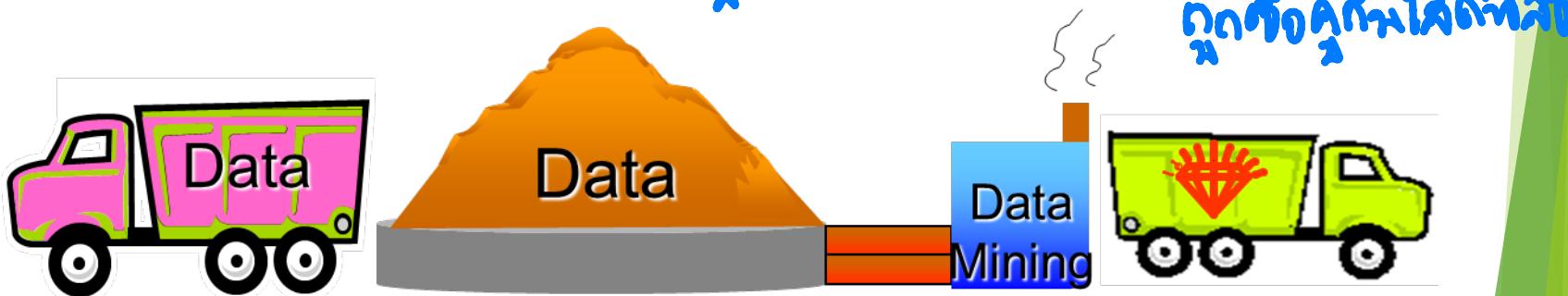


FOUNDATIONS OF BUSINESS INTELLIGENCE:

Data Mining

Data Mining អត្ថបទនៃវិធានការសាស្ត្រ

- ▶ เป็นกระบวนการค้นหาสารสนเทศอุปกรณ์จากฐานข้อมูลขนาดใหญ่ เพื่อนำความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ **เจ้าหน้าที่งานคลัง, ไมโครชิป**
 - ▶ เป็นกระบวนการ (Process) ที่กระทำกับข้อมูลขนาดใหญ่ เพื่อค้นหา **ตัวตั้ง** **เจ้าหน้าที่งานคลัง** รูปแบบของข้อมูล แนวโน้มของข้อมูล และความสัมพันธ์ที่ซ่อนอยู่ใน **ลักษณะ** ชุดข้อมูลนี้โดยการใช้เทคนิคทาง AI, สติติ และคณิตศาสตร์ เพื่อให้ได้ **ผลลัพธ์** สารสนเทศที่เราไม่รู้อุปกรณ์ โดยสารสนเทศที่ได้จะมีเหตุผลและสามารถนำไปใช้ประโยชน์ได้ **สำหรับเจ้าหน้าที่งานคลัง**



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 อีเมล ไปรษณีย์ ไปรษณีย์ จดหมาย จดหมาย

ទី៣ នៃ គ.ផែនក គ.សិទ្ធិភាពកសិករណ៍បង្ហាញនៃការបង់ប្រាក់

2. Risk Analysis and Management

- Forecasting នឹងពេញលាយការអាមេរិកាតា
 → សេវាឌុំបង់ប្រាក់តាមកម្មង់នៃឯកសារដើម្បីពេញលាយនៅក្នុងការបង់ប្រាក់
- Credit Scoring សំរាបក្រប់បានការអនុវត្តន៍យូរការខែត្រីដោយ
 → នូវការថ្លែងទៅលើបង់ប្រាក់ជាបន្ទាន់
- Profile of Attrition គឺជាការគិតចោរតាមការបង់ប្រាក់ដែលបានបង់ប្រាក់
 → ធានាថ្មីនៃការបង់ប្រាក់ដែលបានបង់ប្រាក់

3. Fraud Detection and Management

↳ នៅពេលក្រុងការបង់ប្រាក់ត្រូវបានបង្ហាញដូចតាំង
 សារធម៌ក្នុងការបង់ប្រាក់ដែលបានបង់ប្រាក់ (ខ្លួនមិនមែន)

4. Web Mining

↳ អេឡិចត្រូនការបង់ប្រាក់ដូចតាំង ពេលវេលាដូចជាលីវិន

5. Text Mining

↳ ឯកសារ ឯកសារ, 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

การทำ Text Mining คืออะไร

การทำ Text Mining หรือ การทำเหมืองข้อมูล เป็นเทคนิคการค้นหาความรู้ใหม่จากข้อมูลประเภทข้อมูลที่มีปริมาณมากโดยอัตโนมัติ โดยการสกัดคำ ค้นหารูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในขุดข้อมูลเอกสาร เพื่อให้เกิดความหมายและสามารถนำมาใช้ประโยชน์ได้

เปรียบเทียบง่ายๆ ก็เหมือนกับการ "ขุด" หาเพชรในกองกรายขนาดใหญ่ โดยรายในที่นี้คือข้อมูลข้อมูลจำนวนมหาศาล เช่น ข่าว บทความ ความคิดเห็นในโซเชียลมีเดีย หรือรายงานต่างๆ และเพชรที่เราต้องการหาคือข้อมูลที่น่าสนใจ ความรู้ใหม่ หรือรูปแบบที่ซ่อนอยู่ในข้อมูลเหล่านั้น

ทำไมต้องทำ Text Mining?

- **ข้อมูลข้อมูลมีปริมาณมาก:** ในยุคดิจิทัล ข้อมูลข้อมูลข้อมูลสร้างขึ้นอย่างมหาศาลทุกวัน การทำ Text Mining ช่วยให้เราสามารถจัดการและวิเคราะห์ข้อมูลเหล่านี้ได้อย่างมีประสิทธิภาพ
- **ค้นพบข้อมูลที่ซ่อนอยู่:** ช่วยให้เราค้นพบความสัมพันธ์ รูปแบบ หรือแนวโน้มที่เราอาจมองข้ามไปได้
- **ตัดสินใจได้อย่างมีข้อมูล:** ข้อมูลที่ได้จากการทำ Text Mining สามารถนำไปใช้ในการตัดสินใจทางธุรกิจ การวิจัย หรือการวางแผนเชิงกลยุทธ์ได้
- **เพิ่มประสิทธิภาพในการทำงาน:** ช่วยลดเวลาและทรัพยากรที่ใช้ในการวิเคราะห์ข้อมูลด้วยมือ

เจตนาใช้ Text Mining อย่างไร: พลิกข้อมูลลูกค้าให้เป็นผลลัพธ์เคลื่อนธุรกิจ

เจตนา เป็นสลายการบินที่รู้จักกันดีในเรื่องการนำเทคโนโลยีมาใช้เพื่อปรับปรุงประสบการณ์การเดินทางของลูกค้า หนึ่งในเทคโนโลยีที่เจตนาให้ความสำคัญคือ Text Mining หรือ การท้าเหมือนใจ ข้อความ ซึ่งเป็นเครื่องมือของพลังที่ช่วยให้สามารถบินสามารถวิเคราะห์ข้อมูลเชิงลึกจากข้อความที่ลูกค้าส่งมา ไม่ว่าจะเป็นข้อความในโซเชียลมีเดีย อีเมล หรือแม้แต่การสนทนาน่าผ่านช่องทางต่างๆ

เจตนาใช้ Text Mining ในด้านต่างๆ ดังนี้:

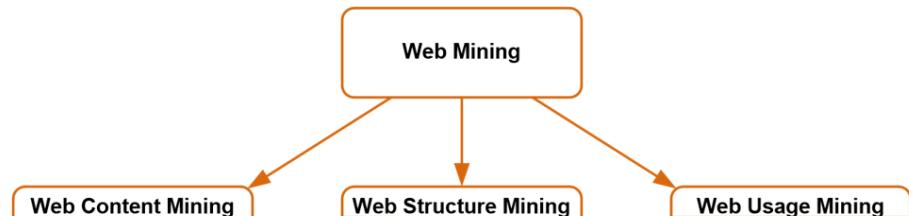
- วิเคราะห์ความคิดเห็นของลูกค้า:** เจตนาใช้ Text Mining เพื่อวิเคราะห์ความคิดเห็นของลูกค้า เกี่ยวกับบริการต่างๆ ของสลายการบิน เช่น การจองตั๋ว การเช็คอิน การบริการบนเครื่องบิน และการจัดการสัมภาระ วิเคราะห์ความรู้สึก (Sentiment Analysis) ช่วยให้เจตนาเข้าใจว่าลูกค้ารู้สึกอย่างไรกับการแต่ละอย่าง และค้นหาจุดแข็งและจุดอ่อนของบริการ
- ค้นหาปัญหาและแนวโน้ม:** Text Mining ช่วยให้เจตนาค้นหาปัญหาที่เกิดขึ้นบ่อย เช่น ปัญหาเกี่ยวกับการจองตั๋ว หรือปัญหาเกี่ยวกับการยกเลิกเที่ยวบิน นอกเหนือจากนี้ ยังช่วยให้เจตนาค้นหาแนวโน้มใหม่ๆ เกี่ยวกับความต้องการของลูกค้า เช่น ลูกค้าต้องการบริการเสริมประเภทใดบ้าง
- ปรับปรุงบริการลูกค้า:** ข้อมูลที่ได้จากการวิเคราะห์ Text Mining ช่วยให้เจตนาสามารถปรับปรุงบริการลูกค้าได้อย่างตรงจุด เช่น การพัฒนาแอปพลิเคชันที่ใช้งานง่ายขึ้น การฝึกอบรมพนักงานให้ตอบสนองต่อปัญหาของลูกค้าได้อย่างรวดเร็ว และการออกแบบผลิตภัณฑ์และบริการใหม่ๆ ที่ตรงกับความต้องการของลูกค้า
- เพิ่มประสิทธิภาพการดำเนินงาน:** Text Mining ช่วยให้เจตนาสามารถเพิ่มประสิทธิภาพการดำเนินงานได้หลายด้าน เช่น การลดต้นทุนในการให้บริการ การเพิ่มความพึงพอใจของลูกค้า และการสร้างความได้เปรียบในการแข่งขัน

ตัวอย่างการนำ Text Mining มาใช้:

- วิเคราะห์ข้อความในโซเชียลมีเดีย:** เจตนาถูกดัดตามข้อความที่ลูกค้าโพสต์เกี่ยวกับสลายการบินในโซเชียลมีเดียต่างๆ เช่น วิดีโอดร็อก แฟชั่น และอินสตาแกรม จากนั้นใช้ Text Mining วิเคราะห์ความรู้สึกของลูกค้า และค้นหาปัญหาที่เกิดขึ้น
- วิเคราะห์เมลจากลูกค้า:** เจตนาวิเคราะห์เมลที่ได้รับจากลูกค้า เพื่อค้นหาคำาณที่พบบ่อย และปัญหาที่ลูกค้าต้องการความช่วยเหลือ
- วิเคราะห์การสนทนาผ่านช่องทางต่างๆ:** เจตนาวิเคราะห์การสนทนาของลูกค้ากับพนักงานผ่านช่องทางต่างๆ เช่น แชทสด และโทรศัพท์ เพื่อค้นหาจุดที่พนักงานสามารถปรับปรุงการให้บริการได้

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 คุณครับ ก็จะได้ หา น้ำหน้าในเว็บไซต์
ที่อยู่ทางล่างไว้ใช้สำหรับ 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

→ หัวข้อ 1. กำหนดวัตถุประสงค์



→ เลือก Model หัวข้อ 2. วิเคราะห์

Data Mining Process

2. Data Preparation

การเตรียมข้อมูล โดยขั้นตอนการเตรียมข้อมูลนี้จะแบ่งออกเป็น 3 ขั้นตอนย่อย ดังนี้

→ **หก步 1B, maxflow ก้าว**

1. **Data Selection** คือการเลือกข้อมูล หรือการได้มาของข้อมูล ซึ่งจะต้องพิจารณาว่าข้อมูลที่จะใช้เป็นข้อมูลอะไร ได้มาจากที่ใด อาจจะได้มาทั้งแหล่งข้อมูลภายใน หรือภายนอกก็ได้ โดยจะเลือกเอาเฉพาะข้อมูลที่จะนำมาทำ Data Mining เท่านั้น
↳ **ตามดูรูป ชีวิตฐานะหก步, ศูนย์กลางก้าว**

เช่น ต้องการทำ Data Mining เพื่อหาลักษณะลูกค้าที่ซื้อรถสปอร์ต ไม่ควรนำรายชื่อพนักงานขายเข้ามาเกี่ยวข้อง สิ่งที่ควรนำมาพิจารณาสำหรับกรณีนี้อาจพิจารณาทั้ง **ข้อมูลลูกค้าและข้อมูลเกี่ยวกับการขาย เช่น เพศ อายุ รายได้ อาชีพ ยี่ห้อรถยนต์ ประเภทรถยนต์ ลักษณะเครื่องยนต์ เป็นต้น**

→ **หก步 2A ก้าว**

2. **Data Preprocessing** เป็นการทำ **ข้อมูลให้มีคุณภาพดี, ทำรูปแบบให้สอดคล้องกัน, ฟื้นฟูค่าทางสถิติ, ไนท์ กะ น้ำหนัก ดาต้า [เข้าคุณภาพให้วันดี]**
→ **หก步 3 ห้าว ก้าว**

→ **หก步 4 ห้าว ก้าว**

3. **Data Transformation** เป็นการสร้างข้อมูลชุดใหม่ที่มาจากการนำข้อมูลชุดเดิม ซึ่งทำเพื่อ **แปลงข้อมูลให้สอดคล้องกับ Model ที่ใช้ได้ เช่น การ Map ค่ามาเพื่อคำนวณเป็นต้น**
↳ **หก步 5 ห้าว ก้าว**

→ **หก步 6 ห้าว ก้าว**

2. Data Preprocessing [ទົກສອງຂໍ້ຕົກ]

- ດູກ່າ outlier (ໂຄສູວພິດໄກຍີ, ທັ່າພິດນາຍີ) ຕື່ຕະຫຼາກຂຶ້ນຫຼຸກເນົາກະຈິປ່ານການເຄີຍການໃຫ້
- ກໍານົດລືບສິ່ງ
- ກໍານົດຈົບຖະລວດໂລກລົ້ອກັກ 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.

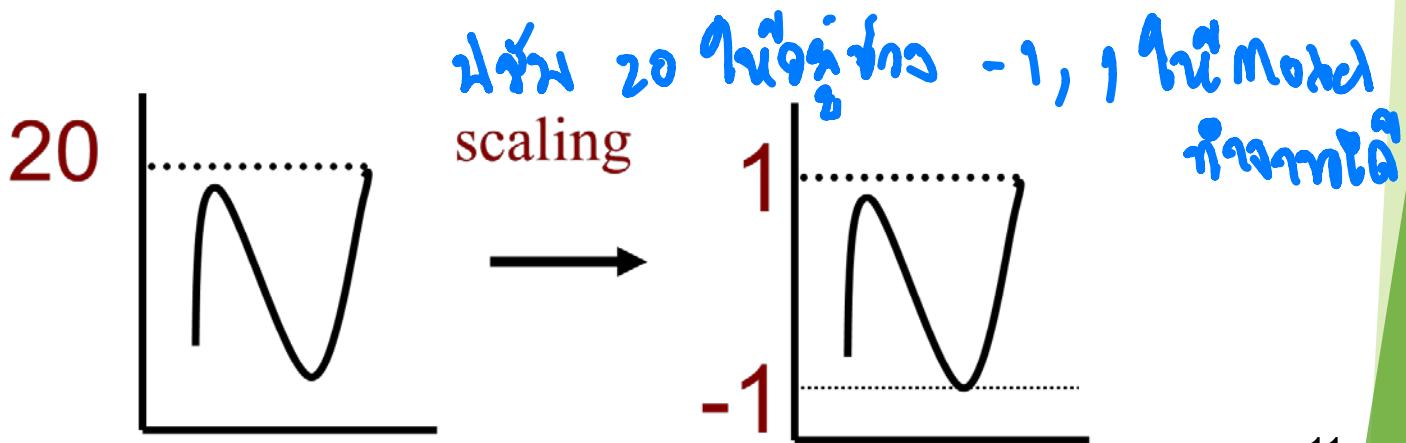
Data Transformation

- ▶ 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 ឃុំជិត្យា → តម្លៃទេរក / តក្ឋៅ

- 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 ให้เหมาะสมกับวัตถุประสงค์ในการทำ



ลี 3 หนนนวѓກ

ສාසඟ

① Predictive Model

③ Link analysis

② Database segmentation

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

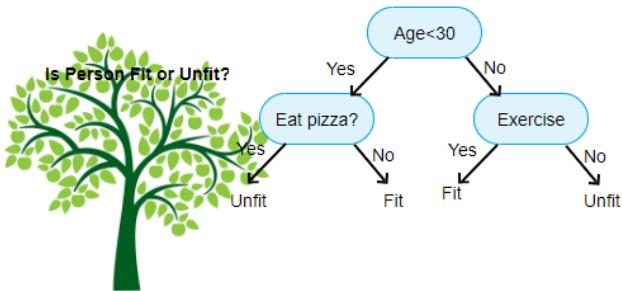
An example of classification

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

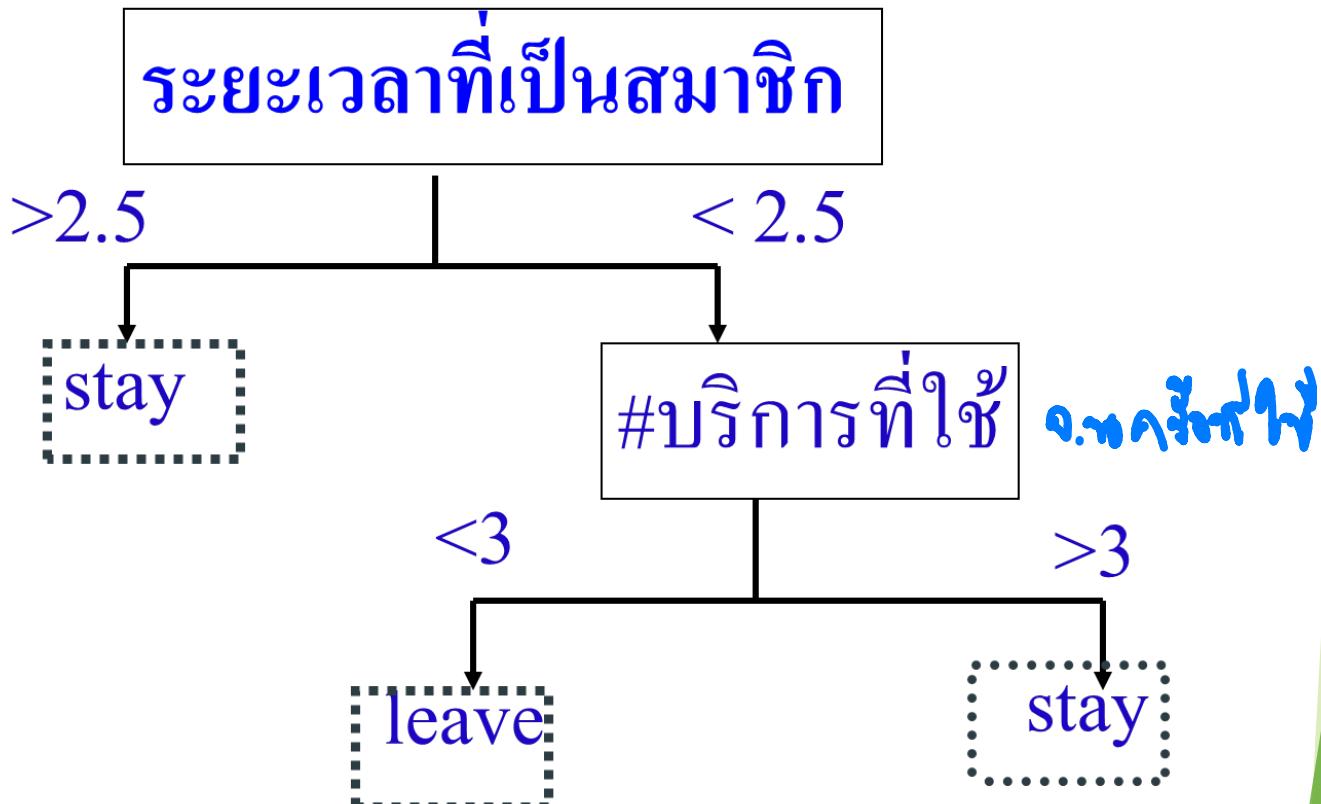
Input

Output

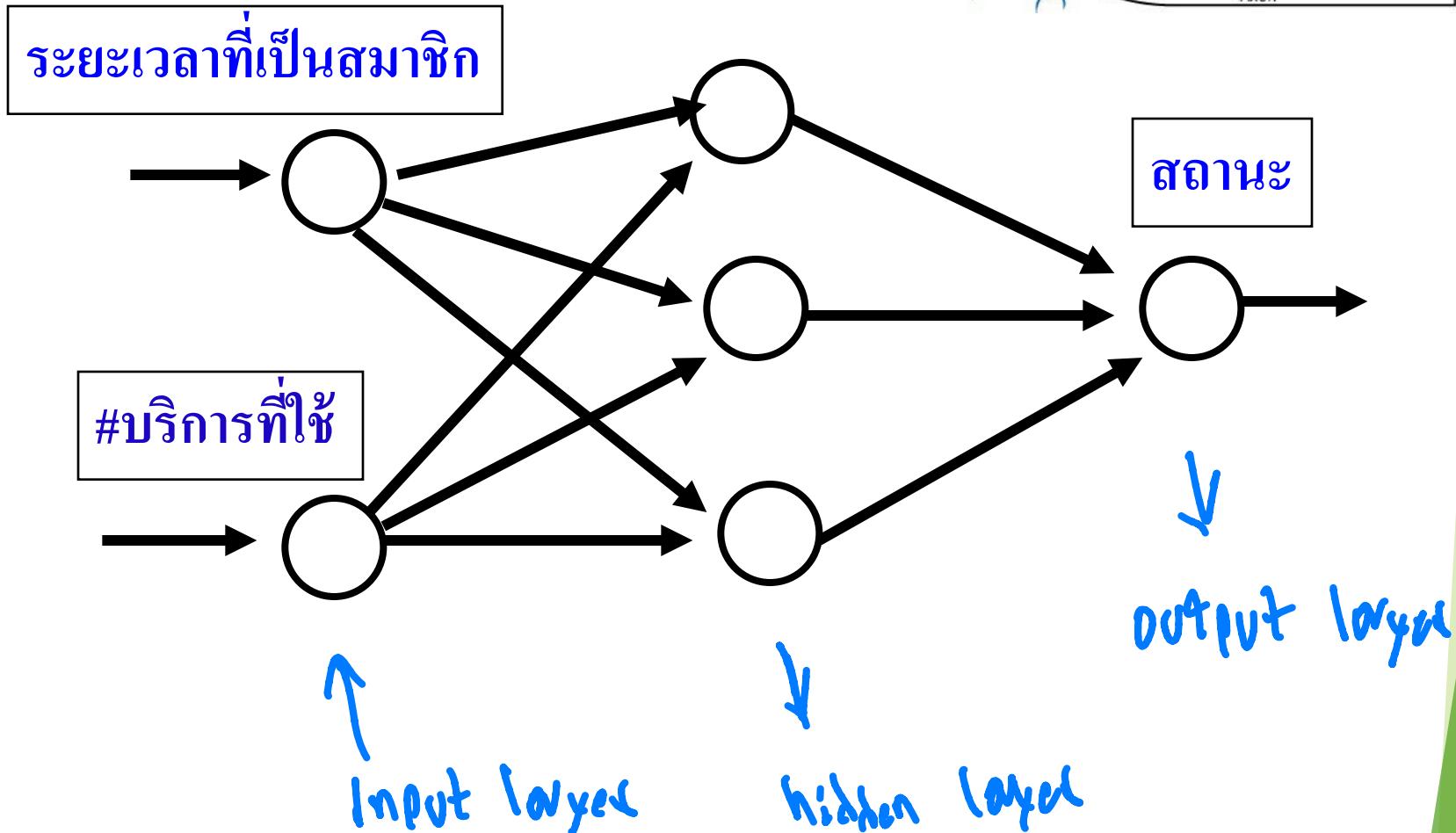
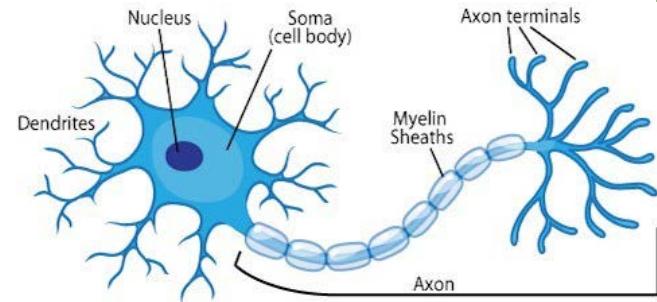
An example of classification by decision tree



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 linear 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

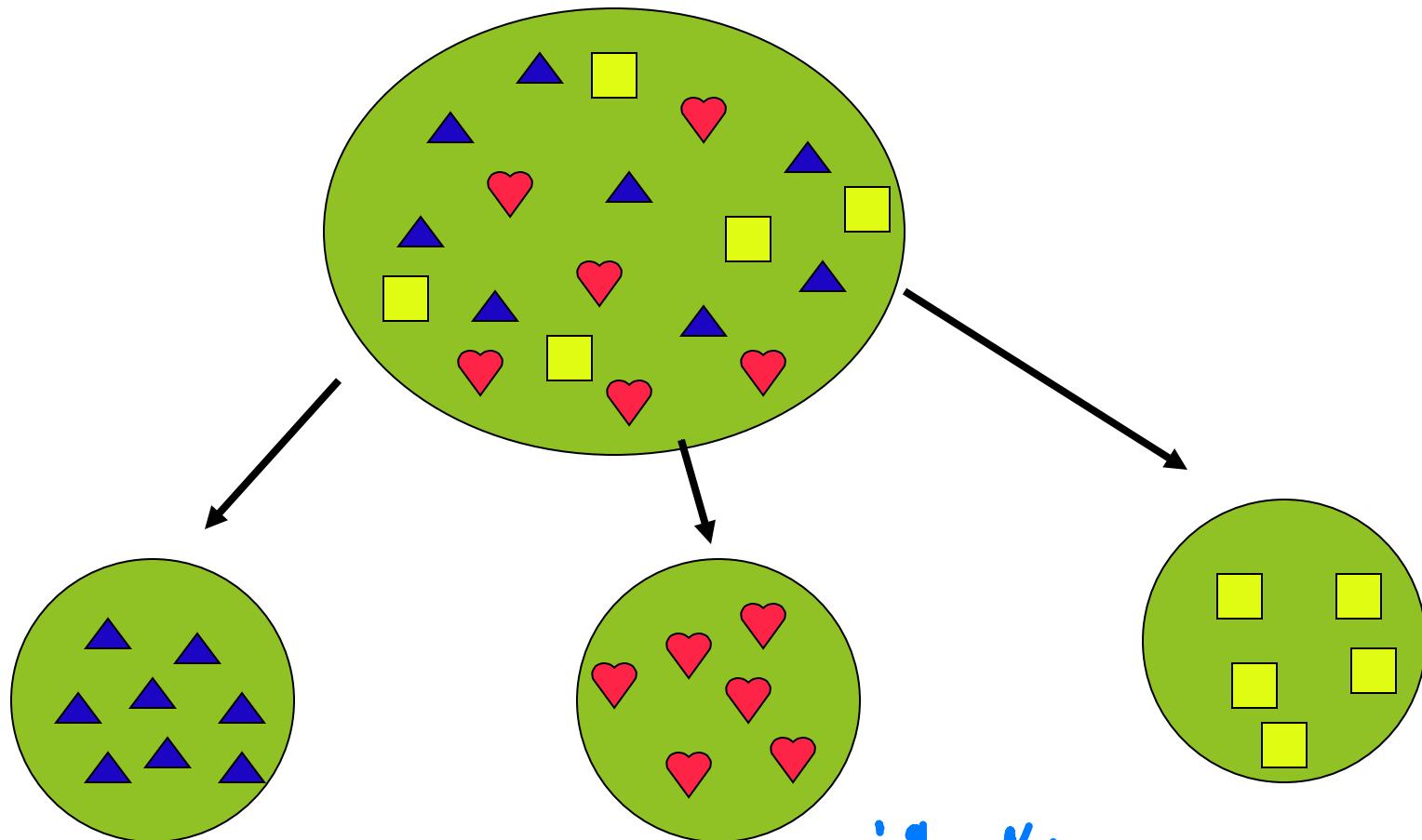
- *ផែន វទ្ធសាស្ត្រ, learning
ពេលកម្រិតបាន ដឹងទូរការ*

• 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

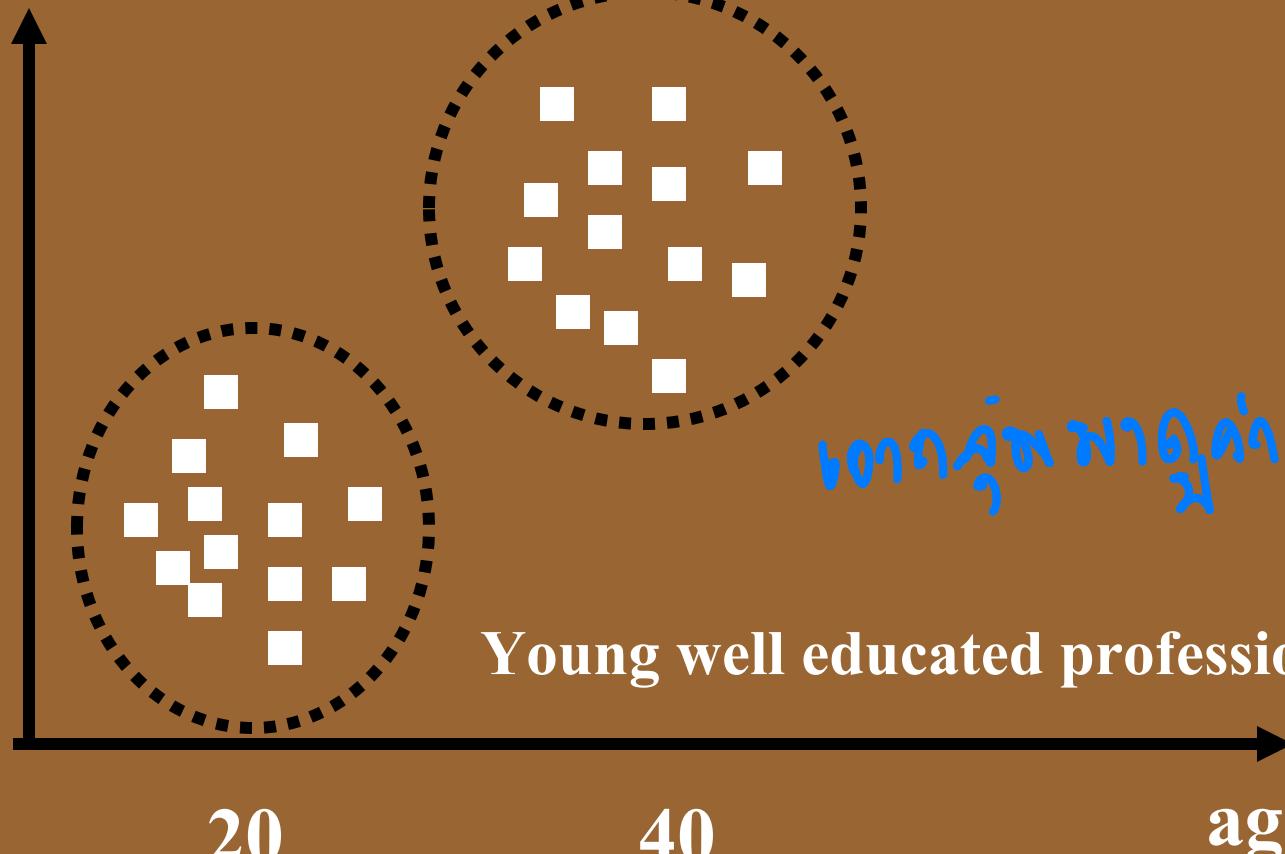
Database segmentation



សេវាសម្រាប់បង្កើតការងារទូទៅជាអ្នកអាជីវកម្ម

Database segmentation

Income (thousands)



Data Mining Process

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

► Associations Discovery  សំណង់ការងារ
នូវគោលការណ៍ ការងារទាំងអស់

► Sequential Pattern Discovery

២ នឹងការកិច្ចការរាយ និងបញ្ជូនការទាំងអស់
ដើម្បីបង្កើតការទាំងអស់

A. ព័ត៌មាន

ផែតាម ពីរធនជាតិ និងពីរធនជាតិ

- A និង B នឹងការកិច្ចការរាយ និងបញ្ជូនការ

- B និង C នឹងការកិច្ចការរាយ និងបញ្ជូនការ

ຕາມຄສນ ປີ (Case Study)

ພາກອອ ໄກສາ ອມ ດິເວລາ

ນີ້ ວັດຖະນາຫຼາດ ກົນ (Virtue Model)

ຕົ້ນ ຂອບເລື່ອດຳ

การทำ Data Mining แบบ Link Analysis หรือ การวิเคราะห์ลิงก์ เป็นเทคนิคหนึ่งในการทำ Data Mining ที่ใช้ในการวิเคราะห์โครงสร้างของข้อมูลที่เชื่อมโยงกัน โดยมุ่งเน้นไปที่การศึกษาความสัมพันธ์ระหว่างหน่วยข้อมูลต่างๆ ผ่านลิงก์ที่เชื่อมโยงกัน เช่น ลิงก์ในเว็บไซต์, การเชื่อมต่อระหว่างผู้คนในเครือข่ายสังคม หรือการเชื่อมต่อระหว่างเอกสารต่างๆ

หลักการทำงานเบื้องหลัง

- **กราฟ (Graph):** ข้อมูลที่เชื่อมโยงกันมักจะถูกแสดงในรูปแบบของกราฟ โดยแต่ละหน่วยข้อมูลจะเป็นจุดยอด (node) และลิงก์ที่เชื่อมโยงกันจะเป็นเส้นเชื่อม (edge)
- **การวิเคราะห์โครงสร้าง:** เราจะวิเคราะห์โครงสร้างของกราฟนี้เพื่อค้นหาแพทเทิรินท์ที่ซ่อนอยู่ เช่น กลุ่มของหน่วยข้อมูลที่เชื่อมโยงกันอย่างใกล้ชิด (community detection), หน่วยข้อมูลที่มีอิทธิพลสูง (influential nodes), หรือเส้นทางที่ลัดที่สุดระหว่างหน่วยข้อมูลสองหน่วย (shortest path)

ประโยชน์ของการทำ Link Analysis

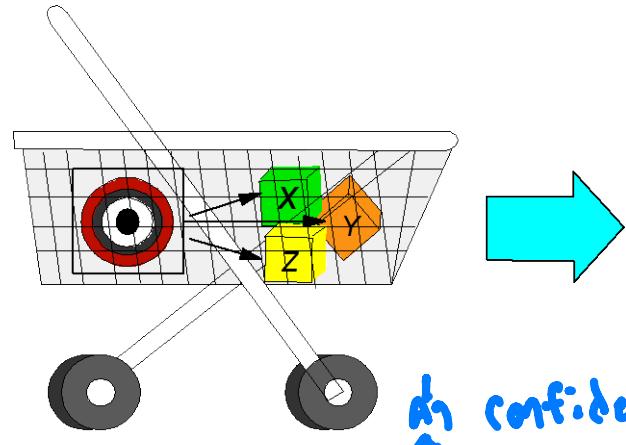
- **ค้นหาความสัมพันธ์ที่ซ่อนอยู่:** ช่วยในการค้นพบความสัมพันธ์ที่ซ่อนอยู่ระหว่างข้อมูลที่เราอาจไม่เคยสังเกตเห็นมาก่อน
- **ทำความเข้าใจโครงสร้างของข้อมูล:** ช่วยให้เราเข้าใจโครงสร้างของข้อมูลได้อย่างลึกซึ้งมากขึ้น
- **ค้นพบกลุ่มของข้อมูลที่คล้ายกัน:** ช่วยในการจัดกลุ่มข้อมูลที่คล้ายกันออกเป็นกลุ่มๆ
- **ค้นหาข้อมูลที่พิเศษ:** ช่วยในการค้นพบข้อมูลที่พิเศษหรือข้อมูลที่อาจเป็นอันตราย
- **ประยุกต์ใช้ในหลายสาขา:** สามารถนำไปประยุกต์ใช้ได้ในหลายสาขา เช่น การตลาดดิจิทัล, การวิเคราะห์เครือข่ายสังคม, การวิเคราะห์ข้อมูลทางชีววิทยา, และการวิเคราะห์ข้อมูลทางเศรษฐศาสตร์

Associations Discovery

การค้นพบสารสนเทศที่มีความสัมพันธ์ที่เชื่อมโยงกัน เช่น ลูกค้าผู้ชายซื้อผ้าอ้อม
เพิ่มเพิ่บ่ายวันศุกร์มีแนวโน้มจะซื้อเบียร์ด้วย

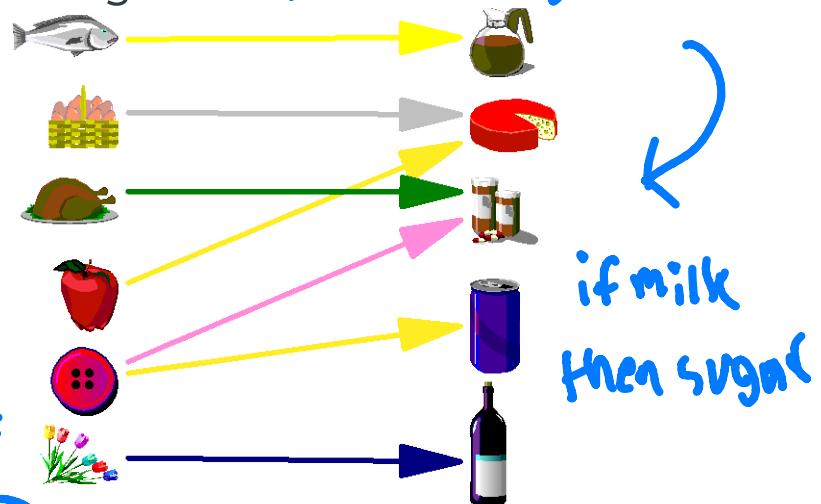
Association rule

Detect sets of attributes that frequently co-occur, and rules among them EX: Milk \rightarrow sugar - 例程ที่หาก if then



ค่า confidence

confidence 98% of people who buy Milk also buy Sugar.



if milk
then sugar

26 support 65% of all grocery shoppers buy both

ค่า support

กฎเชื่อมโยง (Association Rule) คืออะไร?

กฎเชื่อมโยง (Association Rule) เป็นหนึ่งในเทคโนโลยีการขุดข้อมูล (Data Mining) ที่ใช้ในการค้นหาความสัมพันธ์ที่ซ่อนอยู่ระหว่างข้อมูลรายการต่างๆ ในฐานข้อมูลขนาดใหญ่ โดยจะแสดงความสัมพันธ์นี้ในรูปแบบของกฎ "ถ้า...แล้ว..."

:f ... หมาย...

ตัวอย่างง่ายๆ:

สมมติว่าเรามีฐานข้อมูลรายการสินค้าที่ลูกค้าซื้อในชูปอร์มาร์เก็ต เราอาจพบกฎเชื่อมโยงดังนี้:

- ถ้า ลูกค้าซื้อนม แล้ว มีโอกาสสูงที่ลูกค้าจะซื้อบนมปั่นด้วย
- ถ้า ลูกค้าซื้อเบียร์ แล้ว มีโอกาสสูงที่ลูกค้าจะซื้อเนื้อฝรั่งทอดด้วย

องค์ประกอบหลักของกฎเชื่อมโยง:

- Antecedent** (ส่วนก่อนเหตุ): รายการสินค้าหรือเหตุการณ์ที่เกิดขึ้นก่อน
- Consequent** (ส่วนผล): รายการสินค้าหรือเหตุการณ์ที่ตามมา
- Support**: สัดส่วนของธุรกรรมที่ประกอบด้วยทั้ง antecedent และ consequent
- Confidence**: ความน่าจะเป็นที่ consequent จะเกิดขึ้น เมื่อ antecedent เกิดขึ้นแล้ว

ตัวอย่างกฎเชื่อมโยงในทางคณิตศาสตร์:

{นม, ไข่} \rightarrow {ขนมปัง}

- Support**: สมมติว่ามี 100 ใบเสร็จ และมี 20 ใบเสร็จที่ลูกค้าซื้อทั้งนมและน้ำตาล Support ก็คือ 20% (แสดงว่ามี 20% ของใบเสร็จที่ตรงกับกฎนี้)
- Confidence**: สมมติว่าใน 20 ใบเสร็จที่ลูกค้าซื้อทั้งนมและน้ำตาล มี 18 ใบเสร็จที่ลูกค้าซื้อน้ำตาลด้วย Confidence ก็คือ 90% (แสดงว่าเมื่อลูกค้าซื้อทั้งนม มีโอกาส 90% ที่จะซื้อน้ำตาลด้วย)

- Support**: 20% (หมายความว่า มี 20% ของธุรกรรมที่ลูกค้าซื้อทั้งนม ไข่ และขนมปัง)
- Confidence**: 60% (หมายความว่า เมื่อลูกค้าซื้อทั้งนมและไข่ มีโอกาส 60% ที่จะซื้อบนมปังด้วย)

Data Mining Process

4. การวิเคราะห์ผลลัพธ์ที่ได้ (Analysis of Results)

- ▶ เป็นการวิเคราะห์และตีความหมายจากผลที่ได้

ນາທຳ ຂອງ ພາຕິຈຸງກາ ກົດໃນກາລົມຕະຫຼາໄສ

5. การนำสารสนเทศไปใช้ประโยชน์ (Assimilation of Knowledge)

- ▶ การนำความรู้ที่ได้ไปใช้เป็นขั้นตอนสุดท้ายของการบวนการทั้งหมด
- ▶ รวบรวมความเข้าใจในแบบจำลองและการวิเคราะห์ผลลัพธ์ที่ได้ มารวมเข้ากับส่วนความรู้ทางธุรกิจ เพื่อที่จะนำเสนอถึงวิธีการที่จะนำผลที่ได้นี้ไปใช้ให้เกิดประโยชน์

ເພີ້ມໃຫຍ່ການໃຊ້ ແຫ່ງກໍາໄປ

ขั้นตอนที่ 4: การวิเคราะห์ผลลัพธ์ (Analysis of Results)

→ วิเคราะห์ผลลัพธ์, ติดตาม

ในขั้นตอนนี้ เราจะทำการ วิเคราะห์และตีความหมาย จากข้อมูลที่ได้จากการทำ Data Mining โดยจะเน้นไปที่การ ค้นหาความสัมพันธ์ กฎเกณฑ์ หรือรูปแบบ ที่ช่องอยู่ในข้อมูลเหล่านั้น ตัวอย่างเช่น

- การค้นหากฎเชื่อมโยง: เช่น การค้นพบว่าลูกค้าที่ซื้อนมมักจะซื้อขนมปังด้วย
- การสร้างแบบจำลอง: เช่น การสร้างแบบจำลองเพื่อกำหนดพฤติกรรมการซื้อของลูกค้า
- การวิเคราะห์กลุ่มลูกค้า: การแบ่งกลุ่มลูกค้าออกเป็นกลุ่มย่อยที่มีลักษณะคล้ายกัน

เป้าหมายหลักของขั้นตอนนี้คือ:

- เผ้าใจความหมาย: ทำความเข้าใจว่าผลลัพธ์ที่ได้มีความหมายอะไรบ้าง
- ตรวจสอบความถูกต้อง: ตรวจสอบว่าผลลัพธ์ที่ได้มีความถูกต้องและนำไปใช้ได้หรือไม่
- ค้นหาความรู้ใหม่: ค้นพบความรู้ใหม่ๆ ที่ไม่เคยทราบมาต่อจากข้อมูล

ขั้นตอนที่ 5: การนำความรู้ไปประยุกต์ใช้ (Assimilation of Knowledge)

เมื่อเราได้ผลลัพธ์จากการวิเคราะห์แล้ว ขั้นตอนต่อไปคือการ นำความรู้ที่ได้ไปใช้ประโยชน์ ในการตัดสินใจทางธุรกิจหรือแก้ไขปัญหาต่างๆ ตัวอย่างเช่น

- การปรับปรุงกลยุทธ์ทางการขาย: หากพบว่าลูกค้าที่ซื้อผลิตภัณฑ์ A มักจะซื้อผลิตภัณฑ์ B ด้วย เรายาจัดโปรโมชันส่วนลดหรือการซื้อสินค้าสองรายการร่วมกัน
- การตัดแพทเทิลิกันที่ใหญ่: หากพบว่ามีกลุ่มลูกค้าที่มีความต้องการผลิตภัณฑ์ประเภทใหม่ เราอาจ พัฒนาผลิตภัณฑ์ที่เข้ามาเพื่อตอบสนองความต้องการของกลุ่มลูกค้านั้น
- การปรับเปลี่ยนกระบวนการทำงาน: หากพบว่ามีขั้นตอนใดในกระบวนการการทำงานที่มีประสิทธิภาพ เราก็อาจปรับปรุงกระบวนการนั้นให้ดีขึ้น

เป้าหมายหลักของขั้นตอนนี้คือ:

- สร้างมูลค่าเพิ่ม: นำความรู้ที่ได้ไปสร้างมูลค่าเพิ่มให้กับองค์กร
- ปรับปรุงการทำงาน: ใช้ความรู้ที่ได้ในการตัดสินใจทางธุรกิจ ให้อย่างมีประสิทธิภาพ
- แก้ไขปัญหา: ใช้ความรู้ที่ได้ในการแก้ไขปัญหาต่างๆ ที่เกิดขึ้น

สรุป:

ทั้งสองขั้นตอนนี้มีความสำคัญต่อกระบวนการ Data Mining อย่างมาก ขั้นตอนที่ 4 ช่วยให้เราเข้าใจ ข้อมูล และขั้นตอนที่ 5 ช่วยให้เราสามารถนำข้อมูลนั้นไปใช้ประโยชน์ได้จริง การทำงานทั้งสองขั้นตอนนี้จะช่วยให้เราสามารถนำเทคโนโลยี Data Mining ไปประยุกต์ใช้ในการทำงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

Data Mining Process

Mining
Objective ①



Data sources
Databases, flat files,
Complex data



Data Warehouses



2.1 **Data Preparation** 2.2
(Data Selection , Data Preprocessing
and Data Transformation) 2.3



2.2



Data Mining

③



Assimilation of knowledge

⑤



Knowledge

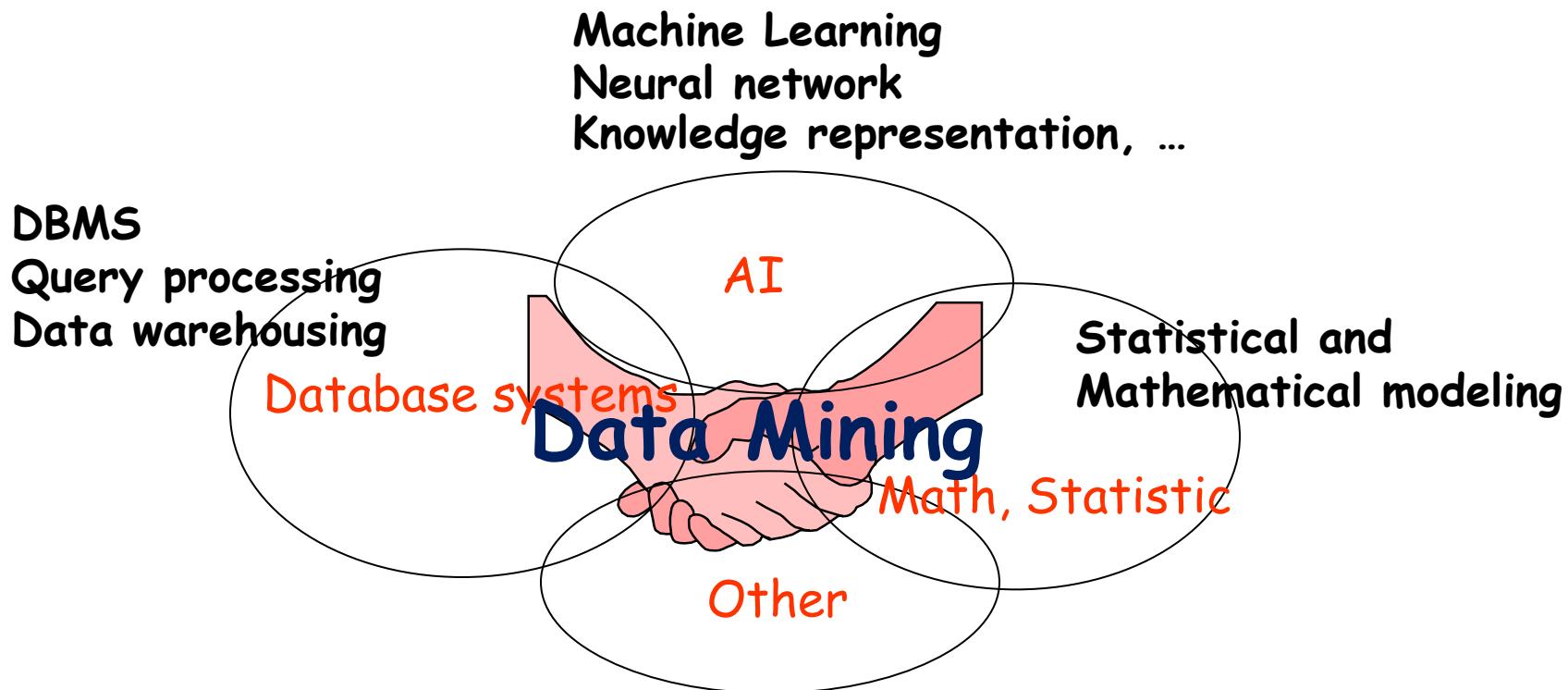
④

Analysis of Results



Data mining is multidisciplinary

ជាសាស្ត្រប្រចាំខ្លួន

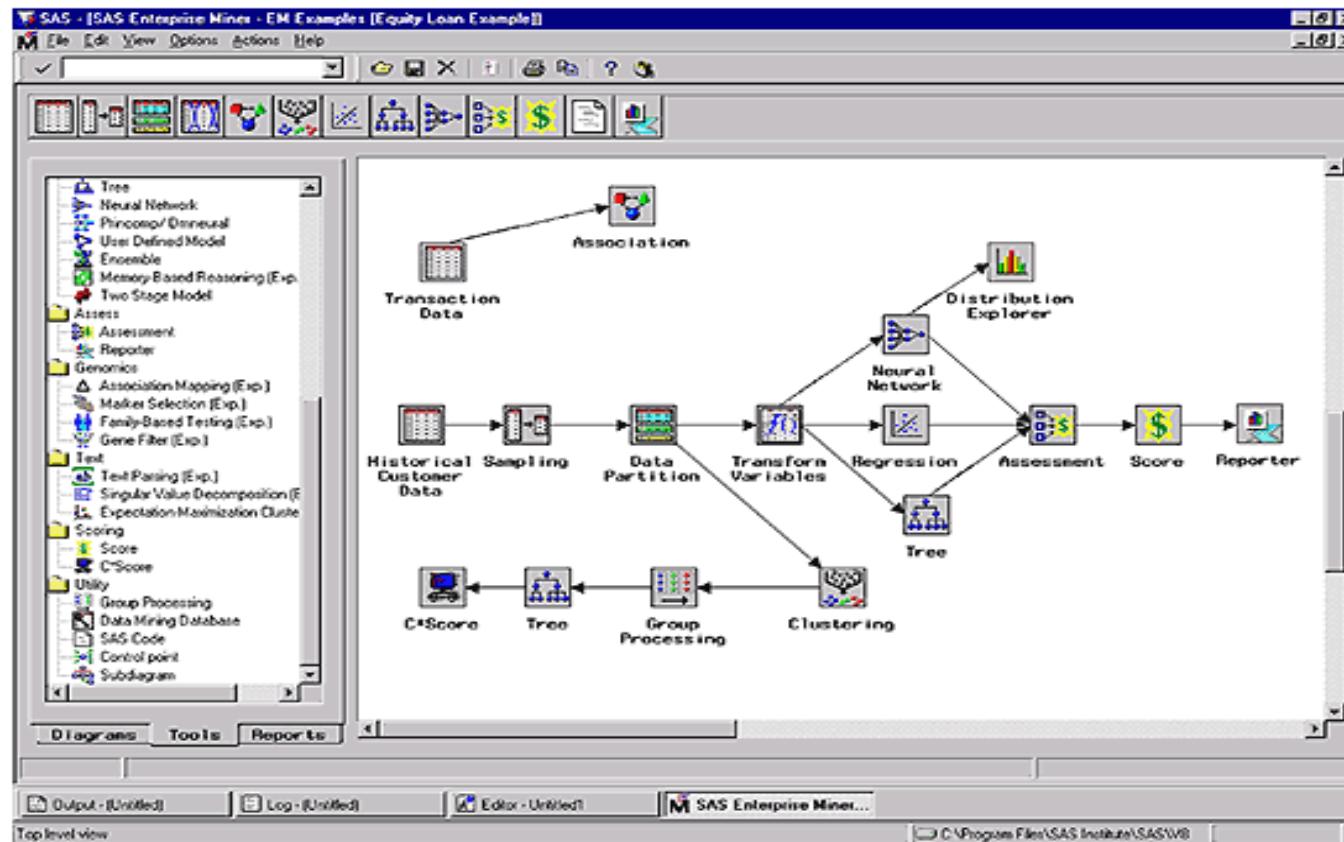


Data Mining Software

- ▶ 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