บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

เทคนิคการทำเหมืองข้อมูล (Data mining) ซึ่งเป็นการสำรวจและวิเคราะห์ข้อมูลที่มีขนาดใหญ่เพื่อค้นหาความสัมพันธ์และรูปแบบหรือกฎที่ซ่อนอยู่และนำความสัมพันธ์เหล่านี้แสดงให้เห็นถึงความรู้ต่าง ๆ ที่มีประโยชน์ และ ช่วยพัฒนาระบบสนับสนุนการตัดสินใจ เพื่อวิเคราะห์ข้อมูล และ หาความเป็นไปได้ของข้อมูลเพื่อวิเคราะห์ข้อมูล ทำนายแนวโน้มช่วยในการแนะแนวในการเลือกคณะที่จะศึกษาต่อในระดับอุดมศึกษาต่อไป

ความหมายของคำที่เกี่ยวข้อง

Data Mining คือ การค้นหาความสัมพันธ์และรูปแบบทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูล แต่ได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมาก โดย Data Mining จะเหมาะกับการแก้ปัญหาบางชนิดเท่านั้น เช่น ปัญหาที่ต้องใช้เหตุผลในการแก้ หรือ ปัญหาที่เกี่ยวข้องกับเศรษฐศาสตร์ และ การเงินการศึกษา เป็นต้น Data Mining มีเทคนิคต่าง ๆ ที่ใช้ในการแก้ปัญหาอยู่หลายเทคนิค ซึ่งจะไม่มีเทคนิคใดเลยที่ สามารถแก้ปัญหาของ Data Mining ได้ทุกปัญหา ดังนั้นความหลากหลายของเทคนิคเป็นสิ่งที่จำเป็นที่จะนำไปสู่วิธีการแก้ปัญหาที่ดีที่สุดของ Data Mining (Berry , Michael J.A. and Linn off , Gordon S,2004 :7)

จุดม่งหมายของ Data Mining

1. การดึงรูปแบบแนวโน้มและกฎเกณฑ์จากข้อมูลเพื่อที่จะประเมินกลยุทธ์ของหน่วยงาน
2. ปรับปรุงความได้เปรียบในการแข่งขันเป็นวิธีการที่นำมาใช้ในด้านการตลาด เช่น การรักษาลูกค้า

ส่วนประกอบของ Data Mining

เครื่องมือในการถาม และ จัดทำรายงาน (Query – and – reporting – tools)

1. อุปกรณ์ด้านปัญญาประดิษฐ์ (Intelligent Agents)
2. เครื่องมือในการวิเคราะห์ข้อมูลหลายมิติ (Multidimensional analysis tool-MDA)

วัฏจักรขั้นตอนการทำงานของ Data Mining (Han , Jiawel and Kamber , Micheline,2001: 29)
วัฏจักรขั้นตอนการทำงานของ Data Mining ประกอบไปด้วย 4 ขั้นตอนหลัก ๆ ดังนี้

1. การระบุโอกาสทางธุรกิจหรือการระบุปัญหาที่เกิดขึ้นกับธุรกิจเป็นการระบุขอบเขตของข้อมูลที่จะนำมาทำการวิเคราะห์เพื่อหาความได้เปรียบทางการตลาดหรือเพื่อนำมาทำการแก้ไขปัญหา

2. ส่วนของ Data Miningเป็นการนำเทคนิคของ Data Mining ไปใช้ถ่ายทอดหรือทำการเปลี่ยนแปลงข้อมูลดิบให้อยู่ในรูปของข้อมูลจะนำไปใช้ได้จริงในทางธุรกิจ

3. การปฏิบัติตามข้อมูล คือ การนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วน Data Mining มาลองปฏิบัติจริงกับธุรกิจ

4. การวัดประสิทธิภาพจากผลลัพธ์ การวัดประสิทธิภาพของเทคนิค Data Mining ที่จะนำมาใช้จากผลลัพธ์ ซึ่งสามารถตรวจสอบได้หลายทาง เช่น วัดจากส่วนแบ่งของตลาด วัดจากปริมาณลูกค้า หรือ วัดจากกำไรสุทธิ

จากทั้ง 4 ขั้นตอนที่กล่าวมาข้างต้น คือ การนำเอา Data Mining ไปใช้กับระบบทางธุรกิจ โดยแต่ละขั้นตอนจะพึ่งพาอาศัยกันผลลัพธ์จากขั้นตอนหนึ่งจะกลายเป็น Input จากอีกขั้นตอนต่อไป ซึ่ง Data Mining จะเปลี่ยนข้อมูลดิบให้เป็นข้อมูลประยุกต์ ดังนั้นการระบุแหล่งข้อมูลที่ถูกต้องจึงเป็นสิ่งที่สำคัญอย่างยิ่งต่อผลลัพธ์ที่ได้จากการวิเคราะห์

งานของ Data Mining (Task of data mining)

ในทางปฏิบัติจริงเดต้าไมนิ่งจะประสบความสำเร็จกับงานบางกลุ่มเท่านั้น และต้องอยู่ภายใต้ภาวะที่จำกัดปัญหาเหมาะสมกับการใช้เทคนิคเดต้าไมนิ่งจะเป็นปัญหาที่ต้องใช้เหตุผลในการแก้ เป็นปัญหาที่เกี่ยวข้องกับเศรษฐศาสตร์และการเงิน ซึ่งจะสามารถจัดรูปแบบของธุรกิจให้อยู่ในรูปแบบของงานทั้ง 6 งานได้ ดังนี้

1. การจัดหมวดหมู่ (Classification)
2. การประเมินค่า (Estimation)
3. การทำนายล่วงหน้า (Prediction)
4. การจัดกลุ่มโดยอาศัยความใกล้ชิด (Affinity Group)
5. การรวมตัว (Clustering)
6. การบรรยาย (Description)

1. การจัดหมวดหมู่ การจัดหมวดหมู่ถือว่าเป็นงานธรรมดาทั่วไปของเดต้าไมนิ่ง เพราะการทำความเข้าใจและการติดต่อสื่อสารต่างๆ ก็เกี่ยวข้องกับการแบ่งเป็นหมวดหมู่การจัดแยกประเภทและการแบ่งแยกชนิดโดยการจัดหมวดหมู่ประกอบด้วยการสำรวจจุดเด่นของวัตถุที่ปรากฏออกมาและทำการกำหนด จุดเด่นนั้นๆ เป็นตัวที่ใช้แบ่งหมวดหมู่ งานในการแบ่งหมวดหมู่คือการบ่งบอกลักษณะโดยการอธิบายจุดเด่นที่เป็นที่รู้จักดีในหมวดหมู่นั้นและเทรนนิ่งเซต (Training Set) ของตัวอย่างในแต่ละหมวดหมู่ ซึ่งมีภาระหน้าที่ในการสร้างโมเดลของบางชนิดที่ไม่สามารถจะจัดหมวดหมู่ของข้อมูลได้ ให้สามารถจัดเป็น หมวดหมู่ได้ ตัวอย่างของการจัดหมวดหมู่ เช่น การจัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำ ระดับกลางและระดับสูงของความเสี่ยงที่จะได้รับเป็นต้น

2. การประเมินค่า การประเมินค่าทางธุรกิจอย่างต่อเนื่องจะก่อให้เกิดผลลัพธ์ที่มีประโยชน์กับธุรกิจ การป้อนข้อมูลที่เรามีอยู่เข้าไป เพื่อใช้ในการประเมินสิ่งต่างๆ ที่จะก่อให้เกิดประโยชน์หรือสำหรับตัวแปรที่เราไม่รู้ค่าแน่นอนเช่น รายได้จากจุดสูงสุดทางธุรกิจ หรือดุลยภาพของบัตรเครดิต ในทางปฏิบัติการประเมินค่าจะถูกใช้ในการทำงานการจัดหมวดหมู่ ตัวอย่างของการประเมินค่าเช่นการประเมินรายได้รวมของครอบครัวหรือการประเมินจำนวนบุตรในครอบครัว

3. การทำนายล่วงหน้า การทำนายล่วงหน้าก็เป็นงานที่มีลักษณะคล้ายกับการจัดหมวดหมู่หรือการประเมินค่า ยกเว้น เพียงแต่จะใช้สถิติการบันทึกของการจัดหมวดหมู่ในการทำนายอนาคตของพฤติกรรมหรือการประเมิน ค่าที่จะเกิดขึ้นในอนาคต ตัวอย่างของงานการทำนายล่วงหน้า เช่น การทำนายการเปลี่ยนแปลงพฤติกรรม ของตลาด หรือการทำนายจำนวนลูกค้าที่จะออกจากธุรกิจของเราใน 6 เดือนข้างหน้า เป็นต้น

4. การจัดกลุ่มโดยอาศัยความใกล้ชิดกันหรือการวิเคราะห์ของตลาด ในการจัดกลุ่มหรือการวิเคราะห์ตลาด คือการตัดสินใจรวมสิ่งที่สามารถไปด้วยกันเข้าไว้ในกลุ่มเดียวกันตัวอย่างของการจัดกลุ่มโดยอาศัยความใกล้ชิดกันหรือการวิเคราะห์ตลาด เช่น การตัดสินใจว่าสิ่งใดบ้างที่จะไปอยู่ด้วยกันอย่างสม่ำเสมอในรถเข็นในซุปเปอร์มาร์เกต

5. การรวมตัว การรวมตัวคืองานที่ทำการรวมส่วนต่างๆ ในแต่ละส่วนที่ต่างชนิดกันให้อยู่รวมกันเป็นกลุ่มย่อยหรือคลัสเตอร์ (Clusters) โดยในแต่ละคลัสเตอร์อาจจะประกอบด้วยส่วนต่างๆที่ต่างชนิดกัน ซึ่งความแตกต่างของการรวมตัวจากการจัดหมวดหมู่คือ การรวมตัวจะไม่พึ่งพาอาศัยการกำหนดหมวดหมู่ล่วงหน้า และไม่ใช้ตัวอย่าง ข้อมูลจะรวมตัวกันบนพื้นฐานของความคล้ายในตัวเอง

6. การบรรยาย ในบางครั้งวัตถุประสงค์ของเดต้าไมนิ่ง คือ ต้องการอธิบายความสับสนของฐานข้อมูลในทางที่จะเพิ่มความเข้าใจในส่วนของประชากรผลิตภัณฑ์หรือขบวนการให้มากขึ้น เทคนิคเดต้าไมนิ่งส่วนใหญ่ต้องการเทรนนิ่งข้อมูลจำนวนมากที่ประกอบด้วยหลายๆ ตัวอย่างเพื่อจะสร้างกฎที่ใช้ในการจัดหมวดหมู่ กฎของความสัมพันธ์ คลัสเตอร์ การทำนายล่วงหน้า ดังนั้นชุดของข้อมูลขนาดเล็ก จะนำไปสู่ความไม่น่าไว้วางใจของผลสรุปที่ได้

ไม่มีเทคนิคใดเลยที่จะสามารถแก้ปัญหาของเดต้าไมนิ่งได้ทุกปัญหาดังนั้นความหลากหลายของเทคนิคจึงเป็นสิ่งที่จำเป็นในการไปสู่วิธีการแก้ปัญหาของเดต้าไมนิ่งได้ดีที่สุด

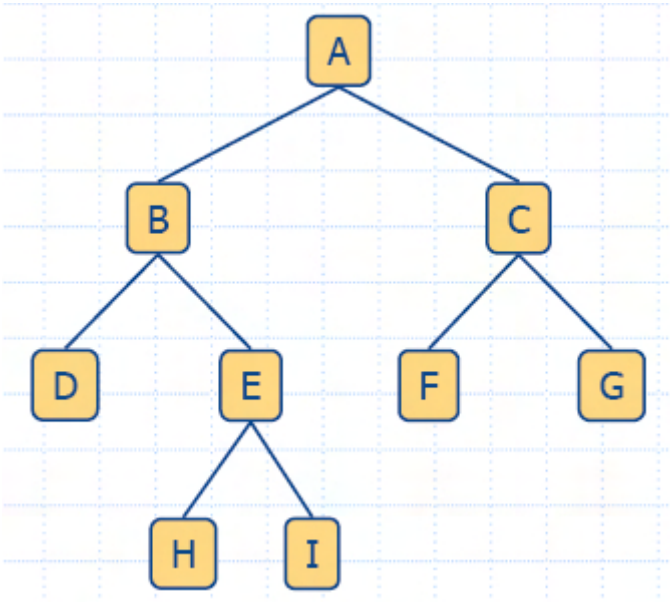
เทคนิคของเดต้าไมนิ่ง

การจำแนกประเภทข้อมูล (Data Classification)

กฤษณะ ไวยมัย , ชิดชนก ส่งศิริ และ ธนาวินท์ รักธรรมานนท์ (2544:135) กล่าวว่าการจำแนกประเภทของข้อมูลเป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ โดยจะนำข้อมูลส่วนหนึ่งมาสอนให้ระบบเรียนรู้ (Training Data) เพื่อจำแนกข้อมูลออกเป็นกลุ่มตามที่กำหนดไว้ ผลลัพธ์ที่ได้จากการเรียนรู้คือ โมเดลจำแนกประเภทข้อมูล (Classifier Model) และจะนำข้อมูลส่วนที่เหลือจากข้อมูลสอนระบบเป็นข้อมูลที่ใช้ทดสอบ (Testing Data) ซึ่งกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้จะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากโมเดลเพื่อทดสอบความถูกต้องและปรับปรุงโมเดลจนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจหลังจากนั้นเมื่อมีข้อมูลใหม่เข้ามา จะนำข้อมูลผ่านโมเดล โดยโมเดลจะสามารถทำนายกลุ่มของข้อมูลได้ ซึ่งโมเดลอาจแสดงในรูปของ โมเดลการพยากรณ์ข้อมูล (Prediction Model) เป็นกระบวนการสร้างโมเดลเพื่อทำนายหาค่าที่ต้องการจากข้อมูลที่มีอยู่ โดยมีกระบวนการสร้างโมเดลคล้ายกับการจำแนกประเภทข้อมูลแต่ต่างกันตรงที่การพยากรณ์ข้อมูลนี้เป็นการพยากรณ์หาค่าที่ต้องการออกมาเป็นตัวเลข ตัวอย่างเช่น หายอดขายของเดือนถัดไป จากข้อมูลการขายทั้งหมดที่ผ่านมา (Berry , Michel J.A. and Linnoff 2000:7) หรือ ทำนายเกรดเฉลี่ยของนักเรียนในปีการศึกษาหน้า จากข้อมูลการลงทะเบียนของนิสิตทั้งหมด (Han , Jiawei and Kamber 2000 : 19) เป็นต้น

ต้นไม้ตัดสินใจ (Decision Tree)

มีลักษณะคล้ายโครงสร้างต้นไม้ ซึ่งประกอบด้วยโหนดภายใน (Internal node) จะแสดงลักษณะ (Attribute) ของข้อมูล โดยที่จุดเริ่มต้นของต้นไม้เรียกว่าโหนดรากแต่ละกิ่งแสดงค่าของคุณลักษณะของแต่ละโหนด และ ลีฟโหนดแสดงกลุ่ม (Class) ซึ่งเป็นผลลัพธ์ที่สามารถแยกแยะได้



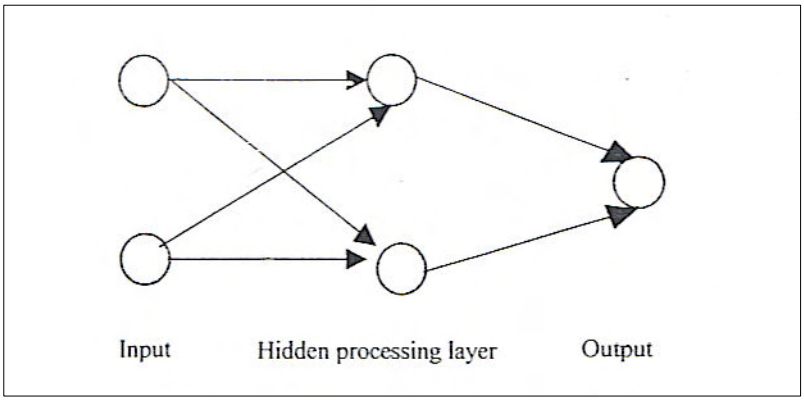
ภาพที่ 1 โครงสร้างต้นไม้ตัดสินใจ(Decision Tree)

กฤษณะ ไชยมัย , ชิดชนก ส่งศิริ และ ธนาวินท์ รักธรรมานนท์ (2544:136-137) ได้ใช้เทคนิคการจำแนกข้อมูลในรูปของต้นไม้ช่วยตัดสินใจมาประยุกต์ใช้ในการช่วยให้นักศึกษาเลือกสาขาวิชาที่เหมาะสม โดยนำข้อมูลของนิสิตที่เรียนดีมาสร้างโมเดลการจำแนกประเภทข้อมูลซึ่งแต่ละโหนดของต้นไม้จะเป็นลักษณะและผลการเรียนในรายวิชาต่าง ๆ ของนิสิต เพื่อจะทำนายว่าลักษณะของนิสิตแต่ละคนจะมีความคล้ายคลึงกับนิสิตที่เรียนดีในสาขาวิชาใดมากที่สุด แต่ผลที่ได้ยังมีความถูกต้องต่ำคือประมาณ 50% จึงได้สร้างโมเดลการจำแนกประเภทข้อมูลสำหรับแต่ละสาขา โดยพิจารณาว่านิสิตเหมาะสมกับสาขานั้นหรือไม่ ซึ่งนิสิตที่มีผลการเรียนที่ดีอาจจะมีหลายสาขาวิชาที่เหมาะสมให้เลือก นิสิตก็สามารถเลือกวิชาที่ดีที่สุดโดยพิจารณาจากสัดส่วนของคลาสปลายทางที่ประกอบด้วย GOOD และ BAD ซึ่งผลการทดสอบมีความถูกต้อง 84.58%

เครือข่ายใยประสาท (Neural Network)

เป็นเทคโนโลยีที่มีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ (Artificial Intelligence: AI) เพื่อใช้ในการคำนวณค่าฟังก์ชันจากกลุ่มข้อมูล วิธีการของเครือข่ายใยประสาทเป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบแล้วฝึก Train ให้ระบบได้รู้จักคิดแก้ปัญหาที่กว้างขึ้นได้ในโครงสร้างของเครือข่ายใยประสาทจะประกอบด้วยโหนด (node) สำหรับ Input-Output และการประมวลผล กระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ input layer, output layer และ hidden layers

การประมวลผลของเครือข่ายใยประสาทจะอาศัยการส่งการทำงานผ่านโหนดต่าง ๆ ใน layer เหล่านี้ ตัวอย่างโครงสร้างแบบใยประสาท



ภาพที่ 2 โครงสร้างเครือข่ายใยประสาท

ฟัซซีลอจิก (Fuzzy Logic)

(มหาวิทยาลัยสุโขทัยธรรมาธิราช 2545: 439) ฟัซซีลอจิก เป็นเทคนิคในทางตรรกศาสตร์หรือการคิดหาเหตุผลแบบหนึ่ง ซึ่งจะเป็นวิธีการในการจัดการกับกฎเกณฑ์ความรู้ที่มีความคลุมเครือของข้อมูลแฝงอยู่ด้วย Fuzzy Logic เป็นศาสตร์ที่พยายามเลียนแบบวิธีการคิดหาเหตุผลของมนุษย์ เป็นวิธีการที่ทำให้คอมพิวเตอร์สามารถทำงานหรือคิดหาเหตุผลได้ภายใต้เงื่อนไขหรือกฎเกณฑ์ที่มีความไม่แน่นอนของข้อมูลระดับหนึ่ง

หลักการสำคัญของ Fuzzy Logic คือการแทนรูปของกฎเกณฑ์ความรู้ในลักษณะ “เงื่อนไข ถ้า……..” ให้ โดยที่ความจริงของเหตุการณ์ที่เกี่ยวข้อง ไม่จำเป็นต้องมีเพียงค่า “จริง” หรือ “เท็จ” เหมือนที่ใช้กันในตรรกศาสตร์แบบปกติ (Classical Logic หรือ Boolean Logic) แต่ค่าความจริงของประเด็นที่สนใจใน Fuzzy Logic นี้ อาจจะมีได้หลายระดับหรือมีขั้นกว่าในระดับต่างๆ เช่นจะมีวิธีการในการเก็บความรู้ที่แทนความหมายของคำว่า “น้อย” “ค่อนข้างน้อย” “ปานกลาง” “ค่อนข้างมาก” “มากที่สุด” เป็นต้น

ตัวอย่างกฎเกณฑ์ความรู้ ที่มีค่าความจริงแบบคลุมเครือแฝงอยู่ อาจจะเป็นกฎเกณฑ์ความรู้ข้อหนึ่งที่นำไปใช้งานด้านการให้สินเชื่อ เช่น ถ้า จำนวนปีที่ทำงาน ค่อนข้างนาน ดังนั้น ความเสี่ยงในการให้สินเชื่อ ค่อนข้างน้อย

Cluster Analysis

กัลยา วานิชย์บัญชา (2546: 125) กล่าวว่า Cluster Analysis เป็นเทคนิคที่ใช้จำแนกหรือแบ่ง Case (คน สัตว์ สิ่งของ หรือ องค์กร ฯลฯ) หรือแบ่งตัวแปรออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป Case ที่อยู่กลุ่มเดียวกันจะมีลักษณะเหมือนกันหรือคล้ายกัน Case ที่อยู่ต่างกลุ่มจะมีลักษณะที่แตกต่างกัน ดังนั้นการพิจารณาเลือกลักษณะตัวแปรที่จะนำมาใช้แบ่งกลุ่ม Case จึงมีความสำคัญ นอกจากนั้น Case ใด Case หนึ่งจะอยู่ในกลุ่มหนึ่งเพียงกลุ่มเดียว

1. ใช้ศึกษาพฤติกรรมการบริโภคของกลุ่มผู้บริโภคที่อยู่ต่างกลุ่มกัน ทำให้สามารถวางกลยุทธ์ทางการตลาดได้อย่างมีประสิทธิภาพมากขึ้น ตัวแปรที่ใช้ในการแบ่งกลุ่ม เช่น อาชีพ อายุ รายได้

2. การแบ่งกลุ่มประเทศด้านสาธารณสุข ซึ่งตัวแปรที่ใช้เช่น จำนวนแพทย์ จำนวนเภสัชกร จำนวนพยาบาล จำนวนเตียง ประเทศที่มีระบบสาธารณสุขคล้ายกันจะอยู่ด้วยกัน

นอกจากนี้ยังมีเทคนิคการจัดกลุ่มอีกหลายวิธี เช่นแบบมีลำดับชั้น (Hierarchy) ซึ่งอาจใช้ Agglomerative hierarchical clustering คือการรวมกลุ่มจากแต่ละหน่วยที่มีอยู่ โดยจัดหน่วยที่คล้ายให้รวมกลุ่มเป็นกลุ่มๆ แล้วดูว่ากลุ่มใดคล้ายกันอีกก็รวมเป็นกลุ่มใหญ่ขึ้น ทำเช่นนี้จนรวมเป็น 1 กลุ่มใหญ่

การค้นหากฎความสัมพันธ์ (Association Rule Discovery)

การค้นหากฎความสัมพันธ์ คือ การค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อช่วยในการวิเคราะห์และตัดสินใจในธุรกิจ ตัวอย่างของการค้นหากฎความสัมพันธ์ เช่น การวิเคราะห์การซื้อสินค้าของลูกค้าเรียกว่า “MarketBasketAnalysis” (Han and Kamber 2001:225)

(กฤษณะ ไวยมัย, ชิดชนก ส่งศิริ, และธนาวินท์ รักธรรมานนท์ 2544:135) กล่าวว่า การค้นหากฎความสัมพันธ์ เป็นการค้นหาความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ที่มีอยู่ เพื่อนำไปใช้ในการวิเคราะห์หรือทำนายปรากฏการณ์ต่างๆ โดยเทคนิคนี้ใช้กันอย่างแพร่หลายในการขายสินค้าหรือการวิเคราะห์ข้อมูลที่เป็นTransaction เป้าหมายของการค้นหากฎความสัมพันธ์ คือ จะแยกและดึงสิ่งที่ซ่อนไว้ในฐานข้อมูลได้อย่างไร และจะหารายการ (Item) ใน Transaction เดียวกันได้อย่างไร ซึ่งจะสามารถบอกได้ว่ารายการใดที่มีแนวโน้มที่จะสามารถพบด้วยกันใน Transaction เดียวกัน การใช้เทคนิคการค้นหากฎความสัมพันธ์สามารถนำมาประยุกต์ใช้กับด้านต่างๆได้หลากหลาย เช่น ด้านการศึกษา กฤษณะ ไวยมัย, ชิดชนก ส่งศิริ, และธนาวินท์ รักธรรมานนท์ (2001:139-141) ได้นำมาประยุกต์ใช้ในการช่วยทำนายเกรดรายวิชาต่างๆในภาคเรียนต่อไปของนิสิตคณะวิศวกรรมศาสตร์ โดยหาความสัมพันธ์ของผลการเรียนในแต่ละวิชาที่ส่งผลต่อกัน ซึ่ง

ทำให้ได้ว่าวิชาใดบ้างที่มีผลต่อวิชาที่ต้องการจะทำนายเกรดล่วงหน้า โมเดลจะทำนายโดยอ้างอิงจากข้อมูลเดิมของนิสิตที่เคยเรียนมาและได้ผลการเรียนเช่นเดียวกับนิสิตคนนั้น

ด้านกฎหมาย กฤษณะ ไวยมัย และธีระวัฒน์ พงษ์ศิริปรีดา (2544 : 143-152) ได้ใช้เทคนิคการค้นหากฎความสัมพันธ์และเทคนิค Data Classification มาประยุกต์ใช้เพื่อการจัดสรรกฎหมายที่เหมาะสมกับการพิจารณาคดีความ โดยนำเทคนิค Data Classification มาสร้างตัวจำแนกข้อมูลจากกฎเกณฑ์ที่ได้จากเทคนิค การค้นหากฎความสัมพันธ์ ตัวจำแนกข้อมูลสามารถนำไปใช้ทำนายคดีความแต่ละคดีว่าควรใช้กฎหมายฉบับใดในการพิจารณา โดยใช้ข้อมูลคดีความของศาลฎีกา ซึ่งประกอบด้วยคดีอาญา และคดีแพ่ง และใช้การตัดคำด้วยพจนานุกรมภาษาไทยเพื่อแบ่งคดีความให้เป็นวลีสั้นๆ โดยใช้เทคนิค Suffix array และการหากฎเกณฑ์และชุดทดสอบจะต้องแบ่งเป็น 2 ชุด คือ ระดับกฎหมาย และระดับมาตรา ผลการทดลองได้ผลลัพธ์ในการทำนายที่ดีกว่าการใช้ Data Classification แบบทั่วไปซึ่งอยู่ในรูปของต้นไม้ช่วยตัดสินใจ (Decision Tree)

ด้านระบบการประกันสุขภาพ (Viveros, Marisa S., Nearhos, John P. and Rothman, Michale J 1996) ได้ใช้เทคนิคการค้นหากฎความสัมพันธ์และ Neural Segmentation มาประยุกต์ใช้กับระบบการประกันสุขภาพ ข้อมูลที่นำมาวิเคราะห์นั้นคือข้อมูลการเรียกใช้สิทธิ (Claim)ในการรักษา จำนวน 6,800,000 เรคอร์ด 120 แอทริบิวต์ และข้อมูลการรักษาพยาบาลของแพทย์ จำนวน 17,000 เรคอร์ด 105 แอทริบิวต์ ย้อนหลัง 5 ปี โดยที่เทคนิคการค้นหากฎความสัมพันธ์นั้นจะใช้ข้อมูลการเรียกใช้สิทธิในการรักษาเพื่อค้นหารูปแบบพฤติกรรมการรักษา โดยคัดเลือกแอทริบิวต์ที่สนใจคือการให้บริการการรักษา ผลจากการวิจัยพบว่ามีข้อสงสัยจากผลลัพธ์ที่ได้คือพบข้อผิดพลาดจากการเรียกใช้สิทธิในสัดส่วนที่มาก ทำให้เสียค่าใช้จ่ายมากถึง $550,000 สำหรับรายการนี้ในระยะเวลาระหว่าง 1 ปี และถ้ามีเหตุการณ์แบบนี้เกิดขึ้นอีกโดยที่ไม่สามารถตรวจพบได้ ก็จะเกิดการสูญเสียเป็นจำนวนมาก ส่วนเทคนิค Neural segmentation จะใช้ข้อมูลทั้ง 2 ฐานข้อมูล เพื่อจัดกลุ่มของการรักษาที่เกิดจากธรรมชาติหรือเกิดจากการปฏิบัติงาน โดยดูจากลักษณะการรักษา

Dick Ng'ambi (2002 : 101-109) ได้ใช้เทคนิคการค้นหากฎความสัมพันธ์ในการทำนายคำถามต่อไปของผู้ถาม โดยใช้ความสัมพันธ์และรูปแบบที่ซ่อนไว้ในรายการ FAQ การค้นพบความสัมพันธ์นี้สามารถใช้คาดเดาคำถามได้ข้อมูล FAQ เป็นข้อมูลของ ITCS (Information Technology and Computing Service) ซึ่งเป็น Help desk ของ University of Cape Town นอกจากนั้นยังสร้างตารางรายงาน Frequently Referenced Question (FRQ) และตารางรายการ Referenced Question (RRQ)

กระบวนการทำนายคำถามและการให้คำตอบกับผู้ถาม มีด้วยกัน 5 ขั้นตอน คือ

1. เลือกคำถามจากFAQและนำคำถามที่ถูกเลือกไปหาในตารางFRQ
2. ค้นหาคำถามที่ถูกถามในตารางRRQก็จะเจอคำถามต่อไปและนับความถี่ที่เจอของแต่ละคำถาม
3. หา Degree of Association
4. แสดงผลการทำนายด้วยความเชื่อมั่น (Confidence)
5. ทำนายคำถามถัดไปได้ ทำให้สามารถที่จะให้คำตอบกับผู้ถามได้ก่อนที่ผู้ถามจะถามคำถามถัดไป

Apriori Algorithm

เป็นขั้นตอนวิธีที่ใช้เทคนิคการสร้าง Frequent item sets ซึ่งเป็นที่รู้จักและมีอิทธิพลต่อการศึกษาในพัฒนาขั้นตอนวิธีอื่น ๆ ในการสร้าง Frequent item sets การทำงานของ Apriori จะมีลักษณะดังนี้ คือ มีการสร้าง Candidate item sets จากการ join กันของ Frequent item sets ในระดับก่อนหน้า และมีการ pruning ผลของการ join กัน ก่อน เพื่อหา Candidate item sets ที่แท้จริงจากนั้นจะทำการอ่านฐานข้อมูล เพื่อนับความถี่ ของ Candidate item sets เหล่านี้ เพื่อหา Frequent item sets (Candidate item sets ที่มีความถี่มากกว่าหรือเท่ากับความถี่ขั้นต่ำ) และจะทำการเช่นนี้ ซ้ำไปจนกว่าจะไม่สามารถสร้างCandidate item sets ได้แล้ว เมื่อได้ Frequent item sets ได้ครบถ้วนแล้วจะนำ Frequent item sets เหล่านี้ไปหา Association Rules ต่อไปโดยทำการสร้างกฎจาก Frequent item sets ทีละคู่ แล้วเลือกเฉพาะกฎที่มีค่าความเชื่อมั่นมากกว่าค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) ให้เป็นกฎที่ยอมรับ และทำซ้ำกระบวนการหา Association Rules ไปจนกว่าจะไม่สามารถสร้างกฎได้แล้ว จึงจบการทำงาน

อรทัศน์ สภานนท์ (2546 : 68) ใช้เทคนิคการสืบค้นกฎความสมพันธ์แบบหลายลำดับชั้นบนฐานข้อมูลขนาดใหญ่โดยเทคนิคที่นิยมใช้ในการสืบค้นปัจจุบันมีอยู่ 2 เทคนิค คือ Apriori Algorithm และ Closed Algorithm ทั้ง 2 เทคนิคนี้แตกต่างกันที่หลักการทำงาน โดย Apriori Algorithm มีการทำงานเป็นแบบล่างขึ้นบน จำนวนรอบการทำงานจึงขึ้นกับขนาดของกลุ่มข้อมูลที่เกิดขึ้นบ่อยที่สืบค้นได้ ในขณะที่ Closed Algorithm มีการทำงานเป็นแบบจำลอง จำนวนรอบการทำงานจึงไม่ขึ้นกับขนาดของกลุ่มข้อมูล ในช่วงเวลาที่ผ่านมาได้มีหลายงานวิจัยที่นำเอา Apriori Algorithm ไปประยุกต์ใช้กับการสืบค้นความรู้ในรูปแบบต่าง ๆ กฎความสัมพันธ์แบบหลายลำดับชั้นก็เป็นหนึ่งในความรู้ที่ได้มีผู้นำเสนอเทคนิคการสืบค้นโดยใช้ Apriori Algorithm เป็นหลักในการทำงาน แต่ยังไม่มีงานวิจัยใดที่นำเอา Closed Algorithm ที่มีงานวิจัยออกมาสนับสนุนแล้วว่าสามารถทำงานได้อย่างมีประสิทธิภาพมากกว่า ไปประยุกต์ใช้กับการสืบค้นกฎความสัมพันธ์แบบ

หลายลำดับชั้นงานวิจัยนี้จึงมีแนวคิดที่จะนำเอา Closed Algorithm มาใช้กับการสืบค้นกฎความสัมพันธ์แบบหลายลำดับชั้น เพื่อให้ได้วิธีการใหม่ที่มีประสิทธิภาพดีกว่าเดิม และเพื่อพิสูจน์ว่า Closed Algorithm สามารถนำประยุกต์ใช้กับการสืบค้นกฎความสัมพันธ์แบบหลายลำดับชั้นได้ นำเอาโครงสร้างข้อมูล Prefix Tree มาประยุกต์ใช้เพื่อเก็บข้อมูลระหว่างการสืบค้น เพื่อลดลงการทำงานในส่วนที่ซับซ้อนให้น้อยลงด้วย จากการทดสอบวิธีการใหม่กับข้อมูลที่สร้างจากโปรแกรม Synthetic Data ซึ่งเป็นโปรแกรมสร้างข้อมูลทดสอบที่เป็นที่ยอมรับในงานวิจัยด้านการทำเหมืองข้อมูล พบว่าวิธีการใหม่สามารถสืบคืนกฎความสัมพันธ์แบบหลายลำดับชั้นได้อย่างมีประสิทธิภาพ ใช้เวลาการทำงานน้อยกว่าวิธีการเดิม และมีอัตราการเพิ่มของเวลาที่ใช้ในการทำงานเมื่อรูปแบบของข้อมูลที่นำมาใช้ทดสอบเปลี่ยนแปลงไปน้อยกว่าวิธีการเดิมอีกด้วย

การแนะแนว

การแนะแนว ตรงกับภาษาอังกฤษว่า Guidance หมายถึงการชี้แนวทางหรือการชี้ช่องทางนั้นเองและนักแนะแนวหลายท่านได้ให้ความหมายของการแนะแนวไว้ในลักษณะต่างๆ ดังนี้

พรอกเตอร์ (Proctor) นักการศึกษาอเมริกันได้ให้ความหมายของคำว่า “การแนะแนว” คือบริการที่จัดขึ้นเพื่อช่วยเหลือเด็กทั้งที่เป็นรายบุคคลและเป็นกลุ่มที่ยังศึกษาอยู่และออกไปแล้ว เพื่อให้เด็กรู้จักการตัดสินใจในการเลือกศึกษาต่อ

บุหงา วิชระศักดิ์มงคล (2545 : 175 -178) ขบวนการที่ช่วยให้นักเรียนรู้จักตนเองและรู้ถึงแนวทางที่จะใช้ความสามารถ ความสนใจ และความถนัดของตนให้เป็นประโยชน์แก่ตนเองและสังคมให้มากที่สุดและช่วยให้สามารถเผชิญความจริงได้อย่างกล้าหาญ สามารถใช้ตัดสินใจเลือกและวางแนวชีวิตในอนาคตของตนได้อย่างฉลาดและถูกต้องในปัจจุบัน หลักสูตรของโรงเรียนมัธยมศึกษาเป็นหลักสูตรแบบกว้าง เปิดโอกาสให้นักเรียนได้เลือกเรียนตามถนัด ความสามารถและความสนใจ ทั้งในปัจจุบันนี้มีปัญหาต่าง ๆ เกิดขึ้นมากมาย บริการแนะแนวจึงเข้ามามีบทบาทเพื่อช่วยให้นักเรียนรู้จักแก้ปัญหาทั้งปรับตนให้เข้ากับสภาพแวดล้อมได้อย่างเหมาะสม

ปิยะนุสรณ์ ชนะชาญชัย (2545 : 38-39) การแนะแนวการศึกษา เป็นกระบวนการให้ความช่วยเหลือนักเรียนให้ได้รับความสำเร็จทางการศึกษาตามความสามารถ ให้รู้จักเทคนิควิธี การเรียนที่ถูกต้อง สามารถปรับตัวเข้ากับระบบการเรียนการสอนของโรงเรียน หลักสูตร และวิชาที่เรียนได้ รู้จักแบ่งเวลาในการเรียน การศึกษาค้นคว้าเพิ่มเติม การใช้ห้องสมุด การเขียนรายงาน ตลอดจนการแก้ปัญหาต่างๆ เกี่ยวกับการเรียน

เปล่งศักดิ์ ชาระ (2545 : 143-155) ได้ศึกษาเปรียบเทียบการเลือกเรียนต่อสายอาชีวศึกษา ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดกรมสามัญศึกษา จังหวัดชลบุรี กลุ่มตัวอย่างที่ใช้ในการศึกษา ได้แก่ นักเรียนที่กำลังศึกษาในชั้นมัธยมศึกษาปีที่ 3 สังกัดกรมสามัญศึกษา จังหวัดชลบุรี ในปีการศึกษา 2545 ที่เลือกเรียนต่อสายอาชีวศึกษา จำนวน 335 คน เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นแบบสอบถาม และสถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ ค่าร้อยละ คะแนนเฉลี่ยความเบี่ยงเบนมาตรฐานการทดสอบค่าที่และการวิเคราะห์ความแปรปรวนทางเดียวผลการวิจัยปรากฏดังนี้

ผลการเปรียบเทียบการเลือกเรียนต่อสายอาชีวศึกษาของนักเรียนมัธยมศึกษาปีที่3 สังกัดกรมสามัญศึกษาจังหวัดชลบุรีปีการศึกษา 2545 จำแนกตามด้านสถานภาพของนักเรียนพบว่าด้านแรงจูงใจใฝ่สัมฤทธิ์ และด้านการคล้อยตามกลุ่มเพื่อนจำแนกตามเพศแตกต่างกันอย่างมีนัยสำคัญทางสถิติ($p<.05$)การรับรู้ข่าวสารทางการศึกษาโดยรวมแตกต่างกันอย่างมีนัยสำคัญเมื่อพิจารณาเป็นรายด้านพบว่าทุกด้านแตกต่างกันอย่างไม่มีนัยสำคัญสถานภาพสมรสของผู้ปกครอง โดยรวมแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อพิจารณาเป็นรายด้าน พบว่า ด้านเจตคติต่อการเรียนแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p<.05$) ส่วนด้านเจตคติต่อการเรียนด้านแรงจูงใจใฝ่สัมฤทธิ์ และด้านการคล้อยตามกลุ่มเพื่อนไม่แตกต่างกันตามรายได้ของผู้ปกครอง โดยรวมแตกต่างกันอย่างมีสำคัญทางสถิติ($p<.05$)และเมื่อพิจารณาเป็นรายด้านพบว่าด้านค่านิยมทางการศึกษาของผู้ปกครองแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p<.05$) ส่วนด้านเจตคติต่อการเรียน ด้านแรงจูงใจใฝ่สัมฤทธิ์ และด้านการคล้อยตามเพื่อนไม่แตกต่างกันอาชีพของผู้ปกครอง โดยร่วมแตกต่างกันอย่างไม่มีนัยสำคัญ และเมื่อพิจารณาเป็นรายด้าน พบว่า ด้านเจตคติต่อการเรียนแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p<.05$) ส่วนด้านค่านิยมทางการศึกษาของผู้ปกครอง ด้านแรงจูงใจใฝ่สัมฤทธิ์ และด้านการคล้อยตามกลุ่มเพื่อนไม่แตกต่างกันระดับการศึกษาของผู้ปกครอง โดยรวมแตกต่างกันอย่างไม่มีนัยสำคัญ และเมื่อพิจารณาเป็นรายด้าน พบว่า ด้านเจตคติต่อการเรียนแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p<.05$) ส่วนทางค่านิยมทางการศึกษาของผู้ปกครอง ด้านแรงจูงใจใฝ่สัมฤทธิ์ และด้านการคล้อยตามกลุ่มเพื่อนไม่แตกต่างกัน