

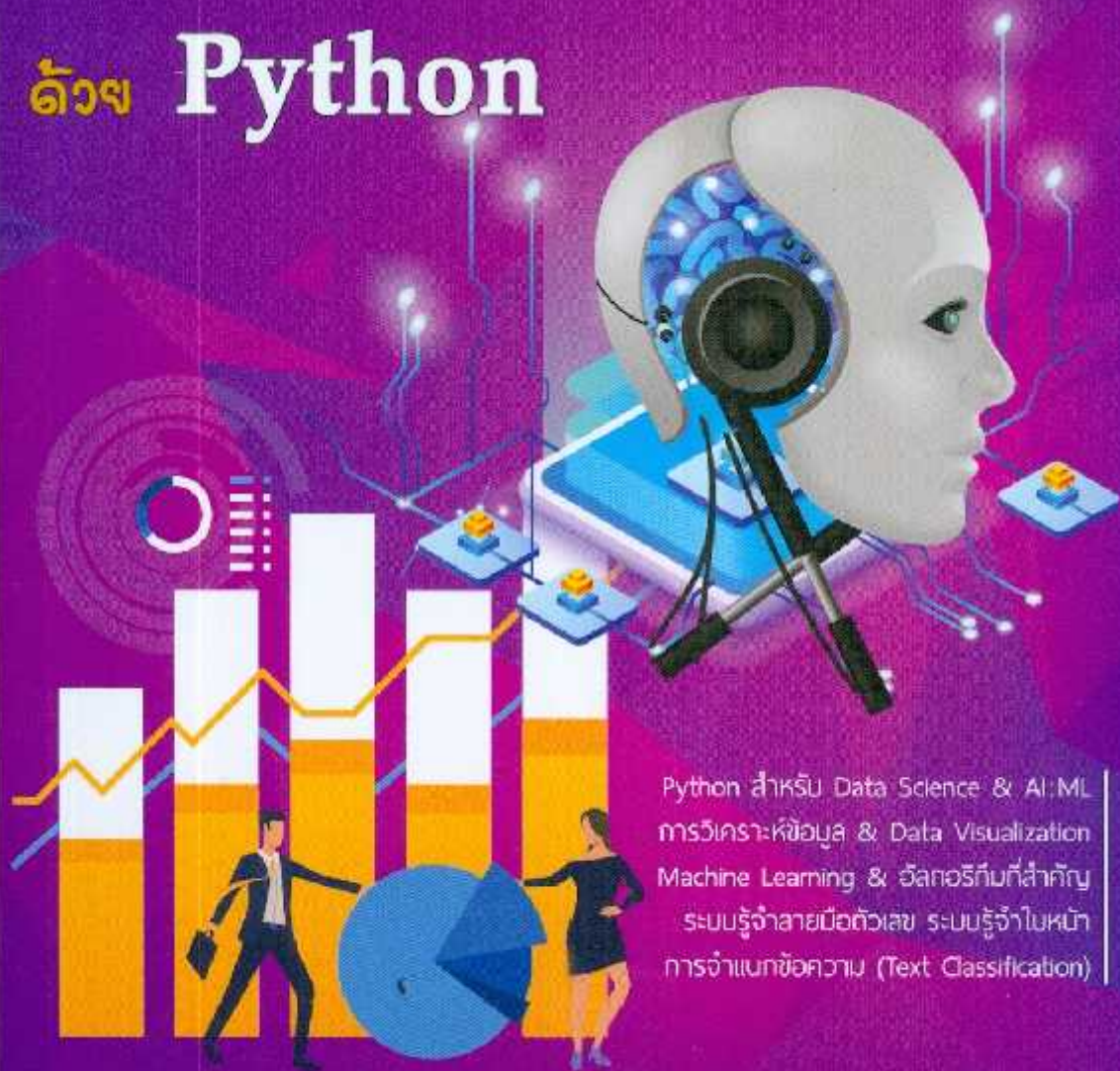
เรียนรู้จาก Workshop

เรียนรู้

Data Science และ:

AI: Machine Learning

ด้วย **Python**



Python สำหรับ Data Science & AI: ML
การวิเคราะห์ข้อมูล & Data Visualization
Machine Learning & อัลกอริทึมที่สำคัญ
ระบบรู้จำลายมือตัวเลข ระบบรู้จำใบหน้า
การจำแนกข้อความ (Text Classification)

01	บทนำ	17
	Data Science คืออะไร	18
	ศาสตร์และความรู้ที่จำเป็น	19
	เครื่องมือที่ใช้ในงาน Data Science	20
	งานของ Data Science ทำอะไรบ้าง	20
	และนำการใช้หนังสือ	24
02	การติดตั้งและเตรียมความพร้อม	25
	Anaconda	26
	Miniconda	31
	รู้จักกับ IPython และ Jupyter	34
	Workshop 1: การใช้งาน Jupyter Lab เบื้องต้น	36
	Workshop 2: โปรแกรมตรวจสอบแพ็คเกจ	43
	Colab	44
03	Python สำหรับ Data Science	45
	Python เบื้องต้น	46
	การเขียนโปรแกรมและการรัน	46
	ตัวแปรและการ Output เบื้องต้น	46
	การคำนวณพื้นฐาน	47
	ชนิดข้อมูล	47
	เงื่อนไขและการเปรียบเทียบ	47
	การเว้นคั่นบรรทัด (Indent)	49
	การรวมรอบ	50
	ฟังก์ชัน	53
	โปรแกรมย่อย Lambda	54
	ข้อมูลแบบชุด	55
	Module	60
	Numpy	61
04	จัดการข้อมูลด้วย Pandas	67
	รู้จักกับ Pandas	68
	Workshop 1: สร้างข้อมูล Series จาก List และ Tuple	70
	Workshop 2: สร้างข้อมูล Series จาก Numpy	71
	Workshop 3: สร้างข้อมูล Series แบบกำหนด Index	71
	Workshop 4: สร้างข้อมูล Series จาก Dictionary	72

Workshop 5: การเข้าถึงข้อมูล	72
Workshop 6: การเข้าถึงช่วงข้อมูล (Slicing)	73
DataFrame	73
Workshop 7: สร้าง DataFrame จาก Tuple หรือ List	74
Workshop 8: สร้าง DataFrame แบบกำหนดชื่อคอลัมน์	75
Workshop 9: สร้าง DataFrame จาก Dictionary	76
Workshop 10: สร้าง DataFrame จาก Series	76
การอ่านไฟล์ข้อมูลและการบันทึก	77
Workshop 11: บันทึก Data Frame เป็นไฟล์ CSV	77
Workshop 12: บันทึกไฟล์ Excel	79
Workshop 13: อ่านไฟล์ Excel ทั้งหมด ทุกคอลัมน์	79
กำหนดตัวเลือกการอ่านไฟล์ Excel	81
Workshop 14: อ่าน Excel บางคอลัมน์และข้าม header	81
Workshop 15: ตั้งชื่อคอลัมน์ใหม่	83
Workshop 16: อ่านไฟล์ Excel ไม่มีหัวคอลัมน์	84
Workshop 17: อ่านไฟล์ CSV	85
Workshop 18: อ่านไฟล์ CSV บางคอลัมน์ บางแถว	86
Workshop 19: อ่านไฟล์ CSV ที่ไม่มี Header	87
การตรวจสอบข้อมูลเบื้องต้น	87
Workshop 20: อ่านข้อมูลหรือตรวจสอบข้อมูลเบื้องต้น	88
ชนิดข้อมูล	89
Workshop 21: ตรวจสอบชนิดข้อมูลในคอลัมน์	89
Workshop 22: ดูสถิติเบื้องต้น	90
Workshop 23: เปลี่ยนชนิดข้อมูลของคอลัมน์	91
Workshop 24: เลือกคอลัมน์และเลือกข้อมูลแบบช่วง	93
Workshop 25: เลือกโดยระบุแถวที่ต้องการ	94
Workshop 26: ค้นหาข้อมูล	95
Workshop 27: เลือกแถวแบบวนรอบ	96
Workshop 28: เลือกแบบมีเงื่อนไข	97
Workshop 29: จัดเรียง	98
Workshop 30: อัปเดตเมทริกซ์ข้อมูล	99
Workshop 31: อัปเดตเมทริกซ์ด้วยค่ากำหนดที่	100
Workshop 32: แก้ไขโดยใช้เงื่อนไขค้นหาข้อความ	101
Workshop 33: เพิ่มและลบคอลัมน์	101

Contents

Workshop 34: เปลี่ยนชื่อคอลัมน์	102
Workshop 35: เพิ่มแถว	102
Workshop 36: ลบแถว	103
Workshop 37: การพ้อยกราฟ	104
Workshop 38: การรวมข้อมูลคอลัมน์และแถว	104
Workshop 39: จัดเข้ากลุ่มด้วย Cut	106

05 สถิติเบื้องต้น

107

สถิติเบื้องต้น	108
ประชากรและกลุ่มตัวอย่าง	108
ตัวแปร พารามิเตอร์ และสถิติ	109
ประเภทของสถิติ	110
ลักษณะข้อมูล	110
มาตรวัดข้อมูล	111
การหาค่าเข้าสู่ศูนย์กลาง หรือค่ากลาง	113
Workshop 1: หาค่าเฉลี่ย สูงสุด ต่ำสุด	113
Workshop 2: หาค่าเฉลี่ย สูงสุด ต่ำสุด จากไฟล์ Excel	114
Workshop 3: มัธยฐาน	117
Workshop 4: มัธยฐาน จากไฟล์ข้อมูล	117
Workshop 5: ฐานนิยม	118
Workshop 6: ฐานนิยม จากไฟล์ข้อมูล	118
การวัดการกระจายข้อมูล	118
Workshop 7: ทวิภาคี	119
การแจกแจงความถี่ (Frequency Distribution)	121
Workshop 8: แจกแจงความถี่ด้วย Pannas	123
Workshop 9: ย่านโพส์	124
Histogram แจกแจงความถี่	124

06 สรุปข้อมูลด้วย Pivot Table

125

ไฟล์ข้อมูล Excel ประเภท Workshop	126
สรุปข้อมูลด้วย Pivot Table	126
Workshop 1: สร้าง Pivot Table ใน Excel	127
Workshop 2: คึงข้อมูลมาทำสรุป	127
Workshop 3: ยอดขายของแต่ละไตรมาส	129
Workshop 4: ตาราง 2 มิติ แถว และ คอลัมน์	131

Workshop 5: ใช้งาน Pandas.....	132
สร้าง Pivot Table ด้วย Pandas.....	134
Workshop 6: Pivot Table หาค่าเฉลี่ยยอดขาย.....	135
Workshop 7: ค่าเฉลี่ยยอดขายแต่ละคนในแต่ละไตรมาส.....	135
Workshop 8: ยอดขายรวมแต่ละไตรมาส.....	136
Workshop 9: ยอดขายรวมเฉลี่ย สูงสุดต่ำสุด แต่ละไตรมาส.....	137
Workshop 10: การเรียงข้อมูล.....	138
Workshop 11: การพล็อตกราฟ.....	139
Workshop 12: ยอดขายรวมของแต่ละคน.....	140
Workshop 13: ยอดขายรวมแต่ละประเทศ.....	140
Workshop 14: ยอดขายแต่ละประเทศ แต่ละไตรมาส 2 มิติ.....	141
Workshop 15: ยอดขายรวมแต่ละคนในแต่ละประเทศ.....	143
Workshop 16: ยอดขายรวมแต่ละคนในแต่ละไตรมาส.....	145
Workshop 17: การกรอง Pivot Table.....	146
Workshop 18: เปรียบเทียบผลรวมของพนักงาน.....	147

07 Data Visualization I..... 149

Workshop 1: พล็อตกราฟฟังก์ชันด้วย matplotlib เบื้องต้น.....	151
Workshop 2: การรันโปรแกรมบน Jupyter.....	152
Workshop 3: บันทึกกราฟเป็นไฟล์.....	153
Workshop 4: กราฟเส้น.....	154
Workshop 5: กราฟข้อมูลหลายตัวแปร.....	154
Workshop 6: กราฟวงกลม.....	155
Workshop 7: Scatter.....	156
Workshop 8: Histogram.....	157
Workshop 9: Boxplot.....	159
การกำหนดรายละเอียดกราฟ.....	159
Workshop 10: กราฟแบบกำหนดรายละเอียด.....	161
Workshop 11: พล็อตกราฟด้วย Pandas.....	162
Seaborn.....	167
Workshop 12: Data Visualization ด้วย Seaborn.....	167
ตกแต่งกราฟ.....	172
Workshop 13: การเปรียบเทียบข้อมูล.....	184
กราฟ Violinplot.....	188

Contents

08 Data Visualization 2..... 193

ไฟล์ข้อมูล.....	194
Workshop 1: โหลดข้อมูลและตรวจสอบเบื้องต้น.....	194
Workshop 2: พล็อตกราฟฟิวเจอร์กระจาย.....	196
Workshop 3: พล็อต Scatter ดูความสัมพันธ์.....	196
Workshop 4: พล็อต Scatter ในกราฟเดี่ยว.....	198
Workshop 5: การใช้ Subplot กับ Scatter.....	198
Workshop 6: วิเคราะห์ผลการให้แสง.....	200
Workshop 7: Jointplot.....	203
Workshop 8: Swarmplot.....	204
Workshop 9: สหสัมพันธ์.....	205
Workshop 10: Heatmap.....	209
Workshop 11: Pairplot.....	210
Workshop 12: Lmplot.....	212

09 การจัดการข้อมูลสูญหาย..... 215

เกี่ยวกับข้อมูลสูญหาย (Missing data).....	216
ประเภทของข้อมูลสูญหาย.....	216
วิธีการจัดการกับข้อมูลสูญหาย.....	219
ไฟล์ข้อมูลสำหรับ Workshop.....	221
Workshop 1: โหลดข้อมูลและตรวจสอบเบื้องต้น.....	221
Workshop 2: เปลี่ยนชนิดข้อมูลให้เหมาะสม.....	224
Workshop 3: หา Missing data ว่างๆ ทั้งหมด.....	224
Workshop 4: ตรวจสอบแต่ละคอลัมน์.....	228
Workshop 5: ตรวจสอบค่าข้อมูลแถวที่มี Missing data.....	229
Workshop 6: การแทนที่ Missing data ด้วยค่าเฉลี่ย.....	230
Workshop 7: การแทนที่ด้วยค่าโดยท่างหนึ่ง.....	231
Workshop 8: การเปลี่ยนคอลัมน์เก็บชนิด String เป็นตัวเลข.....	231
Workshop 9: การแทนที่ด้วยค่าก่อนหน้า.....	233
Workshop 10: การแทนที่ด้วยค่าถัดไป.....	234
Workshop 11: จัดการลบแถวทั้งหมด.....	235
Workshop 12: จัดการลบบางส่วน.....	235
Workshop 13: จัดการลบคอลัมน์.....	236
Workshop 14: ลบแบบมีเงื่อนไข.....	237

Workshop 15: ลบ ค่าที่แหว่งเป็น Missing data	238
Workshop 16: ลบ ค่าที่คล้ายกันเป็น Missing data	239
Workshop 17: กำกับค่า Missing data ตอนโหลดไฟล์	240



10 การทำความสะอาดข้อมูล

ไฟล์ข้อมูล	242
Workshop 1: โหลดข้อมูลและตรวจสอบเบื้องต้น	242
Workshop 2: ยกคอลัมน์ที่ไม่ใช้	244
Workshop 3: เปลี่ยนชื่อคอลัมน์	245
Workshop 4: ตรวจสอบคอลัมน์ประเภท Category	247
Workshop 5: จัดการกับข้อมูลผิดชนิด	252
Workshop 6: การคัดกรองและลบแถวที่	256
Workshop 7: คัดตัวเลขออกจาก Object/String	259
Workshop 8: จัดการค่าทศนิยม	261
Workshop 9: ลบและเปลี่ยนชื่อคอลัมน์	263
Workshop 10: จัดการข้อมูลตัวเลข	264
Workshop 11: จัดการกับ bedroom	269
Workshop 12: จัดการข้อมูลซ้ำ	270
Workshop 13: ตรวจสอบสูตรท้าย	272
Workshop 14: การอ่านไฟล์แบบกำหนดชนิดข้อมูล	274



11 ค่าผิดปกติและการกำจัด

ค่าผิดปกติ	276
ไฟล์ข้อมูล	276
Workshop 1: โหลดข้อมูลและตรวจสอบเบื้องต้น	276
Workshop 2: Data Visualization	278
การหาค่าผิดปกติด้วยวิธี IQR	279
Workshop 3: ค้นหา Outliers ด้วยวิธี IQR	279
Workshop 4: Visualization ด้วย Boxplot	280
Workshop 5: กำจัดค่าผิดปกติเหมือนขอบเขตบน	281
Workshop 6: กำจัดค่าผิดปกติได้ขอบเขตล่าง	282
Workshop 7: ตรวจสอบซ้ำ	283
การใช้ Z-score	284
Workshop 8: การคำนวณ Z-score	284

Contents

Workshop 9: การคำนวณ Z-score โดยใช้ไลบรารี	285
Workshop 10: กำจัดค่าผิดปกติด้วยวิธี Z-score	286
Workshop 11: กำจัดค่าผิดปกติตัวแปร Multivariate	290
Workshop 12: กำจัดค่าผิดปกติตัวแปร Multivariate	291

12 ข้อมูล Time series

ข้อมูลแบบ Time series	296
การใช้ Pandas กับ Time series	297
ไฟล์ข้อมูลตัวอย่าง	298
Workshop 1: สถานข้อมูลและตรวจสอบเบื้องต้น	299
Workshop 2: พล็อตกราฟข้อมูลเบื้องต้น	300
Workshop 3: เปลี่ยนเป็นข้อมูลวินาที	301
Workshop 4: การกำหนดเวลา	304
Workshop 5: พล็อตกราฟช่วงเวลา	306
การหาค่าเฉลี่ยและการรวม	308
Workshop 6: เฉลี่ยข้อมูลด้วย Rolling	308
Workshop 7: เจกี่ยและรวมด้วย Resample	310
Workshop 8: พล็อตกราฟ Resample และ Rolling	311
Workshop 9: กราฟแท่งกับข้อมูล Time series	313
Workshop 10: ข้อมูล Time series ที่มีค่าขาด	314
Workshop 11: ข้อมูล ป.ศ.	317

13 Machine Learning เบื้องต้น

AI และ Machine Learning	322
ประเภทของ Machine Learning	324
Workshop 1: เขียนโปรแกรมไม่ใช้ Machine Learning	325
Workshop 2: เขียนโปรแกรมโดยใช้ Machine Learning	328
Workshop 3: การบันทึกและเรียกใช้งาน Model	333
Workshop 4: ให้นำ Model มาใช้งานไม่โปรแกรม	334
ขั้นตอน Machine Learning	334
Machine Learning ไม่ได้รอบรู้ไปทุกอย่าง	336
คำศัพท์ที่ควรรู้	336

14 Regression

Simple Linear Regression	340
--------------------------------	-----

การประเมินความคลายแม่นยำ Model	341
Workshop 1: ประเมิน Model	342
Workshop 2: การประเมิน Model ด้วยไลบรารี.....	344
Workshop 3: เปรียบเทียบประสิทธิภาพ 2 Model.....	345
Workshop 4: พยากรณ์ผลกำไรโฆษณา TV.....	347
Multiple Linear Regression	351
Workshop 5: Multiple Linear Regression	351
Polynomial Regression	355
Workshop 6: ทำนายค่า GDP.....	356
Workshop 7: ทำนาย GDP ด้วย Polynomial Regression	359
Workshop 8: ทำนายจำนวนผู้ติดเชื้อ COVID 19	362
Overfitting และ Underfitting.....	365

15 Decision Tree..... 367

Workshop 1: Decision Tree อย่างง่าย	368
การคิดตั้งไลบรารีแสดง Decision Tree.....	375
Workshop 2: แสดง Tree ของ Model.....	378
Workshop 3: เพิ่ม Class.....	381
การเข้ารหัส	385
Workshop 4: คัดสินใจจะเล่นเทนนิสหรือไม่	387
Workshop 5: การประเมิน Model.....	393

16 การประเมิน Model..... 395

ตาราง Confusion Matrix	396
Workshop 1: Confusion Matrix แบบ Binary	403
Workshop 2: Confusion Matrix แบบกราฟิก	410
Workshop 3: Confusion Matrix 3 Labels.....	411
การแบ่งข้อมูลสำหรับชุดคลอบ	414
Workshop 4: แบ่งชุดข้อมูลสำหรับ Train Test.....	415
Workshop 5: วิเคราะห์การปล่อยสินค้า	417
Workshop 6: ใช้ข้อมูลจากไฟล์ CSV เป็น Input	424
Workshop 7: ความแปรปรวนผลการประเมิน	427
การแบ่งแบบ Cross Validation	427
Workshop 8: การแบ่ง k-fold	428

Contents

Workshop 9: 5-fold Cross Validation.....	429
--	-----

17 อัลกอริทึมที่สำคัญ..... 431

Naive Bayes Classifier.....	432
Workshop 1: การคำนวณด้วย Naive Bayes เบื้องต้น.....	432
Workshop 2: ระบบแนะนำการป้อนสินค้าชื่อ Naive Bayes.....	436
ข้อมูล Iris.....	438
Workshop 3: โหลดข้อมูล Iris และตรวจสอบเบื้องต้น.....	439
Workshop 4: เตรียมข้อมูล Features และ Label.....	441
Workshop 5: ประเมิน Cross Validation ของ Naive.....	442
Workshop 6: สร้าง Model และประเมินทดสอบ.....	443
Workshop 7: ทำนายสายพันธุ์.....	445
Workshop 8: แอนโตนดูข้อมูลใหม่อยู่ตรงไหน.....	446
Support Vector Machine.....	448
Workshop 9: จำแนกเตี้ยไม่ได้ด้วย SVM.....	449
k-Nearest Neighbors.....	450
Workshop 10: จำแนกเตี้ยไม่ได้ด้วย kNN.....	451
Random Forest.....	452
Workshop 11: Random-Forest.....	452
k-Means Clustering.....	454
Workshop 12: จัดกลุ่มเตี้ยไม่ได้ด้วย k-Means.....	454
Workshop 13: Logistic Regression ทำนายการซื้อสินค้า.....	459

18 ROC AUC และ Threshold..... 465

ROC curve และ AUC.....	466
Workshop 1: โหลดข้อมูล.....	467
Workshop 2: Case 1 ROC AUC ของ Model ที่ที่สุด.....	469
Workshop 3: Case 2 ROC AUC Model แม่นยำ 88%.....	474
Workshop 4: Case 3 ROC AUC Model แม่นยำ 70%.....	475
Workshop 5: Case 4 ROC AUC Model ที่แย่.....	477
ค่า Threshold.....	479
Workshop 6: กำหนดค่า Threshold สำหรับ Model.....	482

19 ระบุจำนวนสายมือเขียน..... 487

ข้อมูลภาพ.....	488
----------------	-----



Workshop 1: โหลดข้อมูลตัวเลขหลายมิติ	488
Workshop 2: แสดงภาพตัวเลข	491
Workshop 3: เตรียมข้อมูลเพื่อ Train และ Test	493
Workshop 4: ทำการ Train และ Test	495
Workshop 5: ดูตัวเลขที่ทำนายผิด	498
Workshop 6: ประเมิน Cross Validation	500
Workshop 7: เปรียบเทียบกับอัลกอริทึมอื่น	500
Workshop 8: เขียนตัวเลขด้วยลายมือ แล้วทำนาย	501
Workshop 9: การสร้าง Dataset ข้อมูลตัวเลข	504

20 การตรวจจับและรู้จำใบหน้า

Computer Vision ตรวจจับและรู้จำใบหน้าคน	508
Workshop 1: โหลดข้อมูลใบหน้าตัวอย่าง	512
Workshop 2: ยူทูปใบหน้าในข้อมูล	514
Workshop 3: เตรียมข้อมูลเพื่อทดลอง Train และ Test	516
Workshop 4: ทำการ Train และ Test	518
OpenCV	521
Workshop 5: OpenCV แสดงภาพ	522
เกี่ยวกับข้อมูลภาพ	523
Workshop 6: OpenCV อ่านค่าสีแต่ละชั้น	524
Workshop 7: เล่น VDO	525
Workshop 8: OpenCV แสดงภาพสดจากกล้อง	526
Workshop 9: ตรวจจับใบหน้าในภาพนิ่ง	528
Workshop 10: ตรวจจับใบหน้าจากกล้องสดๆ หรือ VDO	531
Workshop 11: ระบบรู้จำและแยกแยะใบหน้า	534

21 การลดมิติข้อมูล

การลดมิติข้อมูล	542
การเลือก Feature	543
Workshop 1: คัดกรองด้วยค่า Correlation	544
Workshop 2: คัดกรองตัวแปร Low Variance	548
Workshop 3: เลือก Feature ด้วย Univariate Selection	549
Workshop 4: การเลือกด้วยวิธีเวียนทำซ้ำ	552
การวิเคราะห์องค์ประกอบหลัก PCA	553
Workshop 5: ทำ Standard Scale ข้อมูล Iris	555

Contents

Workshop 6: ลดมิติด้วย PCA.....	558
Workshop 7: นำ PCA ไปใช้ Train และทำนาย.....	563
Workshop 8: ลดมิติ Digit Classification.....	568
Workshop 9: ทดลองกับ Face Recognition.....	577
Workshop 10: การย้อนกลับ PCA.....	579

22 การปรับจูน Model..... 583

ปรับจูน Model เพื่ออะไร.....	584
Workshop 1: เปรียบเทียบผลของ Hyperparameter.....	584
Workshop 2: เปรียบเทียบอัลกอริทึม.....	586
Workshop 3: หาค่าพารามิเตอร์ด้วย Grid Search.....	588
Workshop 4: การนำค่าพารามิเตอร์ไปใช้งาน.....	590
Workshop 5: หาค่าพารามิเตอร์ด้วย Grid Randomized.....	591
Workshop 6: ลอง Grid Search กับ Decision Tree.....	593
Workshop 7: Randomized Search กับ Decision Tree.....	594

23 การแก้ปัญหาความไม่สมดุล..... 597

Text Classification.....	598
การหาค่า Feature ของข้อความ.....	599
Workshop 1: จำนวนหาค่า Features.....	601
Workshop 2: พัดจำนวนข้อความ.....	604
Workshop 3: โหลดและตรวจสอบข้อมูลเบื้องต้น.....	604
Workshop 4: ตรวจสอบ Bag of words.....	608
Workshop 5: ตรวจสอบข้อความในข้อมูล.....	610
Workshop 6: ขั้นตอน Train และ Test.....	610
Workshop 7: ทำนาย.....	611
Workshop 8: การใช้ Pipeline.....	612
Workshop 9: การประเมิน Model.....	614
Workshop 10: เตรียมชุดข้อมูลของเราเอง.....	615

24 Recommender System..... 617

Recommender System.....	618
Workshop 1: วิเคราะห์ความสัมพันธ์ของประวัติ.....	618
Workshop 2: วิเคราะห์ความคล้ายของ 4 ประโยค.....	622
Workshop 3: ระบบแนะนำภาพยนตร์.....	625