

Chapter 1 : Introduction to Data Science

Content

- ✓ Data Science คืออะไร
- ✓ องค์ประกอบของ data science
- ✓ ขั้นตอนการทำ data science
- ✓ ชนิดของข้อมูล
- ✓ 5 ระดับของการใช้ Data Science ในองค์กร
- ✓ บทบาทต่าง ๆ ในสายงาน Data
- ✓ ตัวอย่างการใช้งาน data science

Data Science คืออะไร?

Data Science คืออะไร

กอบเกียรติ สระอุบล (2562 : 18) ได้ให้คำจำกัดความของวิทยาศาสตร์ข้อมูล คือ เป็นศาสตร์ว่าด้วยการนำข้อมูลมาใช้ให้เกิดประโยชน์ในงานด้านต่าง ๆ อาทิ ด้านธุรกิจ โลจิสติก เศรษฐศาสตร์ การเงิน การแพทย์ วิศวกรรม สารสนเทศ สาธารณสุข และอื่น ๆ

Data Science คืออะไร

วรรณะ งามตระกูลชน (2562 : 16-20) วิทยาศาสตร์ข้อมูล หมายถึง การนำข้อมูลมาใช้ประโยชน์ โดยครอบคลุมตั้งแต่ขั้นตอนการเก็บข้อมูล (Collect) การจัดการข้อมูล (Manage) การวิเคราะห์ข้อมูล (Analyze) ไปจนถึงขั้นตอนการนำข้อมูลมาช่วยตัดสินใจ (Decision)

Data Science คืออะไร

ละออ บุญเกษม (2562 : 18) ได้กล่าวว่า วิทยาศาสตร์ข้อมูลเป็นสหวิทยาการที่สร้างขึ้นบนพื้นฐานของสถิติ คณิตศาสตร์ วิทยาการคอมพิวเตอร์ การจัดการ และศาสตร์อื่น ๆ ที่เกี่ยวข้องกับการทำงาน ผ่านกระบวนการทางวิทยาศาสตร์ เพื่อดึงความรู้และข้อมูลเชิงลึกจากข้อมูลที่มีโครงสร้างและไม่มีโครงสร้าง ออกมาเป็นผลิตภัณฑ์ข้อมูล

Data Science คืออะไร

Davy Cielen Arno D. B. Meysman and Mohamed Ali (2016: 1) ได้กล่าวว่า วิทยาศาสตร์ข้อมูล คือ การใช้วิธีที่จะวิเคราะห์ข้อมูลจำนวนมาก และสกัดความรู้ในข้อมูลนั้นออกมา

กิจกรรมใจเราตรงกันไหม

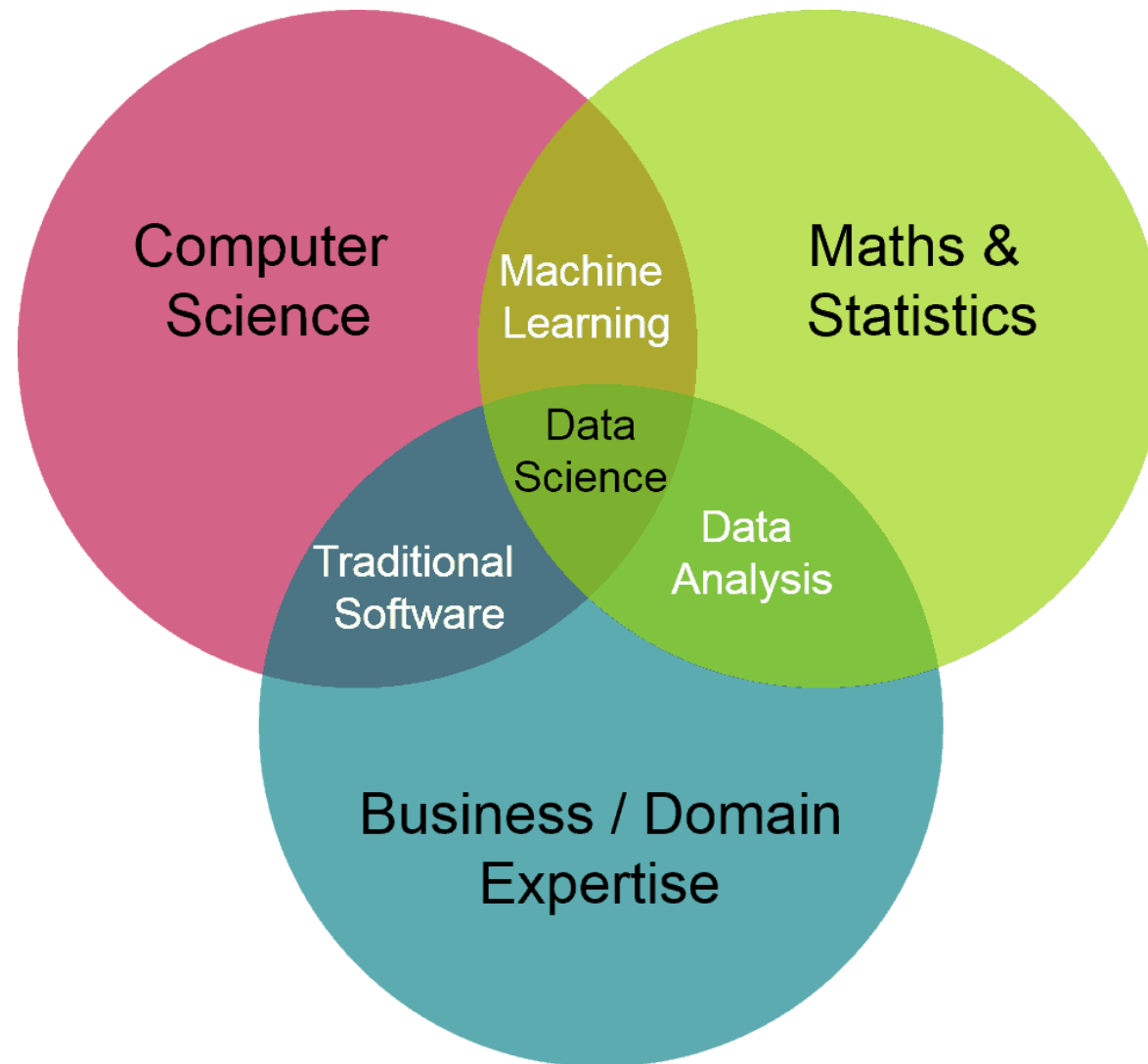
นินยามวิทยาศาสตร์ข้อมูลใน 3 คำ และพิมพ์ใส่ลิงก์ใน QR code/ลิงก์นี้



<https://www.menti.com/aljoxjhx1kjo>

องค์ประกอบของ data science

องค์ประกอบของ Data Science



องค์ประกอบของ Data Science

Computer Science – วิทยาการคอมพิวเตอร์ เช่น การเขียนโปรแกรม, อัลกอริธึม, โครงสร้างข้อมูล (Data Structure) เป็นต้น

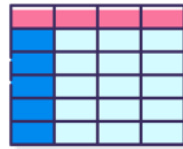
Maths & Statistics – คณิตศาสตร์ และสถิติ

Business / Domain Expertise – ความรู้ด้านธุรกิจ

ชนิดของข้อมูล

ชนิดของข้อมูล (Type of Data)

a) Structured Data



b) Unstructured Data



c) Semi-Structured Data



a) ข้อมูลแบบมีโครงสร้าง (Structured Data)

ข้อมูลลักษณะนี้จะเป็นข้อมูลที่เก็บอยู่ในรูปแบบของตาราง (Table) ซึ่งประกอบด้วยแถว (Row) และคอลัมน์ (Column)

ตัวอย่าง Structured Data

	A	B	C	D	E	F	G	H	I
6	term	site	Popularity	Competition	Competition_paid	Traffic_score	Paid_impression_score	Share_of_voice	Percent_of_traffic
7	woocommerce ???	designil.com	27	40	0	67		15.6731876	8.17067682
8	adobe xd	designil.com	59	66	9	61		0.03968074	2.73210345
9	bootstrap 4	designil.com	60	32	0	61		0.03243026	2.52159489
10	??? woocommerce	designil.com	28	23	1	60		3.40286743	2.33023584
11	font thai	designil.com	23	47	0	60		9.81897263	2.09183006
12	lorem ipsum	designil.com	68	63	0	60		0.012835	1.99960353
13	font ???	designil.com	22	47	0	58		8.73875751	1.60369875
14	infographic	designil.com	55	34	6	58		0.07675551	1.35195664
15	bounce rate ???	designil.com	9	45	0	57		45.14197988	1.32611949
16	google font thai	designil.com	20	24	0	57		5.43834797	1.25140507
17	thai font	designil.com	25	57	0	57		4.65907223	1.13315291
18	woocommerce	designil.com	62	72	28	56		0.02017068	0.94929497
19	??????	designil.com	22	22	5	56		4.33455945	0.91912392
20	bootstrap	designil.com	68	17	0	55		0.00602208	0.84748578
21	getty image ????	designil.com	6	11	0	55		26.7051048	0.77074522
22	jquery	designil.com	62	52	0	55		0.01405276	0.75528394
23	responsive web design	designil.com	42	64	22	54		0.13143857	0.70343553
24	lorem ipsum thai	designil.com	14	44	0	54		8.1528485	0.68265439
25	landing page	designil.com	54	63	42	54		0.02484704	0.62524488
26	thai font for mac	designil.com	11	51	0	54		22.87633838	0.61262878
27	????????????????? html	designil.com	6	20	0	54		31.97742702	0.60416615
28	sukhumvit font	designil.com	12	47	0	53		9.87686444	0.59991033
29	infographic ???	designil.com	10	43	0	53		13.679723	0.58999238
30	????????? wordpress	designil.com	14	46	5	53		10.23742658	0.57643415
31	font thai download free	designil.com	9	51	0	53		27.21822853	0.5711661
32	responsive	designil.com	45	68	1	53		0.11200085	0.56716538
33	wireframe	designil.com	50	59	14	53		0.07768544	0.55139784
34	sukhumvit set font	designil.com	14	55	0	53		5.43973756	0.55101134
35	page speed	designil.com	54	34	1	53		0.01378355	0.54127025
36	????????? ???	designil.com	12	47	0	53		10.61349979	0.53552906
37	????????????????	designil.com	14	40	0	53		7.20789427	0.50002757



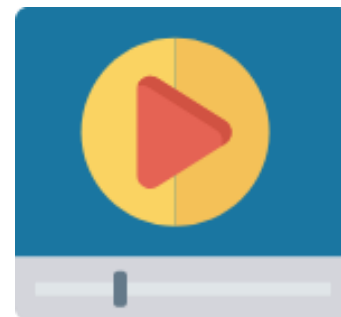
b) ข้อมูลแบบไม่มีโครงสร้าง Unstructured Data



ไฟล์



รูปภาพ



วิดีโอ



เสียง

ตัวอย่าง Unstructured Data

เช่น ข้อความในเว็บไซต์ Twitter แต่ละข้อความก็จะมีกรร่อยเรียงคำพูดที่แตกต่างกันออกไป หรือรูปภาพต่าง ๆ ที่ถูกโพสต์ใน Instagram ก็จะมีรูปและขนาดที่แตกต่างกันเช่นเดียวกัน

c) ข้อมูลแบบกึ่งโครงสร้าง (Semi-Structured Data)

ข้อมูลแบบที่มีโครงสร้างข้อมูลที่มีความยืดหยุ่น สามารถขยายโครงสร้างข้อมูลได้ในอนาคต และเรียกใช้ได้รวดเร็ว

เหมาะสำหรับ: ข้อมูลที่สามารถปรับเปลี่ยนโครงสร้างได้ตลอดเวลา

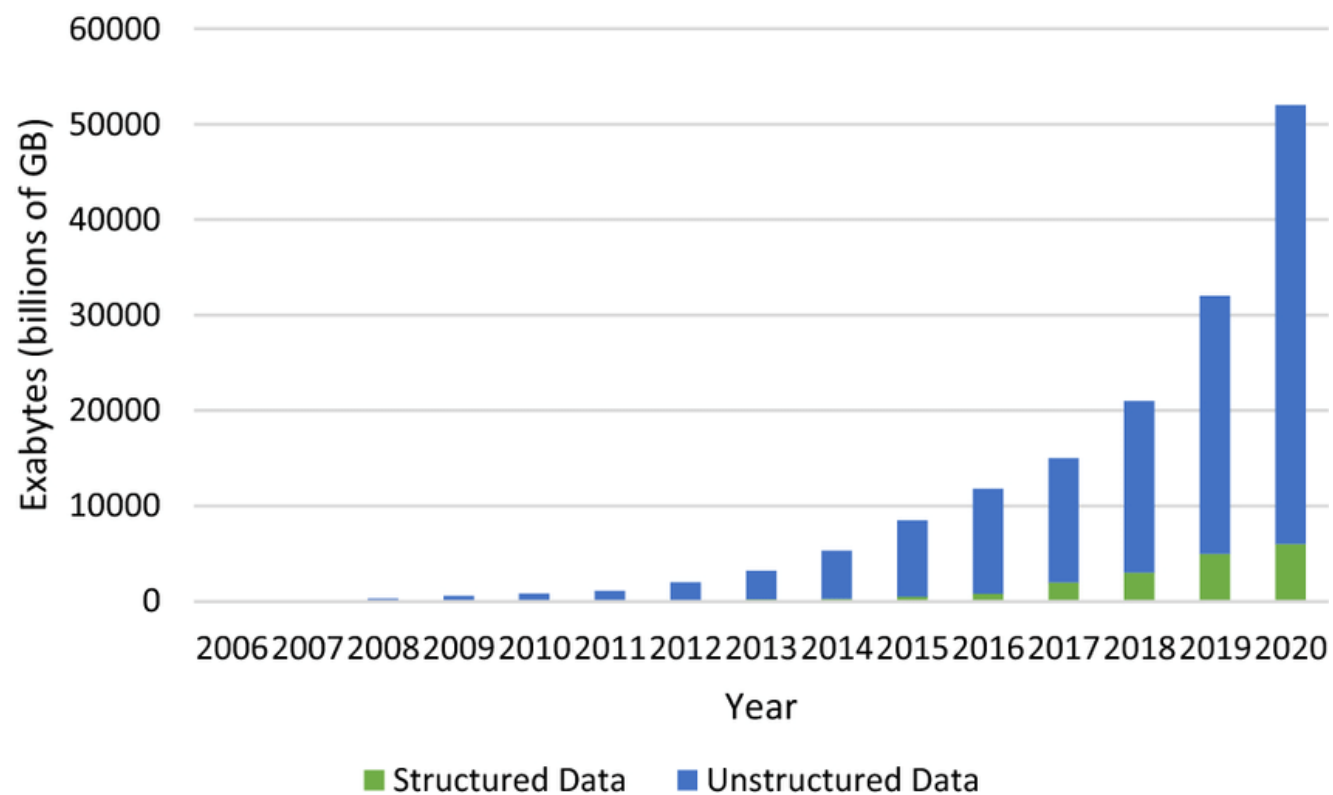


NoSQL



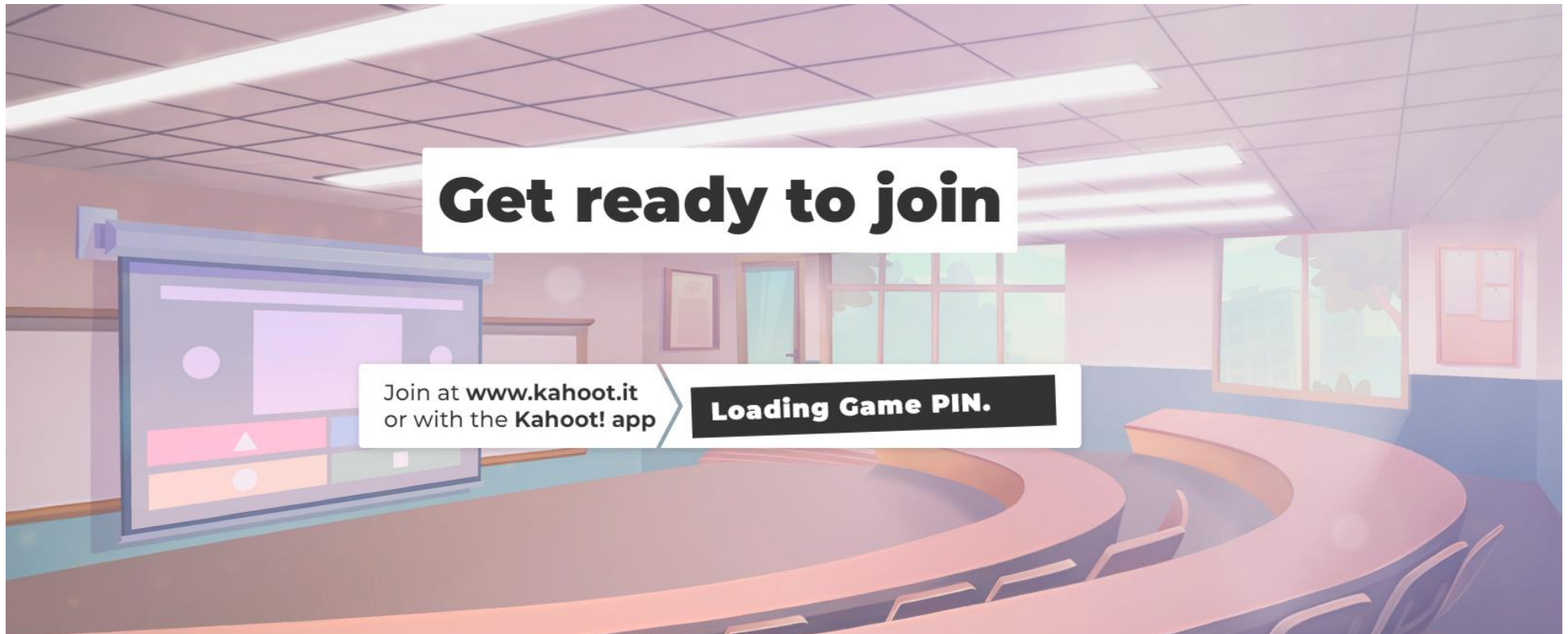
```
{
  "users": [
    {
      "userId": 1,
      "firstName": "Adam",
      "lastName": "Lee",
      "emailAddress": "adam@gmail.com"
    },
    {
      "userId": 2,
      "firstName": "John",
      "lastName": "Doe",
      "phoneNumber": "0406817201",
      "emailAddress": "john.doe@gmail.com"
    }
  ]
}
```

แสดงปริมาณของข้อมูลที่เพิ่มขึ้นในแต่ละปี ซึ่งจะพบว่าข้อมูลส่วนใหญ่เป็นข้อมูลแบบ *Unstructured Data*



กิจกรรม

จงบอกว่าคุณคิดว่าข้อมูลต่อไปนี้เป็นข้อมูลประเภทไหนใน kahoot



ขั้นตอนการทำ Data Science

ขั้นตอนการทำ Data Science ทฤษฎีที่ 1: OSEMN



OSEMN เป็น 5 ขั้นตอนการทำงานด้าน Data Science

1. Obtain เก็บรวบรวมข้อมูล
2. Scrub ทำความสะอาดข้อมูล
3. Explore สำรวจข้อมูล
4. Model ทำโมเดลต่างๆ
5. Interpret การนำเสนอ

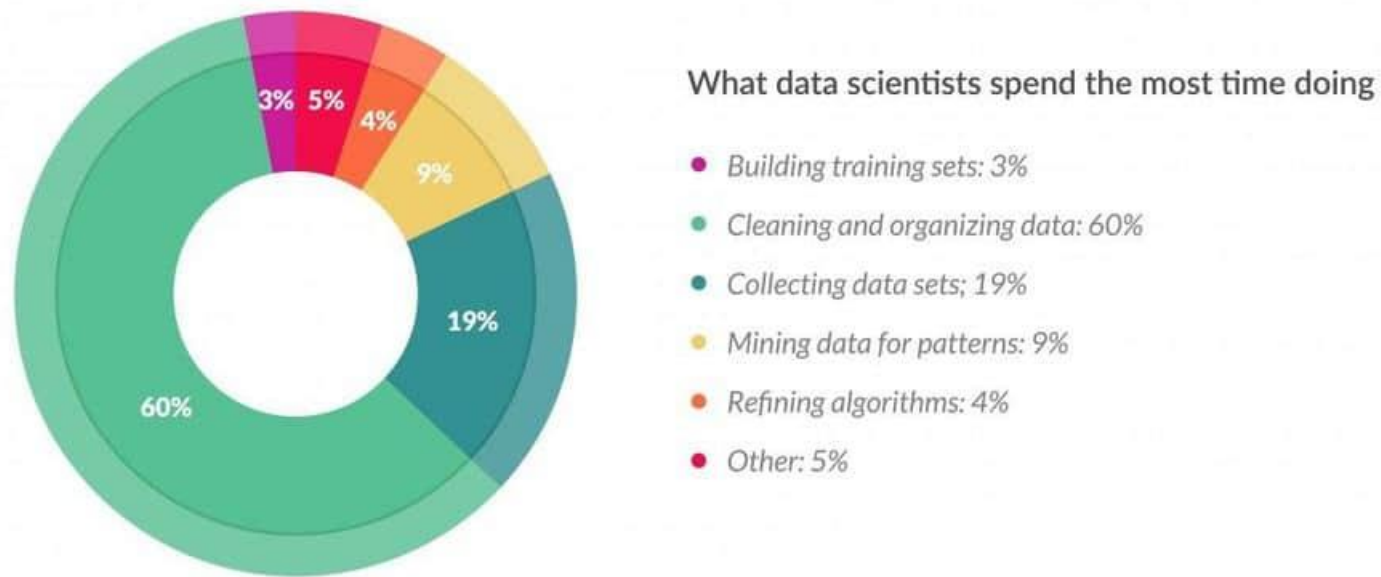
1. Obtain เก็บรวบรวมข้อมูล

การเก็บข้อมูลสามารถทำได้หลายวิธี ซึ่งขึ้นอยู่กับข้อมูลที่เราต้องการเก็บ เช่น ถ้าต้องการเก็บ Log การใช้เว็บไซต์ ก็อาจจะเขียน JavaScript วางไว้ในเว็บไซต์ เพื่อเก็บการกระทำต่าง ๆ ของผู้ใช้ ถ้าต้องการเก็บข้อมูลจากเว็บไซต์อื่น ต้องเขียนโปรแกรมดึงข้อมูลทาง API หรือ Scrape ข้อมูลจากหน้าเว็บไซต์



2. Scrub ทำความสะอาดข้อมูล

การทำทำความสะอาดข้อมูล ปรับหน้าตาให้อยู่ใน format ที่เหมาะสม (จากงานวิจัยของ DJ Patil (2012) พบว่าขั้นตอนการ scrub ใช้เวลานานที่สุดประมาณ 50-80% ของ OSEMN process



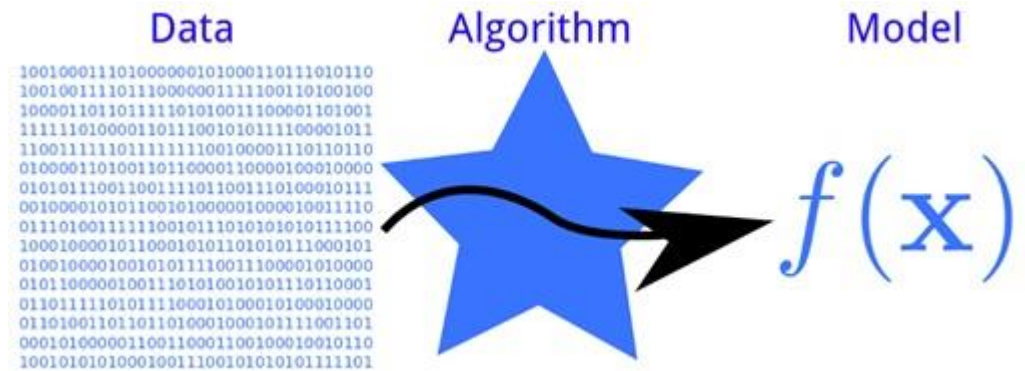
3. Explore สํารวจข้อมูล

คือ การศึกษาข้อมูลเบื้องต้น เช่น ยอดขายกางเกงยีนส์เดือนธันวาคมของ 5 ปีที่ผ่านมาสูงกว่าสินค้าอื่น ๆ เราก็สามารถปรับแผน Marketing ให้เหมาะสมได้



4. Model ทำโมเดลต่าง ๆ สร้าง Model สำหรับนำมาใช้ทำนายผล

เช่น ทำโมเดลต่าง ๆ ด้วย Machine Learning, Deep Learning สำหรับทำนายผล
แนวโน้มยอดขายสินค้าแต่ละตัว เพื่อนำไปใช้ประกอบการวางแผนผลิต เป็นต้น



5. Interpret การนำเสนอ

การนำเสนอผลการวิเคราะห์ข้อมูลต่าง ๆ ทั้งหมดที่ได้ทำมา สรุปผลวิเคราะห์ให้เข้าใจง่าย ๆ เพื่อนำไปเสนอผู้ร่วมงานในฝ่ายบริหาร สามารถแสดงเป็นรูปภาพให้ฝ่ายบริหารเข้าใจง่ายขึ้นได้ ความสามารถด้าน Data Visualization จะมีประโยชน์มาก



ขั้นตอนการทำ Data Science ทฤษฎีที่ 2 : CRISP-DM Process

ขั้นตอนการทำ Data Science ทฤษฎีที่ 2 : CRISP-DM Process

1. Business Understanding ทำความเข้าใจธุรกิจ
2. Data Understanding ทำความเข้าใจข้อมูล
3. Data Preparation เตรียมข้อมูล
4. Modeling สร้างโมเดล
5. Evaluation ประเมินโมเดล
6. Deployment นำไปใช้งาน

Step 1: Business Understanding

เป็นขั้นตอนการทำความเข้าใจธุรกิจ ปัญหาและวัตถุประสงค์ของโครงการจากมุมมองทางธุรกิจ จากนั้นแปลงปัญหาให้อยู่ในรูปของโจทย์สำหรับการวิเคราะห์ข้อมูล และวางแผนการดำเนินงานเบื้องต้น

Step 2: Data Understanding

เป็นขั้นตอนที่จะทำความเข้าใจ ตรวจสอบคุณภาพ และเลือกข้อมูลที่จะเก็บรวบรวมมาว่าจะใช้ข้อมูลใดบ้างในการวิเคราะห์

นอกจากนี้สิ่งที่มีมักจะพบในองค์กรใหญ่ คือ ข้อมูลจะเก็บกระจายอยู่ในฐานข้อมูลหรือตารางต่าง ๆ โดยในขั้นตอนนี้ต้องทำความเข้าใจว่า ข้อมูลแต่ละตัวแปรคืออะไรบ้าง มีเงื่อนไข ในการจัดเก็บหรือข้อยกเว้นอย่างไรบ้าง การเชื่อมโยงข้อมูลหลายๆ ตารางทำได้อย่างไร

Step 3 : Data Preparation

เป็นขั้นตอนที่ใช้เวลามากที่สุดใน 6 ขั้นตอน โดยขั้นตอนนี้จะแบ่งออกเป็น 3 ขั้นตอนย่อย ๆ ได้แก่

1. **Data Selection** เป็นขั้นตอนของการเลือกตัวแปรที่จะนำมาใช้งาน
2. **Data Cleansing** เป็นขั้นตอนการหาว่าข้อมูลมีสิ่งผิดปกติอะไรบ้าง เช่น ข้อมูลมีค่าว่างที่ขาด หายไปหรือเปล่า ถ้ามีแล้วจะต้องแก้ไขอย่างไรดี
3. **Data Transformation** เป็นการแปลงข้อมูลให้เหมาะสมสำหรับการวิเคราะห์ข้อมูลมากขึ้น

Step 4: Modeling

เป็นขั้นตอนการสร้างโมเดล ซึ่งโมเดลมีหลายรูปแบบ เช่น

- ✓ การหาความสัมพันธ์ (Association) ซึ่งเป็นการหาว่ามีการซื้อสินค้าอะไรพร้อมกันบ้าง
- ✓ ระบบแนะนำ (Recommendation System) เป็นการแนะนำสินค้าหรือบริการที่เกี่ยวข้องให้ กับผู้ใช้แต่ละคน
- ✓ การแบ่งกลุ่มข้อมูล (Segmentation and Clustering) เป็นการจัดกลุ่มข้อมูลที่มีลักษณะคล้ายกัน เข้าไว้ในกลุ่มเดียวกัน
- ✓ การจำแนกประเภทข้อมูล (Classification) เป็นการเรียนรู้จากข้อมูลในอดีตและสร้างโมเดล (Model) เพื่อคาดการณ์ว่าจะมีอะไรเกิดขึ้นต่อไปในอนาคต

Step 5: Evaluation

เป็นขั้นตอนการวัดผลลัพธ์ ซึ่งวิธีวัดผลอาจใช้การเปรียบเทียบจากผลลัพธ์ที่ได้ของแต่ละเทคนิคที่ใช้งาน เช่น เปรียบเทียบว่าความถูกต้องของโมเดล ว่าแบบไหน ดีกว่ากัน

ถ้ายังไม่ได้ก็สามารถย้อนกลับไปทำซ้ำในขั้นตอนก่อนหน้าได้อีกเรื่อย ๆ ซึ่งสาเหตุอาจเกิดจากข้อมูลอาจจะยังไม่มากพอ หรือโมเดลที่ใช้อาจจะยังไม่เหมาะสม

Step 6: Deployment

จะนำข้อมูลที่ค้นพบ (Insight) ได้ไปใช้งาน เป็นขั้นตอนของการนำไปใช้งานต่อก็ขึ้นอยู่กับแต่
ละธุรกิจว่า ในรูปแบบไหน เช่น บางธุรกิจอาจต้องการเป็นรายงานสรุปให้ผู้บริหารพิจารณา และ
วางแผนกลยุทธ์ต่อไป

กิจกรรม

พูดชื่อขั้นตอนการทำวิทยาศาสตร์ข้อมูลต่อกัน ใครพูดผิดขั้นตอน ออกมาเต้นหน้าห้อง

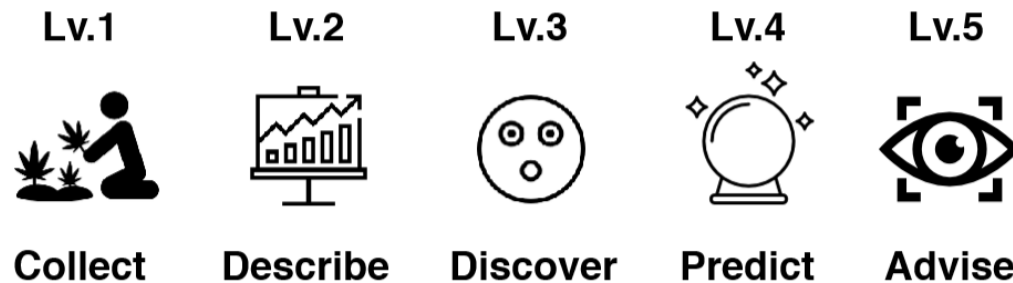


5 ระดับของการใช้ Data Science ในองค์กร

5 ระดับของการใช้ Data Science ในองค์กร

Booz Allen Hamilton ซึ่งเป็นบริษัทให้คำปรึกษาด้านเทคโนโลยีได้ลิสต์ระดับของการใช้ Data Science ในองค์กรเป็น 5 ระดับ เพื่อให้มาดูว่าองค์กรอยู่ในระดับไหน และจะไปทำอะไรต่อได้บ้าง

5 ระดับของการใช้ Data Science



5 ระดับของการใช้ Data Science ในองค์กร

ระดับที่ 1

Collect = ระดับแรก คือ เรายังไม่ทำอะไรเกี่ยวกับ Data เลย ก็ต้องหาวิธี หาเครื่องมือ ในการ “จัดเก็บ” Data ให้ได้ก่อน

5 ระดับของการใช้ Data Science ในองค์กร

ระดับที่ 2

Describe = ระดับที่สอง คือ เรามี Data แล้ว ก็สามารถเอาข้อมูลมาทำตาราง พล็อตกราฟได้ เพื่อ “อธิบาย” Data ที่เรามี สำหรับตอบคำถามทางธุรกิจเบื้องต้น เช่น เดือนไหนขายดี เดือนไหนลูกค้าเยอะ

5 ระดับของการใช้ Data Science ในองค์กร

ระดับที่ 3

Discover = ระดับที่สาม คือ การ “ค้นหาความลับ” จากข้อมูล ข้อนี้จะคล้ายกับ Describe แต่จะเป็นสิ่งที่ไม่เคยคิดมาก่อน ทำให้รู้สึกว้าวได้ ซึ่งต้องใช้เทคนิคมากขึ้นในการค้นหา

เช่น การค้นพบว่าเสื้อกันหนาวขายดี *แปรผันตรง* กับช่วงที่ถุงเท้าขายดี

5 ระดับของการใช้ Data Science ในองค์กร

ระดับที่ 4

Predict = ระดับที่สี่ คือ การ “ทำนาย” อนาคตด้วย Data ในอดีตที่เรามี หรือที่เราเรียกว่า Business Intelligence นั่นเอง

เช่น นำข้อมูลการขายสินค้าทั้งหมด 10 ปีที่ผ่านมา เอามาทำนายยอดขายปีหน้า

5 ระดับของการใช้ Data Science ในองค์กร

ระดับที่ 5

Advise = ระดับที่ห้า คือ การ “**รู้แจ้ง**” ด้วย Data

เป็นระดับที่ระบบจะฉลาดมาก สามารถแนะนำได้ว่า ถ้าอยากทำให้ได้ผลกำไรสูงสุดต้องทำอย่างไร

กิจกรรม

คิดว่ามหาวิทยาลัยเราอยู่ในระดับของการใช้ Data Science ในองค์กร ระดับใด??



<https://www.menti.com/alh32v2gz92k>



บทบาทต่าง ๆ ในสายงาน Data

1. Data Analysts หรือ นักวิเคราะห์ข้อมูล

มีหน้าที่ ในการวิเคราะห์ข้อมูลเชิงธุรกิจ และนำผลลัพธ์ที่ได้มาจัดทำรายงานสรุป โดยแสดงผลในรูปแบบต่างๆ ที่เข้าใจง่ายสำหรับผู้ใช้งาน (Business User) เช่น รูปแบบของชาร์ต (Chart) หรือ แดชบอร์ด (Dashboard) เป็นต้น นอกจากนี้อาจจะช่วยแปลโจทย์ทางด้านธุรกิจให้เป็นหัวข้อทางการวิเคราะห์ข้อมูล เพื่อใช้ประสานงานกับนักวิทยาศาสตร์ข้อมูลสำหรับนำไปวิเคราะห์ข้อมูลในเชิงลึกต่อไปได้อีกด้วย

ทักษะที่สำคัญของ Data Analyst คือ การใช้ซอฟต์แวร์ในการทำได้ เช่น Excel, Tableau, Power BI หรือเขียนโปรแกรมด้วยภาษา SQL ได้ เพื่อแสดงผลการวิเคราะห์เบื้องต้นให้เข้าใจได้ง่าย

2. Data Scientists หรือ นักวิทยาศาสตร์ข้อมูล

มีหน้าที่ วิเคราะห์ข้อมูลเชิงลึกด้วยเทคนิคต่างๆ เช่น วิทยาศาสตร์ข้อมูล (Data Science) เหมืองข้อมูล (Data Mining) การเรียนรู้ของเครื่อง (Machine Learning) โดยอาจจะมีการพัฒนาโมเดล (Model) ทางคณิตศาสตร์ขึ้นมาใหม่ เพื่อให้ตอบโจทย์ทางธุรกิจของผู้ใช้ให้ได้มากที่สุด










ทักษะที่สำคัญของนักวิทยาศาสตร์ข้อมูล คือ ความเข้าใจทางคณิตศาสตร์และสถิติ สามารถเขียนโปรแกรม หรือใช้งานซอฟต์แวร์เพื่อวิเคราะห์ข้อมูลเชิงลึกได้ โดยภาษาโปรแกรมมิ่ง ที่ได้รับความนิยม เช่น การเขียนโปรแกรมด้วยภาษา R, Python และ SQL หรือซอฟต์แวร์ RapidMiner

3. Data Engineering หรือ วิศวกรข้อมูล

ตำแหน่งนี้จะไม่ค่อยได้ติดต่อกับผู้ใช้โดยตรง จะเน้นในส่วนของการจัดการดูแลระบบ เพื่อให้รองรับการทำงานของนักวิทยาศาสตร์ข้อมูลให้มี ประสิทธิภาพทั้งในแง่ของความเร็วในการประมวลผล และการรองรับข้อมูลที่มีปริมาณมากได้

ทักษะที่สำคัญของวิศวกรข้อมูล คือ การใช้เทคโนโลยีในการจัดการข้อมูลต่าง ๆ เช่น SQL ในการจัดการฐานข้อมูล หรือการติดตั้งและจัดการระบบ Hadoop ภายในองค์กร

ความแตกต่างด้านทักษะ และเครื่องมือที่ใช้ในการทำงานของบทบาทต่าง ๆ ในสายงาน Data

 DATA SCIENTISTS A.K.A. STATISTICIANS, DATA MANAGERS	 DATA ENGINEERS A.K.A. DATA ARCHITECTS, DATABASE ADMINISTRATORS	 DATA ANALYSTS A.K.A. BUSINESS ANALYSTS
SKILLS Mathematics, Programming Communication	SKILLS Programming Mathematics Big Data	SKILLS Statistics Communication Business Knowledge
		
TOOLS SQL, Python, R	TOOLS Hadoop, NoSQL, Python	TOOLS Excel, Tableau, SQL
		

มุมมองการใช้ข้อมูลของ 3 คน

Data Engineer



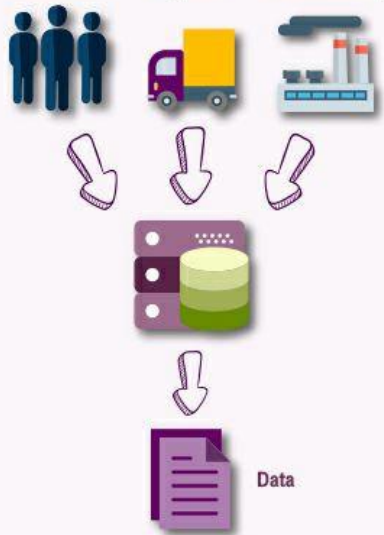
Data Scientist



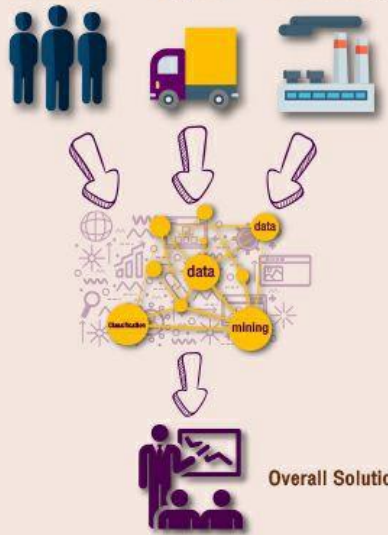
Business and Data Analyst



Customer Logistics Manufacturing



Customer Logistics Manufacturing



Customer Logistics Manufacturing



กิจกรรม

อยากทำงานตำแหน่งไหนกัน ใน 3 ตำแหน่งนี้ เพราะอะไร



pic. <https://www.youtube.com/@cosbit>

ตัวอย่างการใช้ Data Science

ตัวอย่างการใช้ Data Science : Class Cafe

มารุต ชุ่มขุนทด



cr. <https://www.brandbuffet.in.th/>

ตัวอย่างการใช้ Data Science : Class Cafe

เบื้องหลังของ Class Cafe เต็มไปด้วย AI, Machine Learning เข้ามาช่วยในการทำ Data Prediction, Data Analytics เพื่อคาดเดาพฤติกรรมลูกค้าว่าวันนี้จะสั่งเครื่องดื่มอะไร หรือหาแนวโน้มจากปัจจัยต่าง ๆ

เช่น วันนี้อากาศร้อนมาก หรือจะเป็นเครื่องดื่มแบบปั่นดีไหม รวมถึงการวิเคราะห์พื้นที่ในร้านผ่าน Heat Map เพื่อให้เห็นว่า พื้นที่ในจุดใดที่ไม่ค่อยมีคนใช้งาน

ตัวอย่างการใช้ Data Science : Class Cafe

หรือในสาขาเช่นทฤษฎีสถิติวิสัย ที่พบว่า ลูกค้าส่วนใหญ่เป็นกลุ่มครอบครัว ทางร้านก็สามารถจับคู่กับพาร์ทเนอร์ธุรกิจ เช่น โครงการอสังหาริมทรัพย์ หรือสินค้าแม่และเด็ก เข้ามาจัดกิจกรรมใน Class Cafe ได้

ตัวอย่างการใช้ Data Science : Netflix



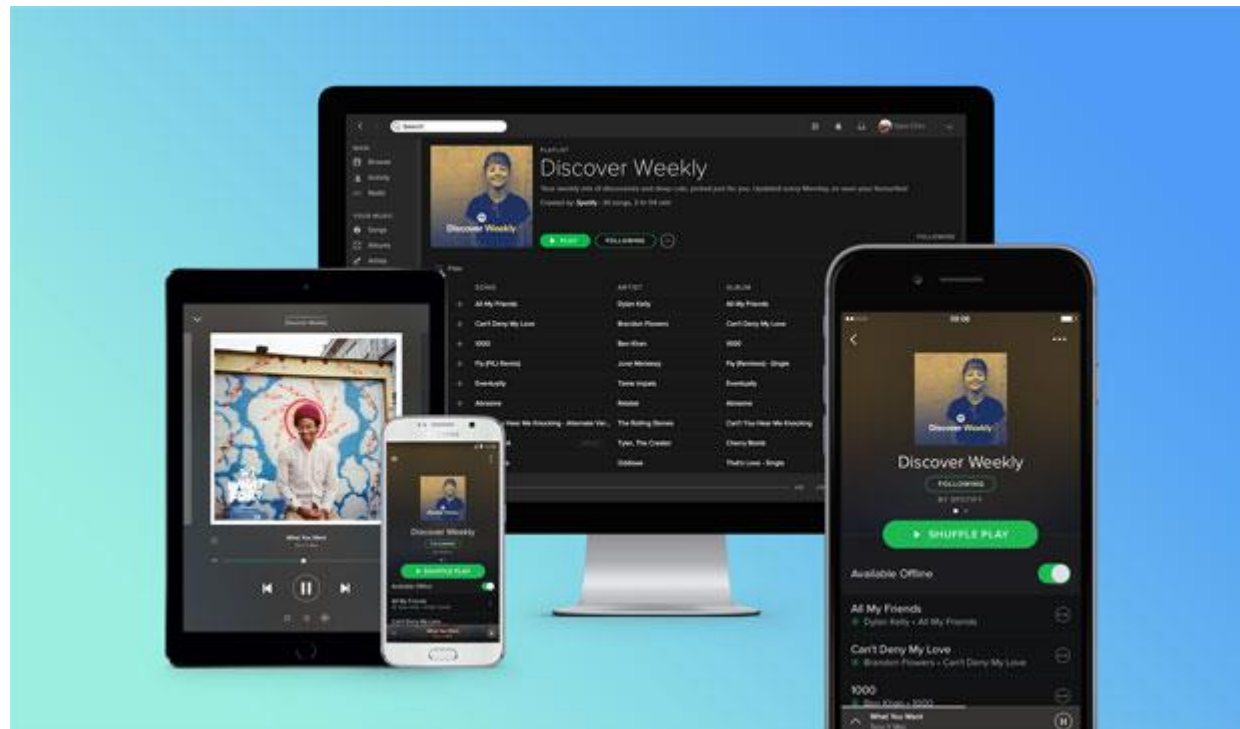
ตัวอย่างการใช้ Data Science : Netflix

ใช้ Big Data ในการวิเคราะห์กลุ่มเป้าหมายในการนำเสนอโฆษณา โดยบริษัทได้รวบรวมข้อมูลจำนวนมากของสมาชิกมากกว่า 100 ล้านรายมาใช้ในการวิเคราะห์ข้อมูล ซึ่งเป็นกุญแจสำคัญที่ทำให้ Netflix ประสบความสำเร็จ

หากเป็นสมาชิกของ Netflix จะคุ้นเคยดีกับวิธีที่ Netflix ส่งคำแนะนำของภาพยนตร์เรื่องต่อไปที่สมาชิกคนนั้นควรดู ขั้นตอนนี้ได้ใช้ข้อมูลจากการค้นหาย้อนหลังของสมาชิก ในการประกอบคำแนะนำได้อย่างแม่นยำ ซึ่งบอกได้ถึงจำนวน % ของภาพยนตร์ว่าตรงต่อความชื่นชอบของสมาชิกแค่ไหน และข้อมูลนี้ทำให้ Netflix สามารถใช้ข้อมูลเชิงลึกได้อย่างเป็นประโยชน์และนำเสนอได้ตรงใจกับลูกค้า

ตัวอย่างการใช้ Data Science : Spotify

Spotify ใช้วิเคราะห์ประเภทของเพลง และพฤติกรรมการฟังเพลงลูกค้า วิเคราะห์ comment และ discussion เกี่ยวกับเพลง เพื่อเข้าใจความต้องการของลูกค้า และวิเคราะห์เสียงเพลง ทำนอง ตัวโน้ต เพื่อแนะนำเพลงให้กับลูกค้า



(Data Science Thailand, 2021)

ตัวอย่างการใช้ Data Science : ระบบแจ้งเตือนการทำงานของระบบชลประทานที่ New York (โครงสร้างพื้นฐาน)



ตัวอย่างการใช้ Data Science : ระบบแจ้งเตือนการทำงานของระบบชลประทานที่ New York (โครงสร้างพื้นฐาน)

Benjamin Brooks, Avishek Kumar และ Syed Ali Asad Rizvi ได้ร่วมมือกับ เมือง Syracuse ที่ New York ในการพัฒนาระบบสนับสนุนการตัดสินใจในการซ่อมแซมการทำงานของระบบชลประทาน

เนื่องจากโครงสร้างระบบชลประทานนี้ถูกสร้างมาตั้งแต่ปี 1900 กว่า ๆ ทำให้การเปลี่ยนโครงสร้างทั้งหมดทำได้ยาก จึงได้มีการค่อย ๆ ซ่อมแซมส่วนที่เสียหายก่อน โดยค่าใช้จ่ายในการซ่อมแซม และเปลี่ยนไปใช้ท่อน้ำใหม่อยู่ที่ประมาณ 1 ล้านดอลลาร์ต่อ 1 ไมล์ แต่ปัญหาคือทางภาครัฐไม่รู้ว่าส่วนไหนควรจะซ่อมก่อน หรือส่วนไหนเสียหายมากที่สุด

ตัวอย่างการใช้ Data Science : ระบบแจ้งเตือนการทำงานของระบบชลประทานที่ New York (โครงสร้างพื้นฐาน)

โครงการนี้เล็งตั้งใจทำระบบสนับสนุนการตัดสินใจในการประเมินท่อน้ำที่มีความเสี่ยงต่อการเสียหาย (Preventive maintenance) เพื่อซ่อมก่อนที่จะแตก โดยใช้ข้อมูลจากทางรัฐตั้งแต่ปี 2004 – 2015 และข้อมูลเกี่ยวกับตัวท่อส่งบริเวณนั้น โดยรายงานออกมาเป็น *Risk score*

ตัวอย่างการใช้ Data Science : Starbucks



ตัวอย่างการใช้ Data Science: Starbucks

สตาร์บัคส์นำเอา AI มาช่วยในการวิเคราะห์ข้อมูลทั้งหมดไม่ว่าจะเป็น location-based ชั้นสูงที่เรียกว่า GIS หรือ Geospatial Information Systems ร่วมกับข้อมูลที่เป็นจำนวนประชากรในละแวกนี้ รายได้ของผู้คนแถวนี้ ปริมาณการเดินผ่านของผู้คนบริเวณนี้ ร้านคู่แข่งทั้งทางตรงและทางอ้อมที่อยู่ใกล้ ๆ และข้อมูลอื่น ๆ อีกมากมาย เท่าที่มีผลต่อรายได้เมื่อเปิดร้าน และผลกำไรในอนาคต

เทคโนโลยี IoT หรือ Internet of Things ที่สตาร์บัคส์เอามาใช้กับเครื่องมือต่าง ๆ ภายในร้านตั้งแต่เครื่องชงกาแฟไปจนถึงเตาอบอาหาร ทำให้ทางสตาร์บัคส์เก็บข้อมูลการทำงานของเครื่องอยู่ตลอดเวลา จนสามารถวิเคราะห์และประเมินล่วงหน้าได้ว่าเครื่องไหนกำลังจะมีปัญหาเล็ก ๆ ที่จะกลายเป็นปัญหาใหญ่ จากนั้นก็จะส่งช่างออกไปจัดการตั้งแต่ยังเป็นปัญหาเล็ก เพื่อไม่ให้เกิดปัญหาใหญ่จนไม่สามารถชงกาแฟให้ลูกค้าได้

ใบงานบทที่ 1

จงค้นหาตัวอย่างการใช้วิทยาศาสตร์ข้อมูลในองค์กรมา 2 ตัวอย่าง เขียนใส่กระดาษที่
อาจารย์ให้ ส่งในคาบหน้า

Reference

Jake VanderPlas. (2016). **Python Data Science Handbook**. the United States of America. : O'Reilly Media, Inc.,

CORALINE CO. LTD. (2017). **Data Engineer, Data Scientist และ Data Analyst ต่างกันอย่างไร.**

Datarockie. (2566). รู้จัก 5 ขั้นตอนการทำงาน Data Science –OSEMN.

<https://datarockie.com/blog/data-science-is-osemn-2/>. กรุงเทพฯ:

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

<https://blog.datath.com/data-science/>