อมูล

บทที่ 1

กฎความสัมพันธ์ (Association rule)

แสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุ ที่เกิดขึ้นพร้อมกัน ตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยง เช่น การวิเคราะห์ข้อมูล การขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (POS) หรือร้านค้าออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่า คนที่ซื้อเทปวีดีโอ มักจะซื้อเทปขาวด้วย ร้านค้าก็อาจจะจัดร้านให้สินค้าสองอย่างอยู่ใกล้กัน เพื่อเพิ่มยอดขาย หรืออาจจะพบว่าหลังจากคนซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็สามารถนำความรู้นี้ไปแนะนำผู้ที่กำลังจะซื้อหนังสือ ก ได้

การจำแนกประเภทข้อมูล (Data classification)

คือ หากกฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หากความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้หรือไม่เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้

แหล่งที่มา: <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

1.5.1

เทคนิคในการทำเหมืองข้อมูล**บทที่ 1****การแบ่งกลุ่มข้อมูล (Data clustering)**

คือ แบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน

การสร้างมโนภาพ (Visualization)

คือ สร้างภาพคอมพิวเตอร์กราฟิกที่สามารถนำเสนอข้อมูลมากมายอย่างครบถ้วนแทนการใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ้อนเร้นเมื่อดูข้อมูลชุดนั้นด้วยจินตทัศน์

1.5.2

กฎความสัมพันธ์ (Association rule)

บทที่ 1

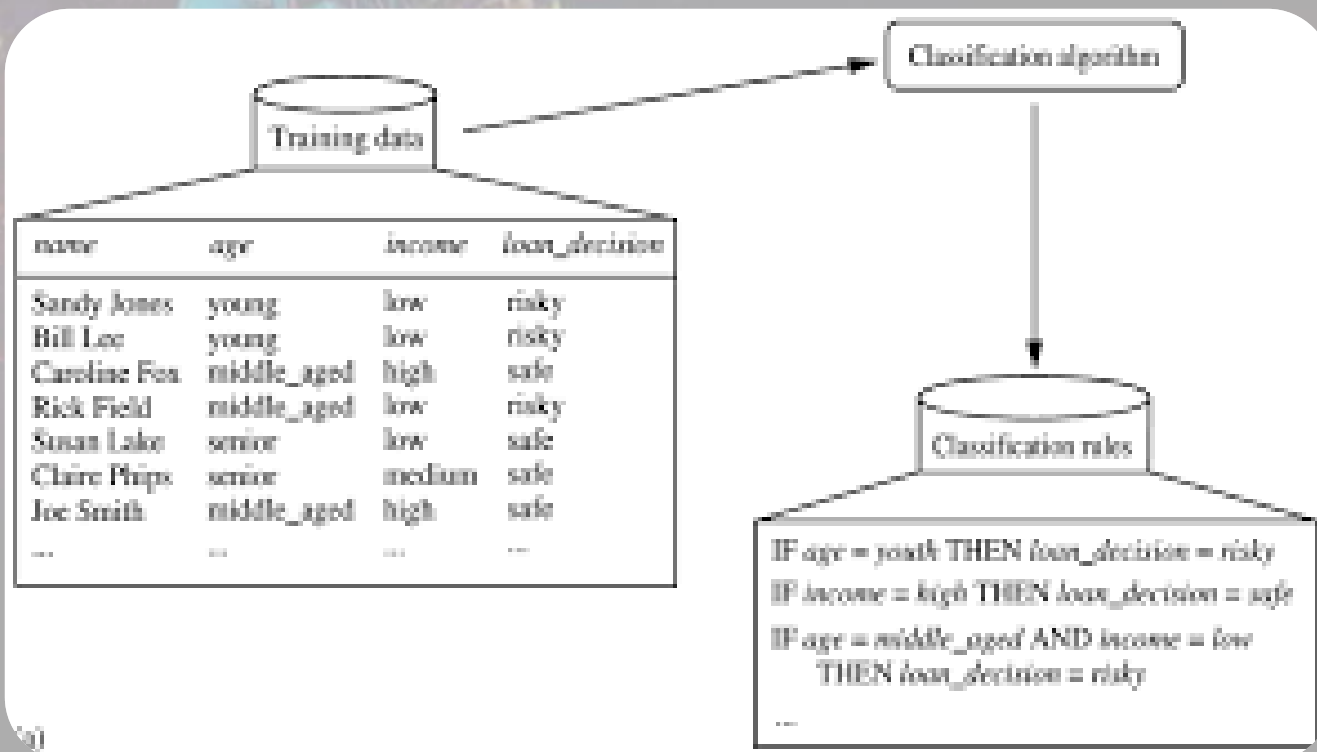
ตะกร้าที่ 1 	  
ตะกร้าที่ 2 	  
ตะกร้าที่ 3 	  
ตะกร้าที่ 4 	 

Transaction ID	Items
1	Apple,Cereal,Diapers
2	Beer,Cereal,Eggs
3	Apple,Beer,Cereal,Eggs
4	Beer,Eggs



Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B \& C \Rightarrow D$	1/5	1/3	5/9

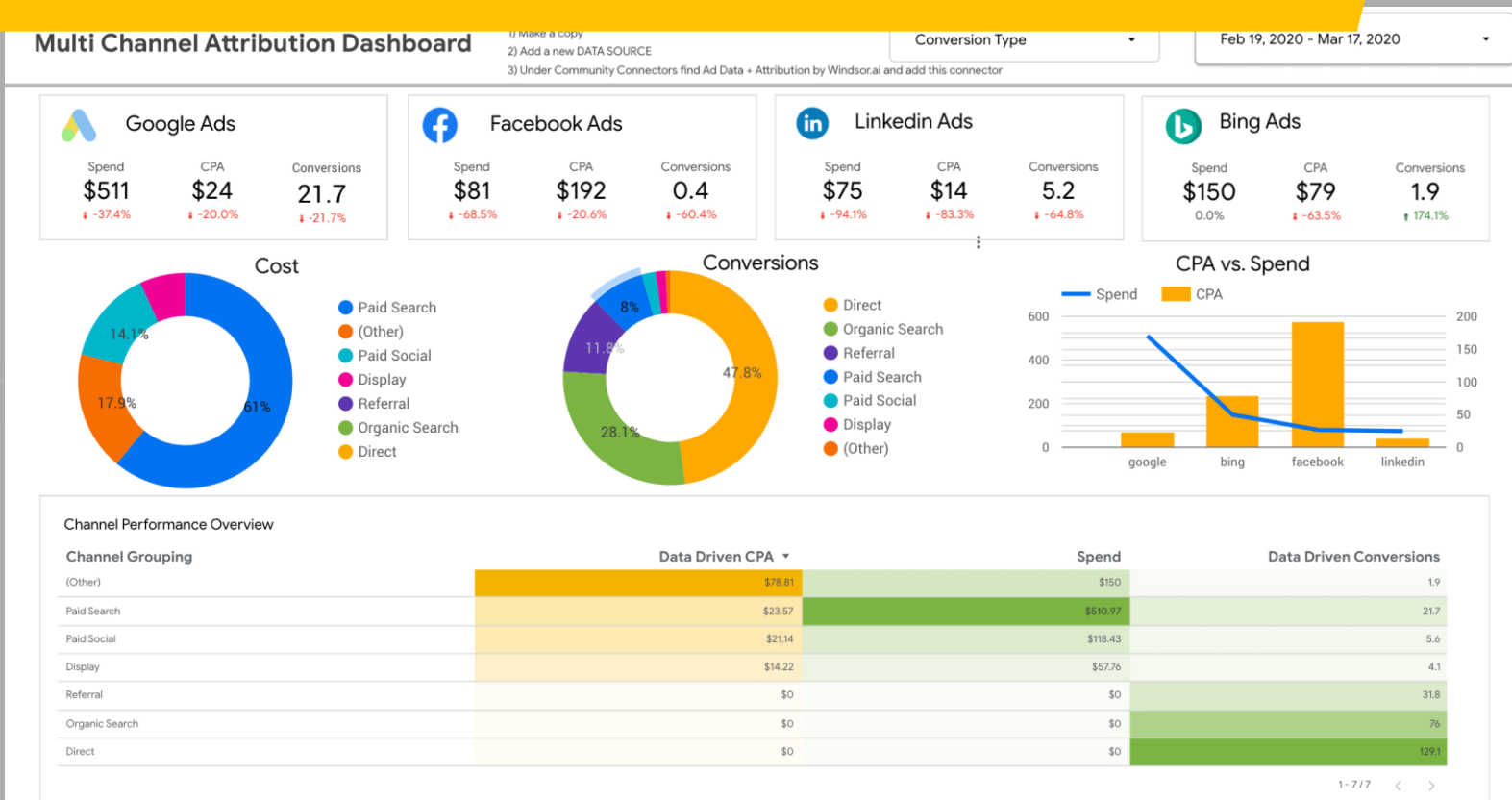
การจำแนกประเภทข้อมูล (Data classification)



1.5.4

การสร้างมโนภาพ (Visualization)

บทที่ 1



1

การทำเหมืองข้อมูลเว็บ (Web Mining)

การทำเหมืองข้อมูลเว็บคือการใช้เทคนิคการทำเหมืองข้อมูลเพื่อค้นหาและสกัดข้อมูล และสารสนเทศจากเอกสารเว็บและบริการบนเว็บโดยอัตโนมัติ เพื่อนำความรู้ที่ได้มาแก้ปัญหาที่ต้องการทั้งทางตรงและทางอ้อม นอกจากนี้ยังได้แบ่งประเภทของการทำเหมืองข้อมูลเว็บโดยพิจารณาจากข้อมูลที่นำมาวิเคราะห์ออกเป็น 3 ประเภท คือ Web Content Mining, Web Structure Mining และ Web Usage Mining

Web Content Mining เป็นการค้นหาข้อมูลที่มีประโยชน์จากข้อมูลที่อยู่ภายในเว็บ เช่น ข้อความ รูปภาพ เป็นต้น โดย Web Content Mining สามารถแบ่งออกเป็น 2 ประเภทตามมุมมองคือ มุมมองทางด้านการสืบค้นสารสนเทศ (Information Retrieval) และมุมมองทางด้านฐานข้อมูล (Database) สำหรับเป้าหมายของ Web Content Mining จากมุมมองของการสืบค้นสารสนเทศคือการทำเหมืองข้อมูลเว็บเพื่อปรับปรุงการหาข้อมูลหรือกรองข้อมูลให้ผู้ใช้โดยพิจารณาจากข้อมูลที่ผู้ใช้อ้างอิงหรือร้องขอ ในขณะที่เป้าหมายของ Web Structure Mining ในมุมมองของฐานข้อมูลส่วนใหญ่พยายามจำลองข้อมูลบนเว็บและรวมข้อมูลนั้นเพื่อให้การสอบถามทำงานดีขึ้นมากกว่าการใช้คำหลักเป็นตัวค้นหาเพียงอย่างเดียว

Web Structure Mining เป็นวิธีการที่พยายามค้นหารูปแบบโครงสร้างการเชื่อมโยงที่สำคัญและซ่อนอยู่ ในเว็บ ซึ่งรูปแบบนี้จะขึ้นอยู่กับรูปแบบการเชื่อมโยงเอกสารภายในเว็บ โดยนำรูปแบบที่ได้มาใช้ในการจัดกลุ่มเว็บเพจและใช้สร้างข้อมูลสารสนเทศที่เป็นประโยชน์ เช่น นำมาใช้ในการปรับโครงสร้างของเว็บให้สามารถให้บริการผู้ใช้ได้อย่างรวดเร็ว

1

การทำเหมืองข้อมูลเว็บ (Web Mining) ต่อ...

Web Usage Mining เป็นวิธีการที่พยายามค้นหาความหมายของข้อมูลที่สร้างจากช่วงการทำงานหนึ่งของผู้ใช้หรือสร้างจากพฤติกรรมของผู้ใช้เรียกอีกชื่อหนึ่งว่า Web Log Mining โดยในขณะที่ Web Content Mining และ Web Structure Mining ใช้ประโยชน์จากข้อมูลจริง หรือข้อมูลพื้นฐานบนเว็บ แต่ Web Usage Mining ทำการค้นหาคำตอบจากข้อมูลการติดต่อสื่อสารระหว่างกันของผู้ใช้ที่ติดต่อกับเว็บ โดย Web Usage Mining ทำการรวบรวมข้อมูลจากบันทึกในการดำเนินการต่างๆ เช่น บันทึกการใช้งานของ Proxy (Proxy Server Log) ข้อมูลการลงทะเบียน (Registration Data) หรือข้อมูลอื่นอันเป็นผลจากการทำงานร่วมกันมาใช้วิเคราะห์ ดังนั้น Web Usage Mining จึงเป็นวิธีการทำงานที่เน้นใช้เทคนิคที่สามารถทำนายพฤติกรรมของผู้ใช้ใน ขณะที่ผู้ใช้ทำงานกับเว็บ กระบวนการทำงานของ Web Usage Mining สามารถแบ่งออกเป็น 2 วิธีคือ

1. ทำการจับคู่ข้อมูลการใช้งานของเครื่องให้บริการเว็บให้อยู่ในรูปแบบของตาราง ความสัมพันธ์ ก่อนที่นำข้อมูลนี้มาปรับใช้กับเทคนิคการทำเหมืองข้อมูลการใช้งานเว็บ
2. ใช้ประโยชน์จากข้อมูลในบันทึกการใช้งานโดยตรงซึ่งจะใช้เทคนิคการเตรียม ข้อมูล (Preprocessing) เพื่อเตรียมข้อมูลก่อนหาคำตอบความสัมพันธ์ (Pattern Discovery) และวิเคราะห์รูปแบบ (Pattern Analysis)

1

การทำเหมืองข้อมูลเว็บ (Web Mining) ต่อ...

Statistical Analysis การวิเคราะห์ทางด้านสถิติ เป็นวิธีการที่จะค้นหาความรู้เกี่ยวกับผู้เยี่ยมชมเว็บไซต์ โดยการวิเคราะห์ส่วนของไฟล์ เป็นวิธีหนึ่งที่สามารถบอกชนิดความแตกต่างเกี่ยวกับการวิเคราะห์ทางสถิติบน ตัวแปร เช่น หน้าที่ถูก เวลาการดูและความยาวของเส้นทางเว็บไซต์ เครื่องมือการวิเคราะห์การจราจรเว็บส่วนมากจะรายงานข้อมูลเกี่ยวกับสถิติ เช่น ความถี่ของหน้าที่เข้าถึงบ่อยๆ เวลาดูเฉลี่ยของหน้าหรือความยาวเฉลี่ยของเส้นทางเว็บไซต์ รายงานนี้อาจจะรวมถึงการวิเคราะห์ข้อผิดพลาดที่จำกัด เช่น การสับหารายการที่ไม่ให้สิทธิ์เข้าถึงหรือการค้นหา url ที่ไม่สมบูรณ์ รูปแบบของความรู้สามารถเป็นประโยชน์อย่างมากสำหรับการปรับปรุงระบบ การเพิ่มความปลอดภัยของระบบ ทำให้สะดวกในการปรับปรุงเว็บไซต์และการสนับสนุนการตัดสินใจทางการตลาด

Association Rule เป็นการค้นหากฎความสัมพันธ์ของข้อมูล โดยค้นหาความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าสองชุดขึ้นไปไว้ด้วยกัน ความสำคัญของกฎทำการวัดโดยใช้ข้อมูลสองตัวด้วยกันคือค่าสนับสนุน (Support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้ หรือเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎที่ใช้มีความถูกต้อง และข้อมูลตัวที่สองที่นำมาใช้วัดคือค่าความมั่นใจ (Confidence) ซึ่งเป็นจำนวนของกรณีที่ถูกถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่กฎสามารถ นำไปใช้ได้ ในการหากฎความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ อัลกอริทึม Apriori

Clustering คือการจัดกลุ่มข้อมูลซึ่งมีลักษณะคล้ายกับการแบ่งประเภทแต่จะไม่เหมือนกัน โดยการแบ่งประเภทจะวิเคราะห์ข้อมูลตามต้นแบบ แต่สำหรับการแบ่งกลุ่มเป็นการวิเคราะห์โดยไม่พิจารณาจัดกลุ่มตามประเภทที่มีหรือที่รู้จัก แต่จะใช้ขั้นตอนวิธีการจัดกลุ่มเพื่อค้นหากลุ่มที่สามารถยอมรับได้เพื่อจัด เข้ากลุ่ม กล่าวคือ กลุ่มของวัตถุมีการสร้างขึ้นโดยเปรียบเทียบกับวัตถุที่มีความเหมือนกันจัดเข้า กลุ่มเดียวกัน

1

การทำเหมืองข้อมูลเว็บ (Web Mining) ต่อ...

Classification เป็นการจัดแบ่งประเภทของข้อมูล โดยหาชุดต้นแบบหรือชุดของการทำงานที่อธิบายและแบ่งประเภทข้อมูล วัตถุประสงค์เพื่อให้สามารถใช้เป็นต้นแบบทำนายประเภทของวัตถุหรือข้อมูลที่ไม่มีการระบุประเภทหรือชนิดของข้อมูล ซึ่งต้นแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยอาจจะเป็นกลุ่มข้อมูลที่มีการระบุประเภทหรือกลุ่มเรียบร้อยแล้ว รูปแบบของต้นแบบแสดงได้หลายแบบเช่น Classification Rules, Decision Trees หรือ Neural Networks เป็นต้น

Sequential Pattern เป็นเทคนิคการสืบค้นรูปแบบลำดับเหตุการณ์ เป็นการหาความสัมพันธ์ของข้อมูลระหว่างทรานแซกชันซึ่งทำให้มีเวลาเข้ามา เกี่ยวข้อง ด้วยรูปแบบลำดับที่น่าสนใจ จะแสดงว่าถ้าเกิดเหตุการณ์นี้แล้วหรือพบกลุ่มของข้อมูลชุดนี้แล้วมีแนวโน้ม ที่จะเกิดเหตุการณ์อะไรหรือพบกลุ่มของข้อมูลใดตามมาจากภายหลัง

Dependency Modeling เป็นแบบจำลองการพึ่งพา เป็นการค้นพบรูปแบบอื่นๆ ที่มีประโยชน์ในการทำเหมืองข้อมูลเว็บ เป้าหมายคือการพัฒนาต้นแบบที่สามารถทำได้ของตัวแทนที่สำคัญของการพึ่งพาท่าม กลางตัวแปรต่างๆ ในโดเมนเว็บ เช่น อาจจะใช้วิธีการสร้างแบบจำลองสถานที่แตกต่างกัน ให้ผู้เยี่ยมชมได้เห็นในขณะที่เลือกซื้อสินค้าทางออนไลน์ มีหลายเทคนิคการเรียนรู้ที่เป็นไปได้ สามารถถูกใช้เป็นแบบของการค้นหาพฤติกรรมของผู้ใช้ เช่น ตัวแบบฮิดเดนมาร์คอฟ (Hidden Markov Model) และแนวคิดข่ายงานความเชื่อเบย์ (Bayesian Belief Network)

ขั้นตอนการทำเหมืองข้อมูลการใช้เว็บไซต์

กระบวนการทำเหมืองข้อมูลการใช้เว็บไซต์ประกอบด้วยกระบวนการ 3 ขั้นตอนด้วยกันคือ

1. การเตรียมข้อมูล (Preprocessing) เป็นขั้นตอนการแปลงข้อมูลเว็บให้อยู่ในรูปแบบที่เหมาะสมสำหรับการค้นหาความสัมพันธ์ เช่น Data Cleaning, Data Filtering
2. การค้นหารูปแบบ (Pattern Discovery) เป็นขั้นตอนที่นำเทคนิคการทำเหมืองข้อมูล เช่น Association Rule Discovery, Sequential Pattern Discovery เพื่อค้นหาความสัมพันธ์หรือรูปแบบจากข้อมูลเว็บที่ผ่านขั้นตอนการเตรียมข้อมูลแล้ว
3. การวิเคราะห์รูปแบบ (Pattern Analysis) เป็นรูปแบบการนำผลลัพธ์จากการค้นหามาทำการวิเคราะห์เพื่อช่วยในการตัดสินใจ หรือการวางแผนทางธุรกิจ

ขั้นตอนการทำเหมืองข้อมูลการใช้เว็บไซต์

ข้อมูลที่ใช้ในการทำเหมืองข้อมูลเว็บไซต์ สามารถนำมาจาก 3 แหล่งด้วยกันดังนี้

1. การรวบรวมข้อมูลจากบันทึกการเข้าใช้งานเครื่องให้บริการเว็บ (Server-Side)
2. การรวบรวมข้อมูลจากบันทึกการเข้าใช้งานเครื่องใช้บริการเว็บ (Client-Side)
3. การรวบรวมข้อมูลจากบันทึกของเครื่องให้บริการการเป็นตัวแทนเชื่อมต่อ (Proxy Server)

การประยุกต์ใช้ Data Mining จะมีได้หลากหลายแต่สามารถจัดกลุ่มกว้างๆ ได้เป็นสองกลุ่ม คือกลุ่มที่ใช้เพื่อการทำนายและกลุ่มที่ใช้เพื่อการอธิบาย

การทำ Data Mining เพื่อการทำนาย เป็นการนำความรู้ที่เรียนรู้มาจากข้อมูลที่มีอยู่เพื่อประโยชน์ในการทำนาย ข้อมูลใหม่ที่จะเกิดขึ้นในอนาคต เช่น จากข้อมูลลูกค้าของแผนกสินเชื่อของธนาคารที่ได้มีการจัดลำดับชั้นของลูกค้า ไว้แล้วว่าใครเป็นลูกค้าชั้นดี ใครเป็นลูกค้าในระดับปานกลาง และใครเป็นลูกค้าที่มักจะผิดนัดชำระหนี้ Data Mining สามารถเรียนรู้จากข้อมูลเหล่านี้และค้นหาโมเดลที่สามารถใช้อธิบายลักษณะของ ลูกค้าชั้นดี ลูกค้าระดับปานกลาง และลูกค้าที่ไม่เป็นที่ต้องการ จากโมเดลที่ได้นี้สามารถนำไปใช้ทำนายลูกค้าใหม่ที่มาขอสินเชื่อได้น่าจะ เป็นลูกค้าประเภทใด

การทำ Data Mining เพื่อการอธิบาย เป็นการค้นหารูปแบบที่น่าสนใจจากกลุ่มข้อมูล รูปแบบนี้มักจะเป็นความสัมพันธ์หรือลักษณะที่เชื่อมโยงกันของข้อมูล การทำแบบนี้ต่างจากแบบแรกตรงที่ผู้ใช้ไม่ได้กำหนดล่วงหน้าว่าจะให้โปรแกรม Data Mining ค้นหา รูปแบบหรือโมเดลของอะไร แต่ให้ค้นหาทุกรูปแบบที่น่าสนใจจากข้อมูล

แบบฝึกหัดท้ายบทที่ 1

จงตอบคำถามต่อไปนี้

1. นิยามของเหมืองข้อมูล อธิบายความหมายของเหมืองข้อมูล (Data Mining) และเปรียบเทียบกับกระบวนการวิเคราะห์ข้อมูลทั่วไป
2. วัตถุประสงค์ของเหมืองข้อมูล เหมืองข้อมูลมีความสำคัญและประโยชน์ต่อองค์กรหรือธุรกิจอย่างไร? ยกตัวอย่างอย่างน้อย 2 กรณี
3. กระบวนการเหมืองข้อมูล อธิบายกระบวนการของเหมืองข้อมูล (เช่น CRISP-DM) อย่างย่อ และอธิบายความสำคัญของแต่ละขั้นตอน
4. ประเภทของปัญหาในเหมืองข้อมูล อธิบายประเภทของปัญหาในเหมืองข้อมูล เช่น การจำแนกประเภท (Classification) การจัดกลุ่ม (Clustering) และการทำนาย (Prediction) พร้อมยกตัวอย่างการใช้งานจริง
5. เครื่องมือที่ใช้ในเหมืองข้อมูล ระบุและอธิบายเครื่องมือหรือซอฟต์แวร์ที่นิยมใช้ในเหมืองข้อมูล พร้อมบอกจุดเด่นของแต่ละเครื่องมือ
6. จริยธรรมในเหมืองข้อมูล คุณคิดว่าการใช้เหมืองข้อมูลมีข้อกังวลด้านจริยธรรมหรือผลกระทบทางสังคมอย่างไร?