



# Data Science & Artificial Intelligence

## Train the trainer

December 14 - 15, 2019

# ลิงค์สู่การอบรมโปรแกรม

<https://bit.ly/2PJ5PII>



[https://colab.research.google.com/drive/12WmaiGS0rMVZFFuiC XKj3hpEIYm\\_BP7C](https://colab.research.google.com/drive/12WmaiGS0rMVZFFuiC XKj3hpEIYm_BP7C)

<https://tinyurl.com/ybn763lq>



<https://colab.research.google.com/drive/1ya6xAxtQd4xuhGwgU27vmU9uwt9mGh1A>

<https://bit.ly/2Lhdc00>



[https://drive.google.com/file/d/1ybO-SLcPl-tc1Lz-f\\_sQDj1aSh\\_Abt3k/view](https://drive.google.com/file/d/1ybO-SLcPl-tc1Lz-f_sQDj1aSh_Abt3k/view)

<https://bit.ly/2rBS6B2>



[https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb#scrollTo=qIKSN\\_z0W1yF](https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb#scrollTo=qIKSN_z0W1yF)

# Link เอกสารประกอบการอบรม

---

- [Machine Learning with Python](#)
- [Python for Data Analysis](#)
- [Python Data Science Handbook](#)
- [Colab Python Data Science](#)
- [Image Processing Neural Network](#)
- [Image Processing HAAR CASCADE](#)
- <https://medium.com/@jimmyspang/data-science-in-a-nutshell-db351fa41c64>
- <https://rapidminer.com/>

# หนังสือที่ใช้ในการสอน

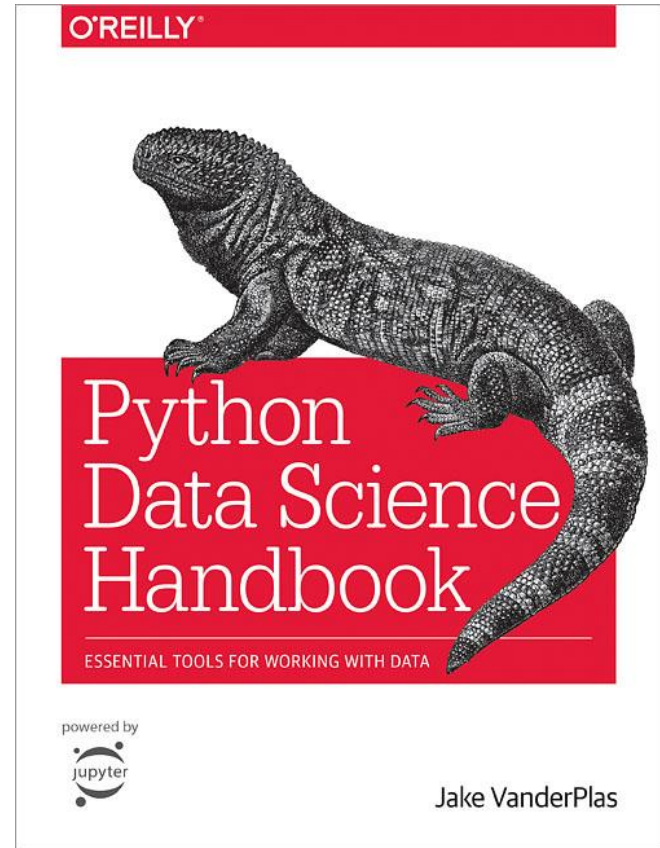
- Python Data Science Handbook

<https://bit.ly/2Lhdc00> 

<https://bit.ly/2rBS6B2>

[https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb#scrollTo=qIKSN\\_z0W1yF](https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb#scrollTo=qIKSN_z0W1yF)

<https://github.com/jakevdp/PythonDataScienceHandbook/tree/master/notebooks>





---

Hello, Google Colaboratory

สวัสดี รู้จัก และใช้งาน กูเกิ้ล โคลแลบ

---

# รู้จัก Google Colab

# รู้จัก Google Colab (1)

---

Google

+

=

colab

2014



Jupyter

# รู้จัก Google Colab (2)



เขียนโปรแกรมภาษาไพทอนบนอินเทอร์เน็ต

- มีทรัพยากรคอมพิวเตอร์ให้ใช้ในการคำนวณฟรี
- เพียงแค่มีอินเทอร์เน็ตก็สามารถใช้งานได้เลย



**NetBeans**



---

# ใช้งาน Google Colab

# ใช้งาน Google Colab (1)

---

1. Google Account (Gmail)
2. Google Colab

การใช้งาน Google Colab เบื้องต้น  
การนำเข้าซอร์สโค้ดจากภายนอก



colab

# การสมัคร Google Account (Gmail)

---

□ เข้าไปที่ [www.google.com/gmail/](http://www.google.com/gmail/)



ความสะดวกและ  
เรียบง่ายของ  
Gmail พร้อมใช้  
งานบนทุก  
อุปกรณ์

เลือก "สร้างบัญชี"

## กรอกข้อมูล

ชื่อ นามสกุล  
aiat eec

ชื่อผู้ใช้ @gmail.com  
aiat.eecth

คุณใช้ตัวอักษร ตัวเลข และจุดได้

รหัสผ่าน ยืนยัน  
..... .....

ใช้อักษร 8 ตัวขึ้นไปที่มีทั้งตัวอักษร ตัวเลข และสัญลักษณ์ผสมกัน

ลงชื่อเข้าใช้แทน

ถัดไป



หมายเลขโทรศัพท์ (ไม่บังคับ)

เราจะใช้หมายเลขโทรศัพท์ของคุณสำหรับรักษาความปลอดภัยของบัญชี และผู้อื่นจะไม่เห็นหมายเลขนี้

ที่อยู่อีเมลสำรอง (ไม่บังคับ)

เราจะใช้ที่อยู่อีเมลนี้เพื่อรักษาบัญชีของคุณให้ปลอดภัย

วัน เดือน ปี  
1 มกราคม 2000

วันเกิดของคุณ

เพศ

ไม่ต้องการตอบ

เหตุผลที่เราขอข้อมูลนี้

กลับ

ถัดไป

## ความเป็นส่วนตัวและข้อกำหนด

กับพาร์ทเนอร์โฆษณาและการวัดผลที่เจาะจงเหล่านี้

### รวมข้อมูล

เรายังรวมข้อมูลนี้ไว้ในบริการต่างๆ ของเราและในอุปกรณ์ต่างๆ ของคุณเพื่อจุดประสงค์เหล่านี้ด้วย เช่น ขึ้นอยู่กับการตั้งค่าบัญชีของคุณ เราแสดงโฆษณาให้คุณรับชม โดยอิงจากข้อมูลเกี่ยวกับสิ่งที่คุณสนใจ ซึ่งเรานำมาจากการใช้ Search และ YouTube ของคุณ และเรา ใช้ข้อมูลจากคำค้นหาหลายล้านล้านรายการเพื่อสร้างโมเดลแก้ไขการสะกดคำที่ใช้ในบริการทั้งหมดของเรา

### คุณสามารถดู

ขึ้นอยู่กับการตั้งค่าบัญชี ข้อมูลบางส่วนอาจจะเกี่ยวข้องกับบัญชี Google ของคุณ เราจะถือว่าข้อมูลเหล่านี้เป็นข้อมูลส่วนบุคคล คุณจะควบคุมวิธีการเก็บรวบรวมและใช้ข้อมูลของเราได้โดยคลิกที่ “ตัวเลือกเพิ่มเติม” ด้านล่าง คุณจะปรับเปลี่ยนการควบคุมหรือถอนคำยินยอมสำหรับในอนาคตได้ทุกเมื่อด้วยการไปที่บัญชีของฉัน (myaccount.google.com)

อ่าน และยอมรับข้อกำหนด

ยกเลิก

ฉันยอมรับ

เขียน

กล่องจดหมาย

ติดตาม

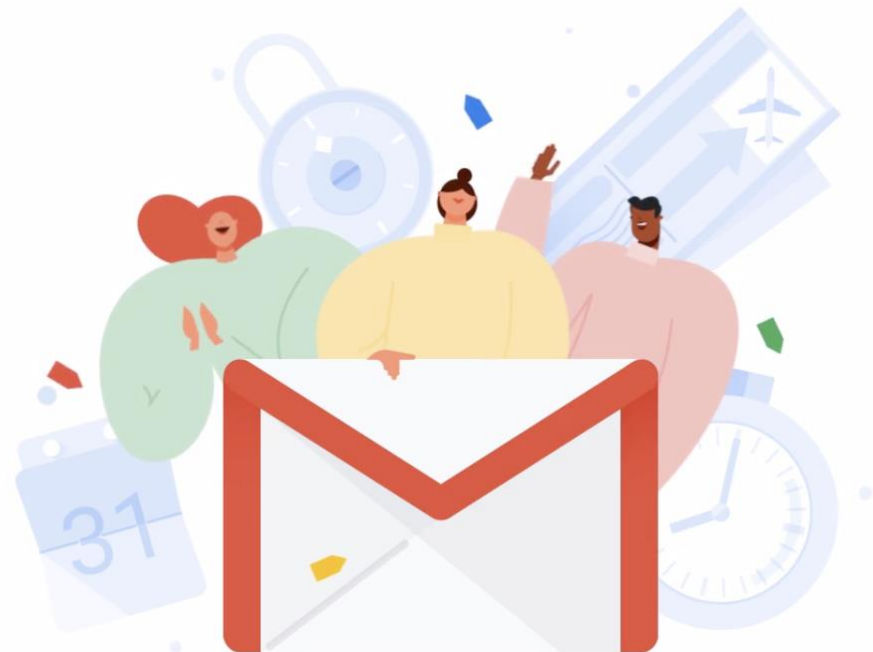
ปิดเสียงเตือนชั่วคราว

ส่งแล้ว

ร่างจดหมาย

เพิ่มเติม

aiat



## ยินดีต้อนรับ

อีเมลโดย Google: ง่าย มีประสิทธิภาพ และปลอดภัย

ถัดไป

ไม่มีแชทล่าสุด  
เริ่มใหม่เลย

Gmail

เขียน

กล่องจดหมาย

ติดตาม

ปิดเสียงเตือนชั่วคราว

ส่งแล้ว

ร่างจดหมาย

เพิ่มเติม

aiat

|                          |   |  |  |
|--------------------------|---|--|--|
| <input type="checkbox"/> | ☆ |  |  |
| <input type="checkbox"/> | ☆ |  |  |
|                          |   |  |  |
| <input type="checkbox"/> | ☆ |  |  |
| <input type="checkbox"/> | ☆ |  |  |
|                          |   |  |  |
| <input type="checkbox"/> | ☆ |  |  |
| <input type="checkbox"/> | ☆ |  |  |

### เลือกมุมมอง

- ค่าเริ่มต้น
- สบาย
- กะทัดรัด

ตกลง

< > ⚙️ 31

ส 23:07

รูปใช้ Gmail

โปรไฟล์

ไม่มีแชทล่าสุด  
เริ่มใหม่เลย



✚ เขียน

📄 **กล่องจดหมาย** 1

★ ติดตาม

🕒 ปิดเสียงเตือนชั่วคราว

▶ ส่งแล้ว

📄 ร่างจดหมาย

👤 aiat +

☐ ↻ ⋮

1-1 จาก 1 < > ⚙️

☐ ☆ Patcharin from Goog. คุณ aiat ยินดีต้อนรับสู่บัญชี Google ใหม่ - ส 23:07

|   |                                    |                                 |
|---|------------------------------------|---------------------------------|
| <p>10%</p> <p>ความคืบหน้าในการติดตั้ง</p> | <p>สร้างบัญชีแล้ว</p>              | <p>เรียนรู้วิธีการใช้ Gmail</p> |
|   | <p>โหลด Gmail มือถือ</p>           | <p>เลือกธีม</p>                 |
|   | <p>นำเข้ารายชื่อติดต่อและอีเมล</p> | <p>เปลี่ยนรูปโปรไฟล์</p>        |

ใช้ไป 0 GB (0%) จาก 15 GB  
จัดการ

ข้อกำหนด · ความเป็นส่วนตัว · นโยบาย  
โปรแกรม

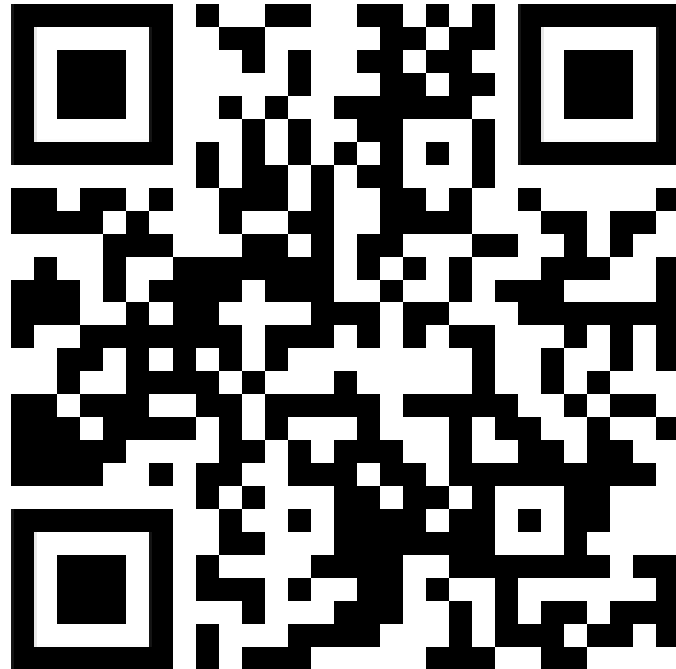
🗨️  
ไม่มีแชทล่าสุด  
เริ่มใหม่เลย

**พร้อมใช้งาน**

# ใช้งาน Google Colab (2)

---

□ [colab.research.google.com/](https://colab.research.google.com/)



EXAMPLES

RECENT

GOOGLE DRIVE

GITHUB

UPLOAD

Filter notebooks



Title

First opened

Last opened



Hello, Colaboratory

0 minutes ago

0 minutes ago



NEW PYTHON 3 NOTEBOOK



CANCEL

# การใช้งาน Google Colab เบื้องต้น

---

□ สร้าง Notebook ใหม่

NEW PYTHON 3 NOTEBOOK ▼

CANCEL



```
|
```



หน้าตาของ Google Colab

# การเชื่อมต่อกับ Server ของ Google Colab



Untitled0.ipynb

COMMENT

SHARE

a

File Edit View Insert Runtime Tools Help

+ CODE + TEXT

↑ CELL ↓ CELL

✓ CONNECTED ▼

EDITING



Connect to hosted runtime

ถ้าสถานะไม่ใช่ **Connected**

- ให้คลิกที่ ▼ ด้านขวาของสถานะ
- เลือก Connect to hosted runtime

สถานะต้องเป็น **Connected** เสมอ

ชื่อไฟล์ (ต้องลงท้ายด้วย .ipynb)

เมนู

สถานะการเชื่อมต่อกับ Server

Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help

COMMENT

SHARE

a

+ CODE + TEXT ↑ CELL ↓ CELL

CONNECT ▾

EDITING

^

เซลล์

เมนูจัดการเซลล์

1. เพิ่มเซลล์โค้ด (Code Cell)
3. เลื่อนเซลล์ที่ถูกเลื่อนอยู่ขึ้น

2. เพิ่มเซลล์ของตัวอักษร (Text Cell)
4. เลื่อนเซลล์ที่ถูกเลือกอยู่ลง



Text Cell

Text Cell

เซลล์ตัวอักษร (Text Cell)



Text Cell



Text Cell

เซลล์ของโค้ด (Code Cell)

```
# Code cell 1|  
print("Hello, World.")
```

```
[ ] # Code cell 2  
print("Hi, World.")
```

- คลิกเลือกที่ Code Cell ที่ต้องการ Run
- คลิกเครื่องหมาย ▶ เพื่อ Run Code Cell นั้น



Text Cell

## การลบเซลล์



```
# Code cell 1  
print("Hello, World.")
```



Link to cell

Delete cell

⌘/Ctrl+M D

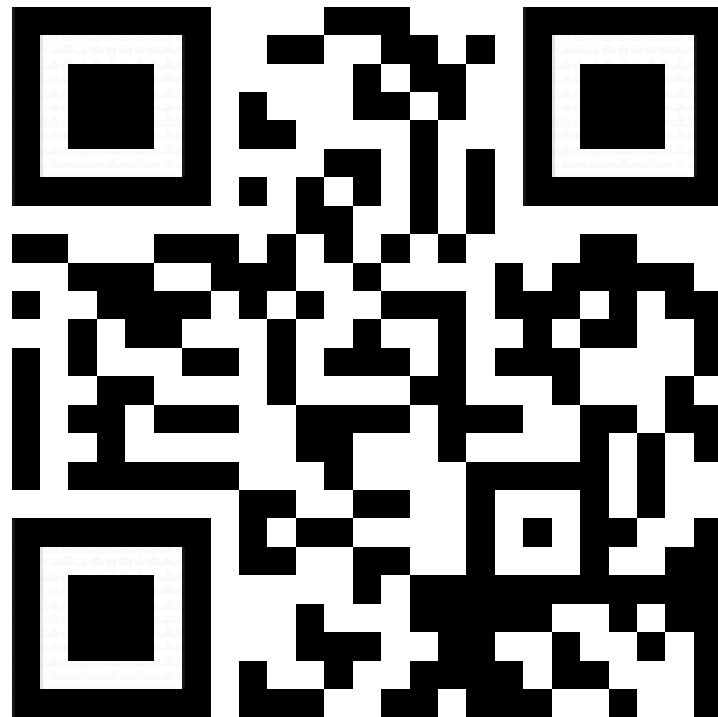
- คลิกเลือกเซลล์ที่ต้องการลบ
- คลิกเลือกที่ จุดสามจุด แนวตั้ง
- คลิก Delete cell เพื่อลบเซลล์

# การนำเข้าซอร์สโค้ดจากภายนอก (กรณีตัวอย่าง)

---

□ ได้รับ URL Link เพื่อมาเปิดใน Google Colab ของตัวเอง

□ <http://bit.ly/2E9wu6S>





# Python101 x Scipy: Tutorial 1 - Welcome to Python World!

## *Tutorial 1 (T1)*

Let's see what we can do with Python.

## ▾ Let's PRINT!

```
▶ print("Hello, world.")
```

```
[ ] print("Hello, %s." % ("Python101"))
```



# Python101 x Scipy Python World!

## Tutorial 1 (T1)

Let's see what we can do with Python.

Connect to hosted runtime

Connect to local runtime...

เชื่อมต่อไปยัง Server ของ Google Colab

- ให้คลิกที่ ▼ ด้านขวาของสถานะ
- เลือก Connect to hosted runtime



# Python101 x Scipy: Tutorial 1 - Welcome to Python World!

*Tutorial 1 (T1)*

Let's see what we can do with Python.

ทดลองรันโค้ดเซลล์

## Let's PRINT!

```
▶ print("Hello, world.")
```

## Warning: This notebook was not authored by Google.

This notebook is being loaded from [GitHub](#). It may request access to your data stored with Google, or read data and credentials from other sessions. Please review the source code before executing this notebook. To prevent this notebook reading state from other sessions, you can reset all runtimes.

Reset all runtimes before running

CANCEL

RUN ANYWAY

เลือก "RUN ANYWAY"

## Reset all runtimes

Are you sure you want to reset all runtimes? State of all runtimes, including all local variables and files, will be lost.

CANCEL

YES

เลือก "YES"





+ CODE

+ TEXT

↑ CELL

↓ CELL

COPY TO  
DRIVE

✓ CONNECTED

✎ EDITING



# Python101 x Scipy: T

## Python World!

*Tutorial 1 (T1)*

Let's see what we can do with Python.

### Let's PRINT!



```
print("Hello, world.")
```



Hello, world.

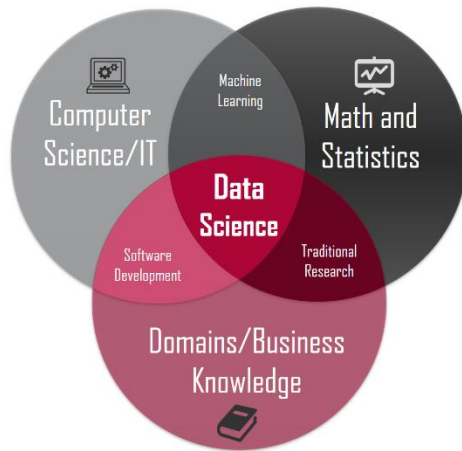
ผลลัพธ์จากโค้ดเซลล์จะถูกแสดงออกมา  
นอกจากนี้ เซลล์ทั้งหมดยังสามารถแก้ไข  
และรันได้อย่างอิสระ



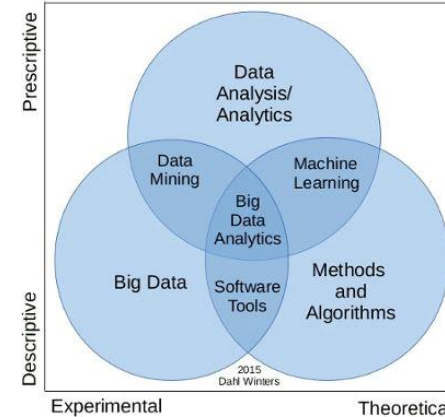
# Data Science (วิทยาการข้อมูล)

[https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science)

- ❑ Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data.
- ❑ Data science is the same concept as data mining and big data: "use the most powerful hardware, the most powerful programming systems, and the most efficient algorithms to solve problems".



The Fields of Data Science



<https://towardsdatascience.com/introduction-to-statistics-e9d72d818745>

<http://www.digitalvidya.com/wp-content/uploads/2017/04/fields-of-data-science-.jpg>

# Data Science vs. Big Data vs. Data Analytics

---

## □ Data Science

- Dealing with unstructured and structured data, Data Science is a field that comprises everything that related to data cleansing, preparation, and analysis.
- Combination of statistics, mathematics, programming, problem-solving, capturing data in ingenious ways, the ability to look at things differently, and the activity of cleansing, preparing and aligning the data.
- In simple terms, it is the umbrella of techniques used when trying to extract insights and information from data.

## □ Big Data

- Big Data refers to humongous volumes of data that cannot be processed effectively with the traditional applications that exist. The processing of Big Data begins with the raw data that isn't aggregated and is most often impossible to store in the memory of a single computer.
- A buzzword that is used to describe immense volumes of data, both unstructured and structured, Big Data inundates a business on a day-to-day basis. Big Data is something that can be used to analyze insights that can lead to better decisions and strategic business moves.
- The definition of Big Data, given by Gartner is, "Big data is high-volume, and high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."

## □ Data Analytics

- Data Analytics the science of examining raw data to draw conclusions about that information.
- Data Analytics involves applying an algorithmic or mechanical process to derive insights. For example, running through a number of data sets to look for meaningful correlations between each other.
- It is used in a number of industries to allow organizations and companies to make better decisions as well as verify and disprove existing theories or models.
- The focus of Data Analytics lies in inference, which is the process of deriving conclusions that are solely based on what the researcher already knows.

# Data Science vs. Big Data vs. Data Analytics

## Data Science VS Big Data VS Data Analytics

DATA IS GROWING FASTER THAN EVER BEFORE.



Each person-  
1.7 megabytes  
created



### WHAT ARE THEY?



**Data Science** is a field that comprises of everything that related to data cleansing, preparation, and analysis.



**Big Data** is something that can be used to analyze insights which can lead to better decision and strategic business moves.



**Data Analytics** Involves automating insights into a certain dataset as well as supposes the usage of queries and data aggregation procedures.

### WHERE ARE THEY USED?

Data Science algorithms are used in industries like:



Big Data is used in industries like:



Data Analytics is used in industries like:



## WHAT ARE THE SKILLS REQUIRED?



### DATA SCIENTIST

- In-depth knowledge in SAS and/or R
- Python coding
- Hadoop platform
- SQL database/coding
- Working with unstructured data

### BIG DATA SPECIALIST

- Analytical skills
- Creativity
- Mathematics and
- Statistical skills
- Computer science
- Business skills

### DATA ANALYST

- Programming skills
- Statistical skills
- Mathematics
- Machine learning skills
- Data wrangling skills
- Communication and Data Visualization skills
- Data Intuition

### DATA SCIENTIST

\$113,436  
per year.

### BIG DATA SPECIALIST

\$62,066  
per year.

### DATA ANALYST

\$60,476  
per year.

<https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>

# Data Science (วิทยาการข้อมูล)

The world is generating data at a higher rate, and so the need of "Data Science" & "Data Analytics" tools increases to analyze and manage this "Big Data".



## Data Science vs. Data Analytics vs. Big Data

WHAT IS DATA SCIENCE?    WHAT IS DATA ANALYTICS?    WHAT IS BIG DATA?

**Data Science** is a field that refers to the collective processes, theories, concepts, tools and technologies that enable the review, analysis and extraction of valuable knowledge and information from raw data.

**Data Analytics (DA)** is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems & software.

**Big Data** refers to voluminous amounts of structured or unstructured data that organizations can potentially mine & analyze for business gains.

### APPLICATION AREAS

1. Digital advertisements
2. Internet Research
3. Recommender System
4. Image/Speech Recognition

1. Gaming
2. Travel
3. Energy Management
4. Healthcare

1. Communication
2. Retail
3. Financial services
4. Education

### TOOLS & LANGUAGES

1. Python
2. SAS
3. SQL

1. R
2. Tableau Public
3. Apache Spark

1. Hadoop
2. NoSQL
3. Hive

### ANNUAL SALARY

Data Scientist  
**\$130,323**

Big Data Specialist  
**\$69,845**

Data Analyst  
**\$62,066**



### Data Scientist

also known as Data Managers, statisticians.

A data scientist will be able to take data science projects from end to end. They can help store large amounts of data, create predictive modelling processes and present the findings.

**Skills:** Mathematics, Programming, Communication

Will use programmes such as: SQL, Python, R

### Data Engineers

also known as database administrators and data architects.

They are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists.

**Skills:** Programming, Mathematics, Big data

Will use programmes such as: Hadoop, NoSQL, and Python

### Data Analysts

also known as business Analysts.

They typically help people from across the company understand specific queries with charts.

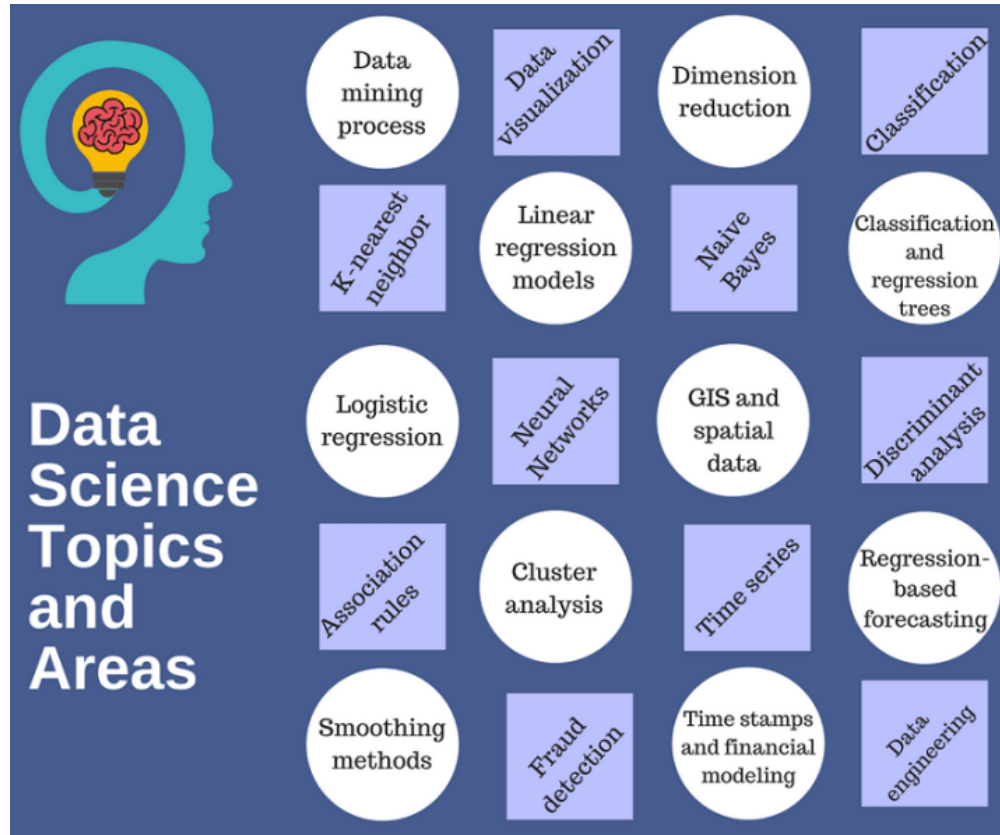
**Skills:** Statistics, Communication, Business knowledge

Will use programmes such as: Excel, Tableau, SQL

<https://elunl/careers-in-data-science-data-analyst-vs-data-engineer-vs-data-scientist/>

<https://i.pinimg.com/originals/41/a5/c7/41a5c71c19a5fb5b379be91d3e5ca4e3.png>

# Data Science Topics (หัวข้อในวิทยาการข้อมูล)



# Modern Data Scientist (นักวิทยาศาสตร์ข้อมูลสมัยใหม่)

## MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

## DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



## PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

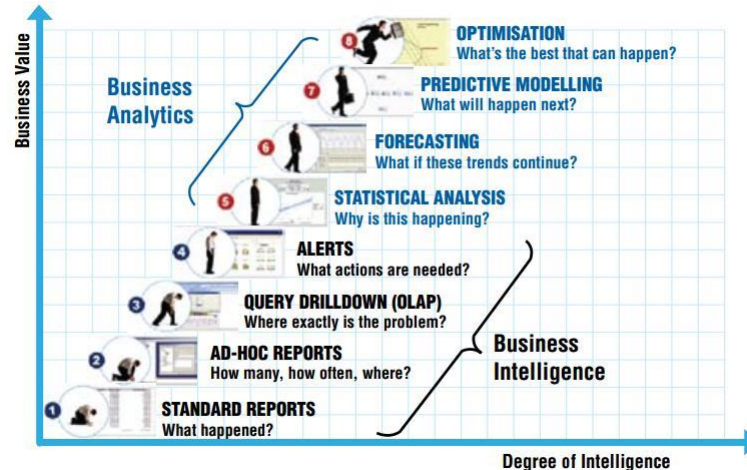
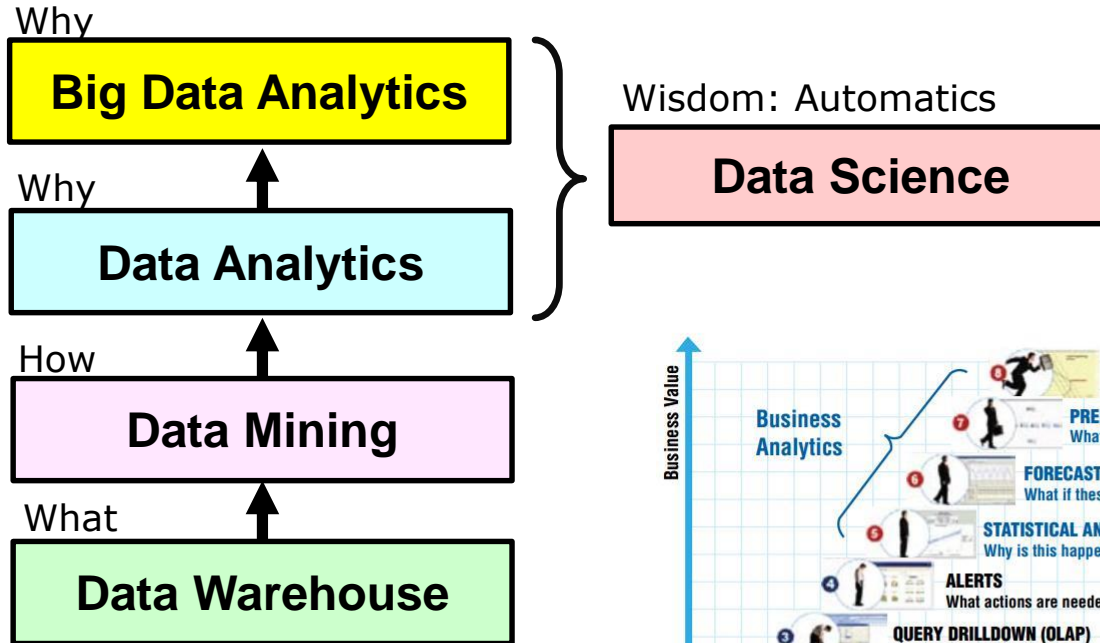
## COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

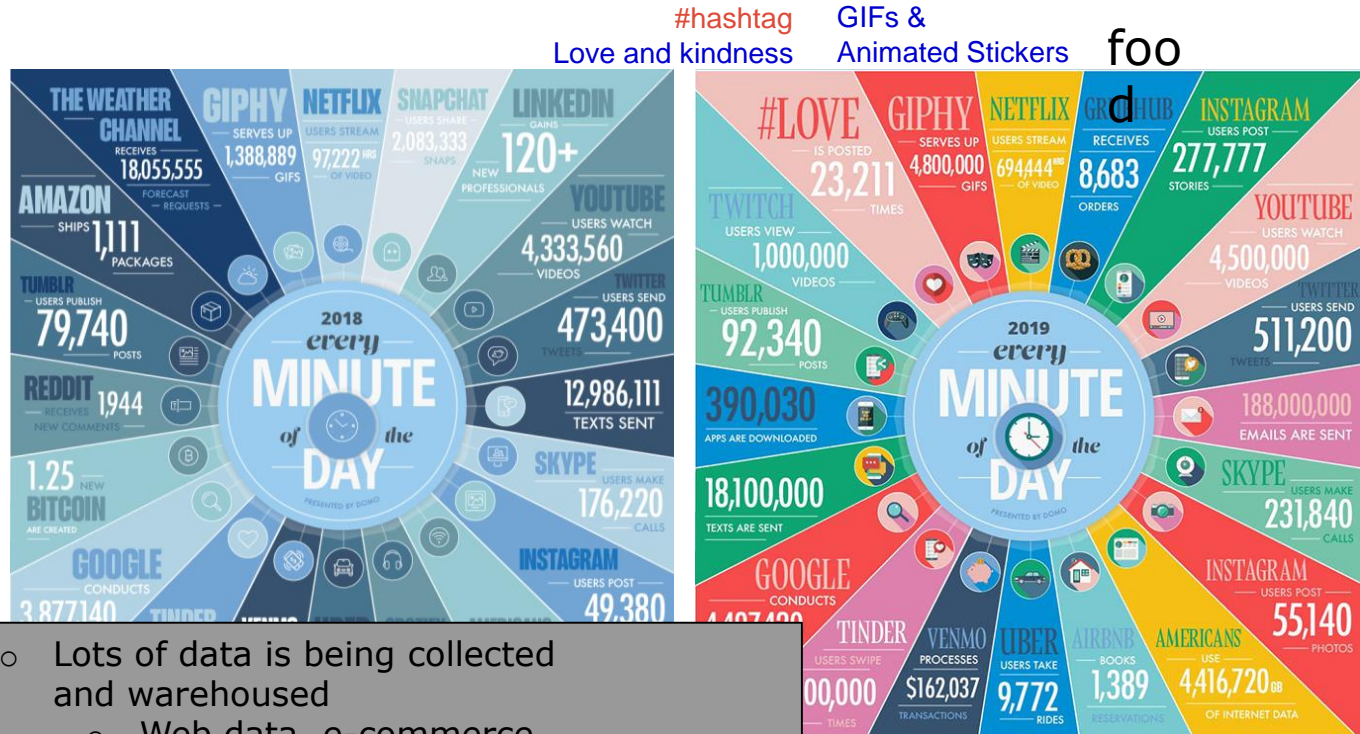
<https://www.themarketingdistillery.com/>



# Data Warehouse vs. Data Mining vs. Data Science vs. Data Analytics vs. Big Data Analytics



# ข้อมูลใหญ่ขนาดไหน (How large data is)



- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - Financial transactions, bank/credit transactions
  - Online trading and purchasing
  - Social Network

Share payments

2019

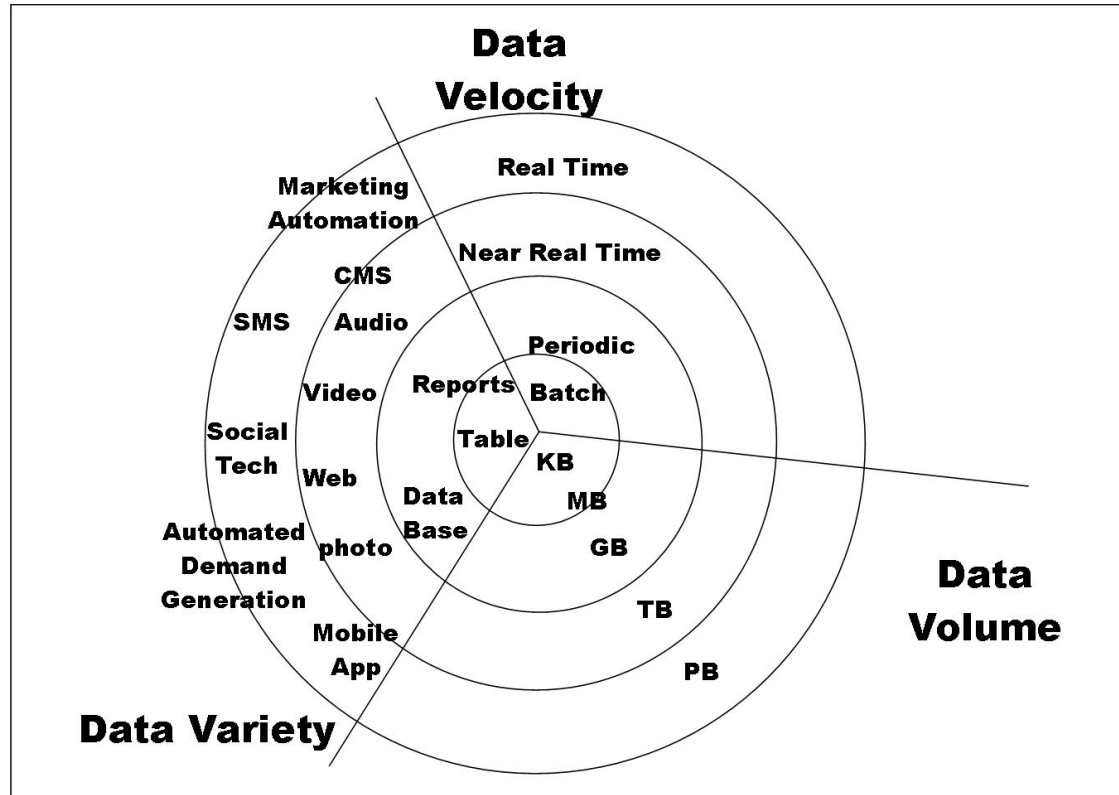
(Source: <https://www.domo.com>)

# Big Data in three views (I)

---

- ◆ Big Data is any data that is expensive to manage and hard to extract value from
  - Volume
    - The size of the data
  - Velocity
    - The latency of data processing relative to the growing demand for interactivity
  - Variety and Complexity
    - the diversity of sources, formats, quality, structures.

# Big Data in three views (II)



# Data types

---

- ❑ Relational Data (Tables/Transaction/Legacy Data)
- ❑ Text Data (Web)
- ❑ Semi-structured Data (XML)
- ❑ Graph Data
- ❑ Social Network, Semantic Web (RDF)
- ❑ Spatiotemporal (Geographical) data
- ❑ Biomedical data
- ❑ Streaming Data
- ❑ Smell
- ❑ Taste
- ❑ Sense of Touch

# Actions to do with data

---

- Aggregation and Statistics
  - Data warehousing and OLAP
- Indexing, Searching, and Querying
  - Keyword based search
  - Pattern matching (XML/RDF)
- Knowledge discovery
  - Data Mining
  - Statistical Modeling
  - Big Data Analytics

# Skills in Data Science

---

- ❑ Mathematics and Applied Mathematics
- ❑ Applied Statistics/Data Analysis
- ❑ Solid Programming Skills (R, Python, Julia, SQL)
- ❑ Data Mining
- ❑ Data Base Storage and Management
- ❑ Machine Learning and discovery

# Data Structure



List vs. Array vs. Tuple



# ลิงค์สู่การอบรมโปรแกรม

<https://bit.ly/2PJ5PII>



[https://colab.research.google.com/drive/12WmaiGS0rMVZFFuiC XKj3hpEIYm\\_BP7C](https://colab.research.google.com/drive/12WmaiGS0rMVZFFuiC XKj3hpEIYm_BP7C)

<https://tinyurl.com/ybn763lq>



<https://colab.research.google.com/drive/1ya6xAxtQd4xuhGwgU27vmU9uwt9mGh1A>

<https://bit.ly/2Lhdc00>



[https://drive.google.com/file/d/1ybO-SLcPl-tc1Lz-f\\_sQDj1aSh\\_Abt3k/view](https://drive.google.com/file/d/1ybO-SLcPl-tc1Lz-f_sQDj1aSh_Abt3k/view)

<https://bit.ly/2rBS6B2>



[https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb#scrollTo=qIKSN\\_z0W1yF](https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb#scrollTo=qIKSN_z0W1yF)

# List vs. Tuple vs. Array

- List1 = [1, 2, 3, 4]
- Tuple1 = (1, 2, 3, 4)
- Array1 = numpy.array([1, 2, 3, 4]) = [ 1 2 3 4 ]

| List vs Tuple  |  |
|--|--|
| A list is a compound data type in Python programming language which can store different type of data and can change elements once created. | A tuple is a compound data type in Python programming language which can store different type of data and cannot change elements once created. |
| Mutability   |  |
| A list is mutable. It can be changed once created.   | A tuple is immutable. It cannot be changed once created.   |
| Enclosing Elements   |  |
| The elements of a list are enclosed in square brackets.  | The elements of a tuple are enclosed in parenthesis.   |
| Speed  |  |
| Iterating through elements in a list is not fast as in a tuple.  | Iterating through elements in a tuple is faster than list.   |

```
Arrays vs Lists

In [4]: from numpy import array

In [5]: cost = array([4,8,12,16,20,100,120,60])

In [6]: divided_cost = cost/2

In [7]: print(divided_cost)
[ 2  4  6  8 10 50 60 30]

In [8]: cost2 = [4,8,12,16,20,100,120,60]

In [9]: divided_cost2 = cost2/2

-----
TypeError                                 Traceback (most recent call last)
<ipython-input-9-479c34722c0c> in <module>()
----> 1 divided_cost2 = cost2/2

TypeError: unsupported operand type(s) for /: 'list' and 'int'

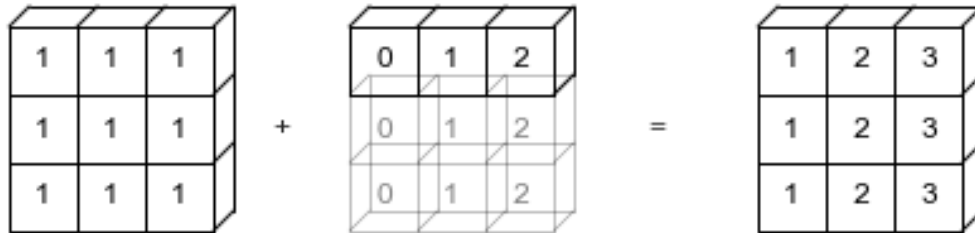
In [ ]:
```

# Computation on Arrays: Broadcasting

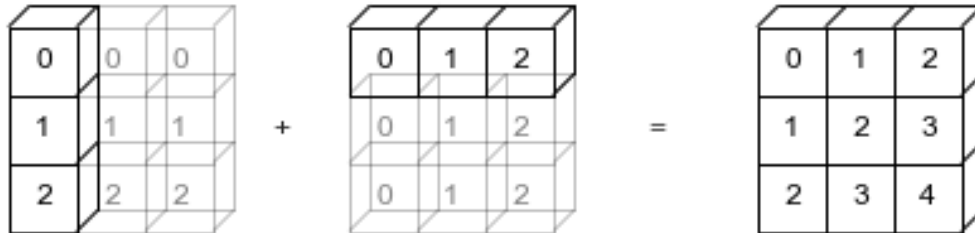
`np.arange(3)+5`



`np.ones((3,3))+np.arange(3)`



`np.arange(3).reshape((3,1))+np.arange(3)`

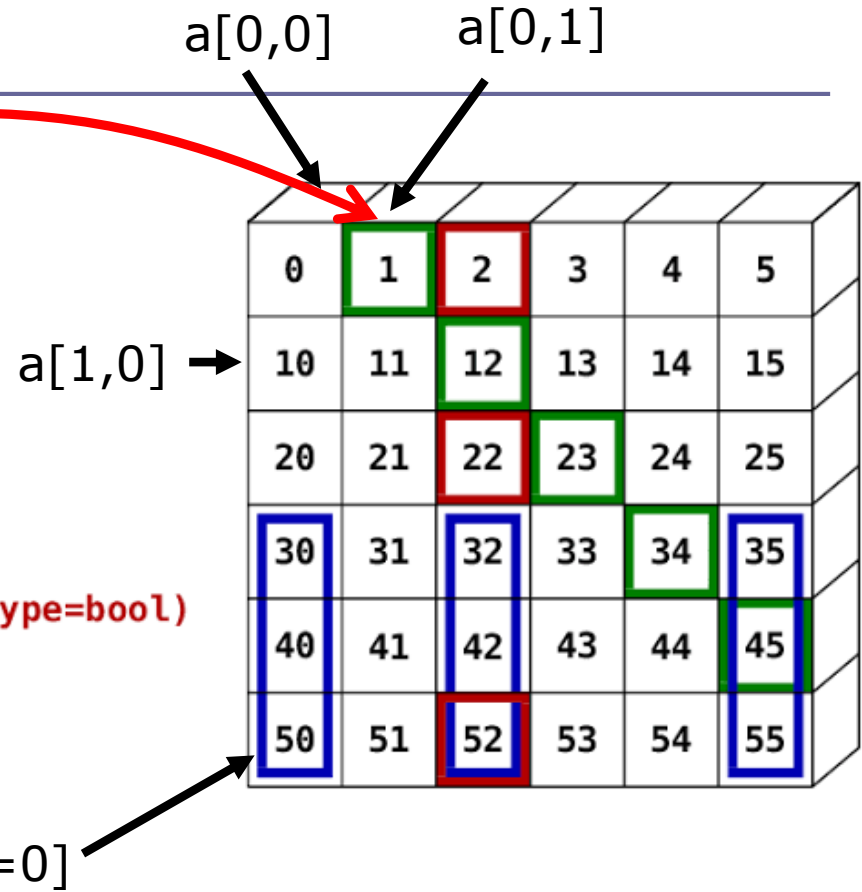


# array

```
>>> a[(0,1,2,3,4), (1,2,3,4,5)]  
array([1, 12, 23, 34, 45])
```

```
>>> a[3:, [0,2,5]]  
array([[30, 32, 35],  
       [40, 42, 45],  
       [50, 52, 55]])
```

```
>>> mask = np.array([1,0,1,0,0,1], dtype=bool)  
>>> a[mask, 2]  
array([2, 22, 52])
```



# Numpy



How to use Numpy

# Introduction to NumPy

---

## 2. Introduction to NumPy.

- [Understanding Data Types in Python](#)
- [The Basics of NumPy Arrays](#)
- [Computation on NumPy Arrays: Universal Functions](#)
- [Aggregations: Min, Max, and Everything In Between](#)
- [Computation on Arrays: Broadcasting](#)
- [Comparisons, Masks, and Boolean Logic](#)
- [Fancy Indexing](#)
- [Sorting Arrays](#)
- [Structured Data: NumPy's Structured Arrays](#)

# Pandas



How to use Pandas

# Data Manipulation with Pandas

---

## 3. Data Manipulation with Pandas

- [Introducing Pandas Objects](#)
- [Data Indexing and Selection](#)
- [Operating on Data in Pandas](#)
- [Handling Missing Data](#)
- [Hierarchical Indexing](#)
- [Combining Datasets: Concat and Append](#)
- [Combining Datasets: Merge and Join](#)
- [Aggregation and Grouping](#)
- [Pivot Tables](#)
- [Vectorized String Operations](#)
- [Working with Time Series](#)
- [High-Performance Pandas: eval\(\) and query\(\)](#)
- [Further Resources](#)




# Pandas – Data Frame

Diagram illustrating a Pandas Data Frame structure. The columns are labeled **Columns** and the rows are labeled **Rows**. The data is organized into a table with 7 rows and 6 columns.

|   | <i>Name</i>     | <i>Team</i>    | <i>Number</i> | <i>Position</i> | <i>Age</i> |
|---|-----------------|----------------|---------------|-----------------|------------|
| 0 | Avery Bradley   | Boston Celtics | 0.0           | PG              | 25.0       |
| 1 | John Holland    | Boston Celtics | 30.0          | SG              | 27.0       |
| 2 | Jonas Jerebko   | Boston Celtics | 8.0           | PF              | 29.0       |
| 3 | Jordan Mickey   | Boston Celtics | NaN           | PF              | 21.0       |
| 4 | Terry Rozier    | Boston Celtics | 12.0          | PG              | 22.0       |
| 5 | Jared Sullinger | Boston Celtics | 7.0           | C               | NaN        |
| 6 | Evan Turner     | Boston Celtics | 11.0          | SG              | 27.0       |

The diagram highlights the structure of the data frame. The word "Columns" is positioned above the table with arrows pointing to the column headers. The word "Rows" is positioned to the left of the table with arrows pointing to the row indices. A pink box labeled "Data" is positioned below the table, with lines connecting it to the data cells in rows 2, 3, 4, and 5, illustrating the concept of data within the frame.



# Matplotlib



How to use Matplotlib

# Visualization with Matplotlib

---

## 4. Visualization with Matplotlib

- [Simple Line Plots](#)
- [Simple Scatter Plots](#)
- [Visualizing Errors](#)
- [Density and Contour Plots](#)
- [Histograms, Binnings, and Density](#)
- [Customizing Plot Legends](#)
- [Customizing Colorbars](#)
- [Multiple Subplots](#)
- [Text and Annotation](#)
- [Customizing Ticks](#)
- [Customizing Matplotlib: Configurations and Stylesheets](#)
- [Three-Dimensional Plotting in Matplotlib](#)
- [Geographic Data with Basemap](#)
- [Visualization with Seaborn](#)
- [Further Resources](#)

# Matplotlib – First Program

---

```
□ import numpy as np
□ import matplotlib.pyplot as plt
□ # Fixing random state for reproducibility
□ np.random.seed(19680801)
□ mu, sigma = 100, 15
□ x = mu + sigma * np.random.randn(10000)
□ # the histogram of the data
□ n, bins, patches = plt.hist(x, 50, density=True, facecolor='g', alpha=0.75)
□ plt.xlabel('Smarts')
□ plt.ylabel('Probability')
□ plt.title('Histogram of IQ')
□ plt.text(60, .025, r'$\mu=100,\ \sigma=15$')
□ plt.xlim(40, 160)
□ plt.ylim(0, 0.03)
□ plt.grid(True)
□ plt.show()
```

# Machine Learning



How to implement machine learning

# Machine Learning with Python

---

## 5. Machine Learning

- [What Is Machine Learning?](#)
- [Introducing Scikit-Learn](#)
- [Hyperparameters and Model Validation](#)
- [Feature Engineering](#)
- [In Depth: Naive Bayes Classification](#)
- [In Depth: Linear Regression](#)
- [In-Depth: Support Vector Machines](#)
- [In-Depth: Decision Trees and Random Forests](#)
- [In Depth: Principal Component Analysis](#)
- [In-Depth: Manifold Learning](#)
- [In Depth: k-Means Clustering](#)
- [In Depth: Gaussian Mixture Models](#)
- [In-Depth: Kernel Density Estimation](#)
- [Application: A Face Detection Pipeline](#)
- [Further Machine Learning Resources](#)

# ศาสตร์ที่เกี่ยวข้องกับปัญญาประดิษฐ์

วิทยาการคอมพิวเตอร์ (Computer Science, CS)  
วิศวกรรมคอมพิวเตอร์ (Computer Engineering, CPE)  
เทคโนโลยีสารสนเทศ (Information Technology, IT)  
วิศวกรรมดิจิทัล (Digital Engineering, DE)

ปัญญาประดิษฐ์  
(Artificial Intelligence, AI)

วิทยาการข้อมูล  
(Data Science, DS)

โครงสร้างพื้นฐานอัจฉริยะ  
(Intelligent Infrastructure Systems, IIS)

การเรียนรู้ของเครื่อง  
(Machine Learning, ML)

การวิเคราะห์ข้อมูล  
(Data Analytics, DA)

ระบบไซเบอร์-กายภาพ  
(Cyber-Physical System, CPS)

การเรียนรู้เชิงลึก  
(Deep Learning, DL)

การวิเคราะห์ข้อมูลขนาดใหญ่  
(Big Data Analytics, BDA)

การประมวลผลแบบกลุ่มเมฆ  
(Cloud computing, CC)

# ศาสตร์ที่ได้รับความสนใจและเกี่ยวข้องกับเอไอ

ปัญญาประดิษฐ์

วิทยาการข้อมูล  
การวิเคราะห์ข้อมูล

อินเทอร์เน็ตของสรรพสิ่ง

หุ่นยนต์และการควบคุม

ชีวภาพและชีววิทยา

Materials &  
Nano Tech



# ปัญญาประดิษฐ์ (AI)

เครือข่ายเส้นประสาท  
(Neural Network)

การค้นหา (Search)

การอนุมาน (Inference)

การให้เหตุผล (Reasoning)

การแก้ปัญหาภายใต้ข้อจำกัด  
(Constraint Satisfaction Problems)

การตัดสินใจ (Decision Making)

การวางแผน (Planning)

การเรียนรู้ (Learning)

เอเจนต์/ตัวแทน (Agent)

ตรรก (Logic)

น่าจะเป็น (Probability)

ความรู้ (Knowledge)

ความคลุมเครือ (Fuzziness)

การประมวลผลสัญญาณ  
(Signal Processing)

การประมวลผลภาพ  
(Image Processing)

การประมวลผลภาษาธรรมชาติ  
(Natural Language Processing)

## การเรียนรู้ของเครื่อง (Machine Learning: ML)

---

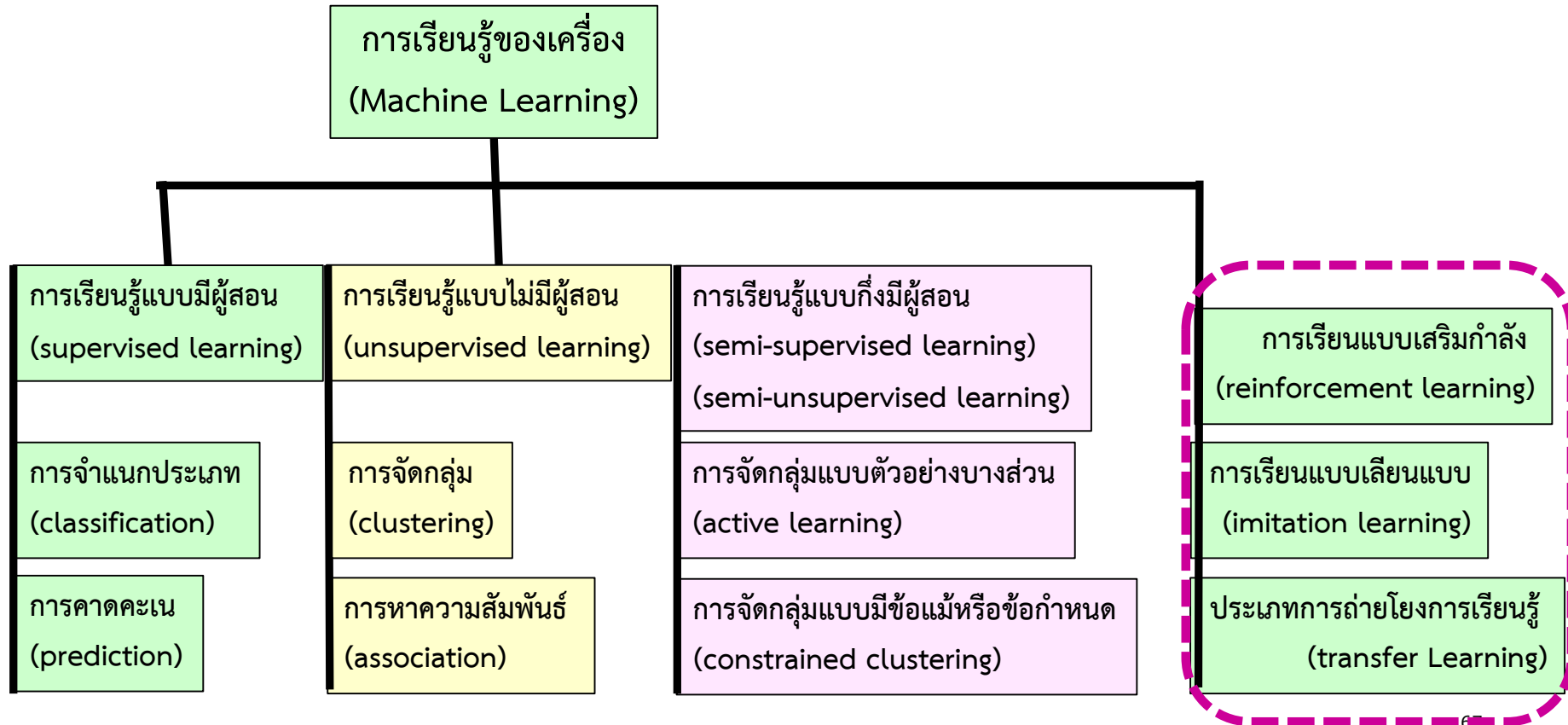
□ สาขาหนึ่งของปัญญาประดิษฐ์

□ การสร้างปัญญาประดิษฐ์ด้วยการใช้ข้อมูล

- สร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้
- แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์

[https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)

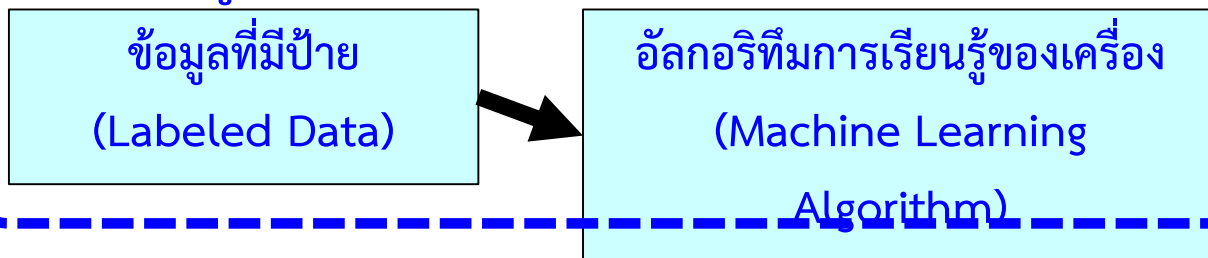
# ประเภทของการเรียนรู้ของเครื่อง (ML Types)



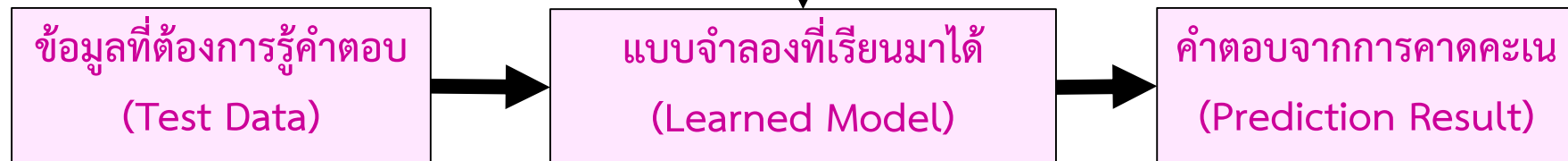
# การทำงานของเครื่องเรียนรู้ของเครื่อง (ML Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ  
การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

## การเรียนรู้แบบจำลอง (Training)



## การใช้แบบจำลองที่เรียนมาได้ (Prediction)

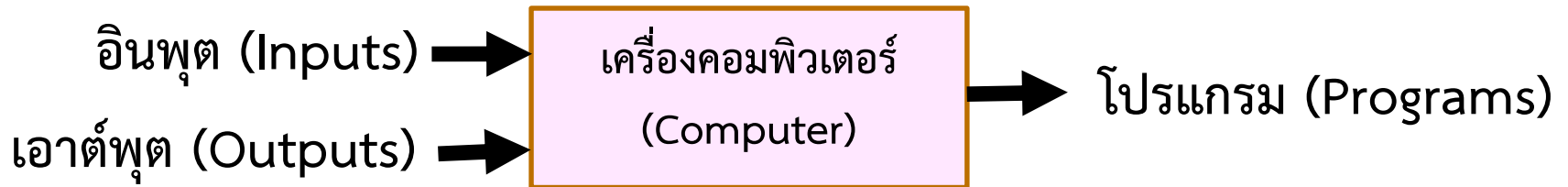


# ความแตกต่างระหว่างการโปรแกรมและการเรียนรู้ของเครื่อง

## โปรแกรมแบบดั้งเดิม (Traditional Programming)



## การเรียนรู้ของเครื่อง (Machine Learning)



# ความแตกต่างระหว่างการโปรแกรมและการเรียนรู้ของเครื่อง

Easy GG !!

Input

Clear Rules



$$\text{BMI} = \frac{(\text{weight in kilograms})}{\text{height in meters}^2}$$

น้ำหนัก (Weight) = 70 kg.  
ความสูง (Height) = 170 cm.

โปรแกรมคอมพิวเตอร์  
(Computer Program)

Output

BMI = 24.2

Super Hard !!

Input

Complex Rules



เราจะโปรแกรมอย่างไรดีนะ งง !?!



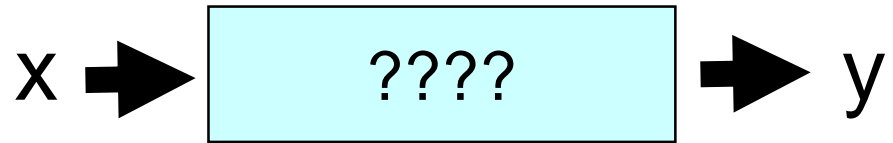
การรู้จำภาพหรือ โมเดลเอไอ  
(Image Recognition and AI Model)

Output

“Cat”

# การเรียนรู้ของเครื่องทำงานอย่างไร

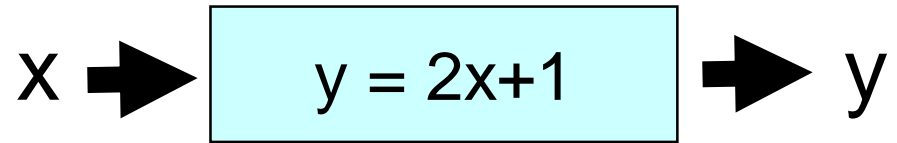
| x (Input) | y (Output) |
|-----------|------------|
| 0         | 1          |
| 1         | 3          |
| 2         | 5          |
| 3         | 7          |



# การเรียนรู้ของเครื่องทำงานอย่างไร

---

| x (Input) | y (Output) |
|-----------|------------|
| 0         | 1          |
| 1         | 3          |
| 2         | 5          |
| 3         | 7          |

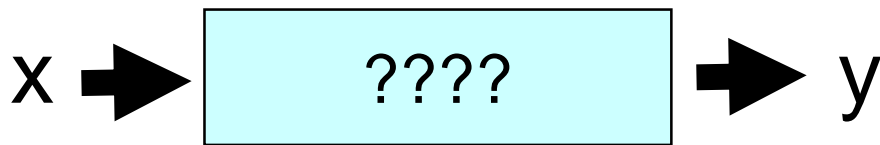




# การเรียนรู้ของเครื่องทำงานอย่างไร

---

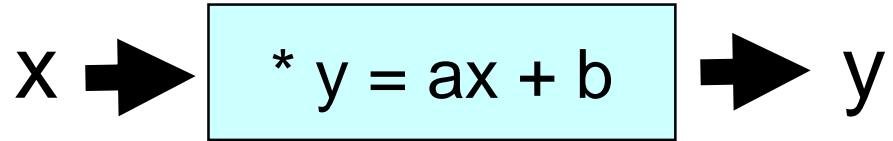
| x (Input) | y (Output) |
|-----------|------------|
| 0         | 1          |
| 1         | 3          |
| 2         | 5          |
| 3         | 7          |



# การเรียนรู้ของเครื่องทำงานอย่างไร

| x (Input) | y (Output) |
|-----------|------------|
| 0         | 1          |
| 1         | 3          |
| 2         | 5          |
| 3         | 7          |

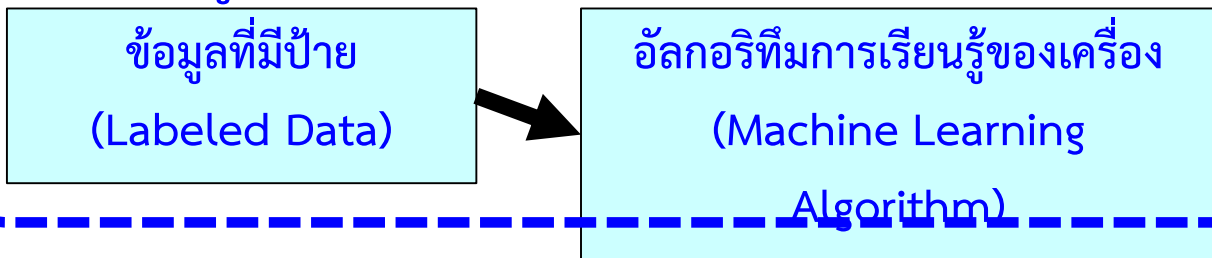
วิศวกรปัญญาประดิษฐ์คิดว่ามันน่าจะเป็นปัญหาเชิงเส้น  
เลยเลือกการวิเคราะห์การถดถอยเชิงเส้น (Linear  
Regression) เป็นอัลกอริทึมของการเรียนรู้



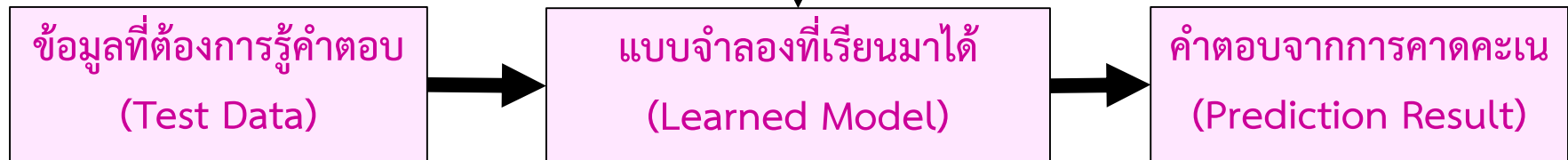
# Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ  
การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

## การเรียนรู้แบบจำลอง (Training)



## การใช้แบบจำลองที่เรียนมาได้ (Prediction)



# การทำงานของเครื่องเรียนรู้ของเครื่อง (ML Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

| x (Input) | y (Output) |
|-----------|------------|
| 0         | 1          |
| 1         | 3          |
| 2         | 5          |
| 3         | 7          |

ระบบจำลอง (Training)

ข้อมูลที่มีป้าย  
(Labeled Data)

อัลกอริทึมการเรียนรู้ของเครื่อง  
(Machine Learning  
Algorithm)

การใช้แบบจำลองที่เรียนมาได้ (Prediction)

ข้อมูลที่ต้องการรู้คำตอบ  
(Test Data)

แบบจำลองที่เรียนมาได้  
(Learned Model)

คำตอบจากการคาดคะเน  
(Prediction Result)

# Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ  
การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

| x (Input) | y (Output) |
|-----------|------------|
| 0         | 1          |
| 1         | 3          |
| 2         | 5          |
| 3         | 7          |

ระบบจำลอง (Training)

แบบจำลองเชิงเส้น

$$y = ax + b$$

ข้อมูลที่มีป้าย  
(Labeled Data)

อัลกอริทึมการเรียนรู้ของเครื่อง  
(Machine Learning  
Algorithm)

การใช้แบบจำลองที่เรียนมาได้ (Prediction)

ข้อมูลที่ต้องการรู้คำตอบ  
(Test Data)

แบบจำลองที่เรียนมาได้  
(Learned Model)

คำตอบจากการคาดคะเน  
(Prediction Result)

# การทำงานของเครื่องเรียนรู้ของเครื่อง (ML Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

| x (Input) | y (Output) |
|-----------|------------|
| 0         | 1          |
| 1         | 3          |
| 2         | 5          |
| 3         | 7          |

แบบจำลอง (Training)

ข้อมูลที่มีป้าย (Labeled Data)

อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning Algorithm)

แบบจำลองเชิงเส้น

$$y = ax + b$$

อัลกอริทึมจะพยายามหาค่า a และ b ที่เข้ากับตัวอย่างที่ดีที่สุด (Learning)

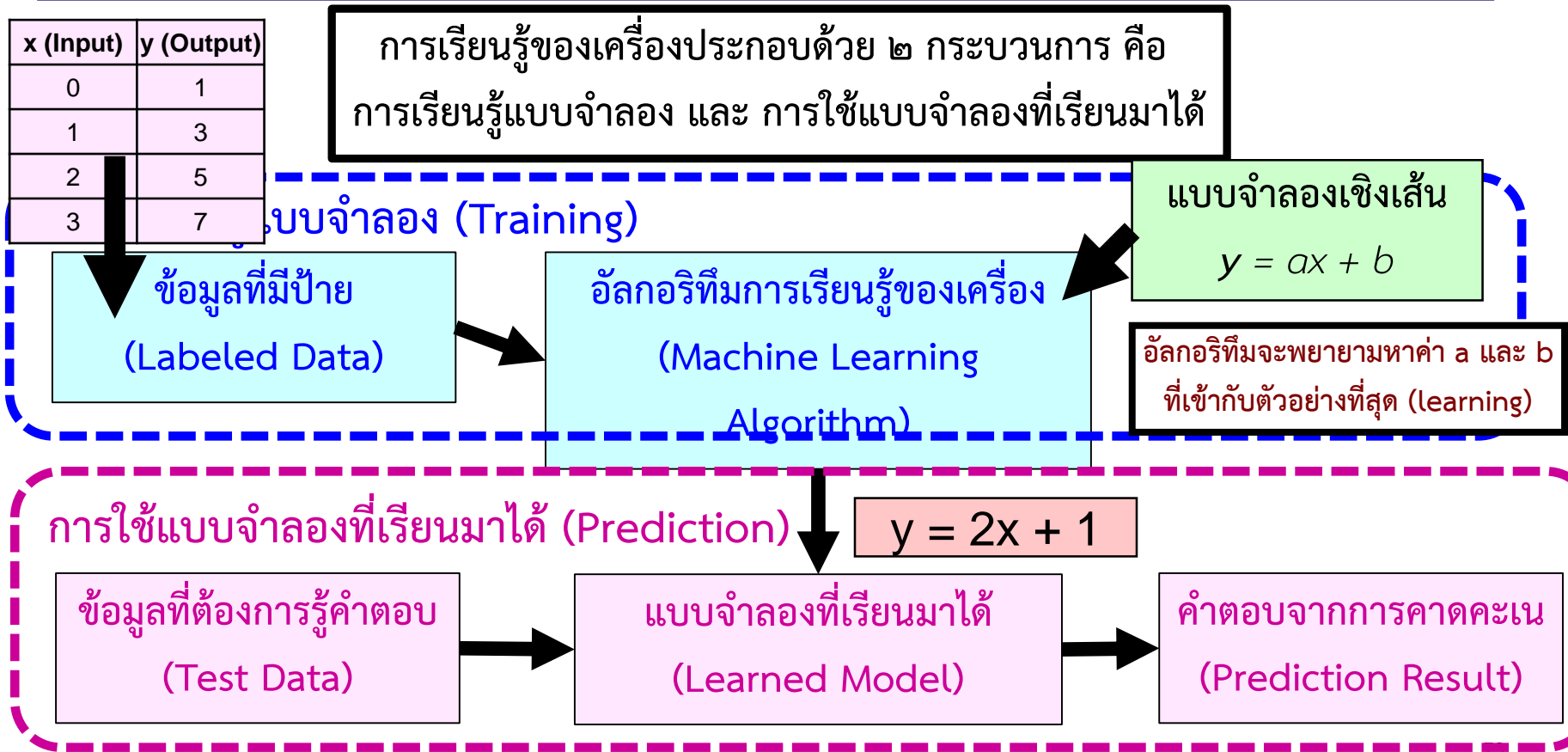
การใช้แบบจำลองที่เรียนมาได้ (Prediction)

ข้อมูลที่ต้องการรู้คำตอบ (Test Data)

แบบจำลองที่เรียนมาได้ (Learned Model)

คำตอบจากการคาดคะเน (Prediction Result)

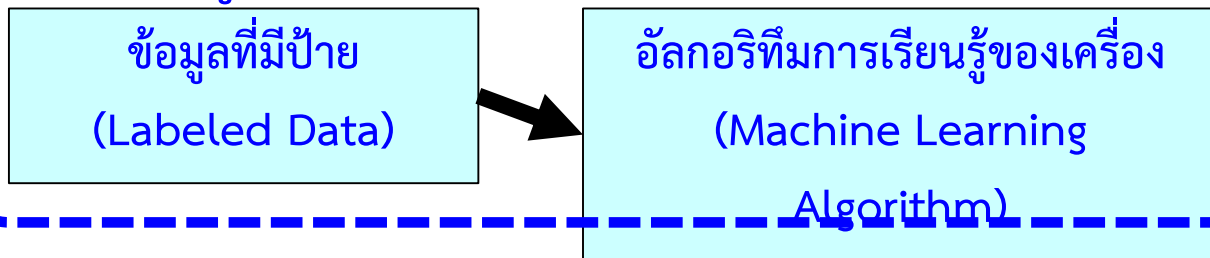
# การทำงานของเครื่องเรียนรู้ของเครื่อง (ML Process)



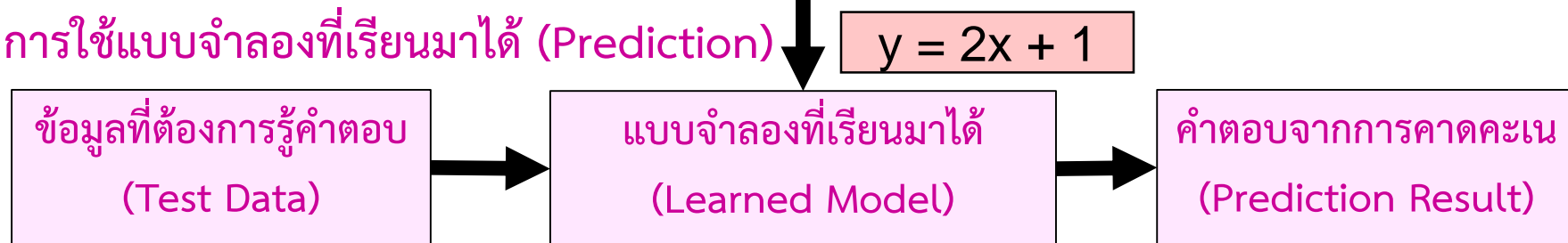
# การทำงานของเครื่องเรียนรู้ของเครื่อง (ML Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ  
การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

## การเรียนรู้แบบจำลอง (Training)



## การใช้แบบจำลองที่เรียนมาได้ (Prediction)

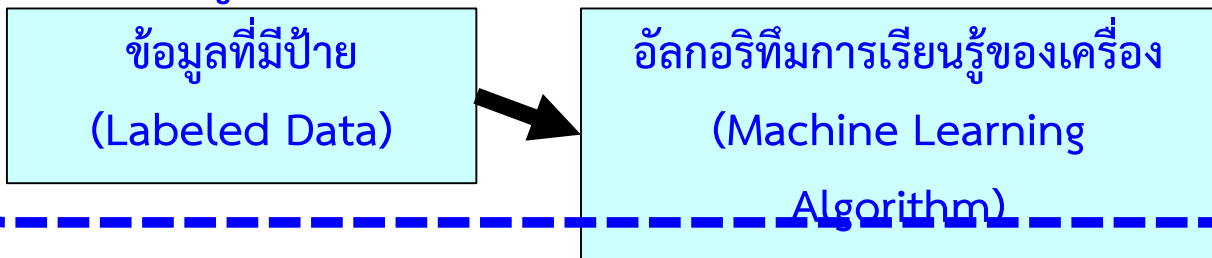




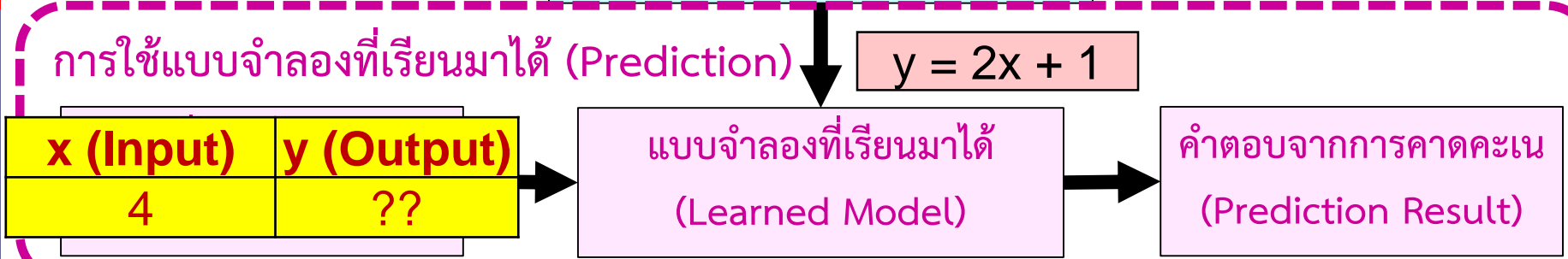
# การทำงานของเครื่องเรียนรู้ของเครื่อง (ML Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ  
การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

## การเรียนรู้แบบจำลอง (Training)



## การใช้แบบจำลองที่เรียนมาได้ (Prediction)



# การทำงานของเครื่องเรียนรู้ของเครื่อง (ML Process)

การเรียนรู้ของเครื่องประกอบด้วย ๒ กระบวนการ คือ  
การเรียนรู้แบบจำลอง และ การใช้แบบจำลองที่เรียนมาได้

## การเรียนรู้แบบจำลอง (Training)

ข้อมูลที่มีป้าย  
(Labeled Data)

อัลกอริทึมการเรียนรู้ของเครื่อง  
(Machine Learning  
Algorithm)

9

## การใช้แบบจำลองที่เรียนมาได้ (Prediction)

$$y = 2x + 1$$

| x (Input) | y (Output) |
|-----------|------------|
| 4         | ??         |

แบบจำลองที่เรียนมาได้  
(Learned Model)

คำตอบจากการคาดคะเน  
(Prediction Result)

# การเรียนรู้แบบมีผู้สอน (supervised learning)

---



# ตัวอย่างการเรียนรู้แบบมีผู้สอน (supervised learning)

---

| Input (X)      | Output (Y)           | Application         |
|----------------|----------------------|---------------------|
| Email          | Spam (0/1)           | Spam Filtering      |
| Audio          | Transcript Text      | Speech Recognition  |
| English        | Thai                 | Machine Translation |
| Ad + User Info | Click (0/1)          | Online Advertising  |
| Image + Radar  | Controlling of a car | Self—Driving Car    |

# ตัวอย่างการเรียนรู้แบบมีผู้สอน (supervised learning)

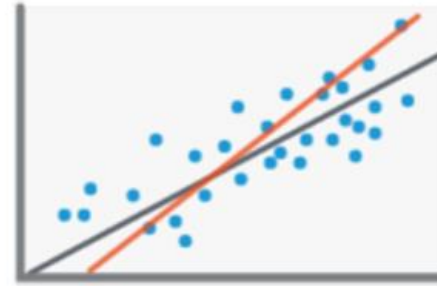
มีการให้คำตอบไว้ (Output data is given)

การวิเคราะห์การถดถอยเชิงเส้น  
(Linear Regression)



คาดคะเนประเภทหรือป้าย

Classification  
(supervised – predictive)



คาดคะเนตัวเลข

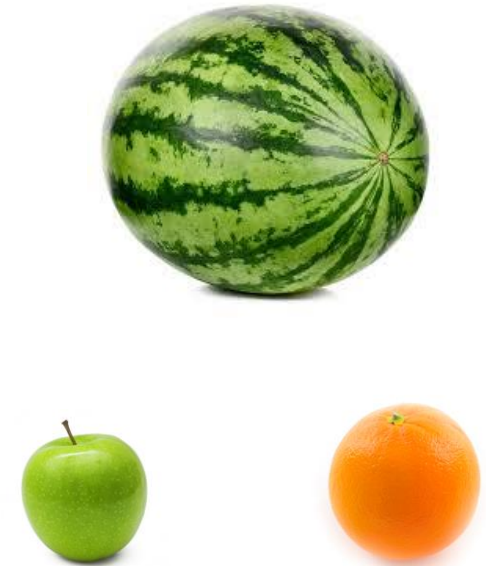
Regression  
(supervised – predictive)

การจำแนกประเภท  
(classification)

การคาดคะเน  
(prediction)

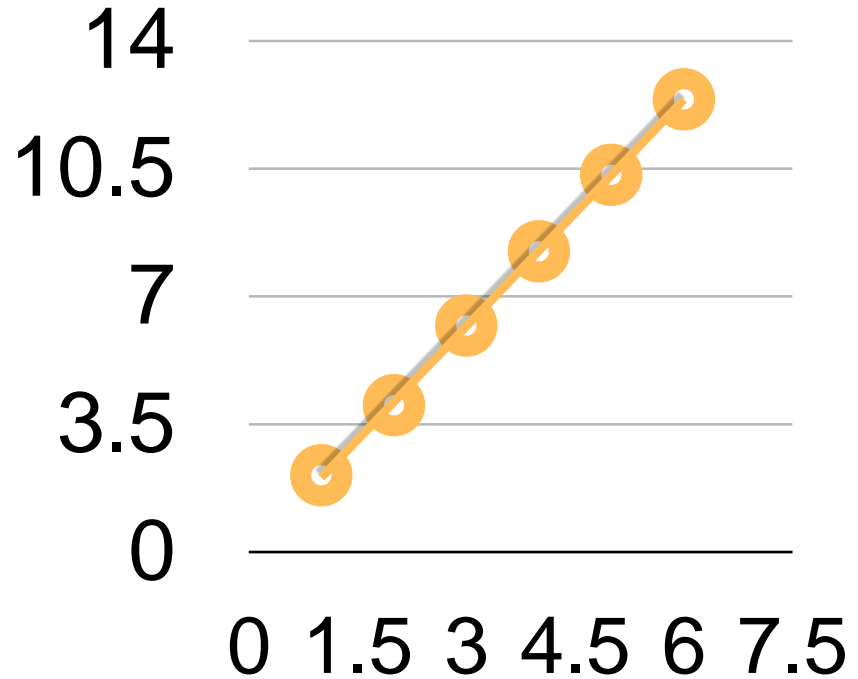
# การจำแนกประเภท (classification)

| Color  | Diameter (cm) | Output Class |
|--------|---------------|--------------|
| Green  | 6             | Apple        |
| Green  | 25            | Water Melon  |
| Green  | 4             | Apple        |
| Orange | 4             | Orange       |
| Green  | 23            | ??           |



# การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression)

| x | f(x) |
|---|------|
| 1 | 2.1  |
| 2 | 4.02 |
| 3 | 6.2  |
| 4 | 8.23 |
| 5 | 10.3 |
| 6 | ??   |



# การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning)

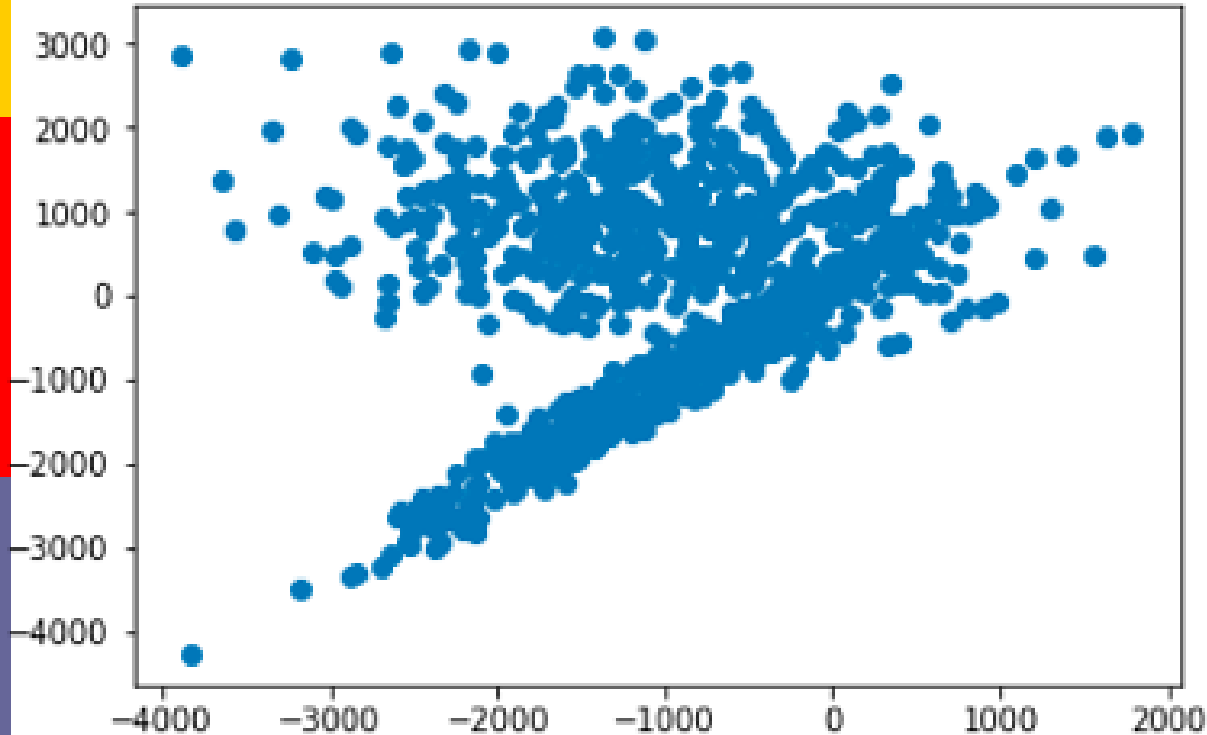
---

ไม่มีตัวอย่างเอาต์พุต





# การจัดกลุ่มข้อมูล (Clustering)



- ในรูปนี้ ข้อมูลมีกี่กลุ่ม
- และอยู่ตรงไหนบ้าง

# การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning)

ไม่มีการให้คำตอบไว้ (Output data is NOT given)



Clustering  
(unsupervised – descriptive)



Anomaly Detection  
(unsupervised – descriptive)

Association  
Analysis

Pepsi

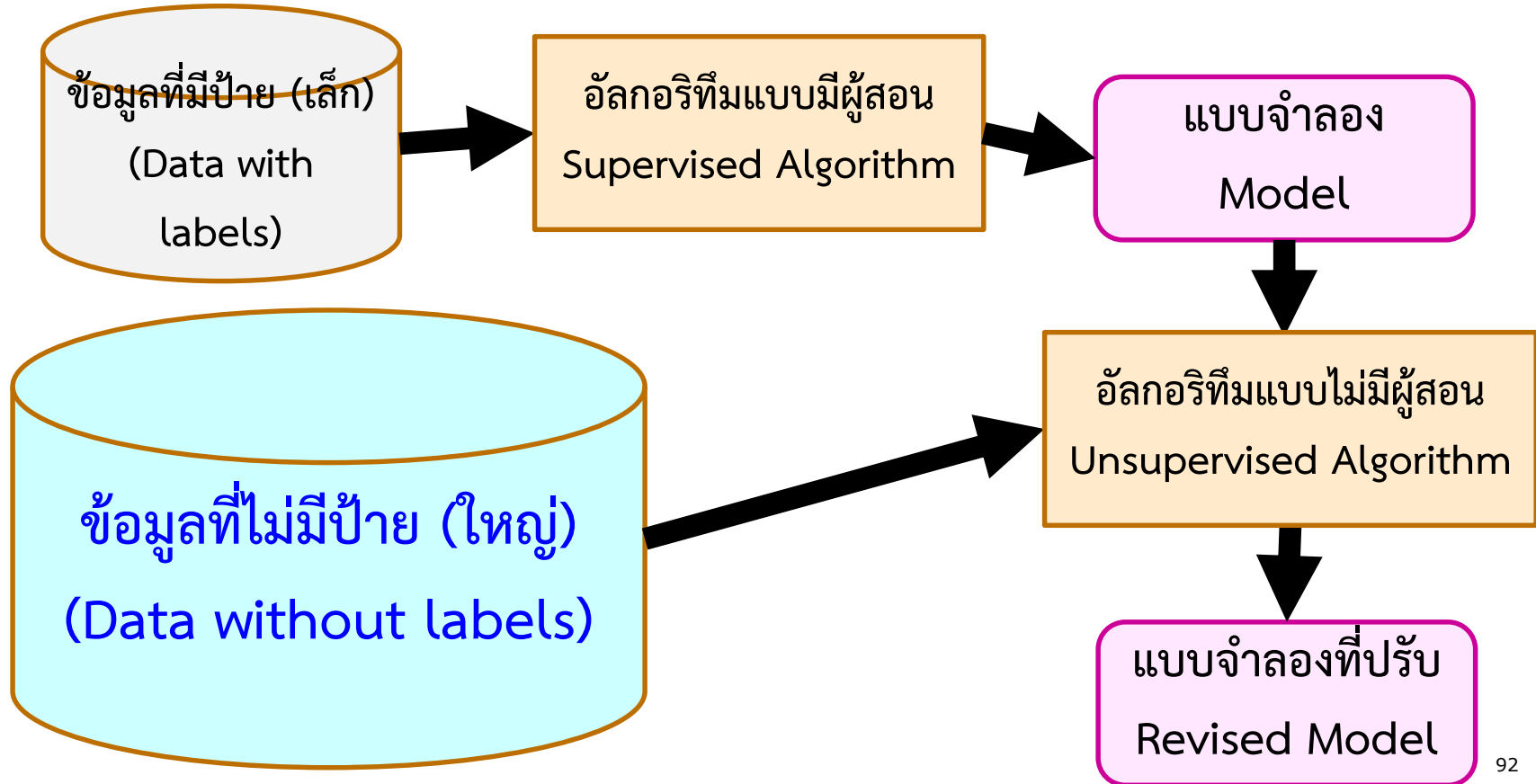


Popcorn

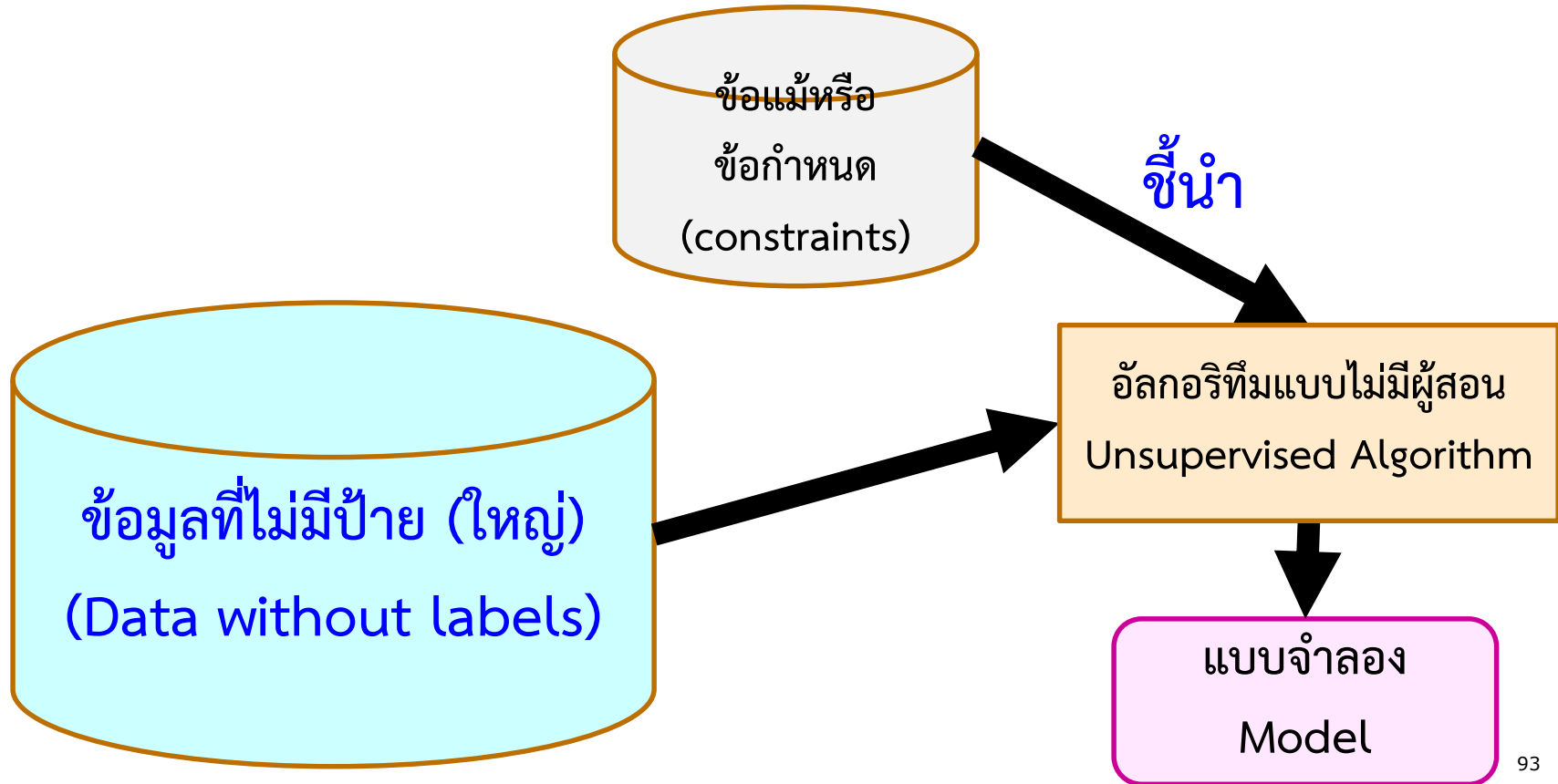
# ตัวอย่างการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning)

| <b>Input (X)</b>           | <b>Looking for</b> | <b>Application</b>           |
|----------------------------|--------------------|------------------------------|
| Customer Profiles          | Spending Behavior  | Market Segmentation          |
| Documents                  | Similarity         | Plagiarism Detection         |
| Credit Card Transaction    | Anomaly            | Fraud Detection              |
| Social Network Information | Groups             | Communities Recognition      |
| Image                      | Image Segments     | Automatic Image Segmentation |

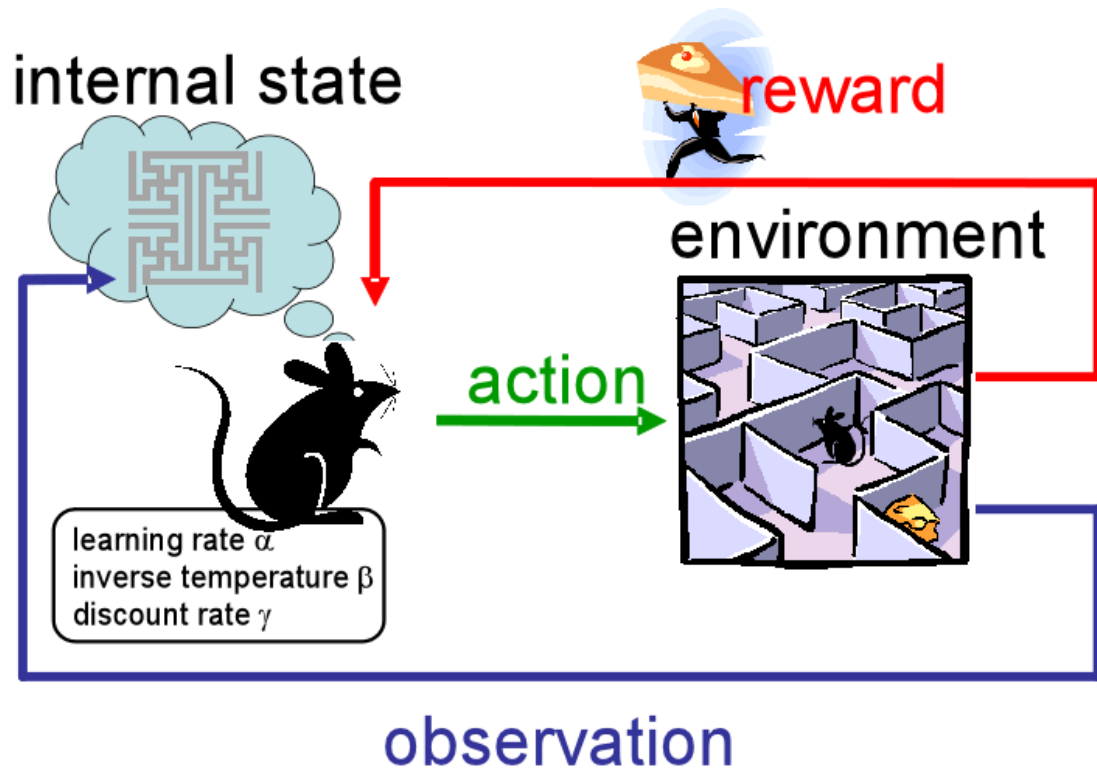
# การเรียนรู้แบบกึ่งมีผู้สอน (semi-supervised learning) <sup>ML</sup>



# การเรียนรู้แบบกึ่งไม่มีผู้สอน (semi-supervised learning)



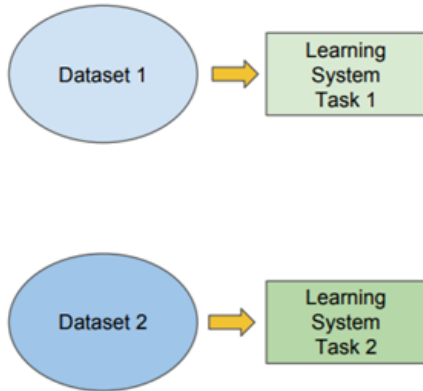
# การเรียนรู้แบบเสริมกำลัง (reinforcement learning)



# การเรียนรู้แบบถ่ายโยงการเรียนรู้ (Transfer Learning)

## Traditional ML

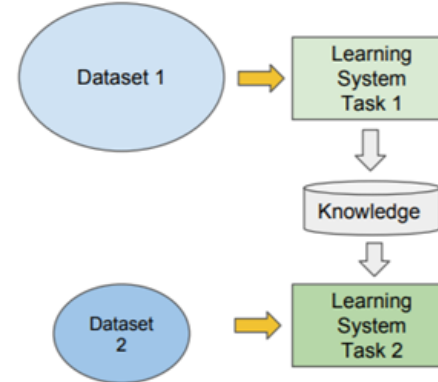
- Isolated, single task learning:
  - Knowledge is not retained or accumulated. Learning is performed w.o. considering past learned knowledge in other tasks



vs

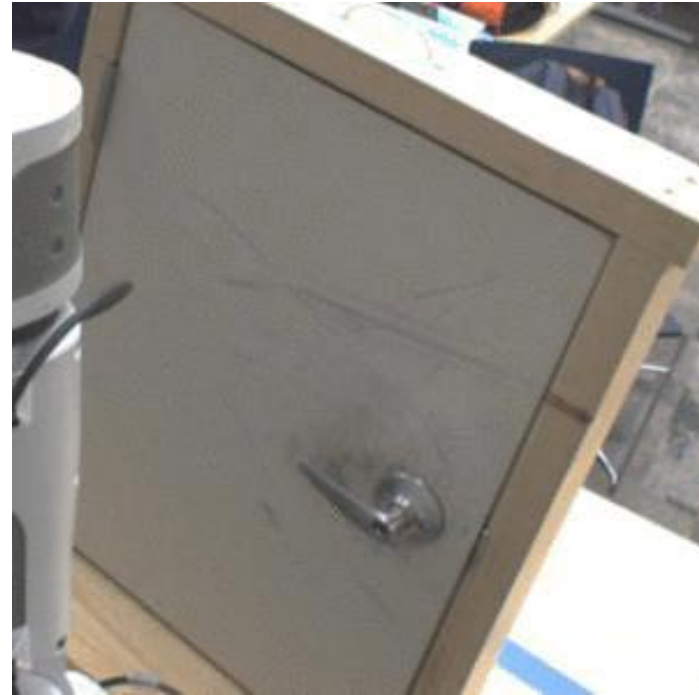
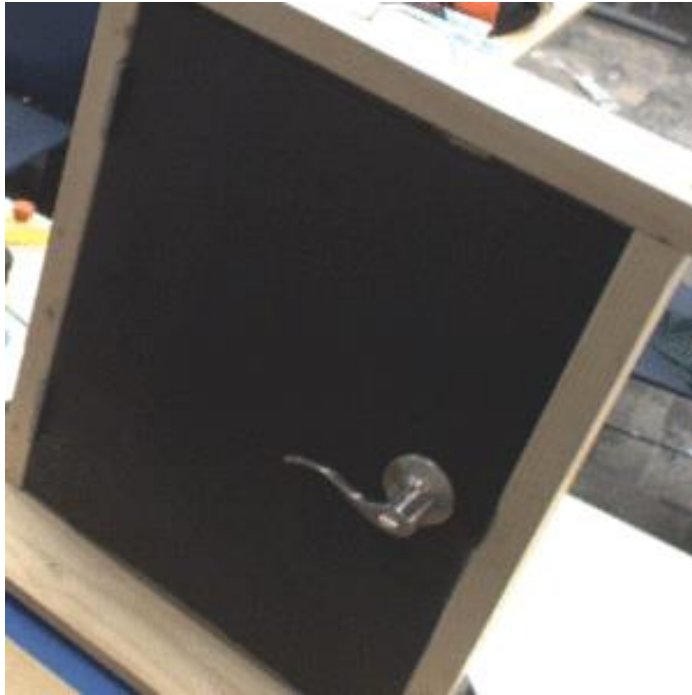
## Transfer Learning

- Learning of a new tasks relies on the previous learned tasks:
  - Learning process can be faster, more accurate and/or need less training data



<https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world->

# การเรียนรู้แบบเลียนแบบ (Imitation learning)



<https://sermanet.github.io/imitation/>

Reinforcement learning

**Self driving car**





# Data Science & Artificial Intelligence

---

## Train the trainer

December 14 - 15, 2019