



สำนักพิมพ์มหาวิทยาลัยนเรศวร  
Naresuan University Publishing House

เทคนิคเหมืองข้อมูล

# DATA MINING TECHNIQUES

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร



ห้องสมุดเทเวศร์



401017225

จรัสศรี รุ่งรัตนอุบล

## คำนำ

เหมืองข้อมูลเป็นหนึ่งในความรู้สมัยใหม่ที่ได้รับความนิยมอย่างมากในปัจจุบัน โดยได้ถูกประยุกต์ใช้ในแทบทุกองค์กร ไม่ว่าจะเป็นภาครัฐและเอกชนเพื่อใช้ในการตัดสินใจ การวางแผนกลยุทธ์ การปรับปรุง การให้บริการและการดำเนินงานต่าง ๆ ในองค์กร เหมืองข้อมูลเป็นการทำงานที่เน้นการค้นหาสารสนเทศหรือองค์ความรู้จากข้อมูลขนาดใหญ่ เพื่อนำสิ่งที่ได้มาให้เป็นประโยชน์ โดยเหมืองข้อมูลเป็นการผสมผสานศาสตร์ทางสถิติ ปัญญาประดิษฐ์ การรู้จำและฐานข้อมูลเข้าด้วยกัน

หนังสือ “เทคนิคเหมืองข้อมูล” เล่มนี้ เป็นเนื้อหาการทำเหมืองข้อมูลโดยมีวัตถุประสงค์เพื่อให้หลักการ เทคนิค และขั้นตอนวิธีที่สำคัญของเหมืองข้อมูล โดยเน้นการนำเสนอแนวคิดและขั้นตอนวิธีของเทคนิคเหมืองข้อมูลต่าง ๆ เช่น เทคนิคต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม การจัดกลุ่มด้วยเคมีน การวิเคราะห์ความสัมพันธ์ เป็นต้น โดยผู้เขียนได้ยกตัวอย่างงานวิจัยที่เกี่ยวกับการประยุกต์ใช้เหมืองข้อมูลที่ผู้เขียน ผู้ร่วมวิจัย และนักศึกษาได้จัดทำร่วมกันเพื่อเป็นแนวทางการประยุกต์ใช้ให้กับผู้อ่าน นอกจากนี้ผู้เขียนได้ใช้โปรแกรมเหมืองข้อมูล เวกา (Weka) ที่พัฒนาโดย University of Waikato ประเทศนิวซีแลนด์ เพื่อนำเสนอผลลัพธ์การทำงานของแต่ละเทคนิคเหมืองข้อมูล โดยโปรแกรมเวกามีรูปแบบการใช้งานง่าย เหมาะกับการใช้งานเพื่อศึกษาเทคนิคเหมืองข้อมูล

ผู้เขียนได้มีโอกาสศึกษา ทำวิจัยและสอนเกี่ยวกับเทคนิคเหมืองข้อมูลตั้งแต่ปี พ.ศ. 2553 และได้รวบรวมหลักการและเทคนิคที่สำคัญของเหมืองข้อมูลไว้ในหนังสือเล่มนี้ ปัจจุบันเทคนิคเหมืองข้อมูลเป็นเนื้อหาที่สอดแทรกไว้ในหลายวิชา เช่น เทคนิคการทำเหมืองข้อมูล (Data Mining Techniques) การทำเหมืองข้อมูลและข้อมูลขนาดใหญ่ (Data Mining and Big Data) การวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics) เป็นต้น ทำให้มีผู้สนใจศึกษาเป็นวงกว้าง หนังสือเล่มนี้เหมาะกับผู้ที่สนใจหลักการพื้นฐานของการทำเหมืองข้อมูลและการทำงานของเทคนิคเหมืองข้อมูลแบบต่าง ๆ โดยผู้เขียนหวังว่าผู้อ่านจะสามารถนำเทคนิคเหมืองข้อมูลไปประยุกต์ใช้งานได้อย่างถูกต้องและมีประสิทธิภาพ และสามารถตีความผลลัพธ์จากการทำเหมืองข้อมูลได้อย่างถูกต้อง โดยไม่มองว่าเหมืองข้อมูลเป็นเพียงโปรแกรมอัตโนมัติที่เมื่อใส่ข้อมูลแล้วจะได้ผลลัพธ์ออกมา

ผู้เขียนขอขอบคุณบิดา มารดา และครูอาจารย์ที่อบรมสั่งสอนให้ความรู้ แรงคิดที่ดีในการทำงานและใช้ชีวิต ตลอดจนกัลยาณมิตรทุกคนที่เป็นกำลังใจให้ผู้เขียนจัดทำหนังสือเล่มนี้จนเสร็จสมบูรณ์

สุดท้ายนี้ ผู้เขียนหวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้จะเป็นประโยชน์ต่อผู้อ่านที่สนใจ หากพบข้อบกพร่องและมีข้อคิดเห็นหรือข้อเสนอแนะประการใด ติดต่อผู้เขียนได้ที่ e-mail : jaratsrir@nu.ac.th ผู้เขียนขอน้อมรับคำแนะนำต่าง ๆ เพื่อนำมาเป็นแนวทางในการปรับปรุงเนื้อหาให้มีความชัดเจนและดียิ่งขึ้นในอนาคตต่อไป

## สารบัญ

บทที่ 1 แนะนำการทำเหมืองข้อมูล (Introduction to Data Mining).....	1
ลักษณะการจัดเก็บข้อมูลและปริมาณข้อมูลในปัจจุบัน.....	2
เหมืองข้อมูลคืออะไร.....	7
ทำไมต้องใช้เหมืองข้อมูล.....	10
กระบวนการทำงานของเหมืองข้อมูล.....	11
ตัวอย่างการประยุกต์ใช้เหมืองข้อมูล.....	20
เครื่องมือและโปรแกรมเหมืองข้อมูล.....	25
บทสรุป.....	26
แบบฝึกหัดท้ายบท.....	27
บทที่ 2 การเตรียมข้อมูล (Data Preprocessing).....	29
ข้อมูลคืออะไร.....	30
ประเภทและลักษณะข้อมูล.....	31
การเตรียมข้อมูล.....	37
มาตรวัดความคล้ายและความต่างของวัตถุ.....	45
บทสรุป.....	53
แบบฝึกหัดท้ายบท.....	54
บทที่ 3 เทคนิคการจำแนก (Classification).....	55
ขั้นตอนการพัฒนาตัวจำแนก.....	56
ขั้นตอนวิธีค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว (K-nearest Neighbor Algorithm).....	59
1. การแปลงข้อมูลเป็นรูปมาตรฐานกับขั้นตอนวิธีค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว.....	64
2. การใช้ส่วนกลับของระยะห่างกับขั้นตอนวิธีค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว.....	65
3. การใช้โปรแกรมเวก้ากับขั้นตอนวิธีค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว.....	69
เทคนิคต้นไม้ตัดสินใจ (Decision Tree).....	73
1. หลักการพัฒนาแบบจำลองการพยากรณ์เพื่อการจำแนกหรือตัวจำแนก.....	73
2. กระบวนการสร้างต้นไม้ตัดสินใจ.....	74
3. การประเมินประสิทธิภาพของโมเดลเพื่อพยากรณ์หรือตัวจำแนก.....	92

4. การแตกกิ่งของลักษณะประจำปัจจัยเข้าแบบต่าง ๆ .....	96
5. การใช้โปรแกรมเวกากับขั้นตอนวิธีต้นไม้ตัดสินใจ .....	101
5.1 การใช้โปรแกรมเวกาเพื่อพัฒนาตัวจำแนกแบบต้นไม้ตัดสินใจกับข้อมูลที่ค่า ลักษณะประจำเต็มหน่วย .....	101
5.2 การใช้โปรแกรมเวกาเพื่อพัฒนาตัวจำแนกแบบต้นไม้ตัดสินใจกับข้อมูลที่ค่าลักษณะ ประจำต่อเนื่อง .....	108
การสร้างกฎเพื่อการจำแนก (Rule Based Classifier) .....	110
1. การคำนวณค่าความครอบคลุมและความถูกต้องของกฎ.....	111
2. หลักการสร้างกฎเพื่อการจำแนก.....	115
3. การใช้โปรแกรมเวกาเพื่อสร้างกฎเพื่อการจำแนก .....	118
ตัวจำแนกเบย์อย่างง่าย (Naïve Bayes Classifier).....	123
1. แนวคิดของตัวจำแนกแบบเบย์.....	124
2. การใช้โปรแกรมเวกากับตัวจำแนกแบบเบย์ .....	132
โครงข่ายประสาทเทียม (Artificial Neural Network).....	135
1. สถาปัตยกรรมของโครงข่ายประสาทเทียม.....	135
2. การใช้โปรแกรมเวกาเพื่อพัฒนาตัวจำแนกด้วยโครงข่ายประสาทเทียม.....	137
บทสรุป.....	143
แบบฝึกหัดท้ายบท .....	144
<b>บทที่ 4 การวิเคราะห์การจัดกลุ่ม (Cluster Analysis).....</b>	<b>147</b>
การวิเคราะห์การจัดกลุ่ม.....	148
การจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering).....	150
การใช้โปรแกรมเวกาเพื่อการจัดกลุ่มแบบลำดับขั้น.....	167
การจัดกลุ่มข้อมูลแบบเคมีน (K-mean Clustering).....	170
การใช้โปรแกรมเวกาเพื่อการจัดกลุ่มแบบเคมีน.....	179
บทสรุป.....	182
แบบฝึกหัดท้ายบท .....	183
<b>บทที่ 5 การวิเคราะห์ความสัมพันธ์ (Association Analysis).....</b>	<b>185</b>
กฎความสัมพันธ์ (Association Rules).....	186


ขั้นตอนวิธีที่ใช้ในการสร้างกฎความสัมพันธ์ .....	187
ขั้นตอนวิธีอะไพริออรี (A priori algorithm).....	198
การเตรียมข้อมูลสำหรับการหากฎความสัมพันธ์.....	202
การใช้โปรแกรมเวก้าเพื่อค้นหากฎความสัมพันธ์.....	203
การตีความหมายของกฎความสัมพันธ์.....	210
สรุปการใช้งานกฎความสัมพันธ์ .....	211
บทสรุป.....	211
แบบฝึกหัดท้ายบท .....	212
<b>บทที่ 6 การพยากรณ์ (Prediction).....</b>	<b>215</b>
การวิเคราะห์การถดถอย (Regression Analysis).....	216
1. การวิเคราะห์การถดถอยอย่างง่าย .....	217
2. การวิเคราะห์การถดถอยเชิงพหุ .....	236
3. เกณฑ์การวัดค่าประสิทธิภาพของตัวแบบพยากรณ์ .....	249
4. การใช้โปรแกรมเวก้ากับการวิเคราะห์การถดถอย .....	250
โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN).....	254
1. โครงสร้างของโครงข่ายประสาทเทียมแบบง่าย .....	256
2. โครงสร้างและองค์ประกอบของโครงข่ายประสาทเทียม .....	258
3. การแปลงข้อมูลเข้าและการแปลงข้อมูลออกหรือผลลัพธ์.....	263
4. หลักการคำนวณค่าน้ำหนักแบบวิธีแพร่ย้อนกลับ (Back Propagation) .....	265
5. การใช้โปรแกรมเวก้ากับโครงข่ายประสาทเทียมแบบ Multilayer Perceptron.....	275
บทสรุป.....	281
แบบฝึกหัดท้ายบท .....	282
<b>บรรณานุกรม.....</b>	<b>284</b>
<b>ดัชนี.....</b>	<b>287</b>

สามารถยืมและติดตามหนังสือใหม่ได้ที่ ระบบห้องสมุดอัตโนมัติ Walai Autolib

<https://lib.rmutp.ac.th/catalog/BibItem.aspx?BibID=๐๐๐๑๘๕๐๖>



**เทคนิคเหมืองข้อมูล = Data mining techniques / จรัสศรี รุ่งรัตนอุบล.**

Author	จรัสศรี รุ่งรัตนอุบล
Published	พฤษภาคม : มหาวิทยาลัยขอนแก่น, 2566
Edition	พิมพ์ครั้งที่ 1
Detail	290 หน้า : ภาพประกอบ ; 26 ซม
Subject	เหมืองข้อมูล
ISBN	9786164262928
ประเภทแหล่งที่มา	 Book



Multi view

View map

สำหรับเพื่อการศึกษาและการอ้างอิง