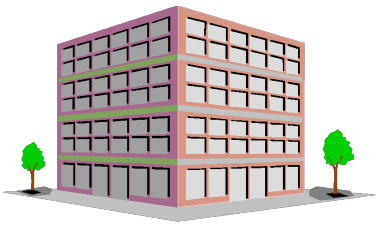
# บทที่ 2 การวิเคราะห์ข้อมูล

วัตถุประสงค์

- เพื่อให้ผู้อ่านเข้าใจความหมายและวัตถุประสงค์ของการวิเคราะห์ข้อมูล
- เพื่อให้ผู้อ่านเข้าใจกระบวนการ รูปแบบและเทคนิคต่างๆของการ วิเคราะห์ข้อมูล และการนำไปประยุกต์ใช้ทางธุรกิจ
- เพื่อให้ผู้อ่านทราบถึงความท้าทายและแนวทางต่างๆ ในการเพิ่มประสิทธิภาพของการวิเคราะห์ข้อมูลที่ซับซ้อน



## 2.1 บทนำ

…

## 2.2 การวิเคราะห์ข้อมูล: หัวใจสำคัญของการวิเคราะห์ธุรกิจ

การวิเคราะห์ข้อมูล (data analytics) เป็นกระบวนการดึงข้อมูลเชิงลึก (insight) ที่มีความสำคัญออกมาจากข้อมูล ถือเป็นการปฏิวัติวิธีการทำงานและการตัดสินใจขององค์กร ช่วยให้บริษัทสามารถเปลี่ยนข้อมูลดิบ ซึ่งอาจจะยุ่งยากและไร้ความหมายในตัวเอง ให้กลายเป็นความรู้ที่สามารถนำไปใช้ประโยชน์ได้ การวิเคราะห์ข้อมูลช่วยให้บริษัทต่าง ๆ สามารถก้าวข้ามจากการเข้าใจข้อมูลเพียงอย่างเดียวไปสู่การตัดสินใจด้วยข้อมูลและมีการวางกลยุทธ์ ซึ่งจะช่วยเพิ่มความสามารถในการแข่งขันและประสิทธิภาพโดยรวมของบริษัท

ข้อมูลเชิงลึก (insight) ที่เป็นความรู้ใหม่นี้สามารถนำไปใช้เพื่อ:

- ลดต้นทุน **(cost reduction)**
    - **ลดการใช้ทรัพยากร:** การวิเคราะห์ข้อมูลช่วยให้บริษัทสามารถระบุและลดขั้นตอนที่ไม่จำเป็นในกระบวนการผลิต ทำให้สามารถลดการใช้ทรัพยากรเช่น วัตถุดิบ หรือพลังงาน ตัวอย่างเช่น บริษัทผลิตเครื่องใช้ไฟฟ้าอาจใช้การวิเคราะห์ข้อมูลเพื่อหาแนวทางในการลดการใช้ไฟฟ้าในกระบวนการผลิต
    - **ลดเวลา:** การวิเคราะห์ข้อมูลช่วยให้สามารถวางแผนการผลิตและจัดการกระบวนการทำงานได้อย่างมีประสิทธิภาพ ทำให้ลดเวลาในการผลิต ตัวอย่างเช่น บริษัทขนส่งใช้การวิเคราะห์ข้อมูล

เพื่อหาทางลัดหรือเส้นทางที่เร็วที่สุดในการส่งสินค้า ทำให้ประหยัดเวลาและค่าใช้จ่ายในการขนส่ง

- **สร้างมูลค่าเพิ่ม (value adding)**
    - **เพิ่มรายได้และกำไร:** การวิเคราะห์ข้อมูลช่วยให้บริษัทสามารถเข้าใจความต้องการของลูกค้าได้ดีขึ้น ทำให้สามารถพัฒนาสินค้าหรือบริการที่ตรงตามความต้องการของลูกค้า ทำให้สามารถเพิ่มรายได้และกำไร ตัวอย่างเช่น บริษัทอีคอมเมิร์ซใช้การวิเคราะห์ข้อมูลลูกค้าเพื่อแนะนำสินค้าที่ตรงใจลูกค้า ทำให้ยอดขายเพิ่มขึ้น
    - **สร้างคุณค่าให้สังคมและสิ่งแวดล้อม:** การวิเคราะห์ข้อมูลช่วยให้บริษัทสามารถพัฒนาผลิตภัณฑ์หรือบริการที่มีประโยชน์ต่อสังคมและสิ่งแวดล้อม ตัวอย่างเช่น บริษัทผลิตรถยนต์ใช้การวิเคราะห์ข้อมูลเพื่อพัฒนารถยนต์ที่ประหยัดพลังงานและปล่อยมลพิษต่ำ ทำให้ช่วยลดปัญหามลพิษทางอากาศ
- **สร้างสรรค์นวัตกรรม (innovation creation)**
    - **ค้นพบปัญหาใหม่ ๆ:** การวิเคราะห์ข้อมูลช่วยให้บริษัทสามารถระบุปัญหาหรือความท้าทายใหม่ ๆ ที่อาจไม่เคยรู้มาก่อน ตัวอย่างเช่น บริษัทเทคโนโลยีใช้การวิเคราะห์ข้อมูลการใช้งานแอปพลิเคชันเพื่อหาปัญหาที่ผู้ใช้เจอ ทำให้สามารถพัฒนาแอปพลิเคชันให้ใช้งานได้ดีขึ้น
    - **สร้างผลิตภัณฑ์ใหม่:** การวิเคราะห์ข้อมูลช่วยให้บริษัทสามารถเข้าใจความต้องการของผู้บริโภคได้ดียิ่งขึ้น ทำให้สามารถพัฒนาผลิตภัณฑ์หรือบริการใหม่ที่ตอบโจทย์ผู้บริโภคได้ ตัวอย่างเช่น บริษัทโทรคมนาคมใช้การวิเคราะห์ข้อมูลการใช้งานของลูกค้าเพื่อพัฒนาบริการอินเทอร์เน็ตความเร็วสูงที่ตอบสนองความต้องการของลูกค้าได้ดียิ่งขึ้น

## 2.3 สามประเภทหลักๆ ของการวิเคราะห์ข้อมูล: เชิงพรรณนา (descriptive analytics), เชิงพยากรณ์ (predictive analytics) และเชิงแนะนำ (prescriptive analytics)

การวิเคราะห์ข้อมูลแบ่งออกเป็นสามประเภทหลักๆ คือ การวิเคราะห์เชิงพรรณนา เชิงพยากรณ์ และเชิงแนะนำ ซึ่งแต่ละประเภทมีบทบาทเฉพาะในการใช้ข้อมูลเพื่อสกัดข้อมูลที่เป็นประโยชน์และชี้นำการตัดสินใจ ดังนี้:

1. **การวิเคราะห์เชิงพรรณนา (descriptive analytics)**
   การวิเคราะห์เชิงพรรณนาคือการสรุปและตีความข้อมูลในอดีตเพื่อทำความเข้าใจสิ่งที่เกิดขึ้นแล้ว โดยให้มุมมองย้อนหลังของข้อมูลและช่วยตอบคำถาม เช่น:
    - ยอดขายในไตรมาสที่แล้วเป็นเท่าไร?
    - อัตราการเติบโตของผู้ใช้งานในช่วงหกเดือนที่ผ่านมาเป็นอย่างไร?
    - ผลิตภัณฑ์ใดได้รับความนิยมมากที่สุดในปีที่ผ่านมา?

   การวิเคราะห์เชิงพรรณนามีความสำคัญในการระบุแนวโน้ม รูปแบบ และความผิดปกติในข้อมูลในอดีต

2. **การวิเคราะห์เชิงพยากรณ์ (predictive analytics)**

   การวิเคราะห์เชิงพยากรณ์ใช้แบบจำลองทางสถิติและเทคนิคการเรียนรู้ของเครื่องเพื่อทำนายเหตุการณ์ในอนาคต โดยช่วยตอบคำถาม เช่น:

   - รายได้ในไตรมาสหน้าจะเป็นเท่าไร?
   - ลูกค้าคนไหนมีแนวโน้มที่จะยกเลิกการสมัครสมาชิก?
   - ผลิตภัณฑ์ใดจะมีความต้องการมากที่สุดในฤดูกาลหน้า?

   การวิเคราะห์เชิงพยากรณ์ใช้เพื่อคาดการณ์แนวโน้มและตัดสินใจเชิงรุกเพื่อคว้าโอกาสหรือบรรเทาความเสี่ยงในอนาคต

3. **การวิเคราะห์เชิงแนะนำ (prescriptive analytics)**

   การวิเคราะห์เชิงแนะนำเป็นการวิเคราะห์ที่ก้าวไปไกลกว่าการพยากรณ์เพื่อแนะนำการกระทำที่เฉพาะเจาะจงที่จะช่วยให้บรรลุเป้าหมายที่ต้องการ โดยช่วยตอบคำถาม เช่น:

   - ต้องทำอะไรเพื่อเพิ่มยอดขาย 10%?
   - กลยุทธ์ที่ดีที่สุดในการลดต้นทุนการผลิตคืออะไร?
   - เส้นทางที่ดีที่สุดสำหรับการจัดส่งเพื่อลดเวลาล่าช้าและต้นทุนคืออะไร?

   การวิเคราะห์เชิงแนะนำมักใช้ผลลัพธ์จากการวิเคราะห์เชิงพรรณนาและเชิงพยากรณ์เพื่อเสนอแนะที่เพิ่มประสิทธิภาพสูงสุดหรือลดต้นทุนให้น้อยที่สุด โดยคำนึงถึงข้อจำกัดและความไม่แน่นอน

การวิเคราะห์ทั้งสามประเภทนี้เชื่อมโยงกันและมักใช้ร่วมกัน การวิเคราะห์เชิงพรรณนาวางพื้นฐานโดยให้ความเข้าใจที่ชัดเจนของอดีต การวิเคราะห์เชิงพยากรณ์ใช้ข้อมูลนี้เพื่อคาดการณ์อนาคต และการวิเคราะห์เชิงแนะนำแนะนำการกระทำที่เป็นรูปธรรมตามการคาดการณ์เหล่านี้ ตัวอย่างเช่น บริษัทอาจใช้การวิเคราะห์เชิงพรรณนาเพื่อวิเคราะห์ผลการขายในอดีต ใช้การวิเคราะห์เชิงพยากรณ์เพื่อคาดการณ์ความต้องการในอนาคต และใช้การวิเคราะห์เชิงแนะนำเพื่อเพิ่มประสิทธิภาพการจัดการสินค้าคงคลังและการวางแผนทรัพยากร

สรุปได้ว่า ทั้งสามประเภทของการวิเคราะห์ข้อมูลช่วยให้บริษัทสามารถก้าวจากการทำความเข้าใจข้อมูลไปสู่การตัดสินใจที่มีความรู้และเชิงกลยุทธ์ เพิ่มความสามารถในการแข่งขันและประสิทธิภาพโดยรวม.

## 2.4 การวิเคราะห์ข้อมูลที่ซับซ้อน (complex data analytics)

การวิเคราะห์ข้อมูลเชิงซับซ้อน (Complex Data Analytics) มุ่งเน้นไปที่การวิเคราะห์ชุดข้อมูลที่ท้าทายและการผสมผสานกันของชุดข้อมูลเหล่านั้น ชุดข้อมูลเหล่านี้มักมีลักษณะที่ทำให้วิธีการวิเคราะห์แบบดั้งเดิมมีประสิทธิภาพน้อยลง

### 2.4.1 ลักษณะของข้อมูลเชิงซับซ้อน:

1. **มิติสูง (High Dimensionality):** ข้อมูลมีลักษณะหลากหลายมาก ทำให้การแสดงผลและวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทำได้ยาก
2. **ความหลากหลาย (Heterogeneity):** ข้อมูลอาจมีโครงสร้าง (เช่น ตาราง), กึ่งโครงสร้าง (เช่น ข้อความ log), หรือไม่มีโครงสร้าง (เช่น ข้อความ, ภาพ) การรวมและวิเคราะห์ข้อมูลในรูปแบบที่หลากหลายนี้ต้องใช้เทคนิคเฉพาะ
3. **ความเร็ว (Velocity):** ข้อมูลอาจถูกสร้างขึ้นแบบเรียลไทม์หรือเปลี่ยนแปลงตลอดเวลา ต้องการการวิเคราะห์แบบเรียลไทม์เพื่อการตัดสินใจที่มีประสิทธิภาพ (เช่น ข้อมูลสตรีม, ข้อมูลจากเซ็นเซอร์ของอุปกรณ์ IoT)
4. **สัญญาณรบกวนและข้อผิดพลาด (Noise and Errors):** ชุดข้อมูลเชิงซับซ้อนสามารถมีสัญญาณรบกวน ความไม่สอดคล้อง และค่าสูญหาย ซึ่งต้องการการทำความสะอาดและการเตรียมข้อมูลอย่างระมัดระวัง
5. **ประเภทข้อมูลที่ซับซ้อน:** เช่น ข้อความ, ภาพ, วิดีโอ, ชุดข้อมูลเวลา (time series)
6. **เฉพาะทาง:** เช่น ข้อมูลชีวสารสนเทศ, ข้อมูลทางการแพทย์, ข้อมูลเชิงพื้นที่ เป็นต้น

### 2.4.2 ความท้าทายของการวิเคราะห์ข้อมูลเชิงซับซ้อน:

- **เทคนิคดั้งเดิมอาจไม่เหมาะสม:** วิธีมาตรฐานเช่น สถิติพื้นฐานหรือการแสดงผลอาจไม่เพียงพอในการจับความซับซ้อนของข้อมูลเชิงซับซ้อน
- **ความสามารถในการขยายขนาด:** การประมวลผลและวิเคราะห์ชุดข้อมูลขนาดใหญ่อาจมีค่าใช้จ่ายสูง ต้องใช้การพัฒนาอัลกอริธึมเฉพาะ
- **การตีความ:** การดึงข้อมูลเชิงลึกที่ชัดเจนและตีความได้จากโมเดลเชิงซับซ้อนอาจเป็นเรื่องท้าทาย

### 2.4.3 เทคนิคสำหรับการวิเคราะห์ข้อมูลเชิงซับซ้อน:

- **การลดมิติ (Dimensionality Reduction):** เทคนิคเช่น PCA สามารถแปลงข้อมูลมิติสูงเป็นข้อมูลที่มีมิติต่ำกว่า
- **การเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก (ML and Deep Learning):** สามารถจัดการความสัมพันธ์ที่ซับซ้อนในข้อมูลและทำการทำนายบนชุดข้อมูลเชิงซับซ้อนได้
- **เทคนิคการแสดงผลขั้นสูง (Advanced Visualization Techniques):** วิธีการเช่น parallel coordinate plots สามารถแสดงผลข้อมูลมิติสูงและระบุความผิดปกติได้อย่างมีประสิทธิภาพ

## 2.5 วิทยาศาสตร์ข้อมูล (data science) และขั้นตอนต่างๆ ของวิทยาศาสตร์ข้อมูล (steps of data science)

ด้วยการทำตามขั้นตอนเหล่านี้ ธุรกิจสามารถใช้ประโยชน์จากวิทยาศาสตร์ข้อมูลเพื่อให้ได้ข้อมูลเชิงลึก ตัดสินใจอย่างมีข้อมูล และสร้างนวัตกรรมใหม่ๆ เพื่อแข่งขันในตลาดได้อย่างมีประสิทธิภาพมากขึ้น

### 2.5.1 ความเข้าใจธุรกิจและวัตถุประสงค์ของธุรกิจ (Business Understanding & Business Objective)

นี่คือขั้นตอนเริ่มต้นที่กำหนดเป้าหมายและวัตถุประสงค์ของธุรกิจ การเข้าใจว่าธุรกิจต้องการบรรลุผลอะไร เป็นสิ่งที่นำทางกระบวนการวิทยาศาสตร์ข้อมูลทั้งหมด โดยต้องระบุคำถามและปัญหาหลักที่การวิเคราะห์ข้อมูลต้องการแก้ไข

### 2.5.2 การเก็บรวบรวมข้อมูลและการรวมข้อมูล (Data Collection & Data Integration)

ในขั้นตอนนี้ ข้อมูลจะถูกรวบรวมจากแหล่งข้อมูลต่างๆ ที่เกี่ยวข้องกับปัญหาธุรกิจ การรวมข้อมูลคือการรวมข้อมูลจากแหล่งต่างๆ เพื่อสร้างชุดข้อมูลที่สมบูรณ์สำหรับการวิเคราะห์

### 2.5.3 ความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนนี้เกี่ยวข้องกับการทำความคุ้นเคยกับข้อมูลที่รวบรวมมา กิจกรรมหลักได้แก่:

1. การแสดงผลหรือสำรวจข้อมูล (EDA: Exploratory Data Analysis): **การวิเคราะห์ข้อมูลเชิงสำรวจ (EDA) ใช้เพื่อสรุปลักษณะหลักของข้อมูล โดยมักใช้วิธีการแสดงผลทางภาพ EDA ช่วยในการเข้าใจการกระจายข้อมูล การระบุข้อมูลที่ผิดปกติ และค้นพบรูปแบบ**
2. การจัดการและวิเคราะห์ข้อมูลพื้นฐาน (Basic Data Manipulation & Analysis): **รวมถึงการดำเนินการพื้นฐานเช่นการกรอง การจัดกลุ่ม และการสรุปข้อมูลเพื่อเข้าใจโครงสร้างและคุณภาพของข้อมูล**

### 2.5.4 การเตรียมข้อมูล (Data Preparation / Data Pre-Processing)

การเตรียมข้อมูลเกี่ยวข้องกับการทำความสะอาดและแปลงข้อมูลดิบให้เป็นรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ ซึ่งอาจรวมถึงการจัดการค่าที่ขาดหายไป การปรับมาตรฐานข้อมูล การเข้ารหัสตัวแปรประเภท และการแยกข้อมูลออกเป็นชุดฝึกอบรมและชุดทดสอบ

### 2.5.5 การสร้างแบบจำลอง (Modeling)

ขั้นตอนนี้เกี่ยวข้องกับการใช้เทคนิควิทยาศาสตร์ข้อมูลต่างๆ กับข้อมูลที่เตรียมไว้:

- **การทำเหมืองข้อมูล (Data Mining)**: นี่คือการสำรวจข้อมูลเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ เทคนิคเช่นการจัดกลุ่ม การเชื่อมโยง และการตรวจจับความผิดปกติถูกใช้งานบ่อยครั้ง
- **การเรียนรู้ของเครื่อง (Machine Learning)**: อัลกอริธึมการเรียนรู้ของเครื่องใช้เพื่อสร้างแบบจำลองการทำนาย แบบจำลองเหล่านี้เรียนรู้จากข้อมูลเพื่อทำการทำนายหรือจัดประเภทข้อมูลใหม่
- **การเรียนรู้เชิงลึก (Deep Learning)**: การเรียนรู้เชิงลึกใช้เครือข่ายประสาทเทียมที่มีหลายชั้นเพื่อจำลองรูปแบบที่ซับซ้อนในข้อมูล

### 2.5.6 การพัฒนาผลิตภัณฑ์ข้อมูล (Data Product Development / Product Dev. from Analytics Results)

ขั้นตอนสุดท้ายเกี่ยวข้องกับการใช้ข้อมูลเชิงลึกและแบบจำลองที่พัฒนาในขั้นตอนก่อนหน้าเพื่อสร้างผลิตภัณฑ์ที่ขับเคลื่อนด้วยข้อมูล ซึ่งอาจรวมถึงการสร้างแดชบอร์ด การพัฒนาระบบแนะนำ หรือการประยุกต์ใช้อื่นๆ ที่ใช้ผลลัพธ์การวิเคราะห์ในการแก้ปัญหาธุรกิจและบรรลุวัตถุประสงค์ของธุรกิจ

## 2.6 การทำเหมืองข้อมูล (data mining)คืออะไร? การทำเหมืองข้อมูลแตกต่างจากวิทยาการข้อมูล (data science) อย่างไร?

การทำเหมืองข้อมูล (data mining) เป็นหนึ่งในเทคนิคของวิทยาศาสตร์ข้อมูล (data science) สำหรับการวิเคราะห์ข้อมูล (data analytics) เป็นกระบวนการประมวลผลข้อมูลที่ใช้ความสามารถในการค้นหาข้อมูลขั้นสูงและอัลกอริธึมทางสถิติเพื่อค้นหารูปแบบและความสัมพันธ์ในฐานข้อมูลขนาดใหญ่ที่มีอยู่เดิม เป็นวิธีการในการค้นพบความหมายใหม่ในข้อมูล โดยรายละเอียดจากภาพมีดังนี้:

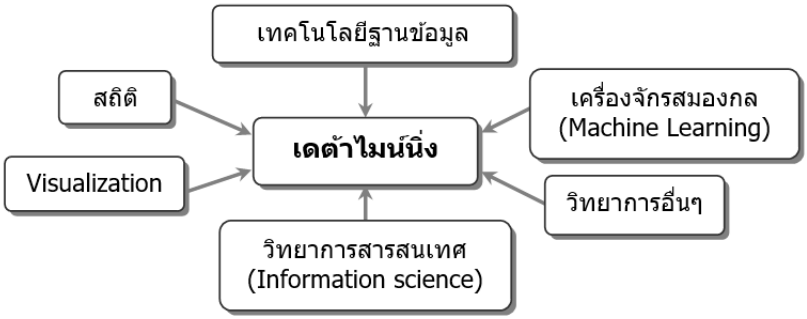
**นิยามจาก WordNet:** การทำเหมืองข้อมูลหมายถึง:

- การประมวลผลข้อมูลโดยใช้ความสามารถในการค้นหาข้อมูลขั้นสูง
- ใช้อัลกอริธึมทางสถิติเพื่อค้นหารูปแบบ (patterns) และความสัมพันธ์ (correlations) ในฐานข้อมูลขนาดใหญ่ที่มีอยู่ก่อนแล้ว
- เป็นวิธีการเพื่อค้นพบความหมายใหม่ในข้อมูล

ดังนั้น การทำเหมืองข้อมูลเป็นเครื่องมือสำคัญที่ช่วยให้ธุรกิจสามารถวิเคราะห์ข้อมูลขนาดใหญ่และค้นพบข้อมูลเชิงลึกที่สามารถนำไปใช้ในการตัดสินใจเชิงกลยุทธ์และการพัฒนาธุรกิจได้

## 2.7 วิทยาการที่เกี่ยวพันกับการทำเหมืองข้อมูล

วิวัฒนาการของวิทยาการแขนงต่างๆ มีความสำคัญและช่วยสนับสนุนระบบการทำเหมืองข้อมูลทั้งสิ้น เนื่องจากการทำเหมืองข้อมูลได้นำหลักการและทฤษฎีจากวิทยาการหลายแขนงมาใช้ร่วมกัน ได้แก่ เทคโนโลยีฐานข้อมูล สถิติ สารสนเทศศาสตร์ การแสดงข้อมูลด้วยภาพ และ การเรียนรู้ของเครื่อง เป็นต้น




### 2.7.1 เทคโนโลยีฐานข้อมูล

เทคโนโลยีฐานข้อมูล (database technology) จากอดีตจนถึงปัจจุบันมีการพัฒนาเทคนิคต่างๆ ตั้งแต่การออกแบบ การเก็บ และการนำข้อมูลมาใช้แสดงผลหรือทำรายงานต่างๆ การพัฒนาดังกล่าวนำมาสู่ศักยภาพในการสะสมข้อมูลจำนวนข้อมูลขนาดมหาศาลในฐานข้อมูล ถึงแม้ว่าจะมีการพัฒนาเทคโนโลยีคลังข้อมูลเพื่อใช้วิเคราะห์ข้อมูลแบบออนไลน์แล้วก็ตาม แต่การวิเคราะห์เจาะลึกเข้าไปในข้อมูลเพื่อค้นหาความรู้ที่ซ่อนอยู่ในแหล่งข้อมูลมหาศาลนั้นยังคงมีการพัฒนาในระดับต่ำ

การทำเหมืองข้อมูลสามารถช่วยทำให้การขุดค้นความรู้ที่ซ่อนอยู่ในเทคโนโลยีฐานข้อมูลเป็นจริงขึ้นมาได้ และการพัฒนาเทคโนโลยีฐานข้อมูลที่มากขึ้นก็เท่ากับส่งเสริมให้การทำเหมืองข้อมูลมีความสำคัญมากขึ้นด้วย

### 2.7.2 สถิติ

ในช่วงหลายศตวรรษที่ผ่านมา มนุษย์ได้ใช้เทคนิคทางสถิติ (statistics) ในการทำความเข้าใจกับความเป็นจริงต่างๆ ของโลก เช่น อัลกอริทึมที่ใช้ทำนาย ทฤษฎีถดถอย (regression) การสุ่มตัวอย่าง (sampling) และการออกแบบการทดลอง (experimental design) ต่างๆ เป็นต้น ปัจจุบันทฤษฎีทางสถิติเหล่านี้ได้ถูกนำมาปรับใช้ทางธุรกิจ แล้ว เช่น โปรแกรม SPSS (xxx) ซึ่งเป็นโปรแกรมที่ใช้ช่วยวิเคราะห์ข้อมูลและคำนวณทางสถิติต่างๆ โดยมีสถิติที่เตรียมไว้ให้ใช้งาน การวิเคราะห์ข้อมูลโดยอาศัยหลักการทางสถิตินั้นนิยมทำกับกลุ่มตัวอย่าง เนื่องจากข้อจำกัดในการเก็บข้อมูลและการประมวลผลข้อมูลที่ซับซ้อนส่งผลกระทบกับเวลาและค่าใช้จ่ายในการทำงาน อีกทั้งต้องอาศัยผู้เชี่ยวชาญด้านสถิติในการแปลความหมายของผลลัพธ์ที่ได้

การทำเหมืองข้อมูลมีความสัมพันธ์อย่างแนบแน่นกับสถิติ เนื่องจากสถิติมีความสำคัญอย่างมากในการช่วยจัดเตรียมอัลกอริทึมพื้นฐานสำหรับการทำเหมืองข้อมูล ทำให้เกิดความเข้าใจและสร้างเป็นกลไกเทคนิควิธีการที่เหมาะสมที่สุดขึ้นมา การทำเหมืองข้อมูลช่วยทำให้การวิเคราะห์ข้อมูลและเตรียมข้อมูลจากฐานข้อมูลขนาดใหญ่เป็นไปได้อย่างง่ายดายโดยไม่จำเป็นต้องอาศัยผู้เชี่ยวชาญด้านสถิติหรือนักสถิติ ลดข้อจำกัดด้านเวลาและค่าใช้จ่ายในการทำงานเมื่อเทียบกับการใช้สถิติแต่เพียงอย่างเดียว

### 2.7.3 สารสนเทศศาสตร์

การพัฒนาอย่างรวดเร็วของสารสนเทศศาสตร์ (information science) ไม่ว่าจะเป็นด้านฮาร์ดแวร์และซอฟต์แวร์ ได้เพิ่มศักยภาพและความสามารถของการประมวลผลและเนื้อที่การจัดเก็บในขนาดที่เล็กและราคาถูกลงอย่างมาก ประกอบกับมีการพัฒนาเทคโนโลยีเครือข่ายและการเชื่อมต่อ โดยนำเครื่องไมโครคอมพิวเตอร์จำนวนมากมาเชื่อมต่อกันในระบบเครือข่ายความเร็วสูง ทำให้ได้ระบบคอมพิวเตอร์สมรรถนะสูงเพิ่มมากขึ้น อีกทั้งทำงานกับข้อมูลและสารสนเทศในปัจจุบันอย่างประสิทธิภาพมากขึ้น

วิทยาการสารสนเทศส่งเสริมให้การจัดการกับสารสนเทศได้ง่ายขึ้นและดีขึ้น แต่ถึงกระนั้นวิทยาการสารสนเทศเพียงศาสตร์เดียวย่อมไม่สามารถก่อให้เกิดการค้นพบความรู้ใหม่ๆ จากฐานข้อมูลได้ การทำเหมืองข้อมูลเป็นแนวทางสมัยใหม่ในการบูรณาการศาสตร์ทั้งหลาย ทั้งทางสถิติ เทคโนยี่ทางคอมพิวเตอร์ทั้งในส่วนของ ฮาร์ดแวร์ ซอฟต์แวร์ ศาสตร์ทางด้านวิทยาการสารสนเทศ และอื่นๆมาเป็นองค์รวมในการทำงานเข้ากันในการหาความรู้อย่างชาญฉลาด

### 2.7.4 การแสดงข้อมูลด้วยภาพ

การแสดงข้อมูลด้วยภาพ (visualization) เป็นการนำเสนอข้อมูลตัวเลขเป็นรูปภาพ กราฟสองมิติแบบต่างๆ และภาพเชิงซ้อนแบบหลายมิติที่มีความสัมพันธ์กัน ทำให้ข้อมูลอยู่ในรูปแบบที่ง่ายต่อการเข้าใจและแปลความหมาย ซึ่งเปรียบได้กับคำพูดที่ว่า “ภาพเพียงภาพเดียวสามารถแทนข้อความได้นับพัน”

โดยทั่วไปการแสดงผลด้วยภาพมักจะแสดงผลข้อมูลที่ผ่านการประมวลผลหรือสรุปมาแล้วและมีจำนวนข้อมูลไม่มากนักเพื่อให้มองเห็นภาพทั้งหมดได้อย่างชัดเจน จึงอาจมองว่าเป็นข้อจำกัดอย่างหนึ่งของวิทยาการแสดงผลด้วยภาพเมื่อเทียบกับการทำเหมืองข้อมูลซึ่งมีกระบวนการในการกระทำกับข้อมูลเพื่อให้เกิดรูปแบบต่างๆ ได้อย่างมากมายและนำวิทยาการแสดงผลด้วยภาพมาต่อยอดเพื่อนำเสนอรูปแบบกฎเกณฑ์ที่ขุดค้นได้ อย่างไรก็ตาม ปัจจุบันได้มีการพัฒนาเทคนิคการแสดงผลด้วยภาพจนอาจถือได้ว่าเป็นเทคนิคที่สำคัญอย่างหนึ่งของการทำเหมืองข้อมูลด้วย เนื่องจากได้นำมาประยุกต์สร้างเป็นโปรแกรมสำหรับค้นพบความรู้จากข้อมูลได้ เช่นโปรแกรม 3DV8 เป็นการนำข้อมูลมาแสดงเป็นรูปภาพที่มีความสัมพันธ์กันแบบหลายมิติเพื่อหาแนวโน้ม คุณลักษณะที่คล้ายหรือต่างกัน จุดที่น่าสนใจคือแต่ละชุดของข้อมูลที่ใช้ทำงานอาจค้นพบสิ่งใหม่ๆ ที่แตกต่างกันก็ได้

### 2.7.5 การเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (machine learning) หรือการทำให้เครื่องจักรเรียนรู้ได้อย่างมนุษย์นั้นมาจากวิทยาการคอมพิวเตอร์สาขาปัญญาประดิษฐ์ (artificial intelligence: AI) โดยมุ่งเน้นไปที่ความพยายามที่จะทำให้คอมพิวเตอร์มีความฉลาด ซึ่งอาจกล่าวแบบเจาะจงได้ว่า กลุ่มนักวิจัยที่ทำให้เครื่องจักรเรียนรู้ได้เองมีความสนใจที่จะพัฒนาโปรแกรมให้มีความสามารถที่จะเรียนรู้สิ่งต่างๆ ได้เองจากการศึกษากลุ่มตัวอย่างที่ให้มา การเรียนรู้ด้วยตัวเองประเภทแรกคือการค้นพบความสามารถที่จะกระทำกับงานบางอย่างให้ได้ เช่นความสามารถที่จะจดจำลักษณะการเขียนหรือลายมือที่เขียนได้ ในบางกรณีการเรียนรู้ใหม่ๆ ได้แสดงออกมาในรูปของกฎเกณฑ์ที่มีการพิสูจน์มาจากตัวอย่างต่างๆ ซึ่งมีโครงข่ายใยประสาท (neural networks) เป็นการเรียนรู้โดยเครื่องจักรประเภทแรกที่ได้รับการพิสูจน์แล้วว่าประสบความสำเร็จอย่างสูง

อาจกล่าวได้ว่าการทำเหมืองข้อมูลแตกสาขาออกมาจากการเรียนรู้โดยเครื่องจักร เนื่องจากการทำเหมืองข้อมูลถูกนำเสนอเป็นครั้งแรกโดยกลุ่มนักวิจัยที่นำวิธีการเรียนรู้โดยเครื่องจักรมาประยุกต์เข้ากับวิทยาการแขนงอื่นๆ ที่นอกเหนือไปจากวิทยาการคอมพิวเตอร์และปัญญาประดิษฐ์ เช่น การควบคุมกระบวนการผลิตทางอุตสาหกรรม และการขายตรง เป็นต้น การค้นพบนี้สนับสนุนอัลกอริทึมและการจดจำถึงรูปแบบหรือกฎเกณฑ์ของข้อมูล รวมไปถึงการ

สร้างกลไกเทคนิคการทำงานของการทำเหมืองข้อมูลซึ่งทำให้ผู้ที่ทำการทำเหมืองข้อมูล ข้อมูลกลายเป็นบุคคลที่ฉลาดและได้รับผลลัพธ์ที่ดีโดยไม่จำเป็นต้องเป็นนักคณิตศาสตร์หรือนักสถิติแต่อย่างใด
แม้ว่าการพัฒนาของการเรียนรู้เครื่องจักรสมองกลจะสามารถสร้างให้เกิดรูปแบบกฎเกณฑ์ที่เป็นความรู้ขึ้นมาได้ก็ตาม แต่มักใช้กับกลุ่มตัวอย่างขนาดเล็กเพื่อให้เกิดการเรียนรู้โดยไม่เน้นที่การวิเคราะห์เจาะลึกกับฐานข้อมูลขนาดใหญ่เพื่อให้ค้นพบความรู้ใหม่ๆ จุดนี้เองที่ทำให้เกิดความแตกต่างระหว่างการเรียนรู้เครื่องจักรสมองกลกับการทำเหมืองข้อมูล และทำให้การทำเหมืองข้อมูลเป็นวิทยาการที่เหมาะกับการขุดค้นความรู้จากฐานข้อมูลขนาดใหญ่ๆ ได้อย่างมีประสิทธิภาพในยุคปัจจุบัน

## 2.8 โจทย์ลักษณะแบบใดบ้างไม่เหมาะกับการประยุกต์ใช้วิทยาการข้อมูล (data science) และการทำเหมืองข้อมูล (data mining)

โจทย์ที่มีวิธีการแก้ไขปัญหาตายตัว เพราะ มีวิธีที่ชัดเจนเป็นลำดับขั้นตอน

การรักษาความเป็นส่วนตัว -> เนื่องจากการทำเหมืองข้อมูลอาจจะมีความกระทบต่อความเป็นส่วนตัวได้สำหรับโจทย์ที่เป็นเรื่องความเป็นส่วนตัว

โจทย์ที่มีข้อมูลขนาดเล็ก เนื่องจากมีหากข้อมูลมีไม่มากพอการทำเหมืองข้อมูลอาจทำให้ได้ข้อมูลที่ผิดพลาด

## 2.9 คำถามท้ายบท

1. การวิเคราะห์ข้อมูล (data analytics) คืออะไร? อธิบายประโยชน์ของการวิเคราะห์ข้อมูลในแง่ของการลดต้นทุน (resource and time reduction) การเพิ่มมูลค่า (value adding) และการสร้างนวัตกรรม (innovation creation) พร้อมยกตัวอย่างประกอบ
2. การวิเคราะห์ข้อมูลที่ซับซ้อน (complex data analytics) คืออะไร? อธิบายคุณสมบัติของข้อมูลที่ซับซ้อน (characteristics of complex data)? อธิบายความท้าทายของการวิเคราะห์ข้อมูลที่ซับซ้อน (complex data analytics)ในแง่ของการลดต้นทุน (resource and time reduction) การเพิ่มมูลค่า (value adding) และการสร้างนวัตกรรม (innovation creation) พร้อมยกตัวอย่างประกอบ
3. อธิบายและเปรียบเทียบความหมายของ Descriptive Analytics, Predictive Analytics, และ Prescriptive Analytics? ยกตัวอย่างกรณีธุรกิจที่คุณสนใจ (your choice of business)? อธิบายการประยุกต์ใช้ Descriptive Analytics, Predictive Analytics, และ Prescriptive Analytics สำหรับกรณีศึกษาที่คุณเลือก
4. วิทยาการข้อมูล (data science) คืออะไร? อธิบายทักษะที่จำเป็นในการเป็นนักวิทยาการข้อมูล (data scientist) ? คุณคิดว่าคุณยังขาดทักษะใดบ้างในการเป็นนักวิทยาการข้อมูลที่ดี?
5. อธิบายขั้นตอนต่าง ๆ ของวิทยาการข้อมูล (data science) พร้อมยกตัวอย่างประกอบแต่ละขั้นตอน

6. การทำเหมืองข้อมูล (data mining)คืออะไร? การทำเหมืองข้อมูลแตกต่างจากวิทยาการข้อมูล (data science) อย่างไร? อธิบายวิทยาการทางคอมพิวเตอร์ (computer science fields) ที่เกี่ยวพันกับการทำเหมืองข้อมูล?
7. โจทย์ลักษณะแบบใดบ้างไม่เหมาะกับการประยุกต์ใช้วิทยาการข้อมูล (data science) และการทำเหมืองข้อมูล (data mining) พร้อมระบุเหตุผล