

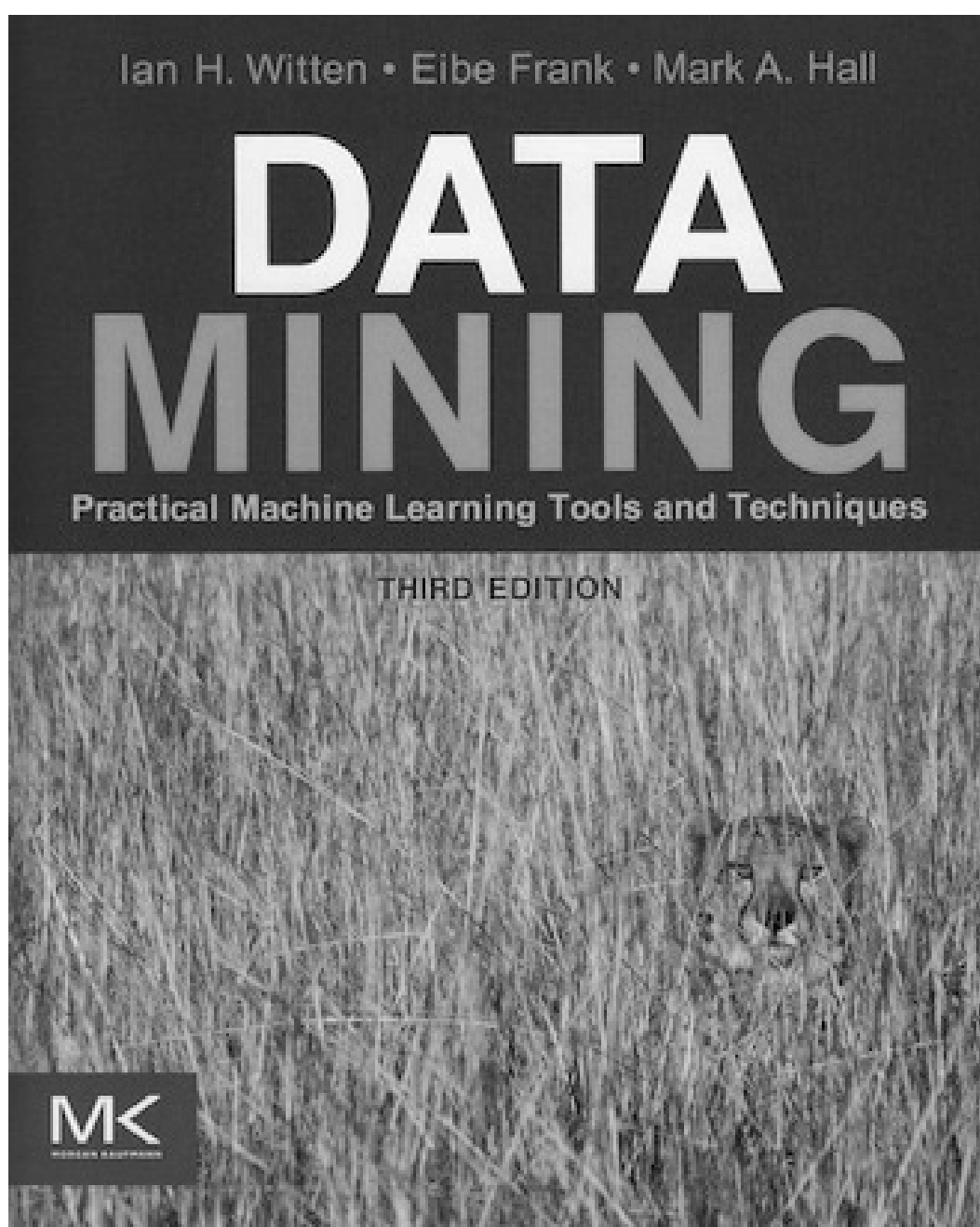


บทวิจารณ์หนังสือ

โดย ดร.อนิรุท สีบสิงห์

คณะบริหารศาสตร์ มหาวิทยาลัยอุบลราชธานี

Data Mining Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers,
(2011), 629 pages, by Ian H. Witten, Eibe Frank and Mark A. Hall



ในปัจจุบันเทคโนโลยีสารสนเทศได้เข้ามามีบทบาทต่อการใช้ชีวิตของมนุษยชาติเป็นอย่างมาก ความเจริญก้าวหน้าทางด้านเทคโนโลยีเป็นสิ่งที่ช่วยอำนวยความสะดวกในการดำรงชีวิตอยู่ของคนในปัจจุบัน ไม่ว่าจะหลับหรือตื่นชีวิตของพวกเราจะต้องติดอยู่กับเทคโนโลยี โดยเฉพาะเทคโนโลยีอินเทอร์เน็ตที่มีการพัฒนาไปอย่างรวดเร็วทำให้ชีวิตเป็นจุดเปลี่ยนที่สำคัญที่ทำให้การดำรงชีวิตอยู่ของมนุษย์เปลี่ยนไป เพราะเป็นเทคโนโลยีที่ทำให้เราสามารถเข้าถึงข้อมูลข่าวสารต่างๆ จากทั่วทุกมุมโลกได้ง่ายขึ้นข้อมูลที่อยู่ในโลกอินเทอร์เน็ตมีความหลากหลาย มีความทันสมัย ข้อมูลที่ได้เป็นข้อมูลที่มีทั้งภาพและเสียง และนอกจากนี้เรายังสามารถที่จะเข้าถึงเนื้อหาที่เป็นข้อคิดเห็นต่าง ๆ จากบุคคลทุกเพศทุกวัน จากทั่วทุกมุมโลก อีกทั้งเรายังสามารถนำเสนอข้อคิดเห็นของเราลงไปได้อีกด้วย ทำให้เกิดสังคมใหม่ที่มีการแบ่งปันข้อมูลข่าวสาร ทำให้เราได้รู้จักเพื่อนๆ กลุ่มใหม่ ๆ จากทั่วทุกมุมโลกจนเกิดคำนิยามใหม่เรียกสังคมเหล่านี้ว่า “สื่อสังคมออนไลน์ (Social Network)” และจากเกิดขึ้นของอินเทอร์เน็ต และสื่อสังคมออนไลน์นี้เองทำให้มีข้อมูลจำนวนมากมายมหาศาลที่วิ่งอยู่บนโลกอินเทอร์เน็ต ไม่ว่าจะเป็นข้อมูลเหล่านั้นจะอยู่ในรูปของไฟล์ข้อมูลข้อความธรรมดา ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย อย่างไรก็ตามมีนักวิจัยหลาย ๆ ท่าน มีการตั้งข้อสังเกตว่าเราจะสามารถสกัดข้อมูลต่าง ๆ ที่อยู่รอบตัวเราที่อยู่บนโลกอินเทอร์เน็ตมาใช้งานได้อย่างไร เพราะข้อมูลที่อยู่ในนั้นมีทั้งข้อมูลที่เป็นประโยชน์ และไม่เป็นประโยชน์ มีความน่าเชื่อถือได้มากน้อยเพียงไร จึงเป็นที่มาของคำว่า “เหมืองข้อมูล (Data Mining)”

เหมืองข้อมูล คือ เทคนิคหรือแนวคิดที่ว่าด้วยการสกัดเอาสารสนเทศหรือข้อมูลที่น่าจะเป็นประโยชน์ออกมาจากข้อมูลดิบ ไม่ว่าจะเป็นอยู่ในรูปแบบของไฟล์ข้อมูล ฐานข้อมูล รูปภาพ เสียง หรือมัลติมีเดีย ออกมาเพื่อสร้างเป็นสารสนเทศหรือองค์ความรู้ (Knowledge) เพื่อที่เราจะนำเอาสารสนเทศหรือองค์ความรู้ที่ได้ไปในด้านต่าง ๆ โดยในหนังสือ Data mining: Practical Machine Learning Tools and Techniques เขียนโดยศาสตราจารย์ Lan H. Witten และ ศาสตราจารย์ Eibe Frank จากภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ ได้แบ่งขั้นตอนในการทำเหมืองข้อมูลออกเป็นขั้นตอนต่าง ๆ ดังต่อไปนี้

1. Data Cleaning เป็นขั้นตอนเป็นขั้นตอนแรกที่ใช้ในการจัดข้อมูลที่ไม่เกี่ยวข้องออกไป รวมถึงการปรับเปลี่ยนข้อมูลที่มีความซ้ำซ้อนหรือผิดพลาดออกไปด้วย
2. Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
3. Data Selection เป็นขั้นตอนที่ใช้ในการเลือกดึงข้อมูลที่น่าจะเป็นข้อมูลที่มีความสำคัญหรือเป็นประโยชน์ออกมาสำหรับการวิเคราะห์ออกมา
4. Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน เช่น การแปลงข้อมูลจากตัวเลขเป็นตัวอักษร หรือการเปลี่ยนแปลงข้อมูลข้อความเป็นตัวเลขเพื่อให้สามารถนำไปใช้ในการคำนวณได้



5. Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่ โดยในขั้นตอนนี้จะมีความสำคัญและเกี่ยวข้องกับการทำเหมืองข้อมูลมากที่สุด

6. Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล

7. Knowledge Representation เป็นขั้นตอนสุดท้าย เป็นการนำเอาโมเดลหรือผลของคำตอบที่ได้ไปใช้เพื่อทำการสนับสนุนการตัดสินใจและใช้ประโยชน์

เทคนิคที่ใช้ในการทำเหมืองข้อมูล (Data mining) นั้น สามารถแบ่งตามรูปแบบการนำไปใช้งานได้เป็น 3 รูปแบบใหญ่ ๆ คือ

1. เทคนิคที่เรียกว่า Classification หรือเทคนิคการจำแนกเป็นเทคนิคหนึ่งที่ได้รับคามนิยมอย่างมาก เพราะเทคนิคนี้เป็นเทคนิคที่ใช้เพื่อการพยากรณ์หรือทำนายผลข้อมูลที่จะเกิดขึ้นในอนาคตโดยอาศัยข้อมูลที่เคยเกิดขึ้นมาก่อนในอดีตเป็นข้อมูลช่วยในการเรียนรู้และจดจำ ซึ่งโมเดลที่ได้จากเทคนิคนี้จะมีคามแม่นยำมากน้อยเพียงไรขึ้นอยู่กับข้อมูลในอดีตที่เคยเกิดขึ้นมาก่อน ยิ่งข้อมูลในอดีตมีจำนวนมากเท่าใด และมีความน่าเชื่อถือมากเท่าใด ตัวโมเดลที่ได้จากการเรียนรู้ก็จะมีคามแม่นยำในการพยากรณ์มากขึ้นตามไปด้วย โดยเทคนิคนี้นำไปประยุกต์ใช้งานทางด้าน การทำนายผลการตอบสนองกับการเปิดตัวสินค้าใหม่ การทำนายยอดขายเมื่อมีการลดราคาสินค้า การทำนายกลุ่มลูกค้าที่น่าจะใช้สินค้าของเรา การคาดการณ์ถึงโอกาสในการชำระหนี้ของลูกค้าว่าสูงเท่าไร? และ ทำนายแนวโน้มของพฤติกรรม ฯลฯ

2. เทคนิคที่เรียกว่า Clustering คือ เทคนิคที่นิยมใช้ในการวิเคราะห์ข้อมูลเพื่อทำการจัดกลุ่มหรือแบ่งแยกข้อมูลที่มีความคล้ายคลึงกันเข้าอยู่ในกลุ่มเดียวกัน และจะแยกข้อมูลที่แตกต่างกันให้ออกห่างจากกัน โดยเราจะแยกแต่ละกลุ่มย่อยที่แยกออกมาว่า Cluster โดย เทคนิค Clustering เป็นวิธีที่สามารถจัดกลุ่มข้อมูลได้โดยไม่ต้องอาศัยการเรียนรู้ (Unsupervised Learning ซึ่งจะแตกต่างจากเทคนิค Classification ซึ่งต้องอาศัยการเรียนรู้ (Supervised Learning) โดยเทคนิค Clustering นี้เป็นเทคนิคที่นิยมถูกนำไปประยุกต์ใช้เพื่อ การส่งเอกสารทำโปรโมชันให้ตรงกับกลุ่มเป้าหมายมากขึ้น หรือ การจัดวาง Positioning ของสินค้าในบริษัทว่าจะกลุ่มเป้าหมายใด ตัวอย่างเช่น : บริษัทขายบ้านแบ่งกลุ่มลูกค้าเพื่อส่งโปรโมชันที่เหมาะสมกับรายได้ของลูกค้าแต่ละกลุ่มไปให้ตรงกลุ่มเป้าหมายมากที่สุด

3. เทคนิคที่เรียกว่า Association หรือการหาความสัมพันธ์ เป็นกระบวนการหนึ่งในการทำ Data Mining ที่ได้รับความนิยมมากโดยเฉพาะงานทางด้านธุรกิจ โดยจะใช้ Association Rules ถูกใช้เพื่อหาความสัมพันธ์ของข้อมูลสองชุดหรือมากกว่าสองชุดขึ้นไปภายในกลุ่มข้อมูลที่มีขนาดใหญ่ ในการหาความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ ขั้นตอนวิธี Apriori ตัวอย่างหนึ่งของ Association Rules ที่ใช้กันก็คือ Market Basket Analysis ที่ใช้ในการหาความสัมพันธ์ของสินค้าที่ลูกค้ามักจะซื้อพร้อมกัน เพื่อใช้ในการจัดรายการส่งเสริมการขาย เนื่องจากใน

ปัจจุบันสภาวะการแข่งขันในธุรกิจค้าปลีกสูงขึ้น เนื่องจากการขยายตัวอย่างรวดเร็วของร้านค้าปลีกที่มีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่องในแต่ละปี ทำให้กลุ่มลูกค้าผู้บริโภคมีตัวเลือกในการซื้อสินค้ามากขึ้น ดังนั้นไม่ว่าจะธุรกิจค้าปลีกขนาดเล็ก ขนาดกลาง หรือขนาดใหญ่ ต่างต้องมีการปรับตัวหาแนวทางในการส่งเสริมการขายให้องค์กรของตัวเองอยู่รอด ส่งผลให้รูปแบบการการค้าขายของผู้ประกอบการร้านค้าได้มีการเปลี่ยนแปลงไปจากเดิมมาก โดยผู้ประกอบการต้องปรับกลยุทธ์ และหาแนวทางส่งเสริมการขายสินค้าของตนเอง ตัวอย่างเช่น การนำเสนอสินค้าและบริการที่เหมาะสมกับความต้องการของลูกค้า และกลุ่มเป้าหมาย การนำเสนอรูปแบบหรือกิจกรรมต่าง ๆ เพื่อสร้างความสนใจดึงดูดลูกค้าในการซื้อสินค้า การปรับปรุงรูปแบบการจัดเรียงสินค้าภายในร้านให้มีความสะดวกในการเลือกซื้อสินค้า รวมไปถึงการบริหารงาน และการควบคุมต้นทุน ซึ่งต้องคำนึงถึงการจัดสินค้าคงคลังให้มีประสิทธิภาพและแนวทางซึ่งกำลังได้รับความนิยมในต่างประเทศ รวมถึงธุรกิจค้าปลีกขนาดใหญ่ในประเทศไทย คือการมองว่า “ในการซื้อสินค้าของผู้บริโภค 1 ครั้ง โดยไม่ต้องจำกัดว่าจะซื้อสินค้าในห้างร้าน หรือส่งผ่านทางไปรษณีย์ หรือการซื้อสินค้าจากร้านค้าเสมือนจริง (Virtual store) บนเว็บไซต์ ปกติเราจะต้องการทราบว่าคุณลูกค้าได้บ้างที่ผู้บริโภคมักซื้อคู่กันหรือร่วมกันเสมอ เพื่อนำไปพิจารณาปรับปรุงการจัดวางสินค้าในร้าน หรือใช้เพื่อหาวิธีวางรูปคู่กันในใบโฆษณาสินค้า” ดังนั้นวิธีการส่งเสริมการขายสินค้าที่ดี คือการจัดวางสินค้าที่มีความเกี่ยวข้องกันไว้ใกล้กันเพื่อเป็นการกระตุ้นให้ผู้บริโภคซื้อสินค้า กลุ่มที่มีความเกี่ยวข้องกันได้ง่ายมากยิ่งขึ้น โดยที่ตัวผู้บริโภคเองอาจจะไม่รู้ตัวด้วยซ้ำว่าทำไมต้องซื้อสินค้านี้ดังกล่าว ซึ่งพฤติกรรมที่เกิดขึ้นนี้เรียกว่า “พฤติกรรมซ่อนเร้นของผู้บริโภค (Hidden Behavior)” อย่างไรก็ตามการที่จะทำให้แนวทางดังกล่าวประสบผลสำเร็จจะประยุกต์ใช้เทคนิคและวิธีการทางด้านเหมืองข้อมูล (Data Mining) ด้วยวิธีการหาความสัมพันธ์ (Association Rule) จากฐานขนาดใหญ่ที่มีอยู่ เพื่อนำไปใช้ในการวิเคราะห์ การซื้อสินค้าของผู้บริโภคที่เรียกว่า “Market Basket Analysis” ซึ่งเป็นการประเมินจากข้อมูลในตารางที่รวบรวมไว้ทุกครั้งที่มีผู้บริโภคมีการซื้อสินค้า ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา ซึ่งจะช่วยให้ทราบว่าสินค้าประเภทใดบ้างที่ผู้บริโภคมักจะซื้อไปพร้อม ๆ กันกับสินค้าชนิดอื่น ๆ ทำให้ผู้ประกอบการสามารถนำผลที่ได้ไปใช้ในการปรับเปลี่ยนวิธีจัดเรียงสินค้าและจัดทำโปรโมชั่นตลอดจนโฆษณาได้ตรงตามความต้องการของกลุ่มลูกค้าได้มากยิ่งขึ้น

โดยในหนังสือเล่มนี้ได้อธิบายให้เห็นว่าเทคนิคการทำเหมืองข้อมูลเป็นเทคนิคที่เริ่มมีบทบาทสำคัญเป็นอย่างมากในปัจจุบันและในอนาคตเทคนิคการทำเหมืองข้อมูลยังจะมีบทบาทต่อการดำเนินธุรกิจมากขึ้นเรื่อย ๆ อย่างไรก็ตามในปัจจุบันการทำเหมืองข้อมูลไม่ได้ถูกจำกัดอยู่เพียงแค่การนำไปใช้ในงานทางด้านธุรกิจเท่านั้น เพราะการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้กับงานทางด้านการแพทย์ เช่นการพยากรณ์อาการของโรค ด้านการทหาร เช่นการหาโอกาสที่จะรบชนะในสงคราม ด้านการศึกษา เช่น การใช้เป็นเครื่องมือที่ช่วยวางแผนการเรียนการสอน หรือการเลือกคณะหรือสาขาวิชาที่เหมาะสมให้กับนักศึกษาที่



จะเรียนต่อในระดับอุดมศึกษา และในยุคเทคโนโลยีในศตวรรษที่ 21 ซึ่งเทคโนโลยีกำลังเปลี่ยนไปเป็นระบบ Cloud Computing และ Big Data ทำให้ข้อมูลที่มีอยู่ในโลกอินเทอร์เน็ตมีจำนวนมหาศาลมากยิ่งขึ้น และจะสูงขึ้นเรื่อย ๆ ดังนั้นเทคนิคการทำเหมืองข้อมูลก็ยิ่งจะมีบทบาทสำคัญมากยิ่งขึ้นเพราะจะเป็นเครื่องมือสำคัญที่จะทำให้ธุรกิจประสบความสำเร็จหรือล้มเหลวก็เป็นได้