



Practical

DATA SCIENCE 1

เล่ม

with RapidMiner Studio



RAPIDMINER

No-Code/Low-Code Data Science Platforms

อธิบายการสร้าง Basic Machine Learning Models
ตั้งแต่ต้นจนจบ เหมาะสำหรับผู้เชี่ยวชาญสาขา
ทางธุรกิจที่อยากเป็น Citizen Data Scientist



ผู้แต่ง ดร.เอกสิทธิ์ พิธรวงศ์พิทยา
บรรณาธิการ กวีพล ศยาจรรย์



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 ประเภทของกินบนเมนู	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 Alerts	39
RapidMiner Notebooks/LupyterHub	43
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

การเตรียมข้อมูลให้มีคุณภาพ ด้วย TURBO PREP

ทำไมข้อมูลที่มีคุณภาพคือสิ่งที่สำคัญที่สุด	66
เพราะเหตุใดข้อมูลจึงมีคุณภาพต่ำ	67
การเตรียมข้อมูลด้วย RapidMiner Turbo Prep	68
ขั้นตอนการ Connect Data	69
ขั้นตอนการสำรวจข้อมูล	75
การเรียงลำดับข้อมูล (Sort)	75
การดูรายละเอียดของข้อมูล (Show Details)	76
ขั้นตอนการแก้ไขข้อมูลด้วยฟังก์ชันปรับพื้นฐานที่จำเป็น	84
การเลือกข้อมูลที่ต้องการใช้ฟังก์ชัน Transform: Filter	85
การลบค่าข้อมูลไม่ตรงด้วยฟังก์ชัน Transform: Replace	91

การลบค่าข้อมูลที่ไม่เป็นตัวเลขด้วยฟังก์ชัน Cleanse: Replace Missing	100
การเชื่อมโยงตารางข้อมูลโดยใช้ฟังก์ชัน Merge/Join	104
การสร้างตารางสรุปข้อมูลโดยใช้ฟังก์ชัน Pivot	112
ขั้นตอนการดู History ใน RapidMiner Turbo Prep	114
ขั้นตอนการ Export ข้อมูลจาก Turbo Prep	115
ขั้นตอนการสร้างไฟล์เอกสารจาก 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: การนำเข้ามาข้อมูล (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
การเรียนรู้แบบมีผู้สอน (Supervised Learning)	161

การเรียนรู้แบบเสริมแรง (Reinforcement Learning).....	151
แนวคิดของการเรียนรู้แบบมีผู้สอน.....	163
การแบ่งข้อมูลเพื่อใช้ทดสอบประสิทธิภาพ.....	164
การประเมินประสิทธิภาพของโมเดล (Performance Metrics for Evaluation of Models).....	165
การประเมินประสิทธิภาพสำหรับ Classification Models.....	166
การประเมินประสิทธิภาพสำหรับ Regression Models.....	168
การสร้าง ML Model ด้วย	
RapidMiner Auto Model.....	170
การสร้าง ML Model ด้วยเทคนิค Decision Tree.....	171
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Decision Tree.....	171
ข้อดี - ข้อจำกัดของเทคนิค Decision Tree.....	172
RapidMiner Auto Model: การสร้างโมเดล	
เทคนิคการสร้างโมเดลด้วย Decision Tree.....	173
RapidMiner Auto Model: การแก้ไขปัญหาข้อมูลที่ไม่สมดุล (Imbalanced Data).....	195
การสร้าง ML Model ด้วยเทคนิค Naive Bayes.....	201
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Naive Bayes.....	201
ข้อดี - ข้อจำกัดของเทคนิค Naive Bayes.....	202
RapidMiner Auto Model: การสร้างโมเดลแบบมีผู้สอน	
การสร้างโมเดลด้วยเทคนิค Naive Bayes.....	200
การสร้าง ML Model ด้วยเทคนิค	
Linear Regression.....	218
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Linear Regression.....	
Linear Regression.....	218
ข้อดี - ข้อจำกัดของเทคนิค Linear Regression.....	219
RapidMiner Auto Model: การสร้างโมเดลแบบมีผู้สอน	
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Linear Regression.....	220
การสร้าง ML Model ด้วยเทคนิค	
Logistic Regression.....	235
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Logistic Regression.....	
Logistic Regression.....	235
ข้อดี - ข้อจำกัดของเทคนิค Logistic Regression.....	236
RapidMiner Auto Model: การสร้างโมเดลแบบมีผู้สอน	
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Logistic Regression.....	237

เทคนิคการเรียนรู้แบบมีผู้สอนขั้นสูง (Advanced Supervised Learning Techniques).....	251
เทคนิค Support Vector Machines (SVM).....	252
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Support Vector Machines.....	254
ข้อดี - ข้อจำกัดของเทคนิค Support Vector Machines.....	254
เทคนิค Random Forest.....	254
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Random Forest.....	256
ข้อดี - ข้อจำกัดของเทคนิค Random Forest.....	257
เทคนิค Gradient Boosted Tree.....	257
ขั้นตอนการสร้างโมเดลด้วยเทคนิค Gradient Boosted Tree (GBT).....	259
ข้อดี - ข้อจำกัดของเทคนิค Gradient Boosted Tree (GBT).....	259
เทคนิค Deep Learning.....	260