

HW Chapter 1

Data Mining คือ การวิเคราะห์ข้อมูลจากข้อมูลจำนวนมาก (big data) เพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ โดยทำการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และหาความน่าจะเป็นที่จะเกิดขึ้น เพื่อให้ได้องค์ความรู้ใหม่ ที่สามารถนำไปใช้ประกอบการตัดสินใจในด้านต่างๆ เช่น ตลาดหลักทรัพย์,ทางธุรกิจ, ทางด้านการแพทย์, ยุทธศาสตร์ทหาร

การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้งานถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล เช่นดังนี้

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง
- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย
- การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
- การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

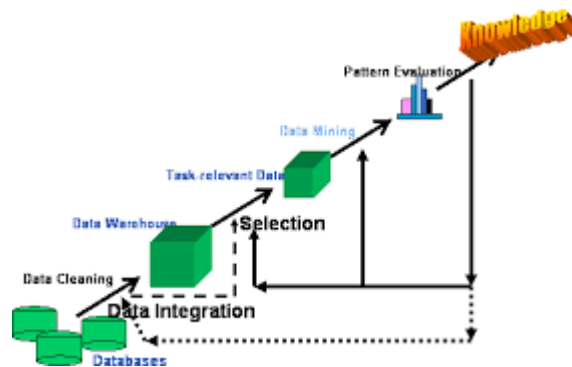
การใช้ Data Mining

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

ขั้นตอนการทำ Data Mining

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ

กระบวนการ Mining (A KDD Process ของ Data)



เป็นกระบวนการในการค้นหาลักษณะแฝงของข้อมูล (Pattern) ที่ซ่อนอยู่ในฐานข้อมูล