

## สรุปบทที่1

ชื่อ นางสาวชนัญชิตา เมธีกุลมานิต รหัสนักศึกษา 623020516-1

### Introduction

#### Data Mining คืออะไร ?

คือ การทำเหมืองข้อมูล

ตัวอย่างเช่น เซ็นเซอร์วัดอุณหภูมิ เซ็นเซอร์น้ำฝน จำนวนคนมาเปิดดูเว็บไซต์(เพื่อดูว่าคนมาดูเว็บไซต์ของเรา ดูอะไรบ้างมากที่สุด ทำให้สามารถทำนายได้ว่ามีกลุ่มลูกค้าแบบไหน จะได้ใช้โฆษณาที่ถูกเป้าหมาย)

-ทำให้การได้ข้อมูลมาจำนวนมากจัดอยู่ในกลุ่มข้อมูลที่เป็นหมวดหมู่มากยิ่งขึ้น

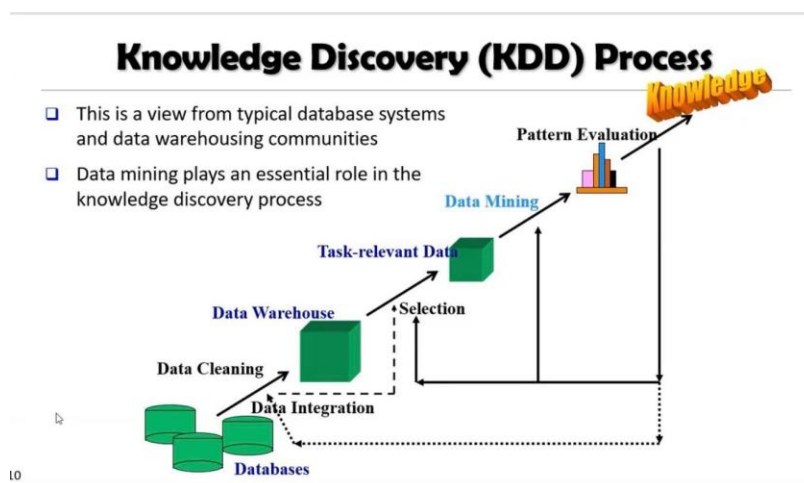
-เป็นการเก็บข้อมูลให้มีประสิทธิภาพ

Knowledge discovery mining in databases KDD (การค้นพบองค์ความรู้จากดาต้าเบสที่มีอยู่)

เราจะเก็บข้อมูลแบบไหนให้มีประสิทธิภาพ !!!

เราจะมีการจัดข้อมูล และนำเอาข้อมูลที่ได้มาจากหลายๆแหล่งนำมารวมกันเพื่อนำไปเก็บใน Data warehouse หลังจากนั้นเอาดาต้าที่สรุปได้ เพื่อนำไปทำ Data mining ใช้เป็นรูปแบบในการสร้างข้อมูล มาวัดผลว่าเป็นองค์ความรู้ที่แท้จริง และนำองค์ความรู้ที่ได้มาไปตรวจสอบ เพื่อนำมาสรุปผล

Data warehouse คือมีข้อมูลอยู่แต่ยังไม่ได้สกัดออกมาเป็นความรู้ แต่ Data mining คือข้อมูลที่ได้รับการสกัดอย่างมีประสิทธิภาพแล้ว



ชื่อ นางสาวชนัญชิตา เมธิกุลมานิต รหัสนักศึกษา 623020516-1

## Data Mining ที่จะเรียนหลักๆ

-Pattern discovery การค้นพบรูปแบบ

-Classification การจำแนกข้อมูล

-Clustering การแบ่งกลุ่มข้อมูล

-Outlier analysis การวิเคราะห์ภายนอก

Column: Attributes , Fields , Features : คำที่ใช้อธิบายคุณสมบัติของข้อมูล

Rows : Record , data point : ข้อมูลแต่ละตัว

## เทคนิคที่จะเรียน

### Association rule

Diaper Beer ใช้เทคนิควิเคราะห์ใบเสร็จของคนที่มาซื้อแป้ง จากข้อมูลที่เขาเก็บมาพบว่าคนที่มาซื้อผ้าอ้อมก็จะซื้อเบียร์ไปด้วย ข้อมูลที่ได้มาก็วิเคราะห์ได้ว่าคนที่มาซื้อแป้งมักจะเป็นคุณพ่อ

คนที่เป็นเจ้าของก็สามารถนำข้อมูลที่ได้มาไปจัดการร้านค้าได้ เช่น จัดร้านให้เบียร์กับผ้าอ้อมอยู่คนละโซนกัน เพื่อที่จะให้เดินผ่านสินค้าหลายๆอย่างได้ หรือตั้งเบียร์กับผ้าอ้อมไว้ใกล้กันเพื่ออำนวยความสะดวกให้กับลูกค้า ใช้ในการเพิ่มยอดขายให้กับเบียร์

### Classification and label prediction

การจำแนกกลุ่ม เช่นเรามีข้อมูลของกรมอุตุนิยมวิทยา มีฝนตกไหม อุณหภูมิเป็นยังไง ก้อนเมฆเยอะไหม เราสามารถนำข้อมูลเหล่านี้มาใช้ทำนายได้ ในวันถัดไปจากข้อมูลที่ได้เก็บได้

### Cluster Analysis

การแบ่งกลุ่มนำข้อมูลที่คล้ายๆกันมาอยู่ด้วยกัน