บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษาและการแสดงผลข้อมูลบนเว็ปไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลเพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

2.2 ทฤษฎี

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

คือ การวิเคราะห์ข้อมูลมุ่งเน้นไปที่วิธีการเชิงกลยุทธ์ในการรับข้อมูลดิบการขุดเพื่อความเข้าใจที่เกี่ยวข้องกับเป้าหมายหลักของธุรกิจและเจาะลึกลงไปในข้อมูลนี้เพื่อแปลงตัวชี้วัดข้อเท็จจริงและตัวเลขเป็นโครงการริเริ่มที่เป็นประโยชน์ต่อการปรับปรุง มีวิธีการต่าง ๆ สำหรับการวิเคราะห์ข้อมูลส่วนใหญ่ขึ้นอยู่กับสองประเด็นหลัก วิธีการวิเคราะห์ข้อมูลเชิงปริมาณและวิธีการวิเคราะห์ข้อมูลในการวิจัยเชิงคุณภาพ การนำข้อมูลจากหลายๆ Source มาประมวลผลวิเคราะห์ด้วยคณิตศาสตร์และสถิติ เพื่อให้ได้ insight ช่วยให้เราตัดสินใจและ Action อะไรบางอย่างเพื่อให้ปรับปรุงกระบวนการทำงาน การตลาด การออกแบบสินค้า เพื่อให้ได้ผลการทำธุรกิจที่ดีขึ้น Data Analytics เป็นส่วนที่สำคัญในการทำธุรกิจในปัจจุบัน คือ การเพิ่มปริมาณของข้อมูลอย่างมหาศาล จากการเปลี่ยนพฤติกรรมของผู้บริโภคที่เปลี่ยนมาใช้งานออนไลน์มากขึ้น ซึ่งทำให้เกิดข้อมูลประเภท Unstructured data ที่ไม่ได้มีการจัดเก็บเป็นรูปแบบชัดเจน ประกอบกับการพัฒนาของเทคโนโลยี และราคาของ Analytics Solution ที่สามารถจับต้องได้มากขึ้น ไม่ว่าจะเป็น Big Data, Business Intelligence เหล่านี้มาประกอบกัน ทำให้หลายธุรกิจตื่นตัวและนำ Data ไปใช้ในการบริหารจัดการภายในองค์กรมากขึ้น โดย Data Analytics นั้นจะแบ่งเป็น 5 ระดับได้แก่

– Descriptive Analytic จะเป็นการวิเคราะห์ข้อมูลเพื่อตอบคำถามที่ว่า ที่ผ่านมาเกิดอะไรขึ้นบ้าง? เราจะคุ้นเคยกับ Analytics ประเภทนี้ค่อนข้างมาก เป็นการทำ Report เช่น การทำรายงานกำไรขาดทุน รายงานด้านบัญชี เพื่อบอกเราว่ามีอะไรเกิดขึ้นบ้าง

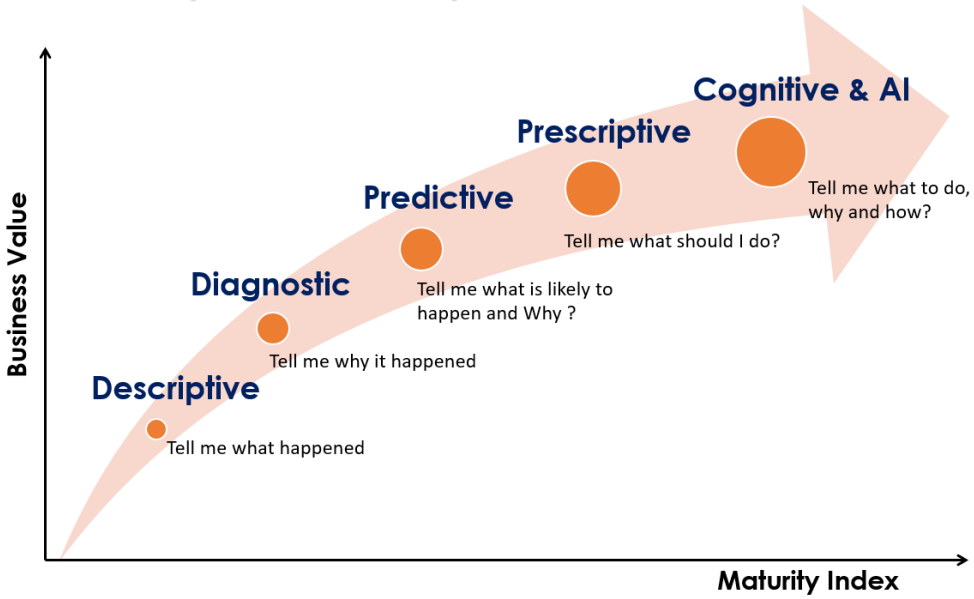
– Diagnostic Analytic เป็นการหาเหตุผลว่าไอ้ที่มันเกิดขึ้นแบบนี้เพราะอะไร (Tell me why it happened) ที่กำไรของเราดีขึ้นเนี่ยเป็นเพราะอะไร เพราะลูกค้าเพิ่มขึ้น หรือว่าเป็นเพราะเราลดต้นทุน ก็เริ่มมีการนำ Business Intelligence เข้ามาใช้งาน เริ่มใช้ Interactive Dashboard สามารถ Drill–down เพื่อให้มีมุมมองของข้อมูลได้หลากหลายมากขึ้น

– Prescriptive Analytic เป็นการวิเคราะห์ข้อมูลที่ไม่ใช่แค่พยากรณ์ว่าจะเกิดอะไรขึ้น แต่จะช่วยบอกเราว่า เราควรทำอะไรดี (Tell me what should I do?) โดยจะบอกด้วยว่าผลลัพธ์ของแต่ละสิ่งที่แนะนำมานั้นมันเป็นอย่างไร ตัวอย่างการใช้งานใน Level นี้คือการใช้ Google Map ในการนำทางนั่นเอง ที่เค้าจะบอกเส้นทางที่สามารถไปถึงจุดหมาย และบอกด้วยว่าเราจะใช้เวลาในการเดินทางในแต่ละตัวเลือกใช้เวลาเท่าไหร่

– Prescriptive Analytic เป็นการวิเคราะห์ข้อมูลที่ไม่ใช่แค่พยากรณ์ว่าจะเกิดอะไรขึ้น แต่จะช่วยบอกเราว่า เราควรทำอะไรดี (Tell me what should I do?) โดยจะบอกด้วยว่าผลลัพธ์ของแต่ละสิ่งที่แนะนำมานั้นมันเป็นอย่างไร ตัวอย่างการใช้งานใน Level นี้คือการใช้ Google Map ในการนำทางนั่นเอง ที่เค้าจะบอกเส้นทางที่สามารถไปถึงจุดหมาย และบอกด้วยว่าเราจะใช้เวลาในการเดินทางในแต่ละตัวเลือกใช้เวลาเท่าไหร่

– Cognitive & AI (Cognitive & Artificial Intelligence) เป็นการนำ AI เทคโนโลยีมาช่วยเพื่อให้วิเคราะห์และบอกเราได้ว่า เราควรจะทำอะไร สาเหตุ และวิธีการทำ เป็นขั้นสูงสุดของ Analytics Maturity ในปัจจุบัน ตัวอย่างการใช้งานคือการวิเคราะห์ข้อมูลเพื่อการสร้าง Self–driving car หรือรถยนต์ไร้คนขับ รวมไปจนถึง Apple Siri, Google Assistant ที่เป็น Companion App นั่นเอง

รูปที่ 2.1 ระดับการวิเคราะห์ข้อมูล



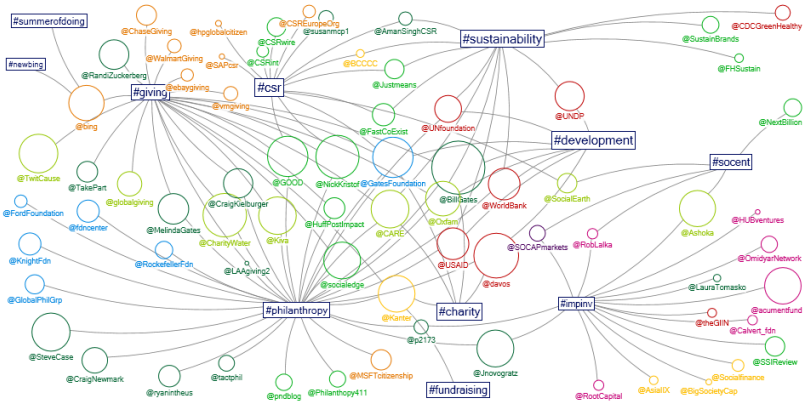
จาก Analytics Maturity จะเห็นได้ว่า ยิ่งเรานำ Data Analytics ไปพัฒนาใช้ในแต่ละ Level มูลค่าของธุรกิจหรือ Business Value จะยิ่งเพิ่มสูงมากยิ่งขึ้น ตอนนี้เราเข้าใจถึงประเภทของ Data Analytics แล้ว ลำดับต่อไปคือการนำไปประยุกต์ใช้ในแต่ละขั้นตอนของการทำธุรกิจเพื่อให้เกิด Impact และเพิ่มมูลค่าให้กับธุรกิจต่อไป

2.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization เป็นเครื่องมือสำคัญที่ช่วยให้เห็นภาพของข้อมูลเชิงลึกที่ผ่านการเก็บข้อมูล การจัดการข้อมูล การวิเคราะห์ข้อมูล แล้วรวบรวมได้เป็นชุดข้อมูล การนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพในความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณ ให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ หากต้องการประสบความสำเร็จในโลกธุรกิจที่มีการแข่งขันสูง ความสามารถในการตัดสินใจที่ถูกต้องในเวลาที่เหมาะสม โดยอาศัยข้อมูลเชิงลึกที่รัดกุมเป็นสิ่งสำคัญ ปัจจุบันองค์กรธุรกิจส่วนใหญ่ในประเทศไทย เริ่มใช้ประโยชน์จากการทำ Data Visualization เพราะว่าเครื่องมือนี้เป็นกุญแจสำคัญในการแสดงผลของข้อมูลให้เข้าใจง่าย ซึ่งจะช่วยให้ทุกแผนกในฝ่ายงานได้ใช้ประโยชน์ ในการหาแนวโน้มและความสัมพันธ์เชื่อมโยง อีกทั้งทำให้ทราบข้อมูลได้เร็วขึ้น ส่งผลให้การทำงานขององค์กรมีประสิทธิภาพเพิ่มขึ้น จนดำเนินการถึงการวางกลยุทธ์ภายในองค์กร และสามารถนำไปใช้กับการทำเว็บไซต์ได้ Data Visualization ช่วยให้การสื่อสารข้อมูล มีความรวดเร็วและเข้าใจง่ายมากขึ้น พื้นฐานของการสร้าง Data Visualization คือ การ Mapping ส่วนข้อมูล กับส่วนของ Graphic เข้าด้วยกัน ซึ่งตอนนี้มีโปรแกรมสำเร็จรูปในการสร้าง Data visualization หลากหลายโปรแกรม มีฟังก์ชันการใช้งานที่เข้าใจง่าย เช่น

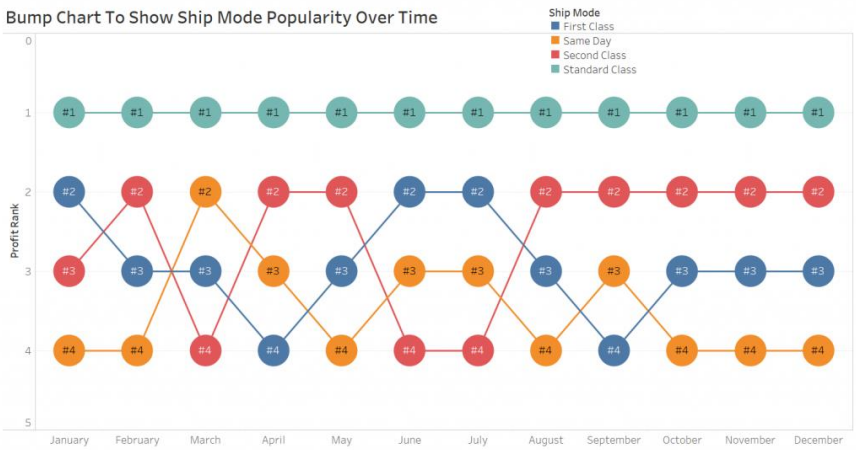
การสร้างฟิลเตอร์ การออกแบบเพื่อให้งานการวิเคราะห์ข้อมูลมีความยืดหยุ่น เป็นต้น ตัวอย่างรูปแบบ Data Visualization ที่แสดงผลให้อย่างสนใจ มีดังนี้

1) กราฟเครือข่าย (Network Graph) ใช้แสดงความเชื่อมโยงของเครือข่ายหรือความสัมพันธ์ในกลุ่ม



รูปที่ 2.2 กราฟเครือข่าย

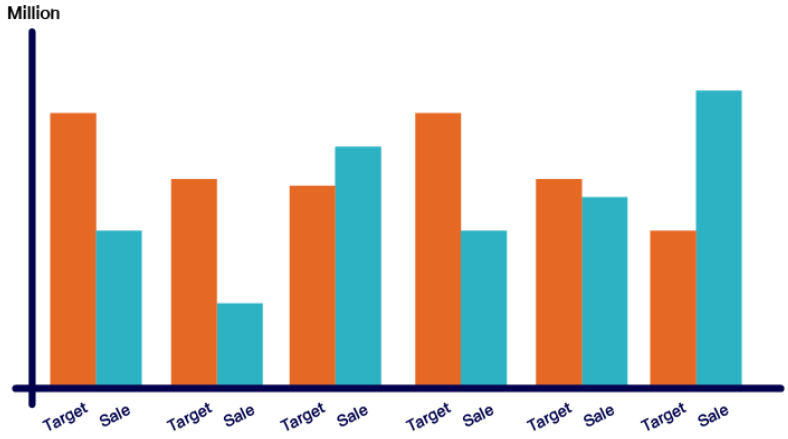
2) RANKING (Bump Chart) ใช้แสดงการเปลี่ยนแปลงลำดับของข้อมูลในหลายช่วงเวลา โดยใช้สีเพื่อช่วยให้อ่านข้อมูลได้ง่ายขึ้น



รูปที่ 2.3 การจัดอันดับ

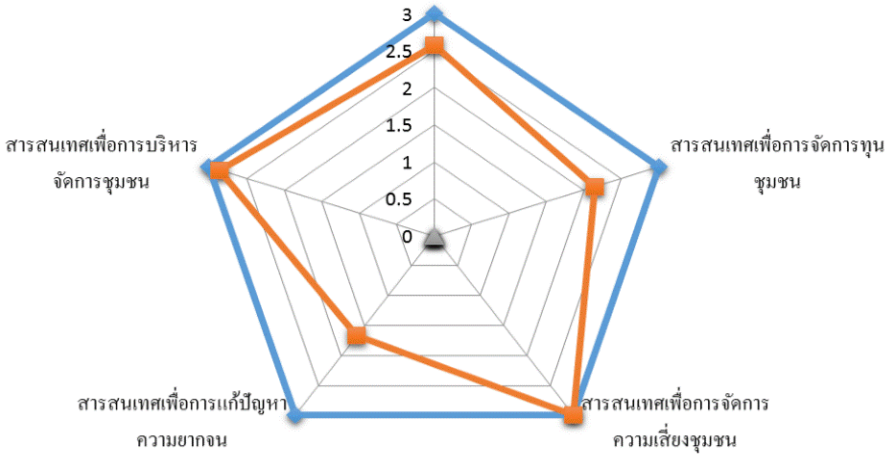
3) แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้ง ที่นิยมแสดงออกมาในรูปแท่งสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับใช้การเปรียบเทียบจำนวนของข้อมูลในแต่ละชุด เช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริงเปรียบเทียบกับเป้าหมายยอดขายที่ตั้งไว้ เป็นต้น ซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท • แผนภูมิแท่งแบบจัดกลุ่ม แผนภูมิแท่งแบบจัดกลุ่ม เหมาะสมกับการนำเสนอข้อมูลที่มีข้อมูลย่อย ๆ อยู่ภายใต้ข้อมูลใหญ่ เป็นการเน้นให้เห็นข้อมูลย่อยนั้น ๆ • แผนภูมิแท่งแบบวางซ้อนกัน แผนภูมิแท่งแบบวางซ้อนกัน เหมาะสมกับการนำเสนอข้อมูล

ที่ต้องการให้เห็นข้อมูลย่อยในแต่ละข้อมูลใหญ่ และยังแสดงให้เห็นสัดส่วนของข้อมูลย่อยต่าง ๆ เหล่านั้นได้ด้วย ใช้แผนภูมินี้เมื่อมีชุดข้อมูล หลายชุดและต้องการเน้นผลรวมทั้งหมด



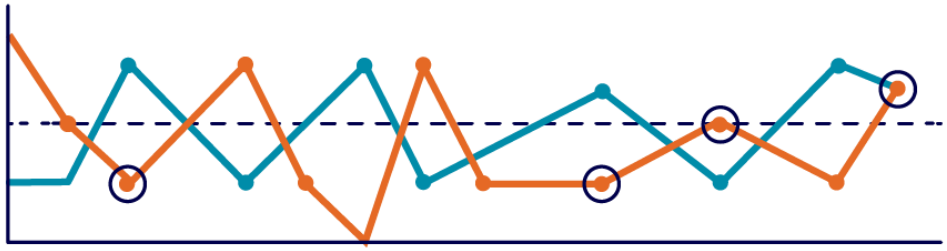
รูปที่ 2.4 แผนภูมิแท่ง

4) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลม จำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความต่อเนื่องของข้อมูล แต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อน จุดแข็งของข้อมูลเช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อนจุดแข็งของแต่ละคน เป็นต้น



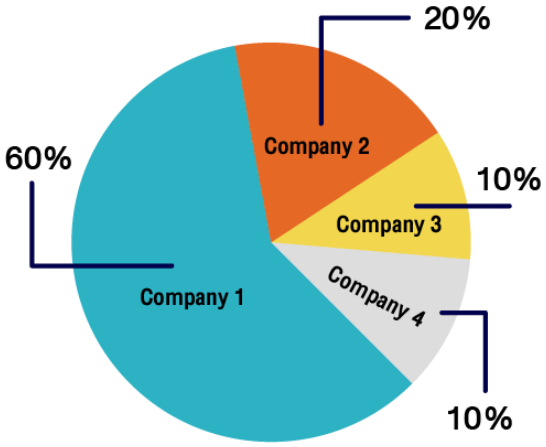
รูปที่ 2.5 แผนภูมิเรดาร์ (Radar Charts)

5) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั่นเองแผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วง
ใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดูแนวโน้ม เป็นต้น



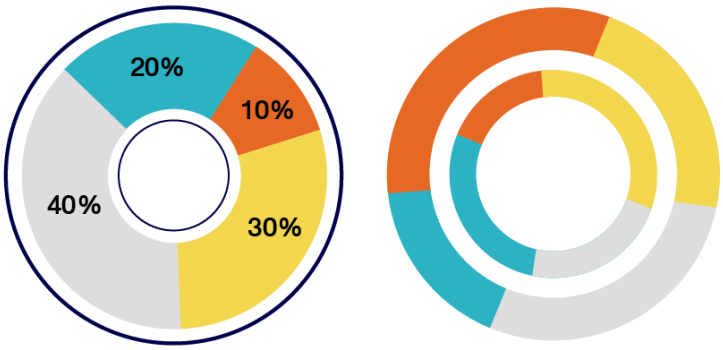
รูปที่ 2.6 แผนภูมิเส้น

6) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยาก เพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น



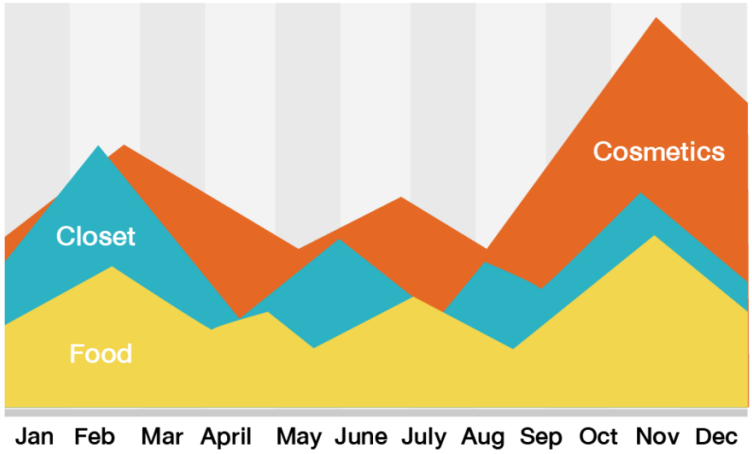
รูปที่ 2.7 แผนภูมิวงกลม

7) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลายๆ ชั้นนั่นเอง



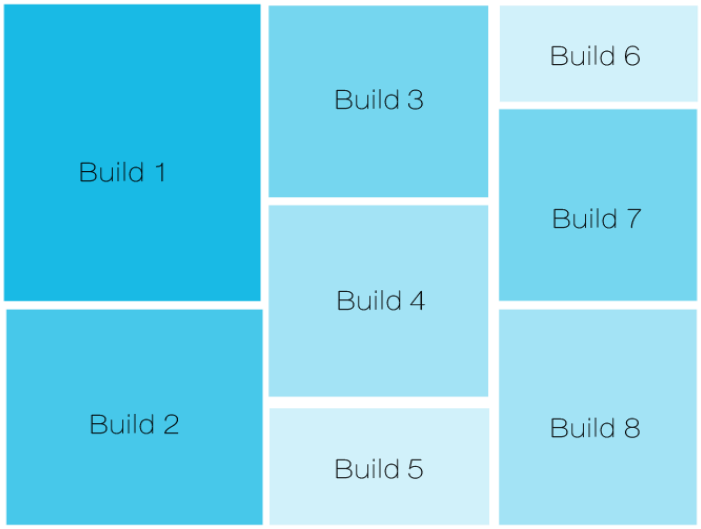
รูปที่ 2.8 แผนภูมิโดนัท

8) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้าย แผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่เดือน ตามหมวดหมู่ต่าง ๆ ไล่ไปเครื่องสำอางค์ เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



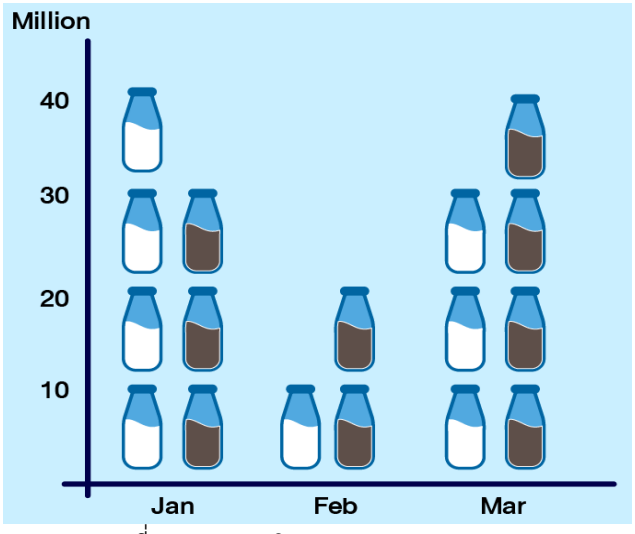
รูปที่ 2.9 แผนภูมิพื้นที่

9) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดงผลได้ในแบบลำดับชั้นเหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้



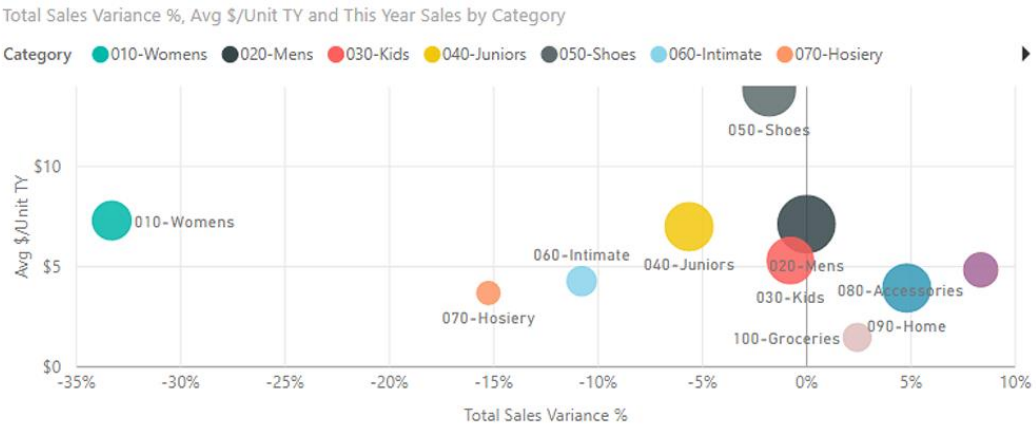
รูปที่ 2.10 แผนภูมิต้นไม้ (Tree Maps)

10) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอน และแกนตั้ง แต่เลือกใช้รูปภาพ หรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ เช่น การแสดงผลจำนวนของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละเดือนซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะนี้ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลง



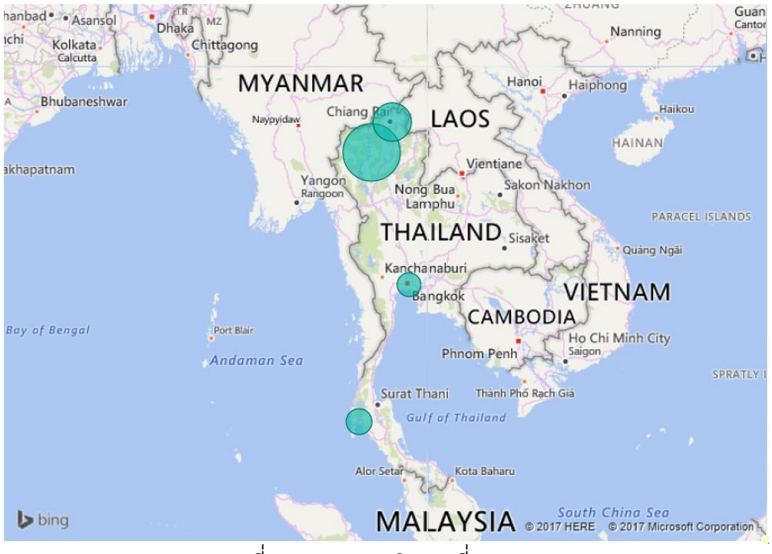
รูปที่ 2.11 แผนภูมิรูปภาพ (Picture Graph)

11) แผนภูมิฟอง (Bubble charts) กราฟประเภทจุดโดยเป็นรูปแบบของแกน X และแกน Y โดยหากเป็น Bubble Chart จะสามารถกำหนดในเรื่องของขนาด (Size) ให้กับจุดได้อีกด้วยและ ยังสามารถกำหนดในเรื่องของความสดของสี (Saturation) ได้อีกด้วยเช่น กราฟแสดงยอดขายและจำนวนหน่วย พร้อมการเปลี่ยนแปลงยอดขายของสินค้าในแต่ละกลุ่ม



รูปที่ 2.12 แผนภูมิฟอง (Bubble charts)

12) แผนภูมิแผนที่ (Maps) สำหรับการนำเสนอข้อมูลที่ต้องอ้างถึงตำแหน่งของประเทศ (Country), ชื่อเมือง (City),จังหวัด (Province) หรือ พิกัด Longitude, Latitude โดยสามารถนำเสนอในรูปแบบของแผนที่ได้ โดย Engine ที่ใช้ในการแสดงผล คือ Bing Map



รูปที่ 2.13 แผนภูมิแผนที่ (Maps)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

หมายถึงข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยากหรือเป็นไปไม่ได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมากเพื่อทำการวิเคราะห์มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วงต้นปีค.ศ. 2000 เมื่อดั๊ก ลานีย์ นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันในขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วย3Vs ดังนี้

VOLUME ปริมาณข้อมูลที่มากเป็นปัจจัยที่มีความสำคัญ ในปริมาณข้อมูลที่มากมายมหาศาลนั้นที่เราจะต้องประมวลผลนั้นเป็นข้อมูลที่ไม่มีโครงสร้าง มีความหนาแน่นของข้อมูลต่ำ และข้อมูลพวกนี้อาจเป็นข้อมูลที่ไม่ทราบค่า เช่น ฟีดข้อมูลของเฟสบุ๊ค ทวีทเตอร์ การคลิ๊กบนเว็บไซต์หรืออุปกรณ์แอพพลิเคชั่นต่าง ๆ หรืออุปกรณ์ที่มีเซนเซอร์ บางองค์การอาจมีข้อมูลให้ประมวลผลเป็นสิบๆเทราไบต์ หรือบางองกรค์อาจมีเป็น ร้อย ๆเพตะไบต์

VELOCITY คือความเร็วของการรรับข้อมูลหรืออาจเป็นการกระทำใด ๆ โดยปกติก็เป็นความเร็วสูงสุดที่ทำการสตีมข้อมูลลงในหน่วยความจำโดยตรงกับการบันทึกข้อมูลลงบนฮาร์ดดิสก์ เครื่องมือหรืออุปกรณ์ที่ใช้กับอินเตอร์เน็ตในสมัยนี้ก็เป็นการทำงานแบบเรียลไทม์หรือเกือบๆจะเรียลไทม์ ซึ่งจะต้องมีการประมวลผลแบบเรียลไทม์

VARIETY คือความหลากหลายของชนิดข้อมูล ข้อมูลในสมัยก่อนมักเป็นพวกข้อมูลที่เป็นโครงสร้างและมีความพอดีกับฐานข้อมูลเชิงสัมพันธ์ ปัจจุบันข้อมูลมีขนาดใหญ่ขึ้นและเป็นข้อมูลแบบไม่มีโครงสร้างหรือกึ่งโครงสร้าง เช่น ข้อมูลแบบตัวอักษร ข้อมูลภาพ ข้อมูลเสียง ซึ่งต้องการการประมวลผลเพิ่มเติม เพื่อที่จะแปลความหมาย และหารายละเอียดคำอธิบายของข้อมูล (meta data)

Big Data ในช่วง 2–3 ปีหลังมานี้ได้เพิ่มมาอีก 2 Vs คือ Value และ veracity ซึ่งคุณค่าและความจริง ซึ่งข้อมูลมันมีค่าอยู่ในตัวของมันเอง แต่มันจะไม่มีประโยชน์เลยถ้าเราค้นไม่พบคุณค่าของมัน และความจริงของข้อมูลและความน่าเชื่อถือว่าเราจะเชื่อถือได้มากแค่ไหน ก็มีความสำคัญเท่าเทียมกันทั้งคุณค่าและความจริงของข้อมูล

ในปัจจุบัน Big Data ได้กลายเป็นทุนหรือทรัพย์สินไปแล้ว ในบริษัทเทคโนโลยีขนาดใหญ่บางบริษัทมูลค่าของบริษัทเกิดมาจากข้อมูลของเขา และในขณะนี้เขาก็วิเคราะห์และประมวลผลเพื่อเพิ่มมูลค่าของข้อมูลให้สูงขึ้นไปอีก และด้วยความก้าวหน้าทางเทคโนโลยีในปัจจุบันส่งเสริมให้ราคาของอุปกรณ์การเก็บข้อมูลและคอมพิวเตอร์ลดลงแบบก้าวกระโดด ทำให้การเก็บข้อมูลง่ายและมีราคาถูก การเก็บข้อมูลและการเข้าถึงข้อมูลขนาดใหญ่สามารถทำได้ง่ายๆและมีราคาถูก ทำให้การตัดสินใจด้านธุรกิจมีความแม่นยำและถูกต้องมากขึ้น

การค้นหามูลค่าของข้อมูล Big Data มันไม่ใช่แค่การวิเคราะห์ธรรมดาเท่านั้น แต่มันต้องมีกระบวนการขั้นตอนทั้งหมด ซึ่งมีตั้งแต่การวิเคราะห์เชิงลึก ความต้องการของธุรกิจ ความสามารถในการถามข้อมูลที่ถูกต้องจากผู้บริหาร การจดจำรูปแบบ การให้ข้อมูลสำหรับสมมุติฐานต่าง ๆ และการทำนายพฤติกรรม เป็นต้น

การวิเคราะห์ Big Data ช่วยให้องค์กรควบคุมข้อมูลของพวกเขาและใช้เพื่อระบุโอกาสใหม่ ๆ ในทางกลับกันนำไปสู่การเคลื่อนไหวทางธุรกิจที่ชาญฉลาดเพื่อการดำเนินงานที่มีประสิทธิภาพมากขึ้น ผลกำไรที่สูงขึ้นและลูกค้าที่มีความสุขมากขึ้น ในรายงาน Big Data ใน บริษัท ขนาดใหญ่โดยผู้อำนวยการฝ่ายวิจัยของ Tom Davenport ให้สัมภาษณ์ว่า มากกว่า 50 ธุรกิจใช้และทำความเข้าใจว่าพวกเขาใช้ Big Data อย่างไร และพบว่าสามารถช่วยเหลือธุรกิจได้ดังต่อไปนี้

- ลดต้นทุน : เทคโนโลยีข้อมูลขนาดใหญ่เช่น Hadoop และการวิเคราะห์บนคลาวด์นำมาซึ่งความได้เปรียบด้านต้นทุนอย่างมีนัยสำคัญเมื่อพูดถึงการจัดเก็บข้อมูลจำนวนมากรวมทั้งสามารถระบุวิธีการทำธุรกิจที่มีประสิทธิภาพมากขึ้น
- เร็วกว่าและตัดสินใจดีกว่า : ด้วยความเร็วของ Hadoop และการวิเคราะห์ในหน่วยความจำรวมกับความสามารถในการวิเคราะห์แหล่งข้อมูลใหม่ๆ ของธุรกิจจะสามารถสร้างข้อมูลได้ทันทีและสามารถดำเนินการต่อได้ทันทีจากการวิเคราะห์นั้น ๆ
- ผลิตภัณฑ์และบริการใหม่ : ด้วยความสามารถในการวัดความต้องการและความพึงพอใจของลูกค้าผ่านการวิเคราะห์นำมาซึ่งสิ่งที่ลูกค้าต้องการ ซึ่งยกตัวอย่างโดย ดาเวนพอร์ท สามารถชี้ให้เห็นว่าด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ของบริษัทต่าง ๆ จะสามารถสร้างผลิตภัณฑ์ใหม่ ๆ เพื่อตอบสนองความต้องการของลูกค้าได้เป็นอย่างดี

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

Data Mining (เหมืองข้อมูล)

คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

- การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง
- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย
- การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
- การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

ทำไมจึงต้องมี Data Mining

ข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลหากเก็บไว้เฉย ๆ ก็จะไม่เกิดประโยชน์ดังนั้นจึงต้องมีการสกัดสารสนเทศหรือการคัดเลือกข้อมูลออกมาใช้งานส่วนที่เราต้องการ
ในอดีตเราได้ใช้คนเป็นผู้สืบค้นข้อมูลต่างๆ ในฐานข้อมูลซึ่งผู้สืบค้นจะทำการสร้างเงื่อนไขขึ้นมาตามภูมิปัญญาของผู้สืบค้น
ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูลเดียวอาจไม่ให้ความรู้เพียงพอและลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูงและมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นที่จะต้องรวบรวมฐานข้อมูลหลาย ๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่า “ คลังข้อมูล” (Data Warehouse) ดังนั้นเราจึงจำเป็นต้องใช้ Data Mining ในการดึงข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะนำข้อมูลนั่นมาใช้งานให้เกิดประโยชน์สูงที่สุด

วัตถุประสงค์ในการใช้ Data Mining

- เพื่อการค้นพบองค์ความรู้ใหม่ในฐานข้อมูล (Knowledge discovery in databases)
- เพื่อการสกัดองค์ความรู้ที่ซ่อนเร้นอยู่ (Knowledge extraction)
- เพื่อจัดการกับข้อมูลในอดีต (Data archeology)
- เพื่อสำรวจข้อมูล (Data exploration)
- เพื่อค้นหา Pattern ของข้อมูลที่ซ่อนอยู่ (Data pattern processing)
- เพื่อใช้ขุดเจาะข้อมูล (Data dredging)
- เพื่อเก็บเกี่ยวผลประโยชน์ให้ได้มาซึ่งสารสนเทศที่มีประโยชน์

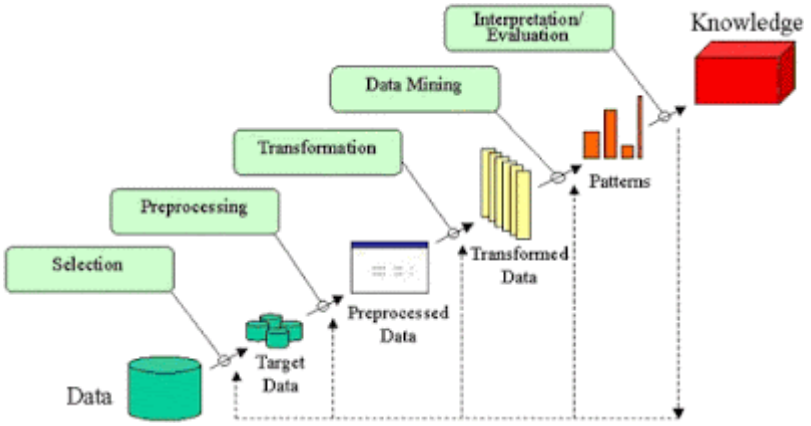
เป้าหมายหลักของ Data Mining

คุณลักษณะและเป้าหมายหลักของ Data Mining คือ ใช้สกลับหรือค้นหา Pattern ของข้อมูลที่ฝังลึกและซ่อนเร้นอยู่ภายในฐานข้อมูลขนาดใหญ่ โดยใช้สถาปัตยกรรม Client–Server (Client/server architecture) ใช้เครื่องมือสมัยใหม่ที่สามารถแสดงผลแบบกราฟฟิก ผู้ใช้สามารถดูข้อมูลแบบเจาะลึก (data drills) และสามารถใช้เครื่องมือในการสอบถามข้อมูลได้อย่างง่ายดาย โดยไม่ต้องอาศัยความชำนาญของ programmer บ่อยครั้งเราอาจค้นพบผลลัพธ์ที่เราไม่คาดหวังมาก่อน เครื่องมือจะทำให้เราใช้งานได้ง่าย ซึ่งเครื่องมือนอกจากจะแสดงผลกราฟิกได้แล้วยังรวม Spreadsheets เอาไว้ด้วย

ขั้นตอนการทำ Data Mining

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล

– Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ



รูปที่ 2.14 ขั้นตอนการทำ Data Mining

ส่วนประกอบการทำ Data Mining

– Database, Data Warehouse, World Wide Web และ Other Info Repositories เป็นแหล่งข้อมูลสำหรับการทำเหมืองข้อมูล

– Database หรือ Data Warehouse Server ทำหน้าที่นำเข้าข้อมูลตามคำขอของผู้ใช้

– Knowledge Base ได้แก่ ความรู้เฉพาะด้านในงานที่ทำจะเป็นประโยชน์ต่อการสืบค้น หรือประเมินความน่าสนใจของรูปแบบผลลัพธ์ที่ได้

– Data Mining Engine เป็นส่วนประกอบหลักประกอบด้วยโมดูลที่รับผิดชอบงานทำเหมืองข้อมูลประเภทต่างๆ ได้แก่ การหากฎความสัมพันธ์ การจำแนกประเภท การจัดกลุ่ม

– Pattern Evaluation Module ทำงานร่วมกับ Data Mining Engine โดยใช้มาตรวัดความน่าสนใจในการกลั่นกรองรูปแบบผลลัทธ์ที่ได้ เพื่อให้การค้นหามุ่งเน้นเฉพาะรูปแบบที่น่าสนใจ

– User Interface ส่วนติดต่อประสานระหว่างผู้ใช้กับระบบการทำเหมืองข้อมูล ช่วยให้ผู้ใช้สามารถระบุงานทำเหมืองข้อมูลที่ต้องการทำ ดูข้อมูลหรือโครงสร้างการจัดเก็บข้อมูล ประเมินผลลัพธ์ที่ได้

ลักษณะเฉพาะของข้อมูลที่มีการใช้ทำ Data Mining

– ข้อมูลขนาดใหญ่ เกินกว่าจะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า หรือโดยการใช้ Database Management System (DBMS) ในการจัดการฐานข้อมูล

– ข้อมูลที่มาจากหลายแหล่ง โดยอาจรวบรวมมาจากหลายระบบปฏิบัติการหรือหลาย DBMS เช่น Oracle , DB2 , MS SQL , MS Access เป็นต้น

– ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดช่วงเวลาที่ทำการ Mining หากข้อมูลที่มีอยู่นั้นเป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาจะต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้นไว้และนำฐานข้อมูลที่บันทึกไว้มาทำ Mining แต่เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลา จึงทำให้ผลลัพธ์ที่ได้จากการทำ

Mining สมเหตุสมผลในช่วงเวลาหนึ่งเท่านั้น ดังนั้นเพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องเหมาะสมอยู่ตลอดเวลาจึงต้องทำ Mining ใหม่ทุกครั้งในช่วงเวลาที่เหมาะสม

– ข้อมูลที่มีโครงสร้างซับซ้อน เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย ข้อมูลเหล่านี้สามารถนำมาทำ Mining ได้เช่นกันแต่ต้องใช้เทคนิคการทำ Data Mining ขั้นสูง

การจำแนกประเภทข้อมูล (Data classification)

หากฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้

การแบ่งกลุ่มข้อมูล (Data clustering)

แบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน

คุณลักษณะของ Data Mining

– การค้นหาข้อมูลโดยอาศัยเทคโนโลยีการทำเหมืองข้อมูล ภายใต้ฐานข้อมูลขนาดใหหรือคลังข้อมูล ซึ่งข้อมูลอาจถูกสะสมมานานหลายปี

– ผู้ใช้งานระบบสารสนเทศไม่จำเป็นต้องมีทักษะในการเขียนโปรแกรม เนื่องจากมีเครื่องมือช่วยค้นหาข้อมูลจากคลังข้อมูลได้อย่างรวดเร็ว

– ผู้ใช้ต้องกำหนดขอบเขตการค้นหาข้อมูลให้ชัดเจนเพื่อความรวดเร็ว

– อาจมีการประมวลผลข้อมูลแบบขนาน เพื่อเพิ่มประสิทธิภาพในการค้นหาและวิเคราะห์ข้อมูล

– เครื่องมือสำหรับทำเหมืองข้อมูลสามารถใช้งานร่วมกับโปรแกรม Spreadsheet และเครื่องมือพัฒนาต่าง ๆ ได้เป็นอย่างดี

ประโยชน์ของ Data Mining

– ช่วยชี้แนวทางการตัดสินใจและคาดการณ์ผลลัพธ์ที่จะได้จากการตัดสินใจ

– เพิ่มความเร็วในการวิเคราะห์ฐานข้อมูลขนาดใหญ่

– ค้นหาส่วนประกอบที่ซ่อนอยู่ภายในเอกสาร รวมถึงความสัมพันธ์ของส่วนประกอบต่าง ๆ ด้วย

– เชื่อมโยงหน่วยงานต่าง ๆ ภายในองค์กร

– การจัดกลุ่มข้อมูล เช่น จัดกลุ่มลูกค้าทั้งหมดของบริษัทประกันภัยที่ประสบอุบัติเหตุลักษณะเดียวกันเพื่อดำเนินการต่าง ๆ ตามนโยบายของบริษัท

2.2.3 ทฤษฎีเกี่ยวกับการ visualization

2.2.3.1 Visualization คืออะไร

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้สามารถเรียนรู้และจดจ าข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวว่า Visualization ก็คือการสร้างมโนภาพของสิ่งต่าง ๆที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการน า เสนอหรือน ามาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ ซึ่งข้อดีของ Visualization มีดังนี้

– ช่วยในการแสดงข้อมูลที่มีปริมาณสูง

– ช่วยในการรับรู้หรือคาดคะเนสิ่งที่อาจจะเกิดขึ้นในอนาคต

– Visualization ไม่เพียงแต่แสดงรายละเอียดข้อมูลในตัวเองเท่านั้นยังสามารถแสดงปัญหาที่เกิดขึ้นได้ด้วย

– ช่วยให้ผู้ใช้สามารถเข้าใจหรือวิเคราะห์ข้อมูลได้สะดวกขึ้นไม่ว่าขอบเขตข้อมูลนั้นจะมีขนาดใหญ่หรือเล็ก

– ช่วยให้ผู้ใช้สามารถสันนิษฐานข้อมูลได้สะดวกขึ้น (Ware, 2004)

ซึ่ง Visualization เป็นมากกว่า วิธีการทาง Computer Visualization เป็นการน าเสนอข้อมูลแบบหนึ่งที่ท าการแสดงข้อมูลให้อยู่ในรูปแบบของ VisualForm ซึ่งอาจจะเป็นการแสดงในรูปแบบของรูปภาพ ,กราฟ หรือ แผนภาพ ซึ่ง ผลลัพธ์ของการท า Visualization คือ การนำข้อมูลต่าง ๆที่ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้สามารถสัมผัสได้ ซึ่งกระบวนการต่าง ๆในการแสดงหรือนำข้อมูลต่าง ๆ ออกมาอาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือ ใช้ในการวิเคราะห์ข้อมูลโดยจุดมุ่งหมายของ Visualization ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ (Diehl, 2007) เพื่อช่วยในการลดช่องว่างระหว่างผู้ใช้และตัวข้อมูล และยังช่วยผู้ใช้สามารถเรียนรู้ ได้อย่างมีประสิทธิภาพ (Keim, 2002)

2.2.3.2วิธีการของ Visualization

การใช้ Visual เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้จะท าขั้นตอนหลักๆอยู่ 3 ขั้นตอน คือ Overview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้ต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้ก็จะท าการตัดสินใจเลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้ก็จะท าการเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็จะอ้างอิงหรือพัฒนาจากขั้นตอนเหล่านี้ซึ่ง Visualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ใช้ต้องการโดยอาจจะใช้หลายๆวิธีการรวมกันเพื่อให้ได้ข้อมูลที่ผู้ใช้ต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามารถนำมาผ่านกระบวนการของ Visualization มีลักษณะต่าง ๆ มากมายดังนี้

ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง,ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi Dimensional Data ได้แก่ Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อข่าวต่าง ๆ และ Web Document, Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document ,Algorithms และ Software ได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลายๆ เช่น

– ประเภทที่แสดงเป็น 2D/3D เช่น แผนภูมิแท่ง และ แผนภูมิxy

– ประเภทที่แสดงข้อมูลในลักษณะภูมิศาตร์ เช่น Parallel Coordinates และ ภาพ Landscape

– ประเภทที่แสดงเป็น Icon–Base เช่น Needle Icon และ Star Icon

– ประเภทที่แสดงข้อมูลแบบ Dense Pixel เช่น Recursive Pattern และ Circle Segment

– ประเภทที่แสดงข้อมูลแบบ Stack เช่น Tree Stamp และ Dimension Stacking

(Dykes, MacEachren & Kraak, 2005,Keim ,2002)

โดยในการศึกษานี้จะนำวิธีการ Visualization ในการแสดงผลข้อมูลในรูปแบบสามมิติ โดยมีวิธีการดังต่อไปนี้

– Interaction techniques เป็นเทคนิคการให้Visualizations โต้ตอบและวิเคราะห์ข้อมูล

โดยตรงและแสดงผลของการวิเคราะห์รูปของ Visualizations เปลี่ยนไปตามวัตถุประสงค์ที่เลือก และสามารถแสดงความสัมพันธ์กับ Visualization อื่น ๆ ซึ่งเป็นวิธีการที่ใช้ในการสำรวจรายละเอียดข้อมูลอื่น ๆ และขณะเดียวกันก็ยังรักษาภาพรวมของข้อมูลไว้ด้วย ซึ่งแนวความคิดจะเป็นการแสดงข้อมูลที่มีระดับสูงในข้อมูลที่สนใจและแสดงรายละเอียดอื่นในระดับต่ำกว่า

– Dynamic Projections เป็นเทคนิคการแสดงผลหน้าจอโดยแสดงตามมิติของข้อมูลต่าง ๆที่ ได้เลือกมา ซึ่งจำนวนที่แสดงผลก็คือจำนวนมิติของข้อมูลนั่นเองโดยลักษณะการแสดงข้อมูลก็อาจจะแสดงในลักษณะสุ่ม, แสดงตามความต้องการของผู้ใช้ หรือ แสดงตามข้อมูลที่มีความสัมพันธ์ด้วย

– Interactive Filtering ในการสำรวจข้อมูลขนาดใหญ่การก าหนดกลุ่มของข้อมูลที่สนใจนั้นเป็นสิ่งที่สำคัญ โดยการก าหนดกลุ่มข้อมูลนั้นสามารถจะกำหนดโดยตรง(Browsing)จากผู้ใช้เองหรือกำหนดจากการ Query ซึ่งการกำหนดแบบ Brow นั้นเป็นสิ่งที่ยากมากสำรับชุดข้อมูลขนาดใหญ่มาก และการก าหนดแบบ Query ก็มักจะได้ผลลัพธ์ที่ไม่เป็นไปตามต้องการ ซึ่ง Interaction Techniques ได้ถูกพัฒนามาเพื่อปรับปรุงการระบุกลุ่มข้อข้อมูลในการสำรวจข้อมูล ซึ่งตัวอย่างก็คือ Magic Lenses แนวคิดพื้นฐานของ Magic Lenses คือการใช้เครื่องมือเหมือนแว่นตาเพื่อสนับสนุนการขยายความข้อมูลโดยตรงใน Visualization

– Interactive Zooming เป็นเทคนิคที่ใช้ใน Application ทั่วไป ซึ่งใช้ในการจัดการข้อมูลที่มีปริมาณสูง ซึ่งเป็นเครื่องมือในการแสดงข้อมูลรายละเอียดแต่ขณะเดียวกันก็ต้องการแสดงข้อมูลภาพรวมทั้งหมดด้วย

– Interactive Linking and Brushing เทคนิคในการรวมการแสดงผลที่มีลักษณะแตกต่างกันให้เข้ามาเป็นเทคนิคเดียวกัน (Keim, 2002)จากวิธีการและประโยชน์ของ Visualization ที่กล่าวมาได้มีการน าไปใช้ในการแก้ปัญหาต่าง ๆ โดยหนึ่งในวิธีการนำไปใช้ก็คือ การน าไปใช้ในการจัดการฐานข้อมูล ซึ่งชื่อว่า THINKBASE ซึ่งช่วยให้การค้นหาข้อมูลที่มีปริมาณที่สูงโดยจะแสดงความสัมพันธ์ของข้อมูลที่ต้องการค้นหาด้วย(Hirsch, Hosking & Grundy, 2009)

ต้นไม้ตัดสินใจเป็นวิธีกำรเรียนรู้ของเครื่องที่นิยมใช้มากที่สุดแบบหนึ่งโดย การจำแนก (classification) ข้อมูลออกเป็นคลาส (class) ต่าง ๆ โดยใช้คุณสมบัติ (attribute) ของข้อมูลในการจำแนกว่าคุณสมบัติใดของข้อมูลที่เป็นตัวกำหนดกำรจำแนกและคุณสมบัติแต่ละตัวของข้อมูลมีกำรวัดความสำคัญอย่างไร

ต้นไม้ตัดสินใจ ประกอบไปด้วย

– โหนดภายใน (internal node) คือ คุณสมบัติต่าง ๆ ของข้อมูล ใช้ในการตัดสินใจว่าข้อมูลจะไปอยู่ในกรณีไหน โดยโหนดภายในที่เป็นโหนดเริ่มต้น เรียกว่า โหนดราก

– กิ่ง (branch, link) เป็นค่าคุณสมบัติหรือเงื่อนไขของคุณสมบัติในโหนดที่ใช้ในการจำแนกข้อมูล ซึ่งโหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่ำคุณสมบัติของโหนดภายในนั้น

– โหนดใบ (leaf node) คือคลาสต่าง ๆ ซึ่งเป็นผลลัพธ์ในการจำแนกข้อมูล

ลักษณะกำรเรียนรู้ของต้นไม้ตัดสินใจ

– ผลกำรเรียนรู้แสดงอยู่ในรูปที่เข้าใจง่าย ซึ่งง่ายต่อการวิเคราะห์คุณสมบัติที่มีผลต่อการจำแนกคลสต่าง ๆ

– แต่ละเส้นทำงจำกโหนดรำกถึงโหนดใบสามารถแสดงให้อยู่ในรูปกฎ IF–THEN ได้

– มีความทนทานต่อข้อมูลรบกวน (noisy data)

– กำรเรียนรู้มีความรวดเร็วเมื่อเทียบกับอัลกอริทึมสำหรับจำแนกชนิดอื่น

– เหมาะแก่กำรน ำไปใช้ในการวิเคราะห์งำนทำงด้ำนธุรกิจ ความเสี่ยงของลูกหนี้

ซึ่งในงานวิจัยนี้ต้องการนำการเรียนรู้ของต้นไม้ตัดสินใจมำใช้ในการวิเคราะห์หาพารามิเตอร์ที่มีผลต่อคุณภาพชิ้นงาน โดยจำแนกคุณสมบัติวัตถุดิบที่นำเข้ามาและก่อให้เกิดคลาสต่าง ๆ โดยสามารถนำเสนอข้อมูลในรูปของกฎได้ เพื่อใช้ ในการหาความสัมพันธ์และคุมพารามิเตอร์ในการผลิตได้

วิธีการเรียนรู้ของต้นไม้ตัดสินใจ

การสร้างต้นไม้ตัดสินใจจะเป็นการค้นหาจากบนลงล่างแบบตะกรำม (top–downgreedy search) โดยเริ่มจำกเลือกคุณสมบัติที่ดีที่สุดในการแยกคลาสมาเป็นโหนดราก เมื่อข้อมูลผ่านการแบ่งที่โหนดรากแล้ว ก็จะหาคุณสมบัติที่ดีที่สุดของข้อมูลที่ถูกแบ่งออกแล้วมาสร้างโหนดลูกของรากนั้นต่อไป และ

จะวนสร้างโหนดลูกและต้นไม้ย่อยของแต่ละกิ่งไปเรื่อย ๆ จนกว่าข้อมูลที่ผ่านการแบ่งแยกนั้นจะจัดอยู่ในคลาสเดียวกัน

การนิยามค่าความดีของคุณสมบัตินั้นแตกต่างกันไปตามอัลกอริทึม เลือกใช้ค่ามาตรฐานเกน (Gain criterion) ของอัลกอริทึม C4.5ซึ่งเป็นอัลกอริทึมที่แพร่หลายที่สุดอัลกอริทึม C4.5

การสร้างต้นไม้ตัดสินใจแบบ C4.5 คล้ายกับอัลกอริทึม ID3 แต่มีการพัฒนาเพิ่มเติม กล่าวคือ จะใช้ค่ามาตรฐานเกนในกำรตัดสินใจเลือกคุณสมบัติที่ใช้เป็นรากหรือโหนดในต้นไม้ โดยการคำนวณค่าเกณฑ์จะเลือกคุณสมบัติที่มีค่าเกนสูงที่สุดมาเป็นรำกหรือโหนดค่าเกนนี้คำนวณได้โดยใช้ความรู้จำกทฤษฎีสำรสนเทศ คือ ค่าสารสนเทศของข้อมูลขึ้นอยู่กับค่าความน่าจะเป็นของข้อมูล ซึ่งสามารถวัดได้ในรูปแบบของบิต (bits) ตามสูตร

ค่าสารสนเทศของ M หรือ ค่าเอนโทรปี ของ M เขียนแทนด้วย I(M) จากสูตร

$$I(M) = \sum_{i}^{n} -P(m_i) log_2 P(m_i)$$

โดย ถ้าให้ชุดของข้อมูล M ประกอบด้วยค่าที่เป็นไปได้ คือ $\{m_1, m_2,...,m_n\}$ ความน่าจะเป็นที่จะเกิดค่า m_i เท่ากับ $P(m_i)$

จากสูตรแสดงให้เห็นว่าค่าสารสนเทศที่น้อยหมายถึงข้อมูลชุดนั้นมีความแตกต่างกันน้อย แต่ถ้าค่าสารสนเทศมากหมายถึงข้อมูลชุดนั้นมีความแตกต่างกันมาก หรือ ประกอบด้วยตัวอย่างหลายพวกที่มีจำนวนใกล้เคียงกันในการเลือกคุณสมบัติที่จะมาเป็นโหนดรากจะใช้ค่ามาตรฐานเกน นั้นคำนวณจากค่าสารสนเทศทั้งหมดของชุดข้อมูลนั้นลบด้วยค่าสารสนเทศหลังจำกเลือกคุณสมบัติใดคุณสมบัติหนึ่งเป็นรากค่าสารสนเทศหลังจากแบ่งคุณสมบัติคำนวณได้ จากค่าผลรวมของผลคูณระหว่างค่าสารสนเทศของแต่ละโหนดกับอัตราส่วนของตัวอย่างในแต่ละกิ่งต่อตัวอย่างทั้งหมดที่โหนดนั้น หรือความน่าจะเป็นไปได้ของแต่ละคุณสมบัติ

$$I_x(T) = \sum_{i=1}^{n} \frac{|t_i|}{|T|} I(t_i)$$

โดยที่ T คือ ข้อมูลเรียนรู้ และ X คือคุณสมบัติที่เป็นโหนด ซึ่งมีค่าเป็นไปได้ n ค่าโหนดที่แบ่งตามคุณสมบัติ X จะสามำราถแบ่งข้อมูล T ออกมาเป็นกิ่ง $\{t_1,t_2,t_3,...,t_n\}$ ตำมค่ำโหนดที่
เป็นไปได้

ค่ามาตรฐานของคุณสมบัติ X สามารถคำนวณได้จากการลบค่าสารสนเทศทั้งหมดที่โหนดนี้กับค่าสารสนเทศที่ได้หลังจากแบ่งด้วยคุณสมบัติ X

$$Gain(X) = I(T) - I_x(T)$$

ในอัลกอริทึม C4.5 ได้เพิ่มกำรใช้ค่ำมำตรฐำนอัตรำส่วนเกน (Gain Ratiocriterion) ในการตัดสินใจเลือกคุณสมบัติที่จะใช้เป็นรากหรือโหนดอีกอย่างหนึ่ง เนื่องจากค่าเกนมาตรฐานจะมีค่าไบแอส(Bias)อย่างมากกับข้อมูลที่ประกอบด้วยคุณสมบัติที่มีค่าเป็นไปได้จำนวนมาก ๆ เช่น ตัวอย่างที่ไม่ซ้ำกัน หรือตัวอย่ำงที่ประกอบด้วยข้อมูลคลาสเดียว การแก้ไขความอคติของค่ามาตรฐานเกนสามารถทำได้โดยการปรับค่ามาตรฐานเกนให้ถูกต้อง โดยใช้ค่าสารสนเทศของการแบ่งแยก (split information) ของคุณสมบัติแต่ละตัว ถ้ำให้ T เป็นชุดข้อมูลแบ่งตามคุณสมบัติ X และได้ชุดของตัวอย่างย่อยในแต่ละกิ่ง n ชุด ตามค่าที่เป็นไปได้ในคุณสมบัติ X ทำการคำนวณค่าสารสนเทศของการแบ่งแยกได้ ดังนี้

$$\text{ค่าสารสนเทศของการแบ่งแยก} = -\sum_{i-1}^{n} \frac{|t_i|}{|T|} \log_2 \frac{|t_i|}{|T|}$$

เมื่อนำค่าสารสนเทศของการแบ่งแยกไปหารค่ามาตรฐานเกนจะได้ค่ามาตรฐานอัตราส่วนเกน ซึ่งช่วยแก้ไขความอคติของค่ามาตรฐานเกนได้โดยทำให้ค่ามาตรฐานอัตราส่วนเกนในการแบ่งด้วยค่าสารสนเทศของการแบ่งแยกคุณสมบัติที่มีการกระจายสูงถูกปรับลดลง ดังนั้นค่ามาตรฐานอัตราส่วนเกนในคุณสมบัติของตัวอย่างที่มีการกระจายตัวของข้อมูลสูงดัง ที่กล่าวมาแล้วจะไม่มีค่าสูงที่สุดเสมอการตัดเล็มต้นไม้ตัดสินใจ

ในกำรสร้างต้นไม้ตัดสินใจจะแบ่งข้อมูลจนกระทั้งได้ข้อมูลที่เป็นคลาสเดียวกันหมด แต่จะเฉพาะเจาะจงกับข้อมูลที่ใช้ในการเรียนรู้เท่านั้น หรือ เรียกว่า ภาวะเหมาะเกินไป ยิ่งต้นไม้ตัดสินใจที่มีขนาดใหญ่ขึ้นจะให้ข้อมูลความถูกต้องบนข้อมูลสอนมากขึ้น แต่เมื่อนำไปใช้งานจริงกลับทำให้ความถูกต้องลดลง

ดังนั้นต้นไม้ควรทำการตัดเล็มเพื่อให้ได้ต้นไม้ขนาดเล็กลง และ ลดความเฉพะเจาะจงกับข้อมูลที่ใช้สอนในอัลกอริทึม C4.5 จะตัดเล็มโดยใช้ค่าความผิดพลำด (error–based pruning)คือมีการรวมต้นไม้ย่อยเข้าด้วยกันเป็นโหนดเดียวกันก็ต่อเมื่อหลังจากรวมแล้วไม่ทำให้ค่าความผิดพลาดเพิ่มขึ้น โดยค่าความผิดพลาดที่ใช้ทดสอบกับข้อมูลที่ไม่เคยเห็นจะใช้ค่าจำกัดบนของการกระจายแบบไบโนเมียล (binomial distribution) ที่ระดับความเป็นอิสระ CF(confidence level)ถ้ำมีข้อมูล N ตัวที่โหนด และมีข้อมูล E ตัวเป็นข้อมูลที่มีคลาสไม่ถูกต้อง ค่าความผิดพลาดในโหนดนี้เขียนได้เป็น $U_{cf}(E,N)$ โดยมีข้อกำหนดที่ว่าขนาดของตัวอย่างสอนเท่ากับขนาดตัวอย่างของข้อมูลที่ไม่เคยเห็น ซึ่งสามารถคาดได้ว่าจะมีจำนวนข้อมูลที่จำแนกผิดพลาดเท่ากับ $N \times U_{cf}(E,N)$ ตัว ซึ่งถ้ำคำนวณจำนวนข้อมูลที่คาดว่าจะจำแนกผิดพลาดของแต่ละกิ่งรวมกันแล้วมากกว่าจำนวนข้อมูลที่คาดว่าจะจำแนกผิดพลำดของโหนดที่แตกกิ่งนั้น ก็จะตัดโหนดที่เป็นลูกในทุกกิ่งของโหนดที่แตกกิ่งนั้นออกให้หมดจนเหลือเฉพาะโหนดที่แตกกิ่งนั้นไว้โหนดเดียว

2.2.5 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.5.1 หลักการออกแบบเว็บไซต์

เว็บไซต์เป็นสื่อที่ได้รับความนิยมอย่ามากบนอินเตอร์เน็ต ซึ่งเว็บไซต์เป็นสื่อทีอยู่ในความควบคุมของผู้ใช้โดยสมบูรณ์ กล่าวคือ ผู้ใช้สามารถตัดสินใจเลือกได้ว่าจะดูเว็บไซต์ใดและจะไม่เลือกดูเว็บไซต์ใดได้ตามต้องการจำทำให้ผู้ใช้ไม่มีความอดทนต่ออุปสรรคและปัญหาที่เกิดจากการออกแบบเว็บไซต์ผิดพลาดถ้าผู้ใช้เห็นว่าเว็บที่กำลังดูอยู่นั้นไม่มีประโยชน์ต่อตัวเขา หรือไม่เข้าใจว่าเว็บไซต์นี้จะใช้งานอย่างไร เขาก็สามารถที่จะเปลี่ยนไปดูเว็บไซต์อื่น ๆ ได้อย่างรวดเร็ว เนื่องจากในปัจจุบันมีเว็บไซต์อยู่มากมาย และยังมีเว็บไซต์ที่เกิดขึ้นใหม่ ๆ ทุกวัน ผู้ใช้จึงมีทางเลือกมากขึ้น และสามารถเปรียบเทียบคุณภาพของเว็บไซต์ต่าง ๆ ได้เองเว็บไซต์ที่ได้รับการออกแบบอย่างสวยงาม มีการใช้งานที่สะดวก ย่อมได้รับความสนใจจากผู้ใช้มากกว่าเว็บไซต์ที่ดูสับสนวุ่นวาย มีข้อมูลมากมายแต่หาอะไรไม่เจอ นอกจากนี้ยังใช้เวลาในการแสดงผลแต่ละหน้านานเกินไป ซึ่งปัญหาเหล่านี้ล้วนเป็นผลมาจากการออกแบบเว็บไซต์ไม่ดีทั้งสิ้น ดังนั้น การออกแบบเว็บไซต์จึงเป็นกระบวนการสำคัญในการสร้างเว็บไซต์ให้ประทับใจผู้ใช้ ทำให้เขาอยากกลับเข้ามาเว็บไซต์เดิมอีกในอนาคต ซึ่งนอกจากต้องพัฒนาเว็บไซต์ที่ดีมีประโยชน์แล้วยังต้องคำนึงถึงการแข่งขันกับเว็บไซต์อื่น ๆ อีกด้วย

2.2.5.2 องค์ประกอบในการออกแบบเว็บไซต์

การออกแบบเว็บไซต์เพื่อให้มีประสิทธิภาพ และสามารถดึงดูดความสนใจของผู้คนได้ดี จะต้องมีองค์ประกอบของเว็บไซต์อย่างครบถ้วน ซึ่งได้แก่

- ความเรียบง่าย เข้าใจง่าย

การออกแบบเว็บไซต์ที่ดี จะต้องเน้นที่ความเรียบง่ายเป็นหลัก โดยเลือกนำเสนอเฉพาะสิ่งที่ต้องการนำเสนอจริง ๆ ในรูปแบบที่หลากหลาย โดยอาจจะเป็นสีสัน กราฟิก ภาพเคลื่อนไหวหรือตัวอักษร ที่สำคัญจะต้องมีการนำเสนอที่ไม่ดูรกหน้าเว็บจนเกินไป เพื่อไม่ให้เกิดความรู้สึกรกสายตาหรือสร้างความเบื่อหน่าย น่ารำคาญให้กับผู้ที่เข้าชมเว็บไซต์ มีตัวอย่างเว็บไซต์ที่มีการออกแบบโดยเน้นความเรียบง่ายได้ดี คือ Apple, Nokia และ Microsoft เป็นต้น

- ความสม่ำเสมอ ไม่สับสน

ควรออกแบบเว็บไซต์ด้วยความสม่ำเสมอ คือจะต้องมีรูปแบบ กราฟิก โทนสีและการตกแต่งต่าง ๆ ให้แต่ละหน้าบนเว็บไซต์มีความคล้ายคลึงกัน และเป็นแนวเดียวกันไปตลอดทั้งเว็บไซต์ ดังตัวอย่างเว็บไซต์ทั่ว ๆ ไปที่จะสังเกตเห็นได้ว่าทุกหน้าของเว็บไซต์นั้น จะเน้นการตกแต่งในรูปแบบเดียวกันทั้งหมด ต่างก็แค่การนำเสนอของแต่ละหน้าเท่านั้น

- สร้างความโดดเด่น เป็นเอกลักษณ์

การออกแบบเว็บไซต์เพื่อให้สามารถสื่อถึงจุดประสงค์ในการนำเสนอเว็บได้ดี จะต้องมีการสร้างความเป็นเอกลักษณ์และจุดเด่นให้กับเว็บไซต์ เพื่อให้สามารถสะท้อนถึงลักษณะขององค์กรได้มากที่สุด

โดยการสร้างเอกลักษณ์ดังกล่าวนั้น อาจใช้ชุดสี รูปภาพ ตัวอักษรหรือกราฟิก นอกจากนี้ก็ต้องขึ้นอยู่กับว่า เป็นเว็บไซต์แบบบทางการหรือไม่ เพื่อจะได้ออกแบบได้อย่างเหมาะสมที่สุด

- เนื้อหาต้องดี ครบถ้วน

เนื้อหาเป็นสิ่งที่สำคัญที่สุดของการสร้างเว็บไซต์ เพราะสิ่งที่ทำให้ผู้คนเกิดความสนใจ และหมั่นติดตามเว็บไซต์เหล่านั้นอยู่เสมอ ก็คือเนื้อหาที่มีความสมบูรณ์และน่าสนใจ นอกจากนี้จะต้องมีการปรับปรุง พัฒนาเนื้อหาบนเว็บให้มีความทันสมัยอยู่เสมอ รวมถึงข้อมูลต้องมีความถูกต้องที่สุด

- ระบบเนวิเกชั่นใช้ง่าย

ระบบเนวิเกชั่นเป็นเสมือนป้ายบอกทางเพื่อให้ผู้ใช้งาน ไม่เกิดความสับสนในขณะใช้งานเว็บไซต์ ซึ่งการออกแบบเนวิเกชั่นก็จะต้องเน้นที่ความเรียบง่าย ใช้งานสะดวก และมีความเข้าใจได้ง่าย ที่สำคัญจะต้องมีตำแหน่งการวางที่สม่ำเสมอเพื่อให้ดูเป็นแนวทางเดียวกัน ทำให้ผู้ใช้งานหรือผู้ชมรู้สึกประทับใจและจดจำเว็บไซต์ได้ง่ายขึ้น ส่วนใครที่มีการนำกราฟิกมาใช้ในระบบเนวิเกชั่น ก็จะต้องเลือกกราฟิกที่สามารถสื่อความหมายได้ดีเช่นกัน

- คุณภาพของเว็บไซต์

เว็บไซต์ที่ดีจะต้องมีคุณภาพ ทั้งสิ่งที่ปรากฏให้เห็นบนเว็บไซต์ ไม่ว่าจะเป็นกราฟิก ชนิดตัวอักษร รูปภาพหรือสีสันที่ใช้ เนื้อหาที่นำมาแสดงผล ซึ่งหากเว็บไซต์มีคุณภาพก็จะสร้างความน่าเชื่อถือ และเป็นจุดเด่นที่ทำให้ผู้คนส่วนใหญ่เกิดความสนใจได้ดี เพราะฉะนั้นห้ามละเลยในส่วนของคุณภาพเด็ดขาด

- ความสะดวกในการเข้าใช้งาน

เว็บไซต์ควรให้ความสะดวกสบายแก่ผู้ใช้งานได้ดี คือจะต้องมีการแสดงผลได้ในทุกระบบปฏิบัติการ ไม่ว่าจะเป็นเว็บเบราว์เซอร์ คอมพิวเตอร์ โน๊ตบุ๊คหรือบนโทรศัพท์มือถือ ที่สำคัญจะต้องมีความละเอียดของการแสดผลและสามารถใช้งานได้โดยไม่มีปัญหาด้วย

- ความคงที่ของการออกแบบ

การออกแบบเว็บไซต์ควรจะมีความคงที่ในการออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกัน และมีการเรียบเรียงเนื้อหาอย่างรอบคอบ ทำให้เว็บมีความน่าเชื่อถือ และดูมีคุณภาพ ช่วยสร้างความประทับใจให้กับผู้ใช้งานได้เป็นอย่างดี

- ความคงที่ของการทำงาน

ระบบการทำงานบนเว็บไซต์จะต้องมีความคงที่ และสามารถใช้งานได้ดี ซึ่งนอกจากการออกแบบระบบการทำงานให้มีความทันสมัยและสร้างสรรค์แล้ว ก็จะต้องหมั่นตรวจสอบอยู่เสมอ เพราะหากระบบการใช้งานมีความผิดปกติก็จะได้แก้ปัญหาได้ทัน นอกจากนี้อาจมีการอัพเดตดีไซน์ให้ทันสมัยขึ้นบ่อย ๆ เพื่อให้ผู้ใช้งานรู้สึกสนุกไปกับการใช้งานเว็บไซต์

ส่วนประกอบสำคัญบนหน้าเว็บเพจที่ต้องมีบนหน้าเว็บเพจ จะมีส่วนประกอบสำคัญที่จำเป็นต้องมีอยู่ 3 ส่วน ได้แก่

- ส่วนหัวของหน้า (Header)

อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซต์ต่อไป ซึ่งส่วนใหญ่ก็มักจะมีการใส่ภาพกราฟฟิคให้ดูสวยงาม สิ่งสำคัญหลักๆ เลย ก็คือ โลโก้ ชื่อเว็บไซต์และเมนูหลักที่สามารถลิ้งค์ไปยังเนื้อหาในหน้าเว็บเพจต่าง ๆ ได้

- ส่วนของเนื้อหา (Body)

อยู่บริเวณตอนกลางของหน้าเว็บ โดยจะแสดงข้อมูลเกี่ยวกับเนื้อหาบนเว็บแบบคร่าวๆ ซึ่งก็จะมีข้อความ กราฟฟิค ตารางข้อมูลหรือวิดีโอประกอบอยู่ และหากมีเมนูแบบเฉพาะกลุ่มก็จะถูกจัดไว้ในหน้านี้เช่นกัน และที่สำคัญเนื้อหาในส่วนนี้ควรจะมีความกระชับ เข้าใจง่าย มีการใช้รูปแบบตัวอักษรแบบเรียบง่ายและเป็นระเบียบ

- ส่วนท้ายของหน้า (Footer)

อยู่ล่างสุดของหน้าเว็บ ซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่าง ๆ เพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูลเจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่าง ๆ เกี่ยวกับการใช้งานเว็บไซต์อย่างถูกต้อง เป็นต้น

2.2.5.3 วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์

การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บ ว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไรต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่มดังต่อไปนี้

สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอบโยนและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

สีโทนเย็น (Cool Colors) เป็นสีแห่งความสุภาพและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้โน้มน้าวจากในระยะไกลได้อีกด้วย

สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

สีสามารถสื่อถึงอารมณ์ความรู้สึกและสื่อความหมายของเว็บไซต์นั้น ๆ ได้อย่างชัดเจน ดังนั้นหากเลือกใช้สีไม่เหมาะสมกับเนื้อหาหรือจุดประสงค์ของเว็บ ก็จะทำให้เว็บดูไม่น่าสนใจ ขาดความน่าเชื่อถือและทำให้ผู้ที่เคยเข้ามาใช้บริการไม่คิดจะกลับมาใช้บริการอีก

2.2.6 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

CSS ย่อมาจาก Cascading Style Sheet มักเรียกโดยย่อว่า "สไตล์ชีต" คือภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบ (หรือ "Style") ของเนื้อหาในเอกสาร อันได้แก่ สีของข้อความ สีพื้นหลัง ประเภทตัวอักษร และการจัดวางข้อความ ซึ่งการกำหนดรูปแบบ หรือ Style นี้ใช้หลักการของการแยกเนื้อหาเอกสาร HTML ออกจากคำสั่งที่ใช้ในการจัดรูปแบบการแสดงผล กำหนดให้รูปแบบของการแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลลัพธ์ของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือต้องการควบคุมให้รูปแบบการแสดงผลเอกสาร HTML มีลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสารภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการกำหนดรูปแบบ (Style) เอกสาร HTML ถูกเพิ่มเข้ามาครั้งแรกใน HTML 4.0 เมื่อปีพ.ศ. 2539 ในรูปแบบของ CSS level 1 Recommendations ที่กำหนดโดย องค์กร World Wide Web Consortium หรือ W3C

ประโยชน์ CSS

ภาษา CSS (Cascading Style Sheets) มีประโยชน์หลายอย่างเลยทีเดียวซึ่งทำให้การพัฒนาเว็บเพจด้วยภาษา HTML เป็นเรื่องที่ง่ายมากขึ้นและการเขียน CSS ที่ดีมีผลดีต่อการดันอันดับเว็บไซต์ ด้วย SEO อย่างไร

– ภาษา CSS จะช่วยในการจัดรูปแบบแสดงผลให้กับภาษา HTML ซึ่งจะช่วยลดการใช้ภาษา HTML ให้น้อยลง โดยเหลือเพียงแต่ส่วนที่เป็นเอกสารที่เป็นภาษา HTML เท่านั้นทำให้มีการแก้ไขและทำความเข้าใจได้ง่ายขึ้น

– ทำให้ขนาดไฟล์ HTML น้อยลงเนื่องจาก ภาษา CSS จะช่วยลงการใช้ภาษา HTML ลงทำให้ขนาดไฟล์นั้นก็เล็กลงไปด้วยเช่นกัน

– ภาษา CSS เป็นภาษา Style Sheets โดย Style Sheets ชุดเดียวสามารถใช้กำหนดรูปแบบการแสดงผลให้เอกสาร HTML ทั้งหน้า หรือทุกหน้ามีผลเหมือนกันได้ จึงทำให้เวลาที่มีการแก้ไขก็จะแก้ไขได้ง่ายขึ้นเพียงแก้ไข Style Sheets ที่ใช้งานเพียงชุดเดียวเท่านั้น

– ทำให้เว็บไซต์มีมาตรฐานเพราะการใช้งาน CSS นั้นจะทำให้การแสดงผลในสื่อต่าง ๆ ถูกปรับเปลี่ยนไปได้อย่างเหมาะสม เช่น การแสดงผลบนหน้าจอ และการแสดงผลในมือถือ

– CSS สามารถที่จะใช้งานได้หลากหลาย เว็บบราวเซอร์ ทำให้การใช้งานนั้นสะดวกมากยิ่งขึ้น

– CSS สามารถกำหนดแยกไว้ต่างหากจากไฟล์เอกสาร HTML และสามารถนำมาใช้ร่วม กับเอกสารหลายไฟล์ได้ การแก้ไขก็แก้เพียง จุดเดียวก็มีผลกับเอกสารทั้งหมด CSS กับ HTML นั้นทำหน้าที่คนละอย่างกัน โดย HTMLจะทำหน้าที่ในการวางโครงร่างเอกสารอย่างเป็นรูปแบบ ถูกต้อง เข้าใจง่าย ไม่เกี่ยวข้องกับการแสดงผล ส่วน CSS จะทำหน้าที่ในการตกแต่งเอกสารให้สวยงาม เรียกได้ว่า HTML คือส่วน coding ส่วน CSS คือส่วน design

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

2.3.1.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

เป็นเทคนิคหนึ่งของ Classification ซึ่งเป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูล โดย Classification นั้นเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining)

การเรียนรู้แบบต้นไม้ตัดสินใจ (อังกฤษ: decision tree learning) เป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูล

ในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (leaf node) จะแสดงประเภทของวัตถุ

ต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่าต้นไม้ตัดสินใจแบบจำแนก (classification trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (continuous values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (regression trees)

ต้นไม้การตัดสินใจในการบริหารธุรกิจ เป็นแผนผังต้นไม้ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และ ทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

2.3.1.2 ลักษณะต้นไม้การตัดสินใจ

ต้นไม้การตัดสินใจจะทำการจัดกลุ่ม (classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (node) ของต้นไม้การตัดสินใจคือตัวแปร (attribute) ต่าง ๆ ของชุดข้อมูล เช่นหากต้องการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณาคือ ทัศนียภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์จากต้นไม้คือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่า ซึ่งแต่ละตัวแปรนั้นก็จะมีค่าของตัวเอง (value) เกิดเป็นชุดของตัวแปร-ค่าของตัวแปร (attribute-value pair) เช่น ทัศนียภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่านั้นก็อาจมีค่าได้เป็นใช่ กับ ไม่ใช่ เป็นต้น การทำนายประเภทด้วยต้นไม้ตัดสินใจ จะเริ่มจากบัพราก โดยทดสอบค่าตัวแปรของบัพ แล้วจึงตามกิ่งของต้นไม้ที่กำหนดค่า เพื่อไปยังบัพลูกถัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอบัพใบซึ่งจะแสดงผลการทำนาย

2.3.1.3 ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย

– โหนด คือ คุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่จุดสูงสุดเรียกว่า โหนดราก (Root Node)

– กิ่ง คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจานวนของกิ่งจะเท่ากับคุณสมบัติของโหนด

– ใบ คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ

2.3.1.4 ขั้นตอนในการสร้าง Decision Tree เพื่อใช้จำแนกข้อมูล มีดังนี้

– เลือก Attribute ที่ทาหน้าที่เป็น Root Node

– จาก Root Node สร้างเส้นเชื่อมโยงไปยังโหนดลูก จานวนเส้นเชื่อมโยง จะเท่ากับจำนวนค่าที่เป็นไปได้ทั้งหมดของ Attribute ที่เป็น root node

– ถ้าโหนดลูกเป็นกลุ่มของข้อมูลที่อยู่ในคลาสเดียวกันทั้งหมด ให้หยุดสร้างต้นไม้ แต่ถ้าโหนดลูกมีข้อมูลของหลายคลาสปะปนกันอยู่ ต้อง สร้าง subtree เพื่อจำแนกข้อมูลต่อไป โดยเลือก subtree มาทาหน้าที่ เป็น root node ของ subtree มาทำซ้ำในขั้นตอนที่ 2.) , 3.)

2.3.2 การแสดงผลแบบ visualization

2.3.2.1 การแสดงผลแบบ visualization ด้วย Tableau Public

Tableau คือ โปรแกรมที่ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ(Data Visualiztion) เพื่อทำให้มีความเข้าใจในข้อมูลมากขึ้น ทำให้สามารถนำข้อมูลจำนวนมากมาวิเคราะห์และมองออกได้อย่างง่ายขึ้น ซึ่งตัวโปรแกรมTableau มีด้วยกันหลายเวอร์ชั่นด้วยกัน เช่น Tableau Desktop ,Tableau Online เป็นต้น แต่ในวันนี้เราจะมาเรียนในโปรแกรม Tableau Public เนื่องจากโปรแกรมนี้เป็นตัวที่ฟรี และสามารถใช้งานได้ดี

ขอดีของ Tableau

Tableau เป็นซอฟต์แวร์ที่ สามารถนำข้อมูลจำนวนมากที่มีหลากหลายในองค์กรมาทำการวิเคราะห์เพื่อเป็น ข้อมูลเชิงธุรกิจช่วยให้ผู้บริหารมีข้อมูลเชิงลึกเพิ่มความสามารถในการ ตัดสินใจอย่างรวดเร็วและชาญฉลาด

คุณสมบัติและจุดเด่นของ Tableau

คุณสมบัติที่โดดเด่นอย่างเห็นได้ชัด คือ ผู้ใช้มีเครื่องมือที่ใช้วิเคราะห์ข้อมูลได้หลากหลายมิติ ไม่ว่าจะเป็น รูปแบบของ Online Processing เป็นต้น

– รูปแบบการใช้งานที่ง่ายเพียงคลิ๊กเมาส์ก็สามารถที่จะเปลี่ยนแปลงรายงานได้ตามความต้องการ

– รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และ สามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot

- รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการนำเสนอต่อผู้บริหาร
- การเข้าถึงหรือการใช้งาน สามารถทำได้โดยง่าย ตั้งแต่ผู้บริหารระดับสูง ถึง พนักงานระดับปฏิบัติการ ทั้งนี้ทั้งนั้น สามารถเพิ่มทักษะและแนวคิดให้กับพนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติและการประยุกต์ใช้งานมากขึ้น
- สิ่งที่สำคัญ คือ Engine ของ Tableau Software นั้น พัฒนาจาก VizQL Technology เป็นเทคโนโลยีที่ช่วยให้เห็นภาพของข้อมูลขนาดใดก็ได้เพียงการลากและวาง (Drag and Drop) โดยใช้ข้อมูลจากฐานข้อมูล ตอบสนองในรูปแบบของ กราฟฟิก
-

2.4 วรรณกรรมที่เกี่ยวข้อง

วทัญญุตา นีลาภาตระกูลและชุติมา เบี้ยวไข่มุข (2562) งานวิจัยนี้การศึกษาเกี่ยวกับเทคนิคการทำ เหมืองข้อมูล เพื่อลดปัญหาอัตราการลาออกของพนักงาน และลดค่าใช้จ่ายที่สูงขึ้นจากการสรรหาพนักงานใหม่ในบริษัทประกันภัยแห่งหนึ่ง ชุดข้อมูลที่ใช้วิเคราะห์เป็น ข้อมูลของพนักงานลาออกและพนักงานที่ยังทำ งานอยู่ ตั้งแต่ปีพ.ศ. 2556–2560 จำ นวน 1,000 รายการ 11 แอตทริบิวต์ เทคนิคกฎความสัมพันธ์ถูกนำ มาใช้หารูปแบบของความสัมพันธ์การลาออกของพนักงาน และ สร้างตัวแบบพยากรณ์การลาออกของพนักงาน โดยใช้เทคนิคการวิเคราะห์5เทคนิคคือต้นไม้ตัดสินใจเทคนิค ซัพพอร์ตเวกเตอร์แมชชีน เทคนิคโครงข่ายประสาทเทียม เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้สุด และเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์เหล่านั้น การทดสอบประสิทธิภาพของตัวแบบใช้วิธีไขว้ ข้อมูลโดยแบ่งข้อมูลออกเป็น 5 ชุดเท่ากัน ผลของกฎความสัมพันธ์พบปัจจัยที่มีความสัมพันธ์กับการตัดสินใจ ลาออก คือ เงินเดือน อายุงาน ความก้าวหน้าในอาชีพ ผลการประเมินการปฏิบัติงาน และความสัมพันธ์กับ หัวหน้างาน ผลการทดสอบตัวแบบพยากรณ์ที่เป็นต้นไม้ตัดสินใจให้ค่าความถูกต้องในการพยากรณ์สูงที่สุดที่ ร้อยละ91.03ในขณะที่ตัวแบบพยากรณ์ที่สร้างจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องเท่ากับ ร้อยละ 90.93 เทคนิคโครงข่ายประสาทเทียม ให้ค่าความถูกต้องเท่ากับ ร้อยละ 90.75 เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้สุด ให้ค่าความถูกต้องเท่ากับ ร้อยละ 89.60 และ 82.10 ตามลำดับ ผลที่ได้จาก การวิจัยสามารถนำผลมาประยุกต์ใช้ในการสร้างแผนกลยุทธ์สำ หรับออกแบบสวัสดิการที่เหมาะสมเพื่อรักษา พนักงานให้คงอยู่กับบริษัท และเป็นแนวทางในการสร้างระบบสารสนเทศในการสนับสนุนการตัดสินใจสำ หรับ งานด้านบริหารทรัพยากรบุคคลต่อไป

วีรวัลย์ ศิรินาม (2561) ปัจจัยที่มีผลต่อภาวะโภชนาการของเด็กในศูนย์พัฒนาเด็กเล็กสังกัด เพื่อศึกษาปัจจัยที่มีผลต่อ ภาวะโภชนาการของเด็กในศูนย์พัฒนาเด็กเล็กสังกัด เทศบาลตำบลบ้านกลาง อำเภอสันป่าตอง จังหวัดเชียงใหม่กลุ่มตัวอย่างที่ใช้ในการศึกษาคือเด็กและผู้ปกครองในศูนย์พัฒนาเด็กเล็ก จำนวน 252 คน เก็บรวบรวมขอ้ มูลด้วยแบบสอบถามและการบันทึกน้ำหนักและส่วนสูงของเด็ก

สถิติที่ใช้ในการวิเคราะห์คือ ค่าความถี่ค่าร้อยละ ค่าเฉลี่ย การวิเคราะห์สหสัมพันธ์และการวิเคราะห์การถดถอยพหุคูณ ผลการวิจัยพบว่า เด็กในศูนย์พัฒนาเด็กมีภาวะโภชนาการต่ำกว่าเกณฑ์มาตรฐาน ร้อยละ 7.1 ภาวะโภชนาการปกติ ร้อยละ 81.8 และภาวะโภชนาการเกินเกณฑ์มาตรฐาน ร้อยละ 11.1 ปัจจัยที่มีผลต่อภาวะโภชนาการของเด็กคือ ปัจจัยความรู้เกี่ยวกบัการรับประทานอาหารของ ครอบครัว ปัจจัยเจตคติเกี่ยวกบัการรับประทานอาหารของครอบครัวและการปฏิบัติเกี่ยวกบัการจัดอาหารของครอบครัว

ปนันดา จันทร์สุกรี และวศิน แก้วชาญค้า (2561) สภาพแวดล้อมทางอาหารภายในมหาวิทยาลัยที่มีอิทธิพลต่อ พฤติกรรมการบริโภคเพื่อสุขภาพของนักศึกษาไทย การศึกษาปัจจัยระดับบุคคลที่มีผลต่อพฤติกรรมการบริโภคเพื่อสุขภาพ พบว่า ในภาพรวมความรู้ ด้านอาหารและโภชนาการมีคะแนนเฉลี่ยสูงที่สุด ในขณะที่ปัจจัยทางจิตวิทยามีคะแนนเฉลี่ยต่ำที่สุด เมื่อ จำแนกตามภูมิภาคพบว่า ด้านความชื่นชอบในรสชาติอาหารที่ไม่ดีต่อสุขภาพ (หวาน มัน เค็ม) นักศึกษาของ มหาวิทยาลัยในภาคกลางมีคะแนนเฉลี่ยสูงที่สุด และนักศึกษาของมหาวิทยาลัยในภาคตะวันออกเฉียงเหนือมี คะแนนเฉลี่ยต่ำที่สุด ด้านความรู้ด้านอาหารและโภชนาการ นักศึกษาของมหาวิทยาลัยในภาคใต้มีคะแนน เฉลี่ยสูงที่สุด และนักศึกษาของมหาวิทยาลัยในกรุงเทพมหานครและปริมณฑลมีคะแนนเฉลี่ยต่ำที่สุด ด้านความตระหนักเกี่ยวกับการบริโภคเพื่อสุขภาพ นักศึกษาของมหาวิทยาลัยในภาคตะวันออกเฉียงเหนือมีคะแนน เฉลี่ยสูงที่สุด และนักศึกษาของมหาวิทยาลัยในภาคกลางคะแนนเฉลี่ยต่ำที่สุด

ปรีชา ลิ้มตระกูล,วิภา เจริญภัณฑารักษ์ และ วิทยา พรพัชรพงศ์ (2559) การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล 2) เพื่อพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจของผู้ใช้งานระบบผ่านเว็บไซต์ โดยใช้ข้อมูลปัจจัยการผลิตมันสำปะหลังจากสำนักงานเกษตรจังหวัดกำแพงเพชร เพื่อสร้างตัวแบบพยากรณ์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (DecisionTree) โดยจะใช้อัลกอริทึม จำนวน 5 ตัว ได้แก่ J48, RandomTree, SimpleCart, NaïveBayes, และLADTree แล้ว ทำการทดสอบตัวแบบการพยากรณ์ด้วยวิธี Cross–validation Test พบว่ายังให้ค่าแม่นยำเพียง 70.96% ซึ่ง ผู้วิจัยเห็นว่ายังไม่ดีพอ จึงได้ปรับปรุงวิธีการทดสอบโดยการแบ่งชุดข้อมูลออกเป็น 2 ส่วนได้แก่ ข้อมูลเรียนรู้ (Training Set)และข้อมูลทดสอบ (Test Set) จำนวน 5 ชุด แล้วทำการสร้างตัวแบบการพยากรณ์ด้วยอัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และ LADTree อีกครั้ง พบว่ามีค่าความแม่นยำสูงขึ้นกว่าวิธี Cross–validation Test ในทุกอัลกอริทึม ดังนั้นผู้วิจัยจึงเลือกตัวแบบการพยากรณ์ที่ให้ค่าความแม่นยำสูงสุด ในส่วนของข้อมูลทดสอบ (Test Set) อัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่ 75.64% อัลกอริทึม SimpleCart ให้ค่าความแม่นยำสูงสุดที่ 80.12% และอัลกอริทึม LADTree ให้ค่าความแม่นยำสูงสุดที่ 89.55% ส่วนอัลกอริทึม RandomTree และอัลกอริทึม Naïve Bayes ให้ค่าความแม่นยำต่ำกว่า 70% จึงไม่นำมา พิจารณา จากนั้นนำอัลกอริทึมทั้งสามตัวที่ให้ค่าความแม่นยำสูงสุดไป

พัฒนาตัวแบบการพยากรณ์ผลผลิตมัน สำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล ระบบสารสนเทศและการสืบค้นข้อมูลในรูปแบบเว็บแอพพลิเคชั่น ซึ่ง ระบบสามารถสืบค้น ปรับปรุง เพิ่มเติม บันทึก และแสดงรายงานข้อมูลมันสำปะหลังได้ และผลการตอบแบบ ประเมินความพึงพอใจผ่านเว็บไซต์ของระบบสารสนเทศดังกล่าว จากเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ใช้งานทั่วไป และผู้ดูแลระบบ รวมทั้งสิ้น 30 คน มีความพึงพอใจในการใช้งานระบบดังกล่าวเฉลี่ย ที่ 91% ซึ่งถือว่าอยู่ในระดับที่ดีมาก

เสกสรรค์ วิลัยลักษณ์และคณะ (2558) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูล และสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตก าแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยใช้ข้อมูล นักเรียนระดับมัธยมศึกษาระหว่างปีการศึกษา 2548 – 2556 เพื่อพัฒนาคลังข้อมูลโดยใช้โครงสร้างแบบสโนว์เฟลกสกีมาและนำเสนอรายงาน จากนั้นใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียน ประกอบด้วย16 คุณลักษณะ มาสร้างตัวแบบพยากรณ์ผลการเรียนโดย ใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Cluster Data) จากนั้นนำไปผ่าน กระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์ เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบ พยากรณ์และเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ผลการวิจัยพบว่า คลังข้อมูลนักเรียนระดับมัธยมศึกษาที่ใช้งานโดย ผู้บริหาร หัวหน้ากลุ่มสาระการ เรียนรู้ และอาจารย์ประจำชั้นมีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในการทำเหมืองข้อมูลพยากรณ์ พบว่า ชุดข้อมูลแบบไม่จัดกลุ่มนำมาคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูง ที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสำหรับการสร้าง ระบบพยากรณ์ผลการเรียนของนักเรียน

วีรวัลย์ ศิรินาม (2561) ปัจจัยที่มีผลต่อภาวะโภชนาการของเด็กในศูนย์พัฒนาเด็กเล็กสังกัด เพื่อศึกษาปัจจัยที่มีผลต่อ ภาวะโภชนาการของเด็กในศูนย์พัฒนาเด็กเล็กสังกัด เทศบาลตำบลบ้านกลาง อำเภอสันป่าตอง จังหวัดเชียงใหม่กลุ่มตัวอย่างที่ใช้ในการศึกษาคือเด็กและผู้ปกครองในศูนย์พัฒนาเด็กเล็ก จำนวน 252 คน เก็บรวบรวมขอ้ มูลด้วยแบบสอบถามและการบันทึกน้ำหนักและส่วนสูงของเด็ก สถิติที่ใช้ในการวิเคราะห์คือ ค่าความถี่ค่าร้อยละ ค่าเฉลี่ย การวิเคราะห์สหสัมพันธ์และการวิเคราะห์ การถดถอยพหุคูณ ผลการวิจัยพบว่า เด็กในศูนย์พัฒนาเด็กมีภาวะโภชนาการต่ำกว่าเกณฑ์มาตรฐาน ร้อยละ 7.1 ภาวะโภชนาการปกติ ร้อยละ 81.8 และภาวะโภชนาการเกินเกณฑ์มาตรฐาน ร้อยละ 11.1 ปัจจัยที่มีผลต่อภาวะโภชนาการของเด็กคือ ปัจจัยความรู้เกี่ยวกบัการรับประทานอาหารของ

ครอบครัว ปัจจัยเจตคติเกี่ยวกบัการรับประทานอาหารของครอบครัวและการปฏิบัติเกี่ยวกบัการจัดอาหารของครอบครัว

สุกัญญา บัวศรีและคณะ (2562) วิจัยนี้มีวัตถุประสงค์เพื่อศึกษาภาวะโภชนาการ กิจกรรมทางกาย และปัจจัยส่วน บุคคลที่มีความสัมพันธ์กับภาวะโภชนาการของนักเรียนอายุ 6–12 ปีในจังหวัดเชียงราย กลุ่มตัวอย่าง จำนวน 1,467 คน สุ่มโดยวิธีแบบแบ่งชั้น น าเสนอข้อมูลโดยใช้จำนวนร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบน มาตรฐาน และสถิติทดสอบความสัมพันธ์Pearson chi-square test ผลการวิจัยพบว่า นักเรียนเป็นเพศหญิงมากที่สุด 784 คน (53.4%) อาศัยอยู่นอกเขต เทศบาลจำนวน 794 คน (54.1%) อยู่ในระดับประถมศึกษาตอนปลาย 777 คน (53%) ระดับ การศึกษาสูงสุดของผู้ปกครองส่วนใหญ่ระดับประถมศึกษา 341 คน (23.2%) รายได้ของผู้ปกครอง ส่วนใหญ่น้อยกว่า 5,000 บาทต่อเดือนจำนวน 910 คน (62.0) ภาวะโภชนาการอยู่ในเกณฑ์ปกติมาก ที่สุด โดยมีน้ำหนักตามเกณฑ์อายุจำนวน 1,081 คน (73.7%) น้ำหนักค่อนข้างน้อยและน้อยกว่าเกณฑ์ จำนวน 158 คน (10.7%) น้ำหนักค่อนข้างมากและมากกว่าเกณฑ์จำนวน 228 คน (15.6%) ส่วนสูงตามเกณฑ์อายุจำนวน 1,083 คน (73.8%) ค่อนข้างเตี้ยและเตี้ย 190 คน (13%) ค่อนข้างสูง และสูงกว่าเกณฑ์จำนวน 194 คน (13.2%) น้ำหนักเทียบกับส่วนสูงสมส่วนจำนวน 1,060 คน (72.2%) ค่อนข้างผอมและผอมจำนวน 179 คน (12.2%) ท้วมเริ่มอ้วน และอ้วนจำนวน 228 คน (15.6%) ระยะเวลาของการมีกิจกรรมทางกายเฉลี่ยรวมเท่ากับ 1,771 นาทีต่อสัปดาห์ เป็นกิจกรรมใน ระดับเบาเป็นส่วนใหญ่ 1,360 นาที กิจกรรมระดับหนักและปานกลาง 287 นาที และ 124 นาที ตามลำดับ กิจกรรมทางกายมากกว่า 420 นาทีต่อสัปดาห์จำนวน 1,413 คน (96.3) ปัจจัยด้านเพศ แหล่งที่อยู่อาศัยและการศึกษาสูงสุดของผู้ปกครองมีความสัมพันธ์กับภาวะโภชนาการอย่างมีนัยสำคัญ ทางสถิติที่ระดับ .05 การแก้ไขปัญหาภาวะ ทุพโภชนาการควรได้รับความร่วมมือจากผู้ปกครอง และมี ระบบติดตาม เฝ้าระวัง ภาวะโภชนาการ มีการเชื่อมโยงระหว่างโรงเรียนและครอบครัวอย่างเป็นระบบ และกระตุ้นให้มีกิจกรรมทางกายระดับหนักเพิ่มมากขึ้น

พฤฒิพงศ์ เพ็งศิริและพันธนา ก้อนเชื้อรัตน์ (2558) นำเสนอการพยากรณ์โดยการประยุกต์ ใช้กระบวนการตัดสินใจด้วยเทคนิคต้นไม้ ตัดสินใจเป็นเทคนิคอย่างหนึ่งในการทำเหมืองข้อมูล ซึ่งอาศยความสัมพันธ์ของปัจจัยข้อมูล นักศึกษาเป็นการบ่งชี้ถึงระดับผลการเรียนของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคล สุวรรณภูมิซึ่งผลลัพธ์ที่ได้นั้นจะส่งผลทำให้เอื้อประโยชน์กับแนวทางในการปรับปรุงบริหาร จัดการและส่งเสริมการวางแผนของหลักสูตรการศึกษาที่เกี่ยวข้อง โดยนำปัจจัยข้อมูลของนักศึกษา ได้แก่ข้อมูลทั่วไป ข้อมูลส่วนตัว ข้อมูลการศึกษา ข้อมูลระดับผลการศึกษาของนักศึกษากลุ่มที่ศึกษา มาหาความสัมพันธ์กบผลการศึกษาที่ได้โดยวิเคราะห์จากระดับผลการเรียน (เกรดเฉลี่ย) ของการเรียนในชั้นปีที่ 1 ผลการทดลองพบว่าปัจจัยข้อมูลของนักศึกษาที่ เหมาะสมในการเรียน มีทั้งหมด 7 ตัวแปร จากข้อมูลนำเข้าทั้งหมด 12 ตัวแปร ทั้งนี้ 7 ตัวแปร มาจากต้นไม้ตัดสินใจที่ได้มา

สามารถสรุปเลือกเฉพาะกิ่งที่มีผลมากที่สุด โดยวัดค่าความ แม่นยำ (Accuracy) ได้ค่าสูงถึง 84.78% ซึ่งถือว่าอยู่ในเกณฑ์สูง ดังนั้นตัวแปรที่น่าจะเป็น ปัจจัยที่ส่งผลกับระดับผลการเรียนที่จบการศึกษาสูงสุดคือ ความสม่ำเสมอการเข้าเรียนในการเรียน

อนันต์ ปินะเต (2559) การรับสมัครบุคคลเข้าศึกษาในระดับปริญญาตรี มหาวิทยาลัยมหาสารคาม มีกระบวนการในการคัดเลือกผู้สมัครที่มี คุณสมบัติตรงตามที่มหาวิทยาลัยกำหนด ซึ่งการแนะแนวการศึกษาจึงเป็นกระบวนการที่สำคัญ ในการประชาสัมพันธ์ ข้อมูลการสมัครให้กับผู้สมัครทราบ อาทิเช่น ข้อมูลการเลือกสาขาวิชา โดยทั่วไปผู้สมัครจะเลือกสาขาวิชาโดยยึดหลัก ตามความชอบ ความรู้สึกของตนเองเป็นส่วนใหญ่ โดยไม่ได้คำนึงถึงความรู้ และทักษะด้านต่าง ๆ ของตนเอง ส่งผลให้ เมื่อเข้ามาศึกษาในสาขาวิชานั้นแล้ว เกิดปัญหาผลการเรียนที่ตกต่ำ ไม่ผ่านตามเกณฑ์ที่มหาวิทยาลัยกำหนด อันเป็นผลทำให้ต้องพ้นสภาพการเป็นนิสิตดังนั้นการประชาสัมพันธ์ข้อมูลเพื่อให้ผู้สมัครได้ใช้ประกอบการตัดสินใจในการเลือก สาขาวิชานั้นจึงเป็นสิ่งสำคัญในกระบวนการแนะแนวการศึกษาต่อ งานวิจัยนี้จึงได้นำเสนอการใช้เทคนิคต้นไม้ตัดสินใจ C4.5 เพื่อค้นหากฎการตัดสินใจในการเลือกสาขาวิชา และนำกฎการตัดสินใจที่ได้มาพัฒนาเป็นระบบสนับสนุนการ ตัดสินใจในการเลือกสาขาวิชา เพื่อให้ผู้ที่จะสมัครเข้าศึกษาในระดับปริญญาตรี ได้ทำการทดลองการเลือกสาขาวิชาก่อนการเลือกสมัครจริง ผู้สมัครจะได้ทราบถึงสาขาวิชาที่เหมาะสมกับความรู้ และทักษะของผู้สมัครเองเมื่อเข้ามาศึกษาใน สาขาวิชานัÊน จากการวิจัยสามารถสร้างเป็นกฎการตัดสินใจได้จำนวนทั้งสิ้น 333 กฎ จากทั้งหมด 51 สาขาวิชาที่ทำการทดลองและสามารถสร้างระบบสนับสนุนการตัดสินใจในการเลือกสาขาวิชาได้

เนตรดาว ตันตรานนท์ (2558) การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความรู้ทางโภชนาการภาวะโภชนาการและภาวะ สุขภาพของผู้สูงอายุในชุมชนบ้านต้นผึ้ง ตำบลสันทรายอำเภอสารภี จังหวัดเชียงใหม่กลุ่มตัวอย่าง คือ ผู้สูงอายุในพื้นที่ดังกล่าวที่ถูกเลือกในการศึกษา จำนวน 60 คน เครื่องมือที่ใช้ในการศึกษาคือ แบบสอบถามผู้สูงอายุ เรื่องความรู้ทางโภชนาการ ภาวะโภชนาการและภาวะสุขภาพของ ผู้สูงอายุ ชุมชนบ้านต้นผึ้ง ตำบลสันทราย อำเภอสารภี จังหวัดเชียงใหม่ มี3 ส่วน คือ1) แบบสอบถามข้อมูล ทั่วไปรวมทั้งการบันทึกน้ำ หนักและส่วนสูงเพื่อประเมินภาวะโภชนาการ 2) แบบวัดความรู้ทาง โภชนาการ 3)แบบสอบถามภาวะสุขภาพ ตรวจสอบความตรงตามเนื้อหาโดยผู้ทรงคุณวุฒิ 3 ท่าน พบว่า มีค่าดัชนีความสอดคล้องระหว่างคำถามกับวัตถุประสงค์มากกว่า 0.5 ทุกข้อทดสอบความเชื่อมั่นโดยใช้สูตรสัมประสิทธิ์อัลฟาของครอนบาค มีค่าเท่ากับ 0.75 วิเคราะห์ข้อมูลด้วยการแจก แจงความถี่ค่าร้อยละค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน

ผลการศึกษาพบว่า

1. ความรู้ทางโภชนาการของกลุ่มตัวอย่างส่วนใหญ่ (ร้อยละ 80.00) อยู่ในระดับมาก
2. ภาวะโภชนาการของกลุ่มตัวอย่างส่วนใหญ่(ร้อยละ 61.70)อยู่ในเกณฑ์สมส่วน

3. ภาวะสุขภาพโดยรวมของกลุ่มตัวอย่างอยู่ในระดับดี(ค่าเฉลี่ย= 2.24) ในขณะที่ ภาวะสุขภาพรายด้านทุกด้านทั้งด้านร่างกาย จิตใจ สังคมและจิตวิญญาณ อยู่ในระดับดีโดยมี ค่าเฉลี่ยอยู่ที่1.65, 2.23, 2.34และ2.73 ตามลำดับ

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM ไปประยุกต์ใช้ในการชี้วัดเกรดเฉลี่ยและการรับประทานอาหาร โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม RapidMiner Studio 9.5.1 ในการสร้างโมเดล Decision Tree เพื่อทำการเปรียบเทียบ และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์