

บทที่ 2

แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลโรคหัวใจ และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

- 2.1.1 แนวคิดเกี่ยวกับฐานข้อมูลขนาดใหญ่ (Big Data)
- 2.1.2 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)
- 2.1.3 แนวคิดเกี่ยวกับธุรกิจอัจฉริยะ (Business Intelligence: BI)
- 2.1.4 แนวคิดเกี่ยวกับเหมืองข้อมูล (Data mining)
- 2.1.5 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

2.2 ทฤษฎี

- 2.2.1 ทฤษฎีเกี่ยวกับระบบจัดการฐานข้อมูล
- 2.2.2 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์
- 2.2.3 ทฤษฎีเกี่ยวกับการออกแบบกราฟิกด้วยโปรแกรม (Photoshop CS6)
- 2.2.4 ทฤษฎีเกี่ยวกับการ visualization
- 2.2.5 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ (Decision Tree)

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

- 2.3.1 โปรแกรม Rapid Miner Studio
- 2.3.2 โปรแกรม Tableau Public
- 2.3.3 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)
- 2.3.4 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับฐานข้อมูลขนาดใหญ่ (Big Data)

ปัจจุบันที่เทคโนโลยีดิจิทัลเติบโตอย่างรวดเร็วมีจำนวนมากขึ้น และได้เข้ามามีส่วนสำคัญกับการใช้ชีวิตประจำวันของผู้คนผ่านอุปกรณ์ดิจิทัลต่าง ๆ เหล่านี้ทั้งในขณะที่ทำงาน และในขณะที่ใช้ชีวิตประจำวันเกิดการสร้างแล้วจัดเก็บข้อมูล มหาศาลขึ้น เช่น เอกสารที่เกิดขึ้นจากการทำงาน การขออนุญาตเครือข่ายสังคมออนไลน์ (Social Network) ไลน์ (Line) ทวิตเตอร์ (Twitter) เฟซบุ๊ก (Facebook)

2.1.1.1 ความหมายของฐานข้อมูลขนาดใหญ่

จิริสิน กิตานุวัฒน์ (2558) กล่าวว่า นิยามของ Big Data คือ ข้อมูลที่มีขนาดและความซับซ้อนของข้อมูลมากอาจได้มาจากแหล่งข้อมูลหลายแหล่งรวมกันจนมีขนาดใหญ่ หรืออาจมาจากแหล่งข้อมูลเดียวกันในปริมาณมาก เช่น ข้อมูลราคาสินค้า มาจากห้างสรรพสินค้าหลายห้างหรือข้อมูลทางอุตุนิยมวิทยาต้องมีข้อมูลหลายแหล่งรวมกัน

สุกิจ คูชัยสิทธิ (2556) กล่าวว่า นิยามของ Big Data หมายถึง ปริมาณข้อมูลที่มีขนาดใหญ่มหาศาลเกินกว่าขีดความสามารถในการประมวลผลของระบบฐานข้อมูลธรรมดาที่จะรองรับได้ปริมาณข้อมูลที่มีขนาดใหญ่มาก ๆ จะมีอัตราการเพิ่มข้อมูลได้อย่างรวดเร็วมากและจะมีรูปแบบที่ไม่มีโครงสร้างหรือกึ่งโครงสร้าง ซึ่งไม่สามารถอยู่ในระบบฐานข้อมูลที่จะจัดเก็บข้อมูลได้

2.1.1.2 คุณลักษณะของฐานข้อมูลขนาดใหญ่

สุกิจ คูชัยสิทธิ (2556) กล่าวว่า Big Data เปรียบเสมือนกลุ่มของก้อนเมฆ (Cloud) ขนาดใหญ่ที่มีข้อมูลจำนวนมากที่รวมกันอยู่ อาทิ ข้อมูลสังคมเครือข่ายข้อมูลการบริการทางเว็บ (Web Server Logs) ข้อมูลจากอุปกรณ์ตรวจตราการจราจร (Traffic Flow Sensors) ข้อมูลภาพถ่ายดาวเทียม (Satellite Imagery) ข้อมูลด้านการกระจายเสียง ข้อมูลธุรกรรมทางธนาคาร (Banking Transaction) ข้อมูลด้านการตลาดการเงิน (Financial Market Data) และข้อมูลการสื่อสารจากโทรศัพท์เคลื่อนที่ (Telemetry from Automobiles) เป็นต้น ดังนั้นคุณลักษณะของบิ๊กดาต้า ได้มีการจัดแบ่งออกเป็น 3 ลักษณะ

1) ปริมาตร (Volume) หมายถึง ข้อมูลที่มีปริมาณมหาศาลซึ่งโครงสร้างข้อมูลของระบบฐานข้อมูลไม่สามารถจัดเก็บข้อมูลได้ปริมาณข้อมูลมหาศาลมีประโยชน์เพื่อเป็นข้อมูลที่ใช้ในการตัดสินใจหรือทำนายอนาคต หรือเพื่อเตรียมการวางแผนการทำงานเชิงรุกทางธุรกิจ

2) ความเร็ว (Velocity) หมายถึง อัตราการเพิ่มขึ้นของข้อมูล ข้อมูลที่เข้าสู่ระบบฐานข้อมูลจะมีอัตราการเพิ่มขึ้นอย่างรวดเร็ว เช่น ข้อมูลที่เกิดขึ้นจากโทรศัพท์เคลื่อนที่ที่ผู้ใช้งานขึ้นเก็บเป็นข้อมูลภาพถ่าย ข้อมูลการพิมพ์การสนทนา ข้อมูลการอัปเดตวิดีโอหรือข้อมูลการอัดเสียงหรือแม้กระทั่งข้อมูลการสั่งซื้อสินค้าการขนส่ง และการบริการต่าง ๆ

3) รูปแบบที่หลากหลาย (Variety) หมายถึง รูปแบบมีความหลากหลายของรูปแบบข้อมูล ซึ่งอาจจะเป็นรูปแบบที่มีโครงสร้างไม่มีโครงสร้าง และกึ่งมีโครงสร้าง เป็นต้น รูปแบบที่ไม่มีโครงสร้างหรือกึ่งโครงสร้างจะไม่เหมือนข้อมูลที่เรাজัดเก็บไว้ในระบบฐานข้อมูล เช่น ข้อความ อีเมลรูปภาพ วิดีโอและเสียง เป็นต้น ซึ่งข้อมูลเหล่านี้มีความซับซ้อนและเชื่อมโยงกัน

2.1.2 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

เป็นศาสตร์ในการวิเคราะห์และประมวลผลข้อมูลด้วยวิธีการทางคณิตศาสตร์และสถิติ นำผลนั้นมาปรับใช้ให้เกิดประโยชน์ ความรู้เรื่องดังกล่าวสำคัญยิ่งขึ้นเมื่อเทคโนโลยีในปัจจุบันมีการเก็บข้อมูลมหาดศาลในระดับ Big Data ที่มากเกินไปจนสามารถใช้งานได้เต็มที่

2.1.2.1 ความหมายการวิเคราะห์ข้อมูล (Data analytic)

Data Analytics นั้นเป็น Business Intelligence อย่างหนึ่ง ซึ่งเป็นศาสตร์ของการใช้ข้อมูลต่าง ๆ จากที่ต่าง ๆ มาร่วมวิเคราะห์รวมกัน เพื่อปรับปรุงธุรกิจ หรือการตลาด ซึ่ง Data Analytics นั้นเป็น Business Analytics อย่างหนึ่ง โดย Business Analytics เป็นวิธีการที่ใช้ข้อมูลเทคโนโลยีด้านสารสนเทศ เข้ามาทำการวิเคราะห์ข้อมูล เพื่อช่วยในการดำเนินธุรกิจ ซึ่ง Business Analytics แบ่งเป็น 3 แบบคือ

1) Descriptive Analytics เป็นรูปแบบการใช้ข้อมูลแบบพื้นฐานที่สุด โดยเน้นการอธิบายว่ากำลังเกิดขึ้น หรืออาจจะเกิดอะไรขึ้น สามารถอธิบายถึงสาเหตุการเกิดต่าง ๆ ได้ว่าทำไม ซึ่ง Descriptive Analytics ตัวอย่างคือรายงานธุรกิจ รายงานด้านการทำ Campaign หรือโฆษณา หรือรายงานผลดำเนินงานที่ผ่านมา เป็นข้อมูลพื้นฐานที่แสดงผลในการทำกิจกรรมต่าง ๆ ให้เราได้รับทราบ

2) Predictive Analytics เป็นรูปแบบการใช้ข้อมูลที่มีความซับซ้อนขึ้นมา โดยจะเป็นการ “พยากรณ์” หรือ “ทำนาย” สิ่งที่กำลังเกิดขึ้น โดยใช้ข้อมูลในอดีต ร่วมกับโมเดลทางคณิตศาสตร์ต่าง ๆ หรือร่วมกับการทำ Data Mining นอกจากนี้ Predictive Analytics ยังทำให้เราสามารถวิเคราะห์หาโอกาส และความเสี่ยงต่าง ๆ ที่จะเกิดขึ้นในอนาคตได้ด้วย

เช่น การรู้เทรนด์ทางการตลาด การพยากรณ์ยอดขายหรือการทำ Campaign ว่าจะมีคนร่วมเท่าไร หรือมี ROI เท่าไร

3) Prescriptive Analytics เป็นรูปแบบการวิเคราะห์ข้อมูลที่มีความซับซ้อนและยากที่สุด เพราะไม่เพียงพยากรณ์หรือทำนายว่าอะไรจะเกิดขึ้น แต่ยังให้คำแนะนำในทางเลือกต่าง ๆ และผลแต่ละทางเลือกว่าจะมี Pros & Cons อย่างไร โมเดลของ Prescriptive Analytics นั้นจะสามารถปรับเปลี่ยนได้ตามข้อมูลที่เพิ่มเติมเข้ามามากขึ้น และ Prescriptive Analytics นี้ยังเป็นการใช้ข้อมูลที่มากที่สุด และเกี่ยวพันกับเรื่อง Big Data เป็นอย่างมาก

2.1.2.2 ประโยชน์ของการวิเคราะห์ข้อมูล (Data analytic)

1) ช่วยให้ระบบสามารถทำงานอัตโนมัติ เช่น การวิเคราะห์ เพื่อหาสิ่งผิดปกติดด้วยตัวเอง

2) การหาคำตอบ เช่น การคาดเดาอนาคต จากข้อมูลสถิติ

3) การหาช่องว่างทางการตลาด

4) เป็นข้อมูลในการสร้าง Product

5) การเปรียบเทียบข้อมูลเพื่อหาช่องว่าง และแนวทางการพัฒนา

2.1.3 แนวคิดเกี่ยวกับธุรกิจอัจฉริยะ (Business Intelligence: BI)

Business Intelligence (BI) หรือระบบธุรกิจอัจฉริยะ เป็นระบบที่ใช้ในการพยากรณ์อนาคตของธุรกิจ ช่วยในการตัดสินใจวางแผนกลยุทธ์เพื่อให้บรรลุตามเป้าประสงค์ BI เป็นเหมือนฐานข้อมูลขนาดใหญ่ ที่เก็บรวบรวมข้อมูลไว้ในลักษณะที่เอื้อต่อการนำข้อมูลไปใช้ในสนับสนุนการตัดสินใจซึ่งจะประกอบไปด้วยระบบข้อมูล และโปรแกรมแอปพลิเคชันด้านการวิเคราะห์มากมายหลายระบบ องค์ประกอบสำคัญของ BI ได้แก่ คลังข้อมูล (Data Warehouse) ดาต้ามาร์ท (Data Mart) การทำเหมืองข้อมูล (Data Mining) เทคนิคการวิเคราะห์ข้อมูลหลายมิติ เพื่อใช้เป็นข้อมูลพื้นฐานสำหรับการนำไปวิเคราะห์และเป็นข้อมูลในการบริหาร สามารถนำข้อมูลที่ได้จากระบบมาใช้เพื่อประเมินประสิทธิภาพการทำงานของบริษัท

2.1.3.1 การนำระบบธุรกิจอัจฉริยะไปใช้

1) การจัดทำประวัติของลูกค้า

2) การประเมินถึงสภาพของตลาด

3) การจัดกลุ่มของตลาด

4) การจัดลำดับทางด้านเครดิต

5) การเพิ่มความสามารถในการทำกำไรของผลิตภัณฑ์

6) การจัดการความเสี่ยงของสินค้าคงคลัง

2.1.3.2 จุดเด่นของธุรกิจอัจฉริยะ

- 1) สามารถดึงข้อมูลจากฐานข้อมูลที่หลากหลายภายในองค์กรมาทำการวิเคราะห์ เช่น Excel, FoxPro, Dbase, Access โดยไม่มีการเขียนโปรแกรมเพิ่มเติมใด ๆ
- 2) ใช้งานง่ายสะดวกสบาย เพียงแค่คลิกเมาส์ก็สามารถเปลี่ยนแปลงรายงานได้โดยไม่ต้องมีการคีย์ข้อมูลใหม่

2.1.4 แนวคิดเกี่ยวกับเหมืองข้อมูล (Data mining)

ศาสตราจารย์คุณหญิงสุชาดา กิระนันท์ (2545) กล่าวว่า Data mining คือ กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์ หรือ สร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยต่าง ๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล การทำเหมืองข้อมูลประกอบขึ้นด้วยการนำกระบวนการทางสถิติ และการเรียนรู้ผ่านระบบคอมพิวเตอร์ เพื่อสร้างตัวแบบกฎเกณฑ์รูปแบบการพยากรณ์และข้อความรู้จากฐานข้อมูลขนาดใหญ่ โดยการทำเหมืองข้อมูลมีขั้นตอนการดำเนินงานหลายขั้นตอนซึ่งต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น วิธีการจัดกลุ่มการค้นหาความสัมพันธ์ การพยากรณ์ เป็นต้น การดำเนินงานมักอยู่ในลักษณะของการสร้างตัวแบบ (modeling) ที่อธิบายความเป็นไปหรือสภาพการณ์หนึ่งที่เกิดขึ้นแล้ว หรือที่ทราบคำตอบ แล้วนำตัวแบบนี้มาใช้อธิบายสถานการณ์ที่ยังไม่เกิดขึ้น หรือที่ไม่ทราบคำตอบ ตัวแบบเหล่านี้อาจเป็นตัวแบบที่เรียบง่ายไปจนถึงตัวแบบที่ยุ่งยากซับซ้อน และอาจใช้การผสมผสานแนวคิดหรือเครื่องมือต่าง ๆ เข้าด้วยกัน เพื่อที่จะสามารถสกัดข้อความรู้ที่อยู่ในข้อมูลขนาดใหญ่ได้โดยใช้เทคโนโลยีคลังข้อมูล (Data Warehouse) เข้ามาช่วยในการจัดการข้อมูลเพื่อเพิ่มประสิทธิภาพของการทำเหมืองข้อมูล ดังนั้น ถ้ามีฐานข้อมูลขนาดใหญ่ที่มีข้อมูลคุณภาพดี เทคโนโลยีการทำเหมืองข้อมูลจะช่วยให้การค้นหาหรือแสวงหาโอกาสทางธุรกิจใหม่ โดยการทำเหมืองข้อมูลจะก่อให้เกิดกระบวนการอัตโนมัติในการค้นพบสารสนเทศหรือข้อความรู้ในฐานข้อมูลขนาดใหญ่ ด้วยการใช่วิธีการ เช่น การพยากรณ์แนวโน้มและพฤติกรรม การบริโภคแบบอัตโนมัติ หรือเกิดกระบวนการอัตโนมัติในการค้นพบรูปแบบที่ไม่เคยรู้จักมาก่อน ด้วยการใช่วิธีการค้นหาเข้าไปในรายละเอียดของฐานข้อมูล เพื่อหารูปแบบที่ซ่อนอยู่ในฐานข้อมูลนั้น

พิพัฒน์ เกียรติกรมรัตน์ (2560) กล่าวว่า Data mining คือ เหมืองข้อมูล คล้ายกับเหมืองแร่ ที่ขุดดินมาเยอะแต่ได้แค่นิดเดียว Data Mining เป็นศาสตร์แขนงหนึ่งทาง AI (Artificial Intelligence) ดังนั้นข้อมูลที่จะพูดถึงต่อไป ท่านที่เคยศึกษาด้าน AI (Artificial Intelligence)

มาบ้างอาจคุ้นเคยเพราะหลายเทคนิคใน AI (Artificial Intelligence) ได้นำมาใช้กับ Data Mining ซึ่งเป็นการค้นหาความสัมพันธ์ และรูปแบบทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูล ซึ่งสัมพันธ์ และรูปแบบเหล่านั้นได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมากที่มีอยู่ Data Mining จะทำการสำรวจ และวิเคราะห์ข้อมูลให้อยู่ในรูปแบบที่เต็มไปด้วยความหมาย และอยู่ในรูปของกฎ โดยความสัมพันธ์หน่วยนี้แสดงให้เห็นถึงความรู้ต่าง ๆ (Knowledge) ที่มีประโยชน์ในฐานข้อมูล ในปัจจุบันองค์กรธุรกิจส่วนใหญ่เผชิญกับ ปัญหาของข้อมูลดิบจำนวนมากแต่ข้อมูลที่ ประยุกต์ใช้ได้มีน้อย Data Mining จึงเป็นสาขาที่คาดว่าจะเป็นที่รู้จัก และนำมาใช้ ประยุกต์ใช้ อย่างแพร่หลายเนื่องจาก Data Mining สามารถดึงความรู้ออกมาจากข้อมูลจำนวน มากที่ถูกเก็บสะสมและซ่อนไว้ ในโลกปัจจุบันธุรกิจต่าง ๆ จะพยายามหาเทคนิคที่สามารถนำ ความสำเร็จมาสู่บริษัท เช่น ในโลกธุรกิจขนาดย่อมจะสร้างความสัมพันธ์กับลูกค้า โดยสังเกต จากความต้องการความชอบและความสนใจของลูกค้า และอาจมีการเรียนรู้ได้จากผลสะท้อน ในอดีต ว่าจะทำอย่างไรให้การบริการลูกค้ามีประสิทธิภาพดีขึ้นในอนาคต เช่น บริษัทที่เป็นผู้ ออกบัตรเครดิต และธนาคารต่าง ๆ จะมีขบวนการที่ใช้ Data Mining ให้เป็นประโยชน์ในการ ตัดสินใจว่าลูกค้ากลุ่มใดเป็นกลุ่มที่ดีทำความเข้าใจลูกค้า ช่วยในการแยกประเภทของลูกค้า และจะทำนายกลุ่มของประชากรที่คาดว่าจะมาเป็นลูกค้าในอนาคต เป็นต้น อย่างไรก็ตามการ เรียนรู้นั้นต้องมากกว่าการเก็บสะสมข้อมูลอย่างตรงไปตรงมา ซึ่งจะทำให้การทำงานเป็นไป อย่างมีประสิทธิภาพวัฏจักรขั้นตอนการทำงานของ Data Mining วัฏจักรขั้นตอนการทำงานของ Data Mining ประกอบไปด้วย 4 ขั้นตอนหลัก ๆ ดังนี้

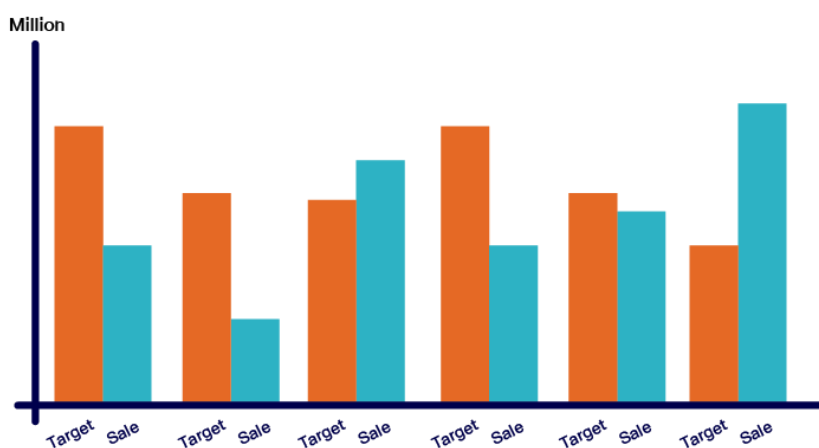
- 1) การระบุโอกาสทางธุรกิจหรือการระบุปัญหาที่เกิดขึ้นกับธุรกิจเป็นการระบุ ขอบเขตของข้อมูลที่จะนำมาทำการวิเคราะห์ เพื่อหาความได้เปรียบทางการตลาดหรือเพื่อนำมา ทำการแก้ไข้ปัญหา
- 2) ส่วนของ Data Mining เป็นการนำเทคนิคของ Data Mining ไปใช้ถ่ายทอดหรือ ทำการเปลี่ยนแปลงข้อมูลดิบให้อยู่ในรูปแบบของข้อมูลที่จะนำไปใช้ได้จริงในทางธุรกิจ
- 3) การปฏิบัติตามข้อมูลคือการนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วน Data Mining มาลองปฏิบัติจริงกับธุรกิจ
- 4) การวัดประสิทธิภาพจากผลลัพธ์การวัดประสิทธิภาพของเทคนิคของ Data Mining ที่จะนำมาใช้จากผลลัพธ์ ซึ่งสามารถตรวจสอบได้หลายทาง เช่น วัดจากส่วนแบ่ง ของตลาด วัดจากปริมาณลูกค้า หรือวัดจากกำไรสุทธิ เป็นต้น

2.1.5 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization คือ การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ ซึ่งคำว่า “ประสิทธิภาพ” ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้ว อาจทำให้เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์เชื่อมโยงได้

การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล ในปัจจุบันเป็นยุคเทคโนโลยีเข้าถึงทุกคน ทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่าย และรวดเร็วมากขึ้น คนที่นำเสนอข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว เพื่อให้การนำเสนอข้อมูลน่าสนใจ และมีประสิทธิภาพสูงสุด ซึ่งเลือกแผนภูมียอดนิยมที่ใช้กันอย่างแพร่หลาย ดังนี้

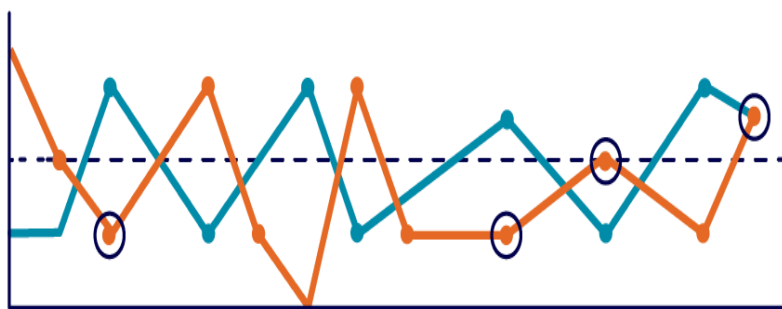
1) แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้งที่นิยมแสดงออกมาในรูปแบบสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับใช้การเปรียบเทียบจำนวนของข้อมูลในแต่ละชุดเช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริงเปรียบเทียบกับเป้าหมายยอดขายที่ตั้งไว้ เป็นต้นซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท แผนภูมิแท่งแบบจัดกลุ่ม แผนภูมิแท่งแบบจัดกลุ่ม เหมาะกับการนำเสนอข้อมูลที่มีข้อมูลย่อย ๆ อยู่ภายใต้ข้อมูลใหญ่เป็นการเน้นให้เห็นข้อมูลย่อยนั้น ๆ แผนภูมิแท่งแบบวางซ้อนกัน แผนภูมิแท่งแบบวางซ้อนกัน เหมาะกับการนำเสนอข้อมูลที่ต้องการให้เห็นข้อมูลย่อยในแต่ละข้อมูลใหญ่ และยังสามารถให้เห็นสัดส่วนของข้อมูลย่อยต่าง ๆ เหล่านั้นได้ด้วย ใช้แผนภูมินี้เมื่อมีชุดข้อมูล หลายชุด และต้องการเน้นผลรวมทั้งหมด



ภาพที่ 2.1 แผนภูมิแท่ง (Bar Charts)

ที่มา : <https://bit.ly/3aXjpLg>

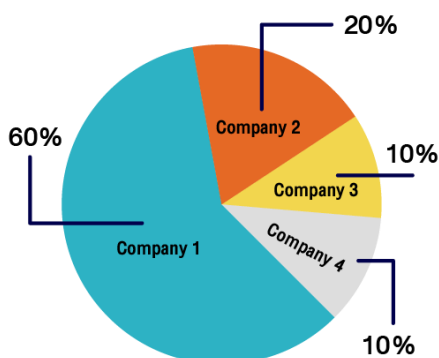
2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั้นเอง แผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วง ใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้เช่น ข้อมูลของยอดขายรายปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดูแนวโน้ม เป็นต้น



ภาพที่ 2.2 แผนภูมิเส้น (Line Charts)

ที่มา : <https://bit.ly/3aXjpLg>

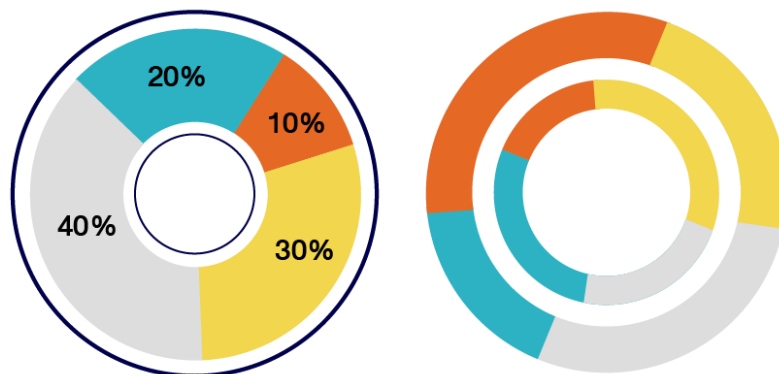
3) แผนภูมिवงกลม (Pie Charts) แผนภูมिवงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกัน อาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยาก เพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูลแสดงส่วนผลสมต่าง ๆ เป็นต้น



ภาพที่ 2.3 แผนภูมिवงกลม (Pie Charts)

ที่มา : <https://bit.ly/3aXjpLg>

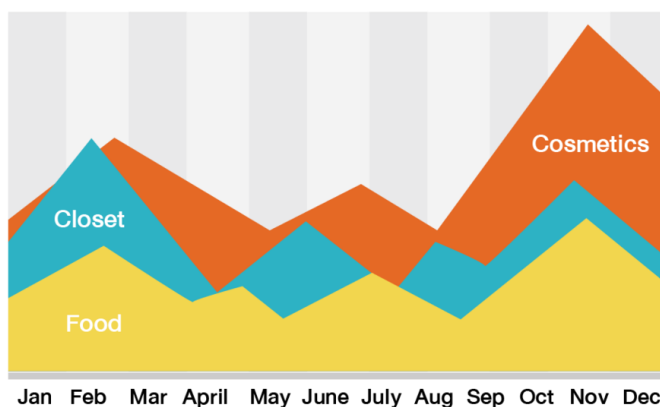
4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลาย ๆ ชั้น



ภาพที่ 2.4 แผนภูมิโดนัท (Doughnut Charts)

ที่มา : <https://bit.ly/3aXjpLg>

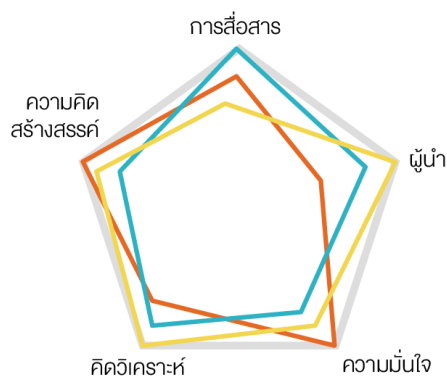
5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่ละเดือน ตามหมวดหมู่ต่าง ๆ ไล่ไป เครื่องสำอาง เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

ที่มา : <https://bit.ly/3aXjpLg>

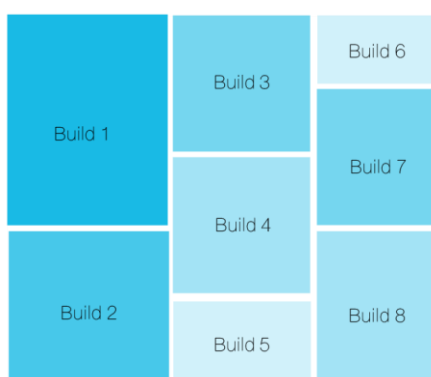
6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อนจุดแข็งของข้อมูลเช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อนจุดแข็งของแต่ละคน เป็นต้น



ภาพที่ 2.6 แผนภูมิเรดาร์ (Radar Charts)

ที่มา : <https://bit.ly/3aXjpLg>

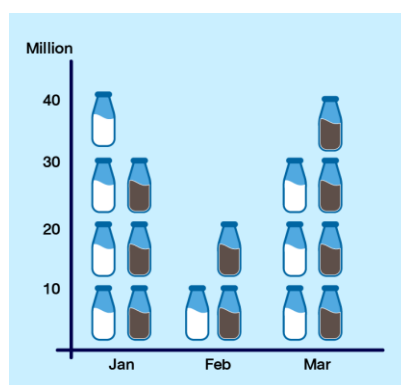
7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ที่แสดงผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่แสดงพื้นที่สีที่แตกต่างกันได้



ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

ที่มา : <https://bit.ly/3aXjpLg>

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอนและแกนตั้ง แต่เลือกใช้รูปภาพ หรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ เช่น การแสดงผลจำนวนของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละเดือนซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง แต่เปลี่ยนจากแท่งเป็นรูปภาพของนม 2 รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะนี้ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคน อาจจะใช้รูปภาพไอคอนคนหรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงินหรือแบงค์แทน ก็จะทำให้คนดูเข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคยอยู่แล้วและยังดึงดูดความสนใจได้มากกว่าการใช้กราฟแท่งสีเหลี่ยม



ภาพที่ 2.8 แผนภูมิรูปภาพ (Picture Graph)

ที่มา : <https://bit.ly/3aXjpLg>

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับระบบจัดการฐานข้อมูล

2.2.1.1 ความหมายระบบจัดการฐานข้อมูล

ระบบจัดการฐานข้อมูล (Database Management System) หรือที่เรียกว่าดีบีเอ็มเอส (DBMS) คือซอฟต์แวร์สำหรับบริหาร และจัดการฐานข้อมูล เปรียบเสมือนสื่อกลางระหว่างผู้ใช้และโปรแกรมต่าง ๆ ที่เกี่ยวข้องกับการใช้ฐานข้อมูล ซึ่งมีหน้าที่ช่วยให้ผู้ใช้เข้าถึงข้อมูลได้ง่ายสะดวกและมีประสิทธิภาพ การเข้าถึงข้อมูลของผู้ใช้ อาจเป็นการสร้างฐานข้อมูล การแก้ไขฐานข้อมูล หรือการตั้งคำถามเพื่อให้ได้ข้อมูลมาโดยผู้ใช้ไม่จำเป็นต้องรับรู้เกี่ยวกับรายละเอียดภายในโครงสร้างของฐานข้อมูล เปรียบเสมือนเป็นสื่อกลางระหว่างผู้ใช้และโปรแกรมต่าง ๆ ที่เกี่ยวข้องกับการใช้ฐานข้อมูลซึ่งต่างจากระบบแฟ้มข้อมูลที่หน้าที่เหล่านี้จะเป็นหน้าที่ของโปรแกรมเมอร์

2.2.1.2 หน้าที่ของระบบการจัดการฐานข้อมูล

- 1) แปลงคำสั่งที่ใช้จัดการกับข้อมูลภายในฐานข้อมูล ให้อยู่ในรูปแบบที่ฐานข้อมูลเข้าใจ
- 2) นำคำสั่งต่าง ๆ ซึ่งได้รับการแปลแล้ว ไปสั่งให้ฐานข้อมูลทำงาน เช่น การเรียกใช้ (Retrieve) จัดเก็บ (Update) ลบ (Delete) เพิ่มข้อมูล (Add) เป็นต้น
- 3) ป้องกันความเสียหายที่จะเกิดขึ้นกับข้อมูลภายในฐานข้อมูล โดยจะคอยตรวจสอบว่าคำสั่งใดที่สามารถทำงานได้ และคำสั่งใดที่ไม่สามารถทำงานได้
- 4) รักษาความสัมพันธ์ของข้อมูลภายในฐานข้อมูลให้มีความถูกต้องอยู่เสมอ
- 5) เก็บรายละเอียดต่าง ๆ ที่เกี่ยวข้องกับข้อมูลภายในฐานข้อมูลไว้ในพจนานุกรมข้อมูล (Data Dictionary) ซึ่งรายละเอียดเหล่านี้มักจะถูกเรียกว่า เมทาดาต้า (Metadata) ซึ่งหมายถึง "ข้อมูลของข้อมูล"
- 6) ดูแลการใช้งานให้กับผู้ใช้ ในการติดต่อกับตัวจัดการระบบแฟ้มข้อมูลได้ โดยจะทำหน้าที่ติดต่อกับระบบแฟ้มข้อมูลซึ่งเสมือนเป็นผู้จัดการแฟ้มข้อมูล (file manager) นำข้อมูลจากหน่วยความจำสำรองเข้าสู่หน่วยความจำหลักเฉพาะส่วนที่ต้องการใช้งาน และทำหน้าที่ประสานกับตัวจัดการระบบแฟ้มข้อมูลในการจัดเก็บ เรียกใช้ และแก้ไขข้อมูล
- 7) ควบคุมการใช้ข้อมูลพร้อมกัน (Concurrency Control) ในระบบคอมพิวเตอร์ที่ใช้อยู่ปัจจุบัน โปรแกรมการทำงานมักจะเป็นแบบผู้ใช้หลายคน (Multi User) จึงทำให้ผู้ใช้แต่ละคนสามารถเรียกใช้ข้อมูลได้พร้อมกัน ระบบจัดการฐานข้อมูลที่มีคุณสมบัติควบคุมการใช้ข้อมูลพร้อมกันนี้ จะทำการควบคุมการใช้ข้อมูลพร้อมกันของผู้ใช้หลายคนในเวลาเดียวกันได้ โดยมีระบบการควบคุมที่ถูกต้องเหมาะสม เช่น ถ้าการแก้ไขข้อมูลนั้นยังไม่เรียบร้อย ผู้ใช้อื่น ๆ ที่ต้องการเรียกใช้ข้อมูลนี้ จะไม่สามารถเรียกข้อมูลนั้น ๆ ขึ้นมาทำงานใด ๆ ได้ ต้องรอจนกว่าการแก้ไขข้อมูลของผู้ที่เรียกใช้ข้อมูลนั้นก่อนจะเสร็จเรียบร้อย จึงจะสามารถเรียกข้อมูลนั้นไปใช้งานต่อได้ ทั้งนี้เพื่อป้องกันไม่ให้เกิดปัญหาการเรียกใช้ข้อมูลที่ไม่ถูกต้อง
- 8) ควบคุมระบบความปลอดภัยของข้อมูลโดยป้องกันไม่ให้ผู้ที่ไม่ได้รับอนุญาตเข้ามาเรียกใช้หรือแก้ไขข้อมูลในส่วนป้องกันเอาไว้ พร้อมทั้งสร้างฟังก์ชันในการจัดทำข้อมูลสำรอง

9) ควบคุมการใช้ข้อมูลในสภาพที่มีผู้ใช้พร้อม ๆ กันหลายคน โดยจัดการเมื่อมีข้อผิดพลาดของข้อมูลเกิดขึ้น

2.2.2 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.2.1 หลักในการออกแบบเว็บไซต์

เว็บไซต์เป็นสื่อที่ได้รับความนิยมอย่างมากบนอินเทอร์เน็ต ซึ่งเว็บไซต์เป็นสื่อที่อยู่ในความควบคุมของผู้ใช้โดยสมบูรณ์ กล่าวคือ ผู้ใช้สามารถตัดสินใจเลือกได้ว่าจะดูเว็บไซต์ใดและจะไม่เลือกดูเว็บไซต์ใด ได้ตามต้องการ จึงทำให้ผู้ใช้ไม่มีความอดทนต่ออุปสรรคและปัญหาที่เกิดจากการออกแบบเว็บไซต์ผิดพลาดถ้าผู้ใช้เห็นว่าเว็บที่กำลังดูอยู่นั้นไม่มีประโยชน์ต่อตัวเขา หรือไม่เข้าใจว่าเว็บไซต์นี้จะใช้งานอย่างไร เขาก็สามารถที่จะเปลี่ยนไปดูเว็บไซต์อื่น ๆ ได้อย่างรวดเร็ว เนื่องจากในปัจจุบันมีเว็บไซต์อยู่มากมาย และยังมีเว็บไซต์ที่เกิดขึ้นใหม่ ๆ ทุกวัน ผู้ใช้จึงมีทางเลือกมากขึ้น และสามารถเปรียบเทียบคุณภาพของเว็บไซต์ต่าง ๆ ได้เองเว็บไซต์ที่ได้รับการออกแบบอย่างสวยงาม มีการใช้งานที่สะดวก ย่อมได้รับความนิยมจากผู้ใช้ มากกว่าเว็บไซต์ที่ดูสับสนวุ่นวาย มีข้อมูลมากมายแต่หาอะไรไม่เจอ นอกจากนี้ยังใช้เวลาในการแสดงผลแต่ละหน้านานเกินไป ซึ่งปัญหาเหล่านี้ล้วนเป็นผลมาจากการออกแบบเว็บไซต์ไม่ดีทั้งสิ้น ดังนั้น การออกแบบเว็บไซต์จึงเป็นกระบวนการสำคัญในการสร้างเว็บไซต์ ให้ประทับใจผู้ใช้ ทำให้เขาอยากกลับมาเยี่ยมชมอีกในอนาคต ซึ่งนอกจากต้องพัฒนาเว็บไซต์ที่ดีมีประโยชน์แล้ว ยังต้องคำนึงถึงการแข่งขันกับเว็บไซต์อื่น ๆ อีกด้วยการออกแบบเว็บไซต์ จึงต้องควรคำนึงถึง

1) ความเรียบง่าย เข้าใจง่าย

การออกแบบเว็บไซต์ที่ดี จะต้องเน้นที่ความเรียบง่ายเป็นหลัก โดยเลือกนำเสนอเฉพาะสิ่งที่ต้องการนำเสนอจริง ๆ ในรูปแบบที่หลากหลาย โดยอาจจะเป็นสีสัน กราฟิก ภาพเคลื่อนไหวหรือตัวอักษร ที่สำคัญจะต้องมีการนำเสนอที่ไม่ดูรกหน้าเว็บจนเกินไป เพื่อไม่ให้เกิดความรู้สึกรกสยดสา หรือสร้างความเบื่อหน่าย นำราคาขายให้กับผู้ที่เข้าชมเว็บไซต์

2) ความสม่ำเสมอ ไม่สับสน

ควรออกแบบเว็บไซต์ด้วยความสม่ำเสมอ คือจะต้องมีรูปแบบ กราฟิก โทนมสี และการตกแต่งต่าง ๆ ให้แต่ละหน้าบนเว็บไซต์มีความคล้ายคลึงกัน และเป็นแนวเดียวกันไปตลอดทั้งเว็บไซต์ ดังตัวอย่างเว็บไซต์ทั่ว ๆ ไปที่จะสังเกตเห็นได้ว่าทุกหน้าของเว็บไซต์นั้นจะเน้นการตกแต่งในรูปแบบเดียวกันทั้งหมด ต่างก็แค่การนำเสนอของแต่ละหน้าเท่านั้น

3) สร้างความโดดเด่น เป็นเอกลักษณ์

การออกแบบเว็บไซต์เพื่อให้สามารถสื่อถึงจุดประสงค์ในการนำเสนอเว็บได้ดี จะต้องมีการสร้างความเป็นเอกลักษณ์และจุดเด่นให้กับเว็บไซต์ เพื่อให้สามารถสะท้อนถึงลักษณะขององค์กรได้มากที่สุด โดยการสร้างเอกลักษณ์ดังกล่าวนั้น อาจใช้ชุดสี รูปภาพ ตัวอักษรหรือกราฟิก นอกจากนี้ก็ต้องขึ้นอยู่กับว่า เป็นเว็บไซต์แบบทางการหรือไม่ เพื่อจะได้ออกแบบได้อย่างเหมาะสมที่สุด

4) เนื้อหาต้องดี ครบถ้วน

เนื้อหาเป็นสิ่งที่สำคัญที่สุดของการสร้างเว็บไซต์ เพราะสิ่งที่ทำให้ผู้คนเกิดความสนใจ และหมั่นติดตามเว็บไซต์เหล่านั้นอยู่เสมอ ก็คือเนื้อหาที่มีความสมบูรณ์และน่าสนใจ นอกจากนี้จะต้องมีการปรับปรุง พัฒนาเนื้อหาบนเว็บให้มีความทันสมัยอยู่เสมอ รวมถึงข้อมูลต้องมีความถูกต้องที่สุด

5) ระบบเนวิเกชัน ใช้งานง่าย

ระบบเนวิเกชัน เป็นเสมือนป้ายบอกทางเพื่อให้ผู้ใช้งาน ไม่เกิดความสับสนในขณะที่ใช้งานเว็บไซต์ ซึ่งการออกแบบเนวิเกชันก็จะต้องเน้นที่ความเรียบง่าย ใช้งานสะดวก และมีความเข้าใจได้ง่าย ที่สำคัญจะต้องมีตำแหน่งการวางที่สม่ำเสมอเพื่อให้ดูเป็นแนวทางเดียวกัน ทำให้ผู้ใช้งานหรือผู้ชมรู้สึกประทับใจ และจดจำเว็บไซต์ได้ง่ายขึ้น ส่วนใครที่มีการนำกราฟิกมาใช้ในระบบเนวิเกชัน ก็จะต้องเลือกกราฟิกที่สามารถสื่อความหมายได้ดีเช่นกัน

6) คุณภาพของเว็บไซต์

เว็บไซต์ที่ดีจะต้องมีคุณภาพ ทั้งสิ่งที่ปรากฏให้เห็นบนเว็บไซต์ ไม่ว่าจะเป็นกราฟิก ชนิดตัวอักษร รูปภาพหรือสีสันทันทีใช้ เนื้อหาที่นำมาแสดงผล ซึ่งหากเว็บไซต์มีคุณภาพก็จะสร้างความน่าเชื่อถือ และเป็นจุดเด่นที่ทำให้ผู้คนส่วนใหญ่เกิดความสนใจได้ดี เพราะฉะนั้นห้ามละเลยในส่วนของคุณภาพเด็ดขาด

7) ความสะดวกในการใช้งาน

เว็บไซต์ควรให้ความสะดวกสบายแก่ผู้ใช้งานได้ดี คือจะต้องมีการแสดงผลได้ในทุกระบบปฏิบัติการ ไม่ว่าจะเป็นเว็บเบราว์เซอร์ คอมพิวเตอร์ โน้ตบุ๊กหรือบนโทรศัพท์มือถือ ที่สำคัญจะต้องมีความละเอียดของการแสดงผล และสามารถใช้งานได้โดยไม่มีปัญหาด้วย

8) ความคงที่ของการออกแบบ

การออกแบบเว็บไซต์ควรมีความคงที่ในการออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกัน และมีการเรียบเรียงเนื้อหาอย่างรอบคอบ ทำให้เว็บมีความน่าเชื่อถือ และดูมีคุณภาพ ช่วยสร้างความประทับใจให้กับผู้ใช้งานได้เป็นอย่างดี

9) ความคงที่ของการทำงาน

ระบบการทำงานบนเว็บไซต์จะต้องมีความคงที่ และสามารถใช้งานได้ดี ซึ่งนอกจากการออกแบบระบบการทำงานให้มีความทันสมัยและสร้างสรรค์แล้ว ก็จะต้องหมั่นตรวจสอบอยู่เสมอ เพราะหากระบบการใช้งานมีความผิดปกติก็จะได้แก้ปัญหาได้ทัน นอกจากนี้อาจมีการอัปเดตดีไซน์ให้ทันสมัยขึ้นบ่อย ๆ เพื่อให้ผู้ใช้งานรู้สึกสนุกไปกับการใช้งานเว็บไซต์

2.2.2.2 รูปแบบโครงสร้างของเว็บไซต์

การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบ และความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure) เป็นโครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยมจัดด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การเรียงลำดับตามตัวอักษร วรรณคดี สารานุกรม หรืออภิธานศัพท์ โครงสร้างแบบนี้ เหมาะกับเว็บไซต์ที่มีขนาดเล็ก เนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปทีละหน้า ทิศทางการเข้าสู่เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรง โดยมีปุ่มเดินหน้า-ถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาของตนเองได้ ทำให้เสียเวลาเข้าสู่เนื้อหา



ภาพที่ 2.9 เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure)

ที่มา : <https://bit.ly/2wxfC8t>

2) เว็บที่มีโครงสร้างแบบลำดับชั้น (Hierarchical Structure) เป็นวิธีที่ดีที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วนต่าง ๆ และมีรายละเอียดย่อย ๆ ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับ แผนภูมิองค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับโครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของ เว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดรวมจุดเดียว คือ โฮมเพจ (Homepage) และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง



ภาพที่ 2.10 เว็บที่มีโครงสร้างแบบลำดับชั้น (Hierarchical Structure)

ที่มา : <https://bit.ly/2wxfC8t>

3) เว็บที่มีโครงสร้างแบบตาราง (Grid Structure) โครงสร้างรูปแบบนี้มีความซับซ้อนมากกว่ารูปแบบที่ผ่านมา การออกแบบเพิ่มความยืดหยุ่น ให้แก่การเข้าสู่เนื้อหาของผู้ใช้ โดยเพิ่มการเชื่อมโยงซึ่งกันและกันระหว่างเนื้อหาแต่ละส่วน เหมาะแก่ การแสดงให้เห็นความสัมพันธ์กันของเนื้อหา การเข้าสู่เนื้อหาของผู้ใช้จะไม่ใช่เป็นลักษณะเชิงเส้นตรง เนื่องจากผู้ใช้สามารถเปลี่ยนทิศทางการเข้าสู่เนื้อหาของตนเองได้



ภาพที่ 2.11 เว็บที่มีโครงสร้างแบบตาราง (Grid Structure)

ที่มา : <https://bit.ly/2wxfC8t>

ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมีลักษณะที่เหมือนกัน และสามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมาบรรจุลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับโครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือกหัวข้อใดก็จะเข้าไปสู่หน้าเนื้อหา (Topic Page)

ที่แสดงรายละเอียดของหัวข้อนั้น ๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้า รายละเอียดของหัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำ โครงสร้างแบบ เรียงลำดับและแบบลำดับชั้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจสร้างความ ยุ่งยากในการเข้าใจได้ และอาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ใช้ได้เข้าใจถึงความสัมพันธ์ ระหว่างเนื้อหา ในส่วนของการ ออกแบบจำเป็นจะต้องมีการวางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้น ได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจเกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4) เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure) โครงสร้างประเภทนี้ จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกันได้หมด เป็นการสร้าง รูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วยตนเอง การเชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกัน ของแต่ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็น รูปแบบที่ไม่มีโครงสร้างที่แน่นอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัด เฉพาะเนื้อหาภายในเว็บนั้น ๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้



ภาพที่ 2.12 เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure)

ที่มา : <https://bit.ly/2wxfC8t>

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือ ไฮเปอร์มีเดีย กับข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้ จะปรากฏอยู่บริเวณใด บริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการ เพื่อเลือกที่จะเข้าไปสู่หน้าใด ๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการ ท่องเที่ยวบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือ ถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอเป็นการยากในการปรับปรุง นอกจากนี้การเชื่อมโยงระหว่าง ข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสน และเกิดปัญหาการคงค้างของหัวข้อ (Cognitive Overhead) ได้

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบกราฟิกด้วยโปรแกรม (Photoshop CS6)

โปรแกรม Adobe Photoshop CS6 คือ โปรแกรมที่สามารถสร้างออกแบบกราฟิกแก้ไขภาพเคลื่อนไหวรวมทั้งการออกแบบหน้าเว็บเพจซึ่งโปรแกรม Adobe Photoshop CS6 มีเครื่องมือเพื่อสนับสนุนการสร้างชิ้นงานประเภทต่าง ๆ ได้แก่ประเภทสิ่งพิมพ์งานนำเสนอ ตลอดจนการออกแบบเว็บเพจดังนั้นโปรแกรม Adobe Photoshop CS6 จึงเป็นโปรแกรมที่มีความนิยมสูงและเหมาะสมกับการสร้างชิ้นงานด้านกราฟิกการแก้ไขภาพและการออกแบบประเภทต่าง ๆ

2.2.3.1 ส่วนประกอบของโปรแกรม Photoshop CS6

- 1) แถบเมนู คือกลุ่มเมนูที่รวบรวมคำสั่งต่าง ๆ เพื่อควบคุมการทำงาน ประกอบด้วยเมนูต่าง ๆ ดังต่อไปนี้ Menu file คือแถบชุดคำสั่งสำหรับการจัดไฟล์ภาพ
- 2) Menu Edit คือแถบชุดคำสั่งที่ทำหน้าที่ในการแก้ไขปรับแต่งภาพชนิดงานต่าง ๆ ที่สร้างขึ้นเช่นการปรับขนาดของภาพการคัดลอกภาพการมองภาพการย้อนกลับการทำงาน เป็นต้น
- 3) Menu Image คือแถบชุดคำสั่งที่ทำหน้าที่จัดการและปรับแต่งไฟล์ภาพ เช่น สีแสงปรับแต่งภาพความคมชัดปรับเปลี่ยนโหมดให้เป็นขาวดำ เป็นต้น
- 4) Menu Layer คือแถบชุดคำสั่งที่ทำหน้าที่จัดการต่าง ๆ ของ Layer เช่น การลบเพิ่ม Layer และการย้ายตำแหน่งลำดับของ Layer เป็นต้น
- 5) Menu Select คือแถบชุดคำสั่งที่ทำหน้าที่จัดการกับการเลือกวัตถุหรือพื้นที่แบบต่าง ๆ เพื่อนำไปใช้ร่วมกับคำสั่งอื่นเช่นเรื่องพื้นที่เพื่อย้ายตำแหน่งเลือกพื้นที่เพื่อเปลี่ยนสี เป็นต้น
- 6) Menu Filter คือแถบชุดคำสั่งที่ใช้ปากและ effect พิเศษให้กับวัตถุ
- 7) Menu 3D คือ แถบชุดคำสั่งที่ใช้ในการสร้างภาพแบบสามมิติ
- 8) Menu View คือคำสั่งที่ใช้ปรับแต่งแก้ไขและแสดงผลให้หน้าจอโปรแกรมเช่น บรรทัดและการย่อ/ขยายของมุมมองภาพ
- 9) Menu Window คือแถบชุดคำสั่งที่เกี่ยวข้องกับการแสดงและการซ่อนของแถบเครื่องมือต่าง ๆ
- 10) Menu Help คือถ้าชื่อคำสั่งที่เกี่ยวข้องกับการแสดงและการสอนของแถบ Adobe Photoshop CS6

2.2.3.2 เมนูของพื้นที่ทำงาน Panel menu Panel (พาเนล) เป็นวินโดว์ย่อย ๆ ที่ใช้เลือกรายละเอียด หรือคำสั่งควบคุมการทำงานต่าง ๆ ของโปรแกรมใน Photoshop มีพาเนลอยู่เป็นจำนวนมาก เช่น พาเนล Color ใช้สำหรับเลือกสี พาเนล Layers ใช้สำหรับจัดการกับเลเยอร์ และพาเนล Info ใช้แสดงค่าสีตรงตำแหน่งที่ชี้เมาส์ รวมถึงขนาด ตำแหน่งของพื้นที่ที่เลือกไว้

2.2.3.3 พื้นที่ทำงาน Stage หรือ Canvas เป็นพื้นที่ว่างสำหรับแสดงงานที่กำลังทำอยู่

2.2.3.4 เครื่องมือที่ใช้งาน Tools panel หรือ Tools box Tool Panel (ทูลพาเนล) หรือกล่องเครื่องมือ จะประกอบไปด้วยเครื่องมือต่าง ๆ ที่ใช้ในการวาด ตกแต่ง และแก้ไขภาพ เครื่องมือเหล่านี้มีจำนวนมาก ดังนั้นจึงมีการรวมเครื่องมือที่ทำหน้าที่คล้าย ๆ กันไว้ในปุ่มเดียวกัน โดยจะมีลักษณะรูปลามเหลี่ยมอยู่บริเวณมุมด้านล่าง เพื่อบอกให้รู้ว่าในปุ่มนี้ยังมีเครื่องมืออื่นอยู่ด้วย

2.2.3.5 สิ่งที่ควบคุมเครื่องมือที่ใช้งาน Tools control menu หรือ Option bar Option Bar (ออปชันบาร์) เป็นส่วนที่ใช้ปรับแต่งค่าการทำงานของเครื่องมือต่าง ๆ โดยรายละเอียดในออปชันบาร์จะเปลี่ยนไปตามเครื่องมือที่เราเลือกจากทูลบ็อกซ์ในขณะนั้น เช่น เมื่อเราเลือกเครื่องมือ Brush (พู่กัน) บนออปชันบาร์จะปรากฏออปชันที่ใช้ในการกำหนดขนาดและลักษณะหัวแปรง, โหมดในการระบายความโปร่งใสของสี และอัตราการไหลของสี เป็นต้น

2.2.4 ทฤษฎีเกี่ยวกับการ visualization

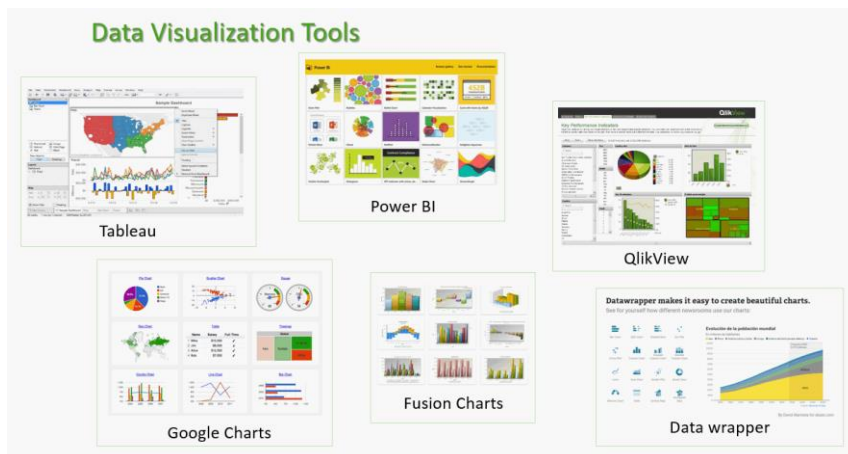
Visualization คือ การจินตนาการ หรือสร้างภาพขึ้นในความคิด ซึ่งเป็นกระบวนการทำให้เกิดความรู้ความเข้าใจสำหรับเรื่องที่จินตนาการยาก วิธีการที่เป็นทางลัดก็คือการสร้างภาพให้เป็นบันไดความคิดไปสู่การใช้ความคิดอีกระดับ บันไดนี้จะช่วยตัดปริมาณข้อมูล ช่วยลดภาระการคำนวณ หรือการนำไปผ่านหลากหลายกระบวนการความคิด เพื่อนำไปสู่คำตอบที่ต้องการได้โดยเร็วและถูกต้อง

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพในความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

Visualization System คือระบบ ซึ่งประกอบด้วยกระบวนการต่าง ๆ ที่ออกแบบมาเพื่อสร้าง รักษา นำไปใช้ และปรับปรุงทัศนสนเทศ เพื่อทำให้เกิดความรู้ความเข้าใจอย่างถูกต้อง รวดเร็ว และได้ผลเป็นอย่างดี

2.2.4.1 Tools Data Visualization

Tools ที่นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qlik View, Google Charts, Fusion Charts, Data wrapper และอื่น ๆ อีกมากมาย



ภาพที่ 2.13 Tools Data Visualization

ที่มา : <https://bit.ly/2Xsu4K3>

2.2.4.2 รูปแบบในการใช้ Data Visualization

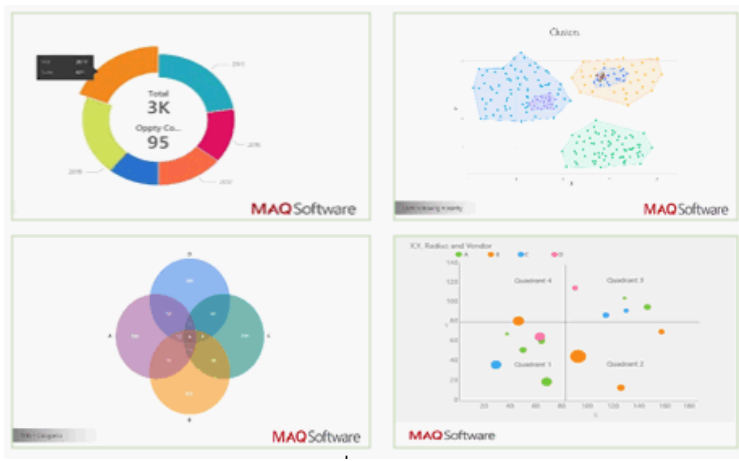
1) การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending) เราใช้กราฟที่แสดงผลแบบทิศทางหรือแนวโน้ม เพื่อนำเสนอข้อมูลให้เห็นจำนวนข้อมูลที่เกิดขึ้นในแต่ละช่วงเวลา (period) รวมถึงเน้นข้อมูลที่ต้องการนำเสนอ เช่น Line Chart, Bar Chart, Radar Chart, Area Chart เป็นต้น



ภาพที่ 2.14 รูปแบบในการใช้ Data Visualization

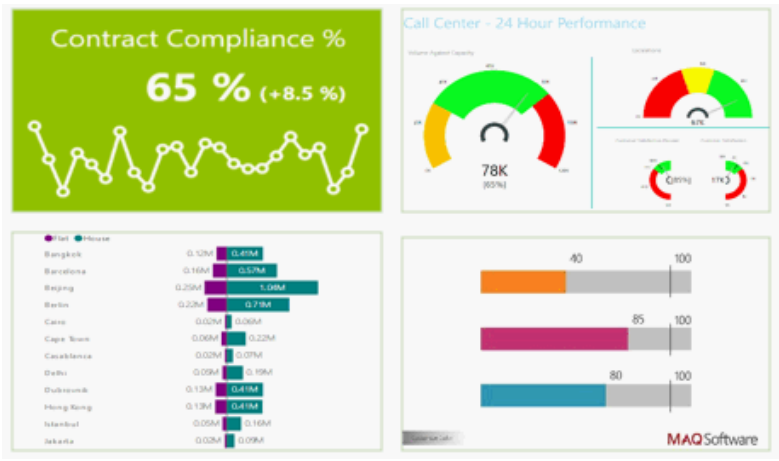
ที่มา : <https://bit.ly/2Xsu4K3>

2) การนำเสนอแบบกลุ่มข้อมูล (Classification) เป็นการนำเสนอโดยนำข้อมูลมาจัดเป็นกลุ่มๆ เช่น Donut Chart, Ring Chart, Pie Chart เป็นต้น



ภาพที่ 2.15 การนำเสนอแบบกลุ่มข้อมูล (Classification)
ที่มา : <https://bit.ly/2Xsu4K3>

3) การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison) เหมาะสำหรับการนำเสนอที่ต้องการเปรียบเทียบผลการดำเนินงาน เช่น เทียบกับปีที่แล้ว (YoY) เปรียบเทียบกับเป้าที่ตั้งไว้ (Target) ซึ่งกราฟที่เหมาะสม และมักนำมาใช้ เช่น KPI Indicator, Bullet Chart, Power BI Card with state เป็นต้น



ภาพที่ 2.16 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)
ที่มา : <https://bit.ly/2Xsu4K3>

4) การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับนำเสนอข้อมูลบนแผนที่ โดยสามารถที่จะนำยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อ Focus กลุ่มลูกค้าในแต่ละพื้นที่ที่เราสนใจ เช่น Globe Map, Google Map, Flow Map เป็นต้น



ภาพที่ 2.17 การนำเสนอรูปแบบแผนที่ (Geographical)

ที่มา : <https://bit.ly/2Xsu4K3>

5) กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics) เราสามารถที่จะใช้ภาษา R หรือ Python ดึงข้อมูลในอดีตมาเพื่อวิเคราะห์เพื่อพยากรณ์อนาคต และนำเสนอข้อมูลที่ได้ในรูปแบบของกราฟ เช่น Association Rules, Clustering, Forecasting Time series, Calculation plot เป็นต้น



ภาพที่ 2.18 กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics)

ที่มา : <https://bit.ly/2Xsu4K3>

2.2.5 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ (Decision Tree)

การเรียนรู้แบบต้นไม้ตัดสินใจ 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 ความเสี่ยงในการลงทุน และผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และ ทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

2.2.5.1 ขั้นตอนวิธีการสร้างต้นไม้การตัดสินใจ

ในปัจจุบันนั้นมีการพัฒนาขั้นตอนวิธี (อังกฤษ: algorithm) ในการสอน (training) ต้นไม้การตัดสินใจมากมาย ซึ่งส่วนมากมาจากวิธีพื้นฐานวิธีหนึ่งซึ่งเป็นการค้นหาแบบละโมภ (อังกฤษ: greedy search) จากบนลงล่าง (top-down) ชื่อว่า ID3 ซึ่งถูกพัฒนาโดย John Ross Quinlan ในปี 1986

เอนโทรปี (Entropy) การสร้างต้นไม้การตัดสินใจจากบนลงล่างด้วยการถามว่าลักษณะใดควรจะเป็นรากของต้นไม้การตัดสินใจต้นนี้ และถามซ้ำ ๆ ไปเรื่อย ๆ เพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด(อังกฤษ: recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดที่สุคนั้นดูจากค่าของลักษณะเรียกว่าเกนส์ความรู้ (Information gain) ก่อนที่จะรู้จักเกนส์ความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ $E(S)$ ดังนี้

$$E(S) = - \sum_{j=1}^n p_S(j) \log_2 p_S(j)$$

เมื่อ

- S คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้น และตัวแปรตามหลาย ๆ กรณี
- $p_S(j)$ คืออัตราส่วนของกรณีใน S ที่ตัวแปรตามหรือผลลัพธ์มีค่า j

โดยสำหรับต้นไม้การตัดสินใจที่มีผลลัพธ์เป็นแค่เพียงค่าตรรกะ (Boolean) ใช้กับไม่ใช่เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปี คือ

$$E(S) = -p_{yes} \log_2(p_{yes}) - p_{no} \log_2(p_{no})$$

เมื่อพิจารณาเอนโทรปีแล้วจะเห็นว่าเอนโทรปีจะมีค่าอยู่ระหว่าง 0 กับ 1 โดยจะมีค่าเป็นศูนย์เมื่อทุก ๆ กรณีมีผลลัพธ์เพียงแบบเดียว เช่น ใช้ทั้งหมด หรือไม่ใช่ทั้งหมด และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น หรือจะพูดอีกนัยหนึ่งก็คือเอนโทรปีจะมีค่ามากขึ้นหากข้อมูลไม่บริสุทธิ์ และจะตัดสินใจได้ว่าผลลัพธ์จะเป็นอะไรเมื่อเอนโทรปีเป็น 0 เท่านั้น

เกนความรู้ (Information Gain) ซึ่งจากการนิยามเอนโทรปีข้างต้น ทำให้เราสามารถนิยามลักษณะของตัวแปรต้นที่ดีได้ โดยตัวแปร A จะเป็นตัวแปรต้นที่ดีก็ต่อเมื่อหากว่าแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุด ๆ มีจำนวนชุดตามจำนวนค่าของ A ที่เป็นไปได้ เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า A เพียงค่าเดียวและค่าเฉลี่ยของเอนโทรปีของชุดข้อมูลที่ถูกแบ่งออก (partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปีหลังจากข้อมูลถูกแบ่งด้วย A ว่าเกนความรู้ของ A นิยามโดย

$$Gain(S, A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v)$$

- เมื่อแทนค่า - S คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลาย ๆ กรณี
- E คือเอนโทรปีของตัวอย่าง
 - A คือตัวแปรต้นที่พิจารณา
 - $value(A)$ คือเซตของค่าของ A ที่เป็นไปได้
 - S_v คือตัวอย่างที่ A มีค่า v ทั้งหมด

จะเห็นว่าหากเกนความรู้ของ A ยิ่งมากแสดงว่าหลังจากแบ่งตัวอย่าง S ด้วย A แล้วในแต่ละชุดที่แบ่งได้จะมี Entropy เข้าใกล้ศูนย์มากยิ่งขึ้น ทำให้ใกล้ที่จะตัดสินใจได้มากขึ้น เกนความรู้จึงเป็นค่าที่ดีที่จะบอกความดีของตัวแปรต้นที่นำมาพิจารณา

2.2.5.2 การใช้ ID3 สอนต้นไม้การตัดสินใจ

เมื่อเราสามารถบอกความดีของตัวแปรต้นได้จึงสามารถนำไปช่วยในการหาต้นไม้การตัดสินใจด้วย ID3 ได้โดยมีกระบวนการดังนี้

- 1) นำตัวแปรต้นที่ยังไม่ถูกนำมาใช้ทั้งหมดมาหาเกนความรู้อยู่
- 2) เลือกตัวที่มีเกนสูงที่สุด
- 3) สร้างต้นไม้ที่มีป้ปรากฏเป็นของตัวแปรต้นตัวนั้น

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม Rapid Miner Studio

เป็นโปรแกรมที่ออกแบบมาสำหรับการวิเคราะห์ข้อมูล การทำเหมืองข้อมูล (Data mining) เป็นกระบวนการ (Process) ที่กระทำกับข้อมูลขนาดใหญ่เพื่อค้นหารูปแบบแนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้นโดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์เพื่อให้ได้สารสนเทศที่เราไม่รู้ออกมา โดยสารสนเทศที่ได้จะมีเหตุผล และสามารถนำไปใช้ประโยชน์ได้ โปรแกรม Rapid Miner สามารถช่วยให้คุณเรียกดูข้อมูล และสร้างแบบจำลองเพื่อระบุแนวโน้มได้อย่างง่ายดาย เมื่อคุณจัดการกับฐานข้อมูลขนาดใหญ่การระบุการเชื่อมต่อระหว่างสองเหตุการณ์อาจเป็นเรื่องยากหรือเป็นไปไม่ได้เนื่องจากมีหลายธุรกิจที่ต้องพึ่งพาข้อมูลที่มีอยู่เพื่อทำการตัดสินใจที่สำคัญ นักวิเคราะห์ข้อมูลจึงใช้แอปพลิเคชันพิเศษเพื่อให้เห็นภาพ และเข้าใจข้อมูล โดยโปรแกรม Rapid MinerStudio มีวัตถุประสงค์เพื่อมอบเครื่องมือที่ใช้งานง่าย และมีประสิทธิภาพสำหรับการวิเคราะห์ข้อมูลจากแหล่งต่าง ๆ ช่วยให้คุณสามารถโหลดข้อมูลที่ต้องการจากไฟล์ข้อความธรรมดา, เอกสาร Office หรือแม้แต่เซิร์ฟเวอร์ฐานข้อมูลเช่น Oracle, MySQL หรือ PostgreSQL

2.3.2 โปรแกรม Tableau Public

Tableau คือ Software ของ Business Intelligence และ Data Analytics เพื่อช่วยให้เราจัดการกับข้อมูลและทำการวิเคราะห์ข้อมูลผ่าน Software โดยเราสามารถที่จะ หมุนแกนข้อมูลต่าง ๆ เพื่อหาความสัมพันธ์ของข้อมูล และสามารถที่จะทำเป็น Dashboard เพื่อนำเสนอในรูปแบบของภาพ เพื่อให้คนที่อ่าน Report เข้าใจได้ง่าย

2.3.2.1 ลักษณะเด่นของ Tableau Public

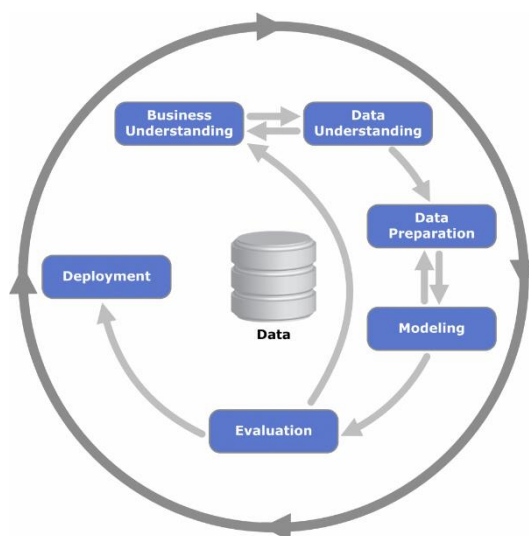
- 1) รูปแบบการใช้งานที่ง่ายเพียงคลิกเมาส์ก็สามารถที่จะเปลี่ยนแปลงรายงานได้ตามความต้องการ
- 2) รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot
- 3) รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการนำเสนอต่อผู้บริหาร
- 4) การเข้าถึงหรือการใช้งาน สามารถทำได้โดยง่าย ตั้งแต่ผู้บริหารระดับสูงถึงพนักงานระดับปฏิบัติการ ทั้งนี้ทั้งนั้นสามารถเพิ่มทักษะ และแนวคิดให้กับพนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติ และการประยุกต์ใช้งานมากขึ้น
- 5) สิ่งที่สำคัญ คือ Engine ของ Tableau Software นั้น พัฒนาจาก VizQL Technology เป็นเทคโนโลยีที่ช่วยให้เห็นภาพของข้อมูลขนาดใหญ่ได้เพียงการลาก และวาง(Drag and Drop) โดยใช้ข้อมูลจากฐานข้อมูล ตอบสนองในรูปแบบของ กราฟฟิก

2.3.3 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP-DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่

- 1) รู้จักและเข้าใจในธุรกิจ (Business understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจ
- 2) สร้างฐานข้อมูลให้ครบ (Data understanding) ขั้นตอนการจัดเก็บ และรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้
- 3) เตรียมข้อมูลให้พร้อมใช้ (Data preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมา และเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้
- 4) จัดทำและเลือกโมเดลที่ใช้ (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์ และสถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ ผสมผสานกัน
- 5) ประเมินผลก่อนตัดสินใจ (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด

6) เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard)



ภาพที่ 2.19 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

ที่มา : <http://dataminingtrend.com/2014/data-mining-techniques/crisp-dm-example/>

2.3.4 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

2.3.4.1 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) และทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

2.3.4.2 ลักษณะต้นไม้การตัดสินใจ

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

2.3.4.3 ขั้นตอนในการสร้าง Decision Tree

- 1) เลือก Attribute ที่ทำหน้าที่เป็น Root Node
- 2) จาก Root Node สร้างเส้นเชื่อมโยงไปยังโหนดลูก จำนวนเส้นเชื่อมโยงจะเท่ากับจำนวนค่าที่เป็นไปได้ทั้งหมดของ Attribute ที่เป็น root node
- 3) ถ้าโหนดลูกเป็นกลุ่มของข้อมูลที่อยู่ในคลาสเดียวกันทั้งหมด ให้หยุดสร้างต้นไม้ แต่ถ้าโหนดลูกมีข้อมูลของหลายคลาสปะปนกันอยู่ ต้อง สร้าง subtree เพื่อจำแนกข้อมูลต่อไป โดยเลือก subtree มาทำหน้าที่ เป็น root node ของ subtree มาทำซ้ำในขั้นตอนที่ 2 และ 3

2.4 วรรณกรรมที่เกี่ยวข้อง

อัจฉรา นุตตะโร (2560) ผลการใช้ตัวแบบเชิงกระบวนการ เพื่อลดความเสี่ยงโรคหลอดเลือดหัวใจ มีวัตถุประสงค์ เพื่อเปรียบเทียบผลการใช้ตัวแบบเชิงกระบวนการ เพื่อลดความเสี่ยงโรคหลอดเลือดหัวใจในข้าราชการกองทัพอากาศ ได้แก่ กระบวนการประเมินความเสี่ยง (risk stratification) การให้ข้อมูลในการปรับเปลี่ยนพฤติกรรม (risk education) การลดความเสี่ยง (risk reduction) และติดตามการลงมือปฏิบัติ (holding on) โดยคัดเลือกกลุ่มตัวอย่างที่มีความเสี่ยงต่อการเกิดโรคหัวใจ ระดับปานกลาง จำนวน 60 คน แบ่งเป็นกลุ่มทดลองและกลุ่มควบคุม กลุ่มละ 30 คน เก็บรวบรวมข้อมูลความเสี่ยงของกลุ่มตัวอย่างก่อนและหลังการใช้ตัวแบบเชิงกระบวนการ นาน 3 เดือน แล้วเก็บข้อมูลอีกครั้งหนึ่งหลังใช้ตัวแบบเชิงกระบวนการนาน 1 ปี เพื่อศึกษาความคงอยู่ของตัวแบบเชิงกระบวนการ วิเคราะห์ข้อมูลโดยใช้สถิติพรรณนา และเปรียบเทียบผลการทดลองด้วยสถิติ t-test

ผลการวิจัยพบว่า 1) ความเสี่ยงต่อการเกิดโรคหลอดเลือดหัวใจของกลุ่มทดลอง หลังใช้ตัวแบบเชิงกระบวนการ เป็นระยะเวลา 3 เดือนและ 1 ปี ต่ำกว่าก่อนใช้ตัวแบบเชิงกระบวนการอย่างมีนัยสำคัญทางสถิติ ($p < .05$) 2) กลุ่มทดลองมีความเสี่ยงต่อการเกิดโรคหลอดเลือดหัวใจ หลังใช้ตัวแบบเชิงกระบวนการนาน 3 เดือน และหลังใช้ตัวแบบเชิงกระบวนการนาน 1 ปี ต่ำกว่ากลุ่มควบคุมอย่างมีนัยสำคัญทางสถิติ ($p < .05$) สรุปได้ว่า บุคคลที่ใช้ตัวแบบเชิงกระบวนการเพื่อป้องกันโรคหลอดเลือดหัวใจ สามารถลดความเสี่ยงต่อการเกิดโรคหลอดเลือดหัวใจได้

ชิตชนก ศรีชัยวงศ์ (2557) ระบบสนับสนุนการตัดสินใจเพื่อวินิจฉัยโรคใบลำไยด้วยเทคนิคต้นไม้ตัดสินใจ การพัฒนาระบบสนับสนุนการตัดสินใจเพื่อวินิจฉัยโรคใบลำไยด้วยเทคนิคต้นไม้ตัดสินใจในลักษณะเว็บแอปพลิเคชัน โดยได้สร้างฐานความรู้จากข้อมูลความรู้และคำแนะนำจากผู้เชี่ยวชาญด้านโรคใบลำไย รวมทั้งความรู้ที่ได้มาจากสังเคราะห์เอกสารงานวิจัยเกี่ยวกับโรคใบลำไยโดยใช้กระบวนการวิเคราะห์เนื้อหา (Content Analysis) แล้วนำมาจัดหมวดหมู่ความรู้ให้เป็นระเบียบกลุ่มตัวอย่างที่ใช้ในการสร้างตัวแบบวินิจฉัยโรคใบลำไย ได้แก่ ข้อมูลใบลำไยจำนวน 225 ระเบียบที่ประกอบด้วย 7 ปัจจัยหลักที่เป็นสาเหตุของโรคใบลำไยทั้งสิ้น 14 โรค ส่วนตัวแบบวินิจฉัยโรคใบลำไยจะใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยใช้อัลกอริทึม C5.0 เพื่อให้ได้มาซึ่งเงื่อนไขในการตัดสินใจ หลังจากนั้นก็นำตัวแบบมาพัฒนาระบบโดยใช้เครื่องมือ ได้แก่ ภาษาPHPในการเขียนเว็บแอปพลิเคชัน และใช้โปรแกรมฐานข้อมูล MySQL

ผลการวิจัยพบว่า 1) ตัวแบบวินิจฉัยโรคใบลำไย มีผลให้ค่าความถูกต้องเท่ากับ 85.3% 2) ระบบที่พัฒนามีผลการประเมินประสิทธิภาพของระบบโดยผู้เชี่ยวชาญโดยรวมอยู่ในระดับดี ($X = 4.10$, $DS = 0.51$) และผู้ใช้งานระบบมีประสิทธิภาพโดยรวมอยู่ในระดับดี ($X = 3.99$, $DS = 0.60$) สรุปได้ว่าระบบที่พัฒนาขึ้นนี้มีความเหมาะสม สามารถนำไปใช้งานได้จริง สามารถช่วยวินิจฉัยโรคที่เกิดขึ้นกับใบลำไย และแนวทางการรักษาโรคใบลำไยได้

วิไลวรรณ นุชศรี (2554) ผลการใช้นวัตกรรมโมเดลหลอดเลือดหัวใจตีบต่อความรู้เรื่องปัจจัยเสี่ยง และการป้องกันโรคหลอดเลือดหัวใจตีบในกลุ่มเสี่ยง มีวัตถุประสงค์เพื่อศึกษาผล การใช้นวัตกรรมโมเดลหลอดเลือดหัวใจตีบที่ประดิษฐ์ขึ้น ประกอบการสอนให้ความรู้เกี่ยวกับ ปัจจัยเสี่ยง และการป้องกันโรคหลอดเลือดหัวใจตีบ โดยใช้การวิจัยกึ่ง ทดลองแบบ เปรียบเทียบก่อน และหลังสองกลุ่มตัวอย่างแบบเฉพาะเจาะจงเป็นประชาชนกลุ่มเสี่ยง จำนวน 60 คน มีอายุระหว่าง 40-60 ปีเข้าร่วมโครงการ แบ่งเป็นกลุ่มควบคุม 30 คน ให้ความรู้แบบปกติ ส่วนกลุ่มทดลอง 30 คน ให้ความรู้โดยใช้นวัตกรรมสิ่งประดิษฐ์ประกอบการ สอน เครื่องมือที่ใช้ในการเก็บ ข้อมูล ได้แก่ แบบประเมินความรู้เกี่ยวกับปัจจัยเสี่ยงและการ ป้องกันโรคหลอดเลือดหัวใจตีบ และแบบประเมินความคิดเห็นเกี่ยวกับโมเดลหลอดเลือดหัวใจ ตีบ วัดความรู้ของกลุ่มตัวอย่างทั้งสองกลุ่มก่อน และหลังการสอน และประเมินความพึงพอใจ ในการใช้โมเดลหลอดเลือดหัวใจตีบหลังการสอนในกลุ่มทดลอง วิเคราะห์ข้อมูลโดยใช้สถิติ บรรยาย Paired t-test และ Independent t-test

ผลการศึกษาพบว่า กลุ่มตัวอย่างทั้งสองกลุ่มมีคะแนนหลังการสอนเพิ่มขึ้นอย่างมี นัยสำคัญทางสถิติ ($P < .001$) และเมื่อเปรียบเทียบคะแนนหลังการสอนระหว่างกลุ่มควบคุมกับ กลุ่มทดลอง พบว่ากลุ่มทดลอง มีคะแนนความรู้เกี่ยวกับปัจจัยเสี่ยงโรคหลอดเลือดหัวใจตีบ และการป้องกันสูงกว่ากลุ่มควบคุมอย่างมีนัย สำคัญทางสถิติ ($p < .01$) ผลการวิจัยครั้งนี้แสดง ว่า นวัตกรรมสิ่งประดิษฐ์โมเดลหลอดเลือดหัวใจตีบสามารถนำไปเป็นอุปกรณ์ ประกอบการ สอนให้ความรู้เกี่ยวกับปัจจัยเสี่ยงโรคหลอดเลือดหัวใจตีบ เพื่อส่งเสริมสุขภาพให้ประชาชนมี สุขภาพหัวใจรู้วิธีป้องกันโรคได้เป็นอย่างดี

สิทธิพงศ์ รัตตะเมธากุล (2559) การศึกษาผู้ป่วยโรคหัวใจด้วยตัวแบบลอการิทึมเชิง เส้นกรณีศึกษาผู้ป่วยโรงพยาบาลเปาโล เมโมเรียล สมุทรปราการ การวิจัยในครั้งนี้มี วัตถุประสงค์หลักของการศึกษาคือ เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรส่วนบุคคล และตัวแปรโรคหัวใจโดยใช้ตัวแบบลอการิทึมเชิงเส้น การศึกษาในครั้งนี้ประกอบด้วยจำนวน ผู้ป่วยที่เป็นโรคหัวใจซึ่งเข้ารับการรักษาที่โรงพยาบาลเปาโล เมโมเรียล สมุทรปราการ ระหว่าง เดือนมกราคม-ธันวาคม 2555 การทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัวแปรใด ๆ

ใช้ค่าโคสแควร์ และวีคราเมอริ์ในการทดสอบ นอกจากนี้ใช้ตัวแบบลอการิทึมเชิงเส้นสองมิติในการประมาณค่าพารามิเตอร์ของตัวแบบใหญ่สุด ซึ่งถือว่าเป็นตัวแบบที่เหมาะสมที่สุดในการศึกษาครั้งนี้

ผลการวิจัยพบว่า ในการศึกษาข้อมูลผู้ป่วยโรคหัวใจจำนวน 95 คน ที่เข้ารับการรักษาที่โรงพยาบาลเปาโล เมโมเรียล สมุทรปราการ ตั้งแต่เดือนมกราคม ถึง ธันวาคม 2555 จากตารางที่ 4 และ 5 พบว่า ผู้ป่วยโรคหัวใจเป็น เพศชายมากกว่าเพศหญิง โดยเป็นผู้ป่วยเพศชายร้อยละ 53.68 ในขณะที่เป็นผู้ป่วยเพศหญิงร้อยละ 46.32 และอายุเฉลี่ยของผู้ป่วยทั้งหมดอยู่ที่ 61 ปีขึ้นไป โดยมีจำนวนผู้ป่วยโรคหัวใจที่มีอายุต่ำกว่าหรือเท่ากับ 40 ปี คิดเป็นร้อยละ 27.37 ส่วนผู้ป่วยโรคหัวใจที่มีอายุ 41-60 ปี คิดเป็นร้อยละ 31.58 และผู้ป่วยโรคหัวใจที่มีอายุตั้งแต่ 61 ปีขึ้นไป คิดเป็นร้อยละ 41.05 ผู้ป่วยส่วนใหญ่จะเป็นโรคหัวใจขั้นรุนแรงอันตรายถึงชีวิต (กลุ่มที่ 3) นั่นคือ หลอดเลือดตีบ แข็งตัวหรือโป่งพอง กล้ามเนื้อหัวใจตาย โรคหัวใจขาดเลือด หัวใจวายเฉียบพลัน เป็นต้น และการตรวจวินิจฉัยโรคหัวใจมาจากการตรวจเพียง 1 แบบเท่านั้น โดยอาจเป็นการตรวจคลื่นแม่เหล็กไฟฟ้าหรือการทดสอบบนสายพานวิ่ง

ไพวรรณ มะละ (2561) การพัฒนาระบบวิเคราะห์ข้อมูลเพียงลมอัจฉริยะเพื่อลดอาการเกิดโรคแผลกตทับ มีวัตถุประสงค์ เพื่อพัฒนาระบบรับ-ส่ง ข้อมูล สำหรับวิเคราะห์ข้อมูลแรงกตทับตามลักษณะท่านอนของผู้ป่วย แบ่งตามอวัยวะทั้ง 9 ส่วน คือ หลังศีรษะ ไหล่ซ้าย ไหล่ขวา ข้อศอกซ้าย ข้อศอกขวา ก้นซ้าย ก้นขวา สันเท้าซ้าย และสันเท้าขวา โดยอาศัยเทคโนโลยี IoT (Internet of things) ร่วมกับ MQTT Protocol (Message Queuing Telemetry Transport Protocol) เพื่อ รับ-ส่ง ข้อมูลจากเพียงลมสู่ระบบฐานข้อมูลแบบ Relational Database วิเคราะห์ข้อมูลด้วยวิธีการ การตัดสินใจแบบโครงสร้างต้นไม้ (Decision Tree) เพื่อทำงานด้านขบวนการวิเคราะห์ และการตัดสินใจ โดยอาศัยหลักการยุบและพองของพื้นที่ผิวสัมผัสของเพียงลมเพื่อลดแรงกตทับ ลักษณะเหมือนการเติมลมสำหรับยางรถยนต์ด้วยมาตรฐานการวัด อัตราความดันลมยาง ปอนด์ / ตารางนิ้ว (PSI : Pounds per Square Inch) โดยแสดงผลในรูปแบบของกราฟกข้อความออกทางหน้าจอ

ผลการวิจัยพบว่า เป็นไปตามวัตถุประสงค์โดยระบบสามารถเก็บข้อมูลแรงดันลมจากเพียงลมมาวิเคราะห์ลักษณะการนอนของผู้ป่วยแสดงผลแรงดันลมจากเพียงลมตามลักษณะการนอนของผู้ป่วยย้อนหลังตามช่วงเวลาที่กำหนด

ศรีเพ็ญ สวัสดิ์มงคล (2553) ปัจจัยที่มีผลต่อพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจของบุคลากรกรมควบคุมโรค มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจของบุคลากรกรมควบคุมโรคกระทรวงสาธารณสุขกลุ่มตัวอย่าง

ได้แก่ บุคลากรกรมควบคุมโรคจำนวน 312 คน ใช้วิธีการสุ่มแบบแบ่งชั้นภูมิเครื่องมือที่ใช้ในการวิจัยคือแบบสอบถามที่ผู้วิจัยสร้างขึ้นวิเคราะห์ข้อมูลโดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูปทางสถิติค่าสถิติที่ใช้ในการวิจัย ได้แก่ ค่าร้อยละค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐานสถิติวิเคราะห์ไคสแควสถิติ Contingency Coefficient ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันพบว่าพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจของบุคลากรกรมควบคุมโรคกระทรวงสาธารณสุขอยู่ในระดับปานกลางปัจจัยทางชีวสังคม ได้แก่ เพศอายุระดับการศึกษาสถานภาพการสมรสรายได้เฉลี่ยต่อเดือนและตำแหน่งหน้าที่ไม่มีความสัมพันธ์กับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจปัจจัยนำ ได้แก่ ความรู้เกี่ยวกับการป้องกันโรคหลอดเลือดหัวใจเจตคติต่อการป้องกันโรคหลอดเลือดหัวใจการรับรู้โอกาสเสี่ยงของการเกิดโรคหลอดเลือดหัวใจการรับรู้ความรุนแรงของการเกิดโรคหลอดเลือดหัวใจการรับรู้ประโยชน์ของการปฏิบัติเพื่อป้องกันโรคหลอดเลือดหัวใจมีความสัมพันธ์ทางบวกกับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 แต่การรับรู้อุปสรรคของการปฏิบัติเพื่อป้องกันโรคหลอดเลือดหัวใจไม่มีความสัมพันธ์กับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจปัจจัยอื่น ได้แก่ ความเพียงพอของทรัพยากรในการป้องกันโรคหลอดเลือดหัวใจมีความสัมพันธ์ทางบวกกับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ปัจจัยเสริม ได้แก่ การได้รับการสนับสนุนจากครอบครัวญาติเพื่อนร่วมงานการได้รับการสนับสนุนจากบุคลากรทางการแพทย์และสาธารณสุข และการได้รับข้อมูลข่าวสารมีความสัมพันธ์ทางบวกกับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกตัวแปร

ผลการวิจัยพบว่า ลักษณะทางชีวสังคมจะเห็นได้ว่าบุคลากรของกรมควบคุมโรคส่วนใหญ่เป็นเพศหญิงร้อยละ 80.4 มีอายุระหว่าง 45-54 ปีและร้อยละ 53.2 มีระดับการศึกษาระดับปริญญาตรีร้อยละ 45.8 สถานภาพการสมรสแล้วร้อยละ 65.4 มีรายได้ต่อเดือน 30,000 บาทขึ้นไปร้อยละ 25.6 และเป็นข้าราชการถึงร้อยละ 77.9 มีพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจอยู่ในระดับปานกลางคิดเป็นร้อยละ 43.9 ส่วนผลการวิเคราะห์ความสัมพันธ์ของปัจจัยทางชีวสังคมกับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจพบว่าเพศอายุระดับการศึกษาสถานภาพสมรสรายได้เฉลี่ยต่อเดือน และตำแหน่งไม่มีความสัมพันธ์กับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจจึงกล่าวได้ว่าปัจจัยทางชีวสังคมเป็นปัจจัยที่เป็นพื้นฐานในการแสดงพฤติกรรมของแต่ละบุคคลหรืออีกด้านหนึ่งปัจจัยนี้เป็นความพอใจของบุคคลที่ได้มาจากการประกอบการเรียนรู้อย่างตนเองซึ่งความพอใจนี้อาจจะไม่มีผลทั้งในทางสนับสนุนหรือยับยั้งในการแสดงออกด้านพฤติกรรมเช่นกันพบว่าอายุระดับการศึกษสถานภาพสมรสรายได้เฉลี่ย

ต่อเดือนไม่มีความสัมพันธ์กับพฤติกรรมการป้องกันโรคหลอดเลือดหัวใจของเจ้าหน้าที่กรม
สนับสนุนบริการสุขภาพกระทรวงสาธารณสุข

วิชญ์วิสิฐ เกษรสิทธิ์ (2562) การลดจำนวนกลุ่มในการจำแนกแบบหลายกลุ่มเป็นสอง
กลุ่มสำหรับการจำแนกการกลับมารักษาซ้ำในโรงพยาบาลของผู้ป่วยโรคเบาหวาน เป็นการ
วิจัยเพื่อเปรียบเทียบประสิทธิภาพการจำแนกประเภทของการกลับมารักษาซ้ำในโรงพยาบาล
ของผู้ป่วยโรคเบาหวานแบบหลายกลุ่ม (multiclass) หรืออเนกนาม (multinomial) และแบบสอง
กลุ่มหรือทวิภาค (binary) จำนวน 2 กรณี โดยใช้เทคนิคการวิเคราะห์การถดถอยลอจิสติกและ
ต้นไม้การตัดสินใจ ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลประวัติการรักษาพยาบาลของผู้ป่วย
โรคเบาหวานจาก Clinical Care at 130 US Hospitals and Integrated Delivery Networks
ตัวแปรเป้าหมายในการจำแนกประกอบด้วยประเภทการนัดหมายให้กลับมารักษาซ้ำใน
โรงพยาบาลของผู้ป่วยโรคเบาหวานจำนวน 3 กลุ่มคือไม่กลับมารักษาซ้ำหรือไม่มีภาวะโรค
กลับมารักษาซ้ำภายใน 30 วัน และกลับมารักษาซ้ำมากกว่า 30 วัน ผลการวิจัยพบว่า
ประสิทธิภาพของการจำแนกประเภทโดยใช้เทคนิคต้นไม้การตัดสินใจแบบทวิภาคจำนวน 2
กรณีมีประสิทธิภาพสูงสุด

ผลการวิจัยพบว่า การเลือกใช้การจำแนกประเภทข้อมูลแบบสองกลุ่มหรือทวิภาคจึง
เป็นเทคนิคการจำแนกที่ทำให้แปลความหมายของค่าในเมตริกซ์สับสนได้ง่ายเนื่องจากกลุ่มของ
ตัวแปรตามมีจำนวน 2 กลุ่มมีการคำนวณค่าของตัววัดดังกล่าวโดยการกำหนดกลุ่มใดกลุ่ม
หนึ่งเป็นกลุ่มอ้างอิงทำให้การอ่านค่าจากตัววัดสามารถใช้เป็นตัวบ่งชี้หรือเกณฑ์การประเมิน
ประสิทธิภาพของตัวแบบหรือผลการตรวจวินิจฉัยอย่างเหมาะสม โดยเฉพาะอย่างยิ่งงานวิจัย
ทางการแพทย์ที่นิยมใช้ค่าจากเมตริกซ์ความสับสนในการแปลผลการวินิจฉัยโรคหรือการ
วินิจฉัยภาวะผิดปกติต่าง ๆ ด้วยการตรวจทางห้องปฏิบัติการ การรายงานผลตรวจวิธีหนึ่ง
คือรายงานว่าผลตรวจเป็นบวกหรือผลตรวจเป็นลบโดยผลการตรวจย่อมมีโอกาสเกิดผลบวก
เท็จ (false positive) หรือผลลบเท็จ (false negative) ขึ้นได้เสมอการแปลความผลจึงต้อง
ระมัดระวังเป็นอย่างยิ่ง และต้องอาศัยความรู้ทางสถิติและทฤษฎีความน่าจะเป็น ค่าผลบวก
เท็จหรือผลลบเท็จจึงเป็นตัวชี้วัดที่มีความสำคัญในการพิจารณาผลการวิจัยหรือผลการ
วินิจฉัยโรคซึ่งจะมีผลต่อประสิทธิภาพการรักษาที่ถูกต้องและเหมาะสมอีกด้วย

แสงเดือน กันทะขู้ (2557) ปัจจัยที่มีอิทธิพลต่อการปฏิบัติกิจวัตรประจำวันของผู้ป่วย
กลุ่มโรคหัวใจขาดเลือดเฉียบพลันหลังจำหน่ายออกจากโรงพยาบาล มีวัตถุประสงค์ของการ
วิจัย เพื่อศึกษาการรับรู้ภาวะสุขภาพ การจัดการกับปัญหาภาวะซึมเศร้า และความวิตกกังวล
ต่อความสามารถการปฏิบัติกิจวัตรประจำวันของผู้ป่วยกลุ่มโรคหัวใจขาดเลือดเฉียบพลันใน

ระยะ 3 และ 6 เดือนหลังจำหน่ายออกจากโรงพยาบาล การออกแบบวิจัยเป็นการวิจัยเชิงทำนายนาย การดำเนินการวิจัย กลุ่มตัวอย่างเป็นผู้ป่วยกลุ่มโรคหัวใจขาดเลือดเฉียบพลันจำนวน 100 คน ที่มารับการรักษาในโรงพยาบาลระดับตติยภูมิแห่งหนึ่งในจังหวัดเชียงราย เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลได้แก่ แบบสอบถามข้อมูลส่วนบุคคล แบบสอบถามการรับรู้เกี่ยวกับความเจ็บป่วย แบบสอบถามการเผชิญความเครียด แบบประเมินสภาวะอารมณ์เศร้า แบบประเมินความวิตกกังวล และแบบประเมินความสามารถในการปฏิบัติกิจวัตรประจำวัน วิเคราะห์ข้อมูลโดยใช้สถิติเชิงพรรณนา สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน และการถดถอยพหุคูณแบบขั้นตอน

ผลการวิจัยพบว่า ภาวะซึมเศร้าสามารถทำนายการปฏิบัติกิจวัตรประจำวันในผู้ป่วยกลุ่มโรคหัวใจขาดเลือดเฉียบพลันหลังจำหน่ายออกจากโรงพยาบาล 3 เดือน ได้ร้อยละ 9.1 ภาวะซึมเศร้าและการจัดการกับปัญหาสามารถทำนายความสามารถในการทำกิจวัตรประจำวันหลังจำหน่ายออกจากโรงพยาบาล 6 เดือนได้ร้อยละ 17.3

ณัฐวดี หงษ์บุญมี (2562) การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลด้วยอัลกอริทึมเหมืองข้อมูลสามแบบคือ โครงข่ายประสาทเทียมการเรียนรู้แบบเบย์และต้นไม้ตัดสินใจเพื่อให้ได้อัลกอริทึมที่มีประสิทธิภาพสูงสุดที่จะถูกนำมาวิเคราะห์หาปัจจัยที่ส่งผลต่อความเสี่ยงการเกิดโรคไฮเปอร์ไทรอยด์โดยการลดการนำเข้าที่ละปัจจัย ซึ่งข้อมูลที่นำมาใช้ในการทดลองเป็นข้อมูลจากโรงพยาบาลในจังหวัดพิษณุโลกจำนวน 323 ชุดข้อมูลข้อมูลสำหรับการวิเคราะห์มีจำนวน 12 ปัจจัย

ผลการวิจัยพบว่า การจำแนกข้อมูลโดยใช้โครงข่ายประสาทเทียมให้ค่าประสิทธิภาพสูงสุดโดยมีค่าความถูกต้อง 82.97% ซึ่งมากกว่าต้นไม้ตัดสินใจและการเรียนรู้แบบเบย์ที่มีค่าประสิทธิภาพความถูกต้อง 79.87% และ 68.11% ตามลำดับผลการค้นหาปัจจัยที่ส่งผลต่อความเสี่ยงโรคไฮเปอร์ไทรอยด์พบว่าปัจจัยลักษณะอาการที่มีความสำคัญคืออารมณ์แปรปรวน และเหนื่อยง่ายส่วนปัจจัยส่วนบุคคลที่มีความสำคัญคือเพศนอกจากการค้นหาปัจจัยแล้วงานวิจัยนี้ยังสามารถนำแบบจำลองการจำแนกข้อมูลที่ได้มาพัฒนาระบบการพยากรณ์ความเสี่ยงโรคไฮเปอร์ไทรอยด์บนสมาร์ตโฟนเพื่อช่วยสนับสนุนการตัดสินใจในส่วนของการวิเคราะห์ความเสี่ยงโรคไฮเปอร์ไทรอยด์ช่วยคัดกรองด้วยตัวเองเบื้องต้นและสามารถแนะแนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป

วาสนา รวยสูงเนิน (2555) การรับรู้อาการ และการตัดสินใจมารับการรักษาในผู้ป่วยกลุ่มอาการโรคหลอดเลือดหัวใจเฉียบพลัน เป็นการวิจัยเชิงบรรยาย เพื่อศึกษาการรับรู้อาการ และการตัดสินใจมารับการรักษาในผู้ป่วยกลุ่มอาการโรคหลอดเลือดหัวใจเฉียบพลันผู้ให้ข้อมูลเป็นผู้ป่วยทั้งรายเก่า และรายใหม่ที่ได้รับการวินิจฉัยว่าเป็นโรคหลอดเลือดหัวใจเฉียบพลัน ซึ่งมารับการรักษาที่โรงพยาบาลขอนแก่น จำนวน 18 ราย เก็บรวบรวมข้อมูลด้วยการสัมภาษณ์เชิงลึกตามแนวคำถาม ข้อมูลที่ได้นำมาวิเคราะห์และสรุปผลเชิงเนื้อหา

ผลการวิจัยพบว่า ผู้ให้ข้อมูลรับรู้อาการของตนเองใน 2 ลักษณะ กล่าวคือ รับรู้อาการตามความรุนแรง และรับรู้อาการตามสาเหตุ และตำแหน่ง ทั้งนี้ผู้ให้ข้อมูลรับรู้ความรุนแรงของอาการแตกต่างกัน 2 กลุ่ม คือ รับรู้ว่าอาการรุนแรงมาก เกิดขึ้นเฉียบพลันทันใด ซึ่งคิดเป็นประมาณ 2 ใน 5 ของผู้ให้ข้อมูลทั้งหมด ส่วนที่เหลือรับรู้อาการว่าไม่รุนแรง ค่อยๆ เกิดขึ้น สำหรับการรับรู้ตามสาเหตุและตำแหน่งของอาการ แบ่งออกเป็น 3 กลุ่ม คือ รับรู้ว่าอาการมีสาเหตุมาจากหัวใจ รับรู้แบบไม่แน่ใจว่าอาการเกิดจากหัวใจ และรับรู้ว่าอาการไม่ได้มีสาเหตุจากหัวใจ ซึ่งผู้ให้ข้อมูลประมาณ 4 ใน 5 รายรับรู้ว่าอาการของตนไม่ได้มีสาเหตุมาจากหัวใจ เนื่องจากอาการที่เกิดคลุมเครือ ไม่ชัดเจน ส่งผลให้ผู้ให้ข้อมูลตอบสนอง และจัดการกับอาการแตกต่างกัน ส่วนกลุ่มที่รับรู้ว่าอาการไม่น่าจะมาจากหัวใจจะพยายามรักษาหรือบรรเทาอาการด้วยวิธีของตนเอง ตามความเข้าใจ และรอประเมินอาการหลังการจัดการนั้น ๆ โดยมีส่วนหนึ่งที่อดทนรอคอยให้อาการหายไปเองโดยไม่คิดว่าต้องรีบมารับการรักษามีผู้ให้ข้อมูลซึ่งเป็นผู้ป่วยเก่าเพียง 1 ราย ที่รับรู้ว่าตนเองจะต้องพบแพทย์โดยด่วน เนื่องจากเคยมีประสบการณ์มาแล้วสำหรับปัจจัยสำคัญที่ทำให้ผู้ให้ข้อมูลตัดสินใจมาโรงพยาบาลคือ ความชัดเจนของอาการ และอาการที่รุนแรงมากขึ้นจนทำให้ รู้สึกเหมือนกำลังจะตาย โดยที่ผู้ให้ข้อมูลที่เป็นผู้ป่วยเก่ายังคงไม่สามารถเชื่อมโยงอาการที่เกิดขึ้นกับโรคที่เป็นอยู่ได้ แม้ผู้ให้ข้อมูลบางรายพยายามจัดการอาการโดยการอมยาใต้ลิ้นตามคำแนะนำ แต่ระยะเวลาที่รอดูอาการจนตัดสินใจมารับการรักษายังขึ้นอยู่กับความรุนแรงของอาการที่รุนแรงมากเป็นหลัก ปัจจัยอื่น ๆ ที่เกี่ยวข้องในการตัดสินใจคือความสามารถในการเดินทางมาโรงพยาบาลด้วยตนเอง การศึกษานี้ยังพบว่า ระบบบริการทางการแพทย์ในการรับส่งผู้ป่วยฉุกเฉินไม่เป็นที่นิยมเพราะความล่าช้าและความเชื่อเกี่ยวกับการเคยใช้รถรับส่งผู้เสียชีวิต ล้วนทำให้ระยะเวลาในการมารับการรักษาค่อนข้างล่าช้า การศึกษานี้ชี้ให้เห็นความสำคัญของการให้ความรู้แก่ประชาชนกลุ่มเสี่ยงต่อ ความเข้าใจในอาการของโรคหลอดเลือดหัวใจเฉียบพลัน โดยเฉพาะอาการที่คลุมเครือ รวมทั้งวิธีการตอบสนองต่ออาการที่ถูกต้อง และการจัดระบบรับส่งผู้ป่วยฉุกเฉินที่ตอบสนองความต้องการของประชาชน

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือ และวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม Rapid Miner Studio 9.5.1 ในการสร้างโมเดล Decision Tree เพื่อนำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่ข้อมูลสารสนเทศนี้บน Web browser โดยการใช้ภาษา HTML ในการพัฒนาหน้าเว็บไซต์สำหรับการ เปิดเผยข้อมูลการวิเคราะห์ข้อมูลโรคหัวใจ เพื่อให้ผู้ใช้งานทราบถึงเกี่ยวกับการวิเคราะห์โรคหัวใจ ได้ผลสรุปที่ ถูกต้อง สะดวก รวดเร็ว เข้าใจง่าย ทันท่วงเวลาที่ต้องการศึกษา