



มหาวิทยาลัยสุโขทัยธรรมาธิราช  
Sukhothai Thammathirat Open University

# สัมมนาเสริม 99711

## วิทยาการข้อมูล&ข้อมูลขนาดใหญ่

วันที่ 9-10 พฤศจิกายน 2567 เวลา 9.00 - 16.00 น.



มหาวิทยาลัยสุโขทัยธรรมาราช  
Sukhothai Thammathirat Open University

## วิทยากร

รองศาสตราจารย์ ณัฐพร เห็นเจริญเลิศ

Email : [nuttaporn.hen@stou.ac.th](mailto:nuttaporn.hen@stou.ac.th)

รองศาสตราจารย์ ดร.วณิชยา ร่มสายหยุด

Email : [walisa.rom@stou.ac.th](mailto:walisa.rom@stou.ac.th)

ติดต่อสาขาวิชา โทร 02-504-8191-3

# หัวข้อการบรรยาย

วันเสาร์ที่ 9 พฤศจิกายน 2567

9.00-12.00 น. บรรยายสรุปเนื้อหา 5 หน่วย

13.00-16.00 น. ฝึกปฏิบัติ การวิเคราะห์เชิงทำนายข้อมูลสำหรับนักวิทยาการข้อมูลโดยใช้ภาษาไพธอน ด้วยเครื่องมือ Colab (กิจกรรมเดี่ยว)


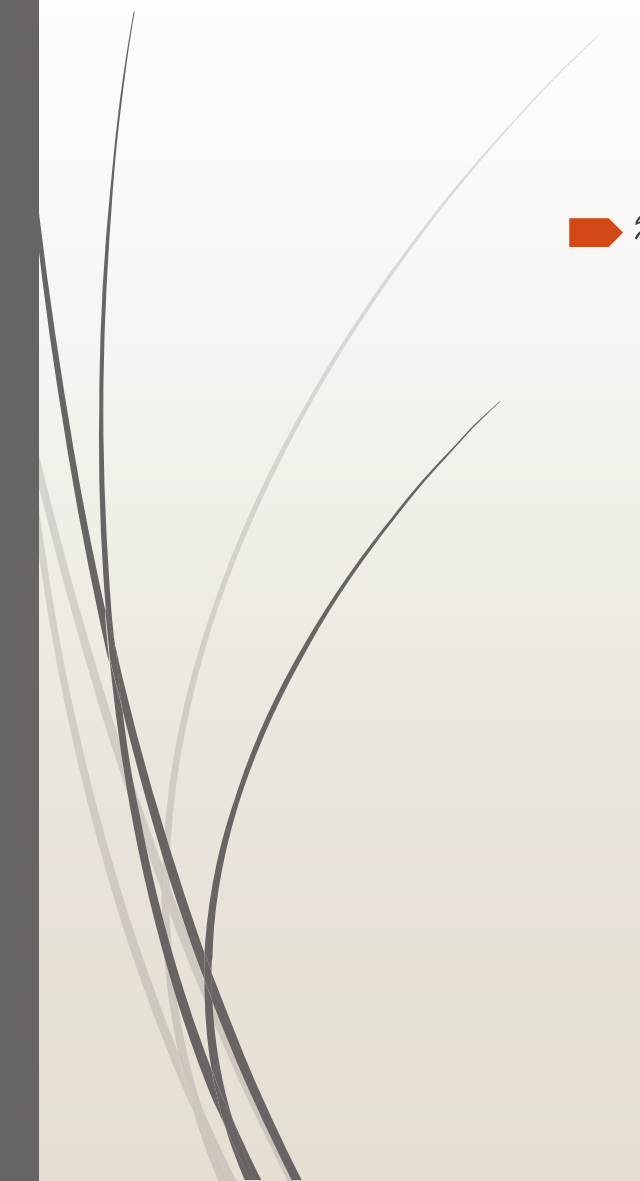
วันอาทิตย์ที่ 10 พฤศจิกายน 2567

9.00-12.00 น. ฝึกปฏิบัติ การวิเคราะห์เชิงทำนายข้อมูลสำหรับนักวิทยาการข้อมูลโดยใช้ภาษาไพธอน ด้วยเครื่องมือ Colab (กิจกรรมกลุ่ม)

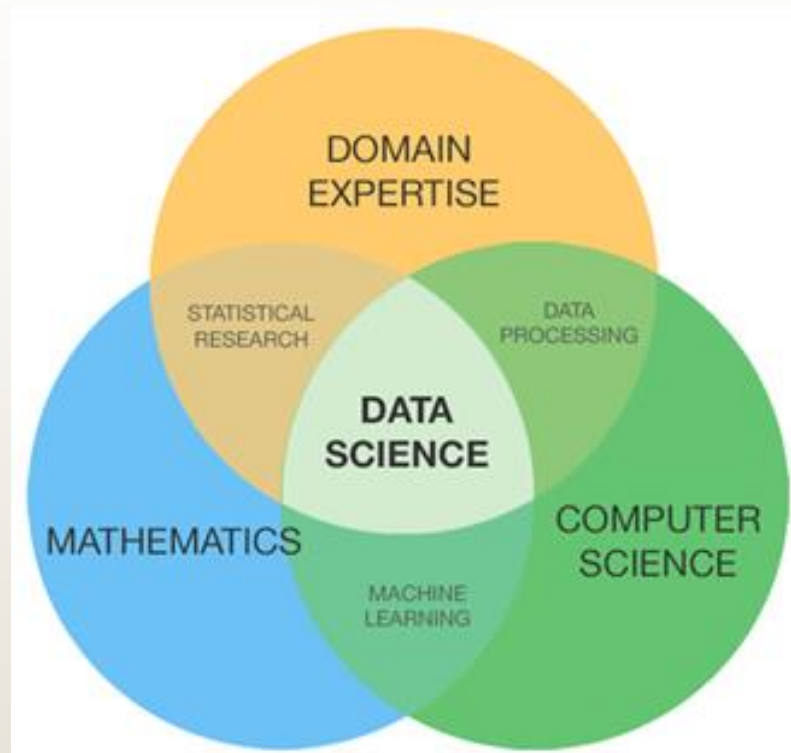
13.00-16.00 น. นักศึกษาแต่ละกลุ่มนำเสนอผลงาน กลุ่มละ 30 นาที จากนั้นเปิดโอกาสให้เพื่อนๆ กลุ่มอื่นๆ ได้ซักถามและตอบคำถาม

## แนวคิดเกี่ยวกับวิทยาการข้อมูล

- ➡ โลกเข้าสู่ยุคของข้อมูลขนาดใหญ่ (big data) ทุกวินาทีมีข้อมูลจำนวนมากเกิดขึ้นและหลงเหลืออยู่ในระบบเครือข่ายคอมพิวเตอร์ อินเทอร์เน็ต และการสื่อสาร
- ➡ ข้อมูลถูกรวบรวม จัดเก็บไว้มาวิเคราะห์เพื่อนำมาใช้ประโยชน์ในรูปแบบต่างๆ
  - ➡ อธิบายเหตุการณ์ ทำนาย เพื่อสนับสนุนการตัดสินใจ วางแผนงาน เพิ่มประสิทธิภาพการดำเนินงาน
- ➡ วิทยาการข้อมูล หรือ วิทยาศาสตร์ข้อมูล (data science) หมายถึง การใช้กระบวนการทางวิทยาศาสตร์ (science) ที่มีระเบียบแบบแผน ตรวจสอบได้ สามารถสกัดองค์ความรู้และความเข้าใจจากข้อมูล เพื่อนำไปใช้ประโยชน์ทั้งในด้านต่าง ๆ

- 
- 
- วิทยาการข้อมูลมีบทบาทสำคัญในแทบทุกด้านของการดำเนินธุรกิจ การกำหนดกลยุทธ์ต่าง ๆ
    - ช่วยในการวิเคราะห์เพื่อคาดการณ์แนวโน้มของสินค้า แนวโน้มการลงทุน
    - ช่วยในการจัดการความเสี่ยงทางการเงิน ตรวจสอบธุรกรรมที่มีความเสี่ยงต่อการทุจริต
    - ช่วยป้องกันการเสียหายของอุปกรณ์ในโรงงานผลิต
    - ช่วยป้องกันการโจมตีทางไซเบอร์ และภัยคุกคามความปลอดภัยในระบบไอที

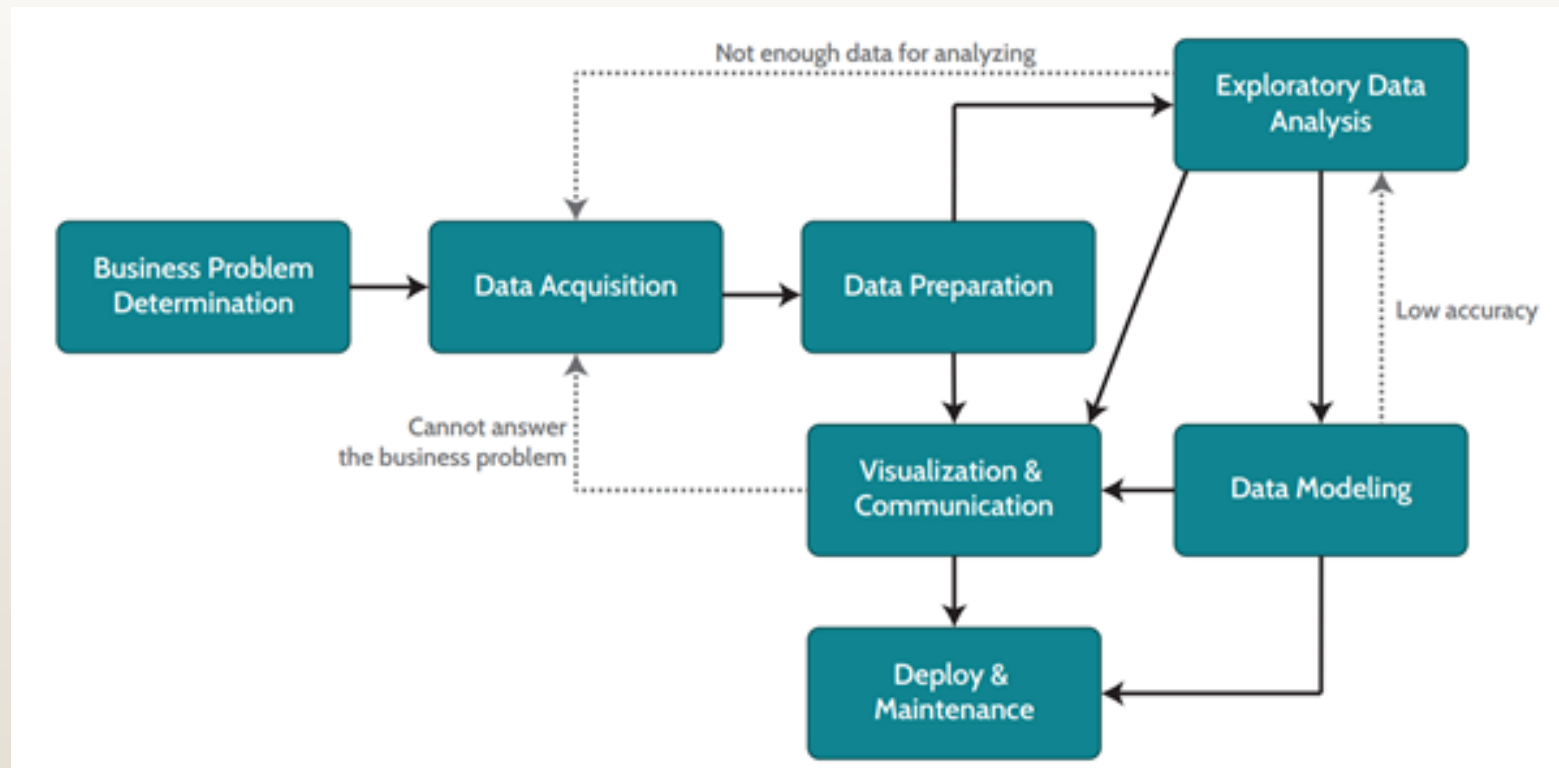
## องค์ความรู้และสาขาที่เกี่ยวข้องกับวิทยาการข้อมูล





แผนภาพเวนน์แสดงองค์ความรู้พื้นฐานของวิทยาการข้อมูล อ้างอิง Shelly Palmer (2015)


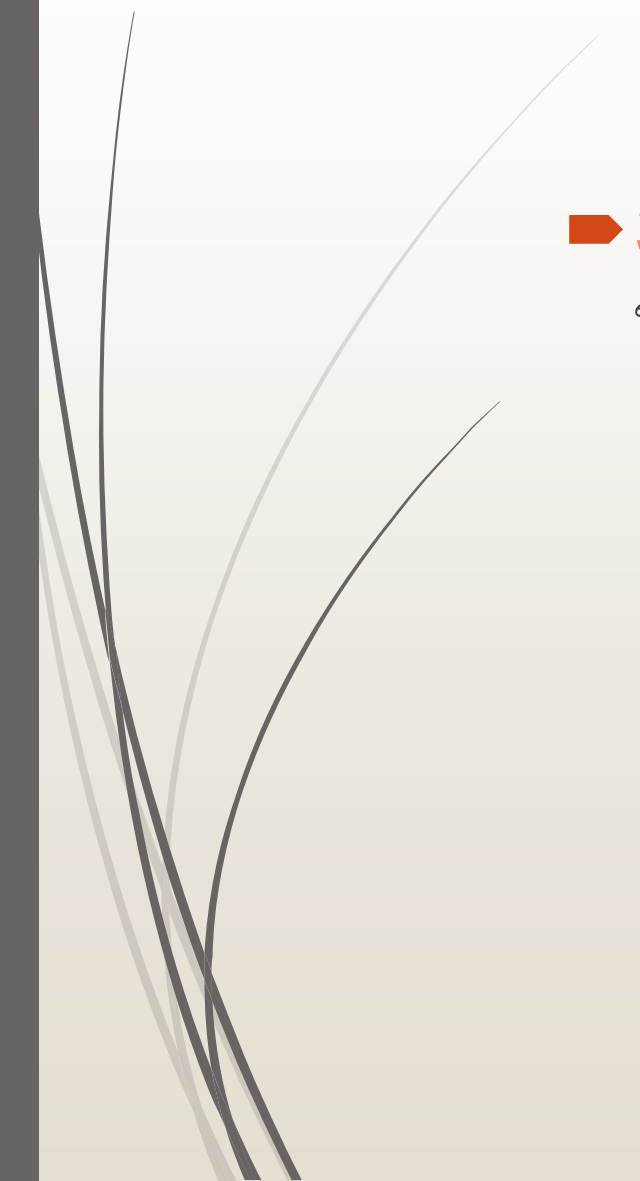
## กระบวนการทางวิทยาการข้อมูล


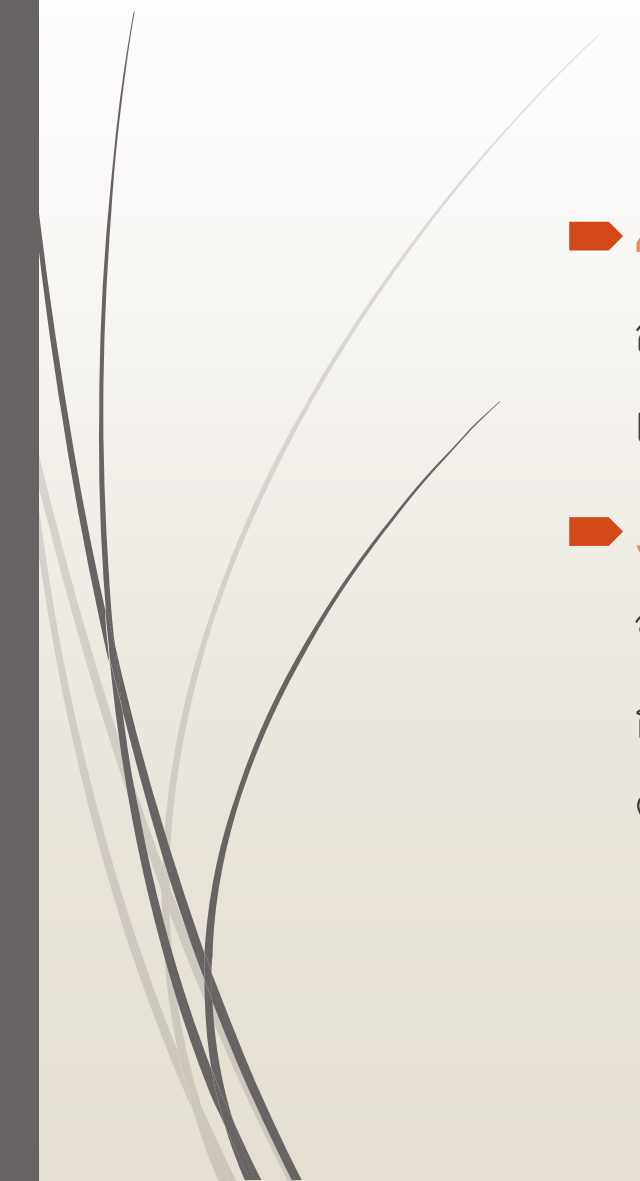
- เป็นกรอบการดำเนินงานทั่วไปในการวิเคราะห์ แก้ปัญหา ค้นหาองค์ความรู้ ตามโจทย์ที่ต้องการวิเคราะห์



- 
- 
- ➡ **1. การกำหนดปัญหา (business problem determination)** ทำความเข้าใจและวิเคราะห์ปัญหา เพื่อให้ได้ซึ่งวัตถุประสงค์หรือเป้าหมายของปัญหาที่ชัดเจนพร้อมทั้งระบุขอบเขตของปัญหา
  - ➡ **2. การเก็บรวบรวมข้อมูล (data acquisition)**
    - ➡ ข้อมูลอาจได้มาจากหลาย ๆ แหล่ง หลากหลายรูปแบบ
    - ➡ ต้องอาศัยการวางแผนและการบริหารจัดการที่ดี
    - ➡ อาจได้มาจากการสัมภาษณ์ การสำรวจ การสังเกต การทดลอง การทบทวนเอกสาร  
การสำมะโน



- 
- 
- ➡ **3. การจัดเตรียมข้อมูล (data preparation)** จัดเตรียมข้อมูลให้พร้อมสำหรับการนำไปวิเคราะห์ ข้อมูลจะถูกตรวจสอบและปรับปรุงคุณภาพ
    - ➡ *การทำความสะอาดข้อมูล (data cleansing)* กระบวนการตรวจสอบ แก้ไข หรือลบรายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล
    - ➡ *การแปลงข้อมูล (data transformation)* การเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมสำหรับการประมวลผล
    - ➡ *การเชื่อมโยงข้อมูล (data combining)* กรณีข้อมูลมาจากหลายแหล่ง ต้องทำการเชื่อมโยงข้อมูลจากหลายแหล่งเข้าด้วยกัน

- 
- 
- ➡ **4. การสำรวจข้อมูล (exploratory data analysis)** กระบวนการตรวจสอบ การกลั่นกรองสำรวจข้อมูลที่ได้มาให้พร้อมต่อการนำไปประมวลผลให้ได้ข้อมูลที่มีคุณภาพ ทำให้เราเข้าใจพื้นฐานเกี่ยวกับข้อมูลชุดนั้น และเป็นการตรวจความผิดพลาดของชุดข้อมูล
  - ➡ **5. การสร้างแบบจำลองข้อมูล (data modeling)** เป็นการนำเอาเทคนิคคณิตศาสตร์และข้อมูลที่จัดเตรียมไว้มาวิเคราะห์ในรูปแบบการพยากรณ์หรือรูปแบบการให้คำแนะนำ โดยส่วนใหญ่จะเป็นการวิเคราะห์ด้วยเทคนิคมากกว่า 1 แบบ เพื่อให้เกิดการประเมินค่าการวิเคราะห์ ความแม่นยำ หรือค่าความผิดพลาด ที่ตอบโจทย์ตามวัตถุประสงค์และมีประสิทธิภาพมากที่สุด

- 
- 
- ➡ **6. การนำเสนอผลการวิเคราะห์ข้อมูล หรือการสร้างภาพข้อมูล (visualization and communication)** ส่วนมากจะถูกนำเสนอโดยใช้ภาพเป็นสื่อกลางเพื่อให้ง่ายต่อการศึกษาและการทำความเข้าใจ เพื่อการสื่อสาร และสร้างการรับรู้ให้กับกลุ่มคนที่เกี่ยวข้อง

- ➡ เช่น แผนภูมิ กราฟต่างๆ แสดงการเปรียบเทียบ หรือแนวโน้มของข้อมูล

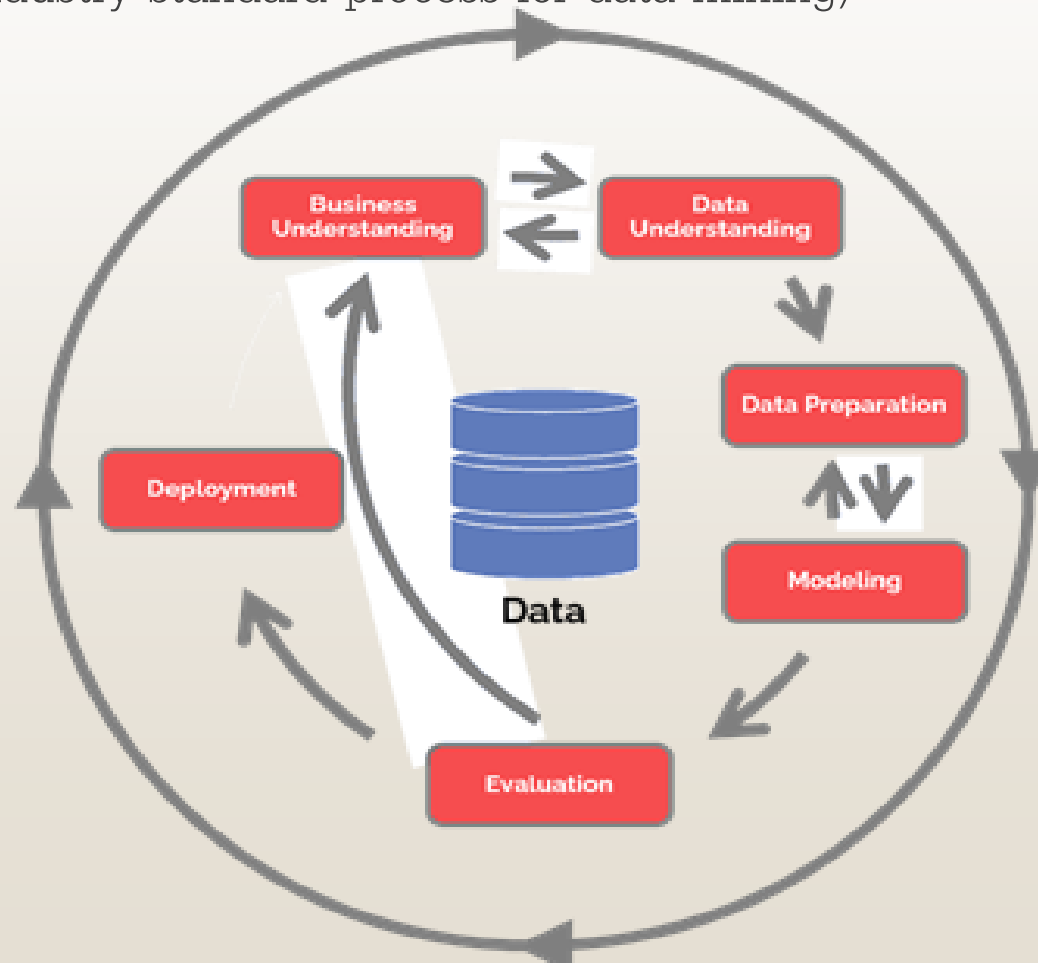
- ➡ การใช้โปรแกรม เครื่องมือทางการสร้างภาพข้อมูลโดยเฉพาะ เช่น dashboard

- ➡ **7. การนำไปใช้และบำรุงรักษา (deploy and maintenance)** เป็นการนำผลการวิเคราะห์ข้อมูลที่ได้รับ การตรวจสอบแบบจำลองให้เหมาะสมกับสถานการณ์ที่อาจเกิดขึ้นจริงในทางปฏิบัติและได้รับการยอมรับ

- ➡ อาจเป็น application ผ่านทางช่องทางต่าง ๆ การ update ข้อมูลให้ทันสมัย


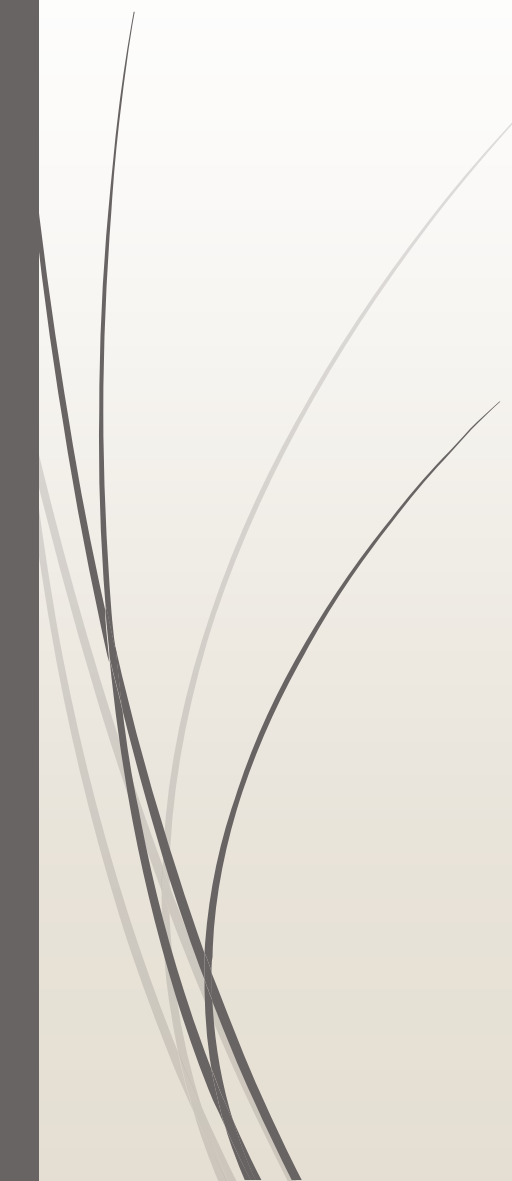
## กระบวนการ CRISP-DM

- การวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล (Data Mining) ด้วยมาตรฐาน CRISP-DM (the cross industry standard process for data mining)



## กระบวนการ CRISP-DM

- **1. business understanding** เป็นการทำความเข้าใจในงาน โอกาส และระบุปัญหาที่จะเกิดขึ้นกับธุรกิจ การกำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์
  - กำหนดวัตถุประสงค์ทางธุรกิจ (determine business objectives)
  - ประเมินสถานการณ์ (assess situation)
  - กำหนดเป้าหมายการทำเหมืองข้อมูล (determine data mining goals)
  - จัดทำแผนโครงการ (produce project plan)

- 
- 
- **2. data understanding** ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์
  - **3. data preparation** ทำการแปลงข้อมูล (raw data) ให้กลายเป็นข้อมูลที่สามารถนำมาช่วยในการวิเคราะห์ต่อไป
    - คัดเลือกข้อมูล (select data)
    - ทำความสะอาดข้อมูล (clean data)
    - สร้างข้อมูล (construct data)
    - รวมข้อมูล (integrate data)
    - จัดรูปแบบข้อมูล (format data)



➤ **4. modeling** การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูล

➤ บางครั้งอาจมีการย้อนกลับไปปรับการเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

➤ การปรับเทคนิค วิธีการการสร้างแบบจำลองที่แตกต่างกันหลายแบบ

➤ **5. Evaluation** การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับวัตถุประสงค์หรือเป้าหมายที่ได้ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด

➤ **6. deployment** การนำเอาข้อมูลที่เป็นผลลัพธ์หรือแบบจำลองที่ได้ มาลองปฏิบัติจริงกับธุรกิจในองค์กร โดยแปลงแนวคิดที่มีให้เกิดเป็นสารสนเทศเพื่อให้ผู้บริหารหรือนักการตลาดเข้าใจ

## หลักการของข้อมูลขนาดใหญ่ (Big Data)

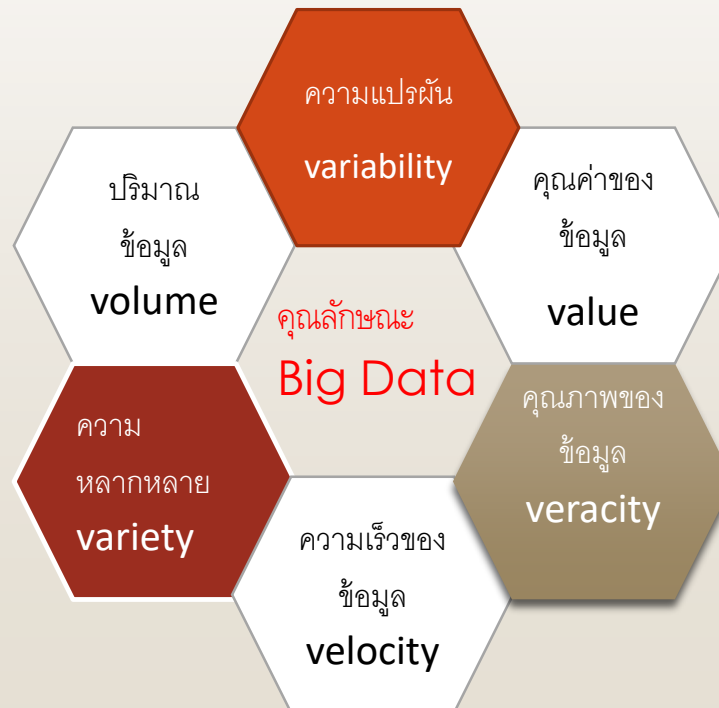
ความก้าวหน้าทางเทคโนโลยี ส่งผลให้เกิดปรากฏการณ์ปริมาณข้อมูลขนาดใหญ่ในเวลาเพียงสั้น ๆ ตั้งแต่ปี ค.ศ. 1990 เป็นต้นมา จาก e-commerce โซเชียลมีเดีย จากการติดต่อสื่อสารระหว่างอุปกรณ์อัตโนมัติต่าง ๆ

**ข้อมูลขนาดใหญ่ (big data)** หรือบิกดาตา หมายถึง ข้อมูลปริมาณมาก มีโครงสร้างข้อมูลที่หลากหลายทุกรูปแบบ มีการเกิดขึ้นและเปลี่ยนแปลงไปได้ตลอดเวลา มีคุณภาพข้อมูลและแหล่งที่มาที่เชื่อถือได้ มีประโยชน์และสามารถนำไปวิเคราะห์ได้หลายรูปแบบ



## คุณลักษณะของข้อมูลขนาดใหญ่

- ระยะแรกเริ่มต้นจากคุณลักษณะ 3V ประกอบด้วย ปริมาณ (volume) ความหลากหลาย (variety) ความเร็ว (velocity) ต่อมามีการนิยาม V ต่าง ๆ เพิ่มเติมตามมา ได้แก่ คุณภาพของข้อมูล (veracity) คุณค่า (value) ความแปรผัน (variability)



# คุณลักษณะของข้อมูลขนาดใหญ่

- **1. ปริมาณข้อมูล (volume)** ข้อมูลในยุคปัจจุบันเกิดขึ้นอย่างมากมาย หลากหลายตลอดเวลา
  - จาก social media, mobile application, sensor ของเครื่องจักรอิเล็กทรอนิกส์อัตโนมัติต่างๆ
  - ปริมาณของข้อมูลขนาดใหญ่ควรมีจำนวนมากกว่าหน่วยเทราไบต์ขึ้นไป เมื่อนำมาวิเคราะห์แล้วจะได้ข้อมูลเชิงลึกที่ตรงกับความเป็นจริง
- **2. ความหลากหลายของข้อมูล (variety)** ข้อมูลขนาดใหญ่มาจากแหล่งต่างๆ มากมาย
  - ข้อมูลที่มีโครงสร้าง ไม่มีโครงสร้าง และกึ่งมีโครงสร้าง


➤ **3. ความเร็วของข้อมูล (velocity)** ข้อมูลที่ถูกสร้างขึ้นอย่างรวดเร็วต่อเนื่องและทันเหตุการณ์

➤ ครอบคลุมทั้งความเร็วของการเกิดข้อมูล การจัดเก็บหรือการเคลื่อนย้ายข้อมูล ความเร็วของการวิเคราะห์ข้อมูล

➤ การประมวลผลแบบ **real time** เพื่อรองรับการทำงานกับข้อมูลที่เข้ามาอย่างต่อเนื่อง

➤ **4 คุณภาพของข้อมูล (veracity)** เกี่ยวข้องกับความถูกต้องของข้อมูล (accuracy of data) ความน่าเชื่อถือของแหล่งที่มาของข้อมูล

➤ มีกระบวนการในการตรวจสอบและยืนยันความถูกต้องของข้อมูล



➡ **5. คุณค่าของข้อมูล (value)** ข้อมูลที่มีประโยชน์ มีคุณค่าต่อการนำไปใช้งาน มีความสัมพันธ์และจะต้องเกี่ยวข้องกับวัตถุประสงค์ทางธุรกิจ

➡ **6. ความแปรผันได้ (variability)** ข้อมูลขนาดใหญ่สามารถเปลี่ยนแปลงรูปแบบไปตามการใช้งาน หรือสามารถคิดวิเคราะห์ได้หลากหลายแง่มุม

➡ จึงต้องมีการออกแบบหรือเข้าใจในสภาพแวดล้อมของการจัดเก็บข้อมูลจึงจะวิเคราะห์ข้อมูลได้ถูกต้อง

# ประโยชน์ของข้อมูลขนาดใหญ่

## ➤ 1. การติดตามและตรวจสอบ (monitoring and tracking application)

- การติดตามความรู้สึกลูกค้าที่มีต่อสินค้า
- การติดตามและตรวจสอบทรัพย์สิน
- การติดตามการใช้ไฟฟ้า
- การบำรุงรักษาเครื่องจักร
- การติดตามหรือเฝ้าระวังโรคระบาดหรือโรคติดต่อ



## ➤ 2. การวิเคราะห์และเจาะลึกเพื่อเข้าใจผลิตภัณฑ์หรือบริการ

- ความเข้าใจลูกค้าเพื่อให้บริการ
- การวิเคราะห์อาชญากรรม
- ด้านสุขภาพ

## ➤ 3. พัฒนาผลิตภัณฑ์หรือบริการใหม่

- ธุรกิจประกันภัยรถยนต์
- การส่งเสริมการขายของธุรกิจค้าปลีก
- การบริการเสนอแนะ

# ประเภทของข้อมูลขนาดใหญ่

- ➡ 1. ข้อมูลที่มีโครงสร้าง (structured data) ข้อมูลที่มีการกำหนดรูปแบบและรายละเอียดของข้อมูลไว้อย่างชัดเจน
- ➡ ฐานข้อมูลแบบสัมพันธ์ (relational database)

customer

Cust_no	Cust_name	Cust_address	Cust_credit
1001	ไพศาลสาร์ทแวร์	54 หมู่ 7 ถ.พิบูลย์สงคราม จ.นนทบุรี	40,000
1002	จินตนาพาณิชย์	115 หมู่ 3 ถ.นครอินทร์ จ. นนทบุรี	30,000
1003	โอฬารซีแพค	39 หมู่ 1 ถ. ติวานนท์ จ.นนทบุรี	40,000
1004	รุ่งเรืองการช่าง	72 หมู่ 5 ถ.แจ้งวัฒนะ จ.นนทบุรี	30,000

order

Order_no	Cust_no	Date_order	Prod_no
2201	1001	12/03/64	T-100
2202	1002	14/03/64	M-006
2203	1003	15/03/64	T-106
2204	1003	16/03/64	T-105

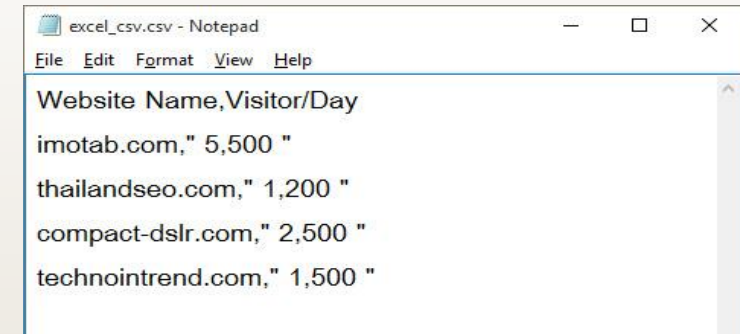
product

Prod_no	Prod_name	Prod_priceperunit
T-100	ท่อน้ำพิวซี 1 นิ้ว	45
T-105	ท่อน้ำพิวซี 2 นิ้ว	85
T-106	ท่อน้ำพิวซี 2 ½ นิ้ว	100
M-006	ถังน้ำ 100 ลิตร	350

➤ 2. ข้อมูลกึ่งโครงสร้าง (semi-structured data) ข้อมูลที่ถูกจัดเก็บอย่างมีรูปแบบในระดับหนึ่ง แต่ไม่ได้กำหนดอย่างชัดเจนนัก

➤ มีการกำหนดแท็ก (tag) หรือเครื่องหมายพิเศษ เพื่อแบ่งแยกรายละเอียด ใช้ค้นหาข้อมูลได้

➤ ซีเอสวี (Comma Separated Value: CSV)



```
excel_csv.csv - Notepad
File Edit Format View Help
Website Name,Visitor/Day
imotab.com," 5,500 "
thailandseo.com," 1,200 "
compact-dslr.com," 2,500 "
technointrend.com," 1,500 "
```

➤ เอกซ์เอ็มแอล (eXtensible Markup Languages: XML)

```
<?xml version="1.0" encoding="UTF-8"?>
- <memberdata>
  <id>001</id>
  <first-name>สุมาลี</first-name>
  <last-name>กิจเจริญ</last-name>
  <address>นนทบุรี</address>
  <email>sumalee@gmail.com</email>
  <mobile>0886732299</mobile>
</memberdata>
```



➤ เจสัน (JavaScript Object Notation: JSON)

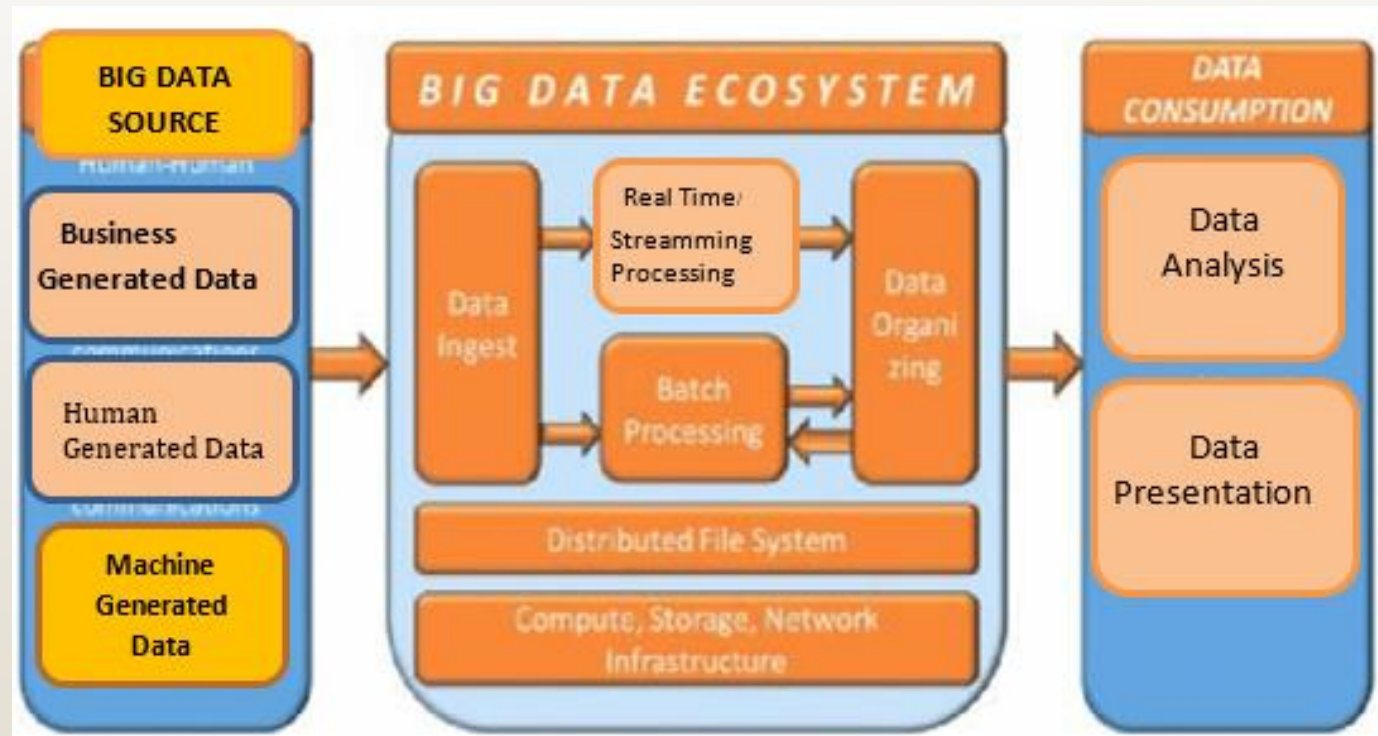
```
{ "plastic products": [  
  { "code" : "001","name":"basket","price":"20"},  
  { "code" : "002","name":"bowl","price":"15"},  
  { "code" : "003","name":"bottle","price":"25"},  
  { "code" : "004","name":"plate","price":"15"}  
]}
```

➤ 3. ข้อมูลที่ไม่มีโครงสร้าง (unstructured data) ข้อมูลที่มีโครงสร้างไม่ชัดเจน หรือไม่สามารถระบุโครงสร้างข้อมูลได้

➤ ข้อความ ภาพ เสียง วิดีโอ sensor data

# สถาปัตยกรรมของข้อมูลขนาดใหญ่

- big data architecture แบ่งเป็น 3 ส่วนใหญ่



## ➡ 1. แหล่งข้อมูลขนาดใหญ่ (big data source)

➡ ข้อมูลที่เกิดจากการทำธุรกิจหรือองค์กร

(business/organization generated data)

➡ ข้อมูลที่เกิดจากคน

(human generated data)

➡ ข้อมูลที่เกิดจากจากเครื่องจักร หรืออุปกรณ์อัตโนมัติต่าง ๆ

(machine generated data)



## ➡ 2. ระบบนิเวศของข้อมูลขนาดใหญ่ (big data ecosystem)

➡ โครงสร้างพื้นฐาน (Infrastructure)

➡ ระบบเพิ่มข้อมูลแบบกระจาย (distributed file system)

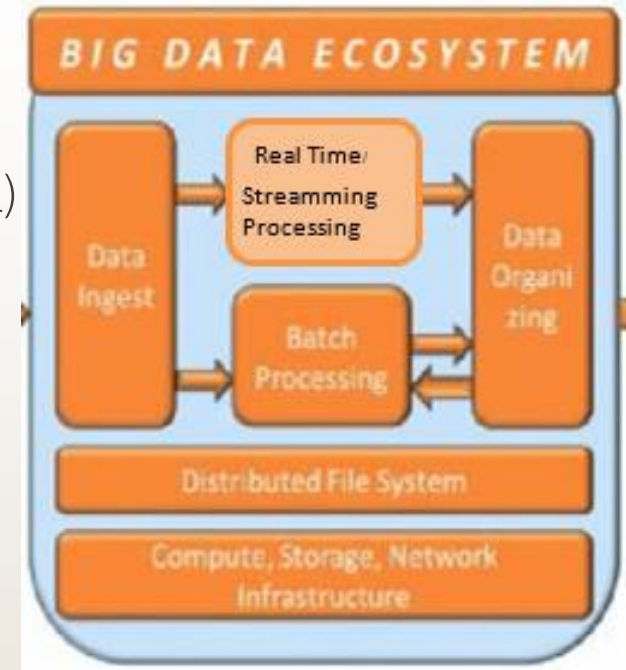
➡ เพิ่มข้อมูลกระจายแบบฮาดูป


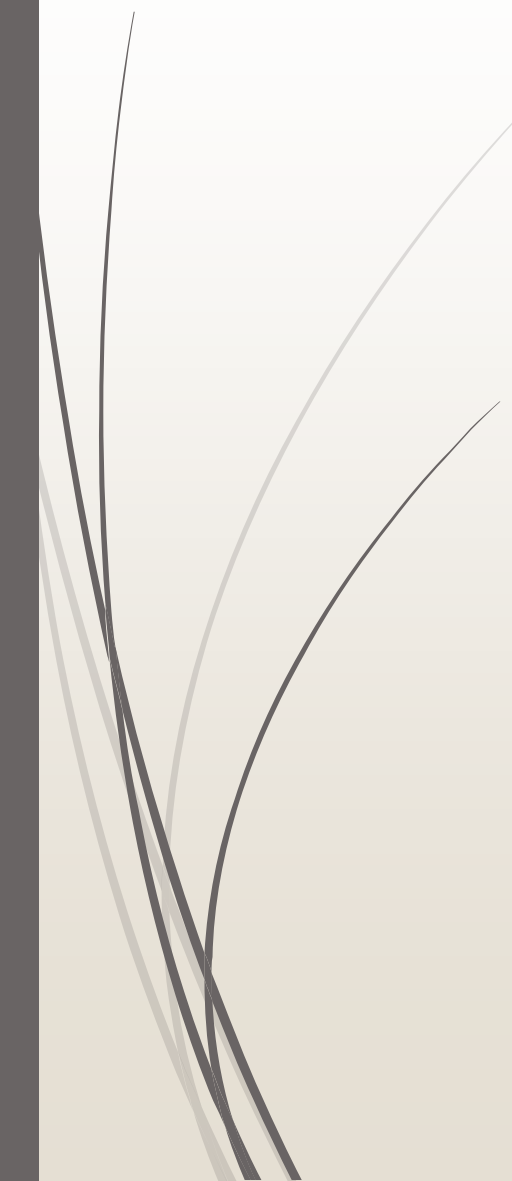
(Hadoop Distributed File System: HDFS)


➡ ส่วนรวบรวมรวมข้อมูลหรือดาตาอินเจสชัน

(Data Ingestion)

➡ ทำหน้าที่รับข้อมูลจากแหล่งข้อมูลต้นทาง เพื่อจัดเก็บที่ระบบเพิ่มข้อมูลแบบกระจายฮาดูป (HDFS) เครื่องมือ เช่น Sqoop, Flume



- 
- 
- ส่วนประมวลผลข้อมูล (data processing)
    - การประมวลผลแบบแบตช์ (batch processing) เช่น Apache Hadoop, MapReduce-M/R
    - การประมวลผลแบบเรียลไทม์/สตรีมมิง (real-time/streaming processing) เช่น อาปาเช่ คอฟฟา (Apache Kafka) อาปาเช่ สโตรม (Apache Storm)
  - ส่วนจัดการข้อมูล (data organizing) การจัดการข้อมูลให้นำมาใช้งานได้ง่าย มีความเกี่ยวข้องกับฐานข้อมูลโนเอสคิวแอล (NoSQL Database)



➡ **3. การนำข้อมูลขนาดใหญ่ไปใช้งาน** (data consumption) ค้นหารูปแบบความสัมพันธ์ของข้อมูล และสิ่งที่เชื่อมโยงกัน เช่น แนวโน้มทางการตลาด ความต้องการของลูกค้า

➡ การวิเคราะห์แบบบรรยายหรือพรรณนา (descriptive analytics)

➡ การวิเคราะห์แบบวินิจฉัย (diagnostic analytics)

➡ การวิเคราะห์แบบทำนายหรือพยากรณ์ (predictive analytics)

➡ การวิเคราะห์แบบกำหนดทางเลือกในการตัดสินใจ (prescriptive analytics)

➡ เทคนิคในการวิเคราะห์ เช่น สถิติ ,เหมืองข้อมูล, Machine Learning, Text Mining

➡ ส่วนของการนำเสนอข้อมูล (data presentation) การทำ data visualization

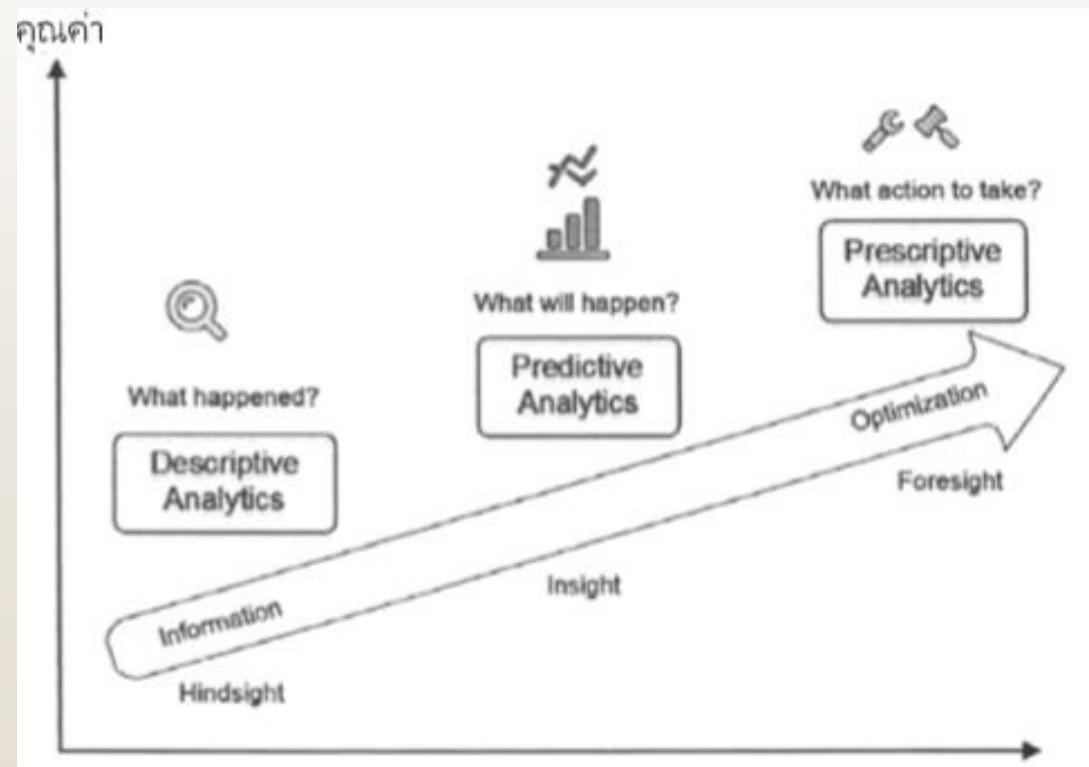
# หลักการและวิธีการวิเคราะห์ข้อมูลขนาดใหญ่


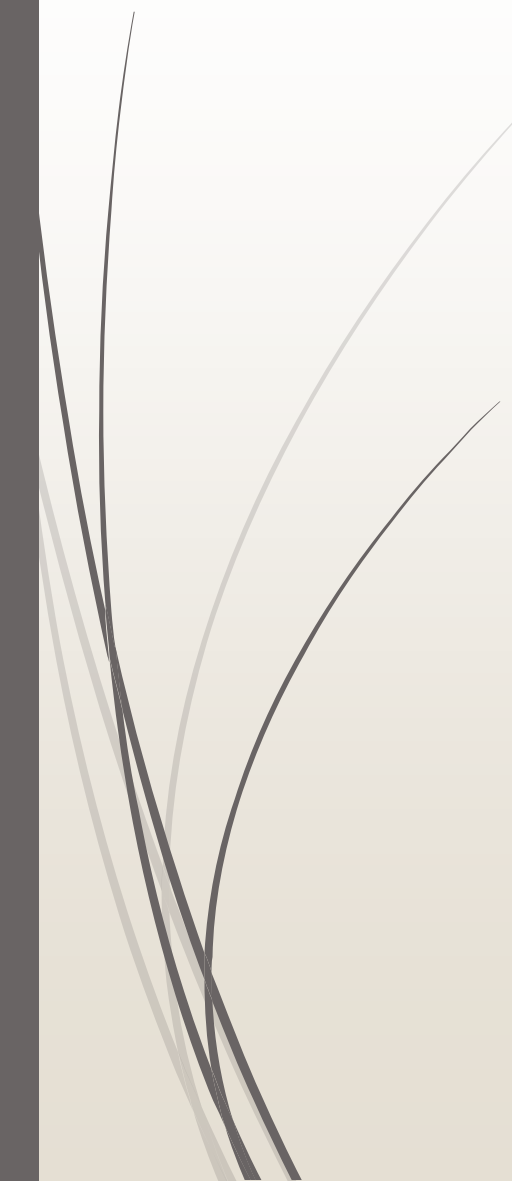
➡ รูปแบบของการวิเคราะห์ข้อมูลขนาดใหญ่ แบ่งเป็น 3 ลักษณะ

➡ การวิเคราะห์เชิงพรรณนา

➡ การวิเคราะห์เชิงพยากรณ์

➡ การวิเคราะห์เชิงแนะนำ





- 
- 
- **1. การวิเคราะห์เชิงพรรณนา (descriptive analytics)** วิเคราะห์ข้อมูลเหตุการณ์หรือปัญหาที่ผ่านมาเพื่อสร้างความเข้าใจภายหลังที่เหตุการณ์นั้นได้เกิดไปแล้ว (hindsight) - สารสนเทศ
  - **2. การวิเคราะห์เชิงพยากรณ์ (predictive analytics)** นำข้อมูลในอดีตมาวิเคราะห์ เพื่อสร้างโมเดลทำนายสิ่งที่จะเกิดขึ้นในอนาคต เพื่อการหยั่งรู้อย่างลึกซึ้ง (insight)
  - **3. การวิเคราะห์เชิงแนะนำ (prescriptive analytics)** รวมความสามารถการวิเคราะห์เชิงพรรณนาและเชิงพยากรณ์ไว้และเพิ่มให้คำแนะนำเกี่ยวกับแนวทางและวิธีปฏิบัติที่เหมาะสมที่สุด (optimization) ในแต่ละแนวทาง




# เทคนิคในการวิเคราะห์ข้อมูลขนาดใหญ่



- **เทคนิคในเชิงสถิติ** (statistical technique) เป็นวิธีพื้นฐานในการวิเคราะห์
  - สถิติเชิงพรรณนา เช่น ความถี่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน
  - สถิติเชิงอนุมาน หาค่าความสัมพันธ์ระหว่างตัวแปร เช่น การวิเคราะห์สหสัมพันธ์ การวิเคราะห์ความแปรปรวน
- **การประมวลผลภาพดิจิทัล** (digital image processing) นำภาพดิจิทัลมาประมวลผลเพื่อการจัดการรูปแบบสารสนเทศของภาพ
  - การรู้จำใบหน้า ลายนิ้วมือ ชิ้นส่วนอุปกรณ์

- 
- 
- **การวิจัยดำเนินการ** (operation research) สร้างแบบจำลองเชิงคณิตศาสตร์ และขั้นตอนช่วยในการตัดสินใจ
    - สถิติ การหาค่าเหมาะสมที่สุด สโตแคสติก (stochastic) ทฤษฎีเกม ทฤษฎีแถวคอย การจำลอง
  - **การประมวลผลภาษาธรรมชาติ** (natural language processing) นำภาษามนุษย์มาใช้งาน
    - การสังเคราะห์เสียง (speech synthesis) การแปลงข้อความให้เป็นเสียง การรู้จำเสียง
  - **การทำvisualไลเซชัน** (visualization) การสร้างภาพ แผนผัง ภาพเคลื่อนไหวแทนข้อความ
    - การพล็อตแบบกระจาย (scatter plot) การแสดงส่วนร่วมกัน (intersection)
  - **เทคนิคการเรียนรู้ของเครื่อง** (machine learning technique) สร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูล และทำนายข้อมูลได้



# การเรียนรู้ของเครื่อง

- อัลกอริทึมแบ่งเป็น 3 ประเภท
  - 1. การเรียนรู้แบบมีผู้สอน (supervised learning)
  - 2. การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning)
  - 3. การเรียนรู้แบบเสริมแรง (reinforcement learning)
- 

- 
- 
- **1. การเรียนรู้แบบมีผู้สอน** สร้างโมเดลในการทำนายโดยใช้ทั้งข้อมูลนำเข้า และข้อมูลผลลัพธ์
    - เปรียบได้กับการเรียนรู้แบบมีครูฝึกทำหน้าที่ให้คำตอบ
    - มีข้อมูลฝึกหัด (training data) มีผลเฉลย หรือคลาสที่เป็นคำตอบที่มีการติดเลเบล (label) ไว้
    - ทดสอบโมเดล โดยมีข้อมูลทดสอบ (test data) เพื่อวัดประสิทธิภาพของโมเดล
    - มักเป็นโมเดลที่ใช้จำแนก (classification model) มีลักษณะข้อมูลแบบไม่ต่อเนื่อง (discrete data)  
เช่น จำแนกสเปกเมมเมลจากเมลปกติ



➡ **2. การเรียนรู้แบบไม่มีผู้สอน** การจัดกลุ่มและแปลผลข้อมูลโดยอาศัยข้อมูลเข้าเพียงอย่างเดียว

➡ ไม่มีการระบุผลลัพธ์ที่ต้องการ (target) ไว้ก่อน

➡ จัดข้อมูลนำเข้าเป็นกลุ่ม (cluster) บนพื้นฐานของความเหมือน และความแตกต่างของรูปแบบข้อมูล

➡ เช่น การลดมิติข้อมูล (dimensionality reduction) การจัดกลุ่มข้อมูล (clustering)

➡ **3. การเรียนรู้แบบเสริมแรง** การเรียนรู้ต่อเหตุการณ์ที่เกิด โดยผู้เรียน (agent) สังเกตเหตุการณ์ เพื่อตัดสินใจการกระทำตามนโยบาย (policy) ที่ตั้งไว้

➡ หลังจากได้กระทำ (action) ผู้เรียนจะได้ รางวัล หรือ การลงโทษ

➡ ผลตอบแทนจะถูกนำไปปรับปรุงนโยบายที่ดีขึ้น ทำให้เกิดการเรียนรู้

➡ กระบวนการปรับปรุงทำในลักษณะวนซ้ำ (iteration) จนได้นโยบายที่เหมาะสมที่สุด

ตารางจำแนกประเภทอัลกอริธึมแบบ Supervised และ Unsupervised learning

	Supervised Learning	Unsupervised Learning
Continuous	<p>รีเกรสชัน (Regression)</p> <ul style="list-style-type: none"><li>• การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression)</li><li>• การวิเคราะห์การถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression)</li><li>• etc.</li></ul>	<p>การจัดกลุ่ม (clustering) และการลดมิติข้อมูล (Dimensionality Reduction)</p> <ul style="list-style-type: none"><li>• K-Mean</li><li>• พีซีเอ (Principle Component Analysis)</li><li>• เอสวีดี (Singular Value Decomposition)</li><li>• etc.</li></ul>
Categorical	<p>การจำแนก (Classification)</p> <ul style="list-style-type: none"><li>• KNN (K-nearest Neighbors)</li><li>• นาอ์ฟเบย์</li><li>• เอสวีเอ็ม (Support Vector Machine)</li><li>• etc.</li></ul>	<p>การวิเคราะห์ความสัมพันธ์ (Association Analysis)</p> <ul style="list-style-type: none"><li>• อาพริออไร (Apriori)</li><li>• เอฟพี-โกรว์ธ (FP-Growth)</li><li>• etc.</li></ul>

