



คณะวิศวกรรมศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

For Preview Only

# การวิเคราะห์ข้อมูล

DATA ANALYTICS



พศ.ดร.รัฐชัย ชาวอุทัย



# TECHNOLOGY SCHOOLBOOK OF ENGINEERING

## ชื่อหนังสือ

ภาษาไทย : การวิเคราะห์ข้อมูล

ภาษาอังกฤษ : Data Analytics

รหัสวิชา : 01076598

ชื่อผู้แต่ง : ผู้ช่วยศาสตราจารย์ ดร.รัฐชัย ชาวอุทัย

ประเมินคุณภาพตำราเรียน : ผ่าน

พิมพ์ครั้งที่ : 1 เดือน พฤษภาคม 2565 จำนวน 100 เล่ม

ISBN : 978-616-338-169-9

วศ.สจล : 294

จำนวนหน้า : 289 หน้า

## ที่ปรึกษาฝ่ายผลิต

ผู้ช่วยศาสตราจารย์ ดร.อนรรฆพล แสนทน

ผู้ช่วยศาสตราจารย์ ดร.ภาสุ พูนภักดี

นายชาคริต เทียนทอง

## ประธานการจัดพิมพ์

น.ส.สุชาดา แดงอินทวัฒน์

## ออกแบบรูปเล่ม

นายทองศักดิ์ ใจชื่นแสน

## จัดทำโดย : งานเทคโนโลยีการศึกษา คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เลขที่ 1 ซอย ฉลองกรุง 1 แขวงลาดกระบัง เขตลาดกระบัง กรุงเทพมหานคร 10520

โทร. 02-329-8000 (ต่อ 3366)

## จัดจำหน่ายโดย : ศูนย์จำหน่ายตำรา คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

[www.facebook.com/technologybookkmitteng](http://www.facebook.com/technologybookkmitteng)

โทร. 02-329-8000 (ต่อ 3492)

## พิมพ์ที่ : ห้างหุ้นส่วนจำกัด มีน เซอร์วิส ซัพพลาย

88/8 ถนนฉลองกรุง แขวงลำปลาทิว เขตลาดกระบัง กรุงเทพฯ 10520

โครงการส่งเสริมการผลิตหนังสือและตำราเพื่อความเป็นเลิศทางวิชาการ  
คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
สงวนลิขสิทธิ์ตามพระราชบัญญัติการพิมพ์

“

หากคุณค่าจากหนังสือเล่มนี้ สามารถเป็นบันไดอีกขั้น  
ให้ผู้อ่านได้ใช้ประโยชน์จากการวิเคราะห์ข้อมูล  
มาพัฒนาตนเอง องค์กร สังคม บ้านเมือง หรือสิ่งแวดล้อม ให้ดีขึ้น  
ผู้เขียนขออุทิศความดีดังกล่าว ให้ครูบาอาจารย์ ครอบครัว มิตรสหาย  
และผู้มีพระคุณต่อผู้เขียนทุกท่าน

”

# คำนำ

จากคำพูดยอดนิยม “ข้อมูลเป็นแหล่งน้ำมันใหม่” (Data is the new oil.) ของ ไคลฟ์ ฮัมบี้ นักคณิตศาสตร์ชาวอังกฤษและผู้ประกอบการด้านการวิเคราะห์ข้อมูล เป็นคำพูดที่สะท้อนความจริงของการแข่งขันทางธุรกิจในยุคปัจจุบัน ซึ่งการแข่งขันเดิมคือยุทธศาสตร์ด้านแหล่งน้ำมัน แต่จากนี้ไปคือการถือครองสินทรัพย์ที่เป็นข้อมูลที่สามารถนำไปต่อยอดสร้างมูลค่าให้กับองค์กร ซึ่งหลายองค์กรระดับโลกที่ประสบความสำเร็จด้านเทคโนโลยี เช่น กูเกิล (Google), เฟสบุ๊ก (Facebook), เน็ตฟลิกซ์ (Netflix) เป็นต้น ล้วนแต่สร้างรายได้เป็นกอบเป็นกำจากการถือครองข้อมูลมหาศาล ดังนั้นไม่ว่าใครก็อยากได้ข้อมูล แต่การมีข้อมูลเพียงอย่างเดียวมันไม่พอ ต้องนำมาวิเคราะห์ประมวลจนได้ผลลัพธ์ที่เป็นผลิตภัณฑ์ที่สร้างมูลค่าหรือสามารถสนับสนุนให้องค์กรตัดสินใจกำหนดทิศทางและกลยุทธ์ได้ถูกต้อง ก็จะมีมูลค่าอย่างสูงกับองค์กร

ด้วยตระหนักในความสำคัญดังที่กล่าวมานี้ หนังสือเล่มนี้จึงได้จัดทำขึ้นเพื่อถ่ายทอดความรู้ด้านการวิเคราะห์ข้อมูล โดยผู้เขียนได้ใช้ความรู้จากหนังสือ ตำรา และบทความวิชาการ ทั้งในและต่างประเทศ รวมถึงประสบการณ์การสอนวิชาด้านการวิเคราะห์ข้อมูลและปัญญาประดิษฐ์ การทำงานวิจัย และการให้บริการทางวิชาการ มาตกผลึกและเรียบเรียงเนื้อหาโดยการดำเนินเรื่องราวตามกระบวนการวิทยาศาสตร์ ข้อมูล ตั้งแต่การรวบรวมข้อมูล, การจัดการประมวลผลข้อมูลให้พร้อมใช้, การสำรวจพฤติกรรมของข้อมูล, การวิเคราะห์ข้อมูลสร้างโมเดลเชิงพยากรณ์ วิเคราะห์จัดกลุ่ม และระบบแนะนำ ที่มีการใช้เทคนิคด้านปัญญาประดิษฐ์มาเป็นส่วนสำคัญ พร้อมทั้งการวัดประสิทธิภาพของโมเดล, และการนำผลการวิเคราะห์ไปใช้งาน รวมถึงกรณีศึกษาที่เป็นงานวิจัยของผู้เขียน โดยเนื้อหาทั้งทฤษฎีและแนวทางการเขียนโปรแกรมภาษาไพธอน เพื่อให้ผู้อ่านสามารถเรียนรู้และฝึกปฏิบัติจนสามารถปฏิบัติงานจริงได้

สุดท้ายนี้ผู้เขียนจึงขอส่งมอบองค์ความรู้ผ่านหนังสือเล่มนี้ สำหรับผู้ที่ต้องการนำศาสตร์ด้านการวิเคราะห์ข้อมูลไปใช้จริง ไม่ว่าจะเป็นด้านธุรกิจ หรือเป็นแหล่งอ้างอิงสำหรับงานวิจัย หรือนำไปพัฒนาสิ่งแวดลอมให้ดีขึ้น อย่างไรก็ตามศาสตร์ด้านการวิเคราะห์ข้อมูลนี้มีความก้าวหน้าอยู่ตลอดเวลา ผู้เขียนยินดีน้อมรับความคิดเห็นจากผู้อ่านเพื่อปรับปรุงเนื้อหาให้ทันสมัยมากขึ้น

รัฐชัย ชาวอุทัย

# สารบัญ

คำนำ .....	5
สารบัญ .....	7
สารบัญรูปภาพ .....	11
สารบัญตาราง .....	15
<b>1. บทนำ .....</b>	<b>17</b>
1.1. การวิเคราะห์ข้อมูล .....	19
1.2. ระดับการวิเคราะห์ข้อมูล .....	30
1.3. กระบวนการของวิทยาศาสตร์ข้อมูล .....	33
1.4. ขอบเขตและโครงสร้างของหนังสือเล่มนี้ .....	36
1.5. สรุปท้ายบท .....	38
1.6. คำถามท้ายบท .....	39
1.7. เอกสารอ้างอิง .....	39
<b>2. การจัดเก็บข้อมูล .....</b>	<b>41</b>
2.1. อินเทอร์เน็ตของสรรพสิ่ง .....	42
2.2. ประเภทข้อมูล .....	43
2.3. ระดับข้อมูล .....	47
2.4. การรวบรวมข้อมูล .....	49
2.5. ตัวอย่างข้อมูลที่ใช้ในหนังสือเล่มนี้ .....	52
2.6. สรุปท้ายบท .....	57
2.7. คำถามท้ายบท .....	58
2.8. เอกสารอ้างอิง .....	58
<b>3. การจัดการประมวลผลข้อมูล .....</b>	<b>61</b>
3.1. เหตุผลที่ต้องจัดการข้อมูล .....	62
3.2. แนวทางการทำความสะอาดข้อมูล .....	63
3.3. แนวทางการแปลงรูปข้อมูล .....	68
3.4. การใช้ไพธอนสำหรับการวิเคราะห์ข้อมูล .....	72

3.5.	โครงสร้างข้อมูลของไพธอน .....	75
3.6.	การใช้ไลบรารีนิ้มไฟ.....	78
3.7.	การใช้ไลบรารีแพนดาส.....	80
3.8.	สรุปท้ายบท .....	93
3.9.	คำถามท้ายบท .....	94
3.10.	เอกสารอ้างอิง .....	94
<b>4.</b>	<b>การสำรวจข้อมูลเบื้องต้น.....</b>	<b>97</b>
4.1.	ทบทวนฟังก์ชันสถิติพื้นฐาน.....	98
4.2.	การสำรวจการกระจายของข้อมูล .....	101
4.3.	ความสัมพันธ์ของข้อมูล .....	104
4.4.	ตัวอย่างการสำรวจข้อมูล .....	107
4.5.	สรุปท้ายบท .....	113
4.6.	คำถามท้ายบท .....	113
4.7.	เอกสารอ้างอิง.....	114
<b>5.</b>	<b>แนวทางการวิเคราะห์ข้อมูล.....</b>	<b>115</b>
5.1.	โมเดล.....	116
5.2.	การเลือกเทคนิคสำหรับสร้างโมเดล.....	118
5.3.	แนวทางการวัดผลโมเดล.....	122
5.4.	กระบวนการสร้างโมเดล .....	126
5.5.	สรุปท้ายบท .....	130
5.6.	คำถามท้ายบท .....	131
5.7.	เอกสารอ้างอิง.....	131
<b>6.</b>	<b>การวิเคราะห์การถดถอย.....</b>	<b>133</b>
6.1.	สมการถดถอยเชิงเส้น .....	134
6.2.	การวัดผลการวิเคราะห์การถดถอย .....	138
6.3.	การเลือกตัวแปรต้นหรือพีเจอร์.....	147
6.4.	การปรับขนาดช่วงข้อมูล.....	149
6.5.	ขั้นตอนการพัฒนาโมเดลสมการถดถอยเชิงเส้น.....	151
6.6.	สรุปท้ายบท .....	156
6.7.	คำถามท้ายบท .....	156

6.8.	เอกสารอ้างอิง.....	157
<b>7.</b>	<b>การวิเคราะห์จำแนกประเภท.....</b>	<b>159</b>
7.1.	รู้จักการวิเคราะห์จำแนกประเภท .....	160
7.2.	การวัดผลการวิเคราะห์จำแนกประเภท .....	161
7.3.	ขั้นตอนการวิเคราะห์จำแนกประเภท .....	174
7.4.	เทคนิคการวิเคราะห์จำแนกประเภท .....	177
7.5.	การเปรียบเทียบประสิทธิภาพของโมเดล.....	198
7.6.	สรุปท้ายบท .....	203
7.7.	คำถามท้ายบท .....	204
7.8.	เอกสารอ้างอิง.....	205
<b>8.</b>	<b>การวิเคราะห์จัดกลุ่ม.....</b>	<b>207</b>
8.1.	รู้จักการวิเคราะห์จัดกลุ่ม .....	208
8.2.	การประเมินผลการวิเคราะห์จัดกลุ่ม.....	209
8.3.	เทคนิคที่ใช้สำหรับการวิเคราะห์จัดกลุ่ม .....	210
8.4.	สรุปท้ายบท .....	222
8.5.	คำถามท้ายบท .....	222
8.6.	เอกสารอ้างอิง.....	223
<b>9.</b>	<b>ระบบแนะนำ.....</b>	<b>225</b>
9.1.	รู้จักกับระบบแนะนำ.....	227
9.2.	การวัดผลระบบแนะนำ.....	231
9.3.	การพัฒนาระบบแนะนำด้วยวิธีพิจารณาการกรองร่วมกัน.....	236
9.4.	การพัฒนาระบบแนะนำแบบผสม.....	241
9.5.	สรุปท้ายบท .....	244
9.6.	คำถามท้ายบท .....	245
9.7.	เอกสารอ้างอิง.....	245
<b>10.</b>	<b>การนำเสนอข้อมูลเชิงภาพ.....</b>	<b>247</b>
10.1.	การใช้ภาพสื่อความหมายข้อมูล.....	248
10.2.	การรับรู้ทางการมองเห็นของมนุษย์ .....	250
10.3.	การแปลงข้อมูลให้เป็นภาพ.....	255
10.4.	แนวทางการนำเสนอข้อมูลด้วยภาพ .....	258

10.5.	สรุปท้ายบท.....	261
10.6.	คำถามท้ายบท.....	261
10.7.	เอกสารอ้างอิง .....	262
<b>11.</b>	<b>กรณีศึกษางานวิจัยด้านการวิเคราะห์ข้อมูล .....</b>	<b>263</b>
11.1.	การตรวจจับถนนชำรุดด้วยเซ็นเซอร์ไร้โจโร .....	264
11.2.	การตรวจใช้เซ็นเซอร์ไมโครเวฟสำหรับตรวจจับการเตะ .....	266
11.3.	ระบบแนะนำนักชีววิทยาเพื่อค้นหาคู่การเกิดเห็ดราบนพืชและสัตว์ .....	269
11.4.	ระบบแนะนำจุดจอดพักรถสำหรับรถ .....	271
11.5.	สรุปท้ายบท.....	274
11.6.	คำถามท้ายบท.....	275
11.7.	เอกสารอ้างอิง .....	275
<b>12.</b>	<b>บทสรุป.....</b>	<b>277</b>
	<b>บรรณานุกรม .....</b>	<b>281</b>
	<b>ประวัติผู้เขียน .....</b>	<b>283</b>



# สารบัญรูปภาพ

รูปที่ 1.1-1 : ทักษะของงานวิทยาศาสตร์ข้อมูล.....	29
รูปที่ 1.2-1 : แผนภาพแสดงระดับการวิเคราะห์ข้อมูล .....	31
รูปที่ 1.3-1 : แผนภาพกระบวนการวิทยาศาสตร์ข้อมูล.....	33
รูปที่ 2.0-1 : ขั้นตอนการจัดเก็บข้อมูลในกระบวนการวิทยาศาสตร์ข้อมูล .....	42
รูปที่ 2.1-1 : แผนภาพทั่วไปของสถาปัตยกรรมอินเทอร์เน็ตของสรรพสิ่ง .....	43
รูปที่ 2.2-1 : รูปภาพแสดงตัวอย่างข้อมูลอายุและส่วนสูงของนักเรียน.....	45
รูปที่ 2.2-2 : คลื่นเสียงสมมติตัวอย่างข้อมูลอายุและส่วนสูงของนักเรียน.....	46
รูปที่ 3.0-1 : ขั้นตอนการจัดการประมวลผลข้อมูลในกระบวนการวิทยาศาสตร์ข้อมูล.....	62
รูปที่ 3.2-1 : คะแนนมาตรฐานในส่วนโค้งปกติของข้อมูล .....	66
รูปที่ 3.3-1 : ตัวอย่างการแปลงข้อมูลลอการิทึม .....	68
รูปที่ 3.3-2 : ตัวอย่างการเข้ารหัสข้อมูล .....	69
รูปที่ 3.3-3 : ตัวอย่างการเข้ารหัสข้อมูลแยกคอลัมน์.....	69
รูปที่ 3.3-4 : ตัวอย่างการระบุช่วงข้อมูล .....	70
รูปที่ 3.3-5 : ตัวอย่างการปรับขนาดช่วงข้อมูลให้ปกติตามค่าต่ำสุดสูงสุด.....	71
รูปที่ 3.3-6 : ตัวอย่างการตัดแบ่งข้อมูล .....	71
รูปที่ 3.4-1 : หน้าจอของกูเกิลโคแอลป์บนเว็บเบราว์เซอร์ .....	73
รูปที่ 3.4-2 : หน้าจอของอะนาคอนดาเนวิเกเตอร์ บนเครื่องคอมพิวเตอร์.....	74
รูปที่ 3.4-3 : หน้าจอของแอปพลิเคชันสเปย์เดอร์บนเครื่องคอมพิวเตอร์ .....	75
รูปที่ 4.0-1 : ขั้นตอนการสำรวจข้อมูลเบื้องต้นในกระบวนการวิทยาศาสตร์ข้อมูล .....	98
รูปที่ 4.3-1 : ตัวอย่างแผนภูมิฮิสโทแกรมของส่วนสูงนักเรียนโรงเรียนแห่งหนึ่ง .....	101
รูปที่ 4.3-2 : รูปแบบแผนภูมิรูปกล่อง .....	103
รูปที่ 4.4-1 : รูปแบบค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน.....	104
รูปที่ 4.4-2 : ตัวอย่างแผนภูมิสแกเตอร์แสดงความสัมพันธ์ระหว่างน้ำหนักกับส่วนสูง .....	106
รูปที่ 5.0-1 : ขั้นตอนการสร้างโมเดลในกระบวนการวิทยาศาสตร์ข้อมูล.....	116
รูปที่ 5.2-1 : โครงสร้างโจทย์การทำงานด้านวิเคราะห์ข้อมูล.....	119
รูปที่ 5.3-1 : แนวทางการแบ่งข้อมูลเพื่อการทดสอบ .....	123
รูปที่ 5.3-2 : แนวทางการวัดผลด้วยวิธีการแบ่งส่วนทดสอบหลายรอบ (ที่ $K=3$ ).....	125

รูปที่ 5.4-1 : แผนภาพแสดงกระบวนการสร้างโมเดล.....	127
รูปที่ 6.0-1 : การวิเคราะห์หัตถด้อยในกระบวนการวิทยาศาสตร์ข้อมูล.....	134
รูปที่ 6.1-1 : แผนภูมิสแกตเตอร์ของตัวอย่างข้อมูลน้ำหนักและส่วนสูงของนักเรียน.....	136
รูปที่ 6.1-2 : ตัวอย่างโมเดลสมการถดถอยเชิงเส้นสำหรับพยากรณ์ส่วนสูงของนักเรียน.....	137
รูปที่ 7.0-1 : การวิเคราะห์จำแนกประเภทในกระบวนการวิทยาศาสตร์ข้อมูล.....	160
รูปที่ 7.2-1 : แผนภาพลำดับการหาค่าวัดผลวิเคราะห์จำแนกประเภท .....	162
รูปที่ 7.2-2 : ตัวอย่างการอธิบายเมตริกซ์คอนฟิวชัน .....	165
รูปที่ 7.2-3 : ตัวอย่างการอธิบายค่าฟรียีชันและรีคอล .....	169
รูปที่ 7.4-1 : กราฟจากสมการถดถอยโลจิสติก.....	178
รูปที่ 7.4-2 : แผนภาพตัวอย่างโมเดลต้นไม้ตัดสินใจ .....	182
รูปที่ 7.4-3 : แผนภาพตัวอย่างการทำงานของเทคนิคเคเนียร์เรสเนเบอร์ .....	192
รูปที่ 7.4-4 : แผนภาพโครงสร้างของเพอร์เซปตรอน .....	194
รูปที่ 7.4-5 : ตัวอย่างฟังก์ชันกระตุ้น.....	195
รูปที่ 7.4-6 : แผนภาพโครงสร้างของโครงข่ายประสาทเทียม.....	196
รูปที่ 8.0-1 : การวิเคราะห์จัดกลุ่มในกระบวนการวิทยาศาสตร์ข้อมูล.....	208
รูปที่ 8.1-1 : ตัวอย่างการจัดกลุ่ม .....	209
รูปที่ 8.3-1 : ตัวอย่างการจัดกลุ่มด้วยเทคนิคเคมีนส์ .....	211
รูปที่ 8.3-2 : ตัวอย่างการจัดกลุ่มที่ซับซ้อนขึ้น .....	219
รูปที่ 8.3-3 : หลักการทำงานของเทคนิคดีปีสแกน .....	220
รูปที่ 8.3-4 : ตัวอย่างผลวิเคราะห์การจัดกลุ่มด้วยเทคนิคดีปีสแกน .....	220
รูปที่ 9.0-1 : ขอบเขตระบบแนะนำในกระบวนการวิทยาศาสตร์ข้อมูล.....	226
รูปที่ 9.1-1 : ตัวอย่างการแนะนำสินค้าในระบบพานิชย์อิเล็กทรอนิกส์.....	228
รูปที่ 9.1-2 : ภาพรวมของระบบแนะนำ.....	228
รูปที่ 9.1-3 : ตัวอย่างการแนะนำไอเท็มตามความเหมือนของข้อมูลผู้ใช้.....	229
รูปที่ 9.1-4 : ตัวอย่างการแนะนำไอเท็มตามพฤติกรรมร่วมกันของการซื้อไอเท็ม.....	229
รูปที่ 9.1-5 : ตัวอย่างการแนะนำไอเท็มตามความคล้ายของข้อมูลไอเท็ม .....	230
รูปที่ 9.1-6 : ตัวอย่างการแนะนำไอเท็มตามองค์ความรู้.....	231
รูปที่ 9.3-1 : ตัวอย่างพฤติกรรมทำให้คะแนนไอเท็มของผู้ใช้ .....	239
รูปที่ 9.4-1 : ตัวอย่างระบบแนะนำของแอปพลิเคชันวิดีโอแห่งหนึ่ง .....	244
รูปที่ 10.0-1 : ขั้นตอนการนำเสนอข้อมูลด้วยภาพในกระบวนการวิทยาศาสตร์ข้อมูล.....	248
รูปที่ 10.1-1 : แผนภูมิเส้นแสดงยอดขายรายเดือนของแต่ละผลิตภัณฑ์ .....	249

รูปที่ 10.1-2 : แผนภูมิแท่งซ้อนสามมิติแสดงยอดขายรายเดือนของแต่ละผลิตภัณฑ์.....	250
รูปที่ 10.2-1 : การมองเห็นของขนาด.....	251
รูปที่ 10.2-2 : ตัวอย่างแผนภูมิที่ใช้ขนาด.....	251
รูปที่ 10.2-3 : ตัวอย่างแผนภูมิแท่งแสดงคะแนนภาพยนตร์.....	252
รูปที่ 10.2-4 : การมองเห็นตำแหน่ง และตัวอย่างแผนภูมิ.....	252
รูปที่ 10.2-5 : การใช้ตำแหน่งบนแผนที่.....	253
รูปที่ 10.2-6 : หลักการของการใช้สี.....	254
รูปที่ 10.4-1 : ตัวอย่างการนำเสนอข้อมูลเชิงภาพแสดงสถิติการใช้เว็บไซต์หนึ่ง.....	258
รูปที่ 11.1-1 : ตำแหน่งที่ตรวจพบถนนชำรุดจากโมเดลต้นไม้การตัดสินใจที่พัฒนาขึ้น.....	266
รูปที่ 11.2-1 : สัญญาณที่ได้รับจากเซ็นเซอร์ไมโครเวฟตามกิจกรรมต่าง ๆ.....	267
รูปที่ 11.4-1 : แผนภาพแสดงลักษณะการเลือกจุดจอดถัดไป.....	272
รูปที่ 12.0-1 : ขั้นตอนการวิเคราะห์ข้อมูลในกระบวนการวิทยาศาสตร์ข้อมูล.....	278

บทที่

# 1. บทนำ

*Introduction*

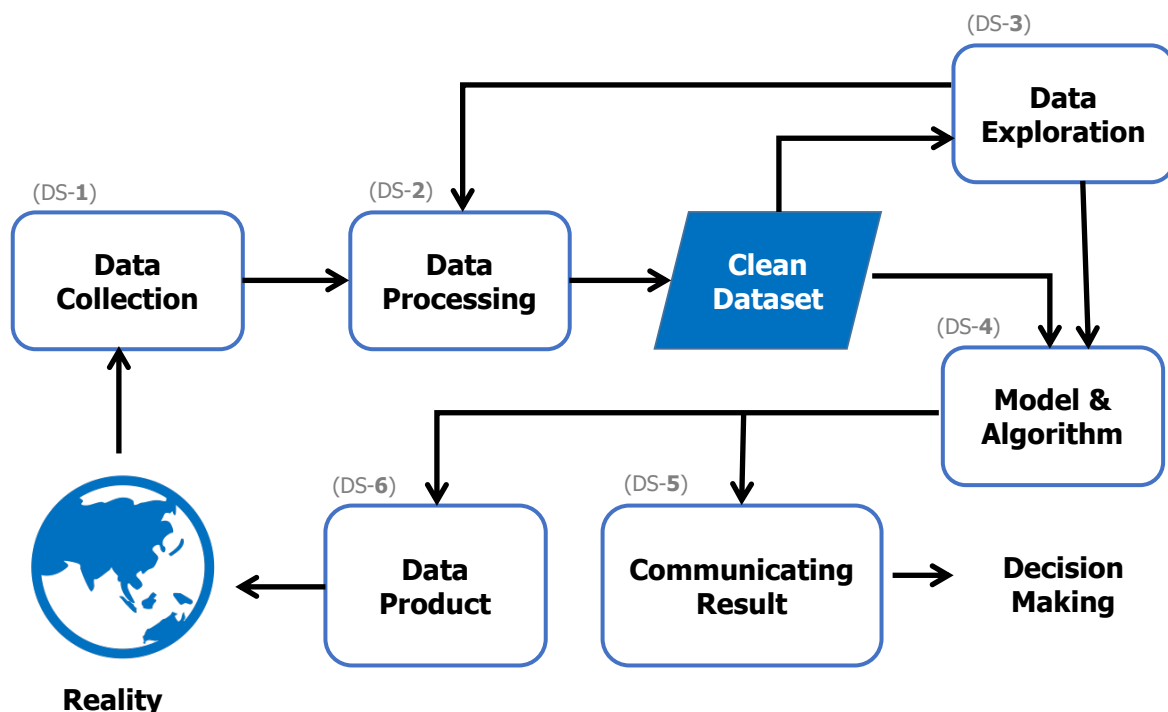
*“Data is the new oil.”*

*- Clive Humby*

อัตโนมัติ, และรวมไปถึงแชทบอทด้วย (Chatbot) เพราะแชทบอทจะพยายามเข้าใจเจตนาของคู่สนทนา แล้วเลือกคำตอบที่ดีที่สุดเพื่อตอบคู่สนทนาให้อัตโนมัติ เป็นต้น

### 1.3. กระบวนการของวิทยาศาสตร์ข้อมูล

กระบวนการของวิทยาศาสตร์ข้อมูลได้ถูกนำมาใช้ดำเนินเนื้อหาในหนังสือการวิเคราะห์ข้อมูลเล่มนี้ โดยมีกระบวนการดังแผนภาพใน **รูปที่ 1.3-1** อ้างอิงจากหนังสือ “Doing data science: Straight talk from the frontline” (O’Neil & Schutt, 2013) ที่เริ่มตั้งแต่ข้อมูลจนถึงผลลัพธ์ที่นำไปสู่การสร้างข้อมูล กลับอีกครั้ง และดำเนินการเช่นนี้อย่างต่อเนื่อง โดยอธิบายในแต่ละขั้นตอนดังต่อไปนี้



รูปที่ 1.3-1 : แผนภาพกระบวนการวิทยาศาสตร์ข้อมูล

#### 1.3.1. สภาพแวดล้อมจริง (Reality)

ข้อมูลทั้งหมดนั้นเริ่มมาจากสภาพแวดล้อมแท้จริง ดังแสดงเป็นสัญลักษณ์รูปโลก **รูปที่ 1.3-1** โดยสื่อให้เห็นว่าทุกกิจกรรมบนโลกมนุษย์รวมถึงจักรวาลล้วนเกิดข้อมูลตลอดเวลา ไม่ว่าจะเป็นความร้อน ความสว่าง การเคลื่อนไหวในทุกกิจกรรม ทุกก้าวเดิน ทุกการเคลื่อนไหว ทุกการสื่อสารของมนุษย์ตั้งแต่อดีตจนถึงปัจจุบันก็ล้วนแต่ทำให้เกิดข้อมูล ซึ่งข้อมูลเหล่านี้จะนำมาใช้ประโยชน์ต่อได้หากได้รับการบันทึกที่ดี

และด้วยความก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยีทำให้เราสามารถบันทึกข้อมูลได้มากขึ้น หลากหลายรูปแบบขึ้น และนำไปทำประโยชน์ได้มากขึ้น

### 1.3.2. การรวบรวมจัดเก็บข้อมูล (Data Collection)

ปัจจุบันเราเก็บข้อมูลได้มีประสิทธิภาพขึ้น เนื่องด้วยความก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยีทำให้ปัจจุบันนี้เรามีอุปกรณ์ในการตรวจจับข้อมูล เช่น เซ็นเซอร์ต่าง ๆ ไม่ว่าจะเป็นเครื่องวัดอุณหภูมิ เครื่องวัดความเข้มแสง เครื่องตรวจจับเสียง เครื่องรับภาพ เครื่องวัดความเร็ว เครื่องวัดความเร่ง เครื่องวัดการหมุน เครื่องวัดสารเคมี และเครื่องวัดต่าง ๆ อีกมากมาย โดยเฉพาะการใช้งานอินเทอร์เน็ตของสรรพสิ่ง (Internet of Things) จะเข้ามามีบทบาทในขั้นตอนนี้ โดยมีนักวิศวกรรมข้อมูลเป็นผู้เล่นสำคัญ ทำให้เราสามารถได้ข้อมูลจากสภาพแวดล้อมจริงมาอยู่ในรูปข้อมูลดิบเพื่อนำไปใช้ต่อในกระบวนการต่อไป ดังขั้นตอนในรูปที่ 1.3-1 (DS-1)

### 1.3.3. การจัดการประมวลผลข้อมูล (Data Processing)

หลังจากที่มีข้อมูลในรูปแบบดิจิทัลแล้ว โดยส่วนใหญ่ยังไม่สามารถใช้ข้อมูลเหล่านั้นได้ทันที เนื่องจากข้อมูลไม่อยู่ในรูปแบบที่พร้อมใช้ จึงต้องเข้าสู่กระบวนการจัดการประมวลผลข้อมูลดังแสดงในรูปที่ 1.3-1 (DS-2) โดยมีกิจกรรมหลักคือ จัดการข้อมูลที่มีค่าว่าง (Missing Data), จัดการค่าผิดปกติ (Outlier), แปลงรูปข้อมูล (Data Transformation), สกัดฟีเจอร์จากข้อมูล (Feature Engineering), ปรับขนาดช่วงข้อมูล (Scaling) เป็นต้น โดยในวงการที่ปฏิบัติงานด้านการวิเคราะห์ข้อมูลให้การทำงานมากกว่าร้อยละ 60 อยู่ที่ขั้นตอนนี้ ดังนั้นจึงต้องเตรียมทั้งกำลังคนและเวลาเพื่อการจัดการประมวลผลข้อมูลให้ดีก่อนนำไปวิเคราะห์

### 1.3.4. ข้อมูลที่พร้อมใช้ (Clean Dataset)

หลังจากที่ข้อมูลผ่านกระบวนการจัดการประมวลผลข้อมูลแล้ว ก็จะได้ข้อมูลที่พร้อมนำไปใช้ในการสร้างโมเดล ซึ่งเป็นข้อมูลที่อยู่ในรูปแบบที่เหมาะสม เช่น อยู่ในรูปแบบตารางที่มีโครงสร้างชัดเจน มีค่าในแต่ละคอลัมน์อยู่ในหน่วยเดียวกัน มีรูปแบบข้อมูลในแต่ละคอลัมน์เป็นแบบเดียวกัน ไม่มีค่าว่างหรือค่าผิดปกติ เป็นต้น

### 1.3.5. การสำรวจข้อมูล (Data Exploration)

ปกติแล้วเมื่อได้ข้อมูลพร้อมใช้แล้วสามารถไปพัฒนาโมเดลต่อได้ แต่ในทางปฏิบัติควรสำรวจข้อมูลเบื้องต้นของชุดข้อมูลเหล่านั้น ก่อนดังขั้นตอนในรูปที่ 1.3-1 (DS-3) เพื่อสังเกตพฤติกรรมของข้อมูล เช่น

ความสัมพันธ์ การกระจายของข้อมูล ขอบเขตช่วงข้อมูล เป็นต้น โดยส่วนใหญ่แล้วจะเป็นการนำเสนอในรูปแบบแผนภูมิต่าง ๆ เช่น แผนภูมิสแกตเตอร์ (Scatter Plot), แผนภูมิฮิสโทแกรม (Histogram Chart), แผนภูมิรูปกล่อง (Box Plot) หรือแผนภูมิอื่น ๆ เป็นต้น เพื่อใช้ทั้งตรวจสอบความสมบูรณ์ของข้อมูล หาแนวโน้มของข้อมูล และเป็นแนวทางในการเลือกใช้เทคนิคในการวิเคราะห์ข้อมูลที่เหมาะสมด้วย

### 1.3.6. การพัฒนาโมเดล (Model & Algorithm)

ขั้นตอนการพัฒนาโมเดลดังขั้นตอนในรูปที่ 1.3-1 (DS-4) เป็นขั้นตอนสำคัญของการวิเคราะห์ โดยใช้เทคนิคด้านการเรียนรู้ของเครื่องจักร (Machine Learning) มาเรียนรู้ชุดข้อมูล จนได้ผลลัพธ์เป็นแบบจำลองทางคณิตศาสตร์ เช่น สมการทางคณิตศาสตร์ กฎทางตรรกศาสตร์ เป็นต้น สำหรับการหาข้อสรุป พยากรณ์ จัดกลุ่ม หรือแนะนำ เป็นต้น ซึ่งขั้นตอนนี้เป็นขั้นตอนที่นักวิทยาศาสตร์ข้อมูลมีบทบาทมากที่สุด โดยพยายามค้นคว้าและพัฒนาให้ได้โมเดลที่เหมาะสมกับโจทย์ที่ได้รับ มีความแม่นยำตามที่ธุรกิจต้องการ มีประสิทธิภาพเหมาะสมกับปริมาณข้อมูลและเทคโนโลยีที่ใช้

หลังจากได้โมเดลที่ดีแล้ว ลำดับต่อไปจะเป็นทางเลือก 2 ทาง ดังแสดงในหัวข้อถัดไป คือ การนำเสนอผลการวิเคราะห์ในรูปแบบการสื่อสาร รายงาน หรือแผนภาพ, หรือการเป็นผลิตภัณฑ์จากข้อมูลซึ่งอาจจะเป็นแอปพลิเคชันที่นำข้อมูลไปใช้งานจริง

### 1.3.7. การนำเสนอผลการวิเคราะห์ (Communicating Result)

ในการนำเสนอข้อมูลผลการวิเคราะห์ดังขั้นตอนในรูปที่ 1.3-1 (DS-5) เป็นจำเป็นการนำเสนอผลการวิเคราะห์มาจัดรูปแบบใหม่เพื่อการสื่อสาร (Communication), การออกรายงาน (Report), หรือการนำเสนอเชิงภาพ (Visualization) เพื่อให้ผู้รับสาร ซึ่งอาจจะเป็นหัวหน้างาน ลูกค้า จนถึงบุคคลทั่วไปเข้าใจผลการวิเคราะห์ที่จะสื่อสารได้อย่างตรงประเด็น รวดเร็ว และสามารถนำไปประกอบการตัดสินใจ (Decision) ได้อย่างมีประสิทธิภาพ โดยในขั้นตอนนี้จะมีนักวิเคราะห์ข้อมูลธุรกิจเข้ามามีบทบาทสำคัญที่ต้องอาศัยทั้งศาสตร์และศิลป์ ความสามารถในการรับรู้ของมนุษย์ และความต้องการของผู้รับสาร เป็นปัจจัยในการออกแบบการนำเสนอ

### 1.3.8. การพัฒนาผลิตภัณฑ์จากการวิเคราะห์ข้อมูล (Data Product)

ผลิตภัณฑ์จากการวิเคราะห์ข้อมูลดังขั้นตอนในรูปที่ 1.3-1 (DS-6) ในที่นี้คือแอปพลิเคชัน ซึ่งใช้ได้ทั้งซอฟต์แวร์หรือฮาร์ดแวร์ ที่ใช้โมเดลจากการวิเคราะห์ข้อมูลเป็นกลไกในการพัฒนา มีการเรียนรู้และปรับปรุงโมเดลเมื่อได้เรียนรู้ชุดข้อมูลใหม่ ๆ โดยมีนักพัฒนามีบทบาทสำคัญในขั้นตอนนี้ ซึ่งผลิตภัณฑ์นี้สามารถเป็นได้ทั้งผลิตภัณฑ์หลักหรือผลิตภัณฑ์ประกอบกับผลิตภัณฑ์อื่น เช่น ระบบแนะนำเว็บไซต์ตาม

คำค้นเป็นผลิตภัณฑ์หลักของระบบค้นหาเว็บ, ระบบการตีความข้อความผลิตภัณฑ์หลักของแชตบอท, ระบบแนะนำสินค้าเป็นผลิตภัณฑ์ประกอบของระบบพาณิชย์อิเล็กทรอนิกส์, ระบบแนะนำเส้นทางเป็นผลิตภัณฑ์ประกอบของแอปพลิเคชันแผนที่ เป็นต้น เมื่อผู้ใช้ได้ใช้งานผลิตภัณฑ์เหล่านี้แล้วตัดสินใจกระทำอย่างใดอย่างหนึ่งก็จะเกิดข้อมูลเข้าสู่สภาพแวดล้อมจริง (Reality) อีกครั้ง ตัวอย่างเช่น ระบบแนะนำเส้นทางของกูเกิลแมพ (Google Map) อาศัยข้อมูลการเดินทางจริงของคนจากสภาพแวดล้อมจริงมาประมวลผล และเลือกเส้นทางที่เหมาะสมให้กับผู้ใช้ เมื่อผู้ใช้เลือกเส้นทางและเดินทางก็จะเกิดข้อมูลในสภาพแวดล้อมจริง และถูกนำไปใช้ประมวลผลให้เกิดประโยชน์กับผลิตภัณฑ์อย่างต่อเนื่อง

## 1.4. ขอบเขตและโครงสร้างของหนังสือเล่มนี้

หนังสือเล่มนี้อธิบายกระบวนการวิเคราะห์ข้อมูลโดยใช้กระบวนการวิทยาศาสตร์ข้อมูล ดังรูปที่ 1.3-1 เป็นแนวทางหลักในการดำเนินการอธิบายเนื้อหา โดยให้ความสำคัญกับเรื่องการพัฒนาโมเดลการวิเคราะห์ข้อมูลที่มีโครงสร้างประเภทตาราง เป็นหลัก และนำเสนอตัวอย่างการเขียนโปรแกรมด้วยภาษาไพธอนกับไลบรารีที่เป็นที่นิยมในการเขียนโปรแกรมวิเคราะห์ข้อมูล หนังสือเล่มนี้จึงเหมาะสำหรับผู้ที่ต้องการเรียนรู้การวิเคราะห์ข้อมูล และมีพื้นฐานการเขียนโปรแกรมมาแล้วโดยเฉพาะภาษาไพธอน โดยเนื้อหาแต่ละบทของหนังสือเล่มนี้มีดังนี้

### บทที่ 1 บทนำ

อธิบายภาพรวมการวิเคราะห์ข้อมูล การตั้งโจทย์ ทีมงานวิเคราะห์ข้อมูล ระดับการวิเคราะห์ข้อมูล และกระบวนการของวิทยาศาสตร์ข้อมูล ซึ่งได้กล่าวในบทนี้

### บทที่ 2 การจัดเก็บข้อมูล

อธิบายวิธีการได้มาของข้อมูล ประเภทของข้อมูล และตัวอย่างข้อมูลที่ใช้ในหนังสือนี้ โดยเนื้อหาครอบคลุมในขั้นตอน สภาพแวดล้อมจริง (Reality) การรวบรวมจัดเก็บข้อมูล (Collecting Data) (DS-1) และที่สำคัญคือการเริ่มอธิบายระดับข้อมูล ประกอบด้วย มาตรฐานนามบัญญัติ (Nominal Scale), มาตรฐานอันดับ (Ordinal Scale), มาตรฐานอัตราส่วน (Interval Scale), และมาตรฐานอัตราส่วน (Ratio Scale) พร้อมตัวอย่างข้อมูล ที่จะเข้ามามีบทบาทในกระบวนการวิเคราะห์ข้อมูลและมีบทบาทสำคัญในหนังสือเล่มนี้ เพื่อใช้พิจารณาการใช้เทคนิคการจัดการประมวลผลข้อมูล การวิเคราะห์ข้อมูล รวมถึงการนำเสนอผลการวิเคราะห์เชิงภาพ



### บทที่ 3 การจัดการประมวลผลข้อมูล

อธิบายการจัดการทำความสะอาดและแปลงรูปข้อมูลข้อมูล โดยเนื้อหาครอบคลุมในขั้นตอน การจัดการประมวลผลข้อมูล (Data Processing) (DS-2) และข้อมูลที่พร้อมใช้ (Clean Dataset) นอกจากนี้ บทนี้เป็นบทที่เริ่มอธิบายการเขียนโปรแกรมภาษาไพธอน (Python) สำหรับการจัดการประมวลผลข้อมูลด้วยไลบรารีนิมโ (Numpy) กับแพนดาส (Pandas) ซึ่งเป็นไลบรารีสำคัญและเป็นที่นิยมสำหรับนักวิเคราะห์ข้อมูล

### บทที่ 4 การสำรวจข้อมูลเบื้องต้น

อธิบายชนิดข้อมูล สถิติพื้นฐาน เช่น การหาค่ากลาง การกระจายของข้อมูล และการหาค่าสหสัมพันธ์ เป็นต้น รวมถึงการสังเกตพฤติกรรมข้อมูลจากค่าสถิติและแผนภูมิต่าง ๆ ประกอบกับตัวอย่างการสำรวจข้อมูล โดยเนื้อหาครอบคลุมขั้นตอน การสำรวจข้อมูล (Exploratory Data Analysis) (DS-3) ดังแสดงในแผนภาพกระบวนการวิทยาศาสตร์ข้อมูล

### บทที่ 5 แนวทางการวิเคราะห์ข้อมูล

อธิบายเรื่องโมเดลการวิเคราะห์ข้อมูล ซึ่งเป็นการเริ่มต้นเข้าสู่เนื้อหาการเรียนรู้ของเครื่องจักร (Machine Learning) ประกอบไปด้วย การเลือกวิธีการสร้างโมเดล แนวทางการวัดผล และกระบวนการสร้างโมเดล โดยเป็นส่วนหนึ่งของขั้นตอน การพัฒนาโมเดล (Model Implementation) (DS-4)

### บทที่ 6 การวิเคราะห์การถดถอย

อธิบายการพัฒนาโมเดลการวิเคราะห์ถดถอย (Regression) สำหรับการพยากรณ์ที่มีผลลัพธ์เป็นตัวเลขต่อเนื่อง โดยเป็นส่วนหนึ่งของขั้นตอน การพัฒนาโมเดล (Model Implementation) (DS-4)

### บทที่ 7 การวิเคราะห์จำแนกประเภท

อธิบายการพัฒนาโมเดลการวิเคราะห์จำแนกประเภท (Classification) สำหรับการพยากรณ์ที่มีผลลัพธ์เป็นกลุ่มประเภท โดยเป็นส่วนหนึ่งของขั้นตอน การพัฒนาโมเดล (Model Implementation) (DS-4)

### บทที่ 8 การวิเคราะห์จัดกลุ่ม

อธิบายการพัฒนาโมเดลการวิเคราะห์จัดกลุ่ม (Clustering) สำหรับการพยากรณ์ที่มีผลลัพธ์เป็นกลุ่มของสิ่งที่มีลักษณะคล้ายกัน โดยงานนี้ถือเป็นส่วนหนึ่งของขั้นตอน การพัฒนาโมเดล (Model Implementation) (DS-4)

## บทที่ 9 ระบบแนะนำ

อธิบายศาสตร์ของการพัฒนาระบบแนะนำ สำหรับการแนะนำสิ่งต่าง ๆ ให้กับผู้ใช้ เช่น แนะนำสินค้าที่น่าจะซื้อโดยการดูบริบทและพฤติกรรมของผู้ใช้ โดยเนื้อหานี้เป็นส่วนหนึ่งของขั้นตอน การพัฒนาโมเดล (Model Implementation) (DS-4) และเป็นกลไกสำคัญของขั้นตอน การพัฒนาผลิตภัณฑ์จากการวิเคราะห์ข้อมูล (Data Product) (DS-6)

## บทที่ 10 การนำเสนอข้อมูลเชิงภาพ

อธิบายศาสตร์และศิลป์ของการนำเสนอข้อมูลเชิงภาพ โดยการทำความเข้าใจความสามารถในการรับรู้ของมนุษย์ และนำสัญลักษณ์ต่าง ๆ มาใช้สื่อความหมายเพื่อให้การนำเสนอมีความเข้าใจง่ายและตรงประเด็น ซึ่งเป็นส่วนสำคัญของขั้นตอน การนำเสนอผลการวิเคราะห์ (Communicating Result) (DS-5)

## บทที่ 11 กรณีศึกษางานวิจัยด้านการวิเคราะห์ข้อมูล

อธิบายกรณีศึกษาจากผลงานวิจัยของผู้เขียน เพื่อให้เห็นภาพการใช้งานจริงของศาสตร์ด้านการวิเคราะห์ข้อมูล

## บทที่ 12 บทสรุป

สำหรับสรุปเนื้อหาทั้งหมดของหนังสือเล่มนี้ใน โดยการดำเนินการสรุปเนื้อหาตามกระบวนการวิทยาศาสตร์ข้อมูล

นอกจากนี้ผู้อ่านสามารถติดตามการเข้าถึงชุดข้อมูลที่ใช้, ตัวอย่างโค้ดการเขียนโปรแกรมภาษาไพธอน, ภาพประกอบสี, และการปรับปรุงเนื้อหาของหนังสือเล่มนี้ได้ที่เว็บไซต์

<https://github.com/Rathachai/DA-LAB>

## 1.5. สรุปท้ายบท

บทนี้กล่าวถึงภาพรวมของการวิเคราะห์ข้อมูล ตั้งแต่ประโยชน์ของการวิเคราะห์ข้อมูลเพื่อกำหนดเป้าหมายสำหรับเป็นแนวทางในการตั้งโจทย์ที่เป็นการตั้งโจทย์เพื่อหาข้อสรุป การตั้งโจทย์เชิงพยากรณ์ และการตั้งโจทย์เชิงแนะนำเพื่อการตัดสินใจ พร้อมกับการกำหนดบทบาทของคนที่ทำงานประกอบด้วยนักวิเคราะห์ข้อมูลธุรกิจ นักวิทยาศาสตร์ข้อมูล นักวิศวกรรมข้อมูล และนักพัฒนา เพื่อให้สามารถตั้งทีมงานพัฒนาการวิเคราะห์ข้อมูล โดยมีระดับการวิเคราะห์ข้อมูลคือ การวิเคราะห์เชิงพรรณนา การวิเคราะห์เชิงพยากรณ์ และการวิเคราะห์เชิงแนะนำ รวมถึงอธิบายกระบวนการของวิทยาศาสตร์ข้อมูลซึ่ง

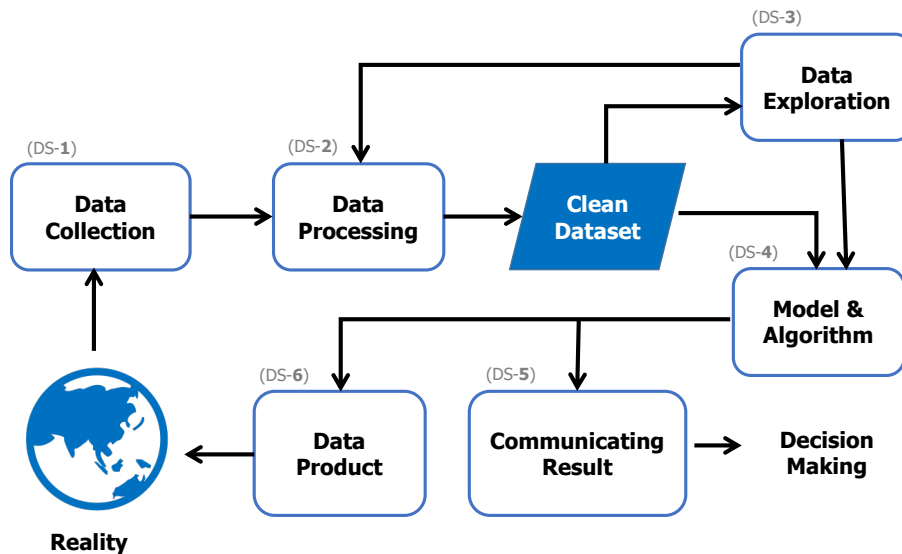
บทที่

## 12. บทสรุป

*Summary*

*“We’re entering a new world in which  
data may be more important than software.”*

*- Tim O’Reilly*



รูปที่ 12.0-1 : ขั้นตอนการวิเคราะห์ข้อมูลในกระบวนการวิทยาศาสตร์ข้อมูล

หนังสือเรียนเล่มนี้ได้บรรยายถึงการวิเคราะห์ข้อมูลตามกระบวนการทางวิทยาศาสตร์ข้อมูล โดยใช้แนวทางกระบวนการทางวิทยาศาสตร์ข้อมูลตามหนังสือ “Doing data science: Straight talk from the frontline” (O’Neil & Schutt, 2013) ดังแสดงในรูปที่ 12.0-1 เป็นแนวทางในการออกแบบโครงสร้างเนื้อหาของหนังสือเล่มนี้โดยใช้ให้แต่ละบททำหน้าที่อธิบายแต่ละขั้นตอน พร้อมใช้โค้ดภาษาไพธอนเพื่อเป็นแนวทางให้ผู้มีพื้นฐานด้านการเขียนโปรแกรม นำไปพัฒนาการวิเคราะห์ข้อมูลได้จริง

### สภาพแวดล้อมจริง (Reality)

สภาพแวดล้อมจริงคือกิจกรรมของสิ่งต่าง ๆ ซึ่งรวมไปถึงกิจกรรมของสิ่งมีชีวิต สิ่งไม่มีชีวิต ทั้งในโลกและนอกโลก ตั้งแต่อดีตถึงปัจจุบันและต่อไปในอนาคตนั้น ได้มีการสร้างข้อมูลตลอดเวลาอย่างต่อเนื่อง และมีข้อมูลจำนวนมากที่ถูกสร้างขึ้นในทุกเสี้ยววินาที โดยขั้นตอนนี้ได้มีการกล่าวถึงในบทที่ 2 การจัดเก็บข้อมูล เพื่อนำไปสู่ขั้นตอนการรวบรวมจัดเก็บข้อมูล

### การรวบรวมจัดเก็บข้อมูล (Data Collection) (DS-1)

ขั้นตอนการรวบรวมจัดเก็บข้อมูลได้อธิบายในบทที่ 2 การจัดเก็บข้อมูล โดยอธิบายถึงการได้ข้อมูลจากสภาพแวดล้อมจริงผ่านการบันทึก โดยเฉพาะอย่างยิ่งการใช้เทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง (Internet of Things: IoT) ที่เข้ามามีบทบาททำให้การจัดเก็บข้อมูลมีประสิทธิภาพมากขึ้น แล้วในบทนี้ยังได้อธิบายถึงประเภทของข้อมูลทั้งข้อมูลที่มีโครงสร้าง และข้อมูลที่ไม่มีโครงสร้าง, การรวบรวมข้อมูล, ชุดข้อมูลที่ใช้ในหนังสือเล่มนี้ นอกจากนี้มีเนื้อหาที่สำคัญคือการอธิบายข้อมูล 4 ระดับ คือ มาตรฐาน

บัญญัติ (Nominal Scale หรือ Classification Scale), มาตราอันดับ (Ordinal Scale), มาตราอันตรภาค (Interval Scale), และมาตราอัตราส่วน (Ratio Scale) เพื่อให้ผู้อ่านสามารถเข้าใจลักษณะข้อมูลได้อย่างดีขึ้น สามารถใช้วิเคราะห์โจทย์ที่ได้รับ และแปลงรูปข้อมูลให้เหมาะสมกับเทคนิคที่เลือกได้

### การจัดการประมวลผลข้อมูล (Data Processing) (DS-2)

ขั้นตอนการจัดการประมวลผลข้อมูล ได้อธิบายในบทที่ 3 การจัดการประมวลผลข้อมูล เนื่องจากการปฏิบัติงานวิเคราะห์ข้อมูลจริงนั้นเป็นการยากที่จะได้ข้อมูลพร้อมใช้ จึงจำเป็นต้องจัดการข้อมูลเพื่อให้ได้ข้อมูลที่พร้อมใช้ โดยอธิบายถึงแนวทางการทำความสะอาดข้อมูล การแปลงรูปข้อมูล และอธิบายการใช้ภาษาไพธอนร่วมกับไลบรารี NumPy และแพนดาส (Pandas) ในการจัดการประมวลผลข้อมูล ซึ่งเป็นพื้นฐานในการเขียนโค้ดการวิเคราะห์ข้อมูลด้วยภาษาไพธอนในบทต่อ ๆ ไป

### ข้อมูลที่พร้อมใช้ (Clean Dataset)

ข้อมูลพร้อมใช้เป็นชุดข้อมูลที่พร้อมนำไปวิเคราะห์ ซึ่งเป็นผลผลิตที่ได้รับจากขั้นตอนการจัดการประมวลผลข้อมูลที่อธิบายไว้ในบทที่ 3 การจัดการประมวลผลข้อมูล

### การสำรวจข้อมูลเบื้องต้น (Exploration) (DS-3)

ขั้นตอนการสำรวจข้อมูลเบื้องต้น ได้อธิบายในบทที่ 4 การสำรวจข้อมูลเบื้องต้น ทบทวนฟังก์ชันสถิติพื้นฐาน เช่น การหาค่ากลาง และค่าการกระจาย รวมถึงอธิบายการใช้แผนภาพเพื่อสำรวจการกระจายข้อมูล และหาความสัมพันธ์ของข้อมูลระหว่างตัวแปรได้ จึงเป็นอีกขั้นตอนที่สำคัญที่จะทำให้ผู้อ่านมีความเข้าใจในพฤติกรรมของข้อมูล และสามารถวางกลยุทธ์ในการวิเคราะห์ข้อมูลในขั้นตอนต่อไปได้

### การพัฒนาโมเดล (Model & Algorithm) (DS-4)

ขั้นตอนการพัฒนาโมเดล ได้อธิบายในบทที่ 5, 6, 7, 8, และ 9 โดยบทที่ 5 แนวทางการวิเคราะห์ข้อมูล ได้อธิบายถึงนิยามของคำว่าโมเดลหรือแบบจำลอง พร้อมทั้งอธิบายแนวทางการวิเคราะห์โจทย์ที่ได้รับเพื่อนำไปสู่การเลือกเทคนิคในการสร้างโมเดลและวิธีการวัดผล รวมถึงกระบวนการและขั้นตอนในการสร้างโมเดลที่มีประสิทธิภาพและน่าเชื่อถือ จากนั้นเมื่อสามารถกำหนดทิศทางการวิเคราะห์ข้อมูลจากโจทย์ที่ได้รับได้แล้ว จึงเข้าสู่บทที่ 6 การวิเคราะห์ถดถอย (Regression Analysis), บทที่ 7 การวิเคราะห์จำแนกประเภท (Classification Analysis), บทที่ 8 การวิเคราะห์จัดกลุ่ม (Cluster Analysis), และบทที่ 9 ระบบแนะนำ (Recommender System) ที่มีการอธิบายแนวคิด เทคนิคการวิเคราะห์สร้างโมเดล วิธีการวัดผลโมเดล และตัวอย่างโค้ดภาษาไพธอน

### การนำเสนอผลการวิเคราะห์ (Communicating Result) (DS-5)

ขั้นตอนการนำเสนอผลการวิเคราะห์ ได้อธิบายในบทที่ 10 การนำเสนอข้อมูลเชิงภาพ โดยมีวัตถุประสงค์เพื่อให้ผู้อ่านได้เข้าใจหลักการนำเสนอผลการวิเคราะห์ข้อมูลโดยใช้แผนภูมิภาพต่าง ๆ ใช้สื่อความหมายแทนข้อความ โดยใช้ความสามารถในการมองเห็นและรับรู้ของมนุษย์มาสนับสนุน และอธิบายหลักการแปลงข้อมูลให้เป็นภาพ เพื่อให้ผู้รับสารสามารถเข้าใจประเด็นสำคัญในสิ่งที่ต้องการสื่อได้อย่างรวดเร็วและตรงประเด็น และสามารถนำไปใช้ตัดสินใจได้ทันที

### การพัฒนาผลิตภัณฑ์จากการวิเคราะห์ข้อมูล (Data Product) (DS-6)

ขั้นตอนการพัฒนาผลิตภัณฑ์จากการวิเคราะห์ข้อมูล ถือเป็นการประยุกต์ใช้โมเดลมาผลิตชิ้นงานจริง โดยบทที่ 9 ระบบแนะนำ ได้วางเส้นทางในการสร้างระบบแนะนำเพื่อให้ผู้อ่านได้แนวคิดไปต่อยอดเป็นฟังก์ชันงานในผลิตภัณฑ์ของตนเอง ซึ่งงานวิจัยที่กล่าวถึงในบทที่ 11 ตัวอย่างงานวิจัยด้านการวิเคราะห์ข้อมูล ได้แสดงให้เห็นถึงการนำโมเดลที่ได้จากการวิเคราะห์นำไปเขียนโปรแกรมเพื่อใช้เป็นระบบย่อยที่ใช้เป็นผลิตภัณฑ์หนึ่งของระบบใหญ่หรือโครงการได้

เพื่อให้เห็นภาพการวิเคราะห์ข้อมูลตั้งแต่ต้นน้ำคือข้อมูล กลางน้ำคือการวิเคราะห์ได้โมเดล และปลายน้ำคือผลงานที่สร้างขึ้น ในบทที่ 11 กรณีศึกษางานวิจัยด้านการวิเคราะห์ข้อมูล ได้อธิบายงานวิจัยของผู้เขียนรวม 4 เรื่อง คือ การตรวจจับถนนชำรุดด้วยเซ็นเซอร์ใจโร, การตรวจใช้เซ็นเซอร์ไมโครเวฟสำหรับตรวจจับการเตะเท้า, ระบบแนะนำนักชีววิทยาเพื่อค้นหาคู่การเกิดเห็ดราบนพืชและสัตว์, และระบบแนะนำจุดจอดพักรถสำหรับรถ ที่ประยุกต์ใช้ศาสตร์ด้านการวิเคราะห์ข้อมูลมาหาคำตอบจากโจทย์ที่ได้รับ

นอกจากนี้ เนื่องจากศาสตร์ด้านการวิเคราะห์ข้อมูลมีความเคลื่อนไหวและความก้าวหน้าตลอดเวลา ผู้อ่านสามารถติดตามการเข้าถึง ชุดข้อมูลที่ใช้ ตัวอย่างโค้ดการเขียนโปรแกรม ภาพสไลด์ และรายการปรับปรุงเนื้อหาของหนังสือเล่มนี้ได้ที่เว็บไซต์

<https://github.com/Rathachai/DA-LAB>

# ประวัติผู้เขียน



## ผู้ช่วยศาสตราจารย์ ดร.รัฐชัย ชาวุธทัย

(Assistant Professor Dr. Rathachai Chawuthai)

อาจารย์ประจำ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

### ประวัติการศึกษา

- ปริญญาเอก สาขา Informatics  
จากมหาวิทยาลัย SOKENDAI, ประเทศญี่ปุ่น (พ.ศ. 2559)
- ปริญญาโท สาขา Information Management  
จากสถาบันเทคโนโลยีแห่งเอเชีย, ประเทศไทย (พ.ศ. 2555)
- ปริญญาตรี สาขาวิศวกรรมคอมพิวเตอร์ (เกียรตินิยมอันดับ 1)  
จากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,  
ประเทศไทย (พ.ศ. 2549)

### ประวัติการทำงาน

- ผู้ช่วยนักวิจัย ณ National Institute of Informatics (3 ปี)
- วิศวกรซอฟต์แวร์อาวุโส ณ บริษัท ปันสาส์นเอเชีย จำกัด (1 ปี)
- วิศวกรซอฟต์แวร์ ณ บริษัท Thomson Reuters (4 ปี)

### รูปแบบงานวิจัย

- การวิเคราะห์ข้อมูล (Data Analytics)
- การวิเคราะห์ข้อมูลจากระบบจีพีเอส (GPS Data Analytics)

### วิชาที่สอน

- การวิเคราะห์ข้อมูลเบื้องต้น (Introduction to Data Analytics)
- การวิเคราะห์ข้อมูลและวิทยาศาสตร์ข้อมูล  
(Data Analytics and Data Science)
- เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence Technology)
- การเรียนรู้ของเครื่องจักร (Machine Learning)



“

ข้อมูลถือเป็นสินทรัพย์ที่สามารถนำไปต่อยอดสร้างมูลค่าได้อย่างมหาศาล แต่การมีข้อมูลแล้วปล่อยให้มันหลับใหลโดยไม่แตะต้องอะไรย่อมไม่เกิดประโยชน์ ดังนั้นจึงต้องปลุกข้อมูลของคุณมาทำงานผ่านเทคนิคการวิเคราะห์ประมวลผลจนสามารถพัฒนาไปเป็นผลิตภัณฑ์หรือได้ผลลัพธ์ที่ช่วยสนับสนุนการตัดสินใจเพื่อกำหนดทิศทางและวางกลยุทธ์ให้เป็นไปอย่างถูกต้องตามเป้าหมายที่วางไว้ จึงจะเกิดมูลค่าอย่างสูง

หนังสือเล่มนี้จะช่วยให้คุณรู้จักการ  
**ปลุกข้อมูลมาทำงาน**  
ผ่านการวิเคราะห์ข้อมูล

”

