

เพิ่มมูลค่าข้อมูลทางธุรกิจด้วย Data mining



วิษณุคุชร์ เมาระพงษ์

คณาจารย์ภาควิชาการคอมพิวเตอร์ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง



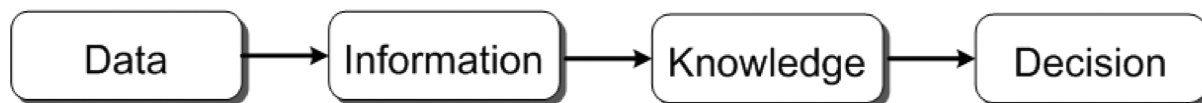
ในปัจจุบันองค์กรที่มีการดำเนินธุรกิจมาเป็นระยะเวลาพอสมควร และมีการใช้เทคโนโลยีสารสนเทศมาช่วยในการจัดเก็บและบริหารข้อมูลนั้น ส่วนใหญ่จะมีข้อมูลปริมาณมาก มีการจัดเก็บไว้โดยที่ไม่ได้นำมาใช้ประโยชน์ อาทิ ข้อมูลลูกค้าที่ซื้อสินค้า หรือใช้บริการ รวมถึงองค์การที่เป็นคู่ค้าและอื่นๆ ซึ่งหากสามารถนำเอาข้อมูลมาประมวลผล วิเคราะห์ และนำเสนอในรูปแบบที่เข้าใจง่าย ก็จะช่วยให้ผู้บริหารสามารถตัดสินใจได้แม่นยำขึ้น และสร้างโอกาส สร้างความได้เปรียบในตลาดที่องค์กรกำลังมองหาช่องทางไปสู่ความสำเร็จอยู่ก็เป็นได้

ระยะเวลาเป็นปัจจัยที่สำคัญประการหนึ่ง ซึ่งหากองค์กรมีการดำเนินงานมาในระยะเวลาที่นานพอสมควรและมีการใช้งานระบบเทคโนโลยีสารสนเทศในการดำเนินกิจกรรมทางธุรกิจ จนมีการจัดเก็บข้อมูลที่เกี่ยวข้องได้ปริมาณมากและเป็นข้อมูลที่วิเคราะห์แล้วสามารถนำมาสร้างเป็นสารสนเทศที่ใช้ประกอบการตัดสินใจสำหรับผู้บริหารได้ก็เหมาะสมที่จะนำมาสร้างให้เกิดประโยชน์อย่างมหาศาลอย่างที่คาดไม่ถึง แทนที่จะเก็บไว้รอเวลาที่จะสูญหายไปอย่างไม่ได้ประโยชน์ตอบแทนกลับมา ทั้งๆ ที่มีการลงทุนเพื่อจัดเก็บไว้มากพอสมควร อย่างไรก็ตาม เนื่องจากข้อมูลที่มีปริมาณมาก วิธีการ หรือเทคนิคในการค้นหาหรือประมวลผลข้อมูลที่ต้องการจากข้อมูลปริมาณมหาศาลนั้น ยังมีเรื่องที่ต้องดำเนินการและลงทุนพอสมควร แต่เชื่อได้ว่าหากทำสำเร็จและเหมาะสมกับองค์การแล้วละก็ จะสามารถเป็นประโยชน์อย่างยิ่งต่อการดำเนินงานในอนาคต เราเรียกเทคโนโลยี หรือเทคนิควิธีการดังกล่าวว่าการทำ **Data mining**

Data mining คืออะไร

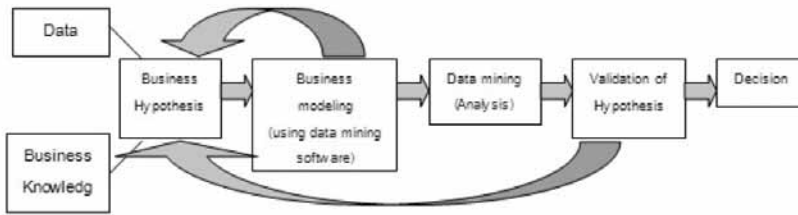
Data mining คือ Application สำหรับวิเคราะห์ประมวลผลข้อมูลที่ถูกออกแบบมาเพื่อใช้งานในระบบสนับสนุนการตัดสินใจของผู้ใช้ เป็น Software ที่มีความสมบูรณ์ทั้งในเรื่องการค้นหา การสร้างรายงาน และการจัดการ ซึ่งเราอาจคุ้นเคยกับระบบ Executive Information System (EIS) หรือระบบข้อมูลสำหรับการตัดสินใจของผู้บริหาร ซึ่งเป็นเครื่องมือหนึ่งที่สามารถค้นหาข้อมูลในฐานข้อมูลขนาดใหญ่ หรือข้อมูลที่เป็นประโยชน์ในการบริหาร เป็นการเพิ่มคุณค่าให้กับฐานข้อมูลที่มีอยู่

ระบบสนับสนุนการตัดสินใจ หรือ Decision Support System (DSS) คือ ระบบที่ช่วยเราตอบโจทย์ที่ว่า ทำอย่างไรให้ข้อมูลที่เราเมืออยู่กลายเป็นความรู้อันมีค่าและสร้างคำตอบของอนาคตได้



1. ในองค์การขนาดกลางถึงขนาดเล็ก กระบวนการทำ Data mining โดยทั่วไปจะเริ่มจากการตั้งสมมติฐานทางธุรกิจตามความรู้และความเข้าใจของ User ที่มีต่อธุรกิจ

3. หลังจากตรวจสอบแก้ไขสมมติฐานในขั้นสุดท้ายแล้ว User ก็ตัดสินใจ



2. ใช้ระบบ Data mining tools โดย user สร้าง model แล้วรันกรองสมมติฐาน ตามด้วยการวิเคราะห์ ซึ่งกระบวนการนี้อาจจะต้องมีการทำซ้ำหลายๆ ครั้ง

ปัจจุบันระบบสนับสนุนข้อมูลในการตัดสินใจได้เข้ามามีอิทธิพลในการรวบรวมข้อมูลและปรับค่าข้อมูลในคลังสินค้า ซึ่งฐานข้อมูลขนาดใหญ่เช่นนี้ จะประกอบไปด้วยข้อมูลจำนวนมาก ยากแก่การค้นหาได้อย่างทันท่วงทีเมื่อต้องการใช้งาน ด้วยวิธีการของระบบฐานข้อมูลทั่วไป (Database Management System: DBMS) ข้อมูลที่เป็นที่สนใจของผู้บริหารองค์กรธุรกิจ ณ วันนี้อาจจะสามารถค้นหาได้ง่ายขึ้นแล้ว ซึ่งเป็นประโยชน์อย่างยิ่งในการค้นหาข้อมูลที่ต้องการในมหาสมุทรข้อมูลเพื่อนำมาเทียบเคียงและดูแนวโน้มและนำข้อมูลที่เป็นขององค์กรส่งกลับให้ผู้บริหารตัดสินใจได้อย่างทันเวลา

นี่คือจุดประสงค์ของ Data mining ที่จะมาช่วยในเรื่องของเทคนิคการจัดการข้อมูล ซึ่งได้มีความพยายามและทดสอบแล้ว ข้อมูลสนับสนุนที่มีอายุย้อนหลังไปถึง 30 ปี ด้วยเทคนิคเดียวกันนี้เราสามารถสืบค้นข้อมูลสำคัญที่ปะปนกับข้อมูลอื่นๆ ในฐานข้อมูลที่ไม่ใช่แค่การสุ่มหา (Knowledge Discovery in Database: KDD) หรือการค้นหาข้อมูลด้วยความรู้และนั่นก็คือ Data mining

ผู้เชี่ยวชาญท่านหนึ่งเคยกล่าวไว้ว่า "Data mining คือ เทคนิคที่ผู้ใช้สามารถปฏิบัติการได้โดยอัตโนมัติกับข้อมูลที่ไม่รู้จัก ซึ่งเป็นการเพิ่มคุณค่าให้กับข้อมูลที่มีอยู่"

จากประโยคข้างต้นมีคำสำคัญอยู่สามคำ คำแรกคือ "อัตโนมัติ" หมายถึง กระบวนการทำงานของ Data mining ซึ่งระบบจะเป็นผู้ทำงานเอง ไม่ใช่ตัวผู้ใช้งาน กระบวนการดังกล่าวจะไม่ให้คำตอบกับปัญหาที่มี แต่จะเป็นทำตัวเป็นศูนย์กลางของข้อมูล คำที่สองคือ "ข้อมูลที่ไม่รู้จัก" หมายถึง เครื่องมือในการค้นหาของ Data mining ซึ่งจะไม่ค้นหาแต่ข้อมูลเก่าและข้อมูลที่ใช้บันทึกเข้าระบบเท่านั้น แต่จะค้นหาข้อมูลใหม่ๆ ด้วยสุดท้ายคือ "เพิ่มคุณค่า" นั้นหมายถึง ผู้ใช้งานไม่ได้เป็นแค่เพียงนักสถิติ (เจ้าหน้าที่ระดับปฏิบัติ นักวิเคราะห์) แต่เป็นได้ทั้งผู้ตัดสินใจ (ผู้บริหาร)

วิวัฒนาการของเทคโนโลยีฐานข้อมูล

ในปี 1960 เทคโนโลยีฐานข้อมูลได้เริ่มพัฒนามาจากระบบ File processing ขึ้นพื้นฐาน การค้นคว้าและพัฒนาระบบฐานข้อมูลก็เริ่มมีมาอย่างต่อเนื่อง

ปี 1970 ได้มีการพัฒนาระบบการเก็บข้อมูลในรูปแบบตารางความสัมพันธ์ (Relational Database System) มีเครื่องมือจัดการโมเดลข้อมูลและมีเทคนิคการใช้ Index (ดัชนี) ในการสืบค้นและการบริหารข้อมูล นอกจากนี้ผู้ใช้อย่างยิ่งได้รับความสะดวกในการเข้าถึงข้อมูลโดยการใช้ภาษาเชิงตรรกะในการเรียกดูข้อมูล (Query Language)

ปี 1980 เทคโนโลยีฐานข้อมูลได้เริ่มมีการปรับปรุงและพัฒนาในการหาระบบจัดการที่มีศักยภาพมากขึ้น ความก้าวหน้าของเทคโนโลยีใน 30 ปีที่ผ่านมา ได้นำไปสู่การจัดเก็บข้อมูลจำนวนมากที่มีความซับซ้อนได้อย่างมีประสิทธิภาพเพิ่มขึ้น

ตั้งแต่ปี 1990 จนถึงปัจจุบัน เราสามารถจัดเก็บข้อมูลได้ในหลายรูปแบบ แตกต่างกันทั้งระบบปฏิบัติการ หรือการจัดเก็บฐานข้อมูล ซึ่งการนำข้อมูลทั้งหมดมารวมและจัดเก็บไว้ในรูปแบบเดียวกันเรียกว่า Data Warehouse เพื่อความสะดวกในการบริหารจัดการต่อไป

เทคโนโลยี Data Warehouse รวมไปถึง Data Cleansing, Data Integration และ On-Line Analytical Processing (OLAP) เป็นเทคนิคการวิเคราะห์ข้อมูลในหลายๆ มิติ นั้นได้ก็เกิดขึ้นตามมาเป็นลำดับ การละเลยข้อมูลควบคู่ไปกับการขาดเครื่องมือวิเคราะห์ข้อมูลที่มีศักยภาพ นำไปสู่สถานการณ์ของการมีข้อมูลมาก แต่กลับมีค่าน้อย (Data rich but information poor) การเติบโตขึ้นอย่างรวดเร็วของข้อมูลจำนวนมากที่สะสมไว้ในฐานข้อมูลขนาดใหญ่ที่เกินกว่ากำลังคนจะสามารถจัดการได้ เป็นผลทำให้มีความจำเป็นที่ต้องมีเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูลและหาความเป็นไปได้ของข้อมูลทั้งหมดที่จะเป็นประโยชน์ออกมา ซึ่งก็คือ Data mining

ทำไมจึงต้องมี Data mining

1. ข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลหากเก็บไว้เฉยๆ ก็จะไม่เกิดประโยชน์ ดังนั้น จึงต้องมีการสกัดเอาสารสนเทศไปใช้งาน การสกัดสารสนเทศ หมายถึง การคัดเลือกข้อมูลในส่วนที่เราต้องการออกมาใช้งาน

2. ในอดีตเราต้องใช้กำลังคนในการค้นข้อมูลต่างๆ ในฐานข้อมูล โดยผู้สืบค้นจะต้องทำการสร้างเงื่อนไขการค้นหาขึ้นมาตามความรู้ ความสามารถของผู้สืบค้น

3. ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูลเดียวอาจไม่ให้ความรู้เพียงพอและลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูงและมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นต้องรวบรวมฐานข้อมูลหลายๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่า "คลังข้อมูล" (Data Warehouse)

ดังนั้น เราจึงจำเป็นต้องใช้ Data mining ในการดึงข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะนำข้อมูลนั้นมาใช้งานให้เกิดประโยชน์สูงสุด