



การทำเหมืองข้อมูล (Data Mining)

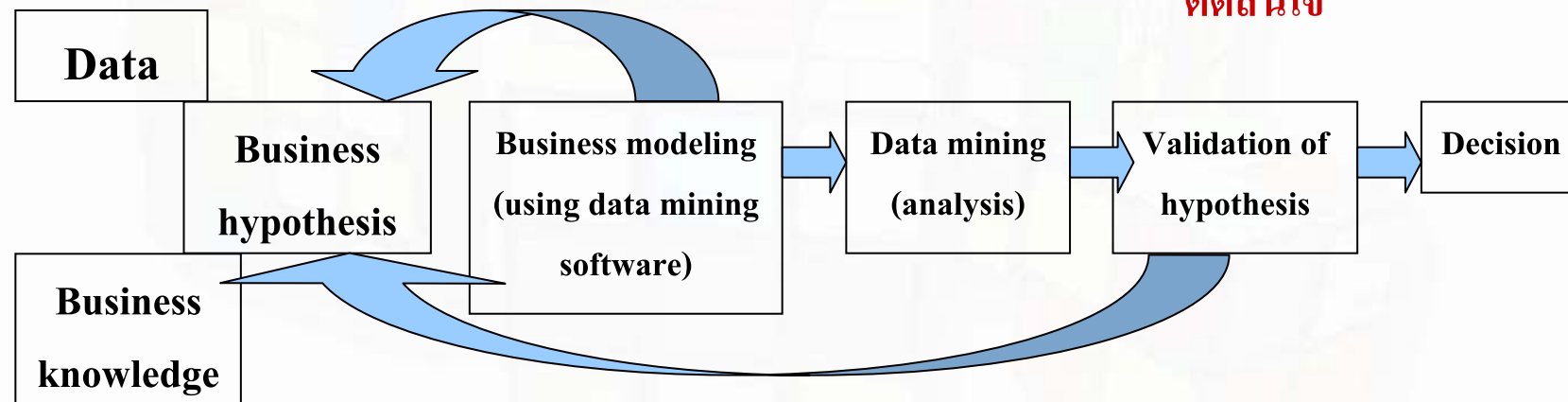
- แนวคิดเกี่ยวกับเหมืองข้อมูล
- เทคนิคการทำ Data Mining
- การประยุกต์ใช้ Data Mining

Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

1. ในบริษัทขนาดกลางถึงขนาดเล็ก ขบวนการทำ data mining โดยทั่วไปจะเริ่มจากการตั้งสมมติฐานทางธุรกิจ ตามความรู้และความเข้าใจของ user ที่มีต่อธุรกิจ

3. หลังจากตรวจสอบ
แก้ไขสมมติฐานในขั้น
สุดท้ายแล้ว user ก็
ตัดสินใจ



2. ใช้ระบบ data mining tools โดย user สร้าง model แล้วกลั่นกรองสมมติฐาน ตามด้วยการวิเคราะห์ ซึ่งขบวนการนี้อาจจะต้องมีการทำซ้ำหลายๆ ครั้ง



DATA MINING คืออะไร

- ปัจจุบันระบบสนับสนุนข้อมูลในการตัดสินใจได้เข้ามามีอิทธิพลในการรวบรวมข้อมูลและปรับค่าข้อมูลในคลังสินค้า ซึ่งฐานข้อมูลขนาดใหญ่นี้จะประกอบไปด้วยข้อมูลเป็นพันๆ ล้านไบต์ ยากแก่การค้นหาได้อย่างทันกาลด้วยวิธี DBMS (Database Management System) โดยทั่วไป ข้อมูลที่เป็นที่สนใจของผู้บริหารธุรกิจวันนี้สามารถจะค้นหาได้ง่ายขึ้นแล้ว ซึ่งจะประโยชน์อย่างยิ่งในการค้นหาข้อมูลที่ต้องการในมหาสมุทรข้อมูลเพื่อนำมาเทียบเคียงและดูแนวโน้ม และนำข้อมูลที่เป็นของบริษัทส่งกลับให้ผู้บริหารตัดสินใจได้อย่างทันกาล



DATA MINING คืออะไร

- Data Mining คือ ชุด software วิเคราะห์ข้อมูลที่ได้ถูกออกแบบมาเพื่อระบบสนับสนุนการตัดสินใจของผู้ใช้ มันเป็น software ที่สมบูรณ์ทั้งเรื่องการค้นหา การทำรายงาน และโปรแกรมในการจัดการ Executive Information System (EIS) หรือระบบข้อมูลสำหรับการตัดสินใจในการบริหาร ซึ่งเป็นเครื่องมือชิ้นใหม่ที่สามารถค้นหาข้อมูลในฐานข้อมูลขนาดใหญ่หรือข้อมูลที่เป็นประโยชน์ในการบริหาร ซึ่งเป็นการเพิ่มคุณค่าให้กับฐานข้อมูลที่มีอยู่



DATA MINING คืออะไร

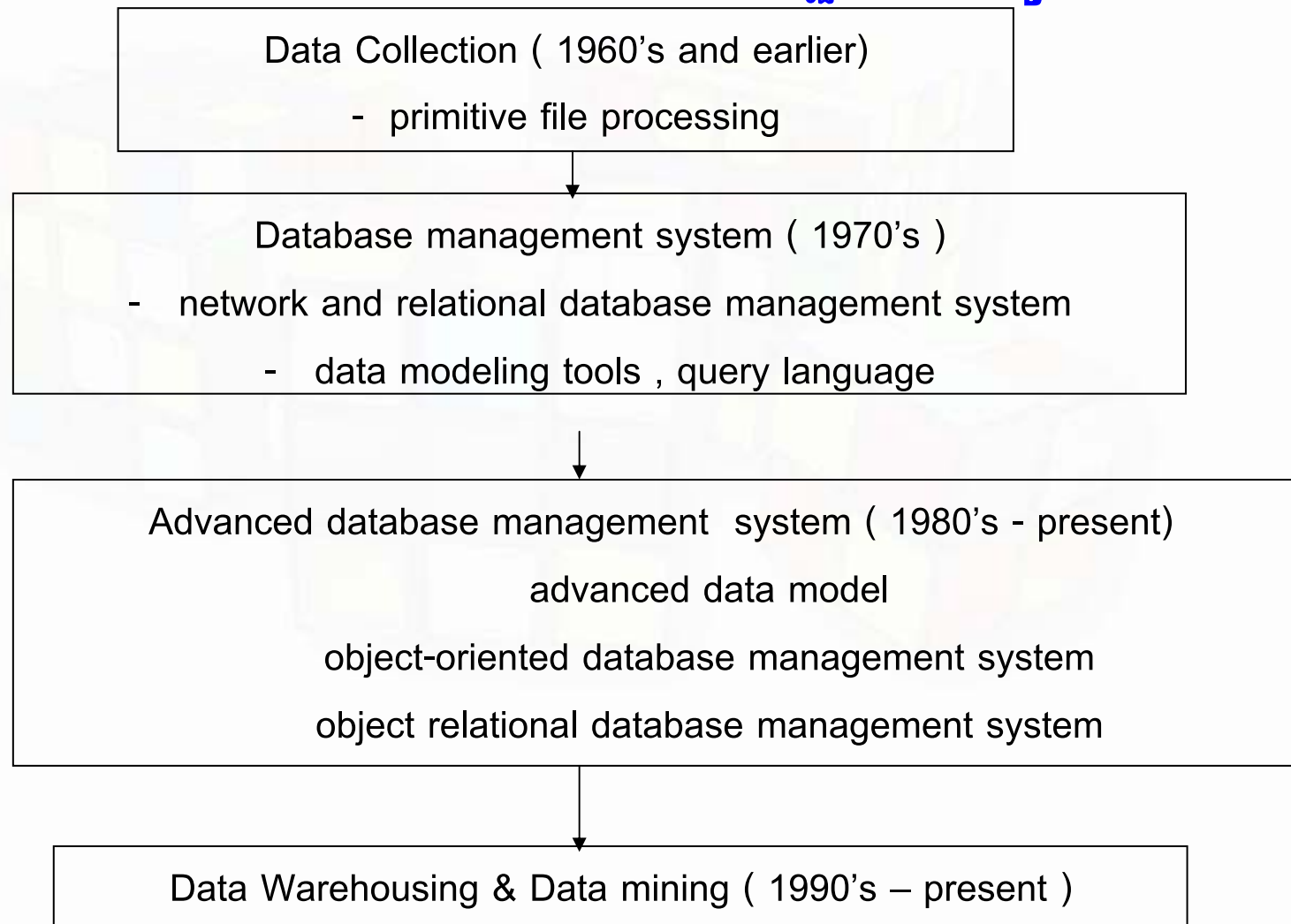
- Data Mining คือการวิเคราะห์ข้อมูล เพื่อแยกประเภท จำแนกรูปแบบ และความสัมพันธ์ของข้อมูลจากคลังข้อมูล และนำสารสนเทศที่ได้ไปใช้ในการตัดสินใจธุรกิจ



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

วิวัฒนาการของเทคโนโลยีฐานข้อมูล





ทำไมจึงต้องมี Data Mining

1. ข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลหากเก็บไว้เฉย ๆ ก็จะไม่เกิดประโยชน์ดังนั้น
จึงต้องมีการสกัดสารสนเทศไปใช้

การสกัดสารสนเทศ หมายถึง การคัดเลือกข้อมูลออกมา ใช้งานในส่วนที่เรา
ต้องการ

2. ในอดีตเราใช้คนเป็นผู้สืบค้นข้อมูลต่าง ๆ ในฐานข้อมูลซึ่งผู้สืบค้นจะทำการ
สร้างเงื่อนไขขึ้นมาตามภูมิปัญญาของผู้สืบค้น



ทำไมจึงต้องมี Data Mining

3. ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูลเดียวอาจไม่ให้ความรู้เพียงพอ และลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูงและมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นที่จะต้องรวบรวมฐานข้อมูลหลาย ๆ ฐานข้อมูล เข้าด้วยกัน เรียกว่า “ คลังข้อมูล ” (Data Warehouse)

ดังนั้นเราจึงจำเป็นต้องใช้ Data Mining ในการดึงข้อมูลจาก ฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะนำข้อมูลนั้นมาใช้งานให้เกิดประโยชน์สูงสุด



ปัจจัยที่ทำให้ Data Mining เป็นที่ได้รับความนิยม

จำนวนและขนาดข้อมูลขนาดใหญ่ถูกผลิตและขยายตัวอย่างรวดเร็ว การสืบค้นความรู้จะมีความหมายก็ต่อเมื่อฐานข้อมูลที่ใช้มีขนาดใหญ่มาก ปัจจุบันมีจำนวนและขนาดข้อมูลขนาดใหญ่ที่ขยายตัวอย่างรวดเร็ว โดยผ่านทาง Internet ดาวเทียม และแหล่งผลิตข้อมูล อื่น ๆ เช่น เครื่องอ่านบาร์โค้ด , เครดิตการ์ด , อีคอมเมิร์ซ



ปัจจัยที่ทำให้ Data Mining เป็นที่ได้รับความนิยม

ข้อมูลถูกจัดเก็บเพื่อนำไปสร้างระบบการสนับสนุนการตัดสินใจ (Decision Support System) เพื่อเป็นการง่ายต่อการนำข้อมูลมาใช้ในการวิเคราะห์เพื่อการตัดสินใจ ส่วนมากข้อมูลจะถูกจัดเก็บแยกมาจากระบบปฏิบัติการ (Operational System) โดยจัดอยู่ในรูปของคลังหรือเหมืองข้อมูล (Data Warehouse) ซึ่งเป็นการง่ายต่อการนำเอาไปใช้ในการสืบค้นความรู้



ปัจจัยที่ทำให้ Data Mining เป็นที่ได้รับความนิยม

ระบบ computer สมรรถนะสูงมีราคาต่ำลง เทคนิค Data Mining ประกอบไปด้วย Algorithm ที่มีความซับซ้อนและความต้องการการคำนวณสูง จึงจำเป็นต้องใช้งานกับระบบ computer สมรรถนะสูง ปัจจุบันระบบ computer สมรรถนะสูงมีราคาต่ำลง พร้อมด้วยเริ่มมีเทคโนโลยีที่นำเครื่อง microcomputer จำนวนมากมาเชื่อมต่อกันโดยเครือข่ายความเร็วสูง (PC Cluster) ทำให้ได้ระบบ computer สมรรถนะสูงในราคาต่ำ



ปัจจัยที่ทำให้ Data Mining เป็นที่ได้รับความนิยม

การแข่งขันอย่างสูงในด้านอุตสาหกรรมและการค้า เนื่องจากปัจจุบันมีการแข่งขันอย่างสูงในด้านอุตสาหกรรมและการค้า มีการผลิตข้อมูลไว้อย่างมากมายแต่ไม่ได้นำมาใช้ให้เกิดประโยชน์ จึงเป็นการจำเป็นอย่างยิ่งที่ต้องควบคุมและสืบค้นความรู้ที่ถูกซ่อนอยู่ในฐานข้อมูลความรู้ที่ได้รับสามารถนำไปวิเคราะห์เพื่อการตัดสินใจในการจัดการในระบบต่างๆ



ประเภทข้อมูลที่สามารถทำ Data Mining

Relational Database เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย entity-relationship (ER) model

Data Warehouses เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่ ๆ เดียวกัน



ประเภทข้อมูลที่สามารถทำ Data Mining

- Transactional Database ประกอบด้วยข้อมูลที่แต่ละทรานแซกชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปแบบ ชื่อลูกค้า และรายการสินค้าที่ลูกค้ารายนั้นซื้อ เป็นต้น
- Advanced Database เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่น ๆ เช่น ข้อมูลแบบ object-oriented , ข้อมูลที่เป็น text file , ข้อมูลมัลติมีเดีย , ข้อมูลในรูปแบบของ web



ลักษณะเฉพาะของข้อมูลที่สามารถทำ Data Mining

- ข้อมูลขนาดใหญ่ เกินกว่าจะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า หรือโดยการใช้ Database Management System (DBMS) ในการจัดการฐานข้อมูล
- ข้อมูลที่มาจากหลายแหล่ง โดยอาจรวบรวมมาจากหลายระบบปฏิบัติการ หรือหลาย DBMS เช่น Oracle , DB2 , MS SQL , MS Access เป็นต้น
- ข้อมูลที่มีโครงสร้างซับซ้อน เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย ข้อมูลเหล่านี้สามารถนำมาทำ Mining ได้เช่นกันแต่ต้องใช้เทคนิคการทำ Data Mining ขั้นสูง



ลักษณะเฉพาะของข้อมูลที่สามารถทำ Data Mining

- ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดเวลาที่ทำการ Mining หากข้อมูลที่มีอยู่นั้นเป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาจะต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้นไว้และนำฐานข้อมูลที่บันทึกไว้มาทำ Mining แต่เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลา จึงทำให้ผลลัพธ์ที่ได้จากการทำ Mining สมเหตุสมผลในช่วงเวลาหนึ่งเท่านั้น ดังนั้นเพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องเหมาะสมอยู่ตลอดเวลาจึงต้องทำ Mining ใหม่ทุกครั้งในช่วงเวลาที่เหมาะสม



คุณลักษณะ Data Mining

- การค้นหาข้อมูลโดยอาศัยเทคโนโลยีการทำเหมืองข้อมูล จะทำการค้นหาซึ่งถูกเก็บที่อยู่มากในคลังข้อมูล โดยบางครั้งอาจเป็นข้อมูลที่ถูกสะสมมาหลายปี
- การทำเหมืองข้อมูลมักใช้สถาปัตยกรรมแบบ Client/Server
- ผู้ใช้งานระบบสารสนเทศที่นำเทคโนโลยีการทำเหมืองข้อมูลมาช่วยจัดการข้อมูลไม่จำเป็นต้องมีทักษะการเขียนโปรแกรม เนื่องจากมีเครื่องมือช่วยค้นหาข้อมูลจากคลังข้อมูลได้อย่างรวดเร็ว



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

คุณลักษณะ Data Mining

- ผู้ใช้ต้องกำหนดขอบเขตและเป้าหมายของข้อมูลที่ต้องการค้นหาให้มีความชัดเจน



เทคนิคต่าง ๆ ของ Data Mining

1. Association rule Discovery

เป็นเทคนิคหนึ่งของ Data Mining ที่สำคัญ และสามารถนำไปประยุกต์ใช้ได้จริงกับงานต่าง ๆ หลักการทำงานของวิธีนี้ คือ การค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ หรือมากจากการวิเคราะห์การซื้อสินค้าของลูกค้าเรียกว่า “ Market Basket Analysis ” ซึ่งประเมินจากข้อมูลในตารางที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา ซึ่งการวิเคราะห์แบบนี้เป็นการใช้ “ กฎความสัมพันธ์ ” (Association Rule) เพื่อหาความสัมพันธ์ของข้อมูล



เทคนิคต่าง ๆ ของ Data Mining

ตัวอย่าง การนำเทคนิคนี้ไปประยุกต์ใช้กับงานจริง ได้แก่ ระบบแนะนำหนังสือให้กับลูกค้าแบบอัตโนมัติ ของ Amazon ข้อมูลการสั่งซื้อทั้งหมดของ Amazon ซึ่งมีขนาดใหญ่มากจะถูกนำมาประมวลผลเพื่อหาความสัมพันธ์ของข้อมูล คือ ลูกค้าที่ซื้อหนังสือเล่มหนึ่ง ๆ มักจะซื้อหนังสือเล่มใดพร้อมกันด้วยเสมอ ความสัมพันธ์ที่ได้จากกระบวนการนี้จะสามารถนำไปใช้คาดเดาได้ว่าควรแนะนำหนังสือเล่มใดเพิ่มเติมให้กับลูกค้าที่เพิ่งซื้อหนังสือจากร้าน ตัวอย่างเช่น

$\text{buys}(x, \text{database}) \rightarrow \text{buys}(x, \text{data mining}) [80\%, 60\%]$

หมายความว่า เมื่อซื้อหนังสือ database แล้วมีโอกาสที่จะซื้อหนังสือ data mining ด้วย 60 % และมีการซื้อทั้งหนังสือ database และหนังสือ data mining พร้อม ๆ กัน 80 %



เทคนิคต่าง ๆ ของ Data Mining

2. เทคนิค Classification & Prediction

เป็นกระบวนการสร้าง model จัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ ตัวอย่างเช่น จัดกลุ่มนักเรียนว่า ดีมาก ดี ปานกลาง ไม่ดี โดยพิจารณาจากประวัติและผลการเรียน หรือแบ่งประเภทของลูกค้าว่าเชื่อถือได้ หรือไม่โดยพิจารณาจากข้อมูลที่มีอยู่

เป็นเทคนิคในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะต่างๆที่ได้มีการกำหนดไว้แล้ว เทคนิคประเภทนี้เหมาะกับการสร้างแบบจำลองเพื่อการพยากรณ์ค่าข้อมูล ในอนาคต แบ่งออกเป็น 2 รูปแบบ



เทคนิคต่าง ๆ ของ Data Mining

2. เทคนิค Classification & Prediction

Prediction

เป็นการทำนายค่าที่ต้องการจากข้อมูลที่มีอยู่ตัวอย่างเช่น หายอดขายของเดือนถัดไปจากข้อมูลที่มีอยู่ หรือทำนายโรคจากอาการของคนไข้ในอดีต เป็นต้น



เทคนิคต่าง ๆ ของ Data Mining

2.1 แบบต้นไม้ (Decision Tree)

เป็นที่นิยมกันมากเนื่องจากเป็นลักษณะที่คนจำนวนมากคุ้นเคย ทำให้เข้าใจได้ง่าย มีลักษณะเหมือนแผนภูมิองค์กร โดยที่แต่ละโหนดแสดง attribute แต่ละกิ่งแสดงผลในการทดสอบ และลีฟโหนดแสดงคลาสที่กำหนดไว้



เทคนิคต่าง ๆ ของ Data Mining

ตัวอย่าง

บริษัทขนาดใหญ่แห่งหนึ่ง ทำธุรกิจอสังหาริมทรัพย์มีสำนักงานสาขาอยู่ประมาณ 50 แห่ง แต่ละสาขามีพนักงานประจำ เป็นผู้จัดการและพนักงานขาย พนักงานเหล่านี้แต่ละคนจะ ดูแลอาคารต่าง ๆ หลายแห่ง รวมทั้งลูกค้าจำนวนมาก บริษัทจำเป็นต้องใช้ระบบฐานข้อมูลที่กำหนดความสัมพันธ์ระหว่างองค์ประกอบเหล่านี้ เมื่อรวบรวมข้อมูลแบ่งเป็นตารางพื้นฐานต่าง ๆ เช่น ข้อมูลสำนักงานสาขา (Branch) ข้อมูลพนักงาน (Staff) ข้อมูลทรัพย์สิน (Property) และข้อมูลลูกค้า



เทคนิคต่าง ๆ ของ Data Mining

ตัวอย่าง(ต่อ)

(Client) พร้อมทั้งกำหนดความสัมพันธ์ (Relationship) ของข้อมูลเหล่านี้ เช่น ประวัติการเช่าบ้านของลูกค้า (Customer_rental) รายการให้เช่า (Rentals) รายการขายสินทรัพย์ (Sales) เป็นต้น ต่อมาเมื่อมีประจักษ์กรรมการผู้บริหารของบริษัท ส่วนหนึ่งของรายงานจากฐานข้อมูลสรุปว่า



เทคนิคต่าง ๆ ของ Data Mining

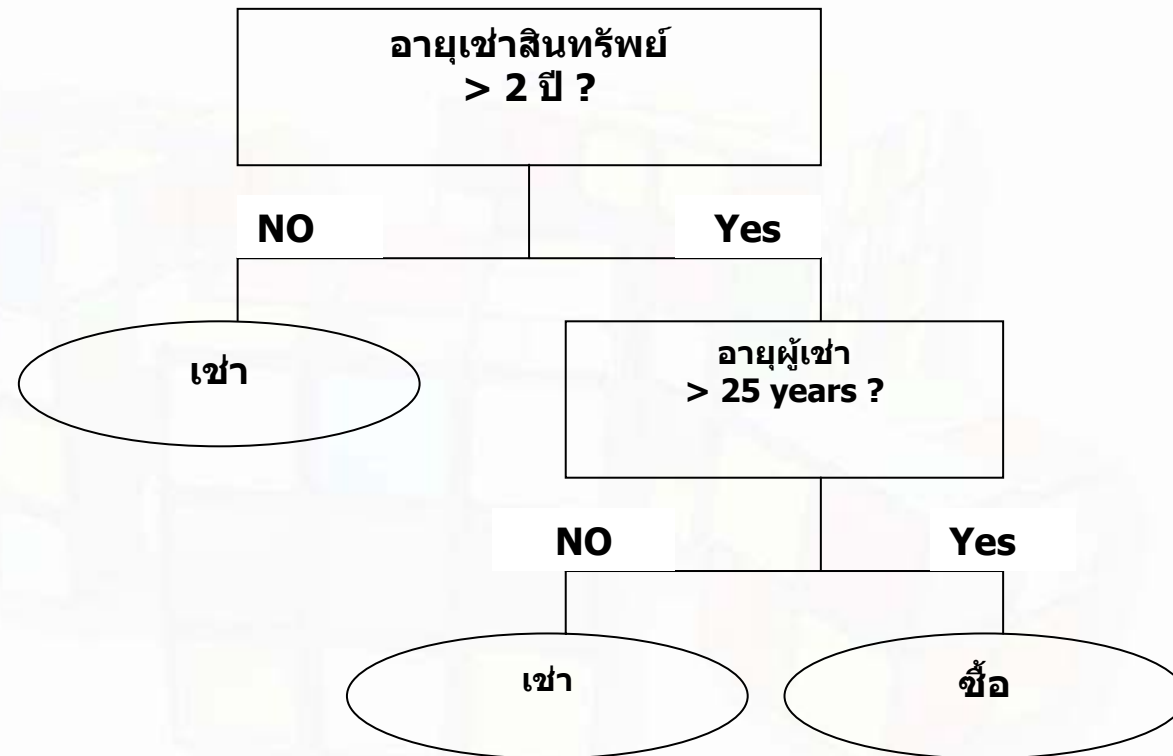
“ 40 % ของลูกค้าที่เข้าบ้านนานกว่าสองปี และมีอายุเกิน 25 ปี จะซื้อบ้านเป็นของตนเอง โดยกรณีเช่นนี้เกิดขึ้น 35 % ของลูกค้าผู้เข้าบ้านของบริษัท”

จากตัวอย่างดังกล่าวสามารถนำมาแสดงให้เห็นถึง Decision Tree สำหรับการวิเคราะห์ว่าลูกค้าบ้านเช่าจะมีความสนใจที่จะซื้อบ้านเป็นของตนเองหรือไม่ โดยใช้ปัจจัยในการวิเคราะห์คือ ระยะเวลาที่ลูกค้าได้เช่าบ้านมา และอายุของลูกค้าได้ดังรูปภาพ



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT



ตัวอย่างของ Decision Tree เพื่อวิเคราะห์โอกาสที่ลูกค้าบ้านเช่าจะซื้อบ้าน



เทคนิคต่าง ๆ ของ Data Mining

จากตัวอย่างพฤติกรรมกรเช่าและซื้อบ้านข้างต้น ลองมาตัวอย่างที่เป็นรูปธรรมมากขึ้น ในตาราง Business_Info แสดงถึงรายการทั้งหมดเกี่ยวกับลูกค้าบ้านเช่าของบริษัท โดยมีรายละเอียดเกี่ยวกับอายุ และระยะเวลาเช่า รวมทั้งการซื้อบ้านของลูกค้าแต่ละราย ดังนี้



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

ตาราง Business_Info

Age	Rent_Period	Buy
23	3	No
36	1.5	No
20	1.5	No
27	2	Yes
20	1	No
50	2.5	Yes
36	1	No
36	2	Yes
22	2.5	no



เทคนิคต่าง ๆ ของ Data Mining

SQL สำหรับ Decision Tree ของตัวอย่างนี้แบ่งเป็น 2 ชุด สำหรับปัจจัยแต่ละอย่าง

1. SQL สำหรับ root node ดังนี้

```
SELECT B.Rent_Period , B.Buy , COUNT(*)
```

```
FROM Business_Info B
```

```
WHERE B.Rent_Period > 2
```

```
GROUP BY B.Rent_Period , B.Buy;
```

ผลลัพธ์ของ SQL นี้คือ



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

Rent_Period	Buy	Yes	No
1	0	2	
1.5	0	2	
2	2	0	
2.5	1	1	
3	0	1	



เทคนิคต่าง ๆ ของ Data Mining

2. SQL สำหรับ node ที่เป็น child ทางขวาของ root คือ

```
SELECT B.Age , B.Buy , COUNT(*)
```

```
FROM Business_Info B
```

```
WHERE B.Age > 25
```

```
GROUP BY B.Age , B.Buy;
```

ผลลัพธ์ของ SQL นี้คือ



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

Rent_Period	Buy	Yes	No
20	0	2	
22	0	1	
23	2	1	
27	1	0	
36	1	2	
50	1	0	

ผลลัพธ์ที่ได้จากแต่ละ โหนดของ Decision Tree เรียกว่า AVC set
(Attribute Value , Class label) จากตัวอย่างข้างต้นจะเห็นว่ามี 2 AVC sets
เพื่อใช้ในการจัดกลุ่มลูกค้า



เทคนิคต่าง ๆ ของ Data Mining

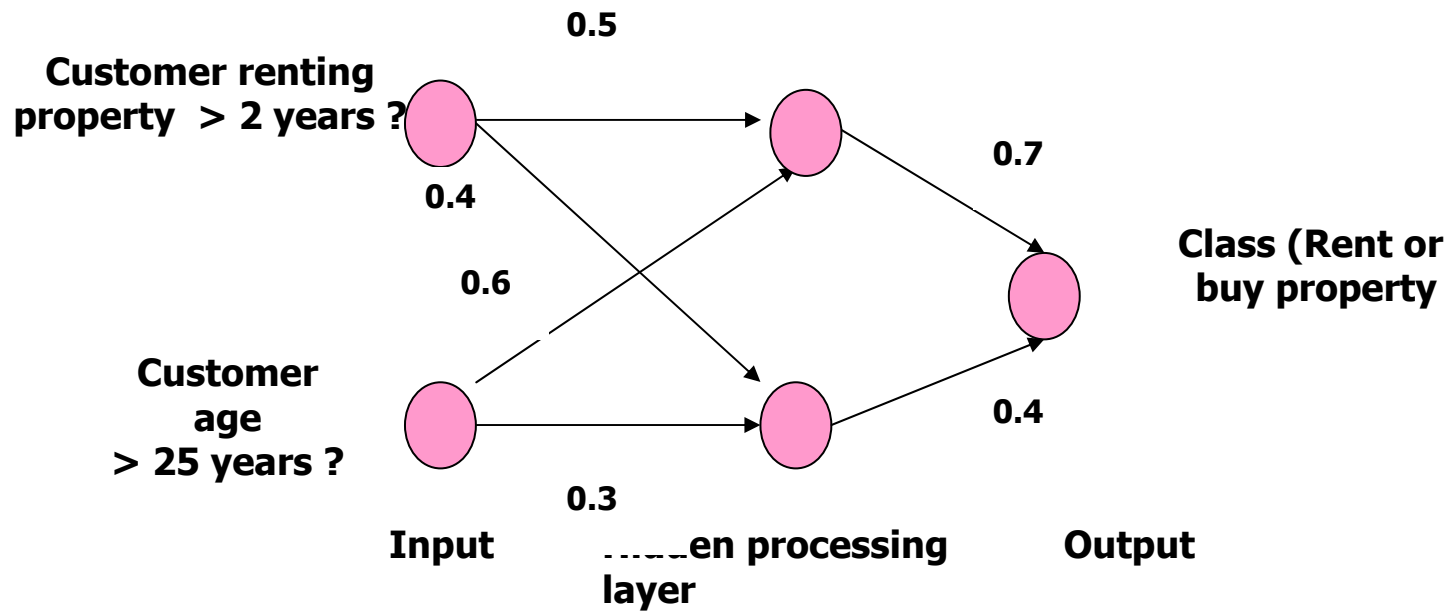
2.2 นิวรอลเน็ต หรือ นิวรอลเน็ตเวิร์ก (Neural Net)

เป็นเทคโนโลยีที่มีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ Artificial Intelligence : AI เพื่อใช้ในการคำนวณค่าฟังก์ชันจากกลุ่มข้อมูล วิธีการของ นิวรอลเน็ต (แท้จริงต้องเรียกให้เต็มว่า Artificial Neural Networks หรือ (ANN) เป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบ แล้วฝึก (train) ให้ระบบได้รู้จักที่จะคิดแก้ปัญหาที่กว้างขึ้นได้ ในโครงสร้างของนิวรอลเน็ตจะประกอบด้วยโหนด (node) สำหรับ Input – Output และการประมวลผลกระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ input layer , output layer และ hidden layers การประมวลผลของนิวรอลเน็ตจะอาศัยการส่งการทำงานผ่านโหนดต่าง ๆ ใน layer เหล่านี้



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT



นิเวศน์เน็ตเพื่อวิเคราะห์การเช่าและซื้อบ้านของลูกค้า



เทคนิคต่าง ๆ ของ Data Mining

3. Database clustering หรือ Segmentation

เป็นเทคนิคการลดขนาดของข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกัน ตัวอย่างเช่น บริษัทจำหน่ายรถยนต์ได้แยกกลุ่มลูกค้าออกเป็น 3 กลุ่ม คือ

กลุ่มผู้มีรายได้สูง ($> \$80,000$)

กลุ่มผู้มีรายได้ปานกลาง ($\$25,000$ to $\$80,000$)

กลุ่มผู้มีรายได้ต่ำ (less than $\$25,000$)



เทคนิคต่าง ๆ ของ Data Mining

และภายในแต่ละกลุ่มยังแยกออกเป็น

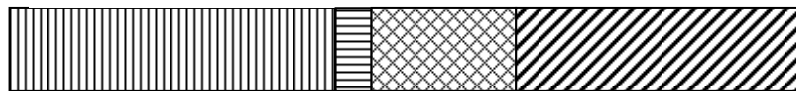
- Have Children
- Married
- Last car is a used car
- Own cars



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

First Segment (High income)



Have Children

Second Segment (Middle income)



Married



Last car is a
used car

Third Segment (Low income)



Own cars

ตัวอย่าง Clustering



เทคนิคต่าง ๆ ของ Data Mining

จากข้อมูลข้างต้นทำให้ทางบริษัทรู้ว่าเมื่อมีลูกค้าเข้ามาที่บริษัทควร
จะเสนอขายรถประเภทใด เช่น ถ้าเป็นกลุ่มผู้มีรายได้สูงควรจะเสนอรถ
ใหม่ เป็นรถครอบครัวขนาดใหญ่พอสมควร แต่ถ้าเป็นผู้มีรายได้
ค่อนข้างต่ำควรเสนอรถมือสอง ขนาดค่อนข้างเล็ก



เทคนิคต่าง ๆ ของ Data Mining

4. Deviation Detection

เป็นกรรมวิธีในการหาค่าที่แตกต่างไปจากค่ามาตรฐาน หรือค่าที่คาดคิดไว้ว่าต่างไปเล็กน้อยเพียงใด โดยทั่วไปมักใช้วิธีการทางสถิติ หรือการแสดงให้เห็นภาพ (Visualization) สำหรับเทคนิคนี้ใช้ในการตรวจสอบ ลายเซ็นปลอม หรือบัตรเครดิตปลอม รวมทั้งการตรวจหาจุดบกพร่องของชิ้นงานในโรงงานอุตสาหกรรม



เทคนิคต่าง ๆ ของ Data Mining

5. Link Analysis

จุดมุ่งหมายของ Link Analysis คือ การสร้าง link ที่เรียกว่า “associations” ระหว่าง recode เดียว หรือ กลุ่มของ recode ในฐานข้อมูล link analysis สามารถแบ่งออกเป็น 3 ชนิด คือ

- associations discovery
- sequential pattern discovery
- similar time sequence discovery



Data Mining และ Data Warehouse

สิ่งสำคัญที่จะต้องทำในการทำ Data Mining ก็คือ การกำหนดข้อมูลที่เหมาะสมในการ mining ดังนั้น Data mining จึงต้องการแหล่งข้อมูลที่มีการจัดเก็บและรวบรวมข้อมูลไว้อย่างดีและมีความมั่นคง

เหตุผลที่ต้องมี Data warehouse ที่มีการจัดเก็บข้อมูลที่ดีสำหรับเตรียมข้อมูลเพื่อทำการ mining ก็คือ



Data Mining และ Data Warehouse

- Data warehouse จะทำการจัดเก็บข้อมูลที่มีความมั่นคงและข้อมูลที่ได้ทำความสะอาดแล้ว ซึ่งการจัดเตรียมและรวบรวมข้อมูลเป็นสิ่งที่จำเป็นสำหรับการ mining ที่ต้องการความแน่ใจในความแม่นยำของ predictive models
- Data warehouse จะเป็นประโยชน์สำหรับการ mining ข้อมูลจากแหล่งข้อมูลหลายๆแหล่งที่ค้นพบมากมายเท่าที่จะเป็นไปได้ ซึ่ง Data warehouse จะบรรจุข้อมูลจากแหล่งข้อมูลเหล่านั้น



Data Mining และ Data Warehouse

- ในการเลือกส่วนย่อยๆของ record และ fields ที่ตรงประเด็น Data mining จะต้องการความสามารถในการ query ข้อมูลของ Data warehouse
- การศึกษาผลที่ได้จากการทำ Data mining จะเป็นประโยชน์อย่างมาก ถ้าหากมีการสืบค้นข้อมูลอย่างมีแบบแผนต่อไปในอนาคต ซึ่ง Data warehouse จะเป็นแหล่งจัดเก็บข้อมูลภายหลังไว้ให้



การประยุกต์ใช้งาน Data Mining

- ธุรกิจค้าปลีกสามารถใช้งาน Data Mining ในการพิจารณาหากกลยุทธ์ให้เป็นที่สนใจกับผู้บริโภคในรูปแบบต่าง ๆ เช่น ที่วางในชั้นวางของจะจัดการอย่างไรถึงจะเพิ่มยอดขายได้ เช่นที่ Midas ซึ่งเป็นผู้แทนจำหน่ายอะไหล่สำหรับอุตสาหกรรมรถยนต์ งานที่ต้องทำคือการจัดการกับข้อมูลที่ได้รับจากสาขาทั้งหมด ซึ่งจะต้องทำการรวบรวมและวิเคราะห์อย่างทั่วถึง



การประยุกต์ใช้งาน Data Mining

- กิจการโทรคมนาคม เช่นที่ Bouygues Telecom ได้นำมาใช้ตรวจสอบการโกงโดยวิเคราะห์รูปแบบการใช้งานของสมาชิกลูกค้าในการใช้งานโทรศัพท์ เช่น คาบเวลาที่ใช้จุดหมายปลายทาง ความถี่ที่ใช้ ฯลฯ และคาดการณ์ข้อบกพร่องที่เป็นไปได้ในการชำระเงิน เทคนิคนี้ยังได้ถูกนำมาใช้กับลูกค้าโทรศัพท์เคลื่อนที่ซึ่งระบบสามารถตรวจสอบได้ว่าที่ใดที่เสี่ยงที่จะสูญเสียลูกค้าสูงในการแข่งขัน France Telecom ได้ค้นหาวิธีรวมกลุ่มผู้ใช้ให้เป็นหนึ่งเดียวด้วยการสร้างแรงดึงดูดในเรื่องค่าใช้จ่ายและพัฒนาเรื่องความรักภักดีต่อตัวสินค้า



การประยุกต์ใช้งาน Data Mining

- การวิเคราะห์ผลิตภัณฑ์ เก็บรวบรวมลักษณะและราคาของผลิตภัณฑ์ทั้งหมดสร้างโมเดลด้วยเทคนิค Data Mining และใช้โมเดลในการทำนายราคาผลิตภัณฑ์ตัวอื่น ๆ
- การวิเคราะห์บัตรเครดิต
 - ช่วยบริษัทเครดิตการ์ดตัดสินใจในการที่จะให้เครดิตการ์ดกับลูกค้าหรือไม่
 - แบ่งประเภทของลูกค้าว่ามีความเสี่ยงในเรื่องเครดิต ต่ำ ปานกลาง หรือสูง
 - ป้องกันปัญหาเรื่องการทุจริตบัตรเครดิต



การประยุกต์ใช้งาน Data Mining

- การวิเคราะห์ลูกค้า
 - ช่วยแบ่งกลุ่มและวิเคราะห์ลูกค้าเพื่อที่จะผลิตและเสนอสินค้าได้ตรงตามกลุ่มเป้าหมายแต่ละกลุ่ม
 - ทำนายว่าลูกค้าคนใดจะเลิกใช้บริการจากบริษัทภายใน 6 เดือนหน้า



การประยุกต์ใช้งาน Data Mining

- การวิเคราะห์การขาย
 - พบว่า 70 % ของลูกค้าที่ซื้อโทรทัศน์แล้วจะซื้อวิดีโอตามมา ดังนั้นผู้จัดการจึงควรมุ่งไปลูกค้าที่ซื้อโทรทัศน์ แล้วจึงส่งเมลไปยังลูกค้าเหล่านั้นเพื่อที่จะเชิญชวน หรือให้ข้อเสนอที่ดีเพื่อให้ลูกค้ามาซื้อวิดีโอในครั้งต่อไป
 - ช่วยในการ โฆษณาสินค้าได้อย่างเหมาะสมและตรงตามเป้าหมาย
 - ช่วยในการจัดวางสินค้าได้อย่างเหมาะสม



การประยุกต์ใช้งาน Data Mining

- Text Mining เป็นการปรับใช้ Data Mining มาอยู่ในรูปของข้อมูลตัวอักษร ซึ่งเป็นรูปแบบของภาษาเครื่อง SDP Infoware ตัวอย่างของงานคือใช้เป็นเครื่องมือตรวจสอบระดับความพึงพอใจของผู้ที่เข้าชมนิทรรศการ โดยผ่านการประมวลผลจากแบบสอบถาม
- e-Commerce
ช่วยให้เข้าใจพฤติกรรมของลูกค้า เช่น ลูกค้ามักเข้าไปที่ web ใดตามลำดับก่อนหลัง
ช่วยในการปรับปรุง web site เช่น พิจารณาว่าส่วนใดของ web ที่ควรปรับปรุง หรือควรเรียงลำดับการเชื่อมโยงในแต่ละหน้าอย่างไรเพื่อให้สะดวกกับผู้เข้าชม



ตัวอย่างการนำเหมืองข้อมูลไปใช้งาน

- การตลาด
 - การทำนายผลการตอบสนองกับการเปิดตัวสินค้าใหม่
 - การทำนายยอดขายเมื่อมีการลดราคาสินค้า
 - การทำนายกลุ่มลูกค้าที่น่าจะใช้สินค้าของเรา
- การเงินการธนาคาร
 - การคาดการณ์ถึงโอกาสในการชำระหนี้ของลูกค้าว่าสูงเท่าไร?
 - ค้นหาลูกค้าขาดคุณภาพ เพื่อหลีกเลี่ยงความเสี่ยงในการปล่อยกู้
 - ค้นหาลูกค้าชั้นดี เพื่อเสนอการปล่อยกู้
 - ทำนายแนวโน้มของพฤติกรรมการใช้บัตรเครดิต
- สถานีโทรทัศน์หรือวิทยุ
 - ค้นหารายการที่ดีและเหมาะสมต่อช่วงเวลาที่สุด เพื่อวางแผนรายการในแต่ละเดือน



ตัวอย่างการนำเหมืองข้อมูลไปใช้งาน (2/4)

- การค้าขาย
 - ทำนายผลกำไรเมื่อลงทุนซื้อสินค้ามาเพื่อขาย
 - ใช้ค้นหาจุดคุ้มทุน
- โรงงานและการผลิต
 - ทำนายอายุการใช้งานเครื่องจักร ตลอดจนเวลาที่เหมาะสมในการเปิดปิดเครื่องจักร
 - ค้นหาปัจจัยสำคัญที่มีผลต่อสถานะการควบคุมการผลิตที่ดีที่สุด
 - คาดการณ์ วันและเวลาที่จำเป็นต่อการซ่อมบำรุงเครื่องจักร
- โรงพยาบาล
 - ทำนายอาการของโรคได้อย่างแม่นยำขึ้น เพื่อกำหนดวิธีการรักษาได้อย่างเหมาะสมที่สุด



ตัวอย่างการนำเหมืองข้อมูลไปใช้งาน (3/4)

- ตลาดหลักทรัพย์
 - ค้นหาปัจจัยที่มีผลต่อการเปลี่ยนแปลงราคาหุ้น
 - ทำนายระดับการแปรผันของหุ้นแต่ละตัว หรือภาพโดยรวมของตลาด
 - ทำนายเวลาที่เหมาะสมต่อการปล่อยหุ้นที่ให้ได้ผลกำไรมากที่สุด
 - ทำนายอนาคตว่าหุ้นตัวใดจะสร้างผลกำไรได้มาก และคุ้มค่าต่อการซื้อ
- ธุรกิจประกันภัย
 - ค้นหาผู้เอาประกันที่มีแนวโน้มที่จะซื้อประกันเพิ่ม
 - แบ่งรูปแบบการเรียกเงินชดเชยจากการประกันภัย เพื่อจัดอันดับความเสี่ยงภัย



ตัวอย่างการนำเหมืองข้อมูลไปใช้งาน (3/4)

- ฮาร์ดแวร์และซอฟต์แวร์คอมพิวเตอร์
 - ค้นหาช่วงเวลาที่เหมาะสำหรับการผลิตชิพคอมพิวเตอร์ตัวใหม่เพื่อป้อนสู่ตลาด
 - การทำนายอายุการใช้งานของ Disk Drive หรือ อุปกรณ์ต่าง ๆ
 - การประเมินศักยภาพของซอฟต์แวร์ เมื่อเทียบกับซอฟต์แวร์ของคู่แข่ง
- กระบวนการใหม่
 - กำหนดงบประมาณที่ใช้ในการเคลื่อนย้ายอาวุธยุทธโปกรณ์ทางทหาร
 - ใช้ทดสอบศักยภาพของกลยุทธ์ทางการทหาร
 - ทำนายทรัพยากรที่ใช้ในสงครามแต่ละครั้ง



การแสดงผลข้อมูลใน ระบบสนับสนุนการตัดสินใจ

- การแสดงผลข้อมูลแบบกราฟิก
- การแสดงผลข้อมูลแบบหลายมิติ

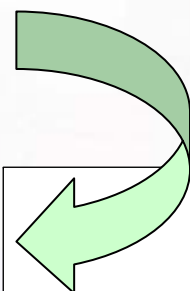


Data Warehousing and Applications

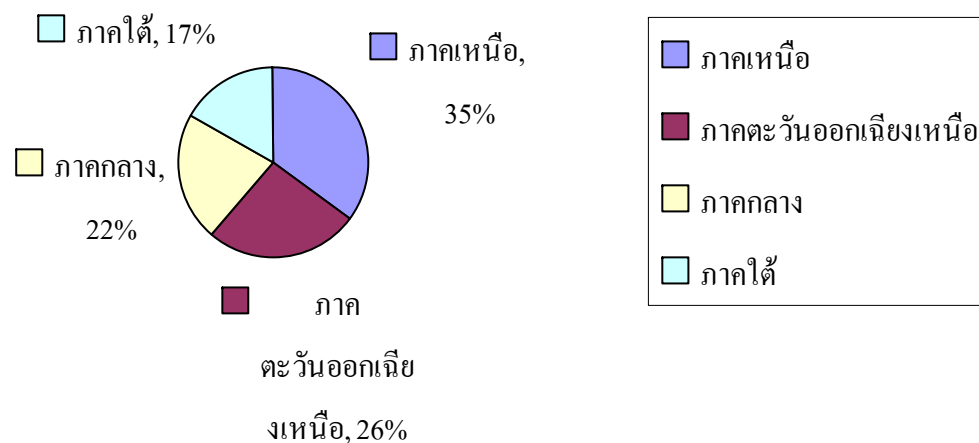
Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบกราฟิก

ฝ่ายขาย	
สรุปยอดขายไตรมาสที่ 1 (หน่วย: พันบาท)	
ตัวแทน	ยอดการขาย
ภาคเหนือ	35%
ภาคตะวันออกเฉียงเหนือ	26%
ภาคกลาง	22%
ภาคใต้	17%



สรุปยอดขายภูมิภาคไตรมาสที่ 1





Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบกราฟิก

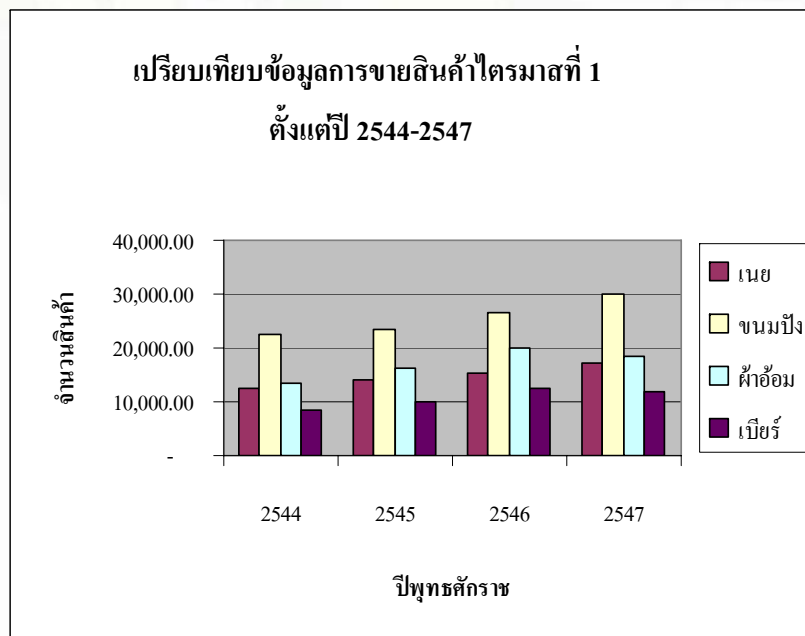
ฝ่ายขาย					
สรุปยอดขายไตรมาสที่ 1 (หน่วย: พันบาท) ของแต่ละปี					
ปี	เนย	ขนมปัง	ผ้าอ้อม	เปียร์	ยอดรวม
2544	12,500	22,500	13,500	8,500	59,544
2545	14,000	23,400	16,250	10,000	66,195
2546	15,250	26,500	20,000	12,500	76,796
2547	17,250	30,000	18,400	11,750	79,947
รวม	59,000	102,400	68,150	42,750	282,482



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบกราฟิก





Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

ประเทศ	สินค้า	ยอดขายปีนี้	ประมาณการณ้ยอดขายปีหน้า
ไทย	วิทยุ	35,000,000.00	42,000,000.00
สิงคโปร์	วิทยุ	72,000,000.00	86,400,000.00
ฟิลิปปินส์	วิทยุ	23,000,000.00	27,600,000.00
มาเลเซีย	วิทยุ	48,000,000.00	57,600,000.00
ไทย	โทรทัศน์	68,000,000.00	81,600,000.00
สิงคโปร์	โทรทัศน์	52,000,000.00	62,400,000.00
ฟิลิปปินส์	โทรทัศน์	45,000,000.00	54,000,000.00



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

ประเทศ	สินค้า	ยอดขายปีนี้	ประมาณการณียอดขายปีหน้า
มาเลเซีย	โทรทัศน์	33,000,000.00	39,600,000.00
ไทย	ตู้เย็น	43,000,000.00	51,600,000.00
สิงคโปร์	ตู้เย็น	51,000,000.00	61,200,000.00
ฟิลิปปินส์	ตู้เย็น	47,000,000.00	56,400,000.00
มาเลเซีย	ตู้เย็น	36,000,000.00	43,200,000.00



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

ประเทศ	สินค้า	ยอดขายปีนี้	ประมาณการณ์ยอดขายปีหน้า
ไทย	ตู้เย็น	43,000,000.00	51,600,000.00
	โทรทัศน์	68,000,000.00	81,600,000.00
	วิทยุ	35,000,000.00	42,000,000.00
ฟิลิปปินส์	ตู้เย็น	47,000,000.00	56,400,000.00
	โทรทัศน์	45,000,000.00	54,000,000.00
	วิทยุ	23,000,000.00	27,600,000.00
มาเลเซีย	ตู้เย็น	36,000,000.00	43,200,000.00
	โทรทัศน์	33,000,000.00	39,600,000.00
	วิทยุ	48,000,000.00	57,600,000.00
สิงคโปร์	ตู้เย็น	51,000,000.00	61,200,000.00
	โทรทัศน์	52,000,000.00	62,400,000.00
	วิทยุ	72,000,000.00	86,400,000.00
รวม		553,000,000.00	663,600,000.00



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

ประเทศ	ข้อมูล	สินค้า			
		ตู้เย็น	โทรทัศน์	วิทยุ	ผลรวมทั้งหมด
ไทย	ยอดขายปีนี้	43,000,000.00	68,000,000.00	35,000,000.00	146,000,000.00
	ประมาณการณียอดขายปีหน้า	51,600,000.00	81,600,000.00	42,000,000.00	175,200,000.00
ฟิลิปปินส์	ยอดขายปีนี้	47,000,000.00	45,000,000.00	23,000,000.00	115,000,000.00
	ประมาณการณียอดขายปีหน้า	56,400,000.00	54,000,000.00	27,600,000.00	138,000,000.00



Data Warehousing and Applications

Lect. Rattana Wadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

ประเทศ	ข้อมูล	สินค้า			
		ตู้เย็น	โทรทัศน์	วิทยุ	ผลรวมทั้งหมด
มาเลเซีย	ยอดขายปีนี้	36,000,000.00	33,000,000.00	48,000,000.00	117,000,000.00
	ประมาณการณ์ยอดขายปีหน้า	43,200,000.00	39,600,000.00	57,600,000.00	140,400,000.00
สิงคโปร์	ยอดขายปีนี้	51,000,000.00	52,000,000.00	72,000,000.00	175,000,000.00
	ประมาณการณ์ยอดขายปีหน้า	61,200,000.00	62,400,000.00	86,400,000.00	210,000,000.00



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

		สินค้า			
ประเทศ	ข้อมูล	ตู้เย็น	โทรทัศน์	วิทยุ	ผลรวมทั้งหมด
รวม	ยอดขายปีนี้	177,000,000.00	198,000,000.00	178,000,000.00	553,000,000.00
	ประมาณการณ์ยอดขายปีหน้า	212,400,000.00	237,600,000.00	213,600,000.00	663,000,000.00



Data Warehousing and Applications

Lect.Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

สินค้า	ข้อมูล	ประเทศ				
		ไทย	ฟิลิปปินส์	มาเลเซีย	สิงคโปร์	ผลรวมทั้งหมด
ตู้เย็น	ยอดขายปีนี้	43,000,000.00	47,000,000.00	36,000,000.00	51,000,000.00	177,000,000.00
	ประมาณการณ์ยอดขายปีหน้า	51,600,000.00	56,400,000.00	43,200,000.00	61,200,000.00	212,400,000.00
โทรทัศน์	ยอดขายปีนี้	68,000,000.00	45,000,000.00	33,000,000.00	52,000,000.00	198,000,000.00
	ประมาณการณ์ยอดขายปีหน้า	81,600,000.00	54,000,000.00	39,600,000.00	62,400,000.00	237,600,000.00



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

การแสดงผลข้อมูลแบบหลายมิติ

		ประเทศ				
สินค้า	ข้อมูล	ไทย	ฟิลิปปินส์	มาเลเซีย	สิงคโปร์	ผลรวมทั้งหมด
วิทยุ	ยอดขายปีนี้	35,000,000.00	23,000,000.00	48,000,000.00	72,000,000.00	178,000,000.00
	ประมาณการณียอดขายปีหน้า	42,000,000.00	27,600,000.00	57,600,000.00	86,400,000.00	213,600,000.00
ผลรวม	ยอดขายปีนี้	146,000,000.00	115,000,000.00	117,000,000.00	175,000,000.00	553,000,000.00
ผลรวม	ประมาณการณียอดขายปีหน้า	175,200,000.00	138,000,000.00	140,400,000.00	210,000,000.00	663,600,000.00



Data Warehousing and Applications

Lect. Rattanawadee Panthong
School of ICT

Reference

กิติ ภัทธีวัฒน์กุล, คัมภีร์ระบบสนับสนุนการตัดสินใจ และระบบผู้เชี่ยวชาญ. กรุงเทพฯ: เคทีพี คอมพ์ แอนด์ คอนซัลท์, 2546