

# รายงานเรื่อง

## การทำเหมืองข้อมูลด้วย Rapidminer Studio 6

### เหมืองข้อมูล (Data mining)

เป็นกระบวนการ (Process) ที่กระทำกับข้อมูลขนาดใหญ่เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้นโดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์เพื่อให้ได้สารสนเทศที่เราไม่รู้ออกมา โดยสารสนเทศที่ได้จะมีเหตุผลและสามารถนำไปใช้ประโยชน์ได้

### เทคนิคต่าง ๆ ของ Data Mining

#### 1. Association rule Discovery

**หลักการทํางาน** คือ การค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่ เพื่อนำไปวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ เช่น การวิเคราะห์การซื้อสินค้าของลูกค้าเรียกว่า “Market Basket Analysis”

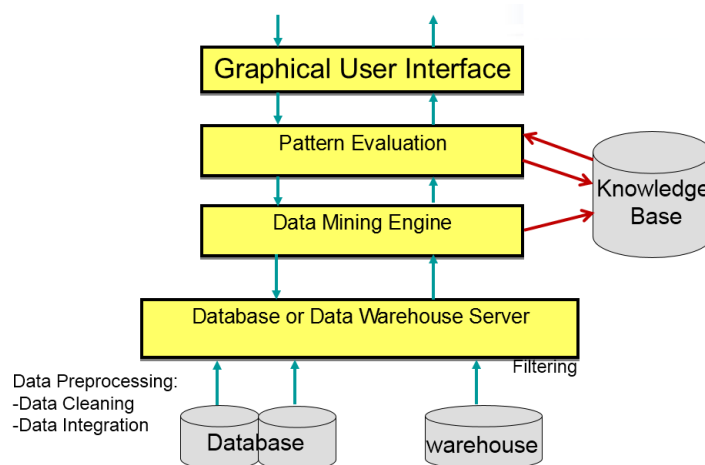
#### 2. Classification & Prediction

เป็นการแบ่งประเภทของข้อมูล โดยจะหากฎเพื่อระบุประเภทวัตถุจากคุณสมบัติของวัตถุ

#### 3. Database Clustering Or Segmentation

เป็นเทคนิคการลดขนาดของข้อมูล ด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกัน เพื่อนำข้อมูลที่ได้ไปวิเคราะห์

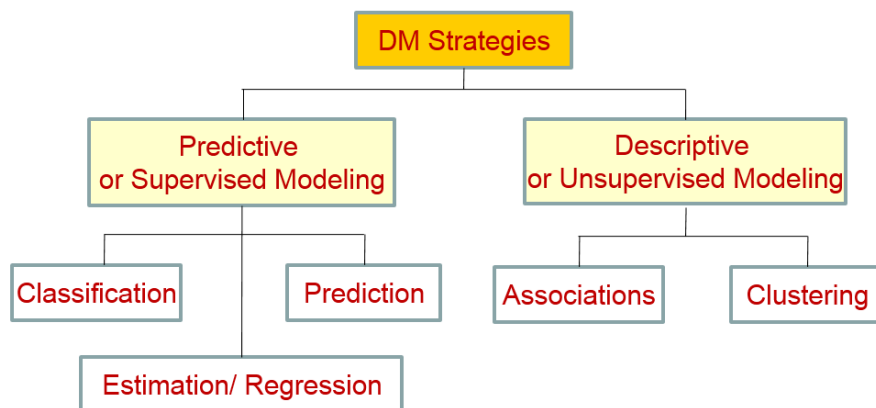
### ส่วนประกอบที่สำคัญของระบบการทำเหมืองข้อมูล



- Database & Data Warehouse เป็นแหล่งที่เก็บข้อมูลอื่นๆ เช่น แฟ้มข้อมูล สเปรดชีท ซึ่งเป็นแหล่งที่เก็บข้อมูลดิบสำหรับการทำเหมืองข้อมูล

- Database/Data Warehouse server ทำหน้าที่นำเข้าสู่ข้อมูลตามคำขอของผู้ใช้
- Knowledge base ได้แก่
  - ความรู้ในงานที่ทำจะเป็นประโยชน์ต่อการชี้แนะทางการสืบค้น หรือการประเมินความน่าสนใจของรูปแบบผลลัพธ์ที่ได้ โดยเป็นความรู้เฉพาะด้านในงานที่ทำ เช่น ความรู้เฉพาะทางการแพทย์ หรือ ดาราศาสตร์ เป็นต้น
- Data Mining Engine เป็นส่วนประกอบหลัก ประกอบด้วยโมดูล (Modules) ซึ่งรับผิดชอบงานทำเหมืองประเภทต่างๆ ได้แก่ การหาความสัมพันธ์ การจำแนกประเภท การจัดกลุ่ม เป็นต้น
- Pattern Evaluation Module ทำงานร่วมกับ Data Mining Engine โดยใช้ค่าขีดแบ่งมาตรฐานวัดความน่าสนใจในการกลั่นกรองรูปแบบผลลัพธ์ที่ได้ ซึ่งใช้ประเมินความน่าสนใจของรูปแบบการทำเหมืองที่ได้
- Graphical User Interface ส่วนติดต่อประสานระหว่างผู้ใช้กับระบบทำเหมือง ซึ่งช่วยให้ผู้ใช้สามารถระบุงานทำเหมืองที่ต้องการ

### กลยุทธ์ในการทำเหมืองข้อมูล



1. แบบจำลองในการทำนาย (Predictive/ Supervised Modeling) เป็นผลลัพธ์ที่สร้างจากการอนุมาน (Inference) ชุดข้อมูลปัจจุบัน เพื่อใช้ในการทำนายประเภทตัวอย่างในอนาคต
2. แบบจำลองในการบรรยาย (Descriptive/ Unsupervised Modeling) ในที่นี้ อาจเป็นการหาความสัมพันธ์ต่างๆ (Association) หรือหาการจัดกลุ่มข้อมูล (Clustering) ซึ่งไม่ได้มีจุดมุ่งหมายเพื่อการทำนาย