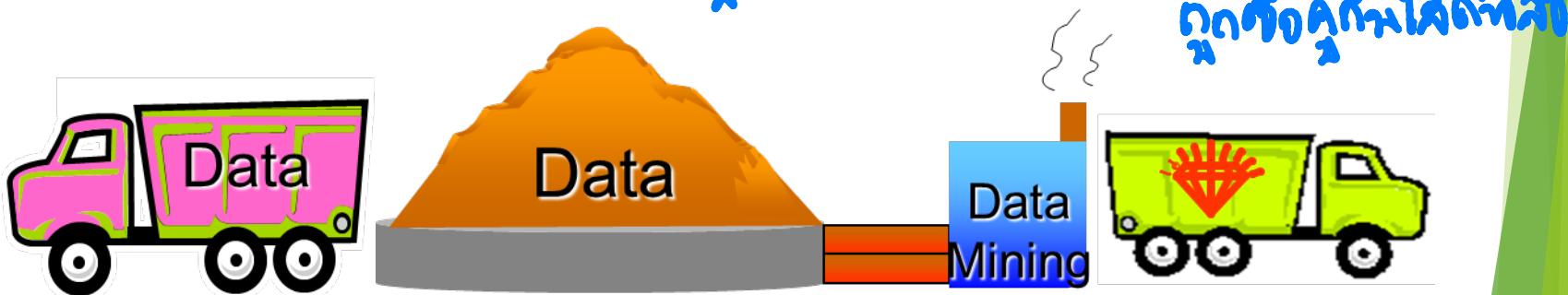


FOUNDATIONS OF BUSINESS INTELLIGENCE:

Data Mining

Data Mining អត្ថបទនៃវិធានការសាស្ត្រ

- ▶ เป็นกระบวนการค้นหาสารสนเทศอุปมาจากฐานข้อมูลขนาดใหญ่ เพื่อ
นำความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ **เจ้าของภาษา, ไมโครซอฟท์**
 - ▶ เป็นกระบวนการ (Process) ที่กระทำกับข้อมูลขนาดใหญ่ เพื่อค้นหา
ตัวต่อตัว จึงสามารถสืบสืบทอดได้
รูปแบบของข้อมูล แนวโน้มของข้อมูล และความสัมพันธ์ที่ซ่อนอยู่ใน **ลักษณะ**
ชุดข้อมูลนี้โดยการใช้เทคนิคทาง AI, สติติ และคณิตศาสตร์ เพื่อให้ได้ **ตัวต่อตัว**
สารสนเทศที่เรามีรู้อุปมา โดยสารสนเทศที่ได้จะมีเหตุผลและสามารถ
นำไปใช้ประโยชน์ได้ **ถ้าหากว่าก็ = ให้ ก็ หัวใจคนตัวนั้น**



Example 例句 วิชาภาษาไทย ภาษาไทยที่ใช้ในภาษาอังกฤษ

พื้นที่ห้องน้ำ ห้องน้ำ

เจ้าบ้านใช้ห้องน้ำ พ่อแม่ใช้ห้องน้ำ ลาก่อน

- จัด shelf ห้องน้ำให้ดี
- ห้องน้ำ

การประยุกต์ใช้ Data Mining

1. Market Management

Purchasing Pattern over Time หรือ รูปแบบการซื้อเปลี่ยนแปลงตามเวลา ในบริบทของการตลาด หมายถึง การศึกษาและวิเคราะห์พฤติกรรมการซื้อของผู้บริโภคในช่วงเวลาที่แยกต่างกัน เพื่อค้นหาแนวโน้มและรูปแบบการเปลี่ยนแปลงที่เกิดขึ้น

■ Purchasing Pattern Over Time

↳ รู้จักลูกค้า ช่วงเวลา ประเภท ความต้องการ

■ Cross Selling

↳ รู้จักลูกค้า ของ ซึ่งไปร่วมกัน

■ Direct Mail Campaign

↳ รู้จักลูกค้า ด้วย จดหมาย ทางไปรษณีย์ ทางอีเมล ทางโทรศัพท์ ทางไปรษณีย์

ទី៣ នៃ គ.ផែនក គ.សិទ្ធិភាពកសិករណ៍បង្ហាញនៃការបង់ប្រាក់

2. Risk Analysis and Management

- Forecasting នឹងពេញលាយការអាមេរិកាតា
 → សេវាឌុំបង់ប្រាក់តាមកម្លោងនិងតម្លៃដែលត្រូវបាន
- Credit Scoring សំរាបក្រប់បានការអនុវត្តន៍ដែលត្រូវបាន
- Profile of Attrition គឺជាការគិតចោរតាមការបង់ប្រាក់ដែលត្រូវបាន

3. Fraud Detection and Management

↳ រាយការណ៍ការបង់ប្រាក់ដែលត្រូវបានបង់ប្រាក់
សារធម៌នៃការបង់ប្រាក់ដែលត្រូវបានបង់ប្រាក់

4. Web Mining

↳ ពេលវេលាដែលការបង់ប្រាក់ត្រូវបានបង់ប្រាក់

5. Text Mining

↳ ឯកសារព័ត៌មាន, ពេលវេលា

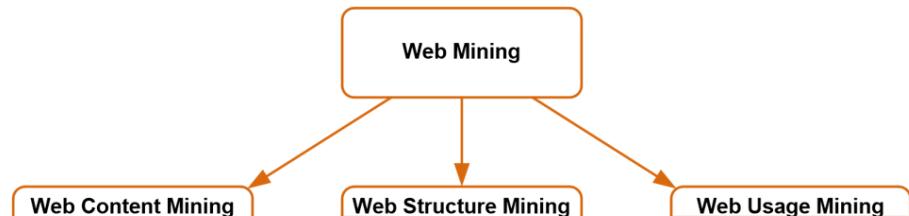
Text mining การทำเหมืองข้อมูลความ

- ▶ สกัดองค์ประกอบที่สำคัญจากชุดข้อมูลขนาดใหญ่ที่ไม่มีโครงสร้าง (unstructured data) และค้นพบ รูปแบบและความสัมพันธ์ และทำการสรุปสารสนเทศได้
- ▶ **Unstructured data:** Word documents, PDF files, text excerpts, and so on
- ▶ **For example:**
 - ▶ การทำเหมืองข้อมูลวิเคราะห์ สำเนาบันทึกการโทรศัพท์ของลูกค้าไปยังศูนย์บริการลูกค้า เพื่อระบุชื่อบริการที่สำคัญ และปัญหาการซ่อมแซม
 - ▶ การแบ่งกลุ่มเอกสารเป็นหมวดหมู่ เพื่อความสะดวกในการใช้งานและการค้นหา
- ▶ **Applications**
 - ▶ Marketing applications: Enables better CRM

Text mining กับ CRM

Web mining

- ▶ Web mining (or Web data mining) is the use of data mining techniques to extract knowledge (discover relationships) from Web data (web document, linkage, or usage)
 - ▶ Data is in HTML, XML, text format
 - ▶ Discovery and analysis of useful patterns and information from web
- ▶ Web content mining
 - ▶ Web structure mining
 - ▶ Web usage mining
- E.g., การทำความเข้าใจพฤติกรรมของลูกค้า, การประเมินประสิทธิภาพของเว็บไซต์



ดู พฤติกรรม คุ้มครอง ภัยคุกคาม ทางไซเบอร์ ที่มาในรูปแบบ
ของภาษา自然语言 processing website

Data Mining Process

1. Objective Determination (กำหนดวัตถุประสงค์ในการทำ Data Mining)
2. Data Preparation (เตรียมข้อมูล)
3. Data Mining (ทำดาต้าไมนิ่ง)
4. Analysis of Results (ทำการวิเคราะห์ผลลัพธ์ที่ได้)
5. Assimilation of knowledge (นำสารสนเทศที่ได้ไปใช้ประโยชน์)

Data Mining Process

1. Objective Determination

- ▶ Business understanding กำหนดวัตถุประสงค์ ที่ต้องการในกระบวนการ Data Mining เช่น
- ▶ เพื่อหาพฤติกรรมผู้บริโภค
- ▶ เพื่อแบ่งกลุ่มลูกค้า
- ▶ เพื่อหาลักษณะลูกค้าที่เปลี่ยนการใช้บริการจากเราไปหาคู่แข่ง เป็นต้น
- ▶ การกำหนดวัตถุประสงค์นั้นจะต้องชัดเจน มิฉะนั้น จะทำให้ตีความผลที่ได้ไม่ถูกต้อง และต้องบอกได้ว่า จะใช้ Data Mining Model ไหนเข้ามาช่วยในการทำ
- ▶ วัตถุประสงค์ที่ได้มานี้อาจมาจากปัญหาที่เกิดขึ้นในองค์กร

๒ หัว A. คือการทำ ขั้นตอน



๑. เก็บข้อมูล ๒. ทำความสะอาด
๓. จัดเรียงข้อมูล

Data Mining Process

2. Data Preparation

การเตรียมข้อมูล โดยขั้นตอนการเตรียมข้อมูลนี้จะแบ่งออกเป็น 3 ขั้นตอนย่อย ดังนี้

→ main DB, warehouse #16

1. **Data Selection** คือการเลือกข้อมูล หรือการได้มาของข้อมูล ซึ่งจะต้องพิจารณาว่าข้อมูลที่จะใช้เป็นข้อมูลอะไร ได้มาจากที่ใด อาจจะได้มาทั้งแหล่งข้อมูลภายใน หรือภายนอกก็ได้ โดยจะเลือกเอาเฉพาะข้อมูลที่จะนำมาทำ Data Mining เท่านั้น

เช่น ต้องการทำ Data Mining เพื่อหาลักษณะลูกค้าที่ซื้อรถสปอร์ต ไม่ควรนำรายชื่อพนักงานขายเข้ามาเกี่ยวข้อง สิ่งที่ควรนำมาพิจารณาสำหรับกรณีนี้อาจพิจารณาทั้งข้อมูลลูกค้าและข้อมูลเกี่ยวกับการขาย เช่น เพศ อายุ รายได้ อาชีพ ยี่ห้อรถยนต์ ประเภทรถยนต์ ลักษณะเครื่องยนต์ เป็นต้น

បច្ចន់ទេរកិយន៍ត សារមុនខោលុយិយន៍ បែនធន
→ ធ្វើ នគរក្រក ៖ → ឈើ ឯកសារអតិថជ្ជកម្ម

2. **Data Preprocessing** เป็นการทําข้อมูลให้มีคุณภาพดี, ทำรูปแบบให้สอดคล้องกัน, เช่น ลบเสียง, ไห้ที่ตัวชี้วัด data [เพื่อคุณให้รับทราบดี]

3. *Data Transformation* เป็นการสร้างข้อมูลชุดใหม่ที่มาจากการแปลงข้อมูลชุดเดิม ซึ่งทำเพื่อ แปลงข้อมูลให้สอดคล้องกับ Model ที่ใช้ได้ เช่น การ Map ค่ามาเพื่อคำนวณเป็นต้น

↳ សំណងក្នុងរាជរដ្ឋបាលនូវនាមត្រូវក្នុង

ନେବ୍ରାକ୍ ପ୍ରକାଶକ୍ତିରେ ମହାକାଵ୍ୟାଳିକାନ୍ତିରେ ଉପରେ

2. Data Preprocessing [ទົກສອງຂໍ້ຕົກ]

- ດູກ່າ outlier (ໂຄສູວພິດໄກຍີ, ທັ່າພິດນາຍີ) ຕື່ຕະຫຼາກຂຶ້ນຫຼຸກເນົາກະຈິປ່ານການເຄີຍການໃຫ້
- ກໍານົດລືບສິ່ງ
- ກໍານົດຈົບຖະລວດຄົດລືອດກັກ ແກ້ໄຂ ແກ້ໄຂ ຫຼາຍ໌ 0,1

Data Transformation

- ▶ เป็นการสร้างข้อมูลชุดใหม่ที่มาจากการนำข้อมูลชุดเดิม ซึ่งทำเพื่อแปลงข้อมูลให้สอดคล้องกับ Model ที่ใช้
 - ▶ **Normalization:** การนำข้อมูลมาแบ่งภายในช่วงที่กำหนด เช่น 0.0-1.0, -1.0-1.0
 - ▶ **Discretization** ▷ แบ่งช่วงๆ ให้เป็นกลุ่มๆ
Involves converting **quantitative variables** into **categorical variables** by dividing the values of the input variable into buckets.
 - ▶ **1 of N Coding**
involves converting **a categorical variable** to numeric representation.

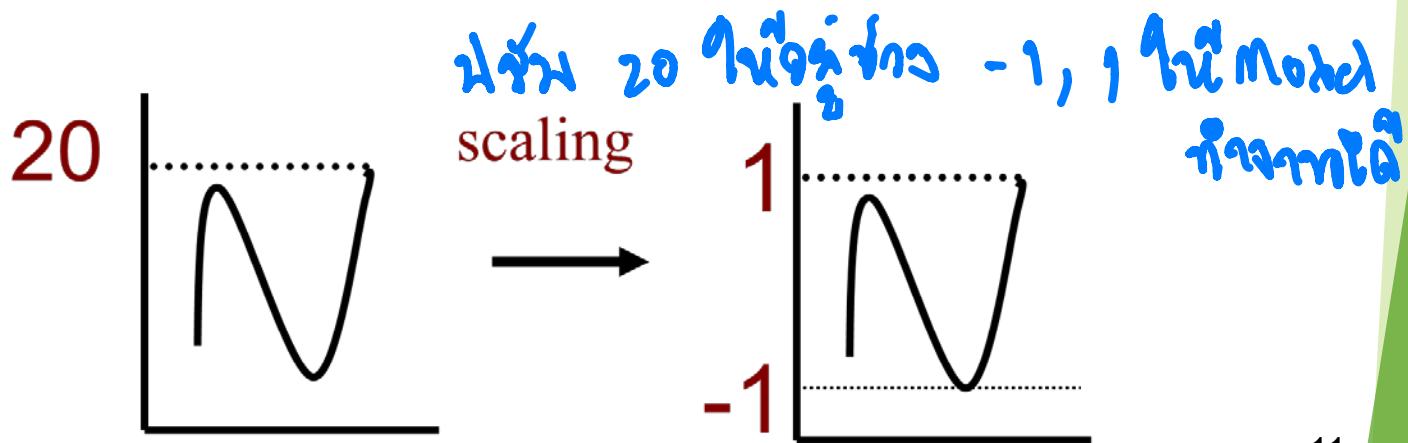
Data Transformation

- ▶ Normalization: การนำข้อมูลมาแบ่งภายในช่วงที่กำหนด เช่น 0.0-1.0, -1.0-1.0

▶ min-max normalization

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A$$

\min_A และ \max_A เป็นค่า Minimum และ Maximum ของแอ็ตทริบิวะ
สามารถใช้ Min-Max Normalization แปลงค่า (Map) V ของ A ไปเป็น
V' บนช่วง [$\text{new_min}_A, \text{new_max}_A$]



Discretization

► Discretization

រៀបចំរាយ ឬចំណាំអង្គភាព → តម្លៃទេរក / តម្លៃទូទៅ

► Involves converting quantitative variables into categorical variables by dividing the values of the input variable into buckets.

► Ex: Income in range \$0-\$9999 = range of 1

\$10,000-\$19,999 = range of 2

One of N coding

- ▶ “One of N coding” involves **converting a categorical variable to numeric representation.**

N = 4.00 ค่า

Ford → 100 0 0

Nissan → 010 0 0

Honda → 001 0 0

Toyota → 000 1 0

BMW → 000 0 1

ค่านะ
เทคโนโลยี

One of N Coding คืออะไร?

One of N Coding หรือที่เรียกอีกอย่างว่า Dummy Coding เป็นวิธีการหนึ่งในการแปลงข้อมูลเชิงหมวดหมู่ (Categorical Data) ให้เป็นตัวเลข ซึ่งเป็นรูปแบบที่คอมพิวเตอร์สามารถนำไปประมวลผลได้ โดยปกติแล้วข้อมูลเชิงหมวดหมู่จะเป็นตัวอักษร เช่น ชื่อยี่ห้อรถยนต์ ในที่นี่คือ Ford, Nissan, Honda, Toyota

ทำไมต้องแปลงข้อมูลเชิงหมวดหมู่เป็นตัวเลข?

- เพื่อให้คอมพิวเตอร์เข้าใจ: คอมพิวเตอร์ทำงานกับตัวเลขได้ดีกว่าตัวอักษร ดังนั้นการแปลงข้อมูลให้เป็นตัวเลขจะทำให้คอมพิวเตอร์สามารถนำข้อมูลไปวิเคราะห์ได้ เช่น การสร้างแบบจำลองทางสถิติ หรือการทำ Machine Learning
- เพื่อใช้ในการวิเคราะห์ข้อมูล: หลายเทคนิคในการวิเคราะห์ข้อมูล เช่น การวิเคราะห์การถดถอย (Regression Analysis) หรือ การวิเคราะห์ความแปรปรวน (Analysis of Variance) ต้องการข้อมูลในรูปแบบตัวเลข

วิธีการทำ One of N Coding

- กำหนดหมวดหมู่: กำหนดหมวดหมู่กั้นหมวดหมู่ที่มีในข้อมูล เช่น ยี่ห้อรถยนต์
- สร้างตัวแปร Dummy: สร้างตัวแปรใหม่สำหรับแต่ละหมวดหมู่ โดยตัวแปรเหล่านี้จะมีค่าเป็น 0 หรือ 1
 - ถ้าข้อมูลตรงกับหมวดหมู่นั้น ตัวแปรจะมีค่าเป็น 1
 - ถ้าข้อมูลไม่ตรงกับหมวดหมู่นั้น ตัวแปรจะมีค่าเป็น 0
- กำหนดจำนวนหลัก: จำนวนหลักที่ใช้ในการแทนแต่ละหมวดหมู่ซึ่งอยู่กับจำนวนหมวดหมู่ทั้งหมด
 - ในตัวอย่างนี้มี 4 หมวดหมู่ ดังนั้นจึงใช้ตัวเลข 4 หลักในการแทนแต่ละหมวดหมู่

ตัวอย่าง

ยี่ห้อรถยนต์	Ford	Nissan	Honda	Toyota
ตัวแปร Dummy 1	1	0	0	0
ตัวแปร Dummy 2	0	1	0	0
ตัวแปร Dummy 3	0	0	1	0
ตัวแปร Dummy 4	0	0	0	1

ส่งออกไปยังชีต

Data Mining Process

3. การทำ Data Mining

- ▶ Data Mining เป็นกระบวนการของการสร้างแบบจำลอง(Model) โดยสร้างแบบจำลองของข้อมูล เพื่อสร้างความเข้าใจในแนวโน้มรูปแบบ และความสัมพันธ์กันของข้อมูล เพื่อใช้ในการทำนายข้อมูลเหล่านั้น โดยจะเลือก Model ให้เหมาะสมกับวัตถุประสงค์
ในการทำ

↓
สืบสาน

Data Mining Process

Model ในการทำ Data Mining

କ୍ଷେତ୍ର ପରିମାଣ ନିର୍ଣ୍ଣାଳୀ

1. Predictive Model

makes predictions about the future based on information from the past and present

ମୁଖ୍ୟ ପରିମାଣ କିମ୍ବା ମୋଡ୍ଯୁଲ ହେଉଥିଲା

① ► Classification เป็นการทำนายว่า สิ่งนั้นควรอยู่ในกลุ่มไหน

► Decision trees

► Neural Networks

କାହାର କାହାର କାହାର

② ► Value prediction เป็นการทำนายค่าที่เป็นตัวเลข

► Linear Regression

କାହାର କାହାର କାହାର କାହାର

Classification

- ▶ the process of assigning new objects to predefined categories or classes

▶ เกิดขึ้นจากการจัดหน่วยเข้ากลุ่ม (classification) ซึ่งเป็นวิธีการจำแนกหน่วยว่าเป็นสมาชิกของกลุ่มใด โดยมีการกำหนดคุณลักษณะไว้ล่วงหน้า

▶ Supervised Learning - ที่มาของงานนี้叫做 **Model**

↳ **ข้อมูลที่แล้วถูกตั้งอย่างชัดเจน**

EX. A classified data of Credit card customers

No	income	Job Position	age	Class
4	A
5	C
6	A
7	B

หาตัวอย่าง
กระบวนการเรียนรู้
ด้วยการฝึกหัด

Supervised Learning និង Prediction នៃការពុគ្គលិត

→ តើវិភាគ គឺ ឬ វារុបេរិសែណា [ផ្សោតការងារ និងជីថិត]

- ▶ Supervised Learning เป็นកម្របវារៈរៀនរូបដើម្បីដោះស្រាយទំនាក់ទំនងនៃការពុគ្គលិតចំណែក។
- ▶ មានការបង្កើតបញ្ជីពីការពុគ្គលិតដែលបានបង្កើតឡើង។
- ▶ ត្រូវបានបង្កើតបញ្ជីពីការពុគ្គលិតដែលបានបង្កើតឡើង។
- ▶ ត្រូវបានបង្កើតបញ្ជីពីការពុគ្គលិតដែលបានបង្កើតឡើង។

No	income	Job Position	age	Class
1	50000	Manager	40	A
2	18000	worker	24	C
3	60000	Manager	45	A
4	35000	supervisor	30	B

Supervised Learning หรือ การเรียนรู้ภายใต้การดูแล เป็นหนึ่งในสาขาหลักของ Machine Learning ที่สอนคอมพิวเตอร์ให้เรียนรู้จากข้อมูลที่มีป้ายกำกับ (labeled data) เพื่อให้สามารถทำนายผลลัพธ์ใหม่ที่ไม่เคยเห็นมาก่อนได้

อธิบายง่ายๆ คือ

เราให้คอมพิวเตอร์ดูภาพแมวและสุนัขจำนวนมากพร้อมกับป้ายกำกับบอกว่าภาพไหนเป็นแมว ภาพไหนเป็นสุนัข เมื่อคอมพิวเตอร์เรียนรู้จากข้อมูลเหล่านี้ไปเรื่อยๆ มันก็จะสามารถ分辨ได้ว่าภาพใหม่ที่มันไม่เคยเห็นมาก่อนนั้นเป็นแมวหรือสุนัข

ตัวอย่างการใช้งาน Supervised Learning ในชีวิตจริง

- การจำแนกประเภทภาพ: เช่น จดจำใบหน้า, สิ่งของ, หรือตรวจหารอยโรคจากภาพ X-ray
- การทำนายค่า: เช่น พยากรณ์อากาศ, ราคารถ, หรือผลการแข่งขันกีฬา
- การแปลภาษา: เช่น Google Translate
- การแนะนำผลิตภัณฑ์: เช่น ระบบแนะนำสินค้านะบินเว็บไซต์อีคอมเมิร์ซ

ขั้นตอนการทำงานของ Supervised Learning

- เตรียมข้อมูล: รวบรวมข้อมูลที่มีป้ายกำกับ เช่น ภาพแมวและสุนัขพร้อมป้ายบอกชนิด
- เลือกโมเดล: เลือกโมเดล Machine Learning ที่เหมาะสมกับปัญหา เช่น Linear Regression, Logistic Regression, Decision Tree, Random Forest, Neural Network
- ฝึกโมเดล: ป้อนข้อมูลที่มีป้ายกำกับให้โมเดลเรียนรู้ เพื่อให้โมเดลสามารถสร้างความสัมพันธ์ระหว่างข้อมูล input กับ output
- ทดสอบโมเดล: ใช้ข้อมูลที่โมเดลไม่เคยเห็นมาก่อนมาทดสอบ เพื่อประเมินความแม่นยำของโมเดล
- ปรับปรุงโมเดล: หากผลลัพธ์ยังไม่เป็นที่น่าพอใจ ก็สามารถปรับปรุงโมเดลได้ เช่น เปลี่ยนโมเดล, ปรับพารามิเตอร์, หรือเพิ่มข้อมูล

Unsupervised Learning หรือ การเรียนรู้แบบไม่มีผู้สอน เป็นอีกหนึ่งสาขาสำคัญของ Machine Learning ที่แตกต่างจาก Supervised Learning ตรงที่ข้อมูลที่นำมาฝึกโมเดลนั้นไม่มีการระบุป้ายกำกับ (unlabeled data) หรือค่าตอบกลับที่ถูกต้องไว้ล่วงหน้า

หมายความว่าอย่างไร?

- ไม่มีค่าตอบกลับที่ถูกต้อง: โมเดลจะต้องค้นหารูปแบบหรือโครงสร้างที่ซ่อนอยู่ในข้อมูลเอง โดยไม่มีตัวอย่างที่ถูกต้องมาให้เปรียบเทียบ
- ค้นหาความสัมพันธ์: โมเดลจะพยายามหาความสัมพันธ์ระหว่างข้อมูลต่างๆ เพื่อจัดกลุ่มข้อมูล หรือลดมิติของข้อมูล

เปรียบเทียบกับ Supervised Learning

คุณสมบัติ	Supervised Learning	Unsupervised Learning
ข้อมูล	มีป้ายกำกับ (labeled data)	ไม่มีป้ายกำกับ (unlabeled data)
เป้าหมาย	ทำนายผลลัพธ์	ค้นหาโครงสร้างในข้อมูล
ตัวอย่าง	จัดประเภทภาพแมวและสุนัข	จัดกลุ่มลูกค้าตามพฤติกรรมการซื้อ

ตัวอย่างการใช้งาน Unsupervised Learning

- การจัดกลุ่มข้อมูล (Clustering): แบ่งข้อมูลออกเป็นกลุ่มย่อยตามความคล้ายคลึงกัน เช่น แบ่งลูกค้าออกเป็นกลุ่มตามพฤติกรรมการซื้อ, แบ่งเอกสารออกเป็นกลุ่มตามหัวข้อ
- การลดมิติของข้อมูล (Dimensionality Reduction): ลดจำนวนมิติของข้อมูลลง เพื่อให้การวิเคราะห์ข้อมูลง่ายขึ้น เช่น การลดมิติของข้อมูลภาพเพื่อใช้ในการจำแนกภาพ
- การค้นหาความผิดปกติ (Anomaly Detection): ค้นหาข้อมูลที่แตกต่างจากข้อมูลล้วนใหญ่ เช่น การตรวจสอบการฉ้อโกงการเงิน

ทำไมต้องใช้ Unsupervised Learning?

- ข้อมูลจำนวนมาก: ในหลายกรณี เราอาจมีข้อมูลจำนวนมาก แต่ไม่มีป้ายกำกับที่เพียงพอสำหรับการใช้ Supervised Learning
- ค้นพบสิ่งใหม่: Unsupervised Learning ช่วยให้เราค้นพบรูปแบบหรือความสัมพันธ์ที่เราไม่เคยรู้มาก่อน
- เตรียมข้อมูลสำหรับการวิเคราะห์เพิ่มเติม: ผลลัพธ์จาก Unsupervised Learning สามารถนำไปใช้เป็นข้อมูลเริ่มต้นสำหรับการวิเคราะห์ข้อมูลเพิ่มเติม เช่น การสร้างกฎการตัดสินใจ

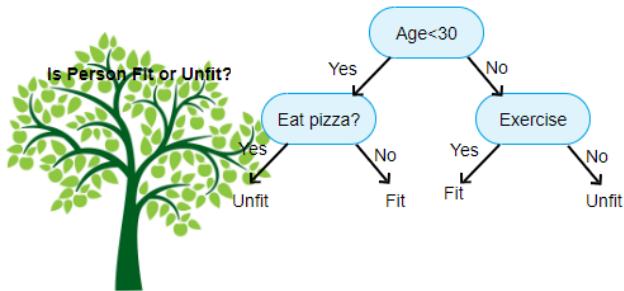
An example of classification

ชื่อ	ระยะเวลาที่เป็นสมาชิก	# บริการที่ใช้	สถานะ
A	>2.5	>3	Stay
B	>2.5	<3	Stay
C	<2.5	<3	Leave
D	<2.5	>3	Stay

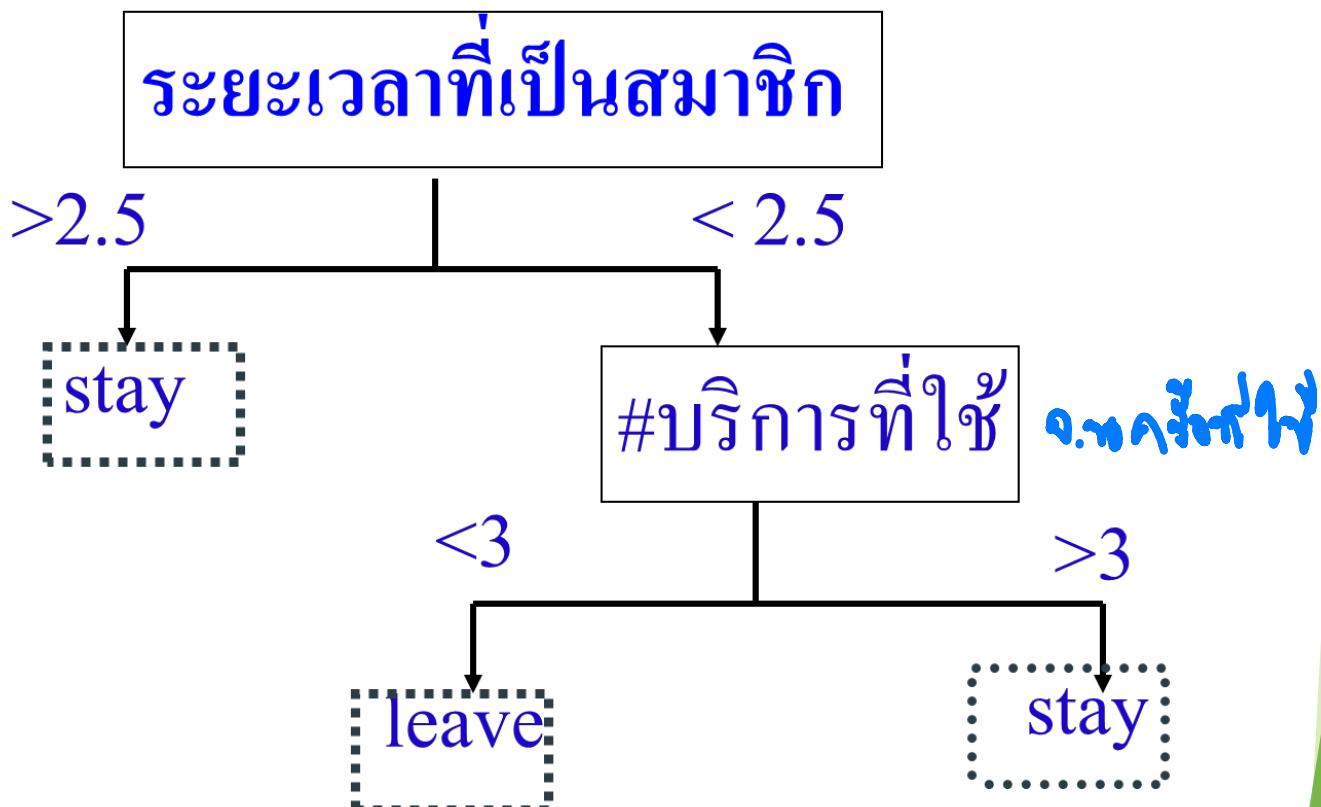
Input

Output

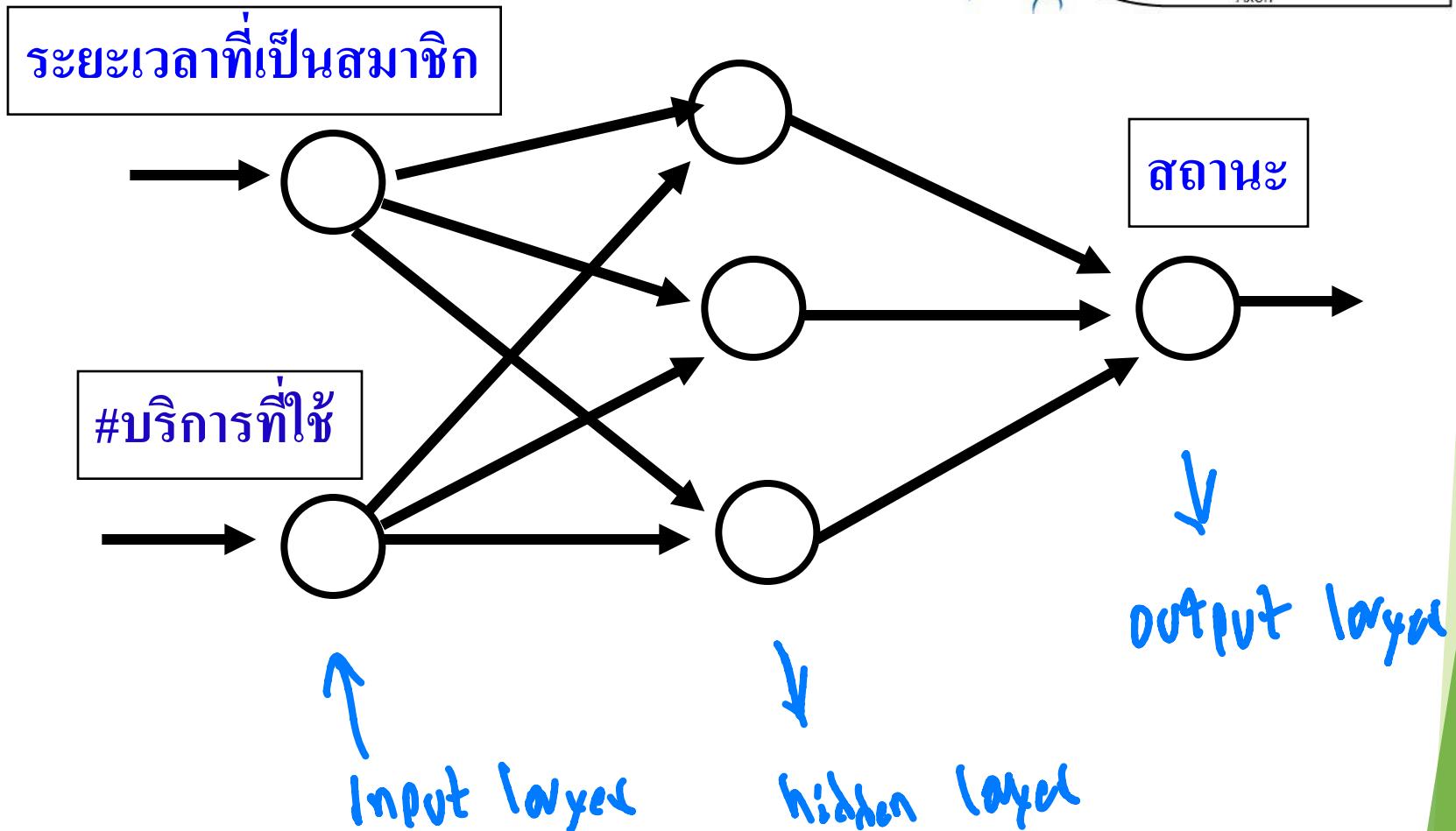
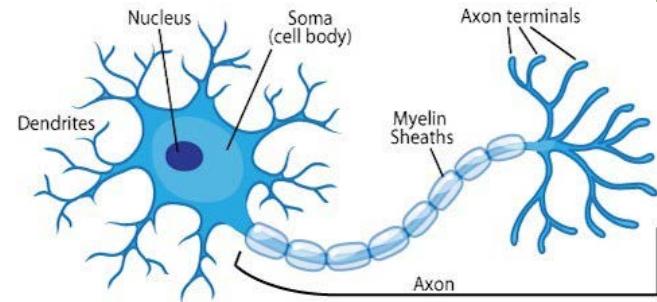
An example of classification by decision tree



ทรรศน์ทางการเรียนรู้
การจำแนก, เรียนรู้ โดย
การทำต้นไม้ ในการจัดกลุ่ม



Example of classification by using neural networks



Data Mining Process

Value Prediction

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

Regression

- ▶ learning a function which maps a data item to a real-valued prediction variable.

EX. A prediction model of car accident probability

No	Vehicle speed (mph)	Alcohol level	age	Probability of Car Accident
1	80	0.02	20	0.80
2	70	0.01	25	0.70
3	55	0.0	30	0.05
4	60	0.1	45	0.26

Data Mining Process

2. Database Segmentation *ការអនុវត្តភាពទិន្នន័យ*

- Clustering

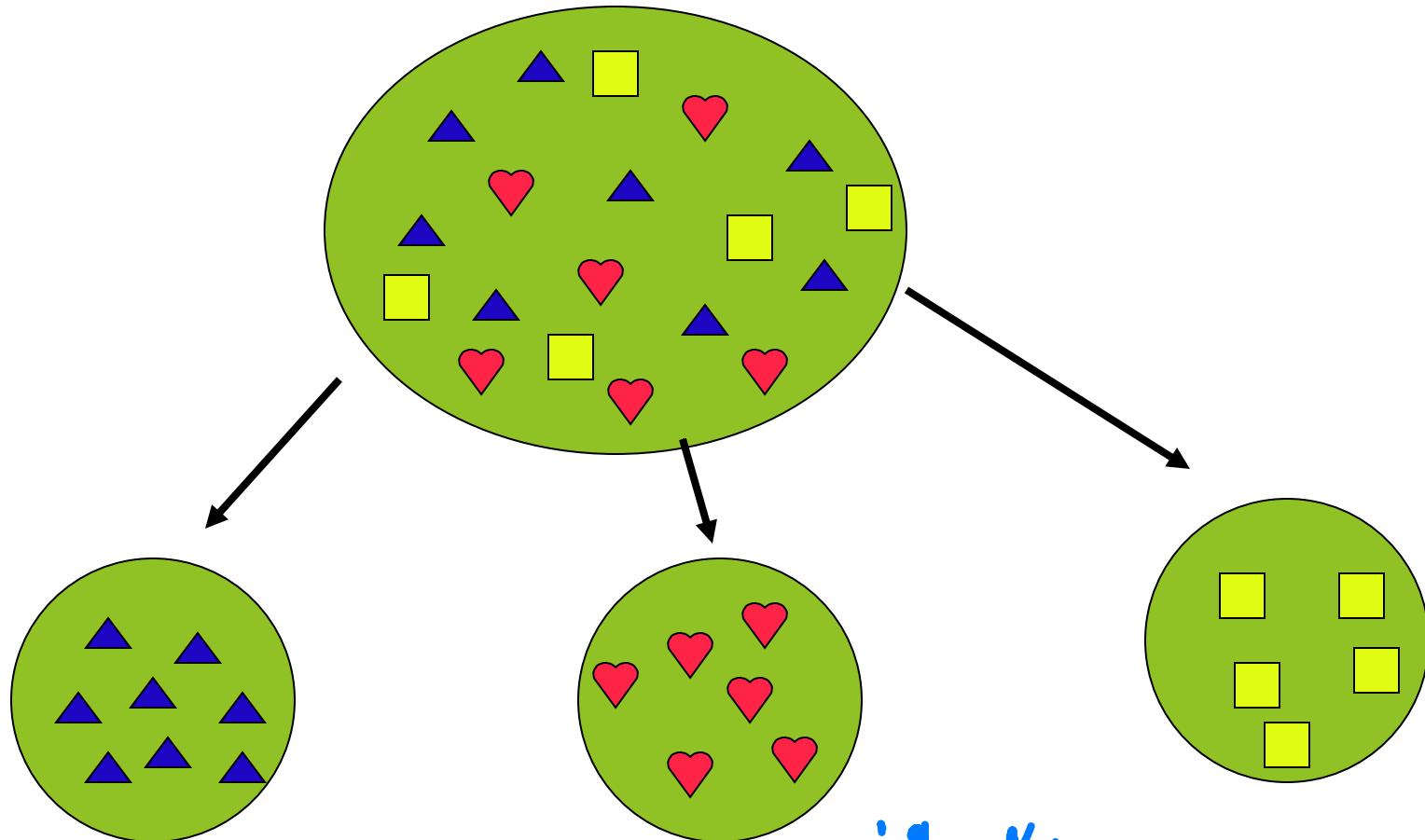
- *ផែន វទ្ធសាស្ត្រ, learning
ពេលកម្រិតបាន ដឹងទូរការ*

• The goal of database segmentation is to partition a database into segments of similar records, that is records that share a number of properties.

• เป็นเทคนิคที่ใช้แบ่งกลุ่มคน สิ่งของ โดยแบ่งคน, สิ่งของ ที่คล้ายกันไว้ด้วยกัน หรืออยู่ในกลุ่มเดียวกัน

• ไม่จำเป็นต้องทราบกลุ่มมาก่อนว่าใครอยู่กลุ่มใด

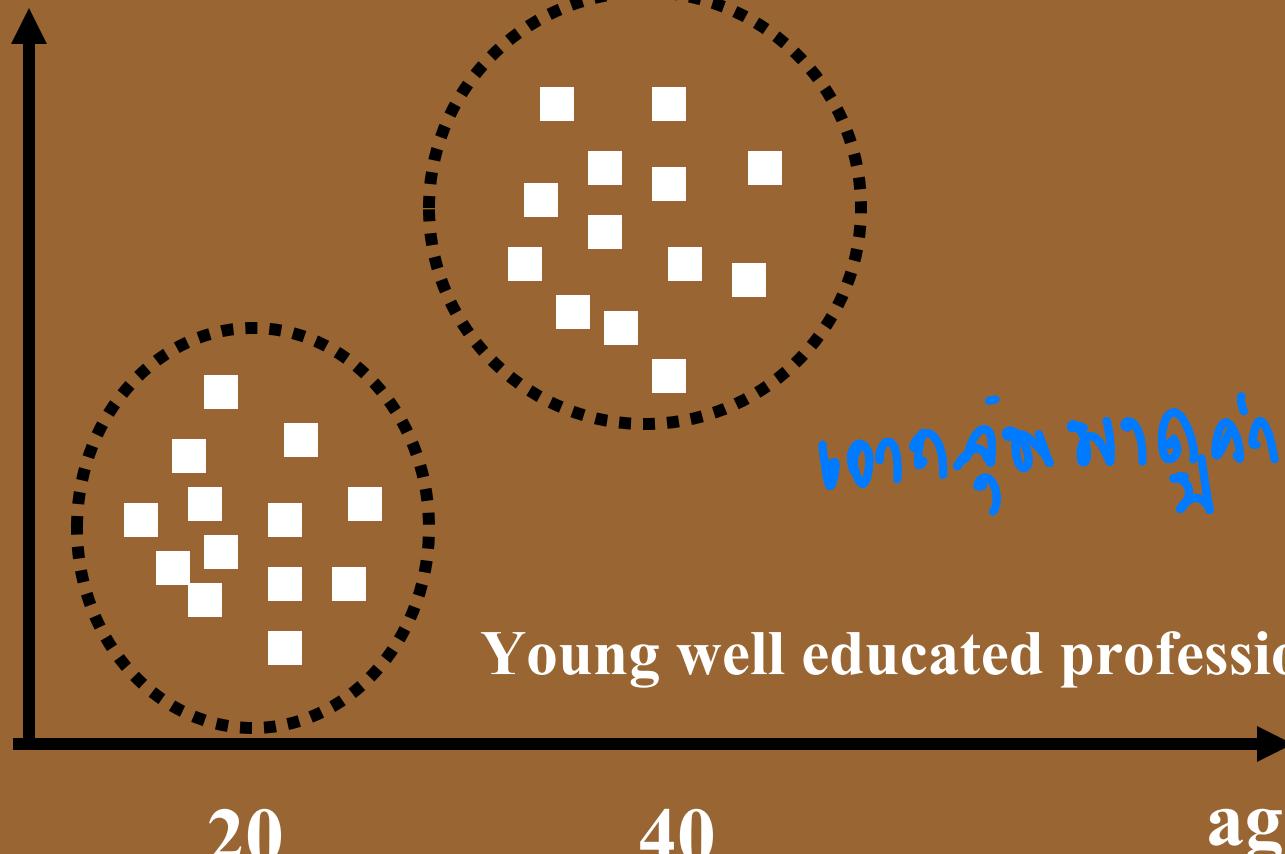
Database segmentation



សេវាសម្រាប់បង្កើតការណ៍ទូទៅជាអាជីវការ

Database segmentation

Income (thousands)



Data Mining Process

3. Link Analysis វិធាយាយន៍ A.S.A សង្គមពីរ
បន្ទាន់លេខា

► Associations Discovery  សំណង់ការងារ
នូវគោលការណ៍ ការងារទូទៅ

► Sequential Pattern Discovery

២ នឹងការកិច្ចការរាយ និងបញ្ជូនការទាំងអស់
តើការបញ្ជូនការទាំងអស់

A. ព័ត៌មាន

ផែតាម ពីរធនជាតិការងារក្នុងប្រព័ន្ធដែល

- A និង S នឹងការកិច្ចការណ៍ ការងារទូទៅ

- S និង A នឹងការបញ្ជូនការទាំងអស់ ការងារទូទៅ

ຕາມຄສນ ປີ (Case Study)

ພາກອອ ໄກສາ ອມ ດິເວລາ

ບໍ່ ຈັດກຸນຫຼຸດຕົກ ກົນ (Part Model)

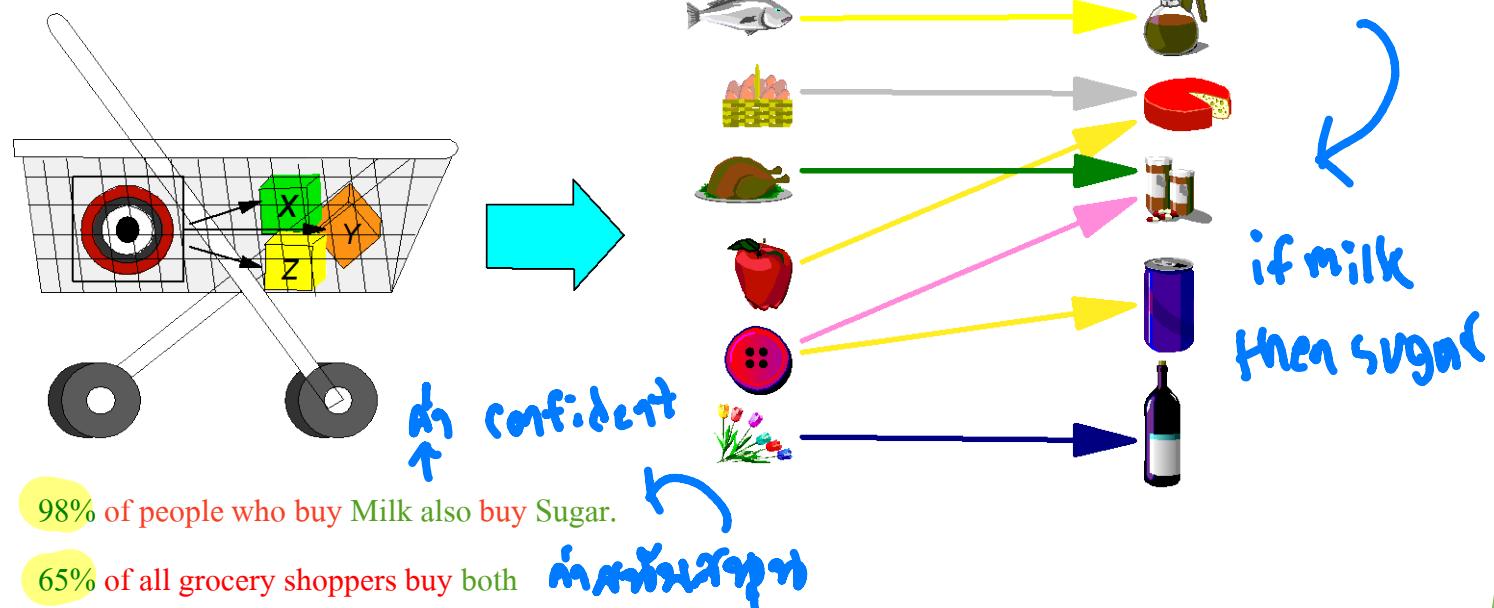
ຕື່ອງ ຂອບເລືອດກ່າ

Associations Discovery

การค้นพบสารสนเทศที่มีความสัมพันธ์ที่เชื่อมโยงกัน เช่น **ลูกค้าผู้ชายซื้อผ้าอ้อม**
เพิ่มเพิ่งบ่ายวันศุกร์มีแนวโน้มจะซื้อเบียร์ด้วย

Association rule

Detect sets of attributes that frequently co-occur, and rules among them EX: Milk \rightarrow sugar - **เมื่อมีสินค้า A ก็จะ if then**



Data Mining Process

4. การวิเคราะห์ผลลัพธ์ที่ได้ (Analysis of Results)

- ▶ เป็นการวิเคราะห์และตีความหมายจากผลที่ได้

ເນັດກາ ຂອງ ພາຕິຈົງຄູວາ ກ່ອນກາເພັດທານີ

5. การนำสารสนเทศไปใช้ประโยชน์ (Assimilation of Knowledge)

- ▶ การนำความรู้ที่ได้ไปใช้เป็นขั้นตอนสุดท้ายของการบวนการทั้งหมด
- ▶ รวบรวมความเข้าใจในแบบจำลองและการวิเคราะห์ผลลัพธ์ที่ได้ มารวมเข้ากับส่วนความรู้ทางธุรกิจ เพื่อที่จะนำเสนอถึงวิธีการที่จะนำผลที่ได้นี้ไปใช้ให้เกิดประโยชน์

ເພົາກາໃຫຍ້ໄດ້ ແຫ່ງກ່ຽວຂ້ອງ

Data Mining Process

Mining
Objective



Data sources

Databases, flat files,
Complex data



Data Warehouses



Data Preparation

(Data Selection , Data Preprocessing
and Data Transformation)



Data Mining



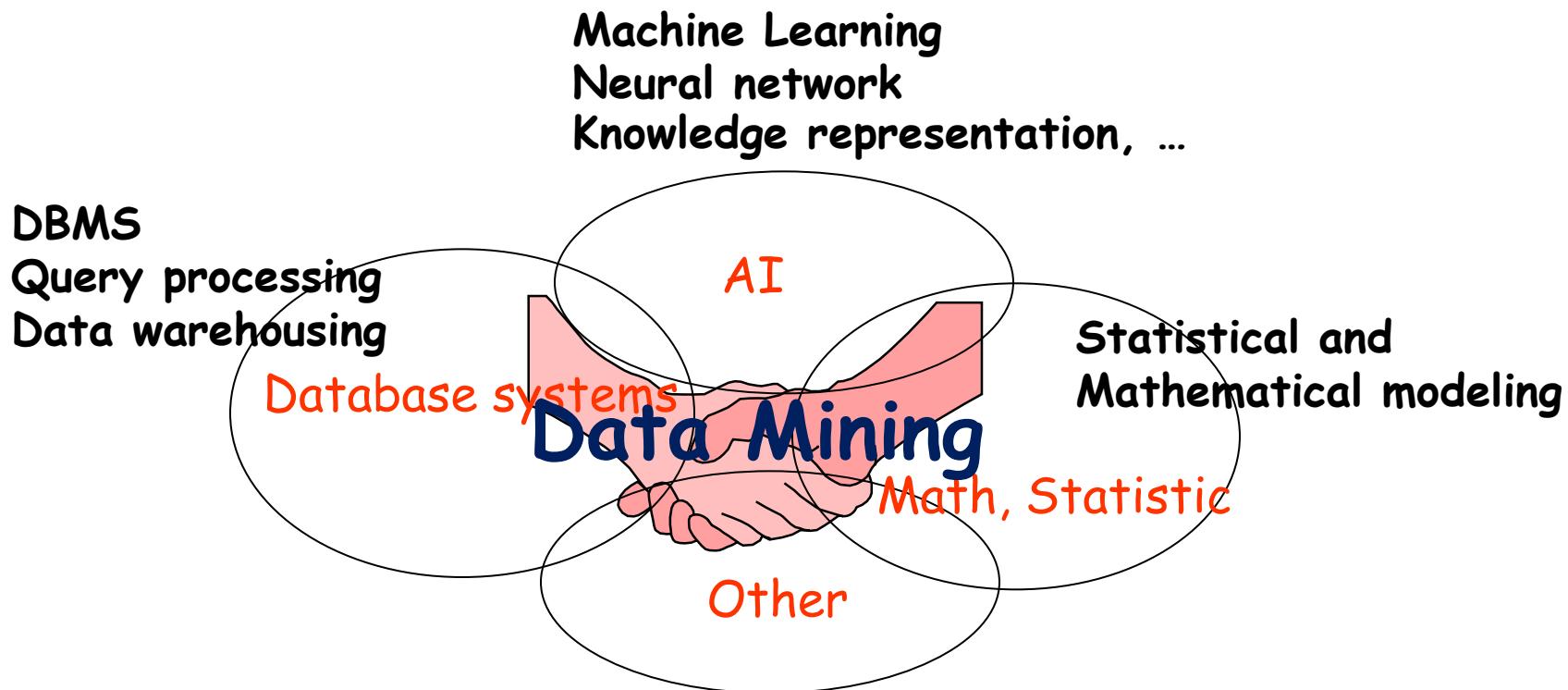
Assimilation of knowledge



Analysis of Results



Data mining is multidisciplinary

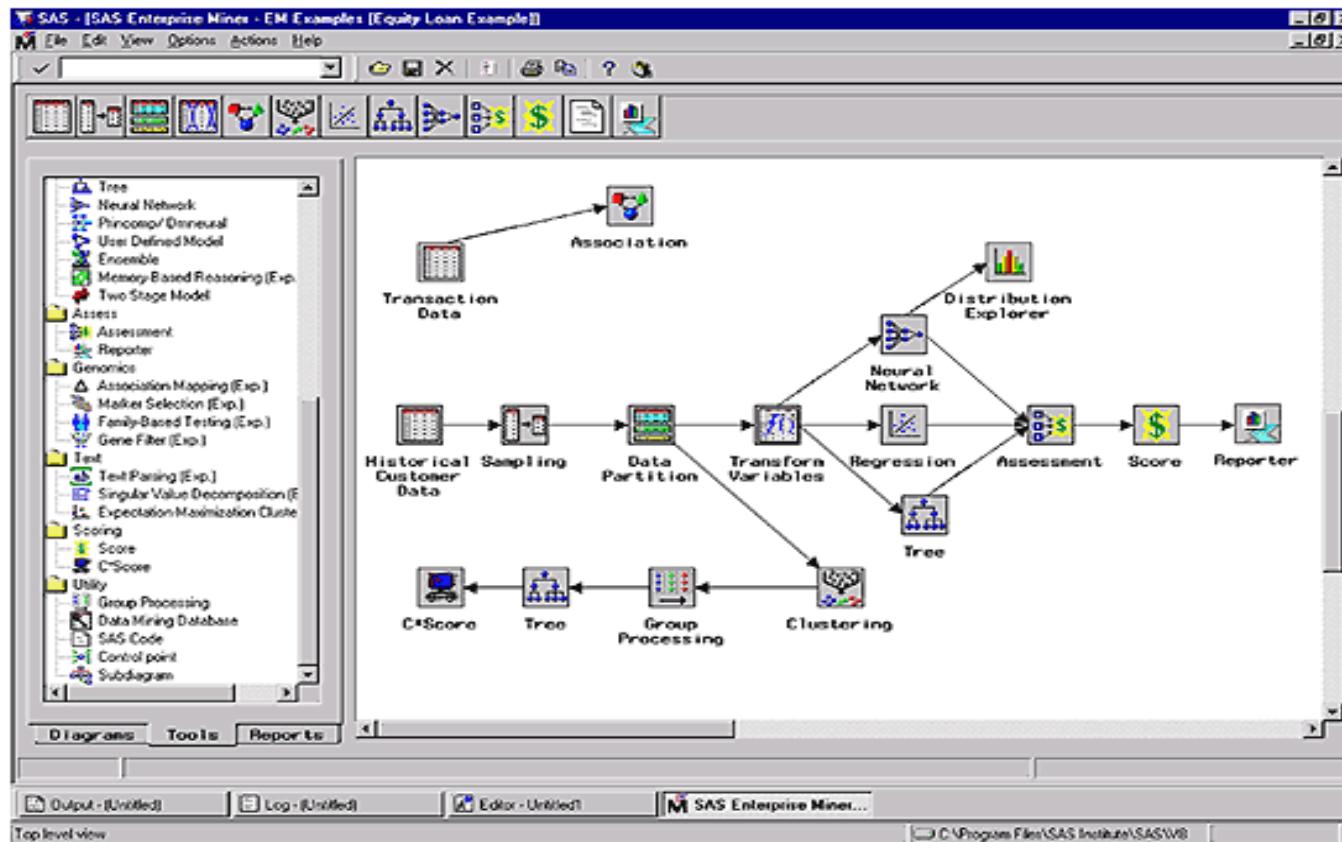


Data Mining Software

- ▶ Software ที่ต่อยอดจากฐานข้อมูล (Integrated data mining into the database)
 - ▶ [DB2 Intelligent Miner](#)
 - ▶ [Microsoft SQL Server Analysis Services - Data Mining](#)
 - ▶ [Oracle Data Mining](#)
- ▶ Software จากบริษัทที่ชำนาญด้านสถิติ
 - ▶ [SAS Enterprise Miner](#)
 - ▶ [IBM SPSS Modeler Professional](#) เดิมคือ SPSS Clementine

Data Mining Software

SAS Enterprise Miner



งานกลุ่ม

- ▶ ทำงานวิจัย หรือ โครงการ หรือ บทความ ที่ทำโดยใช้ Data Mining
- ▶ อ่านแล้วสรุป มาเป็น Data Mining Process
- ▶ ส่งภายในวันที่ 6 ต.ค. 67

↙↙
ผลงาน
ส่งมาทาง
วันที่ 6 ต.ค. 67