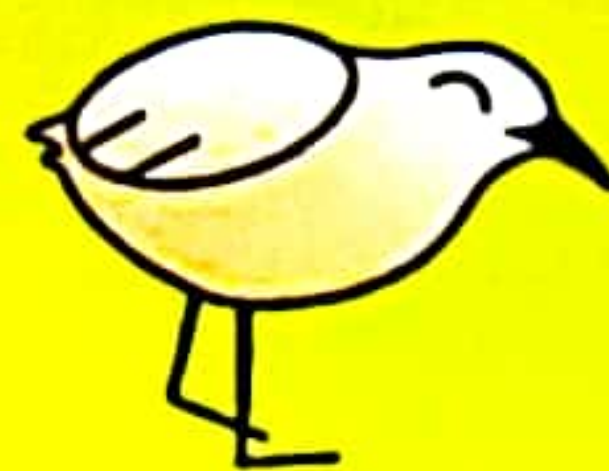


การใช้โปรแกรม WEKA
เพื่อการทำเหมืองข้อมูล

PRACTICAL DATA MINING

WITH



WEKA

การจำแนกข้อมูลด้วย *Decision Tree*, *Naive Bayes*, *kNN*
การแบ่งกลุ่มข้อมูลด้วย *SimpleKMeans*
การหาความสัมพันธ์ด้วย *Apriori*

ทิ พ ว ร ร ณ ศิ ล วั ธ น า นุ ส า น ตี

ได้รับทุนสนับสนุนจากงบประมาณเงินรายได้
ประจำปีงบประมาณ ๒๕๖๓ คณะมนุษยศาสตร์และสังคมศาสตร์
มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี

ได้รับทุนสนับสนุนจาก
มหาวิทยาลัยสงขลานครินทร์
PN.BT-๐๐๑/๖๕

สารบัญ

	หน้า
บทที่ 1 การติดตั้งโปรแกรม WEKA	1
1.1 การทำเหมืองข้อมูลคืออะไร	3
1.2 โปรแกรม WEKA คืออะไร	4
1.3 ข้อดีของโปรแกรม WEKA	6
1.4 ข้อจำกัดของโปรแกรม WEKA	7
1.5 การติดตั้งโปรแกรม WEKA	7
1.6 ความต้องการของระบบ	8
1.7 การดาวน์โหลดโปรแกรม WEKA	8
1.8 คำแนะนำในการดาวน์โหลดโปรแกรม WEKA	8
1.9 คำแนะนำในการติดตั้งโปรแกรม WEKA	12
1.10 สรุป	18
บทที่ 2 แนะนำการใช้โปรแกรม WEKA	19
2.1 การใช้งานโปรแกรม WEKA	21
2.2 ระบบการจัดการแฟ้มเอกสาร WEKA	24
2.3 การทำงาน WEKA Explorer	25
2.4 องค์ประกอบของโปรแกรม WEKA	26
2.5 คำอธิบายแถบเมนู WEKA Explorer	27
2.6 อัลกอริทึมการทำเหมืองข้อมูลยอดนิยม	27
2.7 เริ่มต้นใช้งานโปรแกรม WEKA	28
2.8 การเตรียมข้อมูลเพื่อใช้งานใน WEKA	29
2.9 แหล่งความรู้เพิ่มเติม WEKA	38
2.9 สรุป	39
2.10 ปฏิบัติการท้ายบท	41

	หน้า
บทที่ 3 การเตรียมข้อมูล (Data Preprocessing)	45
3.1 ความสำคัญของการเตรียมข้อมูล	47
3.2 การโหลดข้อมูลเพื่อการเตรียมข้อมูล	49
3.3 การเลือกหรือการกรองแอตทริบิวต์	52
3.4 การแบ่งช่วงข้อมูล (Discretization)	58
3.5 จัดการค่าข้อมูลสูญหาย (Missing Values)	74
3.6 การเปลี่ยนรูปข้อมูล (Data Normalization)	79
3.7 การเปลี่ยนรูปข้อมูลนามกำหนดเป็นรูปข้อมูลตัวเลข (NominalToBinary)	79
3.8 การเปลี่ยนรูปข้อมูลตัวเลขเป็นรูปข้อมูลนามกำหนด (NumericToNominal)	84
3.9 สรุป	88
3.10 ปฏิบัติการท้ายบท	89
บทที่ 4 การจำแนกข้อมูลด้วย Decision Tree	93
4.1 ต้นไม้การตัดสินใจคืออะไร	95
4.2 ต้นไม้การตัดสินใจทำงานอย่างไร	96
4.3 วิธีเลือกแอตทริบิวต์สำหรับการสร้างต้นไม้การตัดสินใจ	97
4.4 โอเวอร์ฟิตติงในต้นไม้การตัดสินใจ	99
4.5 การตัดแต่งกิ่งไม้และวิธีการตัดแต่งกิ่งต้นไม้	99
4.6 ตัวอย่างการสร้างแผนผังการตัดสินใจ	100
4.7 การจำแนกข้อมูลด้วยต้นไม้การตัดสินใจ	101
4.8 การคำนวณค่าเอนโทรปี (Entropy)	103
4.9 การคำนวณค่า Information Gain (IG)	104
4.10 การใช้อัลกอริทึม J48 ใน WEKA	108
4.11 ต้นไม้การตัดสินใจกับข้อมูลชุดทดสอบ (Test set)	121
4.12 ต้นไม้การตัดสินใจกับข้อมูลชุดเรียนรู้ (Training set)	122
4.13 การบันทึกและโหลดแบบจำลอง	141

	หน้า
4.14 รายการผลลัพธ์และการบันทึกแบบจำลอง	144
4.15 การตีความผลลัพธ์	148
4.16 สรุป	151
4.17 ปฏิบัติการท้ายบท	153
บทที่ 5 การจำแนกข้อมูลด้วย Naïve Bayes	165
5.1 Naïve Bayes Classifier ทำงานอย่างไร	167
5.2 การใช้ Naïve Bayes ใน WEKA	175
5.3 การสร้างโมเดลด้วย Naïve Bayes สำหรับข้อมูลชุดเรียนรู้	177
5.4 การสร้างโมเดลด้วย Naïve Bayes สำหรับข้อมูลชุดทดสอบ	182
5.5 สรุป	191
5.6 ปฏิบัติการท้ายบท	193
บทที่ 6 การจำแนกข้อมูลด้วย k-Nearest Neighbors	209
6.1 การทำงานของ k-Nearest Neighbors	211
6.2 ควรใช้ kNN เมื่อใด	215
6.3 ข้อดีข้อจำกัดของการใช้ kNN	215
6.4 สูตรคณิตศาสตร์ที่ใช้ใน kNN	216
6.5 การคำนวณ Euclidean Distance	216
6.6 การคำนวณ Manhattan Distance	216
6.7 การคำนวณ Minkowski Distance	217
6.8 เลือกค่า k อย่างไรให้เหมาะสม	218
6.9 การใช้ kNN ใน WEKA	219
6.10 การจำแนกข้อมูลด้วย kNN ใน WEKA Explorer	220
6.11 การจำแนกข้อมูลด้วย kNN ใน WEKA Experimenter	224
6.12 สรุป	233
6.13 ปฏิบัติการท้ายบท	235

	หน้า
บทที่ 7 การจัดกลุ่มข้อมูลด้วย Simple K-Means	237
7.1 การจัดกลุ่มข้อมูลทำงานอย่างไร	239
7.2 การใช้ K-Means Clustering ใน WEKA	244
7.3 สรุป	261
7.4 ปฏิบัติการท้ายบท	263
บทที่ 8 การหากฎความสัมพันธ์ข้อมูลด้วย Apriori	269
8.1 กฎความสัมพันธ์ข้อมูลคืออะไร	271
8.2 งานทำเหมืองกฎความสัมพันธ์	275
8.3 ขั้นตอนการสร้างกฎความสัมพันธ์	275
8.4 หลักการทำงานของอัลกอริทึม Apriori	276
8.5 ขั้นตอนการทำงานของอัลกอริทึม Apriori	277
8.6 การคำนวณค่า Support และ Confidence	278
8.7 การหากฎความสัมพันธ์ข้อมูลด้วย Apriori ใน WEKA	285
8.8 สรุป	292
8.9 ปฏิบัติการท้ายบท	293
บทสรุป	301
บรรณานุกรม	307
ภาคผนวกชุดข้อมูล	313
ดัชนีคำ	327
ประวัติผู้เขียน	333
คำขอบคุณ	339