บทที่ 2
แนวคิดทฤษฎีเอกสารที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลจากศูนย์ข้อมูลหมอกควันแห่งประเทศไทยเพื่อนำเสนอข้อมูลสารสนเทศบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับระบบ เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด
- 2.1.1 แนวคิดเกี่ยวกับฐานข้อมูลขนาดใหญ่ (Big Data)
- 2.1.2 แนวคิดเกี่ยวกับข้อมูลอนุกรมเวลา (Data Time Series)
- 2.1.3 การวิเคราะห์อนุกรมเวลา (Time Series Analysis)
- 2.1.4 แนวคิดเกี่ยวกับธุรกิจอัจฉริยะ (Business Intelligence: BI)
- 2.1.5 แนวคิดเกี่ยวกับแผงหน้าปัด (Dashboard)
- 2.1.6 แนวคิดเกี่ยวกับการสร้างมโนภาพ (Visualization)

2.2 ทฤษฎี
- 2.2.1 ทฤษฎีเกี่ยวกับระบบจัดการฐานข้อมูล
- 2.2.2 ทฤษฎีเกี่ยวเทคนิควิธีเหมืองข้อมูล
- 2.2.3 ทฤษฎีเกี่ยวออกแบบ (Dashboard)
- 2.2.4 ทฤษฎีเกี่ยวออกแบบ (Visualization)
- 2.2.5 ทฤษฎีเกี่ยวกับการออกแบบกราฟิกด้วยโปรแกรม (Photoshop CS6)
- 2.2.6 ทฤษฎีเกี่ยวกับการสร้างเว็บไซต์

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล
- 2.3.1 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP–DM)
- 2.3.2 ผังงาน (Flowchart)
- 2.3.3 เทคนิควิธีทางเหมืองข้อมูล
- 2.3.4 แบบจำลองอนุกรมเวลา (Time Series)

2.4 วรรณกรรมที่เกี่ยวข้อง

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับฐานข้อมูลขนาดใหญ่

ปจจุบันที่เทคโนโลยีดิจิทัลเติบโตอย่างรวดเร็วมีจำนวนมากขึ้น และได้เขามามีสวนสำคัญกับการใชชีวิตประจำวันของผู้คนผ่านอุปกรณ์ดิจิทัลต่าง ๆ เหลานี้ทั้งในขณะที่ทำงานและในขณะที่ใชชีวิตประจำวันเกิดการสร้างแล้วจัดเก็บข้อมูล มหาศาลขึ้น เชน เอกสารที่เกิดขึ้นจากการทำงาน การขออนุญาตเครือข่ายสังคมออนไลน์ (Social Network) ไลน (Line) ทวิตเตอร์ (Twitter) เฟซบุค (Facebook) รวมถึงข้อมูลจากเซ็นเซอร์ เชน สภาพอากาศ สภาพการจราจร เป็นตน

2.1.1.1 ความหมายของฐานข้อมูลขนาดใหญ่

รัชพร วงศาโรจน์ (2558) กล่าวว่า นิยามทั่วไปBig Data คือ การวิเคราะห์ข้อมูลที่มีลักษณะสำคัญ 4 ข้อ หรือที่เรียกย่อ ๆ ว่า 4Vs (Volume–Velocity–Variety–Veracity) คือ ข้อมูลที่มีปริมาณมากเกินกว่าที่ระบบจะรองรับได้ (Volume) ซึ่งข้อมูลนั้นมีอัตราการเปลี่ยนแปลงแบบ Real–time (Velocity) มีรูปแบบข้อมูลที่หลากหลาย (Variety) และเป็นข้อมูลที่มีความคลุมเครือ (Veracity) ซึ่งเกิดจากการที่ข้อมูลได้มาจากหลายช่องทาง แต่ Big Data จะมีคุณค่าก็ต่อเมื่อนำข้อมูลนั้นมาวิเคราะห์เพื่อดึงข้อมูลเชิงลึก (Insights) มาใช้ให้เป็นประโยชน์

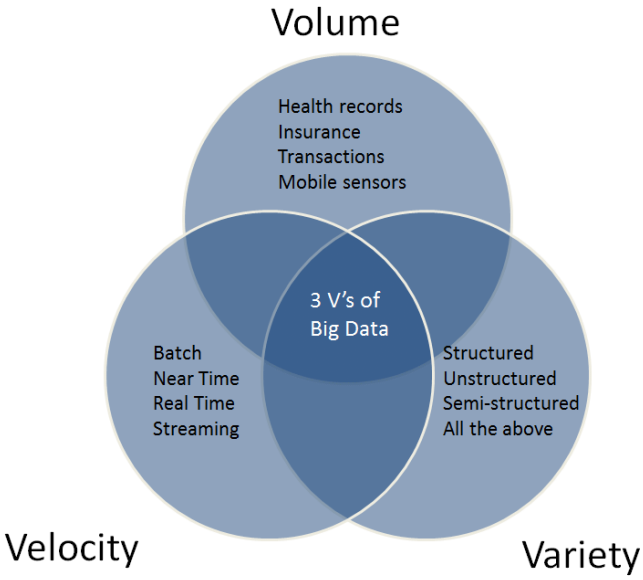
นบส. 2 รุนที่ 10 (2561) กล่าวว่า นิยามของข้อมูลขนาดใหญ่ที่ได้รับการยอมรับกัน คือ ข้อมูลที่มีขนาดใหญ่ (large volume) ที่มีความซับซ้อนหลากหลาย (variety) และความเร็วสูง (high velocity) ซึ่งตองอาศัยวิธีการและเทคโนโลยีที่ก้าวหนาทันสมัยในการจัดเก็บส่งต่อ จัดการ และวิเคราะห์

ข้อมูลขนาดใหญ่มักจะถูกกำหนดด้วย 3 ปัจจัย ได้แก

1) ขนาด (volume) หมายถึง ข้อมูลในปริมาณมหาศาลที่รวบรวมจากแหล่งขอมูลหลากหลายแหล่งและหลากหลายประเภท ขนาด เทระไบต์ (terabyte) และกลายเป็นเพตะไบต (petabyte) หรือเอกซะไบต์ (exabyte) ซึ่งมีอัตราการเติบโตในอัตราก้าวหนา (exponential) เมื่อเทียบกับเวลาที่ผ่านไป ทำให้ระบบฐานข้อมูลที่ใชงานกันอยู่ไม่สามารถประมวลผลได้และจำเป็นตองใช data lake และโปรแกรมฮาดูป (Hadoop) ทำงานประสานกันในการบริหารจัดการ

2) ความหลากหลาย (variety) ซึ่งไม่ได้มีเพียงความหลากหลายของแหล่งข้อมูลหากแต่ยังรวมถึงลักษณะรูปแบบของข้อมูลที่แตกต่างกันไปอันเนื่องมาจากแหล่งข้อมูลที่ต่างประเภทกันจากข้อมูลเดิม ข้อมูลเอกสาร ข อมูลจากการทำงานบนอินเทอร์เน็ต ขอมูลจากเครื่องส่งสัญญาณอารเอฟไอดี (radio frequency identification,RFID) จีพีเอส (geographic positioning system, GPS) ข้อมูลทางการเงิน การใชสื่อสังคมออนไลน์ (social media) ไปรษณีย์อิเล็กทรอนิกส์ (e-mail) ระบบเซ็นเซอร์และมาตรวัดต่าง ๆ (sensor system) มาตรวัดต่าง ๆ ที่บันทึกขอมูลแบบทันที ภาพเคลื่อนไหว (video) และเสียง (audio) เป็นต้นลักษณะของข้อมูลมีทั้งลักษณะที่มีโครงสร้าง (structure) ชัดเจนที่สามารถประมวลผลได้ด้วยวิธีการเดิมไปจนถึงข้อมูลที่เป็นโครงสร้าง (unstructured) ซึ่งมีปริมาณประมาณร้อยละ 80 ถึงร้อยละ 85 ของข้อมูลทั้งหมดในองคกรหรือหน่วยงานต่าง ๆ และมีแนวโนมที่จะเติบโตเร็วกว่าข้อมูลที่มีโครงสร้าง

3. ความรวดเร็ว (velocity) หมายถึง ข้อมูลเกิดขึ้น ได้รับการจัดการและนำถูกไปใชอย่างรวดเร็ว เชน ข้อมูลที่เกิดจากอินเทอร์เน็ต สื่อสังคมออนไลน์ ตัวรับรูมาตรวัด ฯลฯ เหล่านี้สามารถเกิดขึ้นตลอดเวลารวมทั้งระบบการรวบรวมขอมูลที่กระทำได้อย่างทันที (real-time)

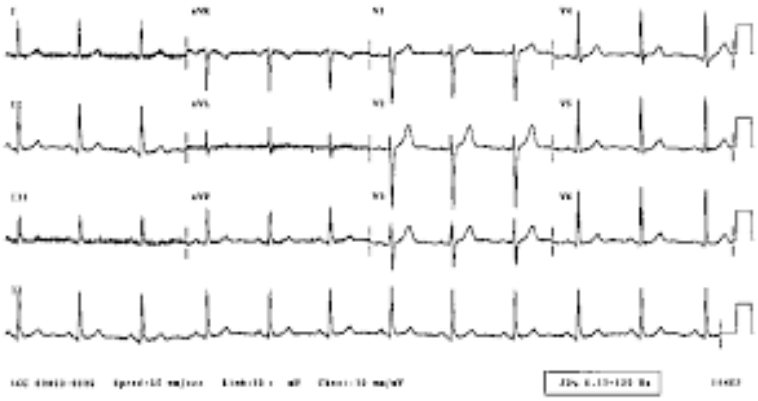


ภาพที่ 2.1 คุณลักษณะ Big Data

ที่มา : Christopher O. Austin, (2559)

2.1.2 แนวคิดเกี่ยวกับข้อมูลอนุกรมเวลา

ข้อมูลอนุกรมเวลาเป็นค่าของข้อมูลที่ถูกเก็บควบคู่กับจุดเวลาที่เกิดขึ้น ดังนั้นทุกจุดข้อมูลในอนุกรมเวลาใด ๆ จะมีเวลากำกับเสมอ ซึ่งข้อมูลลักษณะนี้มักจะพบเห็นได้โดยทั่วไปทั้งใกล้และไกลตัว เช่นข้อมูลอนุกรม เวลาของตลาดหุ้น (stock market) นอกจากนี้ยังมีข้อมูลคลื่นไฟฟ้าสมอง (Electroencephalography หรือ EEG) ข้อมูลคลื่นไฟฟ้าหัวใจ (Electrocardiography หรือ ECG) แสดงในภาพที่ 2.2 ข้อมูลสภาพอากาศในแต่ละวัน หรือแม้แต่ข้อมูลรายได้ของแม่ค้า ที่ได้รับในแต่ละชั่วโมงก็ นับเป็นข้อมูลอนุกรมเวลาเช่นเดียวกัน จะเห็นว่าข้อมูลอนุกรมเวลานั้นเป็นข้อมูลที่พบได้โดยทั่วไปในชีวิตประจำวัน ซึ่งนอกจากจะเป็นข้อมูลในลักษณะที่กล่าวไปข้างต้นแล้ว ข้อมูลอนุกรมเวลายังสามารถนำข้อมูลประเภทอื่นมาเปลี่ยนแปลงเป็นข้อมูลอนุกรมเวลาได้เช่นกัน ซึ่งประโยชน์ของการเปลี่ยนเป็นข้อมูลอนุกรมเวลานั้นคือการทำให้สามารถทำงานได้รวดเร็วยิ่งขึ้นในหลายชุดข้อมูล



ภาพที่ 2.2 อนุกรมเวลาของข้อมูลคลื่นหัวใจ

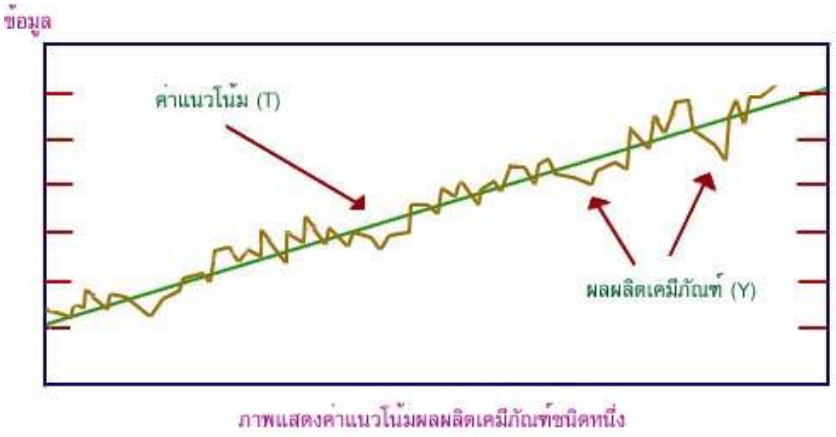
ที่มา : Siamhealth.net (ม.ป.ป.)

อรรณพ กางกั้น (2562) งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพและเลือกเทคนิคที่เหมาะสมในการพยากรณ์และพัฒนาแบบจำลองในการพยากรณ์ปริมาณการใช้น้ำมันดีเซลด้วยเทคนิควิธีเหมืองข้อมูล 3 เทคนิคได้แก่ 1) การถดถอยเชิงเส้น 2) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้นและ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยข้อมูลที่นำมาใช้ศึกษาเป็นข้อมูลอนุกรมเวลารายเดือนปริมาณการใช้น้ำมันดีเซลตั้งแต่วันที่ 1 มกราคม 2549 ถึงวันที่ 31 ธันวาคม 2561 จากผลการทดลองพบว่าชุดข้อมูลปริมาณการใช้น้ำมันดีเซลย้อนหลัง 12 เดือนมีความเหมาะสมในการใช้เป็นชุดข้อมูลเพื่อสร้างแบบจำลองในการพยากรณ์และเมื่อพิจารณาประสิทธิภาพของแบบจำลองกับชุดข้อมูลในปี2561 ที่ใช้เป็นข้อมูลทดสอบ

2.1.3 การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

การวิเคราะห์อนุกรมเวลาเป็นการวิเคราะห์ลักษณะหรือรูปแบบของอนุกรมเวลาโดยสังเกต จากการเปลี่ยนแปลงของข้อมูลในแต่ละช่วงเวลาว่ามีการเปลี่ยนแปลงไปในลักษณะใด มีการเคลื่อนไหวของข้อมูลอย่างไรเนื่องจากอนุกรมเวลาแต่ละประเภทมีลักษณะที่แตกต่างกัน ดังนั้นจึงมีการกำหนดองค์ประกอบของอนุกรมเวลาเป็น 4 ส่วน ดังนี้

1) ส่วนแนวโน้ม (Trend Component) เป็นส่วนหลักที่สังเกตจากข้อมูลในระยะยาว โดยอาจมีการเติบโตหรือถดถอยในอนุกรมเวลา ซึ่งลักษณะของเส้นแนวโน้มนั้นขึ้นอยู่กับอนุกรมเวลา โดยอาจจะเป็นเส้นตรงหรือเส้นโค้งก็ได้ โดยเส้นแนวโน้มที่ถูกลากนั้นจะต้องเรียบไม่มีการเกิดมุมใด ๆ บนเส้น ดังตัวอย่างแสดงในภาพที่ 2.3 ซึ่งแสดงเส้นแนวโน้มที่น่าจะเป็นของข้อมูลอนุกรมเวลาผลผลิตเคมีภัณฑ์



ภาพที่ 2.3 เส้นแนวโน้มผลผลิตเคมีภัณฑ์ชนิดหนึ่ง

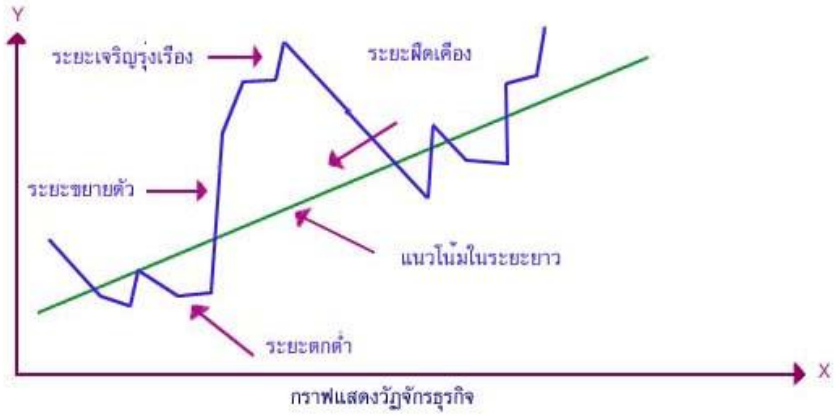
ที่มา : fpo.go.th (ม.ป.ป.)

2) ส่วนฤดูกาล (Seasonal Component) เป็นการเปลี่ยนแปลงของอนุกรมเวลาในช่วงระยะเวลาหนึ่งในรูปแบบเดียวกันช้า ๆ ซึ่งการเกิดของรูปแบบเดียวกันนี้จะเกิดในระยะเวลาสั้น ๆ ซึ่งในรูปที่ 2.3 จะพบลักษณะรูปแบบที่เกิดซ้ากันในช่วงเวลาแต่ละปี 7

ภาพที่ 2.4 แสดงวัฏจักรของธุรกิจ

ที่มา : fpo.go.th (ม.ป.ป.)

3) ส่วนวัฏจักร (Cyclical Component) เป็นการเกิดของเหตุการณ์ที่คล้ายกับส่วนฤดูกาล คือการเกิดรูปแบบซ้า ๆ กันในอนุกรมเวลา แต่จะเกิดในช่วงระยะเวลาที่ยาวกว่า ซึ่งจำเป็นที่จะต้องสังเกตจากอนุกรมเวลาที่มีระยะเวลานานดังแสดงในภาพที่ 2.5



ภาพที่ 2.5 แสดงวัฏจักรของธุรกิจ

ที่มา : fpo.go.th (ม.ป.ป.)

รูปที่ 2.3 แสดงยอดขายรายเดือนของห้างสรรพสินค้าแห่งหนึ่ง 84. ส่วนผิดปกติ (Irregular Component) เป็นการเกิดขึ้นของเหตุการณ์หรือการเปลี่ยนแปลงไปของข้อมูลที่ไม่สามารถคาดการณ์ได้ โดยการเกิดขึ้นของเหตุการณ์นี้ไม่ได้อ้างอิงกับเหตุการณ์ใด ซึ่งเป็นการเปลี่ยนแปลงในเชิงสุ่ม (Random Variation) จากองค์ประกอบของอนุกรมเวลาดังกล่าวทำให้สามารถสร้างแบบจำลองของข้อมูลอนุกรมเวลาได้ โดยมีแบบจำลองของอนุกรมเวลา ดังนี้

1) แบบจำลองผลบวก (Additive Model) เป็นแบบจำลองที่องค์ประกอบทั้ง 4 ส่วนของอนุกรมเวลานั้น ๆ ไม่ขึ้นต่อกัน โดยเมื่อองค์ประกอบใด ๆ มีค่าเปลี่ยนแปลงไป จะไม่ส่งผลกระทบต่อการเปลี่ยนแปลงขององค์ประกอบอื่น ๆ

2) แบบจำลองผลคูณ (Multiplicative Model) เป็นแบบจำลองที่องค์ประกอบทั้ง 4 ส่วนของอนุกรมเวลานั้น ๆ สัมพันธ์กัน โดยเมื่อองค์ประกอบใด ๆ มีค่าเปลี่ยนแปลงไปจะส่งผลให้องค์ประกอบอื่น ๆ มีการเปลี่ยนแปลงตามไปด้วย

2.1.4 แนวคิดเกี่ยวกับธุรกิจอัจฉริยะ (Business Intelligence: BI)

ธุรกิจอัจฉริยะ (Business Intelligence: BI) คือ แนวคิดและเทคโนโลยีสำหรับการสร้างระบบบริหารจัดการข้อมูลการจัดเก็บข้อมูลหลายมิติ และนำเสนอข้อมูลในรูปแบบรายงานตาราง และกราฟ ซึ่งสามารถวิเคราะห์ข้อมูลในระดับเชิงลึกและภาพรวม ทำให้เกิดองค์ความรู้ใหม่หรือตอบคำถามความต้องการจากฐานข้อมูลภายในองค์กร เพื่อใช้สำหรับการวางแผน และสนับสนุนการตัดสินใจให้ผู้บริหารและเจ้าหน้าที่ได้อย่างรวดเร็ว

กระบวนการของธุรกิจอัจฉริยะมี 4 ขั้นตอนที่สำคัญประกอบด้วย

1) การสกัด เปลี่ยนแปลง และถ่ายโอน (ETL) เป็นขั้นตอนการกำจัดข้อมูลที่ไม่เกี่ยวข้องและเปลี่ยนรูปแบบข้อมูลให้เป็นรูปแบบเดียวกันก่อนถ่ายโอนข้อมูลเข้าระบบคลังข้อมูล

2) คลังข้อมูล (Data Warehouse) ทำหน้าที่จัดเก็บข้อมูลตามมิติที่เกี่ยวข้องโดยแบ่งเป็นตารางมิติ (Dimension Table) และตารางข้อเท็จ (Fact Table)

3) การประมวลผลออนไลน์เชิงวิเคราะห์ (OLAP) คือ ประมวลผลข้อมูลโดยการจัดระเบียบข้อมูลและจัดเก็บโครงสร้างข้อมูลแบบมิติ (Multidimensional Data) เพื่อพร้อมสำหรับการวิเคราะห์ข้อมูลแบบซับซ้อน เช่น แบบโดยรวม (Roll Up) แบบละเอียด (Drill Down) แบบแยกส่วน (Slice) และแบบพลิกแกน (Dice)

4) การแสดงผล (Presentation) เป็นขั้นการนำเสนอผลการวิเคราะห์ข้อมูลและสามารถแสดงในรูปแบบแดชบอร์ด(Dashboard) สเปรดชีด (Spreadsheet) และรายงาน

ปัทมา เที่ยงสมบุญ และนิเวศ จิระวิชิตชัย (2561) กล่าวถึงระบบธุรกิจอัจฉริยะที่นำมาใช้เป็นเครื่องมือในการสนับสนุนการตัดสินใจของผู้บริหารโดยทำการเก็บข้อมูลความต้องการของผู้บริหารมาทำการวิเคราะห์และออกแบบรายงานที่สนับสนุนการตัดสินใจได้อย่างมีประสิทธิภาพ โดยการพัฒนาระบบธุรกิจอัจฉริยะที่ต้องอาศัยคลังข้อมูลในการจัดเก็บข้อมูลให้เหมาะสมกับงานเฉพาะด้าน ข้อมูลที่สนใจจะเป็นข้อมูลเกี่ยวกับประวัติธุรกรรมที่ผ่านมาและข้อมูลธุรกรรมที่เกิดขึ้นในปัจจุบัน ซึ่งข้อมูลธุรกรรมทั้งหมดนั้นอาจมาจากหลากหลายแหล่ง ต้องอาศัยเทคนิควิธีในการบูรณาการข้อมูลเข้าด้วยกันเพื่อจัดการกับข้อมูลให้อยู่ในมาตรฐานเดียวกัน และได้มีงานวิจัยที่เกี่ยวกับเครื่องมือธุรกิจอัจฉริยะโดยใช้เครื่องมือที่ชื่อว่าแท็บโบลว์ (Tableau) ที่ช่วยในการสำรวจข้อมูลและหาความสัมพันธ์ของข้อมูลที่ถูกซ่อนอยู่แล้วนำเสนอเป็นข้อมูลรายงาน ซึ่งข้อมูลรายงานที่นำเสนอสามารถปรับเปลี่ยนมุมมองได้ตามต้องการ และสามารถสนับสนุนการตัดสินใจในด้านต่าง ๆ ได้ รวมไปถึงผู้บริหารสามารถวางแผนงานในอนาคตได้ง่ายขึ้น

ปภาดา โพธำอภิชัย และ มณเฑียร รัตนศิริวงศ์วุฒิ (2557) ได้พัฒนาระบบธุรกิจอัจฉริยะเพื่อการวางแผนการผลิตในองค์กร เพื่อสนับสนุนการตัดสินใจของผู้บริหารในการวิเคราะห์แนวโน้มทางการตลาดและมีแนวทางในการดำเนินงานที่ชัดเจน รวมถึงการปรับกลยุทธ์ทางธุรกิจ ในการแข่งขันได้ดีโดยใช้หลักการธุรกิจอัจฉริยะ (Business Intelligence) ประกอบด้วยการสร้างคลังข้อมูลขนาดเล็ก (Data Mart) และนำสารสนเทศที่มีอยู่มาสร้างรายงานอัจฉริยะ (Dashboard) สามารถปรับเปลี่ยนมุมมองในการวิเคราะห์ และตรงตามความต้องการของผู้บริหารและผู้ใช้งาน จากผลการประเมินคุณภาพของระบบโดยผู้เชี่ยวชาญสรุปให้เห็นว่า การพัฒนาระบบธุรกิจอัจฉริยะเมื่อนำมาประยุกต์ใช้ในการวางแผนการผลิตในองค์กรสามารถใช้งานได้จริงและในระดับดี

2.1.5 แนวคิดเกี่ยวกับแผงหน้าปัด (Dashboard)

แผงหน้าปัด (Dashboard) คือการแสดงข้อมูลที่สำคัญที่สุดที่จำเป็นต้องใช้ในการบรรลุวัตถุประสงค์หนึ่งหรือมากกว่าหนึ่ง เป็นการรวมและจัดการภายในหน้าจอเดียวทำให้สามารถดูข้อมูลได้เพียงการมองครั้งเดียว หรือการแสดงภาพผ่านหน้าจอให้ผู้ใช้งาน โดยมีการกำหนดตัวชี้วัดผลการดำเนินงานหลักของธุรกิจ เพื่อให้สามารตัดสินใจได้ แผงหน้าปัดคือการแสดงภาพและการโต้ตอบด้วยเครื่องมือการจัดการประสิทธิภาพซึ่งแสดงผ่านหน้าจอเดียวด้วยสารสนเทศที่สำคัญที่สุดเพื่อให้บรรลุวัตถุประสงค์ของผู้ใช้งานหรือวัตถุประสงค์ขององค์กรโดยให้ผู้ใช้งานสามารถระบุ วิเคราะห์ และสื่อสารออกมาให้ทราบถึงส่วนที่เป็นปัญหาที่ต้องการให้เกิดการกระทำที่ถูกต้องในการแก้ปัญหา (Yigitbasioglu & Velcu, 2012) ดังนั้นจึงสรุปได้ว่าแผงหน้าปัดคือการแสดงข้อมูลที่สำคัญที่สุดที่จำเป็นต้องใช้ โดยต้องสามารถแสดงผ่านหน้าจอเดียวและต้องทำให้ผู้ใช้ดูข้อมูลเพียงครั้งเดียวแล้วสามารถตัดสินใจในเรื่องของธุรกิจนั้น ๆ ได้ การสร้างแผงหน้าปัดต้องบรรลุเป้าหมายดังต่อไปนี้

1) ความสอดคล้อง (Consistency) เพื่อให้เกิดความสอดคล้องระหว่างการวัดผลและกระบวนการที่ใช้ในการวัดผล

2) การตรวจสอบ (Monitoring) เพื่อเป็นการตรวจสอบการดำเนินงาน หรืออาจเป็นทั้งการตรวจสอบและการประเมินว่าใครหรืออะไรที่ทำงานได้ดี และการพัฒนาถึงสิ่งที่ต้องศึกษา

3) การวางแผน (Planning) ใช้ในการวางแผนถึงเป้าหมายและกลยุทธ์ในอนาคตที่ควรจะเป็นจากสถานะปัจจุบันขององค์กร

4) การสื่อสาร (Communication) เพื่อใช้เป็นเครื่องมือสื่อสารการดำเนินธุรกิจต่อหุ้นส่วนสำคัญ และไม่เฉพาะการดำเนินงานแต่รวมไปถึงการวัดเปรียบเทียบมูลค่าขององค์กร ในการดำเนินการแผงหน้าปัดที่มีประสิทธิภาพนั้นไม่จำเป็นต้องให้ผู้ใช้ดูทุกรายละเอียดว่าเกิดอะไรขึ้นบ้าง แต่จะแสดงตัวชี้วัดผลการดำเนินงานหลักที่เป็นตัวเลข ซึ่งเป็นการแสดงยอดสรุปแล้วมีส่วนที่ให้ผู้ใช้สามารถดูข้อมูลเพิ่มเติมได้เอง

Domo Inc, (2555) กล่าวว่า แผงหน้าปัดคือการแสดงสารสนเทศที่สำคัญต่อองค์กรที่เป็นตัวชี้วัดผลการดำเนินงานหลักมาจัดวางภายในหนึ่งหน้าจอ ทำให้ผู้ใช้งานสามารถมองและเข้าใจสารสนเทศได้อย่างรวดเร็วและถูกต้อง เพื่อช่วยสนับสนุนการตัดสินใจของผู้ใช้ส่งผลให้การดำเนินงานภายในองค์กรมีประสิทธิภาพมากขึ้น

วิทยา จิตรภักดี (2553) กล่าวว่า แดชบอร์ด คือ การนำเสนอข้อมูลโดยการเลือกรายงาน หรือกราฟ ที่น่าสนใจมานำเสนอรวมกันไว้ในหน้าจอเดียวกัน ซึ่งจะช่วยให้ผู้บริหารหรือผู้ใช้งานสามารถมองเห็นภาพรวมของข้อมูลที่สำคัญในเรื่องใดเรื่องหนึ่งได้อย่างชัดเจนและรวดเร็ว เช่น การแสดงข้อมูล (Balanced Scorecard) ในมุมมองด้านรายได้เปรียบเทียบกับต้นทุนด้านการตลาด หรือ KPI ในมุมมองต่าง ๆ ตามที่ต้องการ

2.1.6 แนวคิดเกี่ยวกับ การสร้างมโนภาพ (Visualization)

การแสดงข้อมูล (Data Visualization) เป็นสิ่งหนึ่งถูกนำมาใช้แสดงแทนคำพูด เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ไม่ว่าจะเป็นตัวเลข แผนภูมิ กราฟ และอื่น ๆ อีกมากมาย คำว่า (Data) คือ ข้อมูล ส่วน (Visualization) คือ การมองเห็นเมื่อนำมารวมกันแล้วหมายถึง ข้อมูลที่มองเห็นได้ด้วยตาทำให้ข้อมูลในเชิงปริมาณดูน่าสนใจ เข้าใจง่าย เห็นภาพรวมได้ชัดเจน ง่ายต่อการจดจำ และนิยมนำมาใช้ประกอบในการรายงาน การวิเคราะห์ สรุปผลอย่างแแพร่หลาย

Education Learning Initiative ELI (2007) กล่าวว่าแผนภูมิใช้ในการแสดงข้อมูลเชิงปริมาณ ทำให้สามารถมองเห็น แนวโน้มและความสัมพันธ์ของสารสนเทศได้ชัดเจนขึ้น

Browne et al. (2010) กล่าวว่าโดยสรุป จินตทัศน์คือการแสดงผลสารสนเทศให้อยู่ในรูปแบบภาพที่ผู้ใช้มองเห็นแล้วสามารถทำให้ผู้ใช้เข้าใจ สารสนเทศได้ง่ายและรวดเร็วมากขึ้น โดยเห็นถึงแนวโน้มและความสัมพันธ์ของสารสนเทศที่ซ่อนอยู่ ในสารสนเทศที่เดิมอยู่ในรูปแบบข้อความหรือตาราง

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับระบบจัดการฐานข้อมูล

ระบบจัดการฐานข้อมูล (Database Management System) หมายถึง กลุ่มโปรแกรมหรือซอฟต์แวร์ชนิดหนึ่ง ที่สร้างขึ้นมาเพื่อทำหน้าที่บริหารฐานข้อมูลโดยตรง ให้มีประสิทธิภาพมากที่สุดเป็นเครื่องมือที่ช่วยอำนวยความสะดวกให้ผู้ใช้สามารถเข้าถึงข้อมูลได้โดยที่ผู้ใช้ไม่จำเป็นต้องรับรู้เกี่ยวกับรายละเอียดภายในโครงสร้างฐานข้อมูล DBMS นี้ โดยมีซอฟต์แวร์ที่เปรียบเสมือนสื่อกลางระหว่างผู้ใช้และโปรแกรมต่าง ๆ ที่เกี่ยวข้องกับการใช้ฐานข้อมูล เรียกว่าระบบจัดการฐานข้อมูลหรือ DBMS (Data Base Management System) มีหน้าที่ช่วยให้ผู้ใช้เข้าถึงข้อมูลได้ง่ายสะดวกและมีประสิทธิภาพ การเข้าถึงข้อมูลของผู้ใช้อาจเป็นการสร้างฐานข้อมูล การแก้ไขฐานข้อมูล หรือการตั้งคำถามเพื่อให้ได้ข้อมูลมา โดยผู้ใช้ไม่จำเป็นต้องรับรู้เกี่ยวกับรายละเอียดภายในโครงสร้างของฐานข้อมูล

2.2.1.1 หน้าที่ของระบบจัดการฐานข้อมูล มีดังนี้

1) กำหนดมาตรฐานข้อมูล

2) หน้าที่จัดการพจนานุกรมข้อมูล

3) ควบคุมการเข้าถึงข้อมูลแบบต่าง ๆ ของผู้ใช้

4) ดูแลและจัดเก็บข้อมูลเปลี่ยนรูปแบบและการแสดงผลข้อมูลให้มีความถูกต้องแม่นยำ

5) จัดการสำรองข้อมูลและการกู้คืนข้อมูลแฟ้มข้อมูล

6) จัดระเบียบแฟ้มทางกายภาพ (Physical Organization)

7) จัดการด้านความปลอดภัยดูแลรักษาความปลอดภัยของข้อมูลภายในฐานข้อมูล

8) บำรุงรักษาฐานข้อมูลให้เป็นอิสระจากโปรแกรมแอพพลิเคชั่นอื่น ๆ

9) เชื่อมโยงข้อมูลที่มีความสัมพันธ์เข้าด้วยกัน

10) จัดการข้อมูลและจัดสร้างส่วนประสานกับผู้ใช้ในด้านการสื่อสารฐานข้อมูล

2.2.1.2 ฐานข้อมูล (Database) หมายถึง กลุ่มของข้อมูลที่มีความสัมพันธ์กันนำมาเก็บรวบรวมเข้าด้วยกันอย่างมีระบบและข้อมูลที่ประกอบกันเป็นฐานข้อมูลนั้น ต้องตรงตามวัตถุประสงค์การใช้งานขององค์กรด้วยเช่นกัน เช่นในสำนักงานก็รวบรวมข้อมูล ตั้งแต่หมายเลขโทรศัพท์ของผู้ที่มาติดต่อจนถึงการเก็บเอกสารทุกอย่างของสำนักงาน ซึ่งข้อมูลส่วนนี้จะมีส่วนที่สัมพันธ์กันและเป็นที่ต้องการนำออกมาใช้หรืออาจได้มาจากการสังเกต การนับหรือการวัดก็เป็นได้ รวมทั้งข้อมูลที่เป็นตัวเลข ข้อความ และรูปภาพต่าง ๆ ก็สามารถนำมาจัดเก็บเป็นฐานข้อมูลได้ และที่สำคัญข้อมูลทุกอย่างต้องมีความสัมพันธ์กัน

2.2.1.3 ระบบฐานข้อมูล (Database System) ความหมายของระบบฐานข้อมูลคือที่รวมของฐานข้อมูลต่าง ๆ ที่เกี่ยวข้องกันเข้าไว้ด้วยกันอย่างมีระบบ มีความสัมพันธ์ระหว่างข้อมูลต่าง ๆ หรือที่รวมของข้อมูลทั้งหมด ซึ่งอาจจะได้จากการคำนวณหรือประมวลผลต่าง ๆ หรืออาจได้จากการบันทึกข้อมูลโดยผู้ใช้ เช่น ระบบฐานข้อมูลงานทะเบียน นิสิตมหาวิทยาลัย รวมเอาฐานข้อมูลต่าง ๆ เช่น ฐานข้อมูลวิชาเรียน ฐานข้อมูลนิสิตฐานข้อมูล อาจารย์ ผู้สอน และ ฐานข้อมูลหลักสูตร เป็นต้น ซึ่งรวมกันเป็นระบบฐานข้อมูลของงาน ทะเบียนนิสิต หรือฐานข้อมูล ห้างร้านต่าง ๆ ก็จะประกอบด้วย ฐานข้อมูลสินค้าฐานข้อมูล ลูกค้า ฐานข้อมูลระบบบัญชีฐานข้อมูลลูกหนี้ และฐานข้อมูลตัวแทนจำหน่าย เป็นต้น ในระบบ ฐานข้อมูลจะประกอบด้วยแฟ้มข้อมูลหลายแฟ้มที่มีข้อมูลเกี่ยวข้องสัมพันธ์กันเข้าไว้ด้วยกัน อย่างเป็นระบบและเปิดโอกาสให้ผู้ใช้สามารถใช้งานและดูแลรักษาป้องกันข้อมูลได้อย่างมีประสิทธิภาพ

2.2.2 ทฤษฎีเกี่ยวเทคนิควิธีเหมืองข้อมูล

2.2.2.1 การทำเหมืองข้อมูล การทำเหมืองข้อมูล (Data Mining) คือการค้นหาความสัมพันธ์และรูปแบบทั้งหมด ซึ่ง มีอยู่จริงในฐานข้อมูลแต่ได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมาก โดยการทำเหมืองข้อมูลจะเหมาะกับการแก้ปัญหาบางชนิดเท่านั้น มีเทคนิคต่าง ๆ ที่ใช้ในการแก้ไขปัญหาอยู่หลายเทคนิค ซึ่งไม่มีเทคนิคใดสามารถแก้ปัญหาได้ทุกปัญหา ดังนั้นความหลากหลายของเทคนิคเป็นสิ่งที่จำเป็นที่จะนำไปสู่วิธีการแก้ปัญหาที่ดีที่สุดของการทำเหมืองข้อมูล

2.2.2.2 ขั้นตอนการทำงานของการทำเหมืองข้อมูล จากขั้นตอนการทำงานของการทำเหมืองข้อมูล ประกอบไปด้วย 4 ขั้นตอนหลัก ๆ ดังนี้

1) การระบุปัญหาที่เกิดขึ้นกับธุรกิจเป็นการระบุขอบเขตของข้อมูล ที่จะนำมาทำการวิเคราะห์เพื่อหาความได้เปรียบทางการตลาดหรือเพื่อนำมาแก้ไขปัญหา

2) ส่วนของการทำเหมืองข้อมูล เป็นการนำเทคนิคของการทำเหมืองข้อมูลไปใช้ถ่ายทอดหรือทำการเปลี่ยนแปลงข้อมูลดิบให้อยู่ในรูปของข้อมูลจะนำไปใช้ได้จริงในทางธุรกิจ

3) การนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วนการทำเหมืองข้อมูลมาลองปฏิบัติจริงกับธุรกิจ

4) การวัดประสิทธิภาพของเทคนิคการทำเหมืองข้อมูลที่จะนำมาใช้จากผลลัพธ์ เช่น วัด จากส่วนแบ่งของตลาด วัดจากปริมาณลูกค่า หรือ วัดจากกำไรสุทธิ เป็นต้น

ดังนั้นจากทั้ง 4 ขั้นตอนที่ DPU 6 กล่าวมาข้างต้นคือการนำเอาเหมืองข้อมูล (Data Mining) ไปใช้กับระบบทางธุรกิจ โดยแต่ละขั้นตอนจะพึ่งพาอาศัยกันผลลัพธ์จากขั้นตอนหนึ่งจะกลายมาเป็นอินพุต (input) จากอีกขั้นตอนต่อไป ซึ่ง เหมืองข้อมูล (Data Mining) จะเปลี่ยนข้อมูลดิบให้เป็นข้อมูลประยุกต์ ดังนั้นการระบุแหล่งข้อมูลที่ถูกต้องจึงเป็นสิ่งที่สำคัญอย่างยิ่งต่อผลลัพธ์ที่ได้จากการวิเคราะห์

2.2.2.3 งานของการทำเหมืองข้อมูลในทางปฏิบัติจริง การทำเหมืองข้อมูลจะประสบความสำเร็จกับงานบางกลุ่มเท่านั้น และต้องอยู่ภายใต้ภาวะที่จำกัดปัญหาเหมาะสมกับการใช้เทคนิคการทำเหมืองข้อมูล จะเป็นปัญหาที่ต้องใช้เหตุผลในการแก้เป็นปัญหาที่เกี่ยวข้องกับเศรษฐศาสตร์และการเงิน ซึ่งจะสามารถจัดรูปแบบของธุรกิจให้อยู่ในรูปแบบของงานทั้ง 6 งานได้ ดังนี้

1) การจัดหมวดหมู่ (Classification)

การจัดหมวดหมู่ถือว่าเป็นงานธรรมดาทั่วไปของการทำเหมืองข้อมูล เพราะการทำความเข้าใจและการติดต่อสื่อสารต่าง ๆ ก็เกี่ยวข้องกับการแบ่งเป็นหมวดหมู่การจัดแยกประเภทและการแบ่งแยกชนิดโดยการจัดหมวดหมู่ประกอบด้วยการสำรวจจุดเด่นของวัตถุที่ปรากฏออกมา และทำการกำหนดจุดเด่นนั้น ๆ เป็นตัวที่ใช้แบ่งหมวดหมู่งานในการแบ่งหมวดหมู่คือการบ่งบอกลักษณะ โดยการอธิบายจุดเด่นที่เป็นที่รู้จักดีในหมวดหมู่นั้น และชุดข้อมูล เรียนรู้ (Training Set) ของตัวอย่างในแต่ละหมวดหมู่ซึ่งมีภาระหน้าที่ในการสร้างโมเดลของบางชนิดที่ไม่สามารถจะจัดหมวดหมู่ของข้อมูลได้ ให้สามารถจัดเป็นหมวดหมู่ได้ ตัวอย่างของการจัดหมวดหมู่ เช่นการจัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำระดับกลางและระดับสูงของความเสี่ยงที่จะได้รับเป็นต้น

2) การประเมินค่า (Estimation)

การประเมินค่าทางธุรกิจอย่างต่อเนื่องจะก่อให้เกิดผลลัพธ์ที่มีประโยชน์กับธุรกิจการป้อนข้อมูลที่เรามีอยู่เข้าไป เพื่อใช้ในการประเมินสิ่งต่าง ๆ ที่จะก่อให้เกิดประโยชน์หรือ สำหรับตัวแปรที่เราไม่รู้ค่า แน่นอนเช่นรายได้จากการค่าจุดสูงสุดทางธุรกิจ หรือดุลยภาพของบัตร DPU 7 เครดิต ในทางปฏิบัติการประเมินค่าจะถูกใช้ในการทำงานการจัดหมวดหมู่ตัวอย่างของการประเมินค่าเช่นการประเมินรายได้รวมของครอบครัวหรือการประเมินจำนวนบุตรในครอบครัว

3) การทำนายล่วงหน้า (Prediction)

การทำนายล่วงหน้า การทำนายล่วงหน้าก็เป็นงานที่มีลักษณะคล้ายกับการจัดหมวดหมู่หรือการประเมินค่า ยกเว้นเพียงแต่จะใช้สถิติการบันทึกของการจัดหมวดหมู่ในการทำนายอนาคตของพฤติกรรมหรือการประเมินค่าที่จะเกิดขึ้นในอนาคต ตัวอย่างของงานการทำนายล่วงหน้า เช่น การทำนายการเปลี่ยนแปลงพฤติกรรม ของตลาด หรือการทำนายจำนวนลูกค้าที่จะออกจากธุรกิจของเราใน 6 เดือนข้างหน้าเป็นต้น

4) การจัดกลุ่มโดยอาศัยความใกล้ชิด (Affinity Group)

การจัดกลุ่มโดยอาศัยความใกล้ชิดกันหรือการวิเคราะห์ของตลาดงานในการจัดกลุ่มหรือการวิเคราะห์ตลาด คือการตัดสินใจรวมสิ่งที่สามารถไปด้วยกันเข้าไว้ในกลุ่มเดียวกันตัวอย่างของการจัดกลุ่มโดยอาศัยความใกล้ชิดกันหรือการวิเคราะห์ตลาด เช่น การตัดสินใจว่าลูกค้าคนใดจัดอยู่ในกลุ่มค้าประเภทใด

5) การรวมตัว (Clustering)

การรวมตัว การรวมตัวคืองานที่ทำการรวมส่วนต่าง ๆ ในแต่ละส่วนที่ต่างชนิดกันให้อยู่ในรวมกันเป็นกลุ่มย่อย หรือคลัสเตอร์ (Clusters) โดยในแต่ละกลุ่มย่อยอาจจะประกอบด้วยส่วน ต่าง ๆ ที่ต่างชนิดกัน ซึ่งความแตกต่างของการรวมตัวจากการจัดหมวดหมู่ คือ การรวมตัวจะไม่พึ่งพาอาศัยการกำหนดหมวดหมู่ล่วงหน้า และไม่ใช้ตัวอย่างข้อมูลจะรวมตัวกันบนพื้นฐานของความคล้ายในตัวเอง

6) การบรรยาย (Description)

การบรรยาย ในบางครั้งวัตถุประสงค์ของการทำเหมืองข้อมูล คือต้องการอธิบายความสับสนของฐานข้อมูลในทางที่จะเพิ่มความเข้าใจในส่วนของประชากร ผลิตภัณฑ์ หรือขบวนการ ให้มากขึ้น เทคนิคการทำเหมืองข้อมูลส่วนใหญ่ต้องการเทรนนิ่งข้อมูลจำนวนมากที่ประกอบด้วยหลาย ๆ ตัวอย่างเพื่อจะสร้างกฎที่ใช้ในการจัดหมวดหมู่กฎของความสัมพันธ์กลุ่มย่อย การทำนายล่วงหน้า ดังนั้นชุดของข้อมูลขนาดเล็กจะนำไปสู่ความไม่น่าไว้วางใจของผลสรุปที่ได้ไม่มีเทคนิคใดเลย ที่จะสามารถแก้ปัญหาของการทำเหมืองข้อมูลได้ทุกปัญหา

ไม่มีเทคนิคหรือเครื่องมือเพียงชนิดเดียวของการทำเหมืองข้อมูลที่เหมาะสมกับงานทุกชนิด งานในแต่ละชนิดก็จะมีเทคนิคของการทำเหมืองข้อมูลที่แตกต่างกันไปขึ้นอยู่กับชนิดของงาน ดังนั้นความหลากหลายของเทคนิคจึงเป็นสิ่งที่จำเป็นในการไปสู่วิธีการแก้ปัญหาของ Data Mining ได้ดีที่สุด

2.2.2.4 เทคนิคการทำเหมืองข้อมูล

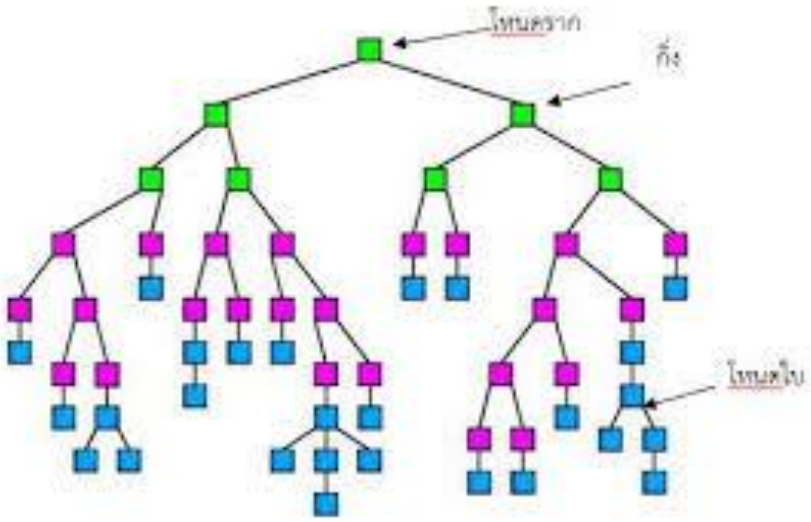
1) การจำแนกประเภทข้อมูล (Data Classification) การจำแนกประเภทข้อมูลเป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ โดยการสร้างกฎเพื่อช่วยในการตัดสินใจจากข้อมูลที่มีอยู่เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของข้อมูลที่ยังไม่เกิดขึ้น โดยการนำเสนอกฎที่ได้จากเทคนิคการจำแนก ประเภทข้อมูลนิยมนำเสนอในรูปแบบของแผนภูมิต้นไม้ซึ่งเรียกว่าต้นไม้ช่วยในการตัดสินใจ DPU 8

2) ต้นไม้ช่วยการตัดสินใจ (decision tree) เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ช่วยการตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ที่แต่ละโหนดแสดงคุณลักษณะ (attribute) แต่ละกิ่งแสดงเงื่อนไขในการทดสอบและโหนดปลาย (leaf node) แสดงกลุ่มที่กำหนดไว้ การแทนต้นไม้การตัดสินใจ (Decision Tree Representation)

– โหนดภายใน (internal node) คือ คุณสมบัติต่าง ๆ ของข้อมูล ซึ่งเมื่อข้อมูลใด ๆ ตกลงมาที่โหนด จะใช้คุณสมบัตินี้เป็นตัวตัดสินใจว่าข้อมูลจะไปในทิศทางใด โดยโหนดภายใน ที่เป็นจุดเริ่มต้นของต้นไม้เรียกว่าโหนดราก

– กิ่ง (Branch, link) เป็นค่าคุณสมบัติของคุณสมบัติในโหนดภายในที่แตกกิ่งนี้ออกมา ซึ่งโหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าคุณสมบัติของโหนดภายในนั้น

– โหนดใบ (leaf node) คือ กลุ่มต่าง ๆ ซึ่งเป็นผลลัพธ์ในการแยกแยะข้อมูล



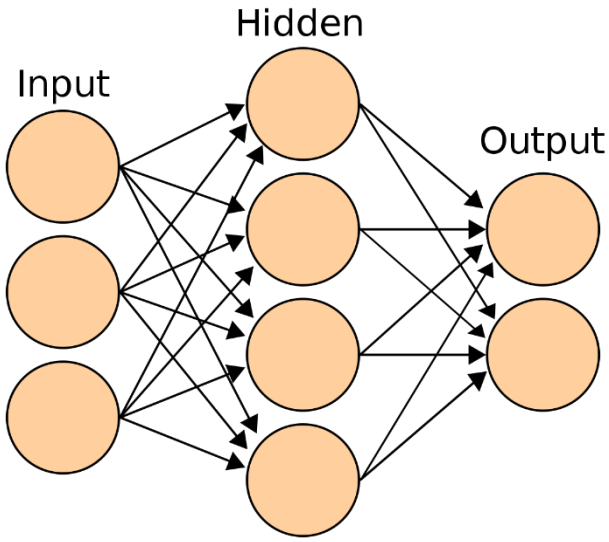
ภาพที่ 2.6 การแทนต้นไม้การตัดสินใจ

ที่มา : สุรวัชร ศรีเปารยะ* และสายชล สินสมบูรณ์ทอง (2560)

3) ตัวจำแนกประเภทเบย์อย่างง่าย (Naïve Bayes Classifier) เป็นตัวจำแนกประเภทแบบหนึ่งที่ใช้งานได้ดี เหมาะกับกรณีของเซตตัวอย่างมีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกันมีการนำตัวจำแนกประเภทเบย์อย่างง่ายไปประยุกต์ใช้งานในด้านการจำแนกประเภทข้อความ (text classification) การวินิจฉัย (diagnosis) และพบว่าใช้งานได้ดีไม่ต่างจากการจำแนกประเภทวิธีการอื่น Naïve Bayes เป็นเทคนิคที่ถูกตั้งชื่อตาม Thomas Bayes นำเสนอโดย Kohavi ในปี 1996 ได้ใช้ทฤษฎี Bayes Theorem ในการแก้ปัญหาเรื่องการจำแนกประเภทข้อมูลที่ใช้แบบจำลองของการทำนาย โดยจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัวกับตัวแปรตาม ทั้งนี้เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ทางทฤษฎี การทำนาย A B C D E F G H I DPU 9 ผลของ Naïve–Bayes จะมีความถูกต้องเมื่อตัวแปรอิสระทั้งหมดมีความเป็นอิสระต่อกัน ซึ่งจะไม่ขึ้นกับตัวแปรอิสระตัวใดตัวหนึ่งข้อจำกัดของอัลกอริทึม Naïve Bayes เทคนิคแบบ Naïve–Bayes จะไม่รองรับข้อมูลที่เป็นค่าต่อเนื่อง (Continuous Data) ดังนั้นตัวแปรอิสระหรือตัวแปรตามที่มีค่าต่อเนื่องจะต้องมีการแบ่งเป็นข้อมูลเป็นช่วง ๆ เช่น อายุ งาน เป็นต้น จะเห็นได้ว่าถ้ามีการแบ่งช่วงของข้อมูลไม่เหมาะสมจะมีผลต่อคุณภาพของแบบจำลองที่สร้างขึ้นได้

4) กฎการตัดสินใจ (Decision rule) สามารถสร้างขึ้นจากต้นไม้การตัดสินใจได้ง่ายโดยจะอยู่ในรูปของกฎ IF...THEN โดยที่ IF{ส่วนของการกำหนดเงื่อนไข} THEN {ส่วนของการ กระทำตามเงื่อนไข} กฎการตัดสินใจจะนำเสนอรูปของ Production Rules ในลักษณะของ IF...Then ที่ใช้ในการแก้ปัญหา สามารถสร้างกฎได้ง่ายและเป็นที่เข้าใจของผู้นำกฎนี้ไปใช้งาน ส่วนประกอบของ กฎ มีด้วยกัน 2 ส่วน ได้แก่ IF ที่จะเรียกว่า Antecedent เป็นส่วนของการกระทำตามเงื่อนไขที่ กำหนดหรืออาจเป็นส่วนของการแสดงบทสรุปว่าจะเป็นอะไรหรือจะได้อะไรเป็นต้น เราสามารถเชื่อมกฎที่สร้างขึ้นด้วยคำเชื่อม AND , OR ซึ่งจะเรียกว่า Production rules

5) เครือข่ายใยประสาท (Neural Network) เป็นเทคโนโลยีที่มาจากงานวิจัยด้าน ปัญญาประดิษฐ์ (Artificial Intelligence: AI) เพื่อใช้ในการคำนวณค่าฟังก์ชั่นจากกลุ่มข้อมูล วิธีการของเครือข่ายใยประสาทเป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบแล้วฝึกให้ระบบได้รู้จักคิดแก้ปัญหาที่กว้างขึ้นได้ในโครงสร้างของเครือข่ายใยประสาทประกอบด้วยโหนด (Node) สำหรับ Input–output และการประมวลผลกระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ Input layer, output layer และ Hidden layers การประมวลผลของเครือข่ายใยประสาทจะอาศัยการส่งการทำงานผ่านโหนด ต่าง ๆ ใน Layers เหล่านี้ ตัวอย่างเครือข่ายใยประสาท



ภาพที่ 2.7 เครือข่ายใยประสาท

ที่มา : wikipedia.org (2563)

6) การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคการแบ่งกลุ่มหน่วยข้อมูลหรือเป็นการแบ่งคน สัตว์ สิ่งของ องค์กร ฯลฯ ออกเป็นกลุ่มย่อยอย่างน้อย 2 กลุ่ม โดยมีหลักเกณฑ์ในการแบ่งดังนี้ “ให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่สนใจเหมือนกันหรือคล้ายกันแต่หน่วยที่อยู่ต่างกลุ่มกันจะมีลักษณะที่สนใจต่างกัน” คำว่าลักษณะที่สนใจอาจจะมีหลาย ๆ ตัวแปร เช่นถ้าสนใจความคิดเห็นทางด้านการเมืองจะมี คำถามหลาย ๆ คำถามด้านการเมือง และจะนำคำตอบเหล่านั้นมาแบ่งกลุ่ม

7) การค้นหากฎความสัมพันธ์ (Association rule discovery) การค้นหาความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนาย ปรากฏการณ์ต่าง ๆ เป็นการประเมินจากข้อมูลในฐานข้อมูลที่รวบรวมไว้ โดยกฎความสัมพันธ์ที่ได้มีส่วนช่วยในกระบวนการตัดสินใจทางด้านธุรกิจ ตัวอย่างของการค้นหากฎความสัมพันธ์ คือ การวิเคราะห์พฤติกรรมของผู้บริโภค โดยการหาความสัมพันธ์ระหว่างสินค้าที่ต่างชนิดกัน

8) K-Nearest Neighbour (K-NN) คือ วิธีการในการจัดแบ่งคลาส เทคนิคนี้จะตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน ("K" ใน K-nearest neighbor) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะ DPU 11 หาผลรวม (Count Up) ของจำนวนเงื่อนไข หรือกรณีต่าง ๆ สำหรับแต่ละคลาสและกำหนดเงื่อนไข ใหม่ ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด

2.2.3 ทฤษฎีเกี่ยวออกแบบ Dashboard

แดชบอร์ด เป็นแบบหน้าเดียว ซึ่งมักเรียกว่าพื้นที่ว่างเปล่าที่ใช้การแสดภาพเพื่อบอกเล่าเรื่องราวหนึ่ง ๆ ได้ เนื่องจากจำกัดอยู่เพียงหนึ่งหน้าแดชบอร์ดที่ออกแบบมาอย่างดีจะประกอบด้วยองค์ประกอบที่สำคัญที่สุดของเรื่องราวเท่านั้น แดชบอร์ดเป็นวิธีที่ยอดเยี่ยมในการตรวจดูธุรกิจของคุณ ในการค้นหาคำตอบและดูเมตริกที่สำคัญที่สุดของคุณอย่างรวดเร็ว การแสดงภาพบนแดชบอร์ดอาจมาจากหนึ่งหรือหลายชุดข้อมูลพื้นฐานและจากรายงานหนึ่งหรือหลายรายงานพื้นฐาน แดชบอร์ดสามารถรวมข้อมูลภายในองค์กรและข้อมูลบนระบบคลาวด์ ให้มุมมองแบบรวมโดยไม่คำนึงถึงตำแหน่งที่ข้อมูลอยู่แดชบอร์ดไม่ได้เป็นเพียงภาพที่สวยงาม เนื่องจากแดชบอร์ดสามารถโต้ตอบได้และมีการอัปเดตไทล์เมื่อมีการเปลี่ยนแปลงข้อมูลพื้นฐาน

แดชบอร์ด (Dashboard) สร้างโดยเครื่องมือใดโปรแกรมง่าย ๆ ในการช่วยทำ แดชบอร์ด มีหลายเครื่องมือ หากเป็นโปรแกรมที่เรามีอาจจะเป็น Excel โดยใช้ความสามารถของ Pivot Table, Pivot Chart แต่หากข้อมูลมีจำนวนมาก อาจจะใช้เครื่องมือในการทำ Business Intelligence (BI) เพิ่มเติมอย่างโปรแกรมอื่น ๆ ตัวอย่างเช่น PowerBI Desktop,Tableau, SimpleKPI, InfoCaptor

2.2.3.1 แดชบอร์ด (Dashboard) สามารถแบ่งออกเป็น 3 ประเภทใหญ่ ๆ คือ

1) Strategic dashboard จะเป็นข้อมูลที่ใช้สำหรับประกอบการวางแผนด้านกลยุทธ์ การวางแผนที่มีการกำหนดวิสัยทัศน์ มีการกำหนดเป้าหมายระยะยาวที่แน่ชัด มีการวิเคราะห์อนาคตและคิดเชิงการแข่งขันที่ต้องการระบบการทำงานที่มีความสามารถในการปรับตัวสูง สำหรับการทำงานในสิ่งแวดล้อมที่มีการเปลี่ยนแปลงอย่างรวดเร็ว เพื่อให้ทันกับการเปลี่ยนแปลงที่เกิดขึ้น ต้องการระบบการทำงานที่คล่องตัว ต้องการดำเนินงานมีประสิทธิภาพสูงในการนำสู่เป้าหมายในอนาคต

2) Analytical dashboard จะเป็นข้อมูลที่ใช้สำหรับประกอบการวิเคราะห์ เป็นการนำข้อมูลการปฏิบัติงานที่รวบรวมได้จากการใช้เครื่องมือ (Instrument) เพื่อให้ได้ผลการวัดที่เรียกว่า “ข้อมูล” มาจัดกระทำหรือหรือจำแนกกลุ่ม จัดประเภทคำนวณค่า สรุปและแนะนำเสนอในรูปแบบที่เหมาะสม

3) Operational dashboards จะเป็นข้อมูลที่ใช้สำหรับประกอบการปฏิบัติงานเป็นการรายงานสรุปข้อมูลทั้งภายนอกและภายในที่เกี่ยวข้องกับองค์การและเป็นสิ่งที่จำเป็นที่จะต้องนำมาเป็นพื้นฐานในการพิจารณาการวางแผนการบริหารองค์การ จำเป็นต้องนำข้อมูลต่าง ๆ มาประกอบการพิจารณา

2.2.3.2 หลักในการออกแบบ แดชบอร์ด 4 ขั้นตอน

1) Users การเข้าใจผู้ใช้ก่อนเป็นอันดับแรก การเข้าใจผู้ใช้นั้นให้ยึดตามหลักการของ Design Thinking จงหา Insight ของผู้ใช้ หรือตอบให้ได้ว่า ทำไมถึงใช้คำถามหลักของการออกแบบ แดชบอร์ด ก็คือ ผู้ใช้จะเอา แดชบอร์ด ไปใช้ทำอะไร เช่น ช่วยตัดสินใจ ติดตามผลการดำเนินงาน ใช้เตือนเมื่อมีสิ่งผิดปกติในอดีต ทำการไปเก็บ Requirement กับผู้ใช้ ถามหาความต้องการของผู้ใช้ อยากดูข้อมูลอะไรบ้าง อยากดูเป็นกราฟแบบไหน ซึ่งบางครั้งอาจไม่ตรงกับความต้องการของผู้ใช้อันนี้เกิดจากการที่เราไม่ได้เข้าใจจริง ๆ ว่าผู้ใช้อยากจะเอาไปใช้ทำอะไร ถ้าเราเข้าใจถึงรูปแบบว่าผู้ใช้จะเอา แดชบอร์ด ไปใช้อย่างไร เราก็จะออกแบบได้ตรงใจกับผู้ใช้มากขึ้น

2) Content เมื่อเข้าใจแล้วว่า ผู้ใช้อยากเอา แดชบอร์ด ไปใช้อย่างไร ก็ต้องวิเคราะห์ เนื้อหา หรือ สิ่งที่ผู้ใช้จะต้องเอาไปใช้นั้น มีอะไรบ้าง ซึ่งก็แบ่งเป็น measures หรือ ตัวเลขที่สนใจ เช่น ยอดขาย จำนวนลูกค้า dimensions หรือ มุมมองที่อยากวิเคราะห์ข้อมูล เช่น ตามช่วงเวลา ตามกลุ่มสินค้า ตามพื้นที่การเลือก measures นั้น สามารถช่วยคิดช่วยออกแบบให้การวัดผลนั้น มีความน่าสนใจ หรือ ตรงประเด็นมากขึ้น จะทำให้ Dashboard นั้น มีความน่าสนใจมากขึ้น

3) Presentation หลังจากที่รู้แล้วว่า จะใช้ measures และ dimensions แบบไหน ต้องทำการเลือกกราฟให้ถูกต้องตามแบบงานหรือรูปแบบของข้อมูลเพื่อสร้างการนำเสนอที่มีเนื้อครบถ้วนและน่าสนใจ

4) Navigation การประกอบกันเป็น Dashboard การจัดวางกราฟเป็นส่วนสำคัญ คือ กราฟที่เป็นเรื่องเดียวกัน จะต้องวางไว้ใกล้ ๆ กัน เพื่อไม่สร้างความยุ่งยากให้กับผู้อ่านและอาจส่งผลให้ไม่เจอความเชื่อมโยง ซึ่งควรจัดวางกราฟให้มี visual hierarchy จากภาพใหญ่ ไปภาพย่อยการแสดงผลข้อมูลด้วยภาพที่คุณเห็นบนแดชบอร์ดเรียกว่าไทล์ และ

จะถูกปักหมุดไปยังแดชบอร์ดโดยรายงานของผู้ออกแบบ ในกรณีส่วนใหญ่การเลือกไทล์ที่จะนำไปยังหน้ารายงานที่การแสดงภาพถูกสร้างขึ้น

2.2.4 ทฤษฎีเกี่ยวกับการออกแบบ Visualization

รูปแบบของจินตทัศน์ (Visualization) รูปแบบของจินตทัศน์ที่ใช้ในการแสดงผลสารสนเทศนั้นมีจำนวนมาก ซึ่งการเลือกใช้ก็ขึ้นอยู่กับสารสนเทศที่ต้องการแสดง รวมถึงผลลัพธ์ที่ต้องการให้เกิดการตอบสนองต่อ สารสนเทศนั้น ๆ ในงานวิจัยนี้มุ่งศึกษาแนวทางการออกแบบแผงหน้าปัดโดยใช้ซอฟต์แวร์ IBM Cognos ดังนั้นจึงมีรูปแบบแผนภูมิที่สามารถสร้างบนซอฟต์แวร์ IBM Cognos ตามที่ได้อธิบายไว้ในคู่มือการใช้งาน IBM Cognos Business Intelligence V10.1 Handbook ดังนี้

1) แผนภูมิแท่ง (Column and Bar Chart) แผนภูมิแท่งใช้ในการแสดงข้อมูลตามช่วงเวลาหรือข้อมูลที่ไม่ต่อเนื่องกัน โดยมีทั้งแผนภูมิแท่งแนวตั้งและแนวนอน ซึ่งสามารถเปรียบเทียบข้อมูลได้ทั้งข้อมูลเดียว หรือหลายข้อมูล เหมาะต่อการดูข้อมูลที่เรียงลำดับตามค่าต่ำสุดหรือสูงสุด นอกจากนี้แผนภูมิแท่ง สามารถแสดงข้อมูลหลาย ๆ ข้อมูลเรียงกันเป็นชั้นภายในแผนภูมิแท่งเดียวกันเป็นแผนภูมิแท่งแบบชั้น (Stack Charts) (Browne et al., 2010) ซึ่งเป็นรูปแบบที่เหมาะกับการแสดงข้อมูลที่มีความสัมพันธ์เป็นชุด เพื่อให้ค่าที่แสดงสามารถเปรียบเทียบกันได้ง่ายเมื่อมีค่าข้อมูลที่ต้องเปรียบเทียบแสดงอยู่ด้วยกัน อาจใช้ในการแสดงข้อมูลตามช่วงเวลาสั้น ๆ ข้อมูลการขายตามหมวดหมู่สินค้า (Kerzner, 2011) สามารถใช้การวัดหลายอย่างได้ภายในแผนภูมิเดียวกันโดยที่ไม่ทำให้ข้อมูลดูหนาแน่นเกินไป สามารถใช้ในการแจ้งเตือนผู้ใช้แผงหน้าปัดถึงค่าที่ผิดปกติไปจากเป้าหมายที่วางไว้ ข้อควรระวังคือ ต้องมั่นใจในความถูกต้องของข้อมูล เนื่องจากถ้าค่าข้อมูลผิดก็จะทำให้แผนภูมิแสดงความสัมพันธ์และแนวโน้มที่ผิด (Smietana, 2010)

2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้นมีความคล้ายคลึงกับแผนภูมิแท่งเพียงแต่จะใช้จุดเป็นตัวบอกค่าของข้อมูลแทนตัวแท่ง แล้วมีการลากเส้นเชื่อมแต่ละจุด จึงเหมาะสำหรับการแสดงข้อมูล แนวโน้มหลาย ๆ แนวโน้มเปรียบเทียบกัน แต่ก็ไม่ควรใช้ในการเปรียบเทียบข้อมูลหลาย ๆ ค่า เพราะจะแยกความแตกต่างระหว่างข้อมูลได้ยากกว่าการแสดงแบบเป็นชั้นในแผนภูมิแท่ง (Browne et al., 2010) เหมาะสำหรับการวิเคราะห์อนุกรมเวลาที่ต้องการที่จะเห็นการเปลี่ยนแปลงที่เกิดขึ้นของข้อมูล หนึ่งหรือมากกว่าหนึ่งข้อมูลการวัดค่าในช่วงเวลา แผนภูมิเส้นยังให้การเปรียบเทียบการวิเคราะห์แนวโน้มใช้ข้อมูลหลายชุดซ้อนกันในหนึ่งแผนภูมิ (Kerzner, 2011) ในแผนภูมิเส้นสามารถใช้จุดบอกตำแหน่งของข้อมูลที่ชัดเจนได้ เนื่องจากถ้าใช้แต่เส้นก็จะทราบเฉพาะแนวโน้มของข้อมูล หรือการใช้จุดอย่างเดียว

ก็จะเห็นแต่ข้อมูล ณ ตำแหน่งนั้น ๆ แต่จะไม่เห็นแนวโน้ม ปัจจุบันจึงมีการใช้แผนภูมิเส้น และจุด (Spline) ด้วยกัน เพื่อให้เห็นทั้งค่าข้อมูลและแนวโน้ม (Smietana, 2010) ควรทำเส้นให้จุดของข้อมูลเด่นว่ามีข้อมูลที่ตำแหน่งใด (Juice Inc., 2009)

3) แผนภูมิพื้นที่ (Area Chart) แผนภูมิพื้นที่สามารถใช้สีในบริเวณพื้นที่ใต้กราฟได้แทนการลากเส้น เพื่อแสดงแนวโน้มของข้อมูลเพื่อใช้ในการเปรียบเทียบ (Browne et al., 2010) การใช้สีเพื่อแสดงพื้นที่ใต้หรือเหนือกราฟของแผนภูมิเส้น ใช้เมื่อต้องการเปรียบเทียบแนวโน้มตามช่วงเวลาของ ข้อมูล (Kerzner, 2011)

4) แผนภูมิจุด (Point Charts) แผนภูมิจุดมีลักษณะคล้ายแผนภูมิเส้นเพียงแต่แสดงเฉพาะตำแหน่งค่าข้อมูล โดยไม่มีการลากเส้นเพื่อแสดงแนวโน้ม แผนภูมิจุดมีประโยชน์สำหรับการแสดง ข้อมูลเชิงปริมาณ โดยใช้จุดแทนข้อมูลตามแกนโดยวางตามลำดับ (Browne et al., 2010)

5) แผนภูมิแบบผสม (Combination Charts) แผนภูมิแบบผสมใช้ในการแสดงค่าข้อมูลหลายค่าโดยใช้แผนภูมิ แท่ง แผนภูมิพื้นที่ หรือแผนภูมิเส้น รวมไว้ในแผนภูมิเดียวกัน มีประโยชน์ในการเน้นความสัมพันธ์ของข้อมูลแต่ละชุด (Browne et al., 2010)

6) แผนภูมิแบบกระจาย (Scatter Plot) แผนภูมิแบบกระจายมักใช้ในการเปรียบเทียบความสัมพันธ์ของข้อมูลระหว่างชุดเพื่อดูความแตกต่างตามตำแหน่งของข้อมูลที่ปรากฏบนแผนภูมิ (Browne et al., 13 2010) ใช้ในกรณีที่ต้องการดูความสัมพันธ์ของข้อมูลมากกว่าค่าของข้อมูลจะเป็นการดูแนวโน้มใน ภาพรวมว่าข้อมูลชุดนี้มีค่าไปในทิศทางใด (Smietana, 2010)

7) แผนภูมิฟอง (Bubble Charts) แผนภูมิฟองมีลักษณะคล้ายแผนภูมิแบบกระจายเพียงแต่จะมีตัววัดเพิ่มขึ้นโดยเป็นขนาดของฟองตามค่าของข้อมูล โดยมักใช้ในการแสดงข้อมูลด้านการเงิน (Browne et al., 2010) แสดงค่าเป็นกลุ่มของข้อมูลด้วยขนาดของฟอง (วงกลม หรือ ทรงกลม) ที่แตกต่างกันตามค่าของข้อมูล ใช้ในการเปรียบเทียบระหว่างข้อมูลเดียวกันแต่ต่างช่วงเวลา หรือเปรียบเทียบ ระหว่างข้อมูลที่แตกต่างกันด้วย (Kerzner, 2011) การเลือกใช้ควรแน่ใจในเรื่องการอธิบายขนาดของฟองสบู่แต่ละฟอง ว่ามีการกำหนดค่าได้ถูกต้องและชัดเจน (Smietana, 2010)

8) แผนภูมิสี่ส่วน (Quadrant Charts) โดยหลักแล้วจะใช้แผนภูมิสี่ส่วนเป็นแกนพื้นหลังให้แผนภูมิฟอง ซึ่งจะแบ่งพื้นที่เป็นสี่ส่วนเท่า ๆ กัน ในการแบ่งส่วนของแผนภูมิอาจใช้สีในการแยกความแตกต่างของแต่ละส่วน ซึ่งขนาดของส่วนที่แบ่งอาจแบ่งย่อยลงไปได้อีกเสมือนเป็นมาตรวัดในแผนภูมิ สามารถใช้แผนภูมิสี่ส่วนในการแบ่งหมวดหมู่เพื่อวางข้อมูลตาม

หมวดหมู่เช่นการวิเคราะห์จุดอ่อนจุดแข็งโอกาส และอุปสรรค (SWOT Analysis) เป็นต้น (Browne et al., 2010)

9) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมใช้ในการแสดงความสัมพันธ์ของข้อมูลในส่วนนั้น เปรียบเทียบกับสัดส่วนทั้งหมดของแผนภูมิ ซึ่งหากต้องการแสดงข้อมูลจริงเป็นแผนภูมิแท่งแบบเป็น ชั้นมากกว่าแผนภูมิวงกลมไม่ใช่ตัวเลือกที่ดีในการแสดงข้อมูลที่มีค่าเป็นศูนย์หรือมีค่าติดลบ (Browne et al., 2010) เป็นแผนภูมิที่มีองค์ประกอบมากกว่าแผนภูมิแบบอื่น โดยเป็นเรื่องยากมากที่จะแยกแยะความแตกต่างของสัดส่วนของข้อมูลกับการแบ่งพื้นที่ตามค่าข้อมูล นอกจากจะเป็นกรณีของชุดข้อมูลขนาดเล็กที่มีความแตกต่างของค่าข้อมูลอยู่มาก และแผนภูมิวงกลมที่ยังก่อให้เกิดปัญหาสำหรับการใส่คำอธิบายเพราะมีทั้งสีของแผนภูมิและรายละเอียดต่าง ๆ ภายใน ซึ่งการใส่คำอธิบายอื่นเพิ่มอาจก่อให้เกิดการสับสนในการดูข้อมูล (Kerzner, 2011) แผนภูมิวงกลมเป็นแผนภูมิสำหรับการแสดงองค์ประกอบข้อมูล ที่มีส่วนประกอบค่อนข้างน้อยและมีเพียงหนึ่งตัวชี้วัดเชิงปริมาณ และการแสดงแผนภูมิวงกลมต้องคิดค่าข้อมูลรวมเป็น 100% และข้อมูลที่ไม่เป็นค่าลบ ใช้การแสดงค่าข้อมูลทั้งหมดด้านนอกแผนภูมิ โดยเรียงค่าข้อมูลตามลำดับที่แสดงในแผนภูมิ (Smietana, 2010)

10) แผนภูมิจุดนำ (Bullet Charts) แผนภูมิจุดนำเป็นรูปแบบหนึ่งของแผนภูมิแท่งเป็นการแสดงตัว วัดที่สำคัญ ใช้ในการเปรียบเทียบตัววัดตั้งแต่ 1 ตัวหรือมากกว่านั้น โดยสามารถวัดโดยเปรียบเทียบ กับสีของพื้นหลังที่จะแสดงตัววัดที่มากขึ้น เช่น ระดับความพึงพอใจที่ดี จนถึงระดับที่ไม่ดี และเนื่องจากเป็นแสดงสารสนเทศที่ไม่จำเป็นต้องใช้พื้นที่มาก จึงสามารถใส่แผนภูมิจุดนำเป็นส่วนหนึ่งในการแสดงรายงานบนแผงหน้าปัด (Browne et al., 2010) จะแสดงการวัดมูลค่าในแกนแนวนอน ซึ่งอาจมีการเพิ่มบริบทข้องข้อมูลเข้าไปโดยการใช้สีเป็นตัวบอกระดับความพึงพอใจ ทำให้มองข้อมูลเข้าใจได้ในครั้งเดียว (Smietana, 2010)

11) แผนภูมิมาตรวัด (Gauge Charts) แผนภูมิมาตรวัดหรืออาจเรียกว่าแผนภูมิหน้าปัด (Dial Charts) หรือแผนภูมิมาตรอัตราเร็ว (Speedometer Charts) มีความคล้ายกับแผนภูมิจุดนำในแง่การเปรียบเทียบตัววัดหลายค่าเพียงแต่ใช้เข็มในการแสดงค่า ซึ่งการอ่านค่าสามารถทำได้ง่ายเหมือนการอ่านค่าจากหน้าปัดและแต่ละค่าที่แสดงจะเปรียบเทียบด้วยช่วงสีที่อยู่บนแผนภูมิ โดยแผนภูมิมาตรวัดนั้นเป็นทางเลือกที่ดีกว่าแผนภูมิจุดนำเมื่อต้องเปรียบเทียบค่ามากกว่าสองค่าขึ้นไป โดยแผนภูมินี้มักถูกใช้ในการแสดงตัวชี้วัดผลการดำเนินงานหลักบนแผงหน้าปัดสำหรับผู้บริหารระดับสูง สำหรับการออกรายงานโดยใช้ PDF และ HTML จะจำกัดอยู่ที่แผนภูมิ และใน Microsoft Excel จะไม่สนับสนุนการแสดงผลแผนภูมิประเภทนี้ (Browne et al., 2010) ใช้ในการบอกสถานะปัจจุบันโดยจะใช้สีแดง ส้ม เหลือง เขียว ในการบอกถึงระดับ

ของปัญหาตั้งแต่มาก (สีแดง) จนถึงระดับปกติ (สีเขียว) สามารถช่วยให้เห็นความสัมพันธ์ระหว่างค่าที่วัด แต่ว่าจะค่อนข้างสิ้นเปลืองเนื้อที่ จึงควรเลือกใช้เมื่อมีพื้นที่มากเพียงพอ (Smietana, 2010)

12) แผนภูมิพาเรโต (Pareto Charts) แผนภูมิพาเรโตเป็นลำดับของหมวดหมู่ที่เรียงจากความถี่มากที่สุดไปจนถึงน้อยที่สุด โดยจะมีเส้นการสั่งสม (Cumulation Line) ซึ่งเป็นการแสดงอัตราส่วนร้อยละยอดรวมสะสมทั้งหมดของแผนภูมิแท่ง โดยแผนภูมิชนิดนี้จะใช้ในการควบคุมคุณภาพ ทำให้สามารถระบุและลดสาเหตุของปัญหาที่เกิดขึ้นตามลำดับความสำคัญของปัญหาและสามารถที่จะสร้างแผนภูมิพาเรโตได้ทั้งก่อนและหลังการแก้ปัญหาเพื่อเป็นการเปรียบเทียบให้เห็นถึงสิ่งที่เปลี่ยนแปลงไป (Browne et al., 2010)

13) แผนภูมิแท่งแบบก้าวหน้าหรือแผนภูมิน้ำตก (Progressive Column Charts or Waterfall Charts) เป็นแผนภูมิที่แสดงการเปลี่ยนแปลงที่เกิดขึ้นต่อจากแผนภูมิแท่ง หรือแผนภูมิแท่งแบบเป็นชั้น โดยมักใช้ในการเปรียบเทียบค่าข้อมูล 1 ข้อมูลต่อข้อมูลทั้งหมด โดยใน Microsoft Excel ไม่สนับสนุนการแสดงผลแผนภูมิประเภทนี้ (Browne et al., 2010) 15

14) แผนภูมิมาริเมกโก (Marimekko Charts) เป็นแผนภูมิชั้นซึ่งความกว้างของชั้นของแท่งแผนภูมิจะเป็น สัดส่วนของข้อมูล 1 ข้อมูลต่อข้อมูลทั้งหมด โดยส่วนที่เป็นความสูงของแต่ละแนวแท่งแผนภูมินั้นจะเป็นอัตราร้อยละของมูลค่ารวมแนวตามลำดับ (Browne et al., 2010)

15) แผนภูมิเรดาร์หรือแผนภูมิแมงมุม (Radar or Spider Charts) แผนภูมิเรดาร์เป็นแผนภูมิที่ใช้ในการเปรียบเทียบข้อมูลหลายค่าตามแกนหลายแกนโดยจะใส่ค่าข้อมูลโดยเริ่มจากตรงกลางที่มีแกนวัดค่าเป็นรัศมี มีประโยชน์ในการเปรียบเทียบข้อมูลหลายชุดและหลายตัววัดและมีประโยชน์ในการดูค่าผิดปกติ (Browne et al., 2010) ข้อมูลจะถูกใส่ค่าตามระยะห่างจากจุดศูนย์กลาง แต่ละจุดข้อมูลที่มีส่วนประกอบในแนวตั้ง คือระยะห่างจากจุดศูนย์กลางของแกน แผนภูมิเรดาร์มีประโยชน์สำหรับการแสดงข้อมูลวัฏจักร เช่นระดับการรับพนักงานรายวันหรือผลรวมรายได้รายเดือนในการออกแบบสามารถวางซ้อนเป้าหมายและเกณฑ์การให้บริบทสำหรับตัวชี้วัดเชิงปริมาณ (Smietana, 2010)

16) แผนภูมิผลได้เสีย (Win–Loss Charts) เป็นแผนภูมิขนาดเล็ก (Microcharts) ที่ใช้ในการวัดค่าที่ตั้งอัตโนมัติ (Default Measure) และค่าได้หรือเสียจากค่าที่ตั้งไว้ ซึ่งแผนภูมิผลได้เสียนั้นเป็นการแสดงผลของเรื่องที่กำหนด และสามารถกำหนดค่าที่ตั้งอัตโนมัติได้เช่นกัน ซึ่งสามารถใช้แผนภูมินี้ในการแสดงจินตทัศน์สำหรับแนวโน้มผลการดำเนินงาน (Browne et al., 2010)

17) แผนภูมิขั้วโลก (Polar Charts) เป็นแผนภูมิวงกลมที่ใช้ค่าข้อมูลและขนาดของมุมเพื่อแสดงสารสนเทศเป็นลักษณะขั้วพิกัด (Browne et al., 2010)

18) เส้นฐานและเส้นแนวโน้ม (Baselines and Trend lines) เส้นฐานและเส้นแนวโน้มนั้นจะแสดงเป็นข้อมูลรายละเอียดเพิ่มเติมบนแผนภูมิ เส้นฐานนั้นสามารถอยู่ได้ทั้งแนวตั้งและแนวนอนที่ตัดผ่านแผนภูมิเพื่อระบุถึงส่วนที่สำคัญของข้อมูลสำหรับเส้นแนวโน้มนั้นใช้เป็นแนวโน้มของข้อมูลเมื่อต้องการพยากรณ์ค่า โดยส่วนมากเส้นแนวโน้มจะเป็นเส้นตรงหรือเส้นโค้งที่ลากผ่านจุดอย่างน้อย 2 จุด เพื่อแสดงแนวโน้มซึ่งสามารถใช้ในแผนภูมิแท่ง แผนภูมิเส้น แผนภูมิพื้นที่ แผนภูมิฟอง และแผนภูมิแบบกระจาย

2.2.5 ทฤษฎีเกี่ยวกับการออกแบบกราฟิกด้วยโปรแกรม Photoshop CS6

โปรแกรม Adobe Photoshop CS6 Adobe Photoshop CS6 เป็นโปรแกรมในตระกูล Adobe ที่ใช้สำหรับ ตกแต่งภาพถ่ายและภาพกราฟฟิก ได้อย่างมีประสิทธิภาพ ไม่ว่าจะเป็นงานด้านสิ่งพิมพ์ นิตยสาร และงานด้านมัลติมีเดีย อีกทั้งยังสามารถ Retouching ตกแต่งภาพและสร้างภาพ ซึ่งกำลังเป็นที่นิยมสูงมาก ในขณะนี้เราสามารถนำโปรแกรม Adobe Photoshop CS6 ในการแต่งภาพ การใส่ Effect ต่าง ๆ ให้กับภาพ และตัวหนังสือ การทำภาพขาวดำและการทำภาพถ่ายเป็นภาพเขียน การนำภาพต่าง ๆ มารวมกัน การ Retouch ตกแต่งภาพ เป็นต้น โปรแกรม Adobe Photoshop CS6 เป็น โปรแกรมที่มีเครื่องมือมากมายเพื่อสนับสนุนการสร้างงานประเภทสิ่งพิมพ์ งานวิดีทัศน์ งานนำเสนอ งานมัลติมีเดีย ตลอดจนงานออกแบบและพัฒนาเว็บไซต์ในชุดโปรแกรม Adobe Photoshop CS6 ดังนั้น โปรแกรม Adobe Photoshop CS6 จึงเป็นโปรแกรมที่มีความนิยมสูงและเหมาะสมกับการสร้างชิ้นงานด้านกราฟฟิกการแก้ไขภาพและการออกแบบประเภทต่าง ๆ ส่วนประกอบสำคัญของ Adobe Photoshop CS6 มีดังนี้

2.2.5.1 ความสามารถพื้นฐานของ Adobe Photoshop

1) ตกแต่งหรือแก้ไขรูปภาพ

2) ตัดต่อภาพบางส่วน หรือที่เรียกว่า crop ภาพ

3) เปลี่ยนแปลงสีของภาพจากสีหนึ่งเป็นอีกสีหนึ่งได้

4) สามารถลากเส้นแบบฟรีสไตล์หรือใส่รูปภาพ สี่เหลี่ยม วงกลม หรือสร้างภาพได้อย่างอิสระ

5) มีการแบ่งชั้นของภาพเป็น Layer สามารถเคลื่อนย้ายภาพได้เป็นอิสระต่อกัน

6) การทำ cloning ภาพ หรือการทำภาพซ้ำในรูปภาพเดียวกัน

7) เพิ่มเติมข้อความ ใส่ effect ของข้อความได้

8) Brush หรือแปรงทาสี ที่สามารถเลือกรูปแบบสำเร็จรูปในการสร้างภาพได้และอื่น ๆ อีกมากมา

2.2.5.2 แถบเมนูคำสั่ง (Menu Bar) เป็นจุดรวบรวมชุดคำสั่งที่ใช้สำหรับเรียกใช้คำสั่งต่าง ๆ เพื่อใช้จัดการไฟล์ภาพหรือตกแต่งภาพ

1) เมนูคำสั่ง (File) รูปแบบการทำงาน สำหรับจัดการกับไฟล์ภาพในลักษณะต่าง ๆ เช่น สร้างไฟล์งานใหม่ เปิดไฟล์ภาพบันทึกไฟล์งานนำเข้าหรือส่งออกไฟล์เพื่อทำงานในลักษณะอื่น ๆ

2) เมนูคำสั่ง (Edit) รูปแบบการทำงาน สำหรับแก้ไขภาพ เช่น ตัด คัดลอก วาง รวมถึงปรับแต่งค่าเบื้องต้นของโปรแกรม

3) เมนูคำสั่ง (Image) รูปแบบการทำงาน สำหรับจัดการภาพ เช่น แก้ไขความสว่างหรือสีของภาพให้สมดุลยิ่งขึ้น รวมถึงใช้สำหรับย่อขยายขนาดภาพ และกำหนดขนาดพื้นที่การทำงานของภาพ

4) เมนูคำสั่ง (Layer) รูปแบบการทำงาน สำหรับจัดการเกี่ยวกับเลเยอร์ เช่น การสร้างเลเยอร์ใหม่ การรวมเลเยอร์ การแปลงเลเยอร์ การจัดการกับ เลเยอร์ของไฟล์ลักษณะต่าง ๆ รวมถึงการจัดการรายละเอียดภาพในเลเยอร์นั้น ๆ

5) เมนูคำสั่ง (Type) รูปแบบการทำงาน สำหรับจัดการและปรับแต่งเกี่ยวกับข้อความ เช่น ปรับแต่งสีข้อความ ปรับแต่งขอบข้อความ หรือการเปลี่ยนข้อความให้เป็นภาพ

6) เมนูคำสั่ง (Select) รูปแบบการทำงาน สำหรับปรับแต่งการเลือกพื้นที่บันทึกและเรียกพื้นที่ที่เลือกมาใช้งาน รวมถึงคำสั่งสำหรับการเลือกพื้นที่เพื่อให้ได้ผลลัพธ์ที่ดียิ่งขึ้น

7) เมนูคำสั่ง (Filter) รูปแบบการทำงาน สำหรับปรับแต่งภาพให้มีรูปแบบที่น่าสนใจยิ่งขึ้น บิด ดัดปรับรูปทรงรูปแบบต่าง ๆให้กับภาพ

8) เมนูคำสั่ง (View) รูปแบบการทำงาน สำหรับเลือกรูปแบบการแสดงผล เช่น ย่อขยายภาพ แสดงไม้บรรทัด เส้นกริด หรือเส้นไกด์

9) เมนูคำสั่ง (Window) รูปแบบการทำงาน สำหรับเลือกเปิดปิดพาเนล รวมถึงกำหนดรูปแบบการแสดงหน้าต่างในแบบต่าง ๆ

10) เมนูคำสั่ง (Help) รูปแบบการทำงาน ใช้สำหรับแสดงความช่วยเหลือเกี่ยวกับการใช้เครื่องมือหรือคำสั่งในรูปแบบต่าง ๆ

2.2.5.3 เมนูของพื้นที่ทำงาน Panel menu Panel (พาเนล) เป็นวินโดว์ย่อย ๆ ที่ใช้เลือกรายละเอียด หรือคำสั่งควบคุมการทำงานต่าง ๆ ของโปรแกรมใน Photoshop มีพาเนลอยู่เป็นจำนวนมาก เช่น พาเนล Color ใช้สำหรับเลือกสี, พาเนล Layers ใช้สำหรับจัดการกับเลเยอร์และพาเนล Info ใช้แสดงค่าสีตรงตำแหน่งที่ชี้เมาส์รวมถึงขนาด/ตำแหน่งของพื้นที่ที่เลือกไว้

2.2.5.4 พื้นที่ทำงาน Stage หรือ Panel เป็นพื้นที่ว่างแสดงงานที่กำลังทำอยู่

2.2.5.5 เครื่องมือที่ใช้งาน Tools panel หรือ Tools box ทูลพาเนล (Tool Panel) หรือ กล่องเครื่องมือจะประกอบไปด้วยเครื่องมือต่าง ๆ ที่ใช้ในการวาด ตกแต่ง และ แก้ไขภาพ เครื่องมือเหล่านี้มีจำนวนมาก ดังนั้นจึงมีการรวมเครื่องมือที่ทำหน้าที่คล้าย ๆ กันไว้ในปุ่มเดียวกัน โดยจะมีลักษณะรูปสามเหลี่ยมอยู่บริเวณมุมด้านล่างเพื่อบอกให้รู้ว่าในปุ่มนี้ยังมีเครื่องมืออื่นอยู่ด้วย

2.2.5.6 สิ่งที่ควบคุมเครื่องมือที่ใช้งาน Tools control menu หรือ Option bar ออปชั่นบาร์ (Option Bar) เป็นส่วนที่ใช้ปรับแต่งค่าการทำงานของเครื่องมือต่าง ๆ โดยรายละเอียดในออปชั่นบาร์จะเปลี่ยนไปตามเครื่องมือที่เราเลือกจากทูลบ็อกซ์ในขณะนั้น เช่น เมื่อเราเลือกเครื่องมือ Brush (พู่กัน) บนออปชั่นบาร์จะปรากฏออปชั่นที่ใช้ในการกำหนดขนาดและลักษณะหัวแปรงโหมดในการระบายความโปร่งใสของสีและอัตราการไหลของสี เป็นต้น

2.2.6 ทฤษฎีเกี่ยวกับการสร้างเว็บไซต์

หลักการออกแบบเว็บไซต์ เว็บไซต์เป็นสื่อที่ได้รับความนิยมอย่างมากบนอินเตอร์เน็ต ซึ่งเว็บไซต์เป็นสื่อที่อยู่ในความควบคุมของผู้ใช้โดยสมบูรณ์ กล่าวคือ ผู้ใช้สามารถตัดสินใจเลือกได้ว่าจะดูเว็บไซต์ใดและจะไม่เลือกดูเว็บไซต์ใดได้ตามต้องการ จึงทำให้ผู้ใช้ไม่มีความอดทนต่ออุปสรรคและปัญหาที่เกิดจากการออกแบบเว็บไซต์ผิดพลาดถ้าผู้ใช้เห็นว่าเว็บที่กำลังดูอยู่นั้นไม่มีประโยชน์ต่อตัวเขา หรือไม่เข้าใจว่าเว็บไซต์นี้จะใช้งานอย่างไร เขาก็สามารถที่จะเปลี่ยนไปดูเว็บไซต์อื่น ๆ ได้อย่างรวดเร็ว เนื่องจากในปัจจุบันมีเว็บไซต์อยู่มากมาย และยังมีเว็บไซต์ที่เกิดขึ้นใหม่ ๆ ทุกวัน ผู้ใช้จึงมีทางเลือกมากขึ้น และสามารถเปรียบเทียบคุณภาพของเว็บไซต์ต่าง ๆ ได้เองเว็บไซต์ที่ได้รับการออกแบบอย่างสวยงาม มีการใช้งานที่สะดวก ย่อมได้รับความสนใจจากผู้ใช้ มากกว่าเว็บไซต์ที่ดูสับสนวุ่นวาย มีข้อมูลมากมายแต่หาอะไรไม่เจอ นอกจากนี้ยังใช้เวลาในการแสดงผลแต่ละหน้านานเกินไป ซึ่งปัญหาเหล่านี้ล้วนเป็นผลมาจากการออกแบบเว็บไซต์ไม่ดีทั้งสิ้น

ดังนั้น การออกแบบเว็บไซด์จึงเป็นกระบวนการสำคัญในการสร้างเว็บไซด์ ให้ประทับใจผู้ใช้ ทำให้เขาอยากกลับเข้ามาเว็บไซด์เดิมอีกในอนาคต ซึ่งนอกจากต้องพัฒนาเว็บไซด์ที่ดีมีประโยชน์แล้ว ยังต้องคำนึงถึงการแข่งขันกับเว็บไซด์อื่น ๆ อีกด้วย

องค์ประกอบของการออกแบบเว็บไซต์ การออกแบบเว็บไซต์ที่มีประสิทธิภาพนั้นต้องคำนึงถึง องค์ประกอบสำคัญดังต่อไปนี้

1) ความเรียบง่าย (**Simplicity**) การออกแบบเว็บไซต์ในรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานอย่างสะดวก

2) ความสม่ำเสมอ (**Consistency**) การสร้างความสม่ำเสมอให้เกิดขึ้นตลอดทั้งเว็บไซต์ โดยอาจเลือกใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ก็

3) ความเป็นเอกลักษณ์ (**Identity**) เว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กร การเลือกใช้ตัวอักษร ชุดสี รูปภาพหรือกราฟฟิกจะมีผลต่อรูปแบบของเว็บไซต์

4) เนื้อหา (**Useful Content**) เนื้อหาในเว็บไซต์ต้องสมบูรณ์และได้รับการปรับปรุงพัฒนาให้ทันสมัยอยู่เสมอ

5) ระบบเนวิเกชั่น (**User–Friendly Navigation**)
ระบบเนวิเกชั่นช่วยไม่ให้ผู้ใช้เกิดความสับสนระหว่างดูเว็บไซต์ จึงเปรียบเสมือนป้ายบอกทางควรทำให้เข้าใจง่ายและใช้งานได้สะดวก

6) คุณภาพของสิ่งที่ปรากฏให้เห็นในเว็บไซต์ (**Visual Appeal**) ลักษณะที่น่าสนใจของเว็บไซต์นั้น ควรมีคุณภาพ เช่น กราฟิกควรสมบูรณ์ไม่มีรอยหรือขอบขั้นบันได้ให้เห็น ชนิดตัวอักษรอ่านง่ายสบายตา มีการเลือกใช้โทนสีที่เข้ากันอย่างสวยงาม เป็นต้น

7) ความสะดวกของการใช้ในสภาพต่าง ๆ (**Compatibility**) สามารถใช้งานได้ดีในสภาพแวดล้อมที่หลากหลาย ไม่มีการบังคับให้ผู้ใช้ต้องติดตั้งโปรแกรมอื่นใดเพิ่มเติมนอกเหนือจากเว็บบราวเซอร์

8) ความคงที่ในการออกแบบ (**Design Stability**) ให้ความสำคัญกับการออกแบบเว็บไซต์ ต้องออกแบบวางแผนและเรียบเรียงเนื้อหาอย่างรอบคอบ

9) ความคงที่ของการทำงาน (**Function Stability**) ระบบการทำงานต่าง ๆ ในเว็บไซต์ควรมีความถูกต้องแน่นอน และตรวจสอบอยู่เสมอ

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 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 ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวมหัวหน้าโครงการหรือที่ปรึกษาด้านการวางระบบวิเคราะห์ข้อมูล จะต้องทำการสัมภาษณ์หรือรับฟังปัญหาความต้องการจากผู้บริหารองค์กรและหน่วยงานต่าง ๆ ที่จะนำผลการวิเคราะห์ข้อมูลไปใช้ประโยชน์ โดยความต้องการทั้งหมดจะนำมาจัดลำดับความสำคัญ และกำหนดวัตถุประสงค์ที่จะนำไปสู่รูปแบบการวิเคราะห์ข้อมูลขององค์กร เช่น ผู้บริหารห้างสรรพสินค้าต้องการรู้ว่าอะไรเป็นเหตุปัจจัยที่ทำให้ลูกค้าเป้าหมายตัดสินใจและเลือกที่จะเข้าห้างไม่ว่าจะเพื่อการจับจ่ายซื้อของใช้ เป็นสถานที่นัดพบ/พักผ่อน หรือหาอาหารรับประทาน ร้านขายสินค้าออนไลน์อยากรู้ว่าผู้คนกำลังให้ความสนใจในสินค้า/บริการประเภทใดอยู่แหล่งข้อมูลออนไลน์ใดที่มีอิทธิพลต่อการตัดสินใจเลือกซื้อสินค้า เป็นต้น

2) สร้างฐานข้อมูลให้ครบ (Data understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูลตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับโดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้ ในอดีตการศึกษาหาแนวโน้มความต้องการตลาด หรือพฤติกรรมผู้บริโภคในการตัดสินใจซื้อสินค้า เป็นเรื่องที่ยุ่งยากและต้องว่าจ้างบริษัทวิจัยสำรวจภาพรวมควบคู่กับการพิจารณารายการสั่งซื้อสินค้าที่เก็บไว้ในฐานข้อมูลของบริษัท แต่ด้วยความก้าวหน้าทางด้านเทคโนโลยีในปัจจุบันและการทำธุรกรรมผ่านเครือข่ายอินเทอร์เน็ตทำให้ข้อมูลมากมายมหาศาลวิ่งผ่านไปมาอยู่ในระบบเว็บไซต์หรือแอพที่เป็นช่องทางในการทำธุรกรรมต่าง ๆ จึงเป็นแหล่งข้อมูลสำคัญ อีกทั้งยังได้ข้อมูลความสนใจของคนที่พร้อมยอมให้อย่างเต็มใจจากห้องแชทต่าง ๆ ที่มีการพูดคุยหารือกัน ปัจจุบันการแกะรอยหรือสะกดรอยตามคนได้ดีที่สุดเกิดขึ้นได้ง่ายมากจากออนไลน์ ไม่ว่าจะเป็นพิกัดตำแหน่งที่อยู่ของเราที่อนุญาตให้แอพต่าง ๆ เข้าถึง

3) เตรียมข้อมูลให้พร้อมใช้ (Data preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก ระบบการรับข้อมูลป้อนเข้าสู่ระบบที่ทันสมัยในปัจจุบันจะลดการคีย์ข้อมูลจากคนให้น้อยที่สุด แต่จะใช้วิธีการ

สแกน การติ๊กเลือก เพื่อลดความผิดพลาดให้น้อยที่สุด เพราะขั้นตอนใช้เวลามากกว่า 50% ของเวลารวมทั้งหมดการลดข้อผิดพลาดของข้อมูลได้มากเท่าใดก็จะยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น

4) จัดทำและเลือกโมเดลที่ใช่ (Modeling) ขั้นตอนการสร้างตัวแบบทาคณิตศาสตร์และสถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ ผสมผสานกัน อาทิ การจำแนก (Classification) การแบ่งกลุ่ม (Clustering) และการสร้างความสัมพันธ์ (Association rule) ในร้านสะดวกซื้อ จะนำข้อมูลการซื้อสินค้าของลูกค้าแต่ละรายมาหาความสัมพันธ์ เช่น คนที่ซื้อเครื่องดื่มแต่ละชนิดมักจะซื้อขนมหรือของกินอะไรร่วมอยู่ด้วย การใช้จ่ายของแต่ละคนจะอยู่ที่ประมาณกี่บาท คนส่วนใหญ่ที่เข้ามาจะซื้อสินค้ากี่ชิ้นต่อคน และเพื่อให้ทราบข้อมูลของผู้ซื้อ ร้านค้ามักจะใช้การออกบัตรเติมเงินที่จูงใจให้ใช้จากส่วนลดหรือสะสมแต้มทำให้สามารถติดตามประวัติการใช้จ่ายได้ง่ายขึ้น ซึ่งปัจจุบันมีการนำกล้องจับภาพผู้ซื้อในการแยกแยะเพศ อายุ และไลฟ์สไตล์ของคน

5) ประเมินผลก่อนตัดสินใจ (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ทั้งนี้อาจต้องกลับไปทบทวนขั้นตอนที่ 2 – 4 ซ้ำอีกครั้ง ในกรณีที่ผลลัพธ์ไม่มีความน่าเชื่อถือเพียงพอ หรือไม่สอดรับกับวัตถุประสงค์

6) เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อมให้ฝ่ายต่าง ๆ นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจต่อไป

2.3.2 ผังงาน (Flowchart)

2.3.2.1 ผังงาน (Flowchart) คือ รูปภาพ (Image) หรือสัญลักษณ์ (Symbol) ที่ใช้เขียนแทนขั้นตอน คำอธิบาย ข้อความ หรือคำพูด ที่ใช้ในอัลกอริทึม (Algorithm) เพราะการนำเสนอขั้นตอนของงานให้เข้าใจตรงกัน ระหว่างผู้เกี่ยวข้อง ด้วยคำพูด หรือข้อความทำได้ยากกว่า

หน้าที่ของผังงาน (Flowchart) คือ แผนภาพแสดงลำดับขั้นตอนการทำงานของ Algorithm, Workflow, Process เป็นเครื่องมือใช้การรวบรวมจัดลำดับความคิด เพื่อให้เห็นขั้นตอนการทำงานที่ชัดเจนและใช้วางแผนการทำงานขั้นแรก โดยสัญลักษณ์ Flowchart แสดงถึงการทำงานลักษณะต่าง ๆ เชื่อมต่อกัน Flowchart ถูกใช้ในการออกแบบ เพื่อช่วยให้เห็นภาพสิ่งที่เกิดขึ้นและช่วยให้เข้าใจกระบวนการทำงานและบางทีอาจช่วยหาข้อบกพร่องภายในงาน

ผังงานแบ่งได้ 2 ประเภท

1) ผังงานระบบ (System Flowchart) คือ ผังงานที่แสดงขั้นตอนการทำงานในระบบอย่างกว้าง ๆ แต่ไม่เจาะลงในระบบงานย่อย

2) ผังงานโปรแกรม (Program Flowchart) คือ ผังงานที่แสดงถึงขั้นตอนในการทำงานของโปรแกรม ตั้งแต่รับข้อมูล คำนวณ จนถึงแสดงผลลัพธ์

ประโยชน์ของผังงาน

1) ทำให้เข้าใจ และแยกแยะปัญหาได้ง่าย (Problem Define)

2) แสดงลำดับการทำงาน (Step Flowing)

3) หาข้อผิดพลาดได้ง่าย (Easy to Debug)

4) ทำความเข้าใจโปรแกรมได้ง่าย (Easy to Read)

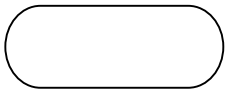
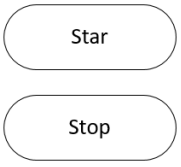
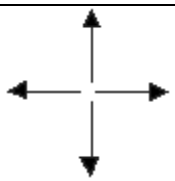
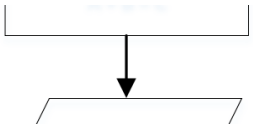
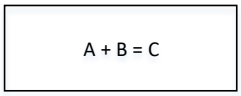
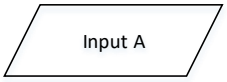
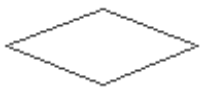
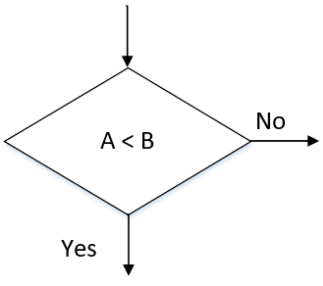
5) ไม่ขึ้นกับภาษาใดภาษาหนึ่ง (Flexible Language)

สัญลักษณ์ Flowchart คือ รูปภาพที่ใช้แทนความหมายการทำงานในลักษณะต่าง ๆ ภายในผังงาน (Flowchart) ประกอบไปด้วย การเริ่มต้น (Start), การจบ (End), การกระทำ (Process), การนำเข้าข้อมูล (Input), การแสดงผลข้อมูล (Output), การตัดสินใจ (Decision), คำอธิบาย (Annotation), จุดเชื่อมต่อ (Connector), ทิศทางการทำงาน (Direction Flow)
สัญลักษณ์เหล่านี้เมื่อถูกนำมาเชื่อมต่อกัน จะกลายเป็น "ผังงาน (Flowchart) ที่แสดงลำดับขั้นตอนการทำงานเพื่อเป็นเครื่องมือในการจัดลำดับความคิดเห็นลำดับขั้นตอนการทำงานที่ชัดเจน

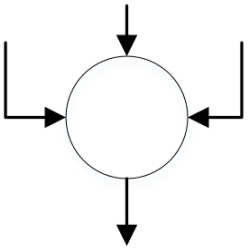
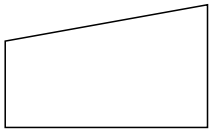
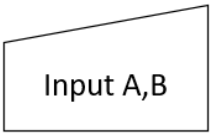
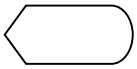
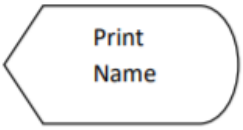
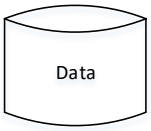
2.3.2.2 การเขียนผังงาน (Flowchart)

ผังงาน คือ แผนภาพที่มีการใช้สัญลักษณ์รูปภาพและลูกศรที่แสดงถึงขั้นตอนการทำงานของโปรแกรมหรือระบบทีละขั้นตอน รวมไปถึงทิศทางการไหลของข้อมูลตั้งแต่แรกจนได้ผลลัพธ์ตามที่ต้องการ

ตารางที่ 2.1 สัญลักษณ์ แผนผังงาน

	จุดเริ่มต้น / สิ้นสุดของโปรแกรม	
	ลูกศรแสดงทิศทางการทำงานของโปรแกรมและการไหลของข้อมูล	
	ใช้แสดงคำสั่งในการประมวลผลหรือการกำหนดค่าข้อมูลให้กับตัวแปร	
	แสดงการอ่านข้อมูลจากหน่วยเก็บข้อมูลสำรองเข้าสู่หน่วยความจำหลักภายใน เครื่องหรือการแสดงผลลัพธ์จากการประมวลผลออกมา	
	การตรวจสอบเงื่อนไขเพื่อตัดสินใจโดยจะมีเส้นออกจารรูปเพื่อแสดงทิศทางการทำงานต่อไป เงื่อนไขเป็นจริงหรือเป็นเท็จ	

ตารางที่ 2.1 (ต่อ) สัญลักษณ์ แผนผังงาน

	แสดงจุดเชื่อมต่อของผังงานภายใน หรือเป็นที่บรรจบของเส้นหลายเส้นที่มาจากหลายทิศทางเพื่อจะไปสู่ การทำงานอย่างใดอย่างหนึ่งที่เหมือนกัน	
	การป้อนข้อมูลผ่านทางแป้นพิมพ์	
	การแสดงผลออกทางจอภาพ	
	ที่เก็บข้อมูล	

2.3.3 เทคนิควิธีทางเหมืองข้อมูล

ในการที่จะเข้าใจว่าการจำแนกข้อมูลคืออะไร ลองพิจารณาตัวอย่างดังต่อไปนี้

1) พนักงานสินเชื่อ ของธนาคารต้องการที่จะทำการวิเคราะห์ข้อมูลเพื่อที่จะทำการศึกษาว่าการกู้ยืมในครั้งหนึ่ง ๆ มีครั้งไหนบ้างที่ ปลอดภัยและครั้งไหนบ้างที่มีความเสี่ยง

2) ผู้จัดการฝ่ายการตลาดของบริษัทขายอุปกรณ์ไฟฟ้าต้องการข้อมูล เพื่อช่วยในการคาดเดาว่า “ลูกค้ามีคุณลักษณะอย่างไรที่จะทำการซื้อคอมพิวเตอร์จากบริษัท”

3) นักวิจัยทางการแพทย์ต้องการที่จะวิเคราะห์ข้อมูลเกี่ยวกับมะเร็งเต้านม เพื่อที่จะทำการทำนายว่าผู้ป่วยควรจะได้รับการ ดูแลด้วยวิธีใดภายใต้วิธีการรักษาทั้ง 3 วิธีที่ได้รับความนิยมอย่างแพร่หลาย

จากตัวอย่างทั้ง 3 ข้างต้นจะต้องการการวิเคราะห์ข้อมูล ซึ่งก็คือการจำแนกข้อมูล (Classification) ที่ซึ่งจะทำการสร้างโมเดลหรือตัว จำแนกข้อมูล (Classifier) เพื่อทำนายหมวดหมู่ของข้อมูล (Categories/Class) อาทิเช่น ปลอดภัย หรือ เสี่ยง ในการวิเคราะห์ข้อมูลสินเชื่อ วิธีการรักษา A หรือ B หรือ C ในการวิเคราะห์ข้อมูลของผู้ป่วยมะเร็งเต้านม เป็นต้น ในส่วนของการทำนายข้อมูล

จากตัวอย่างดังต่อไปนี้ สมมติว่าผู้จัดการฝ่ายการตลาด ต้องการที่จะทำนายหรือคาดเดาว่าลูกค้าคนหนึ่งๆจะทำการจ่ายเงินซื้อสินค้าจากบริษัทเป็นจำนวนเท่าไหร่ การวิเคราะห์ข้อมูลลักษณะนี้จะเป็นส่วนของการทำนายข้อมูลเชิงตัวเลข (Numeric prediction)

สิ่งที่ต้องคำนึงถึงในการจำแนกและการทำนายข้อมูลในการจำแนกและทำนายข้อมูลครั้งหนึ่ง ๆ จะมีหลายปัจจัยที่เราต้องพิจารณาและคำนึงถึง แต่อย่างไรก็ ตาม โดยทั่วไปเราจะต้องพิจารณา 2 ปัจจัยหลัก ดังนี้

2.3.3.1 การจัดเตรียมข้อมูลสำหรับการจำแนกและการทำนายข้อมูล

1) การทำความสะอาดข้อมูล (Data cleansing) จะเกี่ยวข้องกับการประมวลผลข้อมูลเบื้องต้นที่จะ ลบหรือลดข้อมูลที่มีสิ่งรบกวน (noise) ด้วยการประยุกต์ใช้วิธีการปรับเรียบข้อมูลแบบต่าง ๆ และ จัดการกับการขาดหายไปของข้อมูลด้วยการแทนค่าของข้อมูลที่ขาดหายไปด้วยค่าของข้อมูลที่ปรากฏบ่อยที่สุดหรือทำการแทนด้วยค่าของข้อมูลที่มีค่าเชิงสถิติสูงที่สุด เป็นต้น

2) ความเกี่ยวเนื่องของข้อมูล (Relevance analysis) จะทำการตรวจสอบข้อมูลแอทริบิวต่าง ๆ ว่ามี ความเกี่ยวเนื่องหรือซ้ำซ้อนกันมากน้อยเพียงใด ซึ่งโดยปกติของชุดข้อมูลจะมีแอทริบิวที่ซ้ำซ้อนกัน

3) การเปลี่ยนแปลง/เปลี่ยนรูปข้อมูลและการลดจำนวนข้อมูล (Data transformation and reduction) ข้อมูลที่เป็นอินพุตอาจมีช่วงของข้อมูลหรือค่าของข้อมูลที่มีระยะห่างค่อนข้างมาก ดังนั้น เราอาจทำการเปลี่ยนแปลง/เปลี่ยนรูปด้วยวิธีการ normalization ที่จะทำการปรับเปลี่ยนค่า ในแอทริบิวหนึ่ง ๆ ให้อยู่ในช่วงที่กำหนด อาทิเช่น ช่วง –1.0 ถึง 1.0 หรือ ช่วง 0.0 ถึง 1.0 เป็นต้น ซึ่งสามารถลด จำนวนข้อมูลที่ต้องทำการพิจารณาได้ด้วยการประยุกต์ใช้เทคนิคต่าง ๆ เช่น wavelet transformation และ principle component analysis รวมถึงเทคนิคการทำ binning, histogram analysis และ clustering เป็นต้น

2.3.3.2 การเปรียบเทียบประสิทธิภาพวิธีในการจำแนกและทำนายข้อมูล

ในการเปรียบเทียบประสิทธิภาพของวิธีในการจำแนกและทำนายข้อมูลเราจะทำการประยุกต์ใช้เกณฑ์ ดังต่อไปนี้

1) ความถูกต้อง (Accuracy) จะเกี่ยวข้องกับความสามารถของตัวจำแนกข้อมูลที่ถูกสร้างขึ้นที่จะ สามารถจำแนกข้อมูลที่ไม่เคยพบเจอมาก่อนได้อย่างถูกต้อง โดยในการวัดความถูกต้องอาจประเมินได้ จากการใช้ชุดข้อมูลหนึ่ง ๆ (หรือมากกว่าหนึ่งชุดก็ได้)ที่ แยกจากชุดข้อมูลเรียนรู้ (training dataset)

2) ความเร็ว (Speed) จะเกี่ยวข้องกับเวลาที่ใช้ในการคำนวณทั้งในส่วนของการสร้างตัวจำแนกข้อมูล และการจำแนกหรือทำนายข้อมูล

3) ความทนทาน (Robustness) จะเกี่ยวข้องกับความสามารถของตัวจำแนกหรือตัวทำนายข้อมูลที่ จะทำการทำนายได้อย่างถูกต้องจากข้อมูลตั้งต้นที่มีสิ่งรบกวนหรือมีการขาดหายไปของข้อมูล

4) ความยืดหยุ่นต่อปริมาณข้อมูล (Scalability) จะเกี่ยวข้องกับความสามารถในการสร้างตัวจำแนก ข้อมูลหรือตัวทำนายข้อมูลได้อย่างมีประสิทธิภาพเมื่อมีข้อมูลที่ต้องพิจารณาเป็นปริมาณมาก

5) ความสามารถในการเข้าใจ (Interpretability) เกี่ยวเนื่องกับระดับความสามารถที่จะถูกเข้าใจใน ตัวจำแนกหรือทำนายข้อมูลจากผู้ใช้งาน

2.3.4 แบบจำลองอนุกรมเวลา (Time Series)

การพยากรณ์ ในการพยากรณ์อนุกรมเวลาแบบฉบับ ส่วนประกอบที่สำคัญที่มีอิทธิพลต่อการพยากรณ์ คือส่วนประกอบของการเปลี่ยนแปลงตามฤดูกาล แนวโน้ม และการเปลี่ยนแปลงตามวัฏจักร ส่วน การเปลี่ยนแปลงที่ผิดปกตินั้นเราไม่สามารถพยากรณ์ได้ เนื่องจากเป็นตัวแปรสุ่มมีรูปแบบที่ไม่แน่นอน ทำให้ไม่ อาจคาดการณ์ได้ล่วงหน้าว่าจะเกิดอะไรขึ้น ณ เวลาใด และรุนแรงเพียงใด ดังนั้น ค่าพยากรณ์ Y ในหน่วย เวลา t จะหาค่าได้ดังนี้

$$Y = T \times S \times C \times I$$

โดยที่

Y คือ ข้อมูลอนุกรมเวลา

T คือ คือค่าประมาณของส่วนประกอบแนวโน้มของหน่วยเวลา (Trend)

S คือค่าประมาณของส่วนประกอบการผันแปรตามฤดูกาล (Seasonal Variation)

C คือ การผันแปรตามวัฏจักร (Cyclical Variation)

I คือ การผันแปรผิดปกติ (Irregular Variation)

ในการพยากรณ์ค่าส่วนประกอบแนวโน้มและส่วนประกอบการเปลี่ยนแปลงตามฤดูกาล กระทำได้ไม่ยากนัก แต่การพยากรณ์ค่าของส่วนประกอบการเปลี่ยนแปลงตามวัฏจักรนั้นไม่ใช่สิ่งที่ง่ายนัก เพราะการเปลี่ยนแปลงตามวัฏจักรไม่มีรูปแบบ และความยาวของการเปลี่ยนแปลงตามวัฏจักรไม่แน่นอน ใน การพยากรณ์การเปลี่ยนแปลงตามวัฏจักรในระยะสั้นหรืออาจใช้ความคิดเห็นของผู้เชี่ยวชาญในธุรกิจนั้น ๆ เช่น การประมาณอัตราการขยายตัวของธุรกิจโดยใช้ค่าเฉลี่ยของความสูง และความยาวของการเปลี่ยนแปลง ตามวัฏจักรในอดีต เป็นต้น แต่ในทางปฏิบัติก็ไม่ใช่สิ่งที่จะทำได้โดยง่าย และการเปลี่ยนแปลงตามวัฏจักรใน ระยะสั้นก็มีความสำคัญต่อความแม่นยำของค่าพยากรณ์มากพอสมควร ดังนั้นการประมาณการเปลี่ยนแปลง ตามวัฏจักรโดยวิธีการนำเสนอไปในหัวข้อ 4 เป็นแนวทางหนึ่ง ที่อาจใช้ประมาณค่าการเปลี่ยนแปลงตาม วัฏจักรได้อย่างเป็นระบบ นอกเหนือไปจากวิธีหาค่าเฉลี่ยแบบเคลื่อนที่ ค่าประมาณของส่วนประกอบการ เปลี่ยนแปลงตามวัฏจักรที่จะใช้พยากรณ์ในฤดูกาลหน้าจะเป็นค่าในฤดูกาลปัจจุบัน

หากข้อมูลอนุกรมเวลาที่ใช้ในการศึกษาครั้งนี้ ไม่พบการผันแปรตามวัฏจักรและการผันแปรผิดปกติ ของข้อมูลที่ชัดเจน ดังนั้น การพยากรณ์จึงมีเฉพาะองค์ประกอบที่เป็นค่าแนวโน้มและองค์ประกอบที่แสดงค่า การผันแปรตามฤดูกาล ดังนี้

$$Y = T \times S$$

การวัดความแม่นยำของตัวแบบที่นำมาใช้พยากรณ์ ค่าเฉลี่ยเปอร์เซ็นต์ความ คลาดเคลื่อนสมบูรณ์ (Mean Absolute Percentage Error : MAPE) สามารถ คำนวณได้จากสูตร

$$MAPE = \frac{[\sum | A_t - F_t | / A_t] \times 100}{N}$$

ค่าเฉลี่ยเปอร์เซ็นต์ความ คลาดเคลื่อนสมบูรณ์ (MAPE) ยิ่งน้อย หมายถึง การพยากรณ์ยิ่งมีความแม่นยำ

2.4 วรรณกรรมที่เกี่ยวข้อง

อรรณพ กางกั้น (2562) งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพและเลือกเทคนิคที่เหมาะสมในการพยากรณ์และพัฒนาแบบจำลองในการพยากรณ์ปริมาณการใช้น้ำมันดีเซลด้วยเทคนิควิธีเหมืองข้อมูล 3 เทคนิคได้แก่ 1) การถดถอยเชิงเส้น 2) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้นและ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยข้อมูลที่นำมาใช้ศึกษาเป็นข้อมูลอนุกรมเวลารายเดือนปริมาณการใช้น้ำมันดีเซลตั้งแต่วันที่1 มกราคม 2549 ถึงวันที่ 31 ธันวาคม 2561 จากผลการทดลองพบว่าชุดข้อมูลปริมาณการใช้น้ำมันดีเซลย้อนหลัง 12 เดือนมีความเหมาะสมในการใช้เป็นชุดข้อมูลเพื่อสร้างแบบจำลองในการพยากรณ์และเมื่อพิจารณาประสิทธิภาพของแบบจำลองกับชุดข้อมูลในปี2561 ที่ใช้เป็นข้อมูลทดสอบพบว่าแบบจำลองการพยากรณ์ด้วยโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้นมีประสิทธิภาพโดยรวมสูงที่สุดมีค่าเฉลี่ยความ คลาดเคลื่อนสัมพัทธ์เท่ากับ 2.38% รองลงมาคือแบบจำลองการพยากรณ์ด้วยเทคนิคการวิเคราะห์การถดถอยมีค่าเฉลี่ยความ คลาดเคลื่อนสัมพันธ์เท่ากับ 2.66% จากผลการวิจัยแสดงให้เห็นประเด็นสำคัญเกี่ยวกับการใช้ข้อมูลย้อนหลังในการพยากรณ์ปริมาณการใช้น้ำมันดีเซลที่มีความเหมาะสมอยู่ในช่วง 12 เดือนซึ่งอาจจะสอดคล้องกับสถานการณ์การใช้น้ำมันดีเซลที่มีความเปลี่ยนแปลงภายในรอบปีและจากการเปรียบเทียบเทคนิคการเลือกใช้แบบจำลองในการพยากรณ์แสดงให้เห็นว่าการใช้แบบจำลองจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมีความผิดพลาดสูงที่สุดโดยค่า MMRE เท่ากับ2.90% แต่ก็ยังอยู่ในเกณฑ์ต่ำและไม่แตกต่างจากค่า

MMRE ที่ได้จากการสร้างแบบจำลองด้วยเทคนิคโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้นและเทคนิคการวิเคราะห์การถดถอยมากนักดังนั้นจึงขึ้นกับผู้ใช้งานที่จะกำหนดเกณฑ์ความผิดพลาดที่ยอมรับได้

ณรงค์ฤทธิ์ สุคนธสิงห์ (2559) ผลการวิจัย แบบจำลองการพยากรณ์ปริมาณการชำระเงินผ่านธนาคารบนอินเทอร์เน็ตกับ ชุดข้อ มูลในปี2558 สามารถสรุปได้ว่า โดยนำเสนอข้อมูลเป็นรายเดือนพบว่าแบบจำลองการพยากรณ์ด้วยเทคนิควิธีโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น มีค่า MMRE = 8.70% มีค่าประสิทธิภาพ โดยรวมสูงที่สุด ถัดมาได้แก่ แบบจำลองการพยากรณ์ด้วยเทคนิคการถดถอยเชิงเส้น มีค่า MMRE = 16.75% เมื่อพิจารณาจากผลการวิจัยโดยละเอียดจะพบว่า ประเด็น สำคัญ ในการพยากรณ์ปริมาณการชำระเงินผ่านธนาคารบนอินเทอร์เน็ตกับชุดข้อมูลในปี 2558 พบว่าเดือนมกราคม แบบจำลองการพยากรณ์ด้วย วิธีเทคนิคการถดถอยเชิงเส้นและวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยเชิงเส้น มีค่า MRE = 0.07 มีค่าประสิทธิภาพสูงที่สุด เดือนกุมภาพันธ์แบบจำลองการพยากรณ์ด้วยเทคนิควิธีโครงข่ายประสาท เทียมแบบเปอร์เซ็ปตรอนหลายชั้น มีค่า MRE =0.00 มีค่าประสิทธิภาพสูงที่สุด เดือนมีนาคม แบบจำลองการพยากรณ์ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยเชิงเส้น มีค่า MRE = 0.00 มีค่า ประสิทธิภาพสูงที่สุด เดือนเมษายนถึงธันวาคม แบบจำลองการพยากรณ์ด้วยเทคนิควิธีโครงข่าย ประสาทเทียมแบบเปอร์เซ็ปตรอน หลายชั้น มีค่า MRE ระหว่าง 0.01–0.20 มีค่าประสิทธิภาพสูงที่สุด จากผลการวิจัยดังกล่าวจะเห็นว่าแต่ละเดือนค่า MRE ที่มีประสิทธิภาพสูงที่สุดจะเปลี่ยนแปลงตามแบบจำลองการพยากรณ์และข้อมูลย้อนหลัง (Lagged) ดังนั้นสถาบันการเงินสามารถนำผลลพันธ์ของแบบจำลองพยากรณ์ปริมาณการชำระเงินผ่านธนาคารบนอินเทอร์เน็ตไปใช้ประโยชน์โดยเลือกแบบจำลองการพยากรณ์ด้วยวิธีเทคนิคที่มีค่าประสิทธิภาพสูงที่สุดแต่ละเดือนตามผลการวิจัยจะส่งผลให้ การพยากรณ์ปริมาณการชำระเงินผ่านธนาคารบนอินเทอร์เน็ตมีความแม่นยำสูงขึ้น เมื่อพิจารณาผลลัพธ์จากการวิจัยและการนำไปประยุกต์ใช้ในงานด้านแผนการประชาสัมพันธ์เชิงรุก เพื่อกระตุ้น ลูกค้าของสถานบันการเงินให้มาชำระเงินผ่านธนาคารบนอินเทอร์เน็ตมากขึ้น ในเวลาที่ปริมาณการชำระเงินผ่านธนาคารบนอินเทอร์เน็ตลดลง นอกจากนี้ยังพบว่าปริมาณการชำระเงินผ่าน ธนาคารบนอินเทอร์เน็ตมีแนวโน้ม เพิ่มขึ้นเกือบทุกเดือน สถาบันการเงินสามารถนำผลลัพธ์จากการวิจัย ไปวางแผนเทคโนโลยีสารสนเทศในด้านซอฟต์แวร์ ฮาร์ดแวร์ ระบบรักษาความปลอดภัยในการชำระเงิน เพื่อรองรับลูกค้าที่เพิ่มขึ้นในอนาคต

รัตนา สุวรรณวิชนีย์ และ ปราลี มณีรัตน์ (2560) ได้วิจัยศึกษาความต้องการใช้ข้อมูลของ วช. มีความต้องการในลักษณะข้อมูลรูปแบบโต้ตอบปัจจุบันตลอดเวลา สามารถแสดงภาพรวมงานวิจัย จำแนกตามมิติต่าง ๆ ดังนี้ ปี แหล่งทุน กระทรวง หน่วยงาน ด้านการวิจัย นโยบายและยุทธศาสตร์ การวิจัย กลุ่มงานวิจัย และประเด็นงานวิจัยต่าง ๆ รวมถึงสถานภาพนักวิจัย ยังขาดเครื่องมือในการวิเคราะห์ข้อมูลจำนวนมากในระบบ NRMS ให้ได้ทันเวลาและนำเสนอข้อมูลในหลากหลายมิติที่ตรงตามความต้องการใช้ประโยชน์ของผู้ใช้ เพื่อนำเสนอรัฐบาลหรือหน่วยงานนโยบายสำหรับใช้ในการตัดสินใจเชิงนโยบาย ทั้งนี้จึงนำเสนอการพัฒนาระบบรายงานรูปแบบหลายมิติเพื่อสนับสนุนการตัดสินใจเชิงนโยบายของสำนักงานคณะกรรมการวิจัยแห่งชาติ โดยใช้ระบบข่าวกรองทางธุรกิจ (Business Intelligence : BI) งานวิจัยนี้ได้พัฒนาระบบรายงานรูปแบบหลายมิติและแดชบอร์ดในมุมมองต่าง ๆ เพื่อสนับสนุนการตัดสินใจเชิงนโยบาย ทำให้เจ้าหน้าที่ วช. สามารถจัดทำรายงานที่พร้อมนำเสนอผู้บริหารได้อย่างรวดเร็ว และหลากหลายมิติมากขึ้น ทำให้ค้นพบข้อมูลที่น่าสนใจที่เจ้าหน้าที่ วช. ยังไม่เคยนำเสนอมาก่อน เช่น ข้อมูลอายุเฉลี่ยของนักวิจัยจำแนกตามมิติของระดับการศึกษาตำแหน่งทางวิชาการและประเด็นงานวิจัย เป็นต้น ซึ่งการนำเสนอในลักษณะนี้สามารถใช้โปรแกรม Tableau วิเคราะห์และจัดทำรายงานได้อย่างรวดเร็วโดยจากการประเมินความพึงพอใจ พบว่าผู้ใช้งานมีความพึงพอใจต่อระบบรายงาน และสามารถนำไปใช้ประโยชน์ในส่วนของการนำเสนอผู้บริหาร และการวางแผนในเชิงนโยบายได้อย่างรวดเร็วยิ่งขึ้น เครื่องมือที่ใช้พัฒนาระบบรายงานของงานวิจัยนี้ใช้โปรแกรม Tableau ที่สามารถช่วยอำนวยความสะดวกในการเลือกเฉพาะข้อมูลที่สนใจ และนำออกข้อมูลเป็นตารางได้เหมาะสมกับผู้ใช้งานที่ไม่ต้องใช้คำสั่งในการนำออกข้อมูลผ่านระบบ NRMS สามารถดูรายงานและนำออกข้อมูลได้อย่างง่ายดายผ่านเครื่องมือของโปรแกรม Tableau อีกทั้งข้อมูลที่นำออกเป็นข้อมูลแบบทันเหตุการณ์ (real-time) อีกด้วย ทำให้ลดภาระงานของผู้ดูแลระบบในการสร้างคำสั่งเพื่อนำออกข้อมูลผลของงานวิจัยนี้เผยให้เห็นว่า โปรแกรม Tableau Desktop เป็นโปรแกรมที่ใช้งานง่าย ไม่ซับซ้อน นอกจากนี้ยังสามารถเชื่อมต่อกับข้อมูลได้จากหลากหลายแหล่งและหลากหลายรูปแบบ ดังนั้น สามารถนำข้อมูลจากแหล่งอื่นมาร่วมในการวิเคราะห์และจัดทำสารสนเทศได้เช่น ข้อมูลตัวชี้วัดของหน่วยงาน และข้อมูลผลการติดตามการดำเนินงานวิจัยโดยเจ้าหน้าที่ วช. เป็นต้น ซึ่งหน่วยงานอื่น ๆ สามารถนำโปรแกรม Tableau ไปปรับใช้งานภายในองค์กรได้

ศรีสมรัก อินทุจันทร์ยง (2556) กล่าวว่าระบบข่าวกรองทางธุรกิจ คือ เครื่องมือทางด้านเทคโนโลยีสารสนเทศที่ผู้ใช้สามารถนำไปประมวลผล วิเคราะห์ ข้อมูลจำนวนมากที่มา

จากแหล่งข้อมูลหลายแหล่งที่มีทั้งรูปแบบ โครงสร้างข้อมูลที่มีความแตกต่างกัน เพื่อให้เป็นสารสนเทศในรูปแบบที่ผู้ใช้ต้องการได้อย่างมีประสิทธิภาพ โดยเป็นเครื่องมือสนับสนุนการบริหารการตัดสินใจได้ทุกหน้าที่งาน การทำงานของระบบข่าวกรองทางธุรกิจประกอบด้วยองค์ประกอบ 3 ส่วนหลัก ดังนี้ คือ

1) ชุดเครื่องมือในการคัดแยก (Extract) เปลี่ยนแปลง (Transform) และบรรจุ (Load) ในที่จัดเก็บเครื่องมือชุดนี้เป็นที่รู้จักกันในชื่อที่เรียกว่าอีทีแอล (ETL) เครื่องมือชุดนี้จะช่วยทำหน้าที่คัดแยกข้อมูลเฉพาะที่ผู้ใช้ต้องการจากทุกแหล่งข้อมูลมารวมกัน โดยจะทำการทำความสะอาด

2) คลังข้อมูล (Data Warehouse) เป็นที่จัดเก็บข้อมูลนำมาจากแหล่งข้อมูลภายในองค์กร นั่นคือระบบสารสนเทศในระดับปฏิบัติการ แหล่งข้อมูลภายนอกที่ผู้บริหารเห็นว่ามีความจำเป็นต้องใช้ในการทำงานการตัดสินใจของผู้บริหารและข้อมูลส่วนบุคคล (Personnel Data) ข้อมูลเหล่านั้นจะถูกนำมาจัดเตรียมให้อยู่ในรูปแบบที่พร้อมจะทำงานเชิงวิเคราะห์ (Analytical Data) ตามที่ผู้บริหารต้องการได้ คลังข้อมูลจะเป็นฐานข้อมูลสำหรับการวิเคราะห์ด้วยชุดคำสั่งงานต่าง ๆ เช่นการประมวลผลเชิงวิเคราะห์แบบออนไลน์หรือโอแลบ (On-LineAnalytical Processing : OLAP) การทำเหมืองข้อมูล (DataMining) และระบบสารสนเทศอื่น ๆ เป็นต้น

3) ชุดคำสั่งงานเพื่อการวิเคราะห์ จะประกอบด้วยชุดคำสั่งงานหลายชุดคำสั่ง ที่จะทำการวิเคราะห์ในประเด็นที่แตกต่างหลากหลายกันไปผู้ใช้จะเลือกชุดคำสั่งงานตามที่ต้องการมาใช้

Kim Oanh, N. T. Leelasakultum, K. วิศวกรรมและการจัดการสิ่งแวดล้อม, SERD, สถาบันเทคโนโลยีแห่งเอเชีย, ปทุมธานี ได้ผลของงานวิจัยครั้งนี้ตรวจสอบสาเหตุหลักของการเกิดหมอกควันในภาคตะวันตกเฉียงเหนือของประเทศไทยเพื่อแจ้งเตือนและคาดการณ์ล่วงหน้าในกรณีที่ไม่มีข้อมูลล่วงหน้าของอนาตที่จำเป็นสำหรับการเพื่อทำนายและจำลองการเดินทางหมอกควันจึงใช้วิธีคำนวนภูมิอากาศร่วมกับการวิเคราะห์ทางสถิติ โดยโครงการจัดประเภทอุตุนิยมวิทยาอัตโนมัติที่ได้รับการพัฒนาโดยใช้ข้อมูลสถานีอุตุนิยมวิทยาระดับภูมิภาค 8 ปี (2544-2551) ซึ่งจำแนกรูปแบบการสรุปโดยทั่วไปของภาคเหนือของประเทศไทยเป็น 4 รูปแบบ แบบที่ 2 เกิดขึ้นที่ความถี่สูงในเดือนมีนาคมพบว่าเชื่อมโยงกับระดับสูงสุดของฝุ่น PM 10 ในเชียงใหม่ซึ่งเป็นเมืองที่ใหญ่ที่สุดในภาคเหนือของประเทศไทย ลักษณะทั่วไปของรูปแบบนี้คือการปกคุมของระดับความร้อนต่ำกว่าอินเดีย, จีนตะวันตกและภาคเหนือของประเทศไทยที่มีอากาศร้อนแห้งและนิ่งในภาคเหนือของประเทศไทย เดือนมีนาคม 2550 เดือนที่มีเหตุการณ์หมอกควันรุนแรงที่สุดในเชียงใหม่พบว่ามีความถี่สูงในการเกิดรูปแบบที่ 2 ควบคู่

ไปกับการปล่อยมลพิษความเข้มสูงสุดจากการเผาแบบเปิดของชีวมวล จากผลย้อนหลังแสดงให้เห็นว่าในวันที่มีหมอกควันหนาแน่นอากาศเดินทางผ่านบริเวณจุดที่เกิดเพลิงไหม้ก่อนจะเลื่อนตัวมาถึงเชียงใหม่ แบบจำลองการถดถอยแบบขั้นตอนนี้ได้รับการพัฒนาเพื่อทำนายผลในอีก 24 ชั่วโมง ของฝุ่นPM 10 โดยใช้ข้อมูลเดือนกุมภาพันธ์ถึงเมษายนของปี 2550–2552 และทดสอบกับข้อมูลปี 2547–2553 แบบจำลองนี้ดำเนินการออกมาได้ผลอย่างน่าพึงพอใจสำหรับชุดข้อมูลการพัฒนาแบบจำลองและชุดข้อมูลทดสอบซึ่ง แบบจำลองที่พัฒนาขึ้นนี้มีความแม่นยำมากถึง 90% สำหรับการพยากรณ์หมวดหมู่ของฝุ่น PM 10ขั้นตอนการเตือนตอนจะระบุรูปแบบสรุปและคาดการณ์ฝุ่น PM 10 ในเชียงใหม่ล่วงหน้า 24 ชั่วโมง แนวทางนี้จะใช้สำหรับการจัดการเหตุการณ์มลพิษทางอากาศในพื้นที่อื่นที่มีภูมิประเทศซับซ้อนที่มีสภาพคล้ายกันซึ่งผลงานวิจัยนี้เผยให้เห็นถึงการใช้ประโยชน์จากการใช้เครื่องในการวิเคราะห์ข้อมูลย้อนหลังทำให้เกิดพยากรณ์ที่แม่นสามารถบอกสถานะการล่วงหน้าได้

ปภาดา โพธิ์คำอภิชัย และ มณเฑียร รัตนศิริวงศ์วุฒิ (2557) ได้พัฒนาระบบธุรกิจอัจฉริยะเพื่อการวางแผนการผลิตในองค์กร เพื่อสนับสนุนการตัดสินใจของผู้บริหาร ในการวิเคราะห์แนวโน้มทางการตลาดและมีแนวทางในการดำเนินงานที่ชัดเจนรวมถึงการปรับกลยุทธ์ทางธุรกิจ ในการแข่งขันได้ดีโดยใช้หลักการธุรกิจอัจฉริยะ (Business Intelligence) ประกอบด้วยการสร้างคลังข้อมูลขนาดเล็ก (Data Mart) และนำสารสนเทศที่มีอยู่มาสร้างรายงานอัจฉริยะ (Dashboard) แดชบอร์ด คือ การนำเสนอข้อมูลโดยการเลือกรายงาน หรือกราฟที่น่าสนใจมานำเสนอ รวมกันไว้ในหน้าจอเดียวกัน ซึ่งจะช่วยให้ผู้บริหารหรือผู้ใช้งานสามารถมองเห็นภาพรวมของข้อมูลที่สำคัญในเรื่องใดเรื่องหนึ่งได้อย่างชัดเจนและรวดเร็ว เช่น การแสดงข้อมูล Balanced Scorecard ในมุมมองด้านรายได้เปรียบเทียบกับต้นทุนด้านการตลาด หรือ KPI ในมุมมองต่าง ๆ ตามที่ต้องการสามารถปรับเปลี่ยนมุมมองในการวิเคราะห์ และตรงตามความต้องการของผู้บริหารและผู้ใช้งาน จากผลการประเมินคุณภาพของระบบโดยผู้เชี่ยวชาญสรุปให้เห็นว่า การพัฒนาระบบธุรกิจอัจฉริยะเมื่อนำมาประยุกต์ใช้ในการวางแผนการผลิตในองค์กรสามารถใช้งานได้จริงและในระดับดี

อาทิตย์ สิทธิบรรเจิด (2552) กล่าวว่าการสร้างมโนภาพเป็นส่วนประกอบสำคัญในระบบการรับรู้ ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยการใช้ภาพจะทำให้ผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิด และนำไปใช้ในการสนับสนุนการตัดสินใจ

Nawahda (2557) ทำการวิจัยผลกระทบจากการลดความเข้มข้นฝุ่น PM2.5 ภายในประเทศ ประเทศญี่ปุ่น ที่มีผลต่อการเสียชีวิตก่อนวัยอันควร โดยเฉพาะในกลุ่มผู้สูงอายุ (2014) ทำการศึกษาผลกระทบจากการลดความเข้มข้นฝุ่น PM2.5 ต่อการเสียชีวิตก่อนวัยอันควรในประเทศญี่ปุ่น ระหว่างปี ค.ศ. 2549–2552 โดยใช้ข้อมูลการตรวจวัดจากสถานีตรวจวัดมลพิษอากาศ 1,843 แห่งในญี่ปุ่น การคำนวณค่าความเสี่ยงสัมพันธ์ (RR – relative risk) เท่ากับ 1.04 (95 % CI, 1.01– 1.08) ทุก ๆ 10 มคก./ลบ.ม. ที่สูงกว่าค่ามาตรฐานรายปี 10 มคก./ลบ.ม. ขององค์การอนามัยโลก ผลการศึกษา ชี้ว่าคุณภาพอากาศที่ดีขึ้นในช่วงเวลาที่ศึกษาช่วยลดจำนวนผู้เสียชีวิตก่อนวัยอันควรโดยเฉพาะกับกลุ่มเสี่ยง ข้อมูลนี้ใช้ในการทำนายว่าหากทำการลดระดับฝุ่น PM2.5 ลงเหลือ 10 มคก./ลบ.ม. จะช่วยลดจำนวนผู้เสียชีวิต ได้อีก 3,602 ราย ในจำนวนนี้ร้อยละ 77 เป็นกลุ่มผู้มีอายุมากกว่า 75 ปี ผลการศึกษาสรุปได้ว่าการปรับปรุง คุณภาพอากาศสามารถลดจำนวนผู้เสียชีวิตก่อนวัยอันควรได้โดยเฉพาะกับกลุ่มผู้สูงอายุ สรุปได้ว่าจากที่ประเทศญี่ปุ่น ทำการศึกษาและเก็บข้อมูลผลกระทบจากฝุ่น PM2.5 โดยใช้ข้อมูลการตรวจวัดจากสถานีตรวจวัดมลพิษอากาศ 1,843 แห่งในญี่ปุ่น การคำนวณค่าความเสี่ยงสัมพัทธ์ชี้ว่าคุณภาพอากาศที่ดีขึ้นในช่วงเวลาที่ศึกษาช่วยลดจำนวนผู้เสียชีวิตก่อนวัยอันควรได้ดังนั้นการที่เรารู้วิธีรับมือหรือปกกันตนเองการฝุ่นPM2.5 สามารถช่วยลดจำนวนผู้เสียชีวิตก่อนวัยอันควรได้

ปวีณา คำแปง (2558) ได้ศึกษาเรื่อง การออกแบบระบบฐานข้อมูล และพัฒนาระบบฐานข้อมูลด้วยโปรแกรมสำนักงาน Microsoft Excel 2010 และ โปรแกรม Visual Basic for Applications สำหรับ Microsoft Excel ดำเนินการศึกษาในช่วงปี 2558 ระหว่างการพัฒนาฐานข้อมูลผู้วิจัยได้นำเสนอ และให้นักวิชาการสาธารณสุขของศูนย์อนามัยและสำนักงานสาธารณสุขจังหวัด ทดลองใช้โปรแกรม เพื่อรวบรวมความคิดเห็นต่อการปรับปรุงโปรแกรมตลอดจนการพัฒนาผลการศึกษาต้นแบบ ฐานข้อมูลสิ่งแวดล้อมและสุขภาพที่พัฒนาขึ้น ผลวิจัยพบว่า ง่ายต่อการเก็บรวบรวมและประมวลผลข้อมูล และสะดวกต่อการนำสถานการณ์ไปใช้ประโยชน์เพื่อส่งเสริมให้องค์กรปกครองส่วนท้องถิ่นซึ่งเป็นหน่วยงานที่ปฏิบัติงานด้านอนามัย สิ่งแวดล้อมในพื้นที่ และเป็นเจ้าของข้อมูลมีการจัดการข้อมูลอย่างเป็นระบบสรุปการ

ออกแบบระบบฐานข้อมูล และพัฒนาระบบฐานข้อมูลสามารถได้โปรแกรมหลายอย่างทำงานร่วมกันได้เพื่อให้เกิดประโยชน์สูงสุดและประสิทธิภาพในการนำเสนอผลงานวิจัยนี้เผยให้เห็นถึงการใช้ประโยชน์จากการใช้เครื่องมือโปรแกรมสำนักงาน Microsoft Excel 2010 และ โปรแกรม Visual Basic for Applications ใช้งานร่วมกันเพื่อการออกแบบระบบฐานข้อมูล และพัฒนาระบบฐานข้อมูลเกิดเป็นการประมวลผลข้อมูลที่สะดวกง่ายต่อการจัดเก็บและนำเสนอต่อสถานการณ์ต่าง ๆ เกิดเป็นการเสริมประสิทธิภาพให้กับองค์

ปาจรีย์ ทองสนิท (2558) กล่าวว่า ฝุ่นละอองขนาดเล็ก Particle Matter (PM) คือส่วนผสมของอนุภาคที่มีขนาดเล็ก ร่วมกับละอองของเหลว ฝุ่นละอองขนาดเล็กอาจจะมีคุณสมบัติเป็นกรด (เช่นไนเตรดหรือซัลเฟต) เป็นสารเคมีอินทรีย์ (Organic Chemical) เป็นโลหะ เป็นดินหรือฝุ่นผง ก็ได้ ขนาดของฝุ่นละอองขนาดเล็กจะสัมพันธ์กับศักยภาพที่จะทำให้เกิดโรค โดยขนาดที่มีความสำคัญได้แก่ ขนาด 10 Micron หรือเล็กกว่า เนื่องจากสามารถที่ผ่านเข้าไปทางคอหรือจมูกไปถึงหลอดลมและปอด โดยเมื่อสูดอนุภาคเหล่านั้นเข้าไปจะมีผลต่อหัวใจและปอด และส่งผลกระทบรุนแรงต่อสุขภาพ เพื่อที่เห็นภาพว่าฝุ่นละอองขนาดเล็กมีขนาดเท่าไร จะขอยกตัวอย่างขนาดของเส้นผมที่ว่าเล็กจะมีขนาดประมาณ 70 Micron เพราะฉะนั้น PM10 จะมีขนาด เล็กว่าเส้นผมประมาณ 10–28 เท่า สำหรับอนุภาคที่ใหญ่กว่า 10 micron ได้แก่ เศษผง เศษดิน PM ขนาด 2.5 Micron หรือเล็กกว่า (PM2.5) พบได้ในหมอกควัน (Smoke) อนุภาคขนาดนี้อาจจะมาจากแหล่งกำเนิดโดยตรง (Primary Particles) เช่นจากการเผาไหม้ หรืออาจจากการ รวมตัวกันของก๊าซที่ปล่อยจากโรงไฟฟ้า หรือจาก ไอเสียรถยนต์แล้วเกิดปฏิกิริยากับอากาศ ต่อมาจึงรวมตัวกันเป็นอนุภาค (Secondary Particle) โดย PM 2.5 ส่วนใหญ่เป็นประเภท Secondary Particles

วีรนุช ปุยภิรมย์ (2556) กล่าวว่า ฝุ่นละอองขนาดเล็กกว่า 2.5 ไมครอน (PM2.5) เป็นอนุภาคของแข็งหรือ ถึงของแข็ง ที่อยู่ในสภาพกึ่งระเหย (Semi Volatile) ประกอบด้วยอนุภาคส่วนละเอียดปฐมภูมิและอนุภาคส่วนละเอียดทุติยภูมิ ผสมกันอยู่ แต่ส่วนใหญ่ PM2.5 จะเป็นอนุภาคทุติยภูมิ เกิดขึ้นในอากาศ เมื่อแก๊สต่าง ๆ เช่น ซัลเฟอร์ไดออกไซด์ (SO_2) ไนโตรเจนไดออกไซด์ (NO_2) และสารประกอบอินทรีย์ระเหย (VOC_s) ที่เกิดจากการเผาไหม้เชื้อเพลิงถูกเปลี่ยนรูปในอากาศโดยทำปฏิกิริยาทางเคมี และฟิสิกส์ ทำให้เกิดการเปลี่ยนแปลงวัฏภาคจากแก๊สไปอยู่ในรูปของอนุภาคที่อยู่ในอากาศทั้ง อนุภาคส่วนละเอียดปฐมภูมิ และอนุภาคส่วนละเอียดทุติยภูมิ มีช่วงเวลาที่อยู่ในอากาศเป็น ระยะเวลายาวนานเป็นวันหรือสัปดาห์ และสามารถเคลื่อนที่ไปได้ระยะไกล 100–1,000 กิโลเมตร (US.EPA.1999) โดยมีแนวโน้มที่จะเกิดการแพร่กระจายอย่างสม่ำเสมอ ในบริเวณเขตเมือง จึงเป็น การยากที่จะทำการสำรวจ

ย้อนกลับไปสู่แหล่งกำเนิด ฝุ่นละออง (Particle Matter : PM) หมายถึง อนุภาคของแข็งหรือของเหลว ที่มีขนาดเส้นผ่านศูนย์กลางโดยประมาณอยู่ระหว่าง 0.001 ไมครอน (1ไมครอน = 0.000001 เมตร) ซึ่งเป็นขนาดของอนุภาคฝุ่นขนาดเล็กจนถึง 500 ไมครอน ซึ่งเป็นขนาดของทราย หยาบ เวลาที่ อนุภาคมลสารเหล่านี้จะสามารถแขวนลอยอยู่ในบรรยากาศมีค่าตั้งแต่ไม่กี่วินาทีจนถึงหลายๆ เดือน ขึ้นอยู่กับขนาด นอกจากนี้อนุภาคมลสารจะเกิดปฏิกิริยาเคมีกับสารอื่น ๆ ได้ขึ้นอยู่กับชนิดของ อนุภาคมลสารและสารเคมีที่จับอยู่บนอนุภาคมลสาร ทำให้เกิดเป็นสารประกอบที่สามารถกัดกร่อน โลหะหรือเป็นอันตรายต่อพืชต่าง ๆ และยังมีผลกระทบต่อสุขภาพความเป็นอยู่ของมนุษย์อีกด้วย

กลุ่มเวชศาสตร์สิ่งแวดล้อม สำนักโรคจากการประกอบอาชีพและสิ่งแวดล้อมกรมควบคุมโรค กระทรวงสาธารณสุข (2558) กล่าวว่าปัจจัยที่มีผลต่อปัญหาหมอกควันประกอบด้วย สภาพภูมิอากาศ สภาพภูมิประเทศ เช่น อุณหภูมิความชื้น ความกดอากาศ ทิศทางลม พื้นที่แอ่งกระทะ เขตเมืองที่มีอาคารสูง ฯลฯ โดยจะสังเกตได้ว่าในพื้นที่ดังกล่าว จะประสบกับปัญหาหมอกควันรุนแรงกว่าพื้นที่อื่น ๆ เช่น ในพื้นที่ที่เป็นลักษณะแอ่งกระทะที่มีภูเขาล้อมรอบจะมีโอกาสที่จะเกิดปัญหาหมอกควันรุนแรงเพิ่มมากขึ้น และในวันที่มีความกดอากาศสูงหรือไม่มีการพัดผ่านของลม จะทำให้หมอกควันลอยปกคลุมในพื้นที่อย่างยาวนานกว่าวันที่มีอากาศแจ่มใสหรือมีพัดผ่านของลมหรือมีความชื้นในอากาศสูง เช่น ฝนตก ค่าฝุ่นละอองฯ PM10 เกินค่ามาตรฐานวันแรกในวันที่ 1 กุมภาพันธ์ 2558 ที่จังหวัดลำปางการประปาส่วนภูมิภาคแม่เมาะ อำเภอแม่เมาะ จังหวัดลำปาง มีค่าเท่ากับ 127 ไมโครกรัมต่อลูกบาศก์เมตร(ค่ามาตรฐานเท่ากับ 120 ไมโครกรัมต่อลูกบาศก์เมตร) จังหวัดที่มีจำนวนวันของค่าฝุ่นละอองฯ PM10 เกินค่ามาตรฐานสูงสุดเรียงตามลำดับ ได้แก่จังหวัดเชียงราย มีค่าเท่ากับ 381 ไมโครกรัมต่อลูกบาศก์เมตร เมื่อวันที่ 18 มีนาคม 2558 รองลงมาที่จังหวัดแม่ฮ่องสอน มีค่าเท่ากับ 303 ไมโครกรัมต่อลูกบาศก์เมตร เมื่อวันที่ 16 มีนาคม 2558 ที่จังหวัดเชียงใหม่จุดตรวจวัดโรงเรียนยุพราช มีค่าเท่ากับ 299 ไมโครกรัมต่อลูกบาศก์เมตร วันที่ 17 มีนาคม 2558

ดังนั้น วรรณกรรมที่เกี่ยวข้องทั้งหลายที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้จัดทำโครงงานได้เล็งเห็นความสำคัญของการนำเทคโนโลยีและการเลือกเครื่องมือมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลหมอกควันและเผยแพร่ข้อมูลสารสนเทศบน web browser เนื่องจากเทคโนโลยีปัจจุบันได้เข้ามามีส่วนช่วยในการจัดการข้อมูล การนำเสนอ การกระจ่ายข้อมูลข่าวสารและ การดำเนินงานต่าง ๆ ให้มีประสิทธิ รวดเร็วและมีความสะดวกสบายเป็นอย่างมา ให้ผู้ใช้งานยังเข้าถึงข้อมูลได้ง่าย และมีประสิทธิภาพ

2.5 บทสรุป

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลจากศูนย์ข้อมูลหมอกควันแห่งประเทศไทยเพื่อนำเสนอข้อมูลสารสนเทศบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับระบบ และข้อมูลเกี่ยวกับฐานข้อมูลขนาดใหญ่ประกอบด้วย 3 ปัจจัย ขนาด (value) ความหลากหลาย (variety) ความรวดเร็ว (velocity) และทั้งยังมีเอกสารที่เกี่ยวข้องกับการวิเคราะห์อนุกรมเวลาประกอบด้วย 4 ส่วน ส่วนแนวโน้ม (Trend Component) ส่วนฤดูกาล (Seasonal Component) ส่วนวัฏจักร (Cyclical Component) และขั้นตอนการทำเหมืองข้อมูลแบ่งเป็น 4 ขั้นตอน การระบุปัญหาที่เกิดขึ้นกับธุรกิจ ส่วนของการทำเหมืองข้อมูล การนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วนการทำเหมืองข้อมูลมาลองปฏิบัติจริงกับธุรกิจ การวัดประสิทธิภาพของเทคนิคการทำเหมืองข้อมูลที่จะนำมาใช้จากผลลัพธ์