

Master of Science Program in  
Artificial Intelligence for Business Analytics

# Data Science

Week 2: CRISP-DM & Altair AI Studio

Eakasit Pacharawongsakda, Ph.D.

eakasit@datacubeth.ai

Co-Founder and Data Science Team Lead

Cube Analytics Consulting Co., Ltd.



## Introduction

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ตัวอย่างการใช้ Big Data Analytics ในชีวิตประจำวัน

- ลักษณะของข้อมูล

### การจัดเก็บข้อมูล (Data Management)

- ฐานข้อมูล (Database)

#### คลังข้อมูล (Data Warehouse)

- การจัดการข้อมูลขนาดใหญ่ (Big Data)

- ตัวอย่างการประยุกต์ใช้การวิเคราะห์ข้อมูลด้วยเทคโนโลยี Data Science

# Data Storage: Data Warehouse

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- การเก็บข้อมูลในฐานข้อมูลมีข้อดีหลายอย่าง เช่น
  - ทำให้แต่ละฝ่ายหรือแผนก **มองเห็นข้อมูลที่ถูกต้องตรงกัน**
- ฐานข้อมูลส่วนใหญ่จะ **เก็บไว้แยกจากกัน** เป็นแต่ละฝ่ายหรือแผนก เพื่อใช้ตอบโจทย์ของส่วนต้นเอง
- ด้านผู้บริหาร **ต้องการดูภาพรวมของกิจกรรม** จำเป็นต้องดึงข้อมูลจากฐานข้อมูลต่างๆ เหล่านี้มาสรุปให้เห็นเป็น Dashboard หรือ รายงานต่างๆ
  - ค่อนข้างจะ **ใช้เวลานาน**
  - ข้อมูลแยกเป็นหลายๆ ตาราง การ **join ข้อมูลข้ามตารางจะทำงานช้า**
  - ข้อมูลที่มี **ความหลากหลาย** อาจจะเสียบไม่ตรงกับ
- มี concept ของ **คลังข้อมูล (data warehouse)** ที่ **รวบรวมข้อมูล** จากหลายๆ ฐานข้อมูลรวมกัน

# Data Storage: Data Warehouse

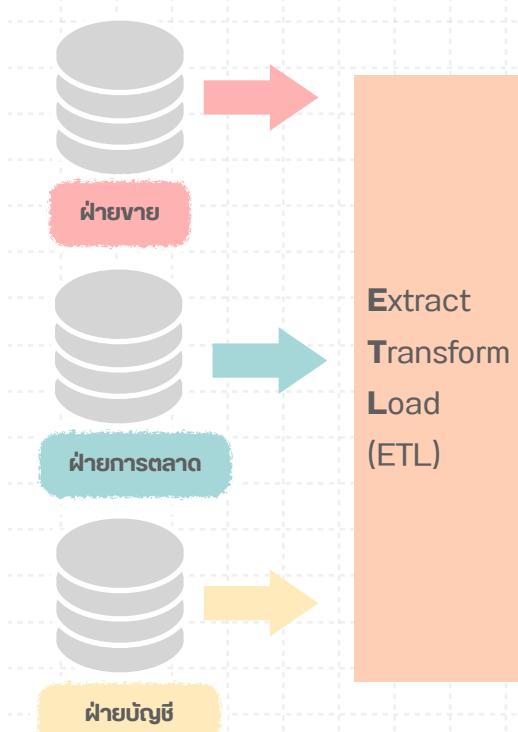
สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



- การเก็บข้อมูลในฐานข้อมูลมีข้อดีหลายอย่างทำให้แต่ละฝ่ายหรือแผนกมองเห็นข้อมูลที่ถูกต้องตรงกัน
- ฐานข้อมูลส่วนใหญ่จะ **เก็บไว้แยกจากกัน** เป็นแต่ละฝ่ายหรือแผนก เพื่อใช้ตอบโจทย์ของส่วนต้นเอง
- ด้านผู้บริหาร **ต้องการดูภาพรวมของกิจกรรม** จำเป็นต้องดึงข้อมูลจากฐานข้อมูลต่างๆ เหล่านี้มาสรุปให้เห็นเป็น Dashboard หรือ รายงานต่างๆ
  - ค่อนข้างจะ **ใช้เวลานาน**
  - ข้อมูลแยกเป็นหลายๆ ตาราง การ **join ข้อมูลข้ามตารางจะทำงานช้า**
  - ข้อมูลที่มี **ความหลากหลาย**
- มี concept ของ **คลังข้อมูล (data warehouse)** ที่ **รวบรวมข้อมูล** จากหลายๆ ฐานข้อมูลรวมกัน

# Data Storage: Data Warehouse

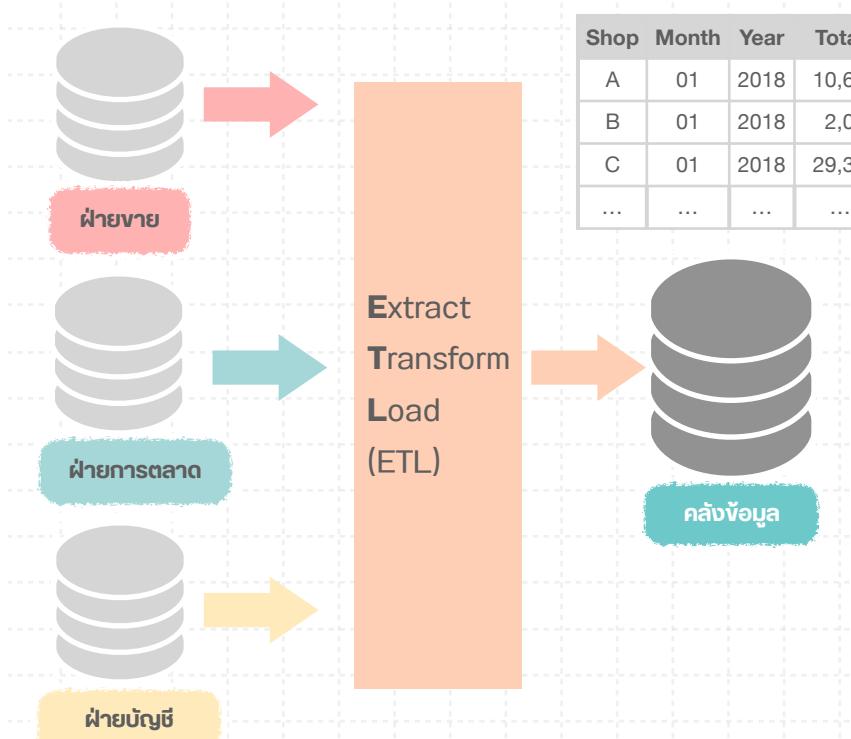
สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



- กระบวนการ ETL เป็น **การดึงข้อมูล (Extract)**, **แปลงข้อมูล (Transform)** และ **บันทึกข้อมูล (Load)**
- เป็นการดึงข้อมูล (Extract) **เฉพาะด้านที่สนใจ** จากฐานข้อมูลต่างๆ marrow กัน
- แปลงข้อมูลให้อยู่ในรูปแบบเดียวกัน** เช่น
  - แปลงข้อมูลเงินเดือนลักษณะให้เป็นเงินบาท
  - แปลงข้อมูลตัวย่อ เช่น F และ M ให้กลายเป็น Female และ Male
- บันทึกข้อมูล (Load) ลงไปในคลังข้อมูล (data warehouse)

# Data Storage: Data Warehouse

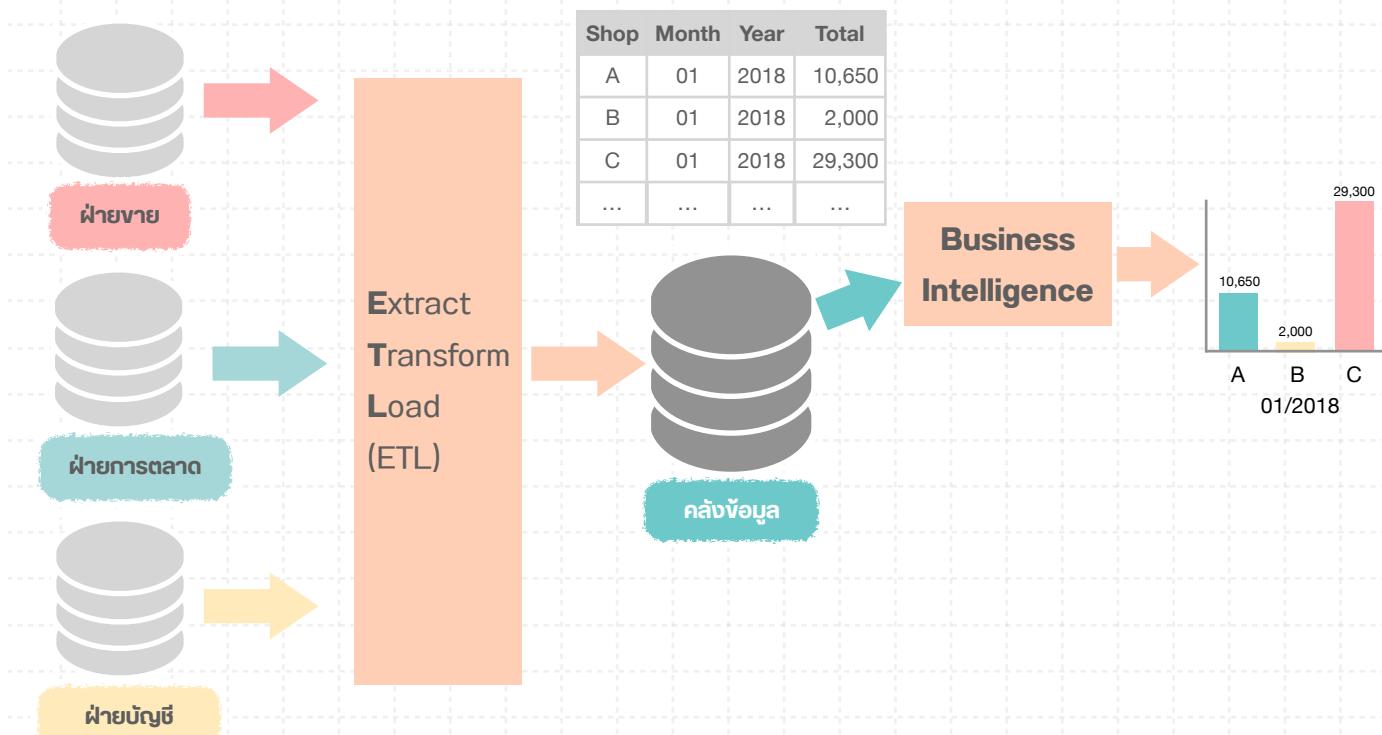
สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



- ข้อมูลที่เก็บไว้ในคลังข้อมูล มักจะ **อยู่ในรูปแบบของ การสรุป (summary)**
- ข้อมูลออกมานะ เช่น
  - ยอดขายในแต่ละเดือน
  - ยอดขายในแต่ละสาขา

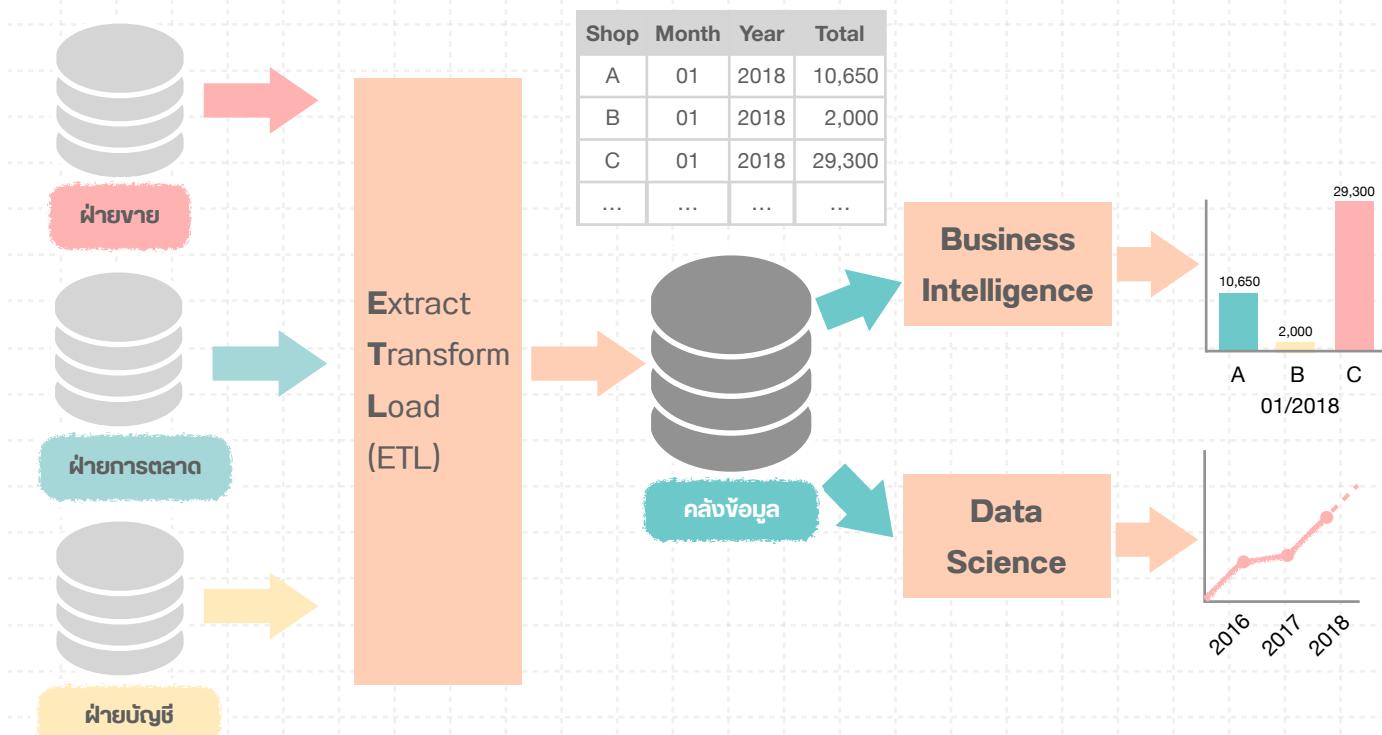
# Data Storage: Data Warehouse

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL



# Data Storage: Data Warehouse

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL



# Data Storage: Data Warehouse

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

## • Database

- **ฐานข้อมูลใช้ในการจัดเก็บข้อมูล** ลดความซ้ำซ้อนของข้อมูล เน้นการจัดเก็บ เพื่อแก้ไข และลบข้อมูล

## • Data Warehouse

- **คลังข้อมูลรวมข้อมูลจากหลายๆ ฐานข้อมูล** แปลงข้อมูลให้มีความเหมือนกัน หมายความว่า สำหรับการเรียกดู (view) เพื่อสร้างรายงานสรุป

## • Data Science/Machine Learning

- เป็นการ**เรียนรู้จากข้อมูลในอดีตเพื่อนำมาสร้างแบบจำลอง** สำหรับใช้ในการพยากรณ์เหตุการณ์ในอนาคต

# Introduction

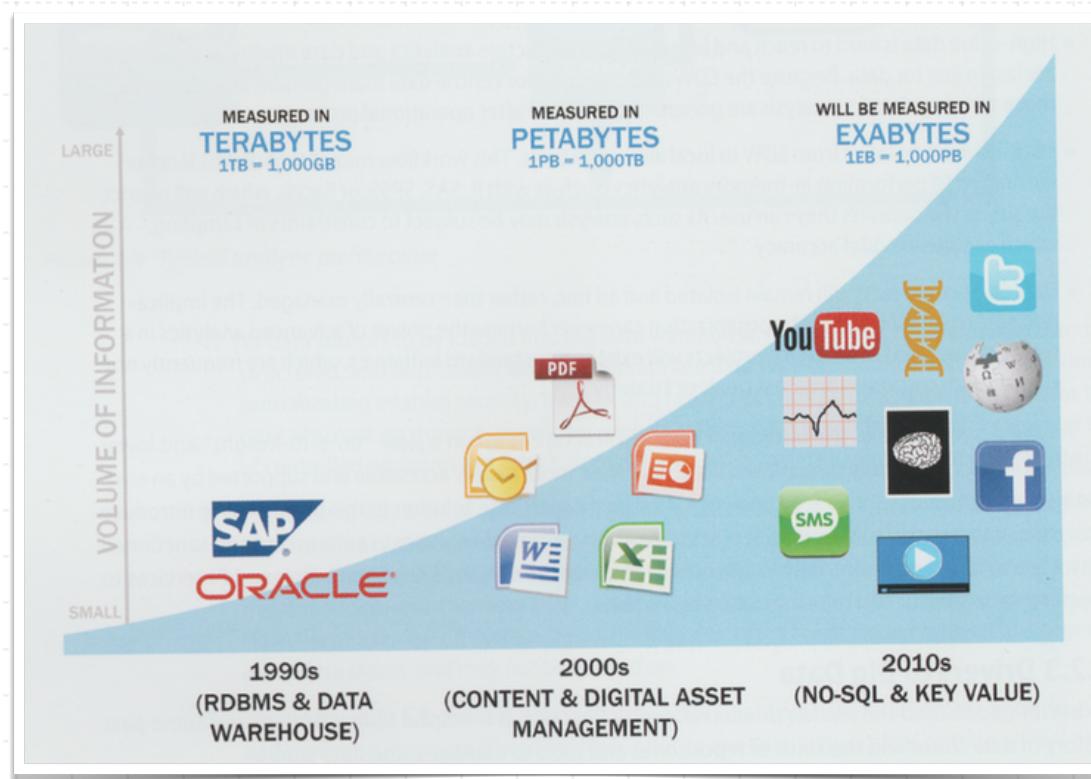
สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

## • ตัวอย่างการใช้ Big Data Analytics ในชีวิตประจำวัน

- ลักษณะของข้อมูล
- การจัดเก็บข้อมูล (Data Management)
  - ฐานข้อมูล (Database)
  - คลังข้อมูล (Data Warehouse)
- **การจัดการข้อมูลขนาดใหญ่ (Big Data)**
- ตัวอย่างการประยุกต์ใช้การวิเคราะห์ข้อมูลด้วยเทคโนโลยี Data Science

# Data Evolutions

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

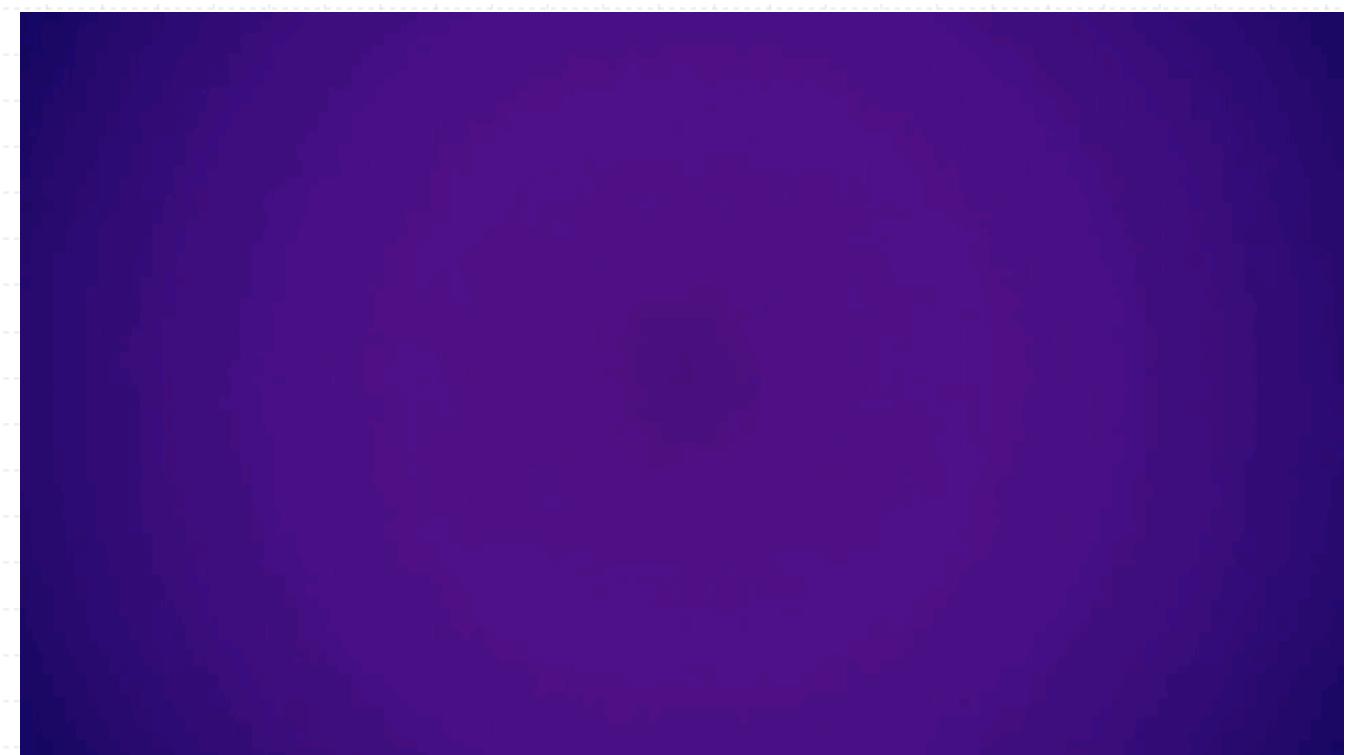


source: Data Science and Big Data Analytics: Discovering, analyzing, visualizing and presenting data [facebook.com/datacube.th](http://facebook.com/datacube.th)

11

## What is Big Data?

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



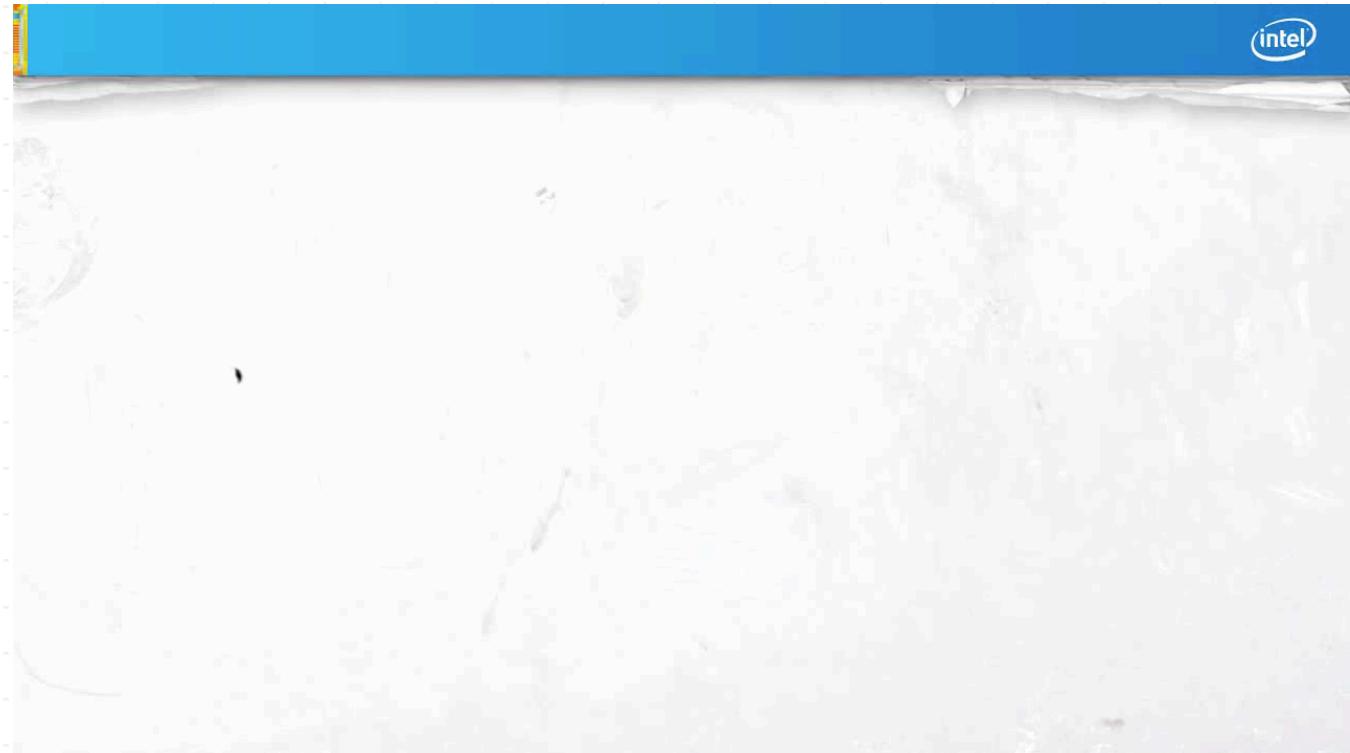
source: <https://www.youtube.com/watch?v=TzxmjbL-i4Y>

<http://facebook.com/datacube.th>

12

# What is Big Data?

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



source: <http://www.intel.com/content/www/us/en/big-data/big-data-101-animation.html#>

<http://facebook.com/datacube.th>

13

# What is Big Data?

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- Big Data ประกอบด้วย 3 V

- **Volume**

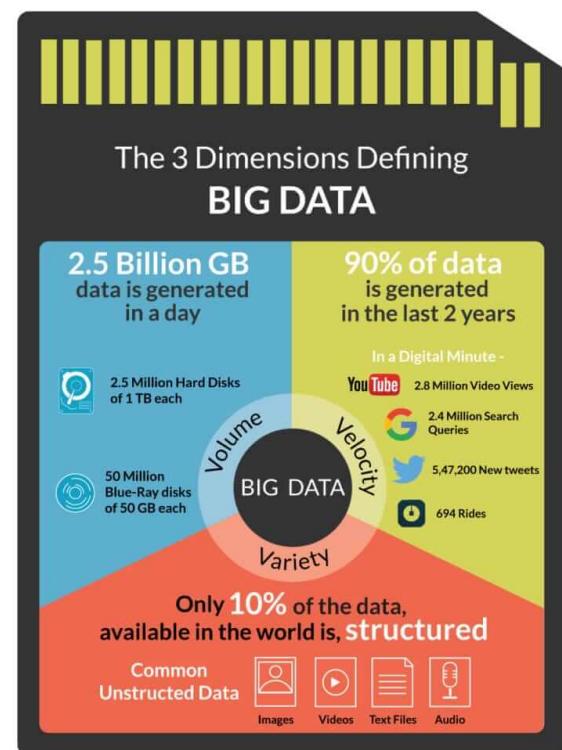
- ข้อมูลมีจำนวนเพิ่มขึ้นอย่างมหาศาล

- **Velocity**

- ข้อมูลเพิ่มขึ้นอย่างรวดเร็ว

- **Variety**

- ข้อมูลมีความหลากหลายมากขึ้น



source: <https://upxacademy.com/beginners-guide-to-big-data/>

<http://facebook.com/datacube.th>

14

# What is Big Data?

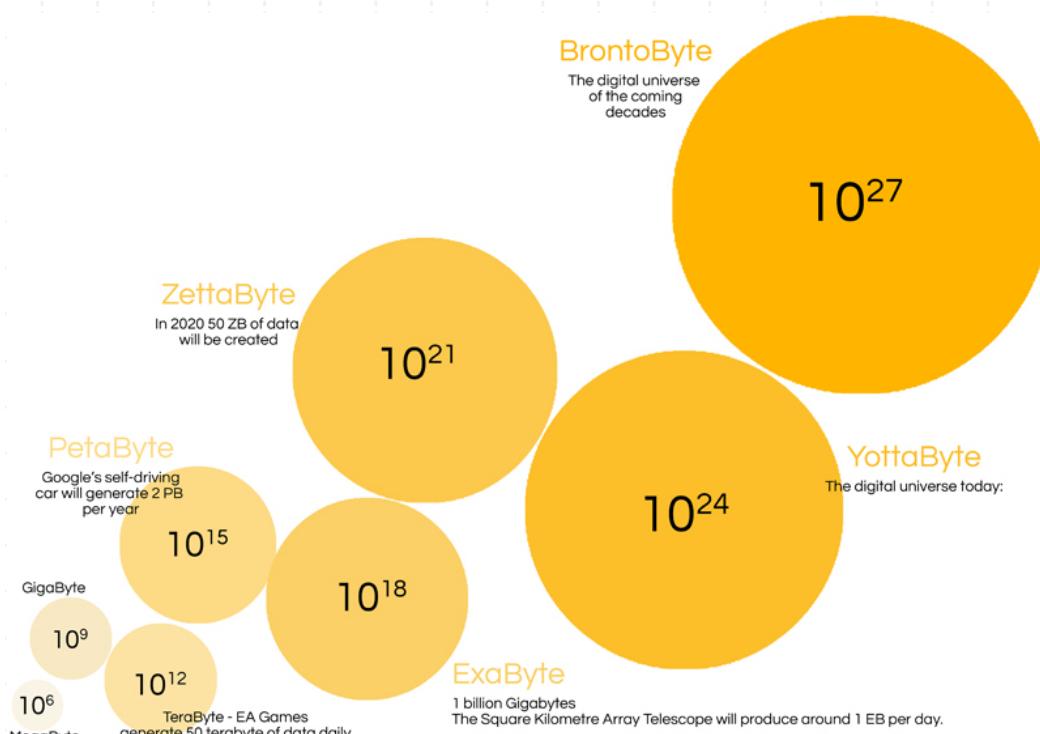
สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- Huge volume of data**

- ข้อมูลมีขนาดใหญ่มากๆ เช่น มีจำนวนเป็นพันล้านแถว (billion row) หรือ เป็นล้านคอลัมน์ (million columns)

## Big Data: Volume

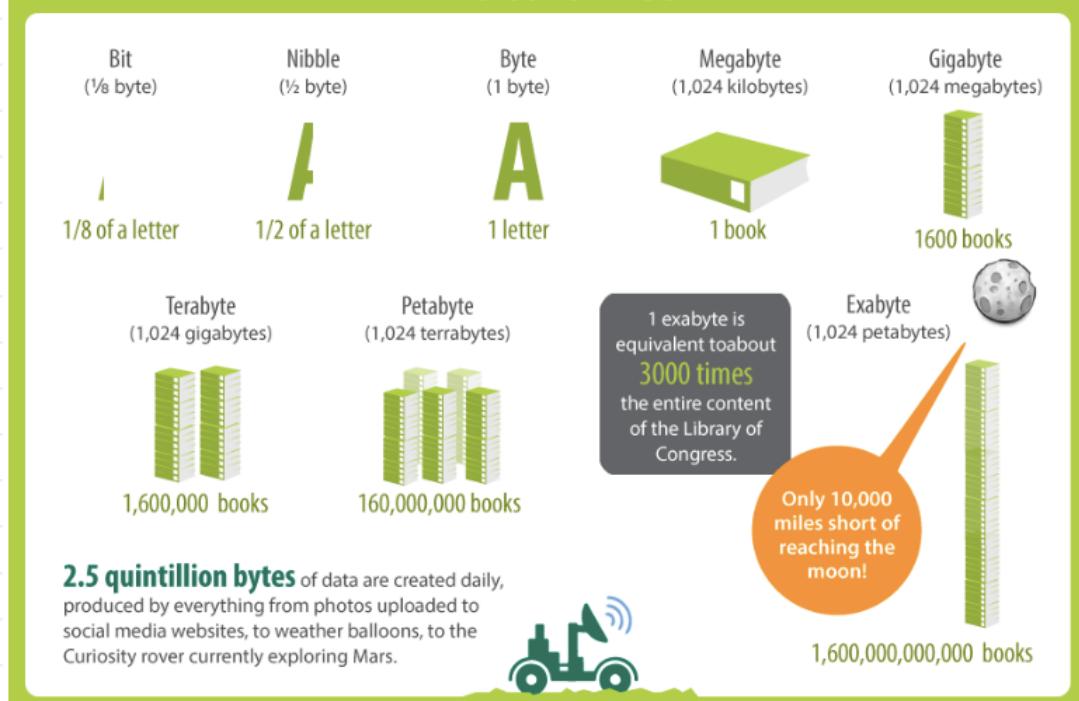
สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL



# Big Data: Volume

สำหรับการสูงสุด M: Sc. in Artificial Intelligence for Business Analytics, KMITL

## Data Sizes



source: <https://www.adeptia.com>  
Copyright © 2020 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

17

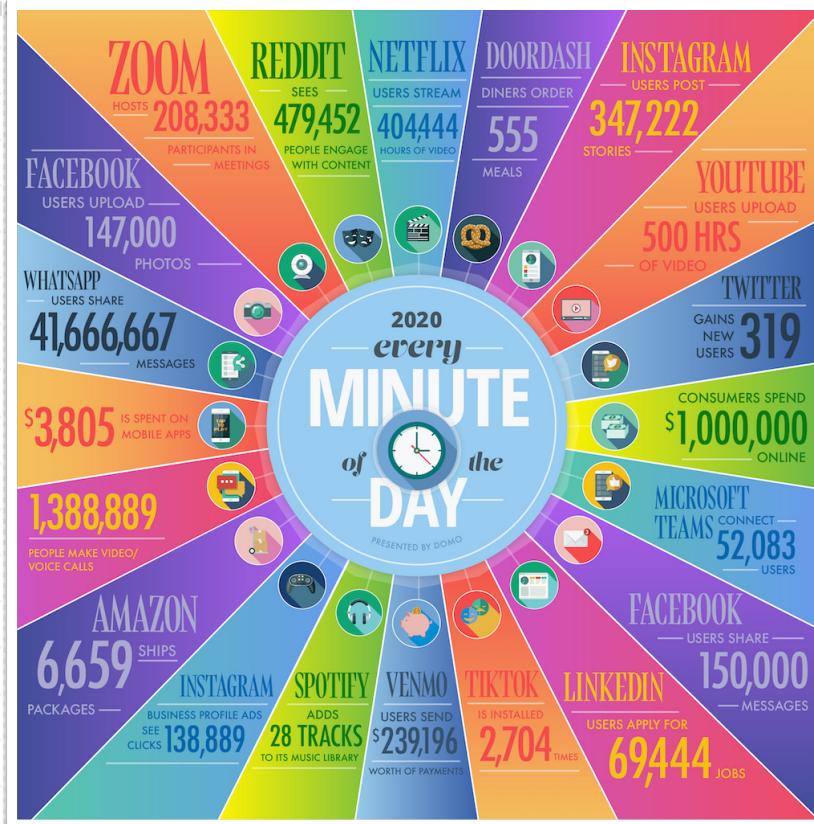
# What is Big Data?

สำหรับการสูงสุด M: Sc. in Artificial Intelligence for Business Analytics, KMITL

### Huge volume of data

- ข้อมูลมีขนาดใหญ่มากๆ เช่น มีจำนวนเป็นพันล้านແడວ (billion row) หรือ เป็นล้านคอลัมน์ (million columns)
- **Speed of new data creation and growth**
  - ข้อมูลเกิดขึ้นอย่างรวดเร็วมากๆ

# Big Data: Velocity



source: <https://web-assets.domo.com/blog/wp-content/uploads/2020/08/20-data-never-sleeps-8-final-01-Resize.jpg>

19

## What is Big Data?

สำหรับนิสิตสายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

### Huge volume of data

- ข้อมูลมีขนาดใหญ่มากๆ เช่น มีจำนวนเป็นพันล้านแถว (billion row) หรือ เป็นล้านคอลัมน์ (million columns)

### Speed of new data creation and growth

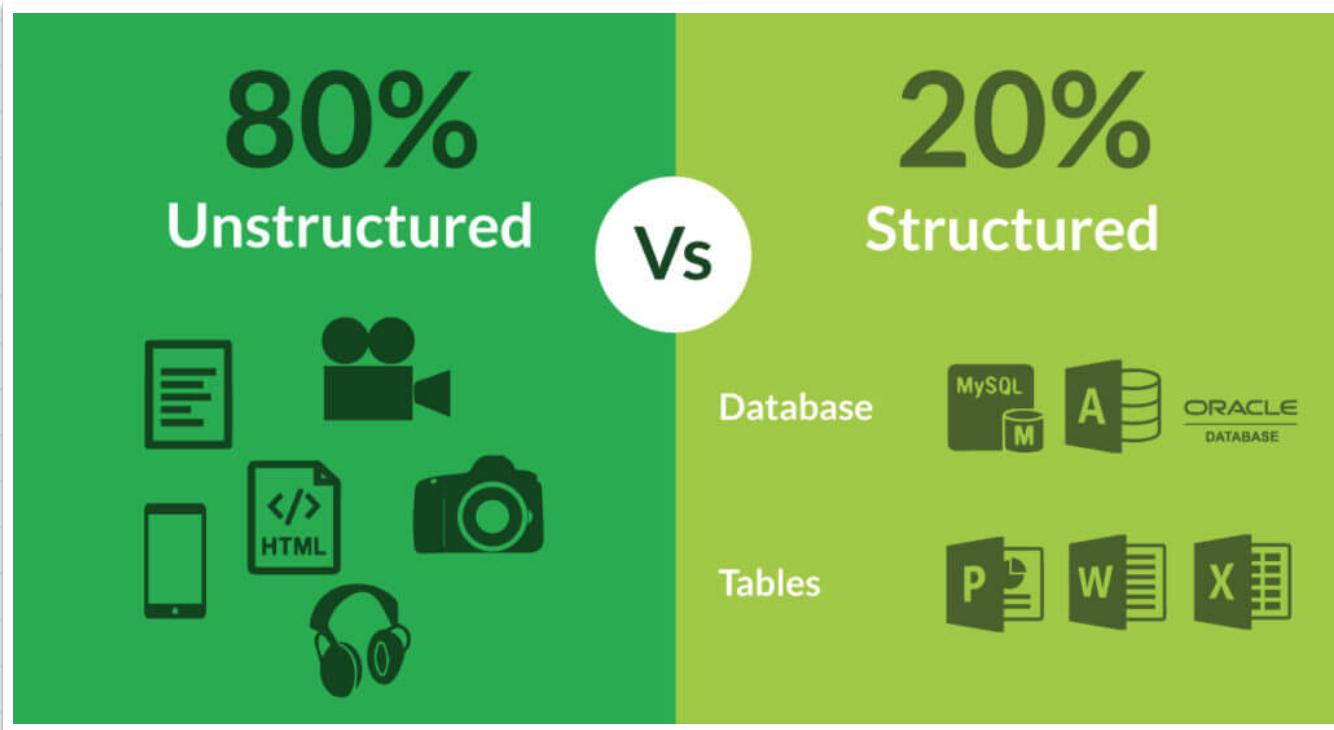
- ข้อมูลเกิดขึ้นอย่างรวดเร็วมากๆ

### Complexity of data types and structures

- ข้อมูลมีความหลากหลาย ไม่ได้อยู่ในรูปแบบของตารางเท่านั้น อาจจะเป็นรูปแบบของข้อความ (text) รูปภาพ (images) หรือ วิดีโอ (video clip)

# Big Data: Variety

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



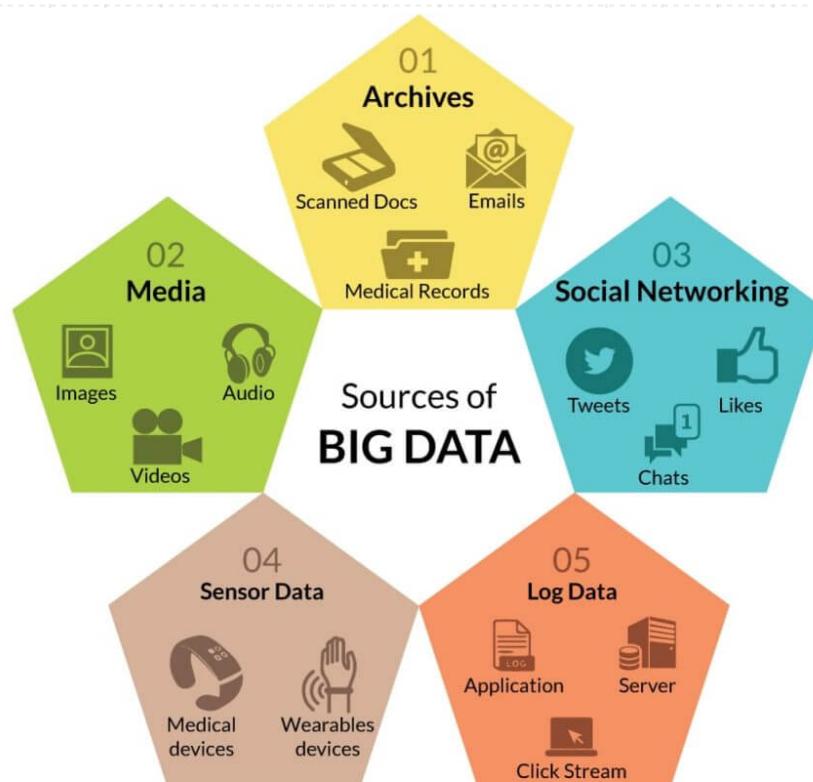
source: <https://upxacademy.com/beginners-guide-to-big-data/>

<http://facebook.com/datacube.th>

21

# Big Data: Variety

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



source: <https://upxacademy.com/beginners-guide-to-big-data/>

<http://facebook.com/datacube.th>

22

# What is Big Data?

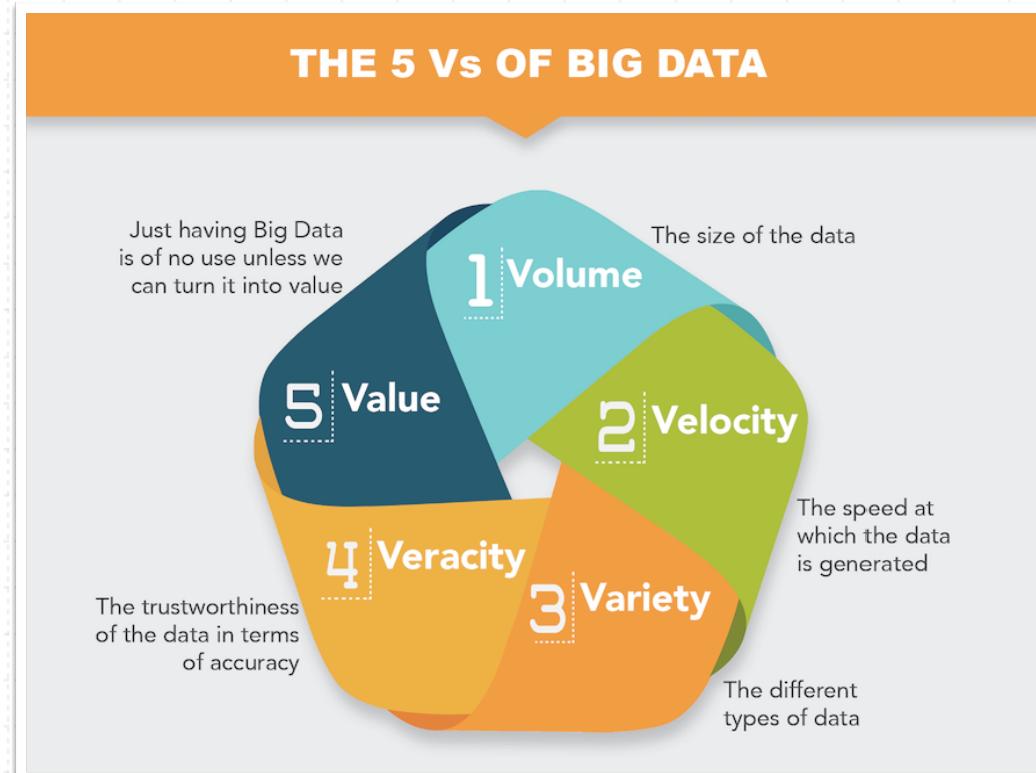
สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

ดังนั้น Big Data จึงหมายความว่า

“ข้อมูลที่มี**หลากหลายรูปแบบ**  
ซึ่งก่อตัวขึ้นอย่าง**รวดเร็ว**จนกลายเป็น  
**ข้อมูลที่มีขนาดมหาศาล**”

# What is Big Data?

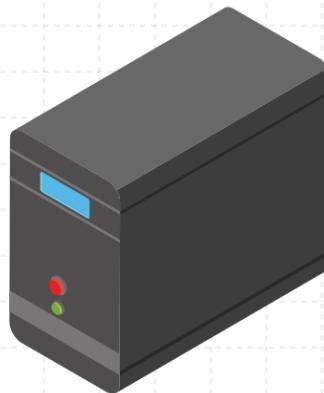
สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL



# Scale Up v.s. Scale Out

สำหรับการอธิบายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

Scale Up



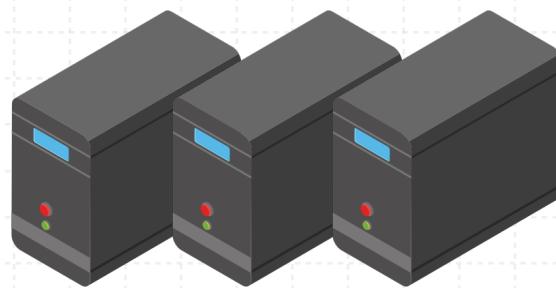
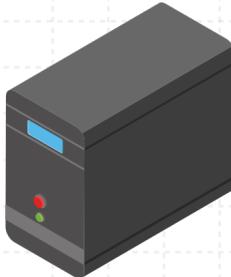
# Scale Up v.s. Scale Out

สำหรับการอธิบายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

Scale Up



Scale Out



# Big Data Technology

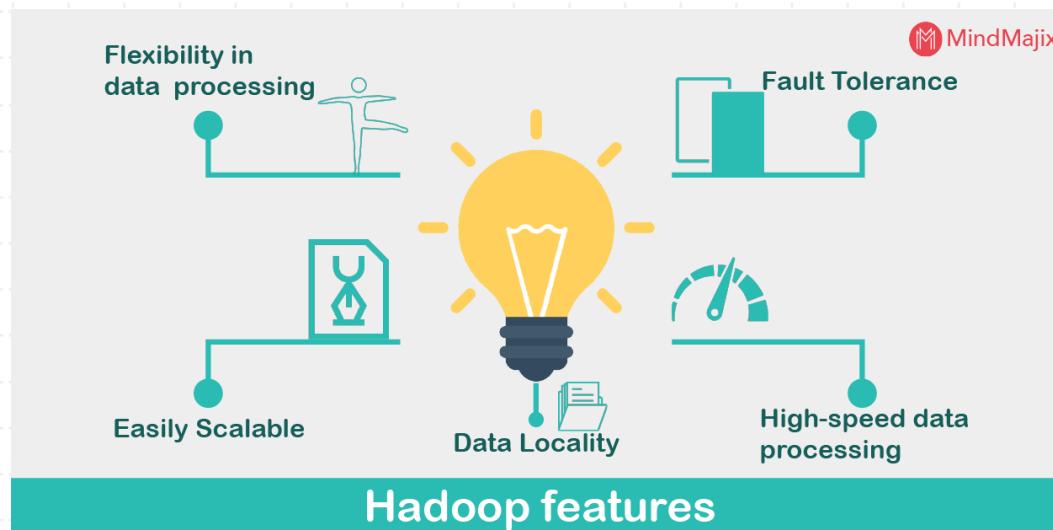
- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- เนื่องจากข้อมูลมีขนาดใหญ่มากจนไม่สามารถประมวลผลด้วยเครื่องเดียว เครื่องคอมพิวเตอร์เครื่องเดียวได้จึงมีเทคโนโลยีที่ช่วยประมวลผลแบบขนาน (parallel) เกิดขึ้น เช่น

- MapReduce** เป็นเทคโนโลยีที่แบ่งข้อมูลและส่งไปประมวลผลในเครื่องคอมพิวเตอร์หลายๆ เครื่องๆ พร้อมกัน พัฒนาขึ้นจากวิศวกรรมของบริษัท Google
- Hadoop** เป็นระบบที่นำแนวคิดของ MapReduce มาพัฒนาต่อยอดร่วมกับการจัดการไฟล์แบบขนาน (**HDFS**) และการทำงานแบบขนานต่างๆ พัฒนาขึ้นจากวิศวกรรมของบริษัท Yahoo!

# Big Data Technology

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- What is Hadoop?

- Hadoop is an open source software framework which is designed to **store** the enormous volumes of data sets in a distributed way on large clusters of the commodity.

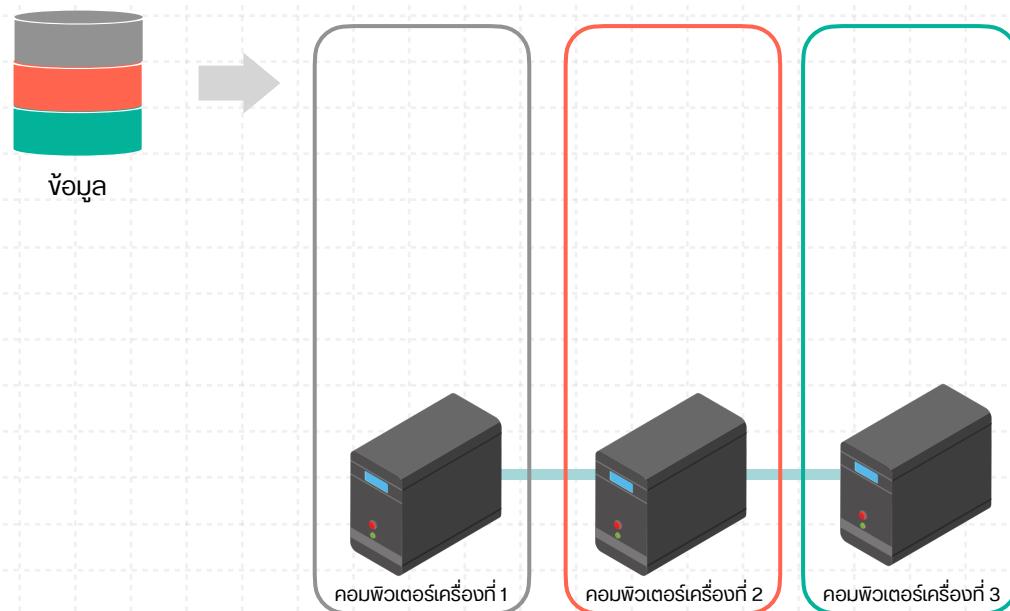


# Big Data Technology

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- How Does Hadoop Work?

ค่อนเซ็ปต์ของ Hadoop

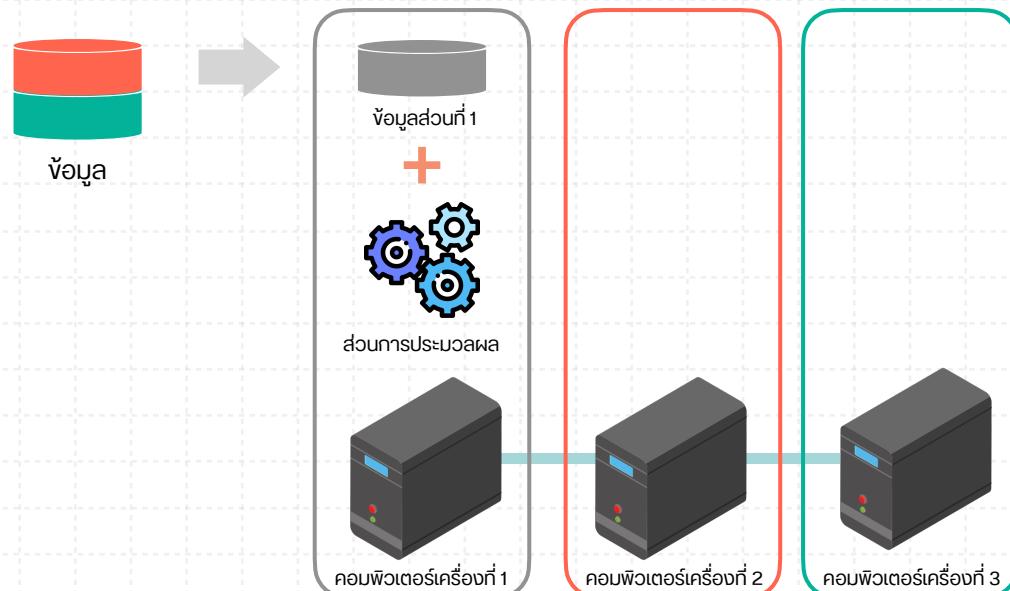


# Big Data Technology

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- How Does Hadoop Work?

ค่อนเซ็ปต์ของ Hadoop

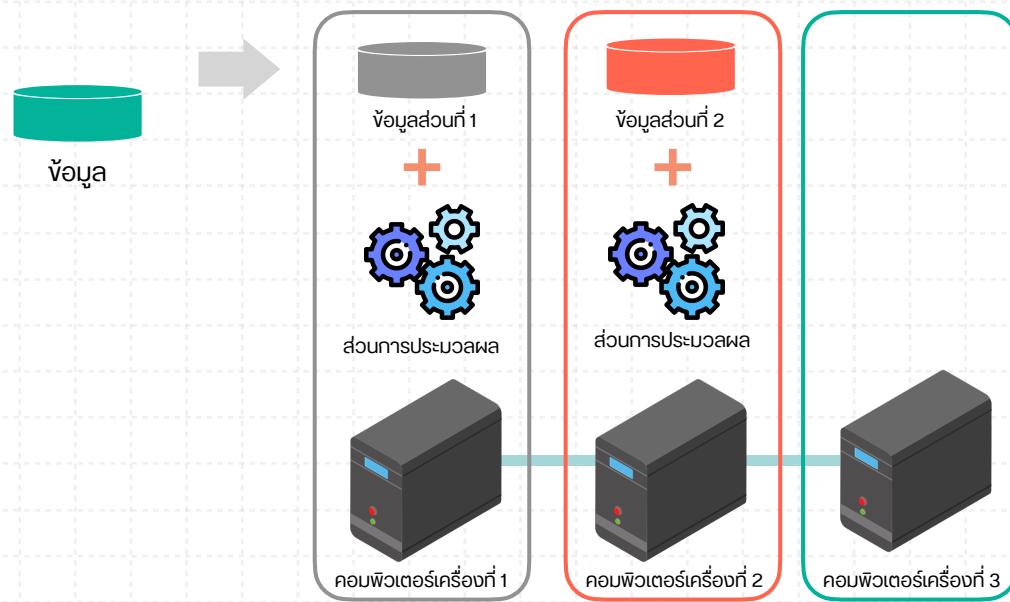


# Big Data Technology

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- How Does Hadoop Work?

ค่อนซึป์ต์ของ Hadoop

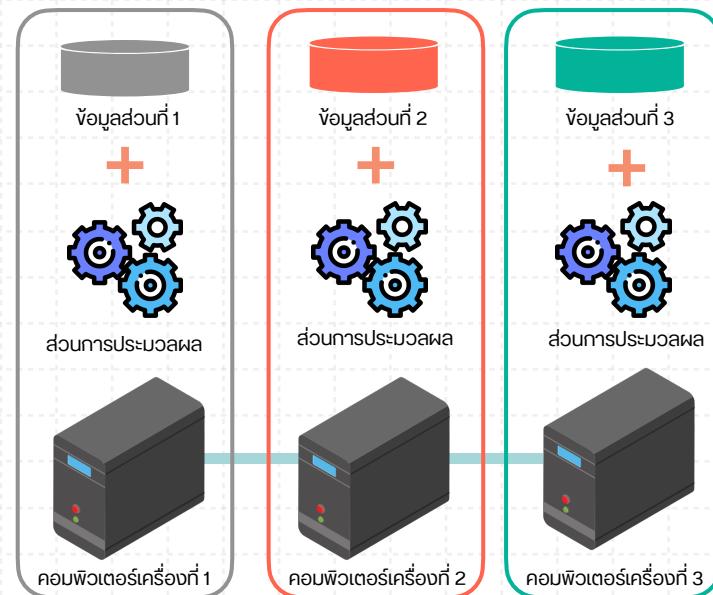


# Big Data Technology

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

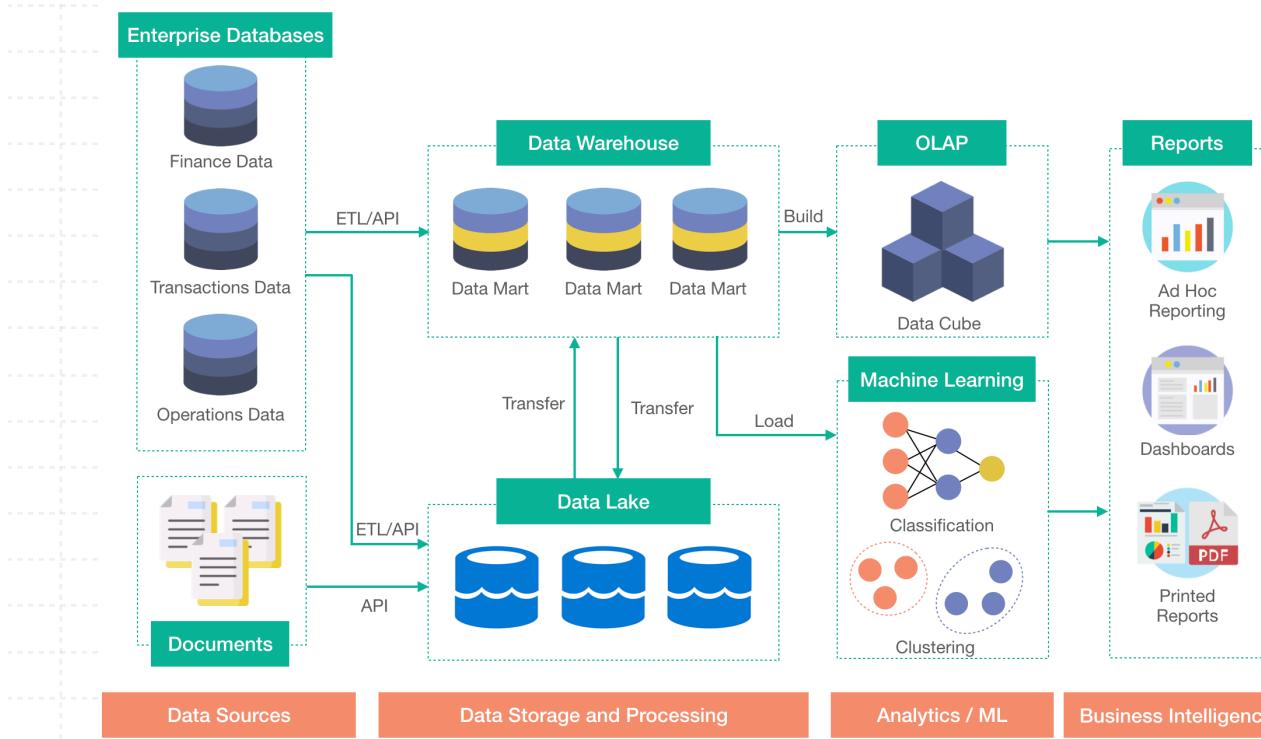
- How Does Hadoop Work?

ค่อนซึป์ต์ของ Hadoop



# Enterprise Big Data Framework

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



## Introduction

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- ตัวอย่างการใช้ Big Data Analytics ในชีวิตประจำวัน

- ลักษณะของข้อมูล
- การจัดเก็บข้อมูล (Data Management)
  - ฐานข้อมูล (Database)
  - คลังข้อมูล (Data Warehouse)
- การจัดการข้อมูลขนาดใหญ่ (Big Data)
- ตัวอย่างการประยุกต์ใช้การวิเคราะห์ข้อมูลด้วยเทคโนโลยี Data Science

# Data Science Use Cases

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



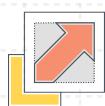
## Customer Lifetime Value

แบ่งแยกลูกค้าตามมูลค่าที่สร้างให้กับบริษัท



## Customer Segmentation

แบ่งกลุ่มลูกค้าตามพฤติกรรมต่างๆ เพื่อจะได้เข้าใจลูกค้าได้มากขึ้น



## Up- and Cross-Selling

แนะนำสินค้า/บริการเพื่อให้ลูกค้าซื้อมากขึ้นหรือแพงขึ้น



## Next Best Action

นำเสนอสินค้า/บริการให้ตรงใจและเหมาะสมกับช่วงเวลาที่ต้องการ



## Propensity to Buy

คาดการณ์ว่าลูกค้าคนใดบ้างที่มีโอกาสซื้อสินค้า/บริการ



## Churn Prevention

คาดการณ์ว่าลูกค้าคนใดบ้างที่มีโอกาสจะยกเลิกการใช้บริการ

**source:** [www.rapidminer.com/](http://www.rapidminer.com/)

Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

35

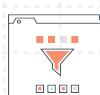
# Data Science Use Cases

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL



## Fraud Detection

ค้นหาและคาดการณ์การใช้งานที่ผิดปกติ



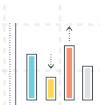
## Risk Management

จัดการกับความเสี่ยงต่างๆ



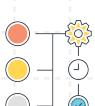
## Demand Forecast

พยากรณ์ยอดขายหรือยอดผลิตสินค้าต่างๆ



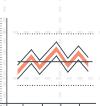
## Price Optimization

ปรับเปลี่ยนราคาของสินค้า/บริการให้สอดคล้อง



## Quality Assurance

ค้นหาปัญหาที่เกี่ยวข้องกับคุณภาพของสินค้าก่อนการจำหน่าย



## Predictive Maintenance

คาดการณ์ความผิดปกติของเครื่องจักรก่อนปัญหาจะเกิดขึ้น

**source:** [www.rapidminer.com/](http://www.rapidminer.com/)

Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

36

# Data Science Use Cases

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

Use-cases	Retail	Banking	Telecommunication	Insurance	Manufacturing
Customer Lifetime Value (CLV)	●	●	●	●	
Customer Segmentation	●	●	●	●	
Up- and Cross-Selling	●	●	●	●	
Next Best Action	●	●	●	●	
Propensity to Buy	●	●	●	●	
Churn Prevention	●	●	●	●	
Fraud Detection		●		●	
Demand Forecast	●		●		●
Quality Assurance					●
Risk Management		●		●	
Price Optimization	●				
Predictive Maintenance					●

# Data Science Applications

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- ตัวอย่างการนำ Data Science ไปใช้งาน



# Data Science Applications

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- คาดการณ์การตั้งครรภ์



- ห้าง Target ทำการวิเคราะห์พฤติกรรมการซื้อสินค้าของลูกค้าเพื่อจัดทำโปรโมชัน
- พบรูปแบบ (pattern) ว่า ลูกค้ามีการซื้อวิตามิน ซื้ออาหารบำรุง หรือ ซื้อตุ๊กตาเพิ่ม เมื่อลูกค้าจะเริ่มตั้งครรภ์
- Target จะส่ง promotion ให้ลูกค้าเหล่านั้น

- source: <http://www.youtube.com/watch?v=f2Kjz4833Y>

Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

39

# Data Science Applications

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

- แนะนำสินค้าที่เกี่ยวข้อง

- Netflix แนะนำภาพยนตร์ที่คล้ายกับที่เคยดูมาท่อง



- source: <https://indianexpress.com/article/technology/social/netflix-partnering-with-whatsapp-in-india-will-send-content-recommendations-5002335/>

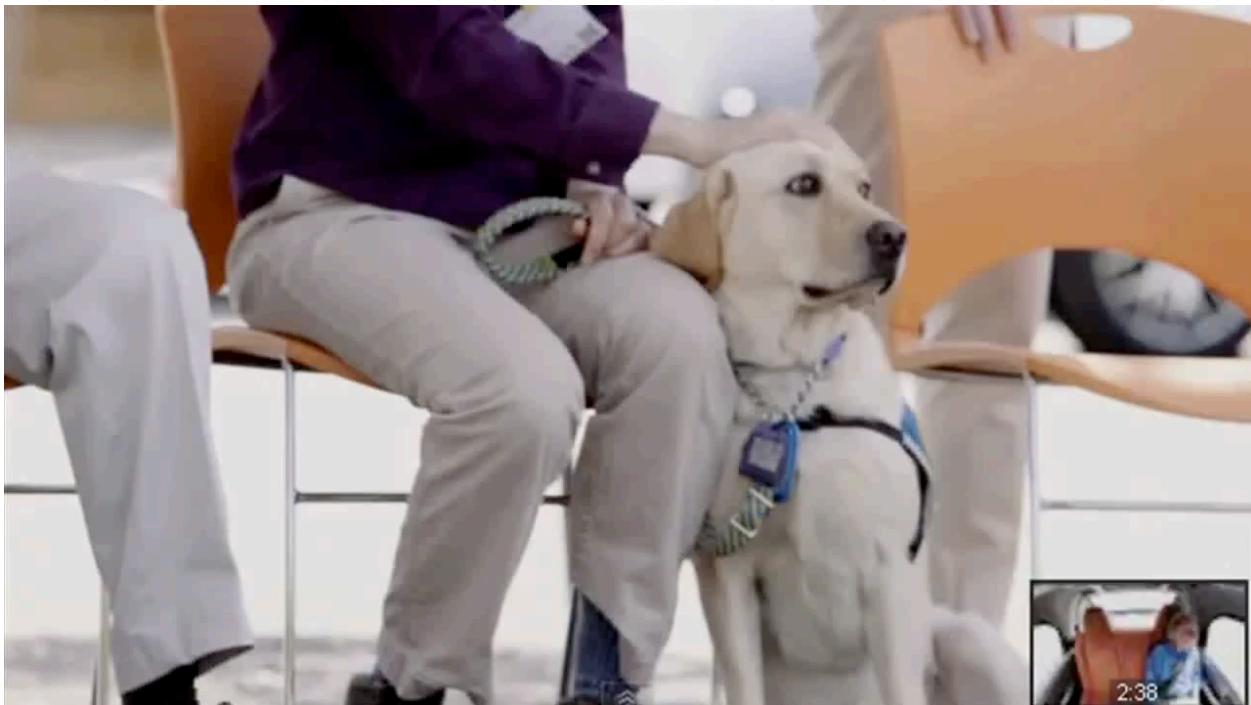
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

40

# Data Science Applications

- สำหรับการussian M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- Google Self-Driving Car



source: <https://www.youtube.com/watch?v=8fjNSUWX7nQ>

<http://facebook.com/datacube.th>

41

# Data Science Applications

- สำหรับการussian M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- แนวโน้มราคาตัวเครื่องบิน

**KAYAK** HOTELS FLIGHTS CARS PACKAGES DEALS Login ☰

Advice: Wait Confidence: 79%  
Prices may fall within 7 days ⓘ

**Stops**

<input checked="" type="checkbox"/> nonstop	\$2089
<input checked="" type="checkbox"/> 1 stop	\$901
<input checked="" type="checkbox"/> 2+ stops	\$1088

LHR ↔ JFK   Jun 1 depart → Jun 5 return   Economy cabin   1 traveler   Change

Sort by: price (low to high) ▾   1095 of 4117 flights show all   Round-trip | Segment NEW

Fly to New York [www.etihad.com/Ne](http://www.etihad.com/Ne)  
Fly to New York w  
Return: 1/4/14/2014  
Adults: 1

↑ Prices may rise within 7 days

82% Confidence: Our model has been 82% accurate on forecasting whether these fares will rise or stay within \$20 of the current price over the next 7 days. The forecast is based on analysis of historical price changes and is not a guarantee of future results.

Time to buy? See the rise and fall of prices over the past 90 days.  
Fare Trend for Flights Departing Dec 25 2012



# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- คาดการณ์การลาออกจากงาน



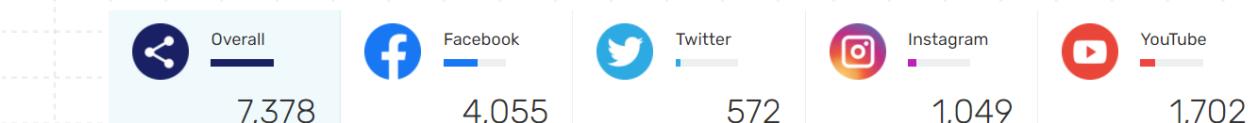
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

43

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- วิเคราะห์ก้าบคติในແນ່ງຕ່າງໆ ຈາກສັນຄມອອນໄລນ໌



## Data Timeline



Reference: การวิเคราะห์การกล่าวถึงแบรนด์ໂກຮ້າພໍານົດໃນປະເທດໄທ ໂດຍ Vivat Boonchun and Surasak Chantarach

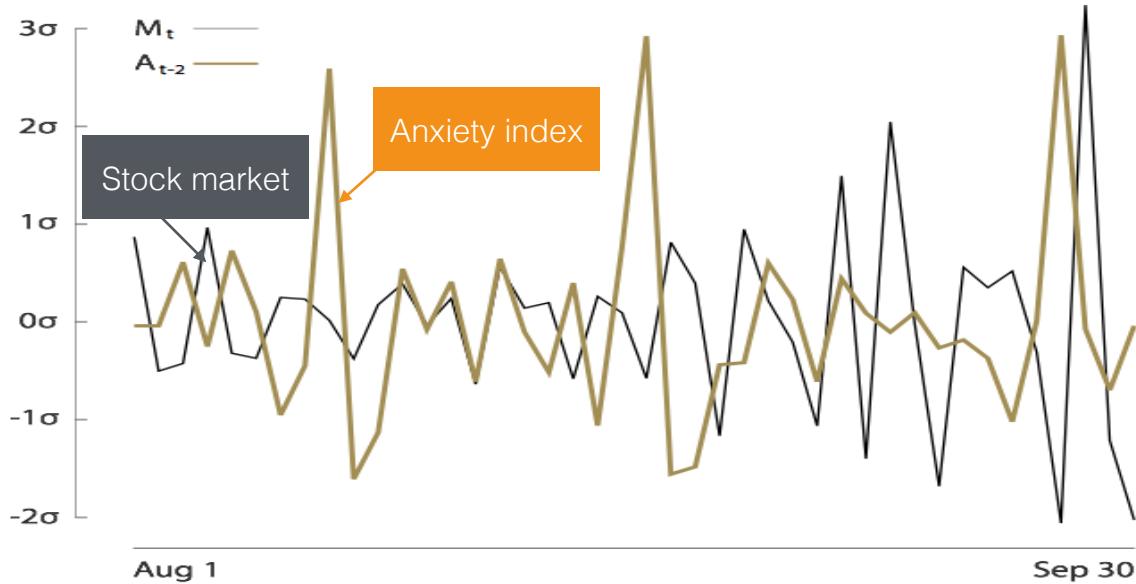
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

44

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- คาดการณ์แนวโน้มของตลาดหุ้น



Reference: Eric Gilbert, Karrie Karahalios: Widespread Worry and the Stock Market. ICWSM 2010

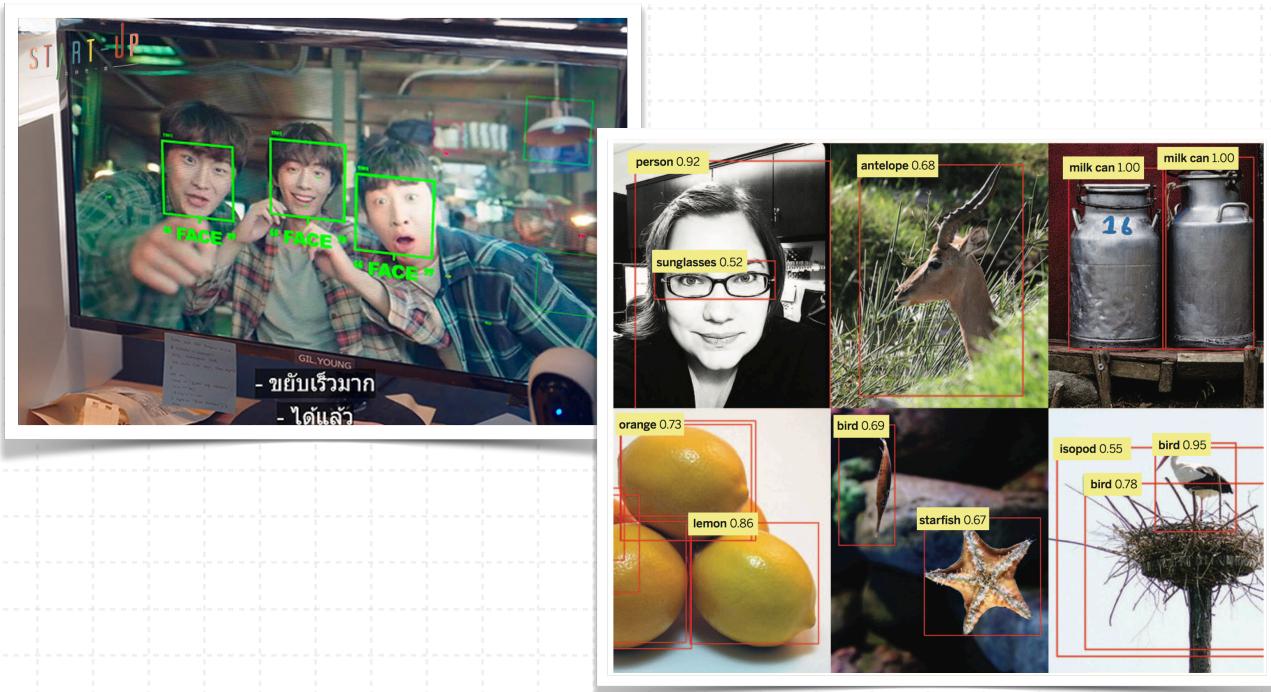
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

45

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- ระบบตรวจจับวัตถุ (Image Recognition)



ที่มา: [https://medium.com/@artisan\\_digital/สอนบันวัตกรรม-ai-เส้นทางพัฒนาธุรกิจชานชาลาจากซีรีส์-start-up-f287ddaa843](https://medium.com/@artisan_digital/สอนบันวัตกรรม-ai-เส้นทางพัฒนาธุรกิจชานชาลาจากซีรีส์-start-up-f287ddaa843)

<https://cs.uwaterloo.ca/~y328yu/mycourses/480-2018/readings/JordanMitchell.pdf>

@acube.th

46

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- ระบบตรวจจับคนที่ไม่ใส่หน้ากากอนามัย



ที่มา: [https://www.youtube.com/watch?v=S4TUM\\_p6H3E&feature=emb\\_logo](https://www.youtube.com/watch?v=S4TUM_p6H3E&feature=emb_logo)

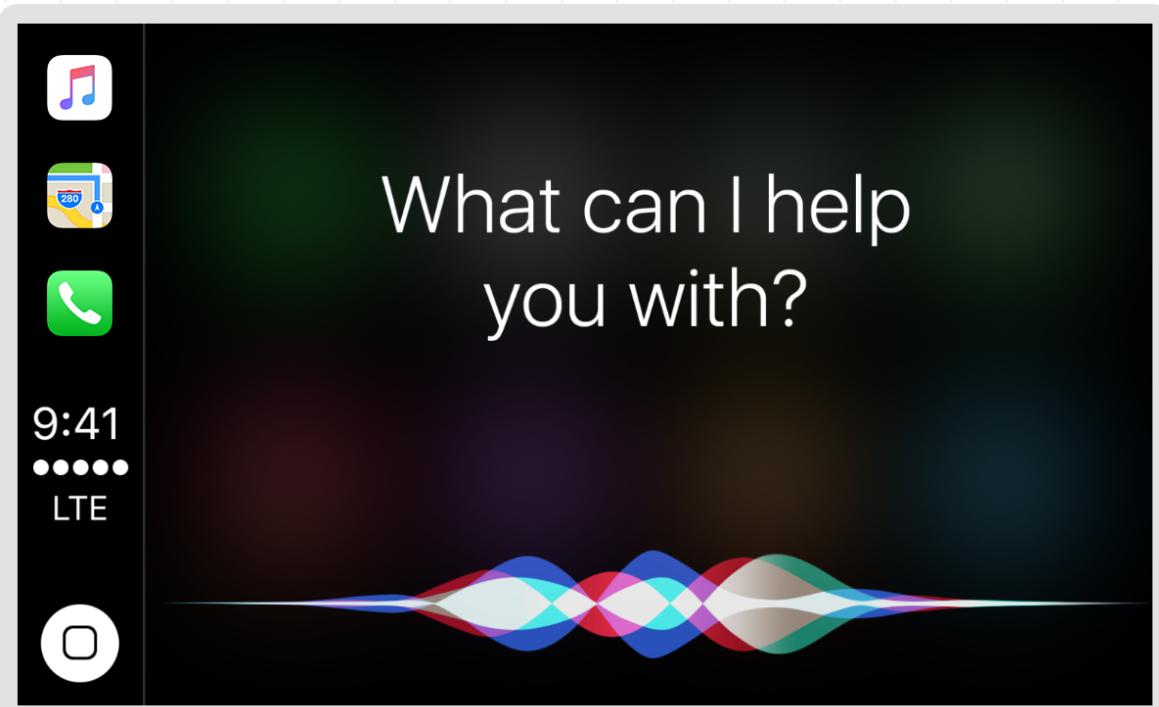
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

47

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- ระบบรู้จำเสียง (Speech Recognition)



ที่มา: <https://developer.apple.com/design/human-interface-guidelines/carplay/interaction/voice/>

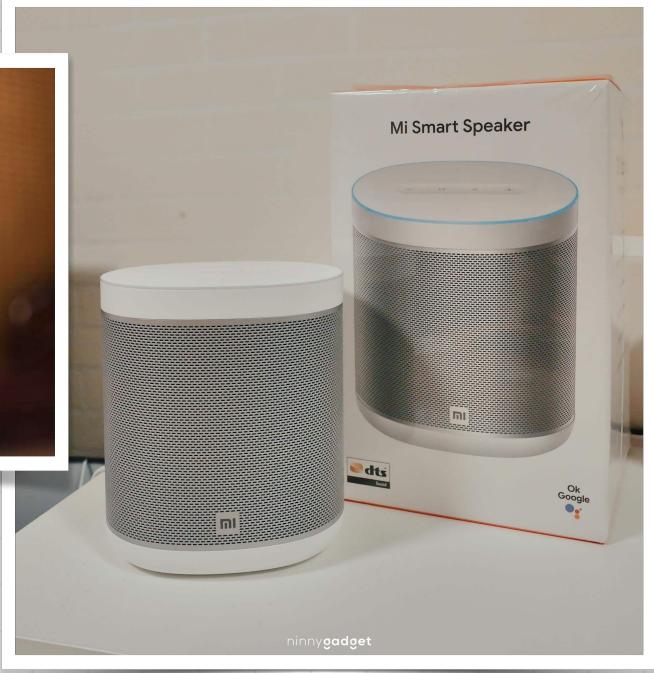
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

48

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL  
 • จำลองอัจฉริยะ



ที่มา: <https://www.zdnet.com/product/amazon-echo/>

<https://ninnygadget.com/review/mi-smart-speaker/>

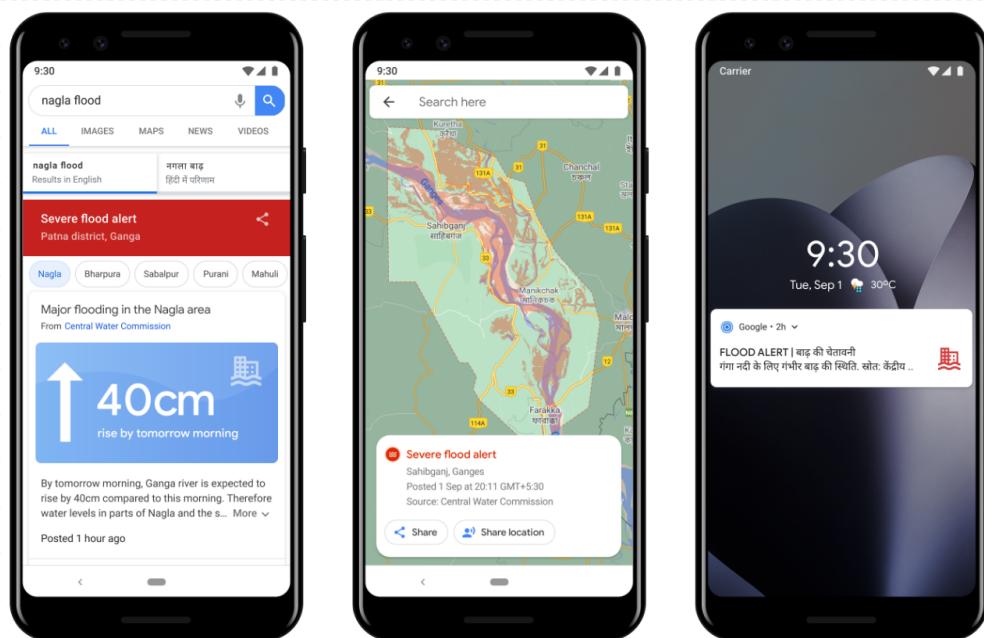
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

49

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL  
 • ระบบคาดการณ์และแจ้งเตือนการเกิดอุตุภัย



ที่มา: <https://blog.google/technology/ai/flood-forecasts-india-bangladesh/>

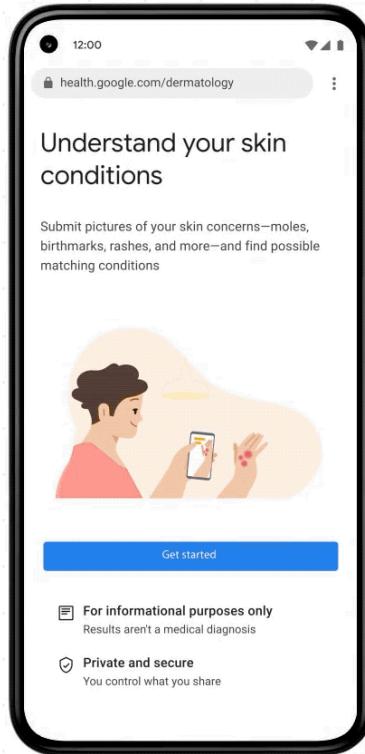
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

50

# Data Science Applications

- สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL
- ระบบคาดการณ์ความผิดปกติบนผิวหนัง



ที่มา: <https://blog.google/technology/health/ai-dermatology-preview-io-2021/>

Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

51

## How to get dataset

สำหรับการบรรยายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

## UCI Machine Learning Repository

เป็นเว็บไซต์ที่รวบรวมข้อมูลที่ใช้ในการทำวิจัยทาง Machine Learning

Latest News:		
09-24-2018: Welcome to the new Repository admins Dheeru Dua and Efi Karra Taniskidou!	04-04-2013: Welcome to the new Repository admins Kevin Bacho and Moshe Lichman!	03-01-2010: Note from donor regarding Netflix data
10-16-2009: Two new data sets have been added.	09-14-2009: Several new data sets have been added.	03-24-2008: New data sets have been added!
06-25-2007: Two new data sets have been added: UJI Dian Chenhaven MACS Company		

Newest Data Sets:		
04-21-2021:  Synchronous Machine Data Set	04-20-2021:  Wikipedia Math Essentials	04-20-2021:  Wikipedia Math Essentials

Most Popular Data Sets (hits since 2007):		
4007794:  Iris	2161539:  Adult	1672204:  Wine

Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

52

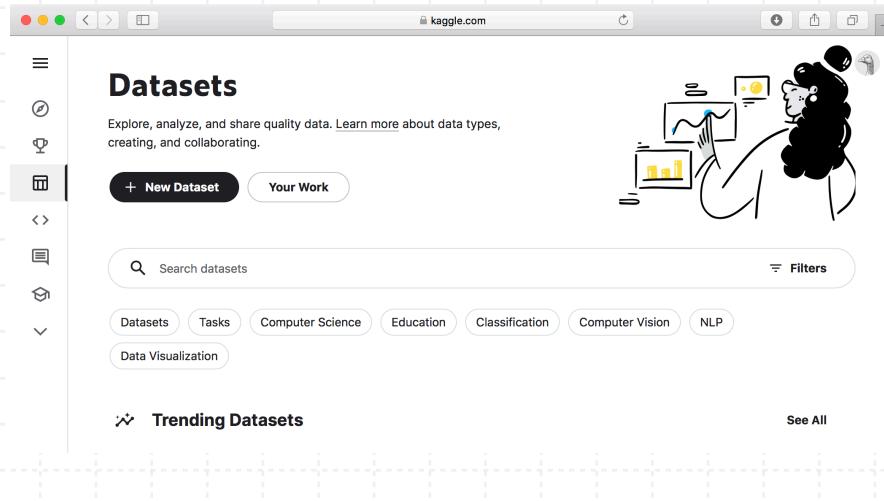
# How to get dataset

สำหรับการอภิปรายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

## kaggle.com

เป็นเว็บไซต์ที่รวมการแข่งขัน (competition)

ทางด้าน Data Science / Machine Learning



The screenshot shows the kaggle.com homepage. The main heading is "Datasets". Below it, a sub-headline says "Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.". There is a "New Dataset" button and a "Your Work" button. To the right is a cartoon illustration of a person working on a computer. Below the main heading is a search bar with the placeholder "Search datasets" and a "Filters" button. A navigation bar below the search bar includes categories like Datasets, Tasks, Computer Science, Education, Classification, Computer Vision, NLP, and Data Visualization. At the bottom, there is a "Trending Datasets" section with a "See All" link.

# How to get dataset

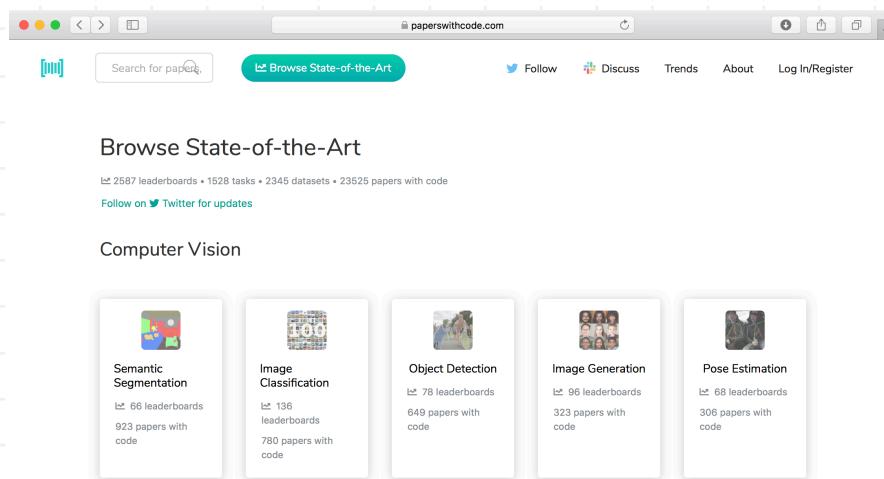
สำหรับการอภิปรายในหลักสูตร M: Sc. in Artificial Intelligence for Business Analytics, KMITL

## paperswithcode.com

เป็นเว็บไซต์ที่รวมงานวิจัยทางด้าน

Machine Learning และมี Code

ที่ใช้ในการทำวิจัยให้ดาวน์โหลดได้โดยครับ



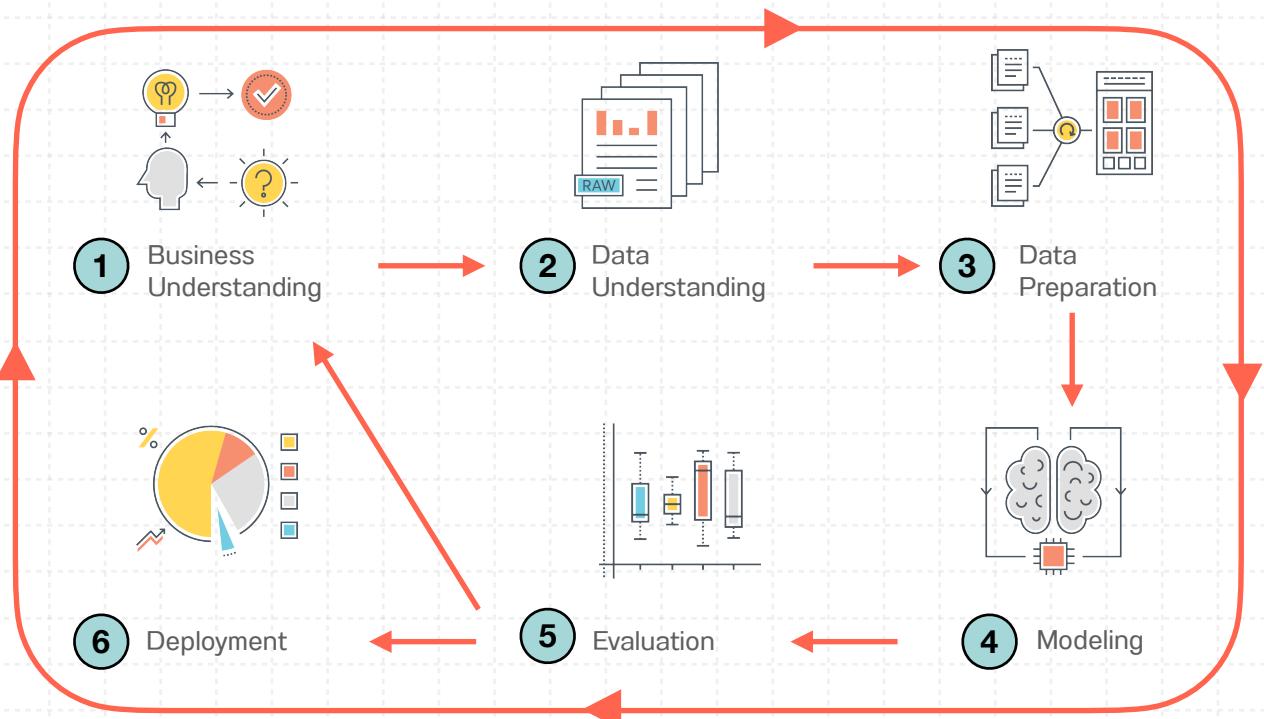
The screenshot shows the paperswithcode.com homepage. The main heading is "Browse State-of-the-Art". Below it, a sub-headline says "2587 leaderboards • 1528 tasks • 2345 datasets • 23525 papers with code". There is a "Follow on Twitter for updates" link. Below the sub-headline, there is a "Computer Vision" section with five sub-sections: Semantic Segmentation, Image Classification, Object Detection, Image Generation, and Pose Estimation. Each sub-section has a thumbnail image, a title, and a brief description of the number of leaderboards, tasks, and datasets.

Task	Leaderboards	Papers with code
Semantic Segmentation	66	923
Image Classification	136	780
Object Detection	78	649
Image Generation	96	323
Pose Estimation	68	306

# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- CRISP-DM ย่อมาจาก **CRoss Industry Standard Process for Data Mining**



# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

STEP	DESCRIPTION
Business Understanding	Define the project.
Data Understanding	Examine the data; identify problems in the data.
Data Preparation	Fix problems in the data; create derived variables.
Modeling	Build predictive or descriptive models.
Evaluation	Assess models; report on the expected effects of
Deployment	Plan for use of models.

# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## • 1. Business Understanding

- ทำความเข้าใจกับปัญหา หรือ โอกาสเชิงธุรกิจ
- ระบุ output หรือเป้าหมายที่ต้องการได้จากการวิเคราะห์ข้อมูล
- ตัวอย่าง เช่น
  - ทำอย่างไรเพิ่มยอดขายให้กับสินค้าชนิดต่างๆ ได้
  - ต้องการแบ่งกลุ่มนักศึกษาออกตามความสนใจ
  - ทำอย่างไรให้ลูกค้ากลับมาซื้อสินค้าได้อีก
  - อยากทำนายปริมาณน้ำฝนที่ตกใน 2 วันถัดไป
  - อยากรู้ว่าลูกค้าคนใดบ้างมีโอกาสป่วยเป็นโรคมะเร็ง

# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## • 2. Data Understanding

- ในขั้นตอนนี้เป็นการ
  - รวบรวมข้อมูลที่เกี่ยวข้อง
  - ข้อมูลถูกต้องนำเชื่อถือ
  - ข้อมูลที่ได้มีปริมาณมากพอหรือยัง
  - ข้อมูลที่ได้มีความเหมาะสม มีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์
- ตัวอย่าง
  - ข้อมูลการซื้อสินค้านั้นแต่ละบุคคล
  - ข้อมูลการลงทะเบียนและผลการศึกษาของนักศึกษา

# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **3. Data Preparation**

- ขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนที่ใช้เวลาบานทึกสุด
- เนื่องจากไม่เดลกที่ได้จากการทำด้วยมือ จึงให้ผลลัพธ์ที่ถูกต้องหรือไม่นัก ขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้ แบ่งออกได้เป็น 3 ขั้นตอนย่อยคือ

- **3.1 ทำการคัดเลือกข้อมูล (Data Selection)**

- กำหนดเป้าหมายก่อนว่าเราจะทำการวิเคราะห์อะไร
- เลือกใช้เฉพาะข้อมูลที่เกี่ยวข้องกับสิ่งที่เราจะทำการวิเคราะห์

# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **3. Data Preparation**

- **3.2 ทำการกรองข้อมูล (Data Cleaning)**

- ลบข้อมูลซ้ำซ้อน
- แก้ไขข้อมูลที่ผิดพลาด
- ข้อมูลผิดรูปแบบ
- ข้อมูลก่าหายไป
- ข้อมูล outlier กี่แปลกแยกจากคนอื่น

ข้อมูลนักศึกษาชั้นปีที่ 1 ปีการศึกษา 2564

รหัส	เพศ	อายุ	ความสูง	น้ำหนัก
64001	ชาย	18	180	70
6402A	ณ		80	35
64123	หญิง	19	150	2500
65002	ช	17	175	90



# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## • 3. Data Preparation

### • 3.3 แปลงรูปแบบของข้อมูล (data transformation)

- เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ ตามอัลกอริทึมของ Machine Learning ที่เลือกใช้



ID	สินค้า	จำนวน	ID	สมุด	ปากกา	ยางลบ	คลิป
1	ปากกา	1	1	-	TRUE	TRUE	TRUE
1	ยางลบ	1	2	TRUE	TRUE	-	-
1	คลิป	10	3	TRUE	TRUE	TRUE	-
2	สมุด	2					
2	ปากกา	2					
3	สมุด	1					
3	ปากกา	3					
3	ยางลบ	2					

ข้อมูลในฐานข้อมูล POS

ข้อมูลสำหรับการหากฎความสัมพันธ์ (Association Rules)

# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## • 3. Data Preparation

### • 3.3 แปลงรูปแบบของข้อมูล (data transformation)

- เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ ตามอัลกอริทึมของ Machine Learning ที่เลือกใช้

เมื่อวันที่ 4 มกราคม 2557 **เฟซบุ๊ก** เปิดตัวหน้าเพวใหม่ชื่อว่า Facebook A Look Back เมื่อผู้ใช้งานคลิกไปยังหน้านี้ก็จะแสดงคลิปวิดีโอเก็บออดเล่าเรื่องราวของผู้ใช้งานคนนั้นๆ เช่น เริ่มเล่น**เฟซบุ๊ก**ครั้งแรกปีไหน, โพสต์แรกบน**เฟซบุ๊ก**, รูปภาพที่ถูกกดไลค์มากที่สุด, รูปภาพที่ถูกแชร์มากที่สุด และ 20 อันดับเรื่องราวต่าง ๆ ที่เกิดขึ้นใน**เฟซบุ๊ก**ก็จะถูกแสดงและรวมไว้ในคลิปวิดีโอนี้

เอกสารที่ 2



ID	เฟซบุ๊ก	รูปภาพ	ไลค์	แชร์	คลิปวิดีโอ
1	4	2	1	1	2
2	...				

ตารางแสดงจำนวนความตื่นต่องแต่ละคำ

# CRISP-DM

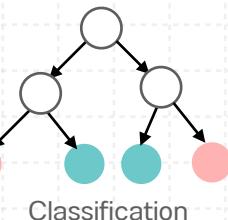
สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **4. Modeling**

- เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคด้าไม่น้อย

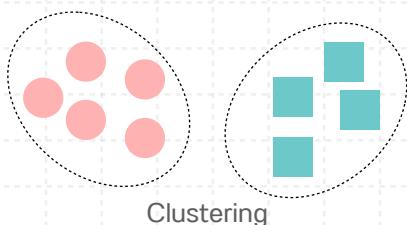
- **Classification**

- สร้างโมเดลจากข้อมูลที่มีอยู่เพื่อทำนายอนาคต
- เช่น ทำนายปริมาณน้ำฝนที่ตกในวันถัดไป



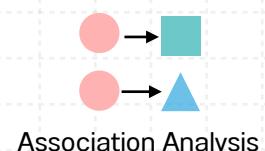
- **Clustering**

- แบ่งข้อมูลหลายๆ กลุ่มตามความคล้ายคลึง
- เช่น แบ่งกลุ่มนักศึกษาตามคะแนนที่ได้



- **Association Analysis**

- หากว่ามีสัมพันธ์ของข้อมูลที่เกิดร่วมกัน
- เช่น คันหาสินค้าที่มีการซื้อร่วมกันบ่อยๆ

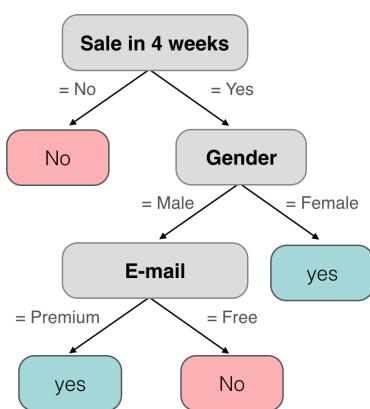


# CRISP-DM

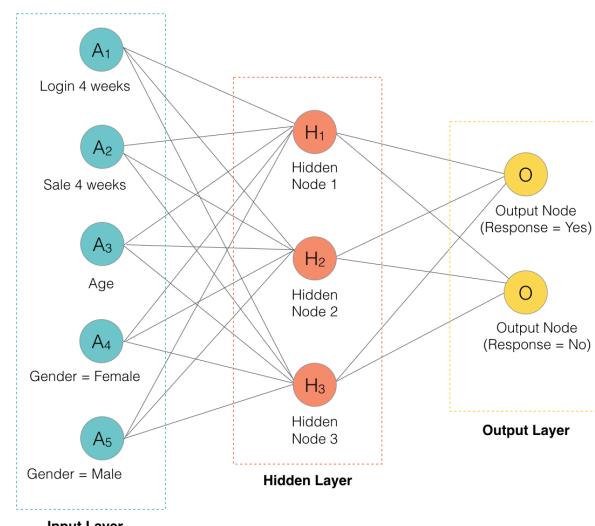
สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **5. Evaluation**

- ประเมินหรือวัดประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลในขั้นตอนก่อนหน้านั้น



VS



โมเดล Decision Tree

โมเดล Neural Network

# CRISP-DM

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## • 6. Deployment

- นำโมเดลที่ได้ หรือ ผลการวิเคราะห์ที่ได้ไปใช้งานจริง

แผนที่คาดการณ์ภัยพิบัติอ่วงหน้า 1 วัน (อุทกภัย/แม่น้ำ/ดินโคลนล้ม/วาตภัย) 

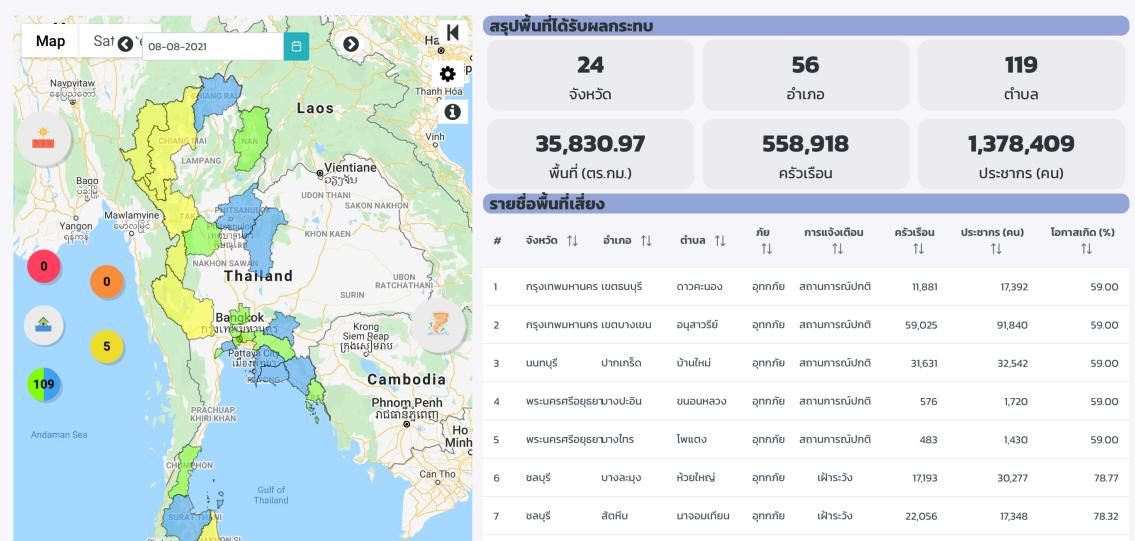


image source: ระบบสนับสนุนการตัดสินใจ กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## • ตัวอย่าง CRISP-DM

- อ้างอิงจากงานวิจัยเรื่อง การใช้เทคนิคดั้งเดิมเพื่อพัฒนาคุณภาพการศึกษาและวิศวกรรมศาสตร์ \*

### 1. Business Understanding

- นิสิตคุณวิศวกรรมศาสตร์ ม.เกษตรศาสตร์ จะเลือกภาควิชาเมื่อในชั้นปีที่ 2
- นิสิตเลือกภาควิชาไม่ตรงกับความสามารถของตนเอง
  - เลือกตามเพื่อน
  - เลือกตามที่ผู้ปกครองแนะนำ
  - นิสิตบางคนได้ผลการเรียนตกต่ำและทำให้ต้องออกจากมหาวิทยาลัยกลางคัน

\* กฤษณะ ไวยนัย, ชิดชนก สังคีธ และ ธนาวิท รัตน์ธนาบานันท์, การใช้เทคนิคดั้งเดิมเพื่อพัฒนาคุณภาพการศึกษาและวิศวกรรมศาสตร์, NECTEC Technical Journal, 3(11), 2001

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 2. Data Understanding

- ข้อมูลนิสิตคณวิศวกรรมศาสตร์ ม.เทคโนโลยีราชมงคลชั่วปี พ.ศ. 2535 - 2542
- นิสิตประมาณ 10,000 คน
- ข้อมูลมีจำนวน 476,085 แถว
- ข้อมูลแบ่งเป็น 2 ส่วน
  - ข้อมูลประวัติส่วนตัวของนิสิต
    - เพศ, ที่อยู่, GPA ระดับมัธยมปลาย, GPA ชั้นปีที่ 1
  - ข้อมูลการลงทะเบียนของนิสิต
    - เกรดวิชาคณิตศาสตร์, เกรดวิชาฟิสิกส์, เกรดวิชาเคมี

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 3. Data Preparation

- คัดเลือกวิชาที่เกี่ยวข้องกับภาควิชาต่างๆ ในคณวิศวกรรมศาสตร์
- แปลงข้อมูลให้เหมาะสมกับการวิเคราะห์

รหัสบัตร	เพศ	ที่อยู่	GPA มัธยมปลาย	...	GPA
37058063	ชาย	กรุงเทพ	2.5	...	2.3
37058167	ชาย	สงขลา	3.4	...	3.3
...	...	...	...	...	...



รหัสบัตร	วิชา	ปีการศึกษา	เกรด
37058063	คอมพิวเตอร์	2537	C
37058063	คณิตศาสตร์	2537	D
...	...	...	...

ข้อมูลประวัติส่วนตัวของนิสิต



ข้อมูลการลงทะเบียนของนิสิต

รหัสบัตร	เพศ	คอมพิวเตอร์	คณิตศาสตร์	...	GPA
37058063	ชาย	LOW	LOW	...	2.3
37058167	ชาย	HIGH	HIGH	...	3.3
...	...	...	...	...	...

ข้อมูลสำหรับการใช้วิเคราะห์

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

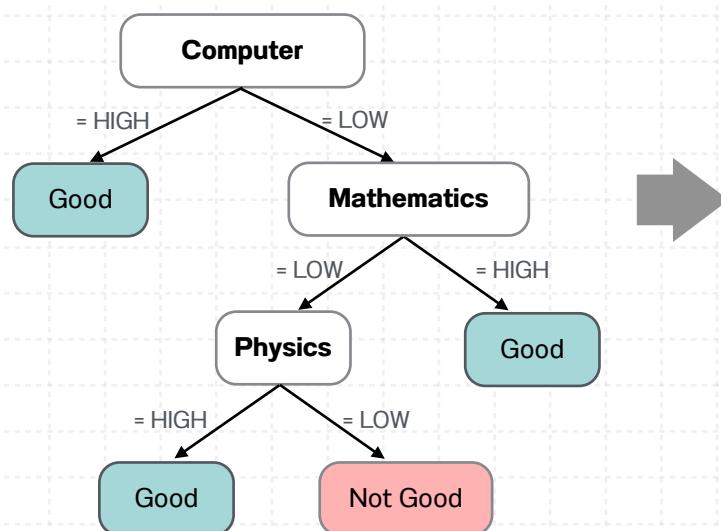
## 4. Modeling

- แบ่งข้อมูลออกเป็น 2 ส่วน คือ
  - 70% ของข้อมูลทั้งหมดใช้ในการสร้างโมเดล
  - 30% ของข้อมูลทั้งหมดใช้ในการทดสอบประสิทธิภาพของโมเดล
- สร้างโมเดลด้วยเทคนิค Decision Tree ซึ่งจะได้โมเดลที่สามารถเข้าใจได้ง่าย
- โมเดลแบ่งแยกตามภาควิชาต่างๆ เช่น ภาควิชาชีวกรรมคอมพิวเตอร์ วิศกรรมไฟฟ้า
- คำตอบ (class) จะแบ่งเป็น 2 ประเภท คือ
  - GOOD** หมายถึง นิสิตเรียนในภาควิชานี้แล้วจบมาได้ GPA อยู่ในช่วง 40% แรก (top 40%)
  - NOT GOOD** หมายถึง นิสิตเรียนในภาควิชานี้แล้วจบมาได้ GPA อยู่ในช่วง 40% จากก้น (bottom 40%)

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 4. Modeling



- **IF Computer is HIGH THEN**  
Graduate is Good
- **IF Computer is LOW AND**  
Mathematics is HIGH **THEN**  
Graduate is Good
- **IF Computer is LOW AND**  
Mathematics is LOW **AND**  
Physics is HIGH **THEN**  
Graduate is Good
- **IF Computer is LOW AND**  
Mathematics is LOW **AND**  
Physics is LOW **THEN**  
Graduate is Bad

เขื่อนไขที่สร้างได้จากโมเดล

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 5. Evaluation

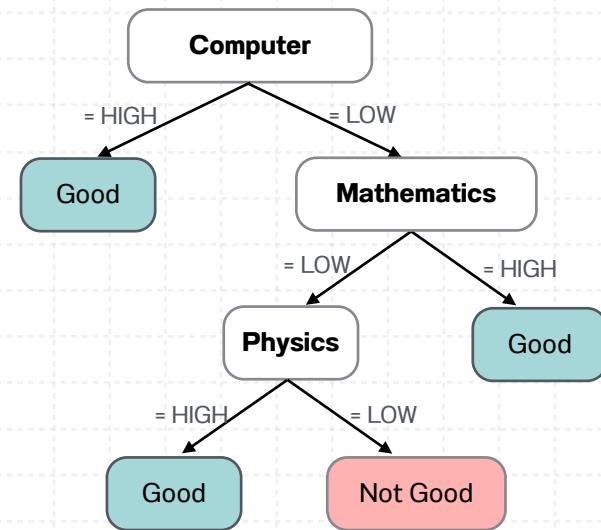
- ทดสอบด้วยข้อมูล 30% ที่แบ่งไว้
- คำนวณค่าความถูกต้อง

รหัสนิสิต	เพศ	คอมพิวเตอร์	คณิตศาสตร์	...	Com Eng
5700123	ชาย	LOW	HIGH	...	??

ข้อมูลของนักศึกษาปีที่ 1 ที่ต้องการได้รับคำแนะนำ

## 6. Deployment

- นำไปแบนเป็นไซต์ชั้นปีที่ 1 ที่กำลังจะเลือกภาควิชา
- พิจารณาจากเกรดตามโมเดลที่สร้างได้



โมเดลของภาควิชาชีวิศวกรรมคอมพิวเตอร์

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 5. Evaluation

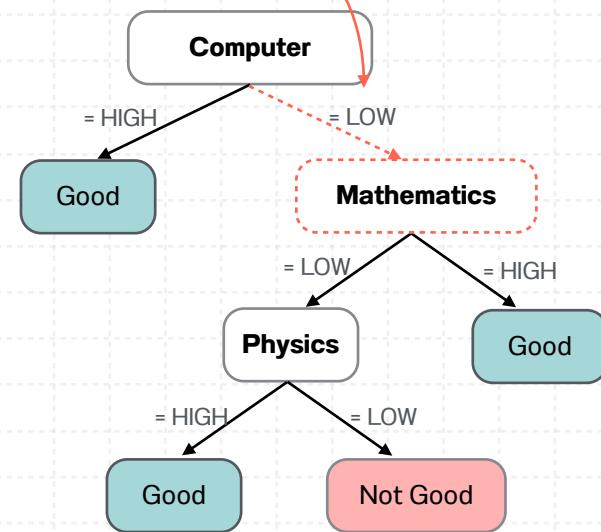
- ทดสอบด้วยข้อมูล 30% ที่แบ่งไว้
- คำนวณค่าความถูกต้อง

รหัสนิสิต	เพศ	คอมพิวเตอร์	คณิตศาสตร์	...	Com Eng
5700123	ชาย	LOW	HIGH	...	??

ข้อมูลของนักศึกษาปีที่ 1 ที่ต้องการได้รับคำแนะนำ

## 6. Deployment

- นำไปแบนเป็นไซต์ชั้นปีที่ 1 ที่กำลังจะเลือกภาควิชา
- พิจารณาจากเกรดตามโมเดลที่สร้างได้



โมเดลของภาควิชาชีวิศวกรรมคอมพิวเตอร์

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 5. Evaluation

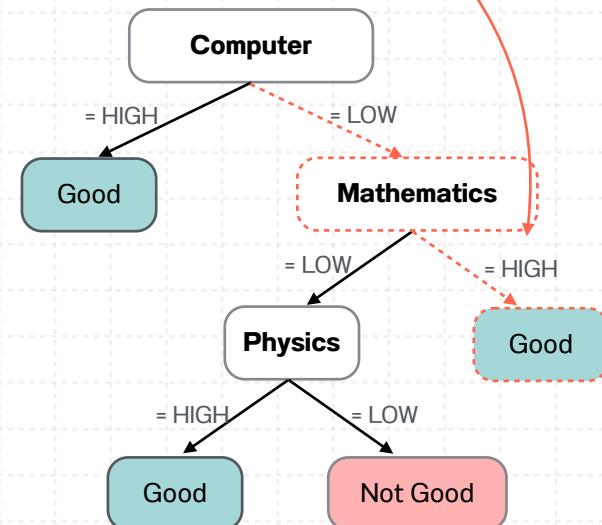
- ทดสอบด้วยข้อมูล 30% ที่แบ่งไว้
- คำนวณค่าความถูกต้อง

รหัสนิสิต	เพศ	คอมพิวเตอร์	คณิตศาสตร์	...	Com Eng
5700123	ชาย	LOW	HIGH	...	??

ข้อมูลของนักศึกษาปีที่ 1 ที่ต้องการได้รับคำแนะนำ

## 6. Deployment

- นำไปແນບັນສົຕ້ນປັກໍ 1 ທີ່ກໍາລັງ  
ຈະເລືອກການວິຊາ
- ພົຈມາຈາດເກຣດຕາມໂນເດລ  
ທີ່ສ້າງໄດ້



ໂນເດລຂອງການວິຊາວິສວກຮຽນຄົມພິວເຕອນ

# CRISP-DM Example 1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 5. Evaluation

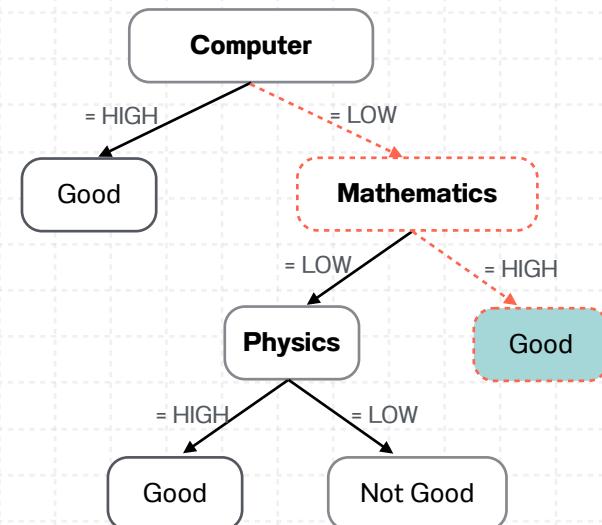
- ทดสอบด้วยข้อมูล 30% ที่แบ่งไว้
- คำนวณค่าความถูกต้อง

รหัสนิสิต	เพศ	คอมพิวเตอร์	คณิตศาสตร์	...	Com Eng
5700123	ชาย	LOW	HIGH	...	Good

ข้อมูลของนักศึกษาปีที่ 1 ที่ต้องการได้รับคำแนะนำ

## 6. Deployment

- นำไปແນບັນສົຕ້ນປັກໍ 1 ທີ່ກໍາລັງ  
ຈະເລືອກການວິຊາ
- ພົຈມາຈາດເກຣດຕາມໂນເດລ  
ທີ່ສ້າງໄດ້



ໂນເດລຂອງການວິຊາວິສວກຮຽນຄົມພິວເຕອນ

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ตัวอย่าง CRISP-DM

## 1. Business Understanding

- บริษัทพัฒนา application บนมือถือแห่งหนึ่งประสบปัญหาว่าผู้ใช้ยกเลิกการใช้บริการของตนเอง (churn)
- อยากรู้ผลการณ์ (predict) ว่าผู้ใช้คนไหนมีโอกาสแยกเลิกการใช้บริการบ้าง

## 2. Data Understanding

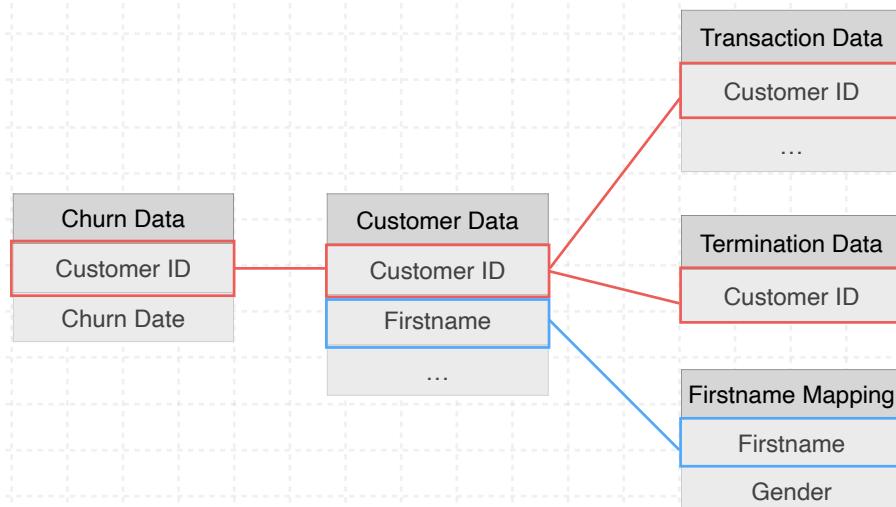
- บริษัทมีข้อมูล 5 ส่วน คือ
  - ข้อมูลของลูกค้า (customer data)** ซึ่งประกอบด้วย ชื่อ วันเกิด และ รหัสไปรษณีย์
  - ข้อมูลการใช้งาน (transaction data)** ซึ่งประกอบด้วย วันเวลาที่ใช้งาน การชำระเงินผ่านช่องทาง ใหม่ และ จำนวนเงินที่ใช้จ่าย
  - ข้อมูลผู้ใช้กี่ยกเลิกการใช้บริการ (churn data)**

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 2. Data Understanding

- ข้อมูล Firstname Mapping** ซึ่งประกอบด้วย ชื่อ และ เพศ (gender)
- ข้อมูล Termination** ซึ่งประกอบด้วย Customer ID



# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 2. Data Understanding

- ข้อมูลของลูกค้า (customer data)

Row No.	id	CustomerId	PostalCode	HashCode	Vorname	Birthday
1	1	3	87213	tKlbadnh	Eva	Apr 13, 1988 4:25:24 PM ICT
2	2	4	38548	RcW2Pb3w	Kunigunde	Jun 21, 1974 4:25:24 PM ICT
3	3	6	44573	akWNQl4e	Notburga	Jun 5, 1985 4:25:24 PM ICT
4	4	7	70936	glrPDLzY	Maximiliane	Jul 26, 1992 4:25:24 PM ICT
5	5	9	49705	3rGPBX98	Dorothea	Aug 4, 1943 4:25:24 PM ICT
6	6	12	42376	XlxhfOlo	Maria	Jul 29, 1960 4:25:24 PM ICT
7	7	16	52245	3ANQ9shn	Rosina	Jun 28, 1964 4:25:24 PM ICT
8	8	17	56625	BDEPLKmG	Susanne	Apr 14, 1989 4:25:24 PM ICT
9	9	20	66713	gQBDZ8Lw	Genoveva	Mar 31, 1948 4:25:24 PM ICT
10	10	21	70052	G1c0kZqQ	Katharina	Nov 16, 1931 4:25:24 PM ICT

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 2. Data Understanding

- ข้อมูลการใช้จ่าย (transaction data)

ExampleSet (4334 examples, 0 special attributes, 4 regular attributes)

Row No.	CustomerId ↑	TransactionValue	PaymentMethod	Date
1	1	11.5081	credit card	Apr 17, 2012 2:05:39 AM ICT
1001	1	374.2178	cheque	Aug 7, 2010 2:05:39 AM ICT
1002	1	11.0981	cash	May 28, 2011 2:05:39 AM ICT
1003	1	0.2493	credit card	Feb 27, 2011 2:05:39 AM ICT
1004	1	0.0266	cash	Oct 8, 2010 2:05:39 AM ICT
1005	1	1.1250	cash	Nov 25, 2010 2:05:39 AM ICT
2	2	330.0703	cheque	Nov 25, 2011 6:58:02 AM ICT
1006	2	0.2298	cheque	Sep 15, 2011 6:58:02 AM ICT
1007	2	23.7319	cheque	Feb 9, 2010 6:58:02 AM ICT
852	3	24.7471	credit card	Feb 15, 2012 5:29:26 PM ICT
3815	3	35.0881	credit card	Jun 13, 2010 5:29:26 PM ICT
3816	3	26.1187	credit card	Dec 12, 2011 5:29:26 PM ICT
3817	3	2.8185	cheque	Dec 4, 2010 5:29:26 PM ICT
3818	3	30.2979	credit card	Jun 12, 2011 5:29:26 PM ICT

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 2. Data Understanding

- ข้อมูลผู้ใช้ที่ยกเลิกการใช้บริการ (churn data)

ExampleSet (96 examples, 0 special attributes, 2 regular attributes)

Row No.	CustomerId	DateChurn
1	16	Apr 3, 2012 6:54:05 AM ICT
2	20	Apr 19, 2013 5:44:51 AM ICT
3	25	Apr 22, 2013 2:22:57 AM ICT
4	27	Nov 22, 2013 4:32:39 PM ICT
5	29	Aug 29, 2011 8:00:12 AM ICT
6	38	Oct 16, 2013 3:44:53 PM ICT
7	41	Nov 8, 2010 3:52:05 AM ICT
8	59	Apr 7, 2012 4:07:46 AM ICT
9	61	Oct 8, 2012 6:33:11 AM ICT
10	67	Dec 2, 2013 4:14:21 AM ICT

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 3. Data Preparation

- แปลงข้อมูลการใช้งาน (transaction data) ให้อยู่ในระดับของผู้ใช้แต่ละคน (user)

ExampleSet (4334 examples, 0 special attributes, 4 regular attributes)

Row No.	CustomerId ↑	TransactionValue	PaymentMethod	Date
1	1	11.5081	credit card	Apr 17, 2012 2:05:39 AM ICT
1001	1	374.2178	cheque	Aug 7, 2010 2:05:39 AM ICT
1002	1	11.0981	cash	May 28, 2011 2:05:39 AM ICT
1003	1	0.2493	credit card	Feb 27, 2011 2:05:39 AM ICT
1004	1	0.0266	cash	Oct 8, 2010 2:05:39 AM ICT
1005	1	1.1250	cash	Nov 25, 2010 2:05:39 AM ICT
2	2	330.0703	cheque	Nov 25, 2011 6:58:02 AM ICT
1006	2	0.2298	cheque	Sep 15, 2011 6:58:02 AM ICT
1007	2	23.7319	cheque	Feb 9, 2010 6:58:02 AM ICT
852	3	24.7471	credit card	Feb 15, 2012 5:29:26 PM ICT
3815	3	35.0881	credit card	Jun 13, 2010 5:29:26 PM ICT
3816	3	26.1187	credit card	Dec 12, 2011 5:29:26 PM ICT
3817	3	2.8185	cheque	Dec 4, 2010 5:29:26 PM ICT
3818	3	30.2979	credit card	Jun 12, 2011 5:29:26 PM ICT

- ยอดรวมค่าใช้จ่ายของแต่ละคน ( $\text{sum}(\text{TransactionValue})$ )
- ค่าใช้จ่ายเฉลี่ยของแต่ละคน ( $\text{average}(\text{TransactionValue})$ )
- วิธีการชำระเงิน ( $\text{PreferredPaymentMethod}$ )
- วันและเวลาที่ใช้งานล่าสุด ( $\text{LastTransaction}$ )
- จำนวนครั้งที่ใช้งาน ( $\text{TotalCount}$ )

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 3. Data Preparation

- แปลงข้อมูลการใช้งาน (transaction data) ให้อยู่ในระดับของผู้ใช้แต่ละคน (user)

CustomerId	Firstname	sum(TransactionValue)	average(TransactionValue)	PreferredPaymentMethod	LastTransaction	TotalCount
3	Eva	119.0702	23.8140	credit card	Feb 15, 2012 5:29:26 PM ICT	5
7	Maximiliane	111.4378	111.4378	credit card	Mar 15, 2012 10:17:03 PM ICT	1
16	Rosina	2.5000	0.5000	credit card	Feb 18, 2011 10:22:00 AM ICT	5
20	Genoveva	488.0992	244.0496	cash	Mar 9, 2011 11:04:47 PM ICT	2
22	Marlene	233.9894	38.9982	credit card	Oct 14, 2010 7:48:11 AM ICT	6
23	Annemarie	432.9481	432.9481	credit card	Aug 14, 2011 4:40:26 PM ICT	1
25	Magdalena	109.6131	54.8066	credit card	Aug 31, 2010 12:27:10 AM ICT	2
29	Barbara	310.9268	44.4181	credit card	Aug 15, 2011 2:23:26 AM ICT	7
32	Barbara	324.2332	324.2332	cash	Jun 6, 2010 3:53:50 AM ICT	1
38	Rosemarie	288.2103	144.1052	credit card	Jul 25, 2010 12:31:48 AM ICT	2
39	Hannelore	271.3465	27.1347	credit card	Nov 25, 2011 10:05:30 AM ICT	10
41	Annemarie	348.7102	174.3551	cash	Aug 4, 2010 12:14:01 AM ICT	2
42	Josephine	431.5721	431.5721	credit card	Oct 12, 2012 2:34:55 AM ICT	1

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 3. Data Preparation

- แปลงข้อมูลการใช้งาน (transaction data) ให้อยู่ในระดับของผู้ใช้แต่ละคน (user)

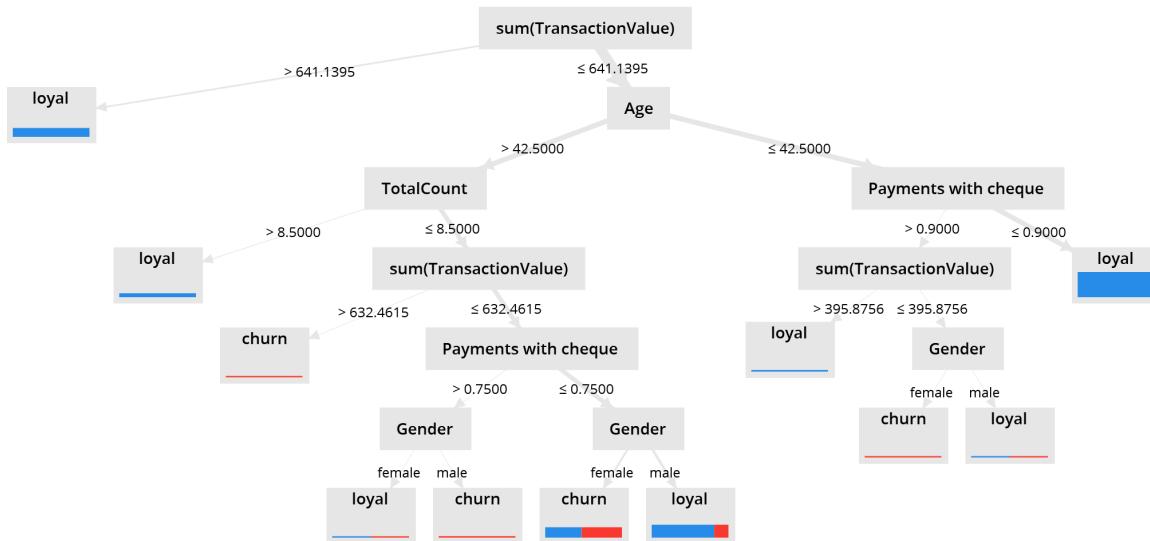
Name	Type	Missing	Statistics	Filter (14 / 14 attributes):
Id <b>CustomerId</b>	Integer	0	Min 3 Max 1000 Average 500.5635	
Label <b>Churn</b>	Nominal	0	Least churn (96) Most loyal (495) Values loyal (495), churn (96)	
Firstname <b>Firstname</b>	Polynomial	0	Least Veronika (0) Most Georg (14) Values Georg (14), Christoph (13), ...[157 more]	
Gender	Polynomial	0	Least female (209) Most male (382) Values male (382), female (209)	
sum(TransactionValue)	Real	0	Min 2.5000 Max 1119.4228 Average 377.0040	
average(TransactionValue)	Real	0	Min 0.2500 Max 1051.1909 Average 150.0511	
PreferredPaymentMethod	Polynomial	0	Least cheque (32) Most credit card (438) Values credit card (438), cash (121), ...[1 more]	
LastTransaction	Date time	0	Earliest date Nov 24, 2009 12:17 AM Latest date Feb 24, 2014 7:17 PM Duration 1553d 18h 59m 37s	
TotalCount	Numeric	0	Min 1 Max 10 Average 4.6531	
Payments with cash	Real	0	Min 0 Max 1 Average 0.2206	
Payments with cheque	Real	0	Min 0 Max 1 Average 0.0912	
Payments with credit card	Real	0	Min 0 Max 1 Average 0.6882	
Age	Integer	0	Min 20 Max 95 Average 45.1032	

# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 4. Modeling

- สร้างโมเดลด้วยเทคโนโลยี Decision Tree ซึ่งจะได้โมเดลที่สามารถเข้าใจได้ง่าย



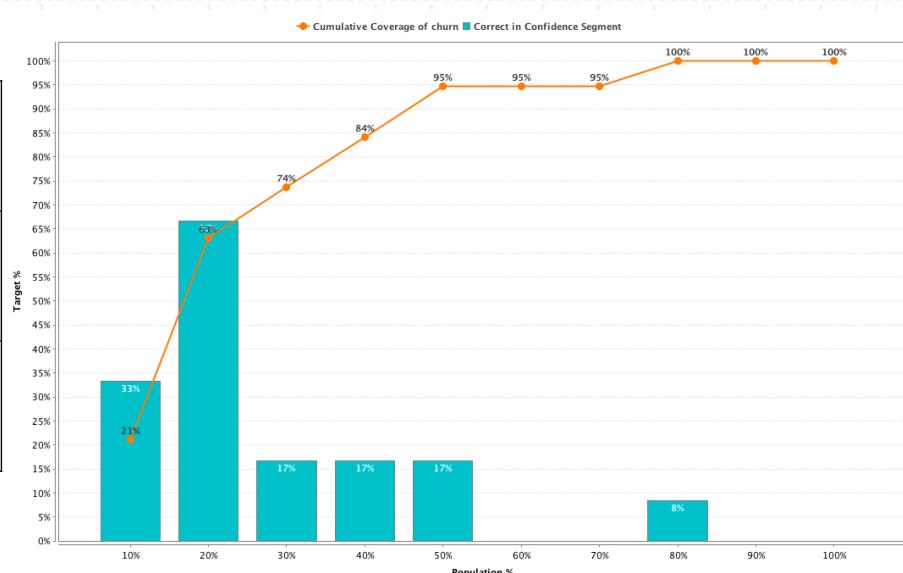
# CRISP-DM Example 2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## 5. Evaluation

- แบ่งข้อมูลเพื่อใช้ในการทดสอบ
- คำนวณค่าความถูกต้อง

	true loyal	true Churn
pred. loyal	375	57
pred. churn	21	20



# CRISP-DM Example 3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## ● 1. Business Understanding

- เป็นโครงการที่จัดทำขึ้นเพื่อ หมู่บ้านหรือพื้นที่ห่างไกลสามารถเข้าถึงบริการของภาครัฐผ่านทางอินเทอร์เน็ตได้
- ต้องสามารถตอบสนองนโยบายรัฐบาลกี่ต้องการให้คุณสมบัติความเร็วอินเทอร์ความเร็วสูงที่ 100 Mbps/50 Mbps
- ซึ่งพื้นที่การดำเนินการได้ถูกกำหนดโดยคุณทำงานพิจารณาทำหน้าที่เป้าหมายให้โครงการยกระดับโครงสร้างพื้นฐานโทรศัพท์มือถือเพื่อขับเคลื่อนเศรษฐกิจของประเทศไทย

ที่มา: การแบ่งกลุ่มหมู่บ้านตามพฤติกรรมการใช้งานอินเทอร์เน็ตประชากร, เอกสารที่พัฒนาศักยภาพ

# CRISP-DM Example 3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## ● 1. Business Understanding

- ต้องการแบ่งกลุ่มหมู่บ้านตามพฤติกรรมการใช้งานของแต่ละหมู่บ้านตามพฤติกรรมการใช้งาน
- ทำให้สามารถวางแผนการเพิ่มจุดส่งสัญญาณหรือการบริการในกลุ่มหมู่บ้านที่มีการใช้งานเยอะๆ ได้อย่างมีประสิทธิภาพ
- เพิ่มการประชาสัมพันธ์ให้กับคนในหมู่บ้านมากขึ้นในกรณีที่เป็นกลุ่มหมู่บ้านที่มีการใช้งานค่อนข้างน้อย

ที่มา: การแบ่งกลุ่มหมู่บ้านตามพฤติกรรมการใช้งานอินเทอร์เน็ตประชากร, เอกสารที่พัฒนาศักยภาพ

# CRISP-DM Example 3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **2. Data Understanding**

- ข้อมูลได้มาจากการ [https://data.go.th/dataset/village\\_internet\\_usage](https://data.go.th/dataset/village_internet_usage) มีตัวแปร

ต่อไปนี้

#	ชื่อตัวแปร	ความหมาย
1	YEAR	ปี
2	MONTH	เดือน
3	CIRCUIT_NO	เลขหมายเลขวงจร
4	VILLAGE	ชื่อหมู่บ้าน
5	SUBDISTRICT	ตำบล
6	PROVINCE	จังหวัด
7	UPLOAD	ปริมาณการนำข้อมูลเข้า (เมกกะไบต์)
8	DOWNLOAD	ปริมาณการนำข้อมูลออก (เมกกะไบต์)
9	NUM_USERS	จำนวนผู้ใช้งาน (ครั้ง)
10	NUM_DEVICES	จำนวนอุปกรณ์
11	LATITUDE	ละติจูดของหมู่บ้าน
12	LONGITUDE	ลองจิจูดของหมู่บ้าน

# CRISP-DM Example 3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **3. Data Understanding**

- เลือกตัวแปรที่จะใช้ในการวิเคราะห์ข้อมูลต่อไป ได้แก่
  - VILLAGE
  - SUBDISTRICT
  - PROVINCE
  - UPLOAD
  - DOWNLOAD
  - NUM\_USERS
  - NUM\_DEVICES

# CRISP-DM Example 3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **4. Modeling**

- ทำการแบ่งกลุ่มหมู่บ้านตามพฤติกรรมการใช้งาน
- โดยทำการแบ่งกลุ่มออกเป็น 5 กลุ่ม ตามค่าตัวแปร
  - UPLOAD
  - DOWNLOAD
  - NUM\_USERS
  - NUM\_DEVICES

ที่มา: การแบ่งกลุ่มหมู่บ้านตามพฤติกรรมการใช้งานอันเตอร์เน็ตประชาธิรัฐ , เอกสิเกิล พิชร Wongศักดา

# CRISP-DM Example 3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- **5. Evaluation**

- ผลการแบ่งกลุ่มได้พฤติกรรมของตัวแทนแต่ละกลุ่มดังนี้

	Group 1	Group 2	Group 3	Group 4	Group 5
UPLOAD	12,013.70	47,747.22	4,238.52	30,364.81	20,118.71
DOWNLOAD	193,306.38	836,339.62	65,563.37	508,271.65	330,744.67
NUM_USERS	1,126.28	3,237.97	582.14	2,115.58	1,577.37
NUM_DEVICES	105.73	180.98	84.02	131.98	117.49
NO. EXAMPLES	7,797	467	7,909	2,808	5,719
LABEL	LOW	VERY HIGH	VERY LOW	HIGH	MEDIUM

ที่มา: การแบ่งกลุ่มหมู่บ้านตามพฤติกรรมการใช้งานอันเตอร์เน็ตประชาธิรัฐ , เอกสิเกิล พิชร Wongศักดา

# CRISP-DM Example 3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

## • 6. Deployment

- วางแผนเพิ่มประสิทธิภาพให้กับหมู่บ้านในกลุ่มที่มีการใช้งานอินเตอร์เน็ต เป็นจำนวนมาก เช่น กลุ่ม VERY HIGH และ กลุ่ม HIGH
- วางแผนประชาสัมพันธ์ให้ผู้ที่อยู่ในกลุ่มหมู่บ้านที่มีการใช้งานน้อย เช่น กลุ่ม LOW และ VERY LOW เพื่อให้มีการใช้งานเพิ่มมากขึ้น
- หาสาเหตุของกลุ่มหมู่บ้านที่มีผู้ใช้งานน้อย