

## บทที่ 5

### สรุปผลการศึกษาและข้อเสนอแนะ

การทำเหมืองข้อมูลเป็นขั้นตอนที่สำคัญอย่างหนึ่งในกระบวนการหาความหมายที่แฝงอยู่ในกลุ่มข้อมูลจำนวนมากที่เก็บไว้ในฐานข้อมูล โดยทำหน้าที่ในการหารูปแบบ หรือความสัมพันธ์ของข้อมูลเหล่านั้น และเนื่องจากในปัจจุบัน บริษัท หรือองค์กรต่างๆ เริ่มให้ความสนใจกับข้อมูลที่ถูกจัดเก็บไว้เพิ่มมากขึ้น โดยมีวัตถุประสงค์เพื่อนำลักษณะเฉพาะที่แฝงอยู่ภายในกลุ่มข้อมูล มาใช้สนับสนุนการตัดสินใจอย่างมีประสิทธิภาพต่อการดำเนินงานที่เป็นประโยชน์ต่อองค์กร ดังนั้นจึงมีการพัฒนา และประยุกต์ใช้เทคโนโลยี รวมถึงระบบในการทำเหมืองข้อมูลอย่างแพร่หลายเพิ่มขึ้น ทั้งในเชิงงานวิจัย และเชิงการค้า ซึ่งในการทำโครงการนี้ได้นำเสนอหลักการและขั้นตอนทั่วไปในกระบวนการค้นหาความหมายที่แฝงอยู่ภายในกลุ่มข้อมูล และการทำเหมืองข้อมูล รวมทั้งประเภทของงานและตัวอย่างวิธีการที่นำมาใช้ในขั้นตอนการทำเหมืองข้อมูลด้วย

ขั้นตอนของการค้นหาความสัมพันธ์ของข้อมูล เป็นหัวใจสำคัญของกระบวนการทำเหมืองข้อมูล แต่ทั้งนี้โมเดลหรือความรู้ที่ค้นหามาได้ จะมีคุณภาพและเป็นประโยชน์เพียงใดนั้นปัจจัยสำคัญที่สุดคือข้อมูล คุณภาพและจำนวนของข้อมูลจะมีผลโดยตรงต่อผลลัพธ์ที่จะได้จากการทำเหมืองข้อมูล คุณภาพของข้อมูลสะท้อนมาจากลักษณะต่างๆ (attribute or feature) ที่ประกอบกันขึ้นเป็นข้อมูลหนึ่งรายการ ถ้าลักษณะที่รวบรวมมาเป็นลักษณะหลักที่จะสามารถตัดสินใจพฤติกรรมของข้อมูล ผลลัพธ์ของการทำการทำเหมืองข้อมูลก็จะต้อง จำนวนของข้อมูลจะเป็นอีกปัจจัยสำคัญที่ช่วยยืนยันว่าผลลัพธ์ที่ได้นั้นมีความเที่ยงตรงสูงเพียงใด ในการทำโครงการนี้ ใช้เวลาส่วนใหญ่ในการจัดเตรียมข้อมูลเพื่อให้มีความสมบูรณ์และให้มีจำนวนของข้อมูลที่เหมาะสม โดยใช้เวลารวมถึง 60 % ของเวลาทั้งหมดในการทำเหมืองข้อมูล

ในทางอุดมคติแล้วจำนวนข้อมูลยิ่งมากเท่าไรจะช่วยให้ผลลัพธ์มีความถูกต้องน่าเชื่อถือมากขึ้นเท่านั้น แต่ในทางปฏิบัติแล้วเนื่องจากทรัพยากรของระบบคอมพิวเตอร์ (ขนาดของหน่วยความจำ และ ความเร็วของหน่วยประมวลผล) มีจำกัด การพยายามลดจำนวนข้อมูลให้สามารถประมวลผลได้ใน เครื่องคอมพิวเตอร์ส่วนบุคคลจึงเป็นแนวทางแก้ปัญหาที่เหมาะสมกว่า ดังนั้นโครงการนี้จึงมุ่งเน้นที่การศึกษาผลของการลดขนาดของข้อมูลว่าเราสามารถลดจำนวนข้อมูลลงได้มากที่สุดถึงจุดใด โดยที่คุณภาพของผลลัพธ์หรือ โมเดลที่ได้จากกระบวนการทำเหมืองข้อมูลยังคงเดิม หรืออยู่ในระดับของความถูกต้องที่ยังยอมรับได้ ผลที่ได้รับตอบแทนจากการลดขนาดของข้อมูลคือ สามารถทำการสกัดหรือขุดค้นได้ในเวลาที่รวดเร็วยิ่งขึ้น หรือในกรณีที่ข้อมูลเริ่มแรกมีขนาด

ใหญ่จนไม่สามารถประมวลผลได้ในคอมพิวเตอร์ส่วนบุคคล การลดขนาดของข้อมูลจะทำให้การทำเหมืองข้อมูลเป็นสิ่งที่เป็นไปได้

ปัจจุบันการทำเหมืองข้อมูล เริ่มได้รับความนิยมในองค์กรต่างๆ มากขึ้น ทั้งนี้เนื่องจากปัจจัยหลายประการ เช่น ปริมาณข้อมูลที่แต่ละองค์กรต้องเกี่ยวข้องกับเพิ่มขึ้น (มีขนาดหน่วยนับเป็น Gigabyte หรืออาจเป็น Terabyte) และนับวันจะเพิ่มขึ้นเรื่อยๆ การทำวิเคราะห์ข้อมูลขนาดใหญ่เหล่านี้เพื่อสร้างคุณค่าให้กับข้อมูลเป็นเรื่องยากแต่ก็จำเป็น ซึ่งการทำเหมืองข้อมูลเป็นเครื่องมือตัวหนึ่งที่เข้ามาช่วยในเรื่องนี้ ความสามารถของหน่วยประมวลผล (CPU) ที่เพิ่มขึ้นในขณะที่ราคาลดลง การเลือกใช้เครื่องมือที่เกี่ยวข้องกับการทำเหมืองข้อมูล ที่ทำงานบนเครื่องคอมพิวเตอร์ส่วนบุคคล (Personal Computer) จะต้องคำนึงถึง คุณสมบัติหลักๆ 4 ประการที่ซอฟต์แวร์จะต้องมี ดังนี้

1. มีเครื่องมือช่วยในการเตรียมข้อมูล (Data preparation facilities) ซึ่งหมายถึงการแปลงข้อมูลจากหลายๆแหล่ง หลายรูปแบบ ให้สามารถนำมาใช้กับ Data Mining ร่วมกันได้ได้อย่างถูกต้อง
2. เปิดโอกาสให้เลือกใช้เทคนิคการวิเคราะห์ต่างๆของ Data Mining ได้ (Selection of Data Mining Operations & Algorithms) โดยส่วนใหญ่แต่ละค่ายจะมี 2-3 แบบให้เลือก
3. ให้ผู้ใช้เลือกซื้อซอฟต์แวร์ได้ตามขนาดข้อมูลและประสิทธิภาพตามต้องการ (product scalability and performance)
4. มีการแสดงผลที่ชัดเจนและน่าสนใจ (Facilities for Visualization of results)