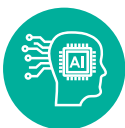




Practical

DATA SCIENCE¹

with RapidMiner Studio



RAPIDMINER

No-Code/Low-Code Data Science Platforms

อธิบายการสร้าง Basic Machine Learning Models
ตั้งแต่ต้นจนจบ เหมาะสำหรับผู้เชี่ยวชาญสาขา
ทางธุรกิจที่อยากเป็น Citizen Data Scientist



WORKSHOP DATASET :
SERAZU.COM

ผู้แต่ง ดร.เอกสิทธิ์ พิชรวงศ์ศักดา
บรรณาธิการ ภัทรพล ภูเขาเจริญ



มีเพียง “ความรู้” เท่านั้นที่มนุษย์ใช้พลิก “โลก”
และเปลี่ยน “ชีวิต” เราจึงสร้างสรรค์ และส่งมอบ “ความรู้”
ในรูปแบบที่ดีกว่า เพื่อให้คนไทย “เรียนรู้” ได้ตลอดชีวิต

Only “Knowledge” can help human
change “The World” and “Their Lives”.
With this truth, it drives us to deliver
“Knowledge” for Thai being able to
“Learn” better everyday.



Practical Data Science with RapidMiner Studio เล่ม 1

Writer	ดร.เอกสิทธิ์ พัชรวงศ์ศักดิ์
Editor	ภีรพล คชาเจริญ
Production Manager	วรพล ณธิกุล
Graphic Designer	ชวณันท์ รัตนะ
Page Layout	สุรัสวดี วงศ์จันทร์สุข
Proofreader	สุนทรี บรรลือศักดิ์
Publishing Coordinators	สุพัตรา อาจปรุ, สุรีย์รัตน์ จิ๋ว
Product Specialist	ศรัณย์ ขาติสุทธิผล

RapidMiner Studio เป็นเครื่องหมายการค้าของบริษัท RapidMiner, Inc. และเครื่องหมายการค้าอื่นๆ ที่อ้างถึงเป็นของบริษัทนั้นๆ

สงวนลิขสิทธิ์ตามพระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537 โดยบริษัท ไอดีซี พรีเมียร์ จำกัด ห้ามลอกเลียนไม่ว่าส่วนใดส่วนหนึ่งของหนังสือเล่มนี้ ไม่ว่าในรูปแบบใดๆ นอกจากจะได้รับอนุญาตเป็นลายลักษณ์อักษรจากผู้จัดพิมพ์เท่านั้น

สร้างสรรค์โดย



ข้อมูลทางบรรณานุกรม

ดร.เอกสิทธิ์ พัชรวงศ์ศักดิ์

Practical Data Science with RapidMiner Studio เล่ม 1

นนทบุรี : ไอดีซี, 2565

264 หน้า

1. โปรแกรมสำหรับระบบการประมวลผลข้อมูลเฉพาะชนิด
I ชื่อเรื่อง

005.37

Barcode 885-916-101-074-6

พิมพ์ครั้งที่ 1 ตุลาคม 2565

ราคา 385 บาท

จัดพิมพ์และจัดจำหน่ายโดย



บริษัท ไอดีซี พรีเมียร์ จำกัด

200 หมู่ 4 ชั้น 19 ห้อง 1901

อาคารจัสตินอินเตอร์เนชั่นแนลทาวเวอร์

ถ.แจ้งวัฒนะ อ.ปากเกร็ด จ.นนทบุรี 11120

โทรศัพท์ 0-2962-1081 (อัตโนมัติ 10 คู่สาย)

โทรสาร 0-2962-1084

สมาชิกสัมพันธ์

โทรศัพท์ 0-2962-1081-3 ต่อ 121

โทรสาร 0-2962-1084

ร้านค้าและตัวแทนจำหน่าย

โทรศัพท์ 0-2962-1081-3 ต่อ 112-114

โทรสาร 0-2962-1084

พิมพ์ที่ บริษัท พงษ์วินการพิมพ์ จำกัด

299-299/1 หมู่ 10 ถ.สุขุมวิท 107 ต.สำโรงเหนือ

อ.เมืองสมุทรปราการ จ.สมุทรปราการ 10270

โทรศัพท์ 0-2399-4525-31 โทรสาร 0-2399-4524



PREFACE

ผมเองอยู่ในแวดวง Data มาเกือบ 20 ปีแล้วครับ เห็นหลายอย่างผ่านไปอย่างรวดเร็ว ตั้งแต่ช่วงที่คนทำงานด้าน Data ไม่มีตำแหน่งงานที่ชัดเจน จนมาถึงยุคปัจจุบันที่หลายองค์กรตระหนักถึงความสำคัญของ Data และอยากหาประโยชน์จากสิ่งเหล่านี้ จึงเกิดตำแหน่งงานใหม่ขึ้นมา เช่น Data Scientist ซึ่งเป็นอาชีพที่ใครหลายคนใฝ่ฝัน แต่การจะเป็น Data Scientist จริงๆ แล้วนั้น จะต้องมีความรู้ความสามารถหลายอย่าง ไม่ว่าจะเป็นความรู้ในเรื่อง Data Science และ Machine Learning ความรู้ด้านการเขียนโปรแกรม และความรู้ทางธุรกิจที่เกี่ยวข้องกับงานที่รับผิดชอบ

ดังนั้น ตำแหน่งนี้จึงดูเหมือนจะขาดแคลนอยู่เป็นจำนวนมาก จนมีแนวทางใหม่เกิดขึ้นที่เรียกว่า Citizen Data Scientist ซึ่งจะเน้นไปที่การนำบุคลากรที่มีความรู้หรือประสบการณ์ทำงานในด้านนั้นๆ เช่น นักบัญชี นักการตลาด มาเรียนรู้เทคนิคการวิเคราะห์ข้อมูลเพิ่มเติม และใช้ซอฟต์แวร์ที่เรียกว่า No Code/Low Code Data Science Platform แทนการเขียนโปรแกรม ซึ่ง RapidMiner ก็เป็นซอฟต์แวร์ในกลุ่ม No Code/Low Code ที่ได้รับการยอมรับและใช้งานกันอย่างแพร่หลายตัวหนึ่งเลยครับ

ผมเองเริ่มต้นศึกษา RapidMiner มาตั้งแต่เวอร์ชัน 6 และจัดอบรมการใช้งานและให้คำปรึกษาเกี่ยวกับ RapidMiner เรื่อยมาจนพบว่า เวอร์ชัน 9 มีฟีเจอร์ใหม่ที่ทำให้เราใช้งานได้มากกว่าแต่ก่อน นั่นคือ Turbo Prep ที่ช่วยให้เราเตรียมข้อมูลได้ง่ายขึ้น และ Auto Model ที่ทำให้เราสร้างและเปรียบเทียบโมเดลทาง Machine Learning ได้เพียงแค่ไม่กี่ขั้นตอนเท่านั้น ด้วยเหตุผลนี้ ผมจึงอยากแชร์ประสบการณ์ใช้งาน RapidMiner Turbo Prep และ Auto Model ด้วยการยกตัวอย่าง Use Case ต่างๆ ไว้ในหนังสือเล่มนี้ครับ

สุดท้ายนี้ ถ้าใครเคยใช้ RapidMiner มาก่อนแล้ว ผมอยากให้ลองอ่านหนังสือเล่มนี้ แล้วคุณจะพบกับความง่ายในการใช้งานมากกว่าเดิม และใครที่เริ่มสนใจทางด้าน Data Science ผมก็ขอแนะนำ RapidMiner ให้เป็นอีกหนึ่งทางเลือกนอกจากการเขียนโปรแกรมต่างๆ ครับ เพราะในตลาดมีบริษัทที่ต้องการความรวดเร็วในการวิเคราะห์ข้อมูลมากกว่า ค่าใช้จ่ายอยู่อีกมากครับ

ดร.เอกสิทธิ์ พัทธวงศ์ศักดิ์

Data Science & RapidMiner Expert

Co-Founder and Data Science Team Lead

Cube Analytics Consulting Co., Ltd.





This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Machine Learning Master Certification

The Master level is the second level after Professional and stands for high proficiency in the topic. Some of the topics tested at this level are:

Complex Predictive Models, Cross Validation and Correct Validation, Parameter Optimization, Model Selection, Feature Engineering, Time Series and Forecasting and Integrating R and Python models.

This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck




Knut Makowski
(Head of Education)


Peter Lee
(CEO RapidMiner)



This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Data Engineering Master Certification

The Master level is the second level after Professional and stands for high proficiency in the topic. Some of the topics tested at this level are:

Loops and Branches, Advanced Text Processing, Exception Handling, Logging, Data Cleansing, Regular Expressions, Macros, Web APIs, and Scripting.

This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck




Knut Makowski
(Head of Education)


Peter Lee
(CEO RapidMiner)



This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Applications & Use Cases Master Certification

The Master level is the second level after Professional and stands for high proficiency in the topic. Some of the topics tested at this level are:

Running Processes, Deploying Models, Model Management and Web Apps.



This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck


Knut Makowski
(Head of Education)


Peter Lee
(CEO RapidMiner)



This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Platform Administration Master Certification

The Master level stands for high proficiency in the topic. Some of the topics tested at this level are:

Platform Overview, RapidMiner Studio Installation, RapidMiner Server Administration, RapidMiner Radoop, RapidMiner Real-Time Scoring, and RapidMiner Marketplace.



This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck


Knut Makowski
(Head of Education)


Peter Lee
(CEO RapidMiner)



EDITOR'S NOTE

ถ้าย้อนกลับไปเมื่อหลายปีก่อนที่ ‘Data Science’ จะกลายเป็น Buzzword ที่น่าตื่นเต้นปลุกกระแส ‘ปัญญาประดิษฐ์’ หรือ ‘AI’ ให้กลับมาโลดแล่นอีกครั้ง หลังจากล้มลุกคลุกคลานมาแล้วหลายครั้ง เนื่องจากองค์ความรู้และเทคโนโลยียังไม่ถึง AI จึงเป็นแนวคิดที่มาก่อนกาล แต่เมื่อเกิด ‘Machine Learning’ และ ‘Deep Learning’ ซึ่งเป็นอัลกอริทึมเรียนรู้เอง สิ่งต่างๆ ที่ไม่สามารถทำได้ในอดีต แต่มันเป็นไปได้แล้วในปัจจุบัน เนื่องจากองค์ความรู้และองค์ประกอบทางเทคโนโลยีในตอนนี้เติบโตอย่างเต็มที่ จนทำให้ AI กลับสู่เส้นทางแห่งความสำเร็จอีกครั้ง

ศาสตร์ Data Science กลายเป็น MegaTrend ที่สร้างโลกเปลี่ยนอนาคต และอาชีพ Data Scientist ก็ถูกยกให้เป็น “The Sexiest Job of the 21st Century” ที่ Harvard Business ให้การยกย่อง กลายเป็น Content สร้างกระแสจากอดีตจนถึงปัจจุบัน

ในยุคแรก ศาสตร์ Data Science ตามหลักการพื้นฐานโดยผู้เชี่ยวชาญในตอนนั้นสรุปแนวคิดไว้ว่า มันประกอบด้วยองค์ความรู้แบบสหสาขา ทั้งคณิตศาสตร์และสถิติ (ศาสตร์), การเขียนโปรแกรม (ทักษะการเขียนโค้ด) และความรู้ทางธุรกิจ/ฟังก์ชันงาน (รู้ดีถึงปัญหาและความท้าทายที่ต้องเผชิญ) ถึงแม้ว่างาน Data Science นั้นจะประกอบด้วยทีม Data ที่มีสมาชิกเชี่ยวชาญกันคนละด้านก็ตาม แต่ Data Scientist และ Data Engineer ก็ยังต้องเป็น Expert ทางเทคนิคอล (Technical) ซึ่งกลายเป็นข้อจำกัด

และด้วย Data Science Process มีหลายขั้นตอน มีรายละเอียดการทำงานที่ซับซ้อน จึงทำให้ต้นทุนการพัฒนาโปรเจกต์สูง ใช้เวลาปลูกบั่นนาน การปรับใช้ให้เกิดประโยชน์อย่างคุ้มค่าในเชิงธุรกิจยาก และที่สำคัญการฟอร์มทีมไม่ง่ายเลยในระยะแรก สิ่งเหล่านี้เป็นเหมือนยาขมที่ขวางทางการเข้าถึง Data Science ของคนในระดับองค์กร และในระดับธุรกิจอุตสาหกรรม

ในที่สุดก็เกิดเทรนด์ใหม่ที่เรียกว่า “Data Science and Machine Learning (DSML) Platforms” ที่เข้ามาเติมเต็มช่องว่าง เปลี่ยนโลก Data Science ดั้งเดิมที่เข้าถึงยาก และต้องใช้ผู้เชี่ยวชาญ มาเป็นแพลตฟอร์มที่เข้าถึงง่ายและมีความยืดหยุ่นมากขึ้น โดย DSML Platforms รองรับทั้ง Expert Data Scientists (Technical) และ Citizen Data Scientists (Nontechnical) และยังเติมขีดความสามารถให้กับ Data Engineers, Developers และ ML Engineers อีกทั้งคนสาย Tech และสาย Non-Tech สามารถทำงานร่วมกันได้บนแพลตฟอร์มเดียวกัน แค่มารู้แอปพลิเคชันซอฟต์แวร์บน DSML Platforms ซึ่งในหนังสือเล่มนี้คือ RapidMiner ที่ Gartner ระบุไว้ในรายงานว่า “RapidMiner is a trustworthy choice.”

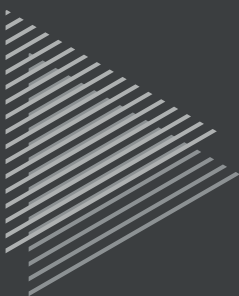
สุดท้ายนี้ ผมต้องขอขอบพระคุณ ดร.เอกสิทธิ์ พัทธวงศ์ศักดิ์ หนึ่งในผู้เชี่ยวชาญสาขา Data Science & RapidMiner Expert ที่ได้สร้างสรรค์หนังสือเชิงปฏิบัติการเล่มนี้ที่สำเร็จได้ยาก เพราะมีอุปสรรคในการเตรียม Datasets ที่เหมาะกับการสร้าง Machine Learning Models ที่สอดคล้องกับ Model Type การอธิบายกระบวนการตั้งแต่ต้นจนจบแบบ Step by Step รวมไปถึงการปรับใช้ในธุรกิจ ตัวอย่างในเล่มเกิดจากประสบการณ์ทั้งในฐานะอาจารย์และที่ปรึกษา หวังว่าผู้อ่านจะได้รับความรู้ตามที่คาดหวัง และอย่าลืมติดตามหนังสือ Practical Data Science with RapidMiner Studio เล่ม 2 (Advanced Model) และ Practical Data Engineer with RapidMiner Studio ในโอกาสต่อไป

กิริพล คชาเจริญ

บรรณาธิการ



CONTENT



Chapter 1

DATA SCIENCE PLATFORMS

ทางเลือกใหม่สำหรับ DATA SCIENTIST

ยุคของการวิเคราะห์ และใช้ประโยชน์จากข้อมูล	14
ข้อมูลจากการซื้อสินค้าออนไลน์ (Online Shopping Data)	14
ข้อมูลจากการใช้งานโทรศัพท์มือถือ (Mobile Data)	14
ข้อมูลจากอุปกรณ์ Internet of Things (IoT Sensors Data)	15
อาชีพทางด้าน Data Science	17
Data Scientist และ: Citizen Data Scientist	19
Data Scientist สาย Depth (ดั้งเดิม)	19
Citizen Data Scientist สาย Simplicity (ทางเลือกใหม่)	21
RapidMiner เครื่องมือสำหรับ Citizen Data Scientist	23
RapidMiner บทบาทของทีมสนับสนุน	24
RapidMiner แพลตฟอร์ม Automated Data Science	25
RapidMiner กับความสามารถทางด้าน Multipersona DSML Platforms	28

Chapter 2

เริ่มต้นการใช้งาน

RAPIDMINER STUDIO 9.10

เริ่มต้นใช้งาน RapidMiner Studio	32
ดาวน์โหลด RapidMiner Studio	32
แนะนำแพลตฟอร์มของ RapidMiner ก่อนเริ่มใช้งาน	36
RapidMiner Studio	37
RapidMiner Go	38

RapidMiner AI Hub	39
RapidMiner Notebooks/JupyterHub	40
เริ่มต้นใช้งาน RapidMiner Studio ครั้งแรก.....	41
การสร้าง Repository ใหม่.....	47

Chapter 3

เรียนรู้กระบวนการ

DATA SCIENCE ในเชิงธุรกิจ

ขั้นตอนการทำ Data Science Project	52
การทำ Data Science ด้วยกระบวนการเหมืองข้อมูล (CRISP-DM).....	53
ตัวอย่างการประยุกต์ใช้ในธุรกิจ (Business Use Case).....	56
ตัวอย่างการประยุกต์ใช้เพื่อจัดกลุ่มลูกค้า (Customer Segmentation)	57

Chapter 4

การเตรียมข้อมูลให้มีคุณภาพด้วย TURBO PREP

ทำไมข้อมูลที่มีคุณภาพคือสิ่งที่สำคัญที่สุด	66
เพราะเหตุใดข้อมูลจึงมีคุณภาพต่ำ	67
การเตรียมข้อมูลด้วย RapidMiner Turbo Prep.....	68
ขั้นตอนการ Connect Data	69
ขั้นตอนการสำรวจข้อมูล	75
การเรียงลำดับข้อมูล (Sort).....	75
การดูรายละเอียดของข้อมูล (Show Details)	78
ขั้นตอนการแก้ไขข้อมูลด้วยฟังก์ชันพื้นฐานที่จำเป็น.....	84
การเลือกข้อมูลที่ต้องการโดยใช้ฟังก์ชัน Transform : Filter.....	85
การแทนค่าข้อมูลในตารางโดยใช้ฟังก์ชัน Transform : Replace	91

การแทนที่ข้อมูลที่หายไปโดยใช้ฟังก์ชัน Cleanse : Replace Missing	100
การเชื่อมโยงตารางข้อมูลโดยใช้ฟังก์ชัน Merge : Join.....	105
การสร้างตารางสรุปข้อมูลโดยใช้ฟังก์ชัน Pivot	112
ขั้นตอนการดู History ใน RapidMiner Turbo Prep.....	114
ขั้นตอนการ Export ข้อมูลจาก Turbo Prep	115
ขั้นตอนการสร้างโฟลเดอร์จาก Turbo Prep.....	117

Chapter 5

การแบ่งกลุ่มลูกค้า CUSTOMER (RFM) SEGMENTATION ด้วย TURBO PREP

แนวคิดการแบ่งกลุ่มลูกค้าในยุค Marketing 5.0.....	122
แนวคิดการแบ่งกลุ่มลูกค้าแบบผสมผสาน.....	123
การแบ่งกลุ่มด้วยวิธี RFM Analysis คืออะไร.....	124
ขั้นตอนการทำ RFM Segmentation ด้วย Turbo Prep.....	127
STEP 1 : การนำเข้าข้อมูล (Data Connect).....	127
STEP 2 : การเตรียมข้อมูล (Data Preparation).....	132
STEP 3 : การแบ่งกลุ่มตามค่า RFM (RFM Segmentation).....	146
STEP 4 : การใช้ประโยชน์จาก RFM Segmentation (Benefits).....	154

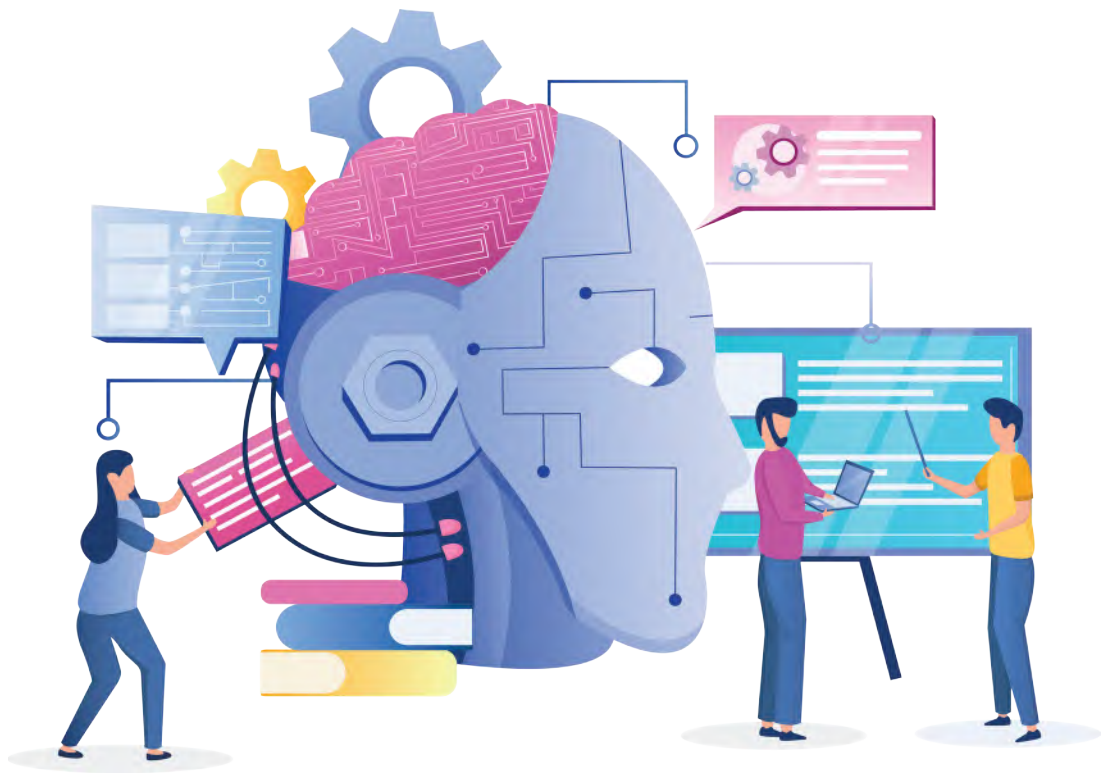
Chapter 6

การสร้าง MACHINE LEARNING ด้วย RAPIDMINER AUTO MODEL

แนวคิดของเทคนิคการเรียนรู้ของเครื่อง (Basic Concepts in Machine Learning)...	158
ประเภทของการเรียนรู้ของเครื่อง (Types of Learning in Machine Learning).....	160
การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning).....	161

การเรียนรู้แบบเสริมแรง (Reinforcement Learning).....	162
แนวคิดของการเรียนรู้แบบมีผู้สอน.....	163
การแบ่งข้อมูลเพื่อใช้ทดสอบประสิทธิภาพ.....	164
การประเมินประสิทธิภาพของโมเดล (Performance Metrics for Evaluation of Models).....	165
การประเมินประสิทธิภาพสำหรับ Classification Models.....	165
การประเมินประสิทธิภาพสำหรับ Regression Models.....	168
การสร้าง ML Model ด้วย RapidMiner Auto Model.....	170
การสร้าง ML Model ด้วยเทคนิค Decision Tree ...	171
สรุปแนวคิดของการสร้างโมเดลด้วย Decision Tree	171
ข้อดี – ข้อจำกัดของเทคนิค Decision Tree.....	172
RapidMiner Auto Model : การสร้างโมเดลพยากรณ์การลาออกด้วย Decision Tree	173
RapidMiner Auto Model : การแก้ปัญหาข้อมูลไม่สมดุล (Imbalanced Data)	195
การสร้าง ML Model ด้วยเทคนิค Naive Bayes	201
สรุปแนวคิดของการสร้างโมเดลด้วยเทคนิค Naive Bayes.....	201
ข้อดี – ข้อจำกัดของเทคนิค Naive Bayes.....	202
RapidMiner Auto Model : การสร้างโมเดลแนะนำรถยนต์	
การขนส่งสินค้าด้วยเทคนิค Naive Bayes	202
การสร้าง ML Model ด้วยเทคนิค Linear Regression.....	218
สรุปแนวคิดของการสร้างโมเดลด้วยเทคนิค Linear Regression.....	218
ข้อดี – ข้อจำกัดของเทคนิค Linear Regression.....	219
RapidMiner Auto Model : การสร้างโมเดลคาดการณ์การใช้	
น้ำมันเพื่อทำความร้อนด้วยเทคนิค Linear Regression	220
การสร้าง ML Model ด้วยเทคนิค Logistic Regression.....	235
สรุปแนวคิดของการสร้างโมเดลด้วยเทคนิค Logistic Regression.....	235
ข้อดี – ข้อจำกัดของเทคนิค Logistic Regression.....	236
RapidMiner Auto Model : การสร้างโมเดลเพื่อคาดการณ์การเป็นโรคเบาหวานด้วยเทคนิค Logistic Regression	237

เทคนิคการเรียนรู้แบบมีผู้สอนขั้นสูง (Advanced Supervised Learning Techniques).....	251
เทคนิค Support Vector Machines (SVM).....	252
ข้อดีของเทคนิค Support Vector Machines.....	254
ข้อจำกัดของเทคนิค Support Vector Machines.....	254
เทคนิค Random Forest.....	254
ข้อดีของเทคนิค Random Forest.....	256
ข้อจำกัดของเทคนิค Random Forest.....	257
เทคนิค Gradient Boosted Tree	257
ข้อดีของเทคนิค Gradient Boosted Tree (GBT).....	260
ข้อจำกัดของเทคนิค Gradient Boosted Tree (GBT).....	260
เทคนิค Deep Learning.....	260





DATA SCIENCE PLATFORMS

ทางเลือกใหม่สำหรับ

DATA SCIENTIST



องค์กรทุกขนาดในทุกอุตสาหกรรม ต่างมุ่งมั่นที่จะขับเคลื่อนด้วยข้อมูลมากขึ้น แต่ด้วยเกิดภาวะการขาดแคลน Data Scientists ที่จะมาตอบสนองได้ทันต่อความต้องการ ธุรกิจจึงสนใจแนวคิดใหม่ ที่จะมาเพิ่มขีดความสามารถให้กับคนที่มืออาชีพเป็นนักวิเคราะห์ข้อมูล ผู้เชี่ยวชาญด้านข้อมูล ผู้ที่มีประสบการณ์ในธุรกิจสาขาต่างๆ มาเรียนรู้แอปพลิเคชันซอฟต์แวร์บน Data Science and Machine Learning (DSML) Platforms โดยเรียกบุคลากรกลุ่มใหม่นี้ว่า “Citizen Data Scientists” และกลุ่ม Data Scientists ซึ่งเป็นผู้เชี่ยวชาญทางเทคนิคก็สามารถใช้ซอฟต์แวร์ DSML Platforms เพื่อ Develop & Deploy โมเดลไปปรับใช้ในองค์กรได้เช่นเดียวกัน



ยุคของการวิเคราะห์และใช้ประโยชน์จากข้อมูล

ในปัจจุบัน เราอยู่ในยุคที่เต็มไปด้วยข้อมูล (Data) ที่หลากหลาย และถูกสร้างขึ้นใหม่มากมาย ในแต่ละวัน อีกทั้ง มีข้อมูลบางประเภทที่ถูกสร้างขึ้นอย่างต่อเนื่องตลอดเวลา (Data Streaming) มีทั้งข้อมูลที่เกิดจากการดำเนินชีวิตประจำวันของพวกเราเอง และจากอุปกรณ์สมาร์ตดีไวส์ (Smart Devices) ต่างๆ หากจะกล่าวถึงแหล่งที่มาของข้อมูลที่สำคัญหลักๆ แล้วมีดังนี้

ข้อมูลจากการซื้อสินค้าออนไลน์ (Online Shopping Data)

ข้อมูลการช้อปปิ้งออนไลน์จะถูกรวบรวมเมื่อมีผู้คนออนไลน์ และเริ่มมีประสบการณ์กับการสั่งซื้อสินค้า ร้านค้าออนไลน์ก็จะเก็บรวบรวมข้อมูลเกี่ยวกับข้อมูลส่วนตัว ประวัติการซื้อ การใช้ยี่ห้อ ผลิตภัณฑ์ราคาถูก โดยเราสามารถจัดเก็บเป็นชุดข้อมูลการช้อปปิ้งออนไลน์ (Online Shopping Datasets) ของลูกค้าแต่ละรายได้

นอกจากนี้ ยังมีข้อมูลที่สำคัญที่มักถูกใช้ในการวิเคราะห์ เพื่อประเมินเว็บไซต์ในด้านต่างๆ เช่น อันดับเพจที่ถูกเข้าชมมากที่สุด อันดับสินค้าที่ถูกซื้อมากที่สุด ลิงก์ที่ลูกค้าชอบคลิก และข้อมูลการช้อปปิ้งออนไลน์ที่น่าสนใจอื่นๆ รวมถึงข้อมูลทางธุรกรรมที่ได้จากผู้ซื้อจริงๆ เช่น องค์กรประกอบการ ใช้งานตะกร้าสินค้า (Shopping Cart), วิธีการชำระเงิน (Payment) และความสัมพันธ์ระหว่างแบรนด์กับลูกค้า (Brand Affinity) เป็นต้น

ข้อมูลจากการใช้งานโทรศัพท์มือถือ (Mobile Data)

ข้อมูลที่อยู่ในรูปแบบของเนื้อหาที่เผยแพร่บนอินเทอร์เน็ต ที่ส่งไปยังอุปกรณ์พกพาต่างๆ เช่น บทความ ข่าวสาร อีเมล วิดีโอสตรีมมิ่ง การประชุมออนไลน์ บริการบนระบบคลาวด์ การเล่นเกมโซเชียลบนแพลตฟอร์มต่างๆ เป็นต้น

ยิ่งเทคโนโลยีเครือข่ายมือถือมีความรวดเร็ว และมีเครือข่ายครอบคลุมพื้นที่มากเท่าไร ก็ยิ่งต้องการสื่อสารข้อมูลมากขึ้นเท่านั้น ในชีวิตประจำวันของคนทั่วโลกทุกวันนี้ จำเป็นต้องพึ่งพาข้อมูลบนมือถือเพื่อทำกิจกรรมทางออนไลน์ต่างๆ ด้วยกันทั้งนั้น ไม่ว่าจะเป็นการศึกษา การค้นหาความรู้ การทำธุรกรรม การเลือกซื้อสินค้า การทำงาน รับส่งอีเมล การพักผ่อน การสื่อสาร ดูหนังฟังเพลง ฟังพอดแคสต์ หรือการขอคำปรึกษาทางการแพทย์ ฯลฯ โดยมี Apps ต่างๆ เป็นตัวกลาง

ด้วยข้อมูลต่างๆ เหล่านี้ถูกสร้างขึ้นตลอด 24 ชั่วโมงแต่ละวันโดยไม่มีวันหยุด จึงทำให้ปริมาณข้อมูลโมบายล์ที่ใช้เพิ่มมากขึ้นอย่างมหาศาลในทุกๆ ปี ฉะนั้น Data Strategy เพื่อเปลี่ยนข้อมูลให้เป็นโอกาสทางธุรกิจ จึงกลายเป็นกลยุทธ์กระแสหลักของการดำเนินธุรกิจในยุคใหม่

ข้อมูลจากอุปกรณ์ Internet of Things (IoT Sensors Data)

เป็นอุปกรณ์ประเภท Smart Devices ที่มีการเชื่อมต่อ และแลกเปลี่ยนข้อมูลกับอุปกรณ์อื่นๆ ผ่านสัญญาณอินเทอร์เน็ต โดยใช้เซนเซอร์ (Sensor) เพื่อตรวจจับหรือวัดค่าต่างๆ แล้วนำข้อมูลที่ได้จาก IoT Sensors ไปวิเคราะห์ อย่างเช่น

- **เซนเซอร์วัดอุณหภูมิ (Temperature Sensors)** ใช้ตรวจจับการเปลี่ยนแปลงของอุณหภูมิ มีประโยชน์ในการคอนโทรลอุณหภูมิในโรงงาน เช่น

ในโรงงานแปรรูปอาหารทะเลแช่แข็ง การรักษาคุณภาพเพื่อคงความสดใหม่ของวัตถุดิบ ก่อนถึงโรงงานแปรรูปเป็นสิ่งที่สำคัญมาก จะต้องควบคุมอุณหภูมิภายในห้องเย็นให้ได้ตามมาตรฐาน

หรือการวัดอุณหภูมิในดิน ซึ่งเป็นปัจจัยสำคัญที่ส่งผลต่อการเจริญเติบโตและการพัฒนาการของพืช เซนเซอร์ช่วยให้เราเห็นความผันผวนของอุณหภูมิได้ง่าย และสามารถแก้ไขได้อย่างรวดเร็วก่อนที่จะเกิดปัญหา เป็นต้น



- **เซนเซอร์วัดค่าทางชีวภาพ (Biomedical Sensors)** ถูกใช้ในอุตสาหกรรมทางการแพทย์ ในการเก็บข้อมูลเพื่อวิเคราะห์ปัญหาด้านสุขภาพ หากตรวจพบข้อมูลที่บ่งบอกถึงความผิดปกติของร่างกาย ก็จะได้เข้ารับการรักษาเสียตั้งแต่เนิ่นๆ หรือใช้เพื่อติดตามการรักษาผู้ป่วย เรามักพบ Biomedical Sensors ได้ในอุปกรณ์สวมใส่ไว้กับร่างกาย (Wearable) เช่น Smart Watch ที่นอกจากจะใช้ดูเวลาแทนนาฬิกาแล้ว ยังมีฟังก์ชันอื่นๆ อีกมากมาย เช่น ใช้วัดความดันโลหิต วัดอัตราการเต้นของหัวใจ เป็นต้น โดยมีการทำงานผ่าน Apps

นอกจากนี้ ยังมีเซนเซอร์อีกมากมาย เช่น เซนเซอร์วัดค่าทางเคมี (Chemical Sensors), เซนเซอร์วัดความชื้น (Humidity Sensors), เซนเซอร์วัดความดัน (Pressure Sensors) และเซนเซอร์วัดแสง (Light Sensor) ฯลฯ

ด้วยการพัฒนาอย่างรวดเร็วของเทคโนโลยีทางสารสนเทศและคอมพิวเตอร์ ทำให้เราสามารถเก็บข้อมูลได้มากขึ้นด้วยค่าใช้จ่ายที่ถูกลง แต่ถ้าเราเก็บข้อมูลไว้มากมายแล้วไม่สามารถนำมาสร้างคุณค่าใดๆ ได้อย่างคุ้มค่า ก็คงไม่เกิดประโยชน์มากนัก ดังนั้น หลายๆ องค์กรจึงอยากนำข้อมูลเหล่านี้มาทำการวิเคราะห์อย่างจริงจัง แล้วหาวิธีใช้ประโยชน์จากข้อมูลเชิงลึกที่ได้รับ ซึ่งการวิเคราะห์ข้อมูลนี้อาจจะเป็นแบบง่ายๆ โดยใช้การหาค่าทางสถิติ อย่างเช่น

- ยอดขายรวมในแต่ละสาขา
- จำนวนผู้เข้าชมเว็บไซต์เฉลี่ยแต่ละเดือน ในช่วงไตรมาสที่หนึ่งของปี 2022
- การวิเคราะห์ในลักษณะของการคาดการณ์ล่วงหน้า เช่น ในอีก 3 เดือนข้างหน้า ยอดขายโทรศัพท์มือถือ iPhone จะเป็นจำนวนกี่เครื่อง



อาชีพทางด้าน Data Science

เนื่องจาก Data Science ไม่ใช่งานที่สามารถทำได้ด้วยคนเดียว จำเป็นต้องมีทีมสนับสนุน จึงทำให้เกิดอาชีพใหม่ขึ้นมาอย่างน้อย 3 อาชีพ ซึ่งมีบทบาทที่แตกต่างกันครับ ใครทำหน้าที่อะไร และต้องมีทักษะความเชี่ยวชาญอะไรเป็นพิเศษ ผมสรุปไว้ให้ดังนี้



รูปที่ 1 - 1 แสดงตำแหน่งงานที่เกี่ยวข้องกับ Big Data

http://insidebigdata.com/wpcontent/uploads/2016/04/springboard_infographic.jpg

- **Data Analyst** หรือ **นักวิเคราะห์ข้อมูล** จะมีหน้าที่ในการวิเคราะห์ข้อมูลเชิงธุรกิจ และนำผลลัพธ์ที่ได้มาจัดทำรายงานสรุป โดยแสดงผลในรูปแบบต่างๆ ที่เข้าใจง่ายสำหรับผู้ใช้งาน (Business User) เช่น รูปแบบของชาร์ต (Chart) หรือแดชบอร์ด (Dashboard) เป็นต้น นอกจากนี้ อาจจะช่วยแปลโจทย์ทางด้านธุรกิจให้เป็นหัวข้อทางการวิเคราะห์ข้อมูล เพื่อใช้ประสานงานกับนักวิทยาศาสตร์ข้อมูล สำหรับนำไปวิเคราะห์ข้อมูลในเชิงลึกต่อไปได้อีกด้วย



ทักษะที่สำคัญของนักวิเคราะห์ข้อมูล คือ การใช้ซอฟต์แวร์ในการทำ Self-Service Analytics ได้เช่น Excel, Tableau, Power BI, Google Data Studio หรือเขียนโปรแกรมด้วยภาษา SQL ได้ เพื่อแสดงผลการวิเคราะห์เบื้องต้นให้เข้าใจได้ง่าย

- **Data Scientist** หรือ **นักวิทยาศาสตร์ข้อมูล** จะมีหน้าที่วิเคราะห์ข้อมูลเชิงลึกด้วยเทคนิคต่างๆ เช่น วิทยาศาสตร์ข้อมูล (Data Science), เหมืองข้อมูล (Data Mining), การเรียนรู้ของเครื่อง (Machine Learning) โดยอาจจะมีการพัฒนาโมเดล (Model) ทางคณิตศาสตร์ขึ้นมาใหม่ เพื่อให้ตอบโจทย์ทางธุรกิจของผู้ใช้ได้มากที่สุด

ทักษะที่สำคัญของนักวิทยาศาสตร์ข้อมูล คือ ความเข้าใจทางคณิตศาสตร์และสถิติ สามารถเขียนโปรแกรม หรือใช้งานซอฟต์แวร์เพื่อทำการวิเคราะห์ข้อมูลเชิงลึกได้ โดยภาษาโปรแกรมมิ่งที่ได้รับความนิยม เช่น การเขียนโปรแกรมด้วยภาษา R, Python และ SQL หรือซอฟต์แวร์ RapidMiner

- **Data Engineer** หรือ **วิศวกรข้อมูล** ตำแหน่งนี้จะไม่ค่อยได้ติดต่อกับผู้ใช้โดยตรง จะเน้นในส่วนของการจัดการดูแลระบบ เพื่อให้รองรับการทำงานของนักวิทยาศาสตร์ข้อมูลให้มีประสิทธิภาพทั้งในแง่ของความเร็วในการประมวลผล และการรองรับข้อมูลที่มีปริมาณมากได้

ทักษะที่สำคัญของวิศวกรข้อมูล คือ การใช้เทคโนโลยีในการจัดการข้อมูลต่างๆ เช่น SQL ในการจัดการฐานข้อมูล หรือการติดตั้งและจัดการระบบ Hadoop ภายในองค์กร

Data Scientist และ Citizen Data Scientist

การคาดการณ์อนาคตคือ การวิเคราะห์ข้อมูลในระดับที่ซับซ้อนที่เป็นส่วนหนึ่งของแนวทาง **วิทยาศาสตร์ข้อมูล** หรือ **Data Science** เป็นศาสตร์ที่กำลังเป็นกระแสนิยมในปัจจุบัน ธุรกิจและอุตสาหกรรมด้านต่างๆ ทั่วโลกต่างตื่นตัวกับการขับเคลื่อนด้วยเทคโนโลยีทางด้าน Data Science อย่างเห็นได้ชัด มองเห็นศักยภาพจนเป็นที่ประจักษ์แล้วว่า มันสร้างให้เกิดคุณค่าต่อธุรกิจได้อย่างแท้จริงจากตัวอย่างเคสมากมาย และเป็นกลยุทธ์สำคัญต่อความอยู่รอดในอนาคต

Data Scientist สาย Depth (ดั้งเดิม)

ด้วยเหตุผลข้างต้น จึงมีความต้องการบุคลากรทางด้าน **วิทยาศาสตร์ข้อมูล** หรือ **Data Science** เป็นจำนวนมาก แต่การพัฒนาบุคลากรทางด้าน Data Science ก็ไม่ได้ง่ายครับ เพราะเป็นเรื่องใหม่ที่ต้องอาศัยองค์ความรู้และทักษะหลายสาขาประกอบกัน โดยส่วนผสมขององค์ความรู้ทาง Data Science ที่ถูกอธิบายอย่างเป็นรูปธรรมเคยถูกกล่าวไว้โดย Drew Conway ดังแสดงในรูปที่ 1 - 2 เมื่อปี ค.ศ. 2010

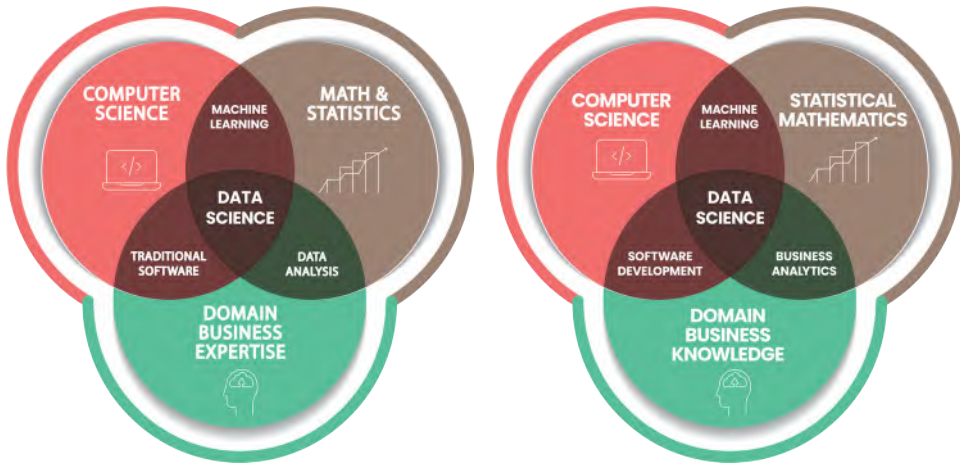


รูปที่ 1 - 2 แผนภาพ Data Science Venn Diagram โดย Drew Conway

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>



ต่อมาได้เกิดแผนภาพเพื่อใช้อธิบาย Data Science ด้วยมุมมองใหม่ๆ ที่ดูเป็นทางการกว่าต้นฉบับอีกหลายแบบ แต่ความหมายก็คล้ายคลึงกัน ผมขออธิบายสรุปในภาพรวมไว้ดังนี้



รูปที่ 1 – 3 แผนภาพ Data Science Venn Diagram แบบอื่นๆ

- **Computer Science** ความรู้ทางด้านวิทยาศาสตร์คอมพิวเตอร์ หรือไอที เช่น เข้าใจองค์ประกอบและระบบการทำงานของคอมพิวเตอร์ การสื่อสารข้อมูลบนเครือข่าย การเขียนโปรแกรมทาง Data Science โดยใช้ภาษาที่จำเป็น เช่น Python, R หรือ SQL เป็นต้น หรือการใช้ซอฟต์แวร์สำเร็จรูปประเภท No Code/Low Code
- **Math and Statistics** ความรู้ทางด้านคณิตศาสตร์และสถิตินั้น เป็นรากฐานที่สำคัญของการวิเคราะห์ข้อมูล เนื่องจากข้อมูลที่ถูกนำมาใช้ในกระบวนการวิเคราะห์จะเป็นข้อมูลทางสถิติไม่ว่าจะเป็นข้อมูลเชิงปริมาณ (Quantitative Data) และข้อมูลเชิงคุณภาพ (Qualitative Data) ไปจนถึงเทคนิคการพัฒนาโมเดลการเรียนรู้ของเครื่อง (Machine Learning)
- **Domain/Business Knowledge** การเข้าใจลักษณะธุรกิจอย่างลึกซึ้งซึ่งเป็นเรื่องที่สำคัญมากหากขาดความรู้ความเข้าใจในธุรกิจอย่างถ่องแท้ ก็ยากที่จะพัฒนาโปรเจกต์ทาง Data Science ให้ตอบโจทย์ หรือแก้ปัญหาให้กับธุรกิจได้ มีโอกาสล้มเหลวสูง ซึ่งข้อนี้ทำให้บริษัทหาคนที่เหมาะสมได้ยากมากทีเดียวครับ

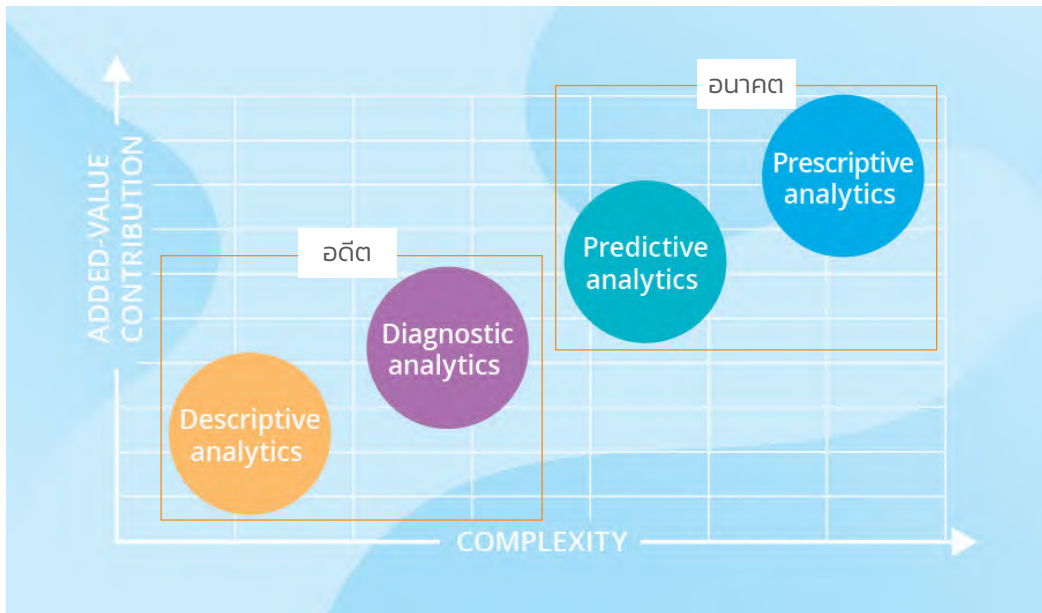
ด้วยอุปสรรคสำคัญในการสรรหาบุคลากร จึงเกิดแนวทางใหม่คือ การเลือกคนที่มีคุณสมบัติเหมาะสม โดยไม่จำกัดว่าจะเป็นบุคลากรภายในหรือภายนอกองค์กร หรือจะเป็นสายเทคนิคอล (Technical) หรือสายธุรกิจในฟังก์ชันงานต่างๆ (Nontechnical) ก็ตาม ให้มาเรียนรู้เรื่องการวิเคราะห์ข้อมูล และการใช้ซอฟต์แวร์แอปพลิเคชันทางด้าน Data Science and Machine Learning (DSML) Platforms เพิ่มเติม โดยเน้นผลลัพธ์ทางธุรกิจเป็นหลัก อาจไม่ต้องถึงขั้นมีสเกิลสูงทางด้านการเขียนโปรแกรมก็ได้ เพราะสามารถใช้ซอฟต์แวร์สำเร็จรูปแทนได้ เราเรียก Data Scientist ประเภทนี้ว่า **“Citizen Data Scientist”**

Citizen Data Scientist สาย Simplicity (ทางเลือกใหม่)

Citizen Data Scientist คำคำนี้ถูกกล่าวขึ้นโดย Gartner ที่ต้องการสื่อถึงกลุ่มคนที่เหมาะจะผันตัวตัวเองมาทำงานด้านนี้ โดยธุรกิจมีส่วนช่วยเพิ่มศักยภาพด้วยเครื่องมือ และการสนับสนุนที่จำเป็นในการเปลี่ยนสายงานให้มีสเกิลด้าน Data Science โดยกลุ่มคนที่กล่าวถึง เช่น นักวิเคราะห์ข้อมูลขั้นสูง (Data Analytics Professionals), ผู้เชี่ยวชาญด้านข้อมูล (Data Professional) หรือผู้เชี่ยวชาญด้านข้อมูลในสาขาต่างๆ (Business/Domain Experts) เช่น การขาย การตลาด การเงิน หรือทรัพยากรบุคคล ฯลฯ

บุคลากรเหล่านี้ต่างก็มีความรู้เกี่ยวกับปัญหา หรือความท้าทายทางธุรกิจที่ต้องเผชิญอยู่แล้ว พอได้รู้จักการใช้ซอฟต์แวร์สำเร็จรูปที่มีความฉลาด (DSML Platforms) ก็สามารถนำข้อมูลที่มีอยู่แล้วมาวิเคราะห์ แล้วสร้างโมเดลเชิงคาดการณ์ถึงสิ่งที่อาจเกิดขึ้นในอนาคต (Predict) หรือแนะนำสิ่งที่ควรทำ (Prescribe) ตามระดับความซับซ้อนของการวิเคราะห์

หากเป็นเมื่อก่อนเราต้องพึ่งพานักวิทยาศาสตร์ข้อมูล นักสถิติ หรือนักคณิตศาสตร์เท่านั้น ทุกวันนี้ แนวคิด Citizen Data Scientist จึงกลายเป็นกระแสหลักจนอาจกล่าวได้ว่า **“The Age of the Citizen Data Scientist”**



รูปที่ 1 - 4 การทำความเข้าใจการวิเคราะห์ข้อมูลในระดับที่ซับซ้อนและคุณค่าที่ได้รับ

เครดิตภาพ : www.scnsoft.com/blog/4-types-of-data-analytics

อยากให้เข้าใจว่า Citizen Data Scientist ไม่ได้จะเข้ามาแทน Data Scientist ดังเดิมแต่อย่างใด แต่คือส่วนที่เข้ามาเติมเต็มให้ทีม Data Science มีความยืดหยุ่นมากขึ้น เพราะการใช้แพลตฟอร์มหรือซอฟต์แวร์สำเร็จรูปนั้นทำได้ง่าย รวดเร็ว และประหยัด ที่สำคัญคือ หาคูคลองที่เหมาะสมได้ง่ายกว่า ช่วยให้ธุรกิจสามารถปรับเปลี่ยนเพื่อเปลี่ยนแปลงด้วย Data Science ได้จริงมากขึ้นครับ

ด้วยแนวคิดดังกล่าว ผมจึงตั้งใจเขียนหนังสือเล่มนี้ขึ้นมา เพื่อให้โอกาสกับคนที่อยากทำงานด้าน Data Science ได้เข้าใจแนวคิดการวิเคราะห์ข้อมูลต่างๆ ทาง Machine Learning โดยเรียนรู้จากการลงมือปฏิบัติจริงโดยใช้ซอฟต์แวร์ที่ชื่อว่า **RapidMiner** ซึ่งถูกกล่าวถึงในรายงาน 2 ฉบับของ Gartner ว่า “RapidMiner is a trustworthy choice.”

อ้างอิง : <https://rapidminer.com/downloads/2022-gartner-market-guides/>



RapidMiner เครื่องมือสำหรับ Citizen Data Scientist

RapidMiner คือ **Data Science Platform** ที่มีส่วนผลักดันให้เกิดการมีส่วนร่วมในงาน Data Science มากขึ้น มีวิธีการใช้งานที่เหมาะสมสำหรับทุกคนทุกระดับทักษะ แพลตฟอร์มนี้มีขั้นตอนการดำเนินงานตามวงจรที่ช่วยให้งานสามารถดำเนินไปได้อย่างรวดเร็ว เริ่มตั้งแต่การเตรียมข้อมูล (Data Prep) ไปจนถึงการใช้งานโมเดล (Model Ops)



รูปที่ 1 – 5 RapidMiner : Data Science Lifecycle



RapidMiner อนาคตของทีมสนับสนุน

บทบาทในภาพรวมของผู้ที่มีส่วนเกี่ยวข้องหลักๆ มีดังนี้



- **Data Science Experts :** แม้ว่าบุคลากรกลุ่มนี้จะมีความพร้อมในการดำเนินโปรเจกต์ซับซ้อนของงาน Data Science อยู่แล้ว แต่ RapidMiner เป็นเหมือนทางเลือกที่ช่วยให้งานบางตัวสำเร็จเร็วขึ้น เพื่อจะได้เอาเวลาที่มีค่าไปทุ่มให้กับงานที่มีมูลค่าสูงกว่า หรือแม้กระทั่งการให้ความช่วยเหลือกลุ่มคนที่เขียนโค้ดไม่ได้ที่มี ซึ่งเป็นประโยชน์ต่อธุรกิจอย่างแท้จริง



- **Information Technology (IT) :** เมื่อโมเดลถูกสร้างสำเร็จแล้ว ฝ่ายไอทีมีบทบาทสำคัญในการมอนิเตอร์ว่า โมเดลสามารถดำเนินการได้อย่างเหมาะสม และเชื่อมต่อกับระบบที่มีอยู่โดยไม่กระทบต่อความปลอดภัย เพื่อให้องค์กรได้รับมูลค่าทางธุรกิจเพิ่มขึ้นจากโมเดล RapidMiner มีมาตรฐานความปลอดภัยระดับสูง โดยแอดมินสามารถควบคุมการเข้าถึงข้อมูลและโครงการต่างๆ ได้อย่างเต็มที่ เพื่อให้เป็นไปตามนโยบายและข้อบังคับขององค์กร



- **Domain Experts :** บุคลากรเหล่านี้ต่างก็มีความรู้เกี่ยวกับปัญหาทางธุรกิจที่ต้องเผชิญอยู่แล้ว โดยจะต่อยอดด้วยการสร้างโมเดลเชิงคาดการณ์ (Predictive Models) ที่จะช่วยให้การตัดสินใจที่สำคัญมีประสิทธิภาพมากยิ่งขึ้น คนกลุ่มนี้มักเขียนโค้ดไม่ได้ จึงสามารถใช้ RapidMiner เป็นเครื่องมือสร้างการทำนายผลลัพธ์ในอนาคตได้

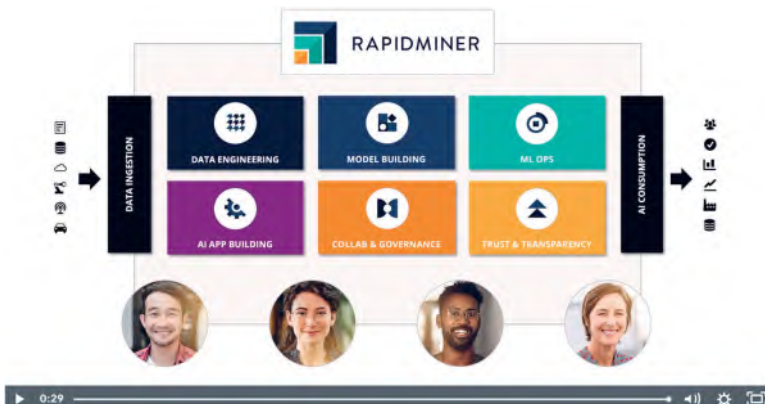


- **Leadership :** ผู้ที่มีหน้าที่กำหนดข้อมูลที่เป็น “Core Business” ของกลยุทธ์ธุรกิจ ให้การสนับสนุนเพื่อให้ผู้ใช้ทุกระดับที่มีทักษะต่างกัน สามารถเข้าถึง Data Science ได้ และสร้างวัฒนธรรมการตัดสินใจ โดยใช้ข้อมูลให้เกิดขึ้นด้วยการสร้าง “Data-First Culture” ไม่ปล่อยให้เกิดการเปลี่ยนแปลงโปรเซสที่สำคัญ โดยใช้लगสังหรณ์หรือคาดเดา ช่วยให้พนักงานมีข้อมูลเชิงลึกที่จำเป็นสำหรับการตัดสินใจ

กลุ่มที่กล่าวมาคือ ผู้ที่มีบทบาทสำคัญในการริเริ่ม Data Science ให้เกิดขึ้น และองค์กรยังต้องเพิ่มขีดความสามารถให้กับพนักงานในส่วนอื่นๆ ทั้งทั้งองค์กรในการปรับใช้ให้เกิดประโยชน์ในด้านต่างๆ

RapidMiner แพลตฟอร์ม Automated Data Science

RapidMiner เป็นเคอร์เนล (Kernel), โอเพ่นซอร์ส และเฟรมเวิร์ค (Framework) ที่มีความยืดหยุ่น โดยซัพพอร์ต Applications, Data Sources, Systems, Coding Languages, Machine Learning Libraries, เทคโนโลยี Cloud ที่หลากหลาย และเป็นเครื่องมือที่พยายามปรับให้เข้ากับความต้องการที่เปลี่ยนแปลงตลอดเวลาของ Data Science Community



รูปที่ 1 - 6 คลิปแสดงภาพรวมวิธีที่ RapidMiner Platform ที่เป็น Hub สำหรับ Data Science ทุกอย่าง

<https://rapidminer.com/platform/integrations/>