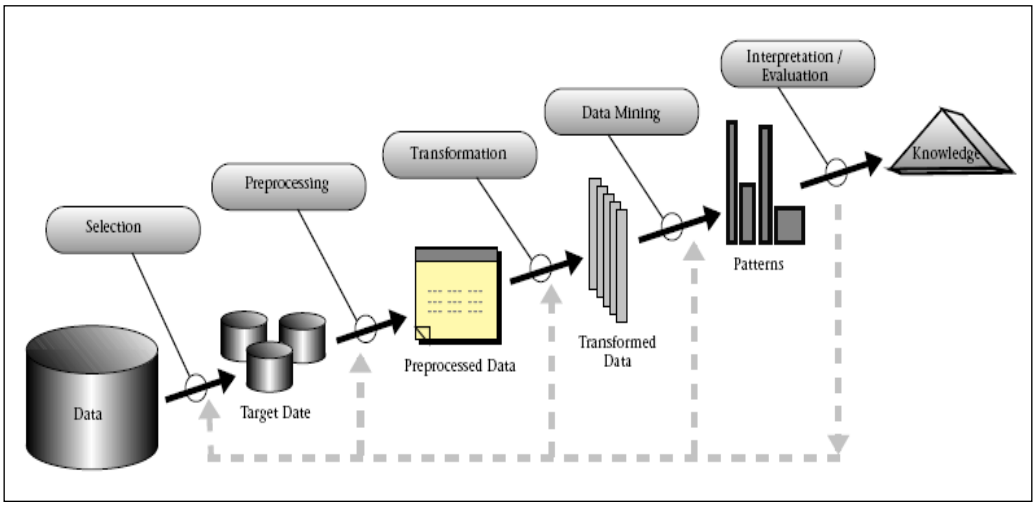
บทที่ 6
การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล เป็นขั้นตอนหนึ่งในการค้นหาความรู้ในฐานข้อมูล (Fayyad *et al.*, 1996) ซึ่งประกอบด้วยการเลือกขั้นตอนวิธีที่จะใช้ในการวิเคราะห์ พร้อมทั้งการวิเคราะห์ข้อมูลเพื่อค้นหารูปแบบที่ซ่อนอยู่ในข้อมูลเหล่านั้น (Olson *et al.*, 2008) โดยขั้นตอนการทำงานทั้งหมดสามารถแสดงดังภาพประกอบ 6.1



ภาพประกอบ 6.1 กระบวนการค้นหาความรู้ในฐานข้อมูล (Fayyad *et al.*, 1996)

การทำเหมืองข้อมูลเป็นส่วนสำคัญส่วนหนึ่งสำหรับการค้นหาความรู้ในฐานข้อมูล จุดประสงค์ของการทำเหมืองข้อมูล คือ การดึงรูปแบบแนวโน้มและกฎเกณฑ์จากข้อมูลเพื่อนำมาพัฒนาหรือประยุกต์ใช้กับองค์กร หรือเป็นตัวตัดสินในการปรับปรุงการดำเนินงานหรือวิธีการขององค์กร ผลลัพธ์ที่ได้จากการทำเหมืองข้อมูล คือ ได้ความรู้หรือสิ่งที่เป็นประโยชน์ใหม่ๆ เพื่อระบุเหตุผลที่สนับสนุนการตัดสินใจ และเข้าใจถึงรูปแบบของข้อมูล (Taniar, 2007) ซึ่งการทำเหมืองข้อมูลมีกระบวนการทำงานเพื่อให้ได้มาซึ่งความรู้ มีขั้นตอนการทำงานที่สำคัญๆ ดังนี้

1) การเลือกข้อมูล (Data Selection)

การเลือกข้อมูล คือ การระบุแหล่งของข้อมูลที่มีและดึงเอาข้อมูลออกมาใช้สำหรับการวิเคราะห์เบื้องต้นเพื่อนำไปสู่ขั้นตอนการเตรียมข้อมูล ซึ่งต้องทำความเข้าใจถึงลักษณะของข้อมูล (Understanding Data) ในฐานข้อมูลว่าข้อมูลแต่ละตัวแปรนั้นมีความหมายว่าอย่างไร ต้องมีคำอธิบายที่ชัดเจนเกี่ยวกับชนิดของข้อมูล ค่าที่เป็นไปได้ แหล่งกำเนิดของข้อมูล รูปแบบของข้อมูล เพื่อใช้ในการกำหนดเป้าหมาย (Target Data) ของผลลัพธ์หรือความรู้ที่ต้องการจากฐานข้อมูลเหล่านั้นสำหรับในกระบวนการค้นหาความรู้ในฐานข้อมูล โดยเลือกเฉพาะข้อมูลที่คาดว่าเกี่ยวข้องกับ

เป้าหมายที่ต้องการจากฐานข้อมูลเพื่อใช้เป็นข้อมูลนำเข้า (Input) สู่ขั้นตอนต่อไปของกระบวนการค้นหาความรู้ในฐานข้อมูล ซึ่งการเลือกข้อมูลนั้นจะแตกต่างกันไปตามวัตถุประสงค์ของแต่ละงานที่ได้กำหนดไว้ตั้งแต่ต้น

2) การเตรียมข้อมูล (Data Preprocessing)

การเตรียมข้อมูลเป็นขั้นตอนที่สำคัญขั้นตอนหนึ่ง เนื่องจากต้องให้ความสำคัญในการตรวจสอบข้อมูลและทำความสะอาดข้อมูล (Cleansing Data) เพื่อให้ได้มาซึ่งข้อมูลนำเข้าที่มีคุณภาพ หากข้อมูลนำเข้าไม่มีคุณภาพแล้วข้อมูลนำออก (Output) หรือผลลัพธ์ที่ได้นั้นจะดีได้อย่างไร โดยขั้นตอนการเตรียมข้อมูลสามารถแยกออกเป็น 3 วิธีการดังนี้

(1) การตรวจสอบขอบเขตของข้อมูล คือ กำจัดข้อมูลที่ไม่ถูกต้องหรือข้อมูลที่ไม่สอดคล้องออกไป เช่น ข้อมูลเงินเดือนต้องเป็นตัวเลข ห้ามเป็นตัวอักษรหรือสัญลักษณ์ เป็นต้น

(2) การกำจัดข้อมูลสูญหาย (Missing Data) หรือข้อมูลที่ไม่สมบูรณ์ เช่น แก้ไขข้อมูลที่เป็นค่าว่าง (Null) เพื่อให้มีค่าหรือตัดข้อมูลนั้นออกไป เพื่อไม่ให้มีผลกระทบกับข้อมูลอื่นๆ เป็นต้น

(3) การกำจัดข้อมูลที่ผิดปกติ (Noisy Data) คือ ค่าที่มีความแตกต่างไปจากข้อมูลอื่นๆ ในฐานข้อมูลอย่างชัดเจน โดยในการตรวจสอบเพื่อกำจัดข้อมูลเหล่านี้อาจใช้วิธีการทางสถิติมาประกอบการตัดสินใจ

3) การลดรูปหรือแปลงข้อมูล (Data Transformation)

การลดรูปหรือแปลงข้อมูลนำเข้า โดยข้อมูลที่ได้ผ่านการทำความสะอาดแล้วจะถูกแปลงให้อยู่ในรูปแบบของข้อมูลที่พร้อมจะถูกวิเคราะห์ ซึ่งภายในขั้นตอนนี้เป็นขั้นตอนที่สำคัญมาก เนื่องจากต้องการความถูกต้องเพื่อให้อยู่ในรูปแบบที่อำนวยต่อการทำงานสำหรับการค้นหาความรู้ในฐานข้อมูลและนำไปสู่ความรู้ที่ตั้งเป้าหมายไว้อย่างถูกต้อง

4) การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล จะประกอบด้วยขั้นตอนของการเลือกเทคนิคของการทำเหมืองข้อมูล (Data Mining Techniques) เป็นขั้นตอนที่สำคัญ เนื่องจากเทคนิคที่เลือกนั้นต้องคำนึงถึงลักษณะข้อมูลที่เรามีอยู่และผลลัพธ์หรือข้อมูลนำออกที่เราต้องการว่ามีรูปแบบอย่างไร เพื่อที่จะเลือกใช้เทคนิคที่เหมาะสมในการค้นหาความรู้ในฐานข้อมูลตามที่ได้วางแผนไว้ เมื่อเลือกเทคนิคของการทำเหมืองข้อมูลแล้ว ขั้นตอนต่อไปคือการเลือกขั้นตอนวิธี (Selecting Algorithm) ที่จะใช้ในการค้นหาความรู้ในฐานข้อมูล โดยการเลือกขั้นตอนวิธีนั้นต้องคำนึงถึงลักษณะของข้อมูลนำเข้า เวลาที่ใช้ในการประมวลผล หรือแม้กระทั่งทรัพยากรที่มีอยู่ เพื่อให้ได้ขั้นตอนวิธีที่เหมาะสมและไม่ก่อให้เกิดปัญหาในกระบวนการของการประมวลผลเพื่อใช้ในการวิเคราะห์ ยกตัวอย่างเช่น หากต้องการความรู้ที่อยู่ในรูปของกฎความสัมพันธ์ อาจเลือกใช้เทคนิคการทำเหมืองข้อมูลที่เรียกว่า ความสัมพันธ์

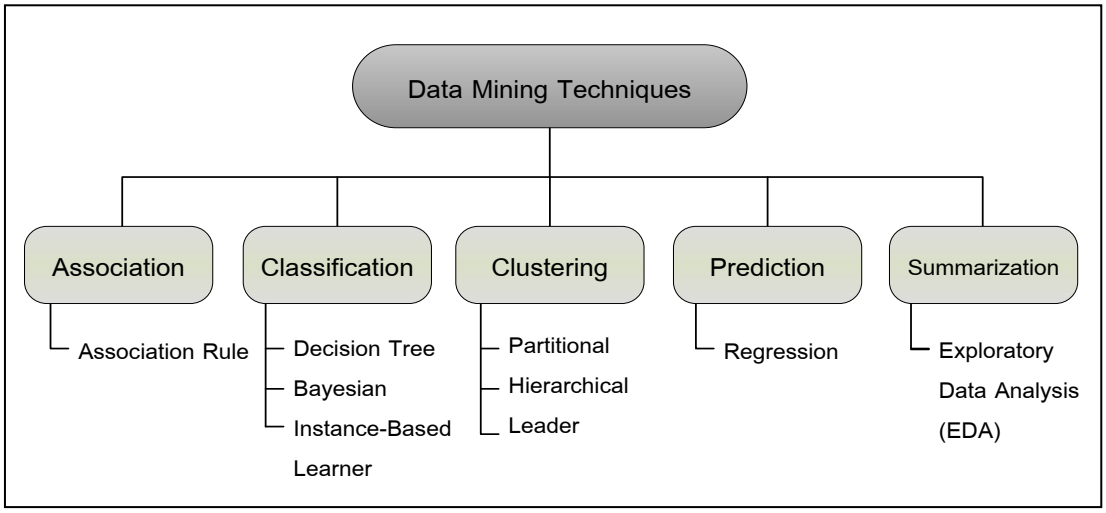
(Association) และเลือกขั้นตอนวิธี Apriori (Apriori Algorithm) ในการประมวลผลเพื่อให้ได้กฎความสัมพันธ์ เป็นต้น

5) การแปลความหมายของผลลัพธ์ (Data Interpretation)

ข้อมูลนำออกหรือผลลัพธ์ที่ได้หลังจากการประมวลผลในขั้นตอนการทำเหมืองข้อมูลนั้นอาจอยู่ในรูปที่ไม่สามารถเข้าใจได้ง่าย จึงต้องมีการแปลความหมายหรือตีความของข้อมูลนำออกที่ได้นั้นให้อยู่ในรูปของความรู้ที่สามารถเข้าใจและนำไปใช้ได้ง่าย จึงต้องมีขั้นตอนของการแปลความหมายข้อมูล โดยข้อมูลที่แปลความหมายแล้วอาจอยู่ในรูปของตาราง กราฟ หรือรูปภาพ เป็นต้น ซึ่งหลังจากในขั้นตอนนี้แล้วผลลัพธ์ที่ได้มีความผิดพลาดหรือไม่ตรงกับเป้าหมายที่ตั้งไว้ก็อาจย้อนกลับไปทำงานในหัวข้อที่ 2.4.1 ถึง 2.4.5 เพื่อให้ได้ความรู้ที่ถูกต้องและมีประโยชน์มากที่สุด และเมื่อทำการแปลข้อมูลนำออกให้อยู่ในรูปแบบที่พร้อมนำไปใช้งานแล้ว ในขั้นตอนนี้ก็สามารถนำความรู้ที่ได้ไปใช้ให้เกิดประโยชน์ต่อไป โดยความรู้ที่ได้อาจนำไปใช้ในการสนับสนุนการตัดสินใจ พัฒนาองค์กรให้มีคุณภาพ หรือสร้างผลิตผลิตภัณฑ์ใหม่ๆ จากความรู้ที่ได้ เป็นต้น

6.1 เทคนิควิธีการทำเหมืองข้อมูล (Data Mining Techniques)

การเลือกเทคนิควิธีการทำเหมืองข้อมูลเป็นขั้นตอนที่สำคัญ (Mitra and Acharya, 2003) เนื่องจากการทำเหมืองข้อมูลมีวิธีการต่างๆ ที่ใช้ในการแก้ปัญหาอยู่หลายวิธีการ ซึ่งจะไม่มีวิธีการใดเลยที่สามารถแก้ปัญหาของทำเหมืองข้อมูลได้ทุกปัญหา ดังนั้นความหลากหลายของวิธีการเป็นสิ่งจำเป็นที่จะนำไปสู่การแก้ปัญหาที่ดีที่สุดของการทำเหมืองข้อมูล เพื่อให้ได้ความรู้ตามที่ต้องการจากการค้นหาความรู้ในฐานข้อมูลเหล่านั้น โดยสามารถสรุปเทคนิควิธีการทำเหมืองข้อมูลดังภาพประกอบ 6.2 พร้อมทั้งอธิบายได้ดังนี้



ภาพประกอบ 6.2 เทคนิควิธีการทำเหมืองข้อมูล

1) ความสัมพันธ์ (Association)

การค้นหาความสัมพันธ์ของข้อมูลเป็นเทคนิคหนึ่งของการทำเหมืองข้อมูล โดยหลักการทำงาน คือ การหาความสัมพันธ์ของข้อมูลภายในกลุ่มข้อมูล เพื่อใช้ลักษณะของข้อมูลหนึ่งในการบอกถึงลักษณะที่จะเกิดขึ้นกับข้อมูลอีกตัวหนึ่ง ซึ่งอาจจะเป็นการหาความสัมพันธ์ของข้อมูลในกลุ่มเดียวกัน ยกตัวอย่างเช่น การระบุว่าในกลุ่มของลูกค้าที่ซื้อน้ำอัดลมแล้วจะซื้อขนมคบเขี้ยวด้วยนั้นมีโอกาสเกิดขึ้นร่วมกัน หรืออาจจะเป็นการหาความสัมพันธ์ของตัวแปรระหว่างกลุ่มข้อมูล ยกตัวอย่างเช่น ในทุกๆ ครั้งที่ฝนตกหนักขึ้นจะทำให้มีปริมาณน้ำในแม่น้ำเพิ่มขึ้น เป็นต้น โดยลักษณะของการหาความสัมพันธ์นั้นอาจแบ่งได้เป็น 3 กลุ่ม ดังนี้

(1) การหาความสัมพันธ์ระหว่างข้อมูล (Association Discovery) ยกตัวอย่างเช่น การวิเคราะห์หาความสัมพันธ์ของการซื้อสินค้าของลูกค้า (Market Basket Analysis) เพื่อใช้ในวางแผนสำหรับการจัดทำโปรโมชั่น (Liao and Chen, 2004) หรือการพัฒนาสินค้าให้ตรงตามความต้องการของลูกค้า (Liao *et al*., 2003) เป็นต้น

(2) การหาความสัมพันธ์ในลักษณะที่เป็นลำดับของข้อมูล (Sequential Pattern Discovery) ยกตัวอย่างเช่น การระบุความเกี่ยวเนื่องกันของการซื้อสินค้าของลูกค้า โดยมีจุดมุ่งหมายที่จะเข้าใจถึงพฤติกรรมการซื้อสินค้าของลูกค้าในลักษณะระยะยาว (Long Term) ยกตัวอย่างเช่น ลูกค้าที่ซื้อเครื่องคอมพิวเตอร์ไปแล้วมีแนวโน้มที่จะกลับมาซื้อเครื่องพิมพ์ในเวลาต่อมา

(3) การหาความสัมพันธ์ของข้อมูลกับช่วงเวลาใดๆ (Similar Time Sequence Discovery) ใช้สำหรับการค้นหาความเกี่ยวเนื่องกันระหว่างกลุ่มของข้อมูล 2 กลุ่มที่มีผลต่อกัน โดยจะมีช่วงเวลาเข้ามาเกี่ยวข้องด้วย ยกต้วอย่าง เช่น เมื่อใดก็ตามที่ ยอดขายสินค้าเครื่องดื่มแอลกอฮอล์สูงขึ้น ยอดขายอาหารคบเขี้ยวจะสูงขึ้นตาม

ผลลัพธ์ที่ได้จากการค้นหาความสัมพันธ์นั้นสามารถนำไปใช้ในการวิเคราะห์สนับสนุนหรือพัฒนางานที่เกี่ยวข้องกับฐานข้อมูลนั้นๆ ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น ซึ่งวิธีที่ได้รับความนิยมและเป็นที่รู้จัก คือ การค้นหากฎความสัมพันธ์ โดยขั้นตอนสำคัญสำหรับการค้นหากฎความสัมพันธ์นั้น คือ การค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย

2) การจัดหมวดหมู่ (Classification)

การจัดหมวดหมู่ข้อมูล คือ การจัดกลุ่มให้กับแต่ละข้อมูลในฐานข้อมูล เป็นการแบ่งกลุ่มข้อมูลในฐานข้อมูลตามลักษณะเด่นของหมวดหมู่นั้นๆ โดยมีการระบุค่าหรือลักษณะที่เป็นไปได้ของข้อมูลภายในแต่ละกลุ่ม การจัดหมวดหมู่จะให้ผลลัพธ์เป็นค่าที่แน่นอน ยกตัวอย่างเช่น การจัดกลุ่มของผู้ป่วยตามผลของการใช้ยารักษาเพื่อระบุรูปแบบการรักษาให้กับผู้ป่วยใหม่ที่เข้ารับการรักษา

เป็นต้น โดยวิธีการที่เป็นที่รู้จักอย่างแพร่หลายสำหรับการจัดหมวดหมู่ คือ ต้นไม้ตัดสินใจ (Decision Tree) เนื่องจากผลลัพธ์ที่ได้จากการจัดหมวดหมู่สามารถทำความเข้าใจได้ง่าย

3) การรวมกลุ่ม (Clustering)

การรวมกลุ่ม คือ การแบ่งข้อมูลเป็นแบบกลุ่มๆ โดยการรวมกลุ่มข้อมูลดังกล่าวนั้นได้จากการพิจารณาคุณสมบัติในหลายๆมิติของข้อมูล ซึ่งหากถ้ารายการในข้อมูลมีลักษณะคล้ายคลึงกันเป็นกลุ่มเดียวกันได้ก็จะรวมเข้าด้วยกัน หลักการทำงานของการรวมกลุ่ม คือ พยายามมองหาความเหมือนและความแตกต่างภายในกลุ่มของข้อมูลและแบ่งกลุ่มต่าง ๆ ออกเป็นส่วน ๆ ยกตัวอย่างเช่น หากองค์กรต้องการทราบความเหมือนที่มีในกลุ่มลูกค้าของตนเอง เพื่อจะสามารถเข้าใจลักษณะเฉพาะของกลุ่มลูกค้าเป้าหมายขององค์กร และสามารถสร้างกลุ่มของลูกค้าเพื่อที่องค์กรจะสามารถขายสินค้าได้ในอนาคตนั้น องค์กรก็สามารถใช้วิธีการของการรวมกลุ่มเพื่อทำการแยกกลุ่มของลูกค้าออกเป็นกลุ่มๆ ได้ เป็นต้น โดยวิธีการที่เป็นที่รู้จัก คือ การรวมกลุ่มแบบ *K*-means (*K*-means Clustering) การทำงานของ *K*-means จะเป็นการกำหนดเพียงจำนวนกลุ่มที่ต้องการเท่านั้น แล้วผลลัพธ์ที่ได้ข้อมูลจะถูกแบ่งออกเป็นกลุ่มๆ ตามจำนวนกลุ่มที่กำหนด

4) การทำนายล่วงหน้า (Prediction)

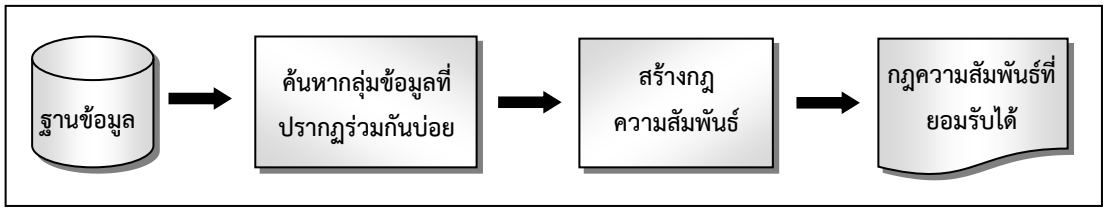
การทำนายล่วงหน้าเป็นเทคนิคที่เหมือนกับการจัดหมวดหมู่ แต่มีความแตกต่างตรงที่ว่าการทำนายล่วงหน้านั้น รายการข้อมูลที่ถูกแยกเพื่อจัดลำดับนั้น เกิดขึ้นตามลักษณะการทำนายพฤติกรรมในอนาคตหรือการทำนายค่าที่จะเกิดขึ้นในอนาคต ข้อมูลในอดีตจะถูกสร้างเป็นต้นแบบ (Model) ขึ้นมาเพื่อทำนายหรืออธิบายสิ่งที่จะเกิดขึ้นในอนาคต ตัวอย่างเช่นการทำนายว่าหน้าร้อนในปีหน้าอุณหภูมิสูงสุดจะเป็นกี่องศาเซลเซียส (Degree Celsius) หรือปริมาณน้ำในแม่น้ำจะเพิ่มขึ้นเป็นเท่าใดหากฝนตกหนักขึ้น 10% เป็นต้น โดยวิธีการหนึ่งที่ใช้ในการทำนายล่วงหน้า คือ สมการถดถอย (Regression) หลักการทำงานจะใช้ความรู้ทางสถิติมาช่วยในการสร้างสมการขึ้นมาจากข้อมูลที่มีอยู่ ผลลัพธ์ที่ได้จะอยู่ในรูปของสมการเพื่อใช้ในการทำนาย

5) การสรุป (Summarization)

การสรุป คือ การนำข้อมูลที่มีอยู่ในฐานข้อมูลมาสรุปให้อยู่ในรูปแบบที่สามารถเข้าใจถึงลักษณะของข้อมูลให้ง่ายขึ้นโดยใช้ความรู้ทางสถิติเข้ามาผสมผสานกับกราฟิก (Graphic) เช่น กราฟที่อยู่ในรูปของสามมิติเพื่อให้สามารถเห็นลักษณะ แนวโน้ม หรือการกระจายตัวของข้อมูลได้อย่างชัดเจน บางครั้งอาจใช้เทคนิคการสรุปเพื่อตรวจสอบความถูกต้องหรือการสูญหายของข้อมูลก่อนการค้นหาความรู้ในฐานข้อมูล โดยวิธีการที่ใช้ในการสรุปนั้น ยกตัวอย่างเช่น วิธีการ Exploratory Data Analysis (EDA)

6.2 การค้นหากฎความสัมพันธ์ (Association Rules)

การค้นหากฎความสัมพันธ์ของข้อมูลในฐานข้อมูล (Rygielski *et al*., 2002; Agrawal *et al*., 1993; Berberidis *et al*., 2005) ได้มีการพัฒนาขึ้นครั้งแรกโดยนักวิจัยจากศูนย์วิจัย IBM (International Business Machines Corporation) ประเทศสหรัฐอเมริกา มีจุดประสงค์เพื่อการค้นหาความสัมพันธ์ที่น่าสนใจซึ่งซ่อนอยู่ในข้อมูล โดยการนำไปใช้ที่เห็นได้ชัดเจนสำหรับการค้นหาความสัมพันธ์ คือ การค้นหากฎความสัมพันธ์ของการซื้อสินค้าของลูกค้าว่าจะซื้อสินค้าใดบ้างร่วมกันในรถเข็นในซูเปอร์มาร์เก็ต (Market Basket Analysis) เพื่อทำความเข้าใจถึงพฤติกรรมการซื้อสินค้าของลูกค้า ยกตัวอย่างเช่น เมื่อลูกค้าซื้อนมแล้วจะซื้อขนมปังด้วย เป็นต้น การค้นหากฎความสัมพันธ์นั้นมีขั้นตอนการทำงานที่สำคัญที่สุด คือ การค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยซึ่งจะใช้เวลาในการทำงานนานที่สุด ดังนั้นในการเลือกขั้นตอนวิธีจึงต้องเลือกขั้นตอนวิธีที่เหมาะสมกับลักษณะของข้อมูลในฐานข้อมูล เพื่อลดระยะเวลาและเนื้อที่หน่วยความจำในการทำงาน โดยทั่วไปการค้นหากฎความสัมพันธ์สามารถแสดงดังภาพประกอบ 6.3 แบ่งการทำงานออกเป็น 2 ขั้นตอน คือ ขั้นตอนการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยทั้งหมด โดยหาได้จากกลุ่มข้อมูลในฐานข้อมูลที่เกิดขึ้นร่วมกันในแต่ละรายการข้อมูล โดยกลุ่มข้อมูลเหล่านั้นจะต้องมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่ผู้ใช้กำหนดจึงจะถือว่าเป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย และขั้นตอนการนำกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยมาสร้างกฎความสัมพันธ์ โดยกฎความสัมพันธ์จะเป็นที่ยอมรับได้หากตรวจสอบค่าความเชื่อมั่นของกฎนั้นแล้วมีค่ามากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่ผู้ใช้กำหนด



ภาพประกอบ 6.3 กระบวนการค้นหากฎความสัมพันธ์

6.3 การค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย (Frequent Itemsets)

การค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยเป็นขั้นตอนที่สำคัญสำหรับการค้นหากฎความสัมพันธ์ เนื่องจากต้องค้นหากลุ่มข้อมูลที่อาจเป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยเพื่อนำไปสร้างกฎความสัมพันธ์เป็นจำนวนมาก ซึ่งขั้นตอนนี้จะใช้เวลานานและเนื้อที่ในหน่วยความจำนวนมาก จึง

ทำให้ต้องมีการเลือกขั้นตอนวิธีที่เหมาะสมกับลักษณะข้อมูลในฐานข้อมูล ทรัพยากรของเครื่องคอมพิวเตอร์ที่มีอยู่ หรือผลลัพธ์ของข้อมูลที่ต้องการจากการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย

6.4 ตัวอย่างการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย

จากฐานข้อมูลการซื้อสินค้าในตารางที่ 6.1 แต่ละตัวอักษรภาษาอังกฤษ หมายถึงชิ้นข้อมูลที่ใช้แทนชื่อสินค้าที่ลูกค้าซื้อในแต่ละรายการข้อมูลซึ่งมีทั้งหมด 5 รายการข้อมูล การค้นหากฎความสัมพันธ์จะเริ่มจากการนับความถี่ของชิ้นข้อมูลแต่ละชิ้นในฐานข้อมูล เพื่อหาค่าสนับสนุนว่ามากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำหรือไม่ และเพื่อหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยต่อไป ยกตัวอย่างเช่น มีการซื้อสินค้า C ทั้งหมด 4 รายการข้อมูล กล่าวได้ว่ามีจำนวนความถี่เท่ากับ 4 ซึ่งหากกำหนดค่าสนับสนุนขั้นต่ำเป็น 0.4 (40%) สามารถคำนวณค่าสนับสนุนได้จากจำนวนความถี่ของสินค้า C หารด้วยจำนวนของรายการข้อมูลทั้งหมดในฐานข้อมูล ดังสมการที่ (6.1)

$$\text{ค่าสนับสนุน } C = \frac{4}{5} = 0.8 \qquad (6.1)$$

จะได้ว่า สินค้า C เป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย เนื่องจากมีค่าสนับสนุนเป็น 0.8 (80%) ซึ่งมากกว่าค่าสนับสนุนขั้นต่ำที่กำหนดไว้ คือ 0.4 (40%)

ตารางที่ 6.1 ฐานข้อมูลการซื้อสินค้า

รหัสรายการ (TID)	**ชิ้นข้อมูล (Items)**
001	C D E F G I
002	A C D E L
003	A B D E G
004	A C D H
005	A C D J

6.5 ตัวอย่างการสร้างกฎความสัมพันธ์

การสร้างกฎความสัมพันธ์นั้น จำนวนของชิ้นข้อมูลที่ใช้จะต้องประกอบด้วย 2 ชิ้นข้อมูลขึ้นไปในการสร้างกฎความสัมพันธ์ ยกตัวอย่างเช่น สินค้า A และ D ที่ซื้อร่วมกันมีทั้งหมด 4 รายการข้อมูล กล่าวได้ว่ามีจำนวนความถี่เท่ากับ 4 จะได้ว่า สินค้า A และ D เป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย สามารถนำมาสร้างกฎความสัมพันธ์ได้ ยกตัวอย่างเช่น หากลูกค้าซื้อสินค้า A แล้วจะซื้อสินค้า D ด้วย ($A \rightarrow D$) หรือหากลูกค้าซื้อสินค้า D แล้วจะซื้อสินค้า A ด้วย ($D \rightarrow A$) โดยคำนวณค่าความเชื่อมั่นได้ดังสมการ (6.2) และ (6.3) ตามลำดับ

$$\text{ค่าความเชื่อมั่นของกฎ } A \rightarrow D = \frac{4}{4} = 1.0 \qquad (6.2)$$

$$\text{ค่าความเชื่อมั่นของกฎ } D \rightarrow A = \frac{4}{5} = 0.8 \quad (6.3)$$

แต่กฎดังกล่าวที่สร้างขึ้นจะเป็นกฎความสัมพันธ์ที่สามารถยอมรับได้ก็ต่อเมื่อมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่กำหนดไว้ หากกำหนดค่าความเชื่อมั่นขั้นต่ำเป็น 0.5 (50%) ซึ่งกฎ A→D และกฎ D→A มีค่าความเชื่อมั่นไม่น้อยกว่าค่าความเชื่อมั่นขั้นต่ำที่กำหนดไว้ นั่นคือ 1.0 และ 0.8 ตามลำดับ ดังนั้นจะได้กฎความสัมพันธ์ที่สามารถยอมรับได้ว่า หากลูกค้าซื้อสินค้า A แล้วมีโอกาสที่จะซื้อสินค้า D ด้วยมีความเชื่อมั่นเป็น 100% หรือหากลูกค้าซื้อสินค้า D แล้วมีโอกาสที่จะซื้อสินค้า A ด้วยมีความเชื่อมั่นเป็น 80%