

บทที่ 2

แนวคิด ทฤษฎีเอกสารที่เกี่ยวข้อง

จากการศึกษาเอกสารงานการวิเคราะห์ข้อมูลรายการเครื่องดื่มน้ำที่ลูกค้าซื้อพร้อมกัน โดยใช้โมเดล Association Rule ด้วยการใช้อัลกอริทึมความสัมพันธ์ Apriori Algorithm เพื่อจะสามารถมาทำการพยากรณ์ปริมาณวัตถุดิบที่ต้องนำมาใช้ในแต่ละวันเพื่อความพอดีเหมาะสมของปริมาณวัตถุดิบในแต่ละวันและเพื่อให้ได้ข้อมูลแนวทางในการจัดโปรโมชันคู่แข่งเครื่องดื่ม ผู้ศึกษาพบว่ามีการเอกสาร แนวความคิดและโครงการต่าง ๆ ที่เกี่ยวข้องกับการศึกษาสรุปได้ ดังนี้

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic)

2.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

2.2.6 ทฤษฎีเกี่ยวกับการ Visualization

2.2.7 ทฤษฎีเกี่ยวกับกฎความสัมพันธ์ ด้วย Apriori Algorithm

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม Rapid Miner Studio

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic) Data analytics เป็นศาสตร์ของการวิเคราะห์ข้อมูลต่าง ๆ จาก Big data เพื่อช่วยใน ด้านธุรกิจ หรือตามวัตถุประสงค์อื่น ๆ ที่ต้องการซึ่งเริ่มต้นจากความสามารถของเราในการนำ ข้อมูลเหล่านั้นมาให้อยู่ในรูปแบบที่พร้อมจะประมวลผลได้ โดยเทคโนโลยีหรือชุดคำสั่งและ แบบจำลองที่สร้างขึ้น เพื่อนำข้อมูลที่ได้ผ่านการวิเคราะห์แล้วมาใช้ หรือแปลความหมายโดย บุคคลที่ได้รับการฝึกอบรมการใช้เทคโนโลยีเหล่านั้น รูปแบบของการวิเคราะห์ข้อมูลสามารถ แบ่งได้ดังนี้

1) การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive analytics) เป็นการวิเคราะห์เพื่อแสดงผลของรายการทางธุรกิจ เหตุการณ์ หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้นในลักษณะที่ง่ายต่อการเข้าใจ หรือต่อการตัดสินใจ ตัวอย่างเช่น รายงานการขาย รายงานผลการดำเนินงาน

2) การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้นปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์กันของสิ่งที่เกิดขึ้น ตัวอย่างเช่น ความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่ช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

3) การวิเคราะห์แบบพยากรณ์ (Predictive analytics) เป็นการวิเคราะห์เพื่อพยากรณ์สิ่งที่กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้น โดยใช้ข้อมูลที่ได้เกิดขึ้นแล้วกับแบบจำลองทางสถิติ หรือเทคโนโลยีปัญญาประดิษฐ์ต่าง ๆ (Artificial intelligence) ตัวอย่างเช่น การพยากรณ์ยอดขาย การพยากรณ์ผลประชามติ

4) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive analytics) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด เป็นทั้งการพยากรณ์สิ่งต่าง ๆ ที่จะเกิดขึ้น ข้อดี ข้อเสีย สาเหตุ และระยะเวลาของสิ่งที่เกิดขึ้น และการให้คำแนะนำทางเลือกต่าง ๆ ที่มีอยู่ และผลของแต่ละทางเลือก

2.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization) การนำข้อมูลดิบมาเปลี่ยนเป็นกราฟ แผนภูมิ หรือแม้กระทั่งวิดีโอที่ช่วยอธิบาย ปริมาณ ตัวเลข และช่วยให้คุณได้ ข้อมูลเชิงลึกจากข้อมูลเหล่านั้น การใช้ Data Visualization 9 เปลี่ยนแนวคิดของเราเกี่ยวกับการสร้างคุณค่าให้กับข้อมูล ทำให้เราค้นพบรูปแบบใหม่ ๆ และ มองเห็นแนวโน้มของข้อมูลมากขึ้น

หากต้องการประสบความสำเร็จในโลกธุรกิจที่มีการแข่งขันสูง ความสามารถในการตัดสินใจที่ถูกต้องในเวลาที่เหมาะสม โดยอาศัยข้อมูลเชิงลึกที่รัดกุมเป็นสิ่งสำคัญ ปัจจุบันองค์กรธุรกิจส่วนใหญ่ในประเทศไทย เริ่มใช้ประโยชน์จากการทำ Data Visualization เพราะว่าการเครื่องมือนี้เป็นกุญแจสำคัญในการแสดงผลของข้อมูลให้เข้าใจง่าย ซึ่งจะช่วยให้ทุกแผนกในฝ่ายงานได้ใช้ประโยชน์ ในการหาแนวโน้มและความสัมพันธ์เชื่อมโยง อีกทั้งทำให้ทราบข้อมูลได้เร็วขึ้น ส่งผลให้การดำเนินงานขององค์กรมีประสิทธิภาพเพิ่มขึ้น จนดำเนินการถึงการวางกลยุทธ์ภายในองค์กร และสามารถนำไปใช้กับการทำเว็บไซต์ได้ Data Visualization ช่วยให้การสื่อสารข้อมูล มีความรวดเร็วและเข้าใจง่ายมากขึ้น



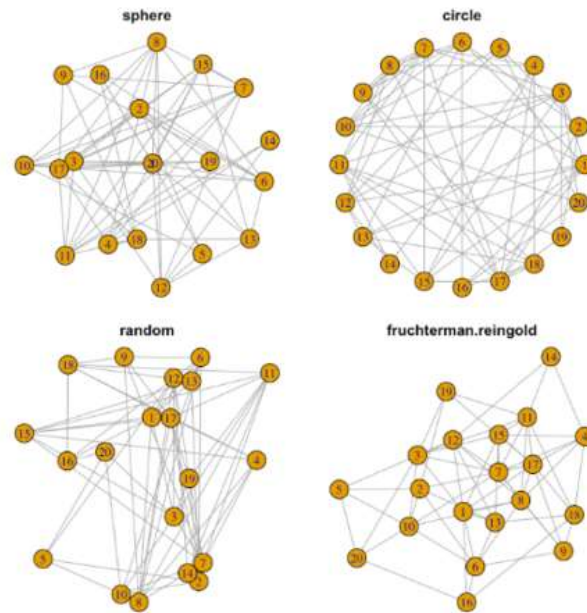
ภาพที่ 2.1 แสดง Data Visualization

(ที่มา : <https://vizzlo.com/data-viz-guide>)

พื้นฐานของการสร้าง Data Visualization คือ การ Mapping ส่วนข้อมูลกับส่วนของ Graphic เข้าด้วยกัน ซึ่งตอนนี้มีโปรแกรมสำเร็จรูปในการสร้าง Data Visualization หลากหลาย โปรแกรมมีฟังก์ชันการใช้งานที่เข้าใจง่าย เช่น การสร้างฟิลเตอร์ การออกแบบเพื่อให้งานการวิเคราะห์ข้อมูลมีความยืดหยุ่นเป็นต้น ตัวอย่างรูปแบบ Data Visualization ที่นิยมใช้กันมีดังนี้

1) NETWORK/FLOW

Network Graph ใช้แสดงความเชื่อมโยงของเครือข่ายหรือความสัมพันธ์ในกลุ่ม

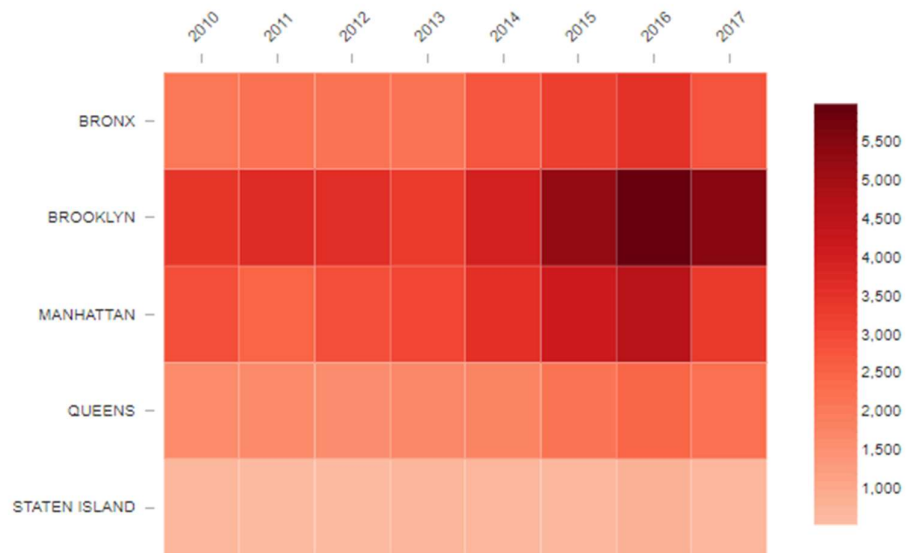


ภาพที่ 2.2 แสดง Network Graph

(ที่มา : <https://www.r-graph-gallery.com/network.html>)

2) Relationship

Heatmap ใช้แสดงรูปแบบความสัมพันธ์ของข้อมูล โดยจะแสดงออกมาในรูปแบบ ของ “สี” ซึ่งแต่ละสีจะบ่งบอกถึงระดับความถี่ของพฤติกรรม แต่ใช้อ่านค่าความแตกต่าง เล็กน้อย ได้ยากที่

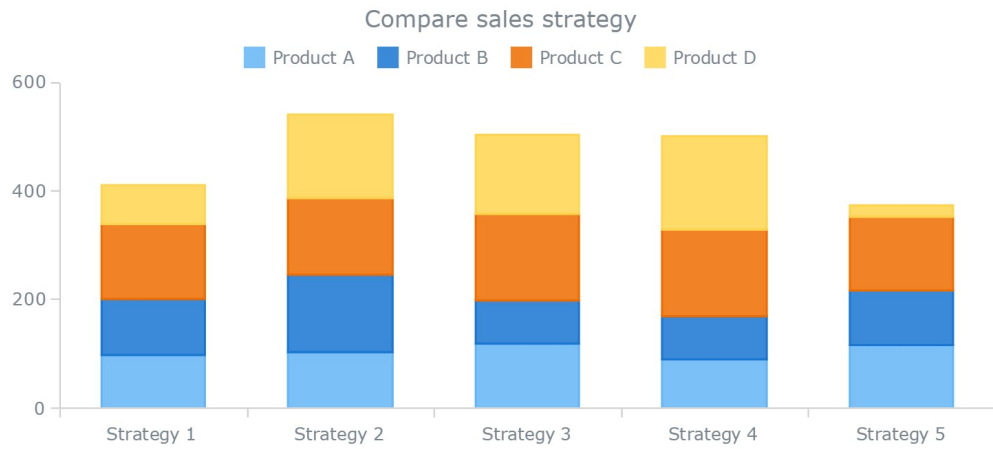


ภาพที่ 2.3 แสดง Heatmap

(ที่มา : <https://www.displayr.com/how-to-create-a-heatmap-in-displayr/>)

3) Comparison

Stacked Bar Chart ใช้เปรียบเทียบค่าผลรวมและสัดส่วนจากข้อมูลหลายกลุ่ม อาจอ่านสัดส่วนยากเมื่อมีข้อมูลหลายกลุ่มมากเกินไป

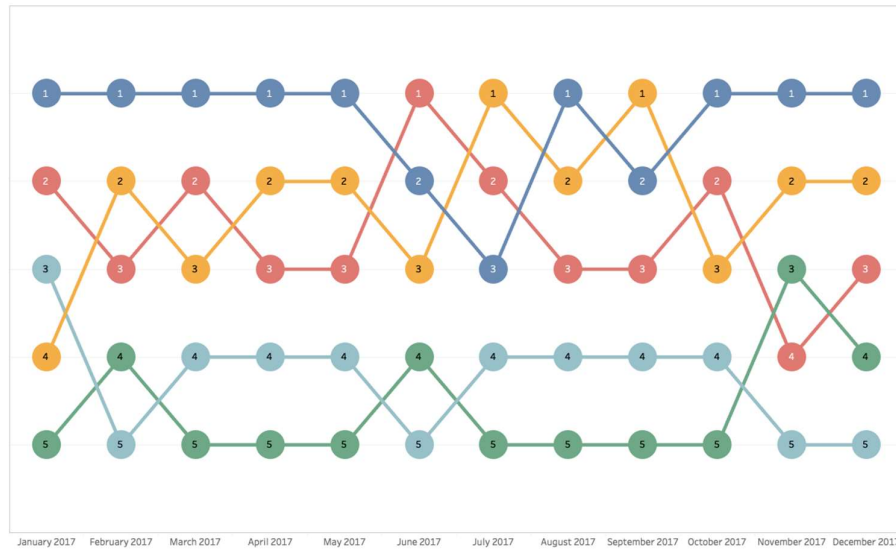


ภาพที่ 2.4 แสดง Stacked Bar Chart

(ที่มา : <https://github.com/kalkih/mini-graph-card/issues/79>)

4) Ranking

Bump Chart ใช้แสดงการเปลี่ยนแปลงลำดับของข้อมูลในหลายช่วงเวลา โดยใช้สี เพื่อช่วยให้อ่านข้อมูลได้ง่ายขึ้น

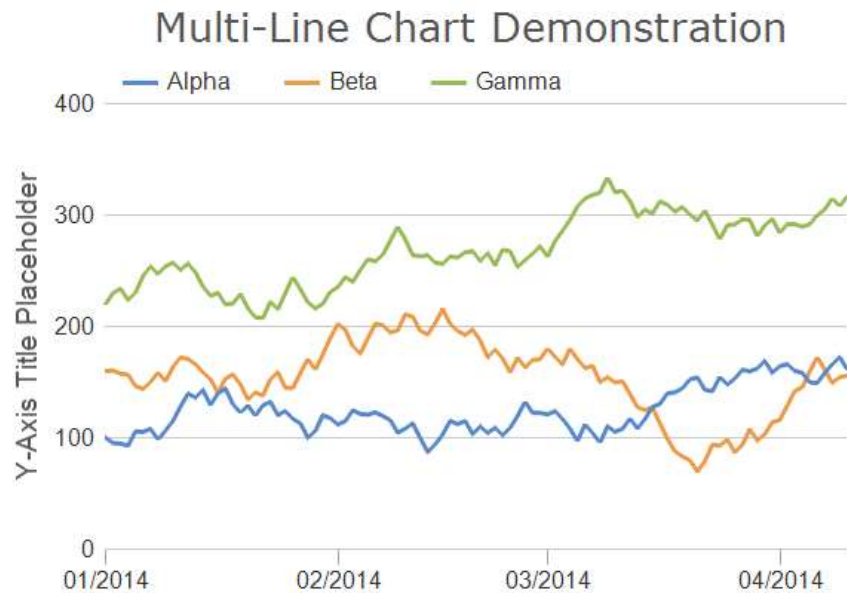


ภาพที่ 2.5 แสดง Bump Chart

(ที่มา : <https://creativewithdata.com/2019/04/24/how-to-create-a-bump-chart-in-tableau/>)

5) Time-Series

Line Chart ใช้เปรียบเทียบข้อมูลเพื่อดูแนวโน้ม (Trends) โดยอาจจะเทียบกับมิติ ของ เวลา time

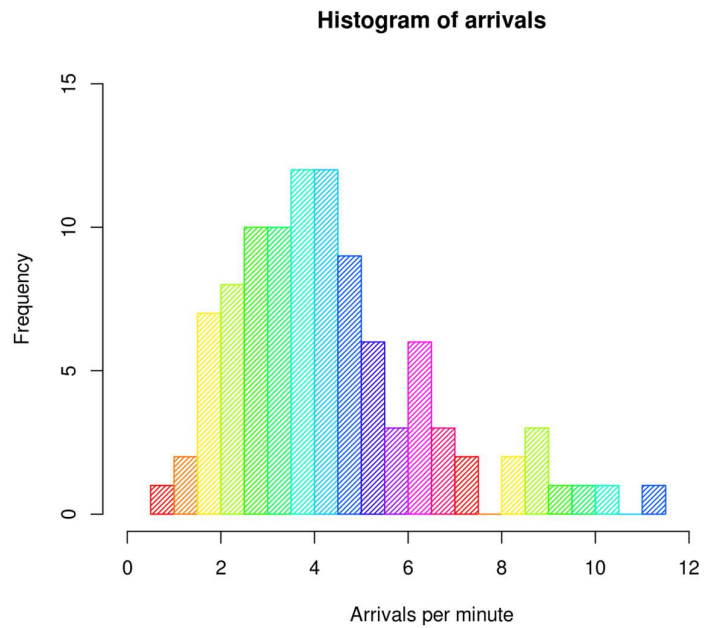


ภาพที่ 2.6 แสดง Line Chart

(ที่มา : https://www.advsofteng.com/gallery_line.html)

6) Distribution

Histogram กราฟแท่งแบบเฉพาะที่แสดงความสัมพันธ์ระหว่างข้อมูลเป็นหมวดหมู่ที่เรียกว่าชั้นข้อมูลกับความถี่ของข้อมูล เพื่อดูการกระจายของข้อมูล ลดระยะห่างระหว่างแท่งเพื่อให้เห็น “รูปทรง” ของข้อมูลชัดเจน

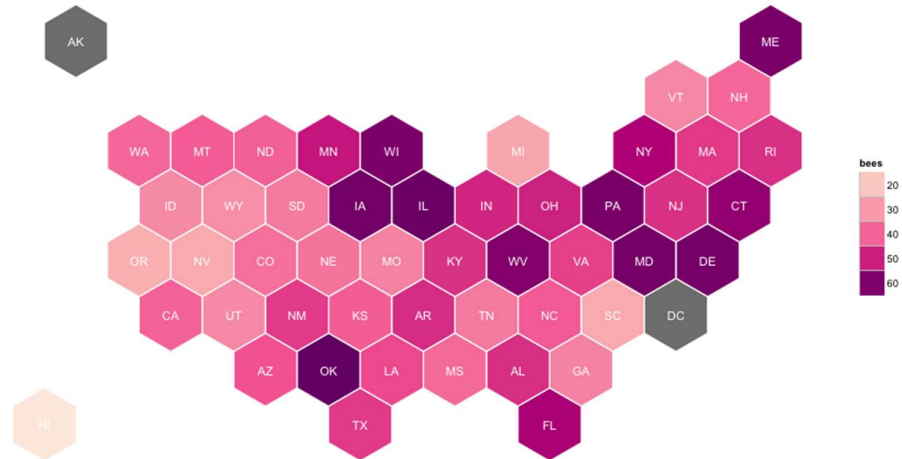


ภาพที่ 2.7 แสดง Histogram

(ที่มา : <https://en.wikipedia.org/wiki/Histogram>)

7) Spatial

Hexbin Map เป็นการแสดงผลข้อมูลด้วยรูปทรงหกเหลี่ยมที่มีขนาดเท่ากัน โดย กำหนดสีแทนระดับความถี่ของจำนวนข้อมูล ส่วนมากนำไปใช้ลงบนตำแหน่งทางภูมิศาสตร์ เพื่อให้เห็นภาพได้ชัดเจน

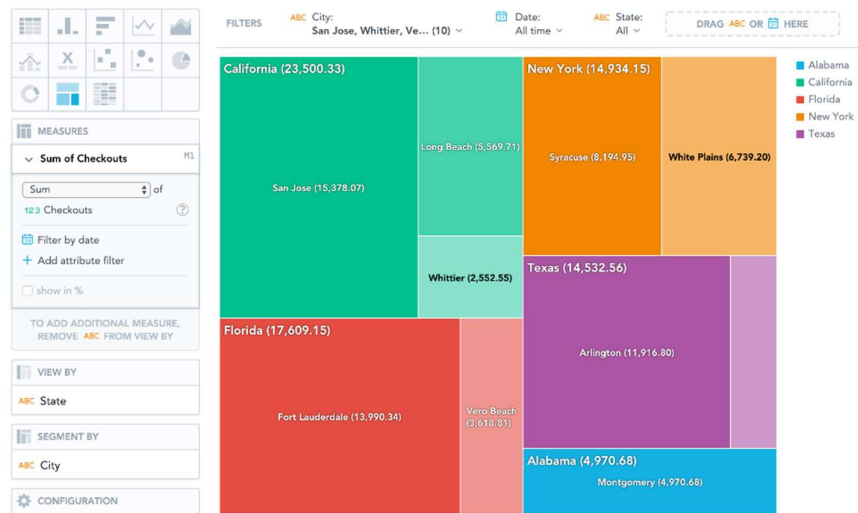


ภาพที่ 2.8 แสดง Hexbin Map

(ที่มา : <https://rud.is/b/2015/05/14/geojson-hexagonal-statebins-in-r/>)

8) Composition

Treemap ใช้แสดงสัดส่วนของข้อมูลแบบมีลำดับชั้น (hierarchy) แสดงเป็นรูป สีเหลี่ยม โดยจะแสดงทั้งในแง่ของสีเดียวกันจัดเป็นกลุ่มเดียวกัน อาจอ่านสัดส่วนยากเมื่อมี กลุ่มขนาดเล็กจำนวนมาก



ภาพที่ 2.9 แสดง Treemap

(ที่มา : <https://help.gooddata.com/doc/enterprise/en/dashboards-and-insights/analytical-designer/visualize-your-data/insight-types/treemaps>)

จากที่ได้กล่าวไปจะเห็นได้ว่าข้อมูลแบบ Data Visualization มีพลังทางการสื่อสารอย่างมาก เพราะสามารถแสดงผลและเพิ่มความสนใจได้เป็นอย่างดี หากองค์กรได้ลองนำการแสดงผลข้อมูลแบบ Data Visualization มาใช้ก็จะช่วยในการจัดการระดับสายงานแผนกอื่น ๆ ในบริษัท จะส่งผลให้การวิเคราะห์ข้อมูลและตัดสินใจในสายงานต่าง ๆ ได้ง่าย สามารถช่วยให้มีประสิทธิภาพที่ดีต่อธุรกิจและทีม หรือนำมาใช้ในหน้าแรกของเว็บไซต์ เพื่อทำการโปรโมทโปรโมชันหรือข้อเสนอทางการตลาดของธุรกิจ แล้วจะพบว่า Data Visualization ช่วยให้เกิดการตอบสนองกับข้อความมากขึ้น

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่ (Big Data)

Big Data หรือข้อมูลขนาดใหญ่มหาดศาล ที่เกิดขึ้นแต่ละวัน เช่น ข้อมูลจากการทำธุรกิจ ข้อมูลจากการซื้อขายสินค้า ข้อมูลคนผ่านเข้าออกสถานีรถไฟ หรือ ข้อมูลคนที่ใช้ บริการของ Facebook ข้อมูลที่มีปริมาณมหาศาลจนซอฟต์แวร์หรือฮาร์ดแวร์ธรรมดา นั้นไม่ สามารถที่จะจัดการหรือวิเคราะห์ได้อย่างมีประสิทธิภาพ ดังนั้นต้องมีเทคโนโลยีและสถาปัตยกรรมไอทีรุ่นใหม่ ซึ่งอาจมาในรูปแบบซอฟต์แวร์ ที่สามารถรองรับการจัดการเก็บ การจัดการ กรองเลือกข้อมูล การวิเคราะห์ แสดงผล ซึ่งสามารถนำไปประยุกต์ใช้งานให้เป็นประโยชน์ต่อองค์กรต่าง ๆ สำหรับภาคเอกชนทั่วไปก็ต้องนำมาใช้ให้เกิดประโยชน์ต่อองค์กรมากที่สุด วัตถุประสงค์การใช้งานเพื่อองค์กรของตนเป็นหลัก แต่นอกจาก Big Data จะมีความสำคัญกับภาคเอกชนแล้วสำหรับภาครัฐนั้นก็มีความสำคัญอย่างยิ่งเช่นกัน เพราะรัฐจำเป็นต้องใช้ข้อมูลของ Big Data มากก่อให้เกิดประโยชน์ต่อประชาชนทุกคน เพื่อจะกระจายความช่วยเหลือไปถึงทุกที่ พร้อมทั้งสามารถนำข้อมูลที่มีการวิเคราะห์ไปใช้ประโยชน์ในการบริหาร วางแผนกำหนดทิศทางในด้านต่าง ๆ อย่างมีประสิทธิภาพที่สูงขึ้น คำว่า Big Data มักจะสื่อถึงชุดข้อมูลที่มีขนาดใหญ่ ซึ่งอาจจะแบ่งย่อยออกเป็นข้อมูล 3 รูปแบบคือ

- 1) ข้อมูลเชิงโครงสร้าง (structured data) คือตารางข้อมูลที่มีการจัดเรียง อย่างมีรูปแบบชัดเจนและเป็นระเบียบ สามารถนำมาใช้วิเคราะห์ได้เลย เช่น ข้อมูลที่เก็บไว้ใน โปรแกรม spreadsheet อย่าง Microsoft Excel
- 2) ข้อมูลกึ่งโครงสร้าง (semi-structured data) คือข้อมูลที่ถูกรวบรวมอย่างมี รูปแบบในระดับหนึ่งและข้อมูลที่สามารถค้นหา (search) หรือแท็ก (tag) ได้ เช่น เว็บไซต์ที่มี การระบุชื่อเพจ คำสำคัญในเพจ และวันที่อัปเดตข้อมูล
- 3) ข้อมูลที่ไม่มีโครงสร้างแน่นอน (Unstructured data) มักจะหมายถึงข้อความยาวๆ ข้อความเสียง รูปภาพ และวิดีโอ

คุณลักษณะของ Big Data ที่มีประสิทธิภาพ มี 6 ประการประกอบด้วย

- 1) ปริมาณมาก (Volume) ซึ่งได้จากการที่องค์กรต่าง ๆ รวบรวมข้อมูลจาก หลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย ซึ่งสามารถเป็นได้ทั้งรูปแบบ Online และ Offline ซึ่งส่วนมากแล้วจะมี ปริมาณมากกว่าหน่วย TB (Terabyte) ขึ้นไป

2) มีการเปลี่ยนแปลงอย่างรวดเร็ว (Velocity) ด้วยเทคโนโลยีที่ล้ำหน้าขึ้นทุก วัน ความเร็วในการได้รับข้อมูลที่เพิ่มขึ้นและต้องได้รับการจัดการในเวลาที่เหมาะสม ข้อมูล เหล่านี้ในแบบเรียลไทม์ ส่งผ่านข้อมูลกันอย่างต่อเนื่อง (Real-time) จนทำให้การวิเคราะห์ต่างๆ เกินขีดข้อจำกัด หรือไม่สามารถจับรูปแบบหรือทิศทางของข้อมูลได้

3) หลากหลายประเภทหรือแหล่งที่มา (Variety) หมายถึงรูปแบบของข้อมูลที่ แตกต่างกันอย่างออกไป นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึง เอกสารข้อความ ทั้งในรูปแบบ ตัวอักษร วิดีโอ รูปภาพ ไฟล์ต่าง ๆ และหลากหลายแหล่งที่มา เช่น Social Network ต่าง ๆ อีเมล วิดีโอ เสียง ข้อมูลหุ่น และธุรกรรมทางการเงิน

4) ยังไม่ผ่านการประมวลผล (Veracity) ข้อมูลที่มีระดับคุณภาพปะปนกันไป ยังไม่ผ่านการ Process ให้อยู่ในรูปแบบของข้อมูลที่สามารถใช้สร้างประโยชน์ต่อองค์กร ซึ่งเป็น ข้อมูลที่ยังไม่สมบูรณ์จึงยังไม่สามารถนำไปประกอบการตัดสินใจได้

5) คุณค่า (Value) หมายถึง ข้อมูลมีประโยชน์และมีความสัมพันธ์ในเชิงธุรกิจ ซึ่งต้องเข้าใจก่อนว่าไม่ใช่ทุกข้อมูลจะมีประโยชน์ในการเก็บและวิเคราะห์ ข้อมูลที่มีประโยชน์ จะต้องเกี่ยวข้องกับวัตถุประสงค์ทางธุรกิจ เช่นถ้าต้องการเพิ่มขีดความสามารถในการแข่งขัน ในตลาดของผลิตภัณฑ์ที่ขาย ข้อมูลที่มีประโยชน์ที่สุดน่าจะเป็นข้อมูลผลิตภัณฑ์ของคู่แข่ง

6) ความแปรผันได้ (Variability) หมายถึง ข้อมูลสามารถในการเปลี่ยนแปลง รูปแบบไปตามการใช้งาน หรือสามารถคิดวิเคราะห์ได้จากหลายแง่มุม และรูปแบบในการ จัดเก็บข้อมูลก็อาจจะต่างกันออกไปในแต่ละแหล่งของข้อมูล

2.2.1.2 การจัดการข้อมูลขนาดใหญ่ การรวบรวมข้อมูลมาจัดเก็บ (Storage) เป็นการรวบรวมข้อมูลของจาก หลากหลายทั้งที่มา และการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบ ดั้งเดิม ETL (Extract Transform and Load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิค วิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาด เทราไบต์ และเพตะ 16 ไบต์ ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล จัดรูปแบบ ให้เหมาะสมสำหรับการใช้ในการ วิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์ นั้น ๆ

1) การประมวลผล (Processing) ข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ On-premises หรือแบบ Cloud ขึ้นกับความต้องการ หรือความสะดวกในการใช้ ซึ่งเรา

สามารถใช้และประเมินผลได้เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูล หรือข้อมูลบางอันต้องการความยืดหยุ่นสูง และไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยมกันเป็นอย่างมาก

2) การวิเคราะห์ (Analyst) การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่าง และชัดเจนในชุดข้อมูลที่มีอยู่ในการสำรวจข้อมูลยังทำให้ค้นพบสิ่งใหม่ แสร้งสิ่งที่ค้นพบใหม่ ๆ ต่อผู้อื่น สร้างรูปแบบจำลองข้อมูลด้วยการเรียนรู้ของเครื่องจักร และปัญญาประดิษฐ์ AI และนำข้อมูลเหล่านั้นไปใช้งาน

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม เทคนิคต่าง ๆ ในการจำแนกประเภท ของข้อมูล อาทิ เช่น การจำแนก ข้อมูลด้วยการสร้างต้นไม้ตัดสินใจ (Decision tree classifier) การจำแนกข้อมูลด้วยเบย์เซียนและเบย์เซียนปิลิฟเน็ตเวิร์ค (Bayesian classifier and Bayesian belief networks) การจำแนกข้อมูลด้วยกฎ (Rulebased classifiers) การจำแนกข้อมูลด้วยโครงข่าย ประสาทเทียมและการส่งค่าย้อนกลับ (Neural network and backpropagation) การจำแนกข้อมูลจากกฎความสัมพันธ์ของข้อมูล (Classification based on association rule mining) การค้นหาเพื่อนบ้าน ใกล้สุด k อันดับ (k-nearest-neighbor) และทำการศึกษาเกี่ยวกับการทำนายข้อมูล ที่จะประกอบไปด้วยการถดถอยเชิงเส้นตรง (linear regression) และการถดถอยที่ไม่เป็นเส้นตรง (Nonlinear regression) การทำเหมืองข้อมูล มีการใช้เทคนิคหลายแบบในการวิเคราะห์ข้อมูล โดยมากใช้ 3 เทคนิคใหญ่ๆ คือ Classification โดยสร้างโมเดลจากข้อมูลที่มีอยู่เพื่อทำนายเหตุการณ์ที่จะเกิดขึ้น ในอนาคต Clustering แบ่งข้อมูลเป็นหลายๆ 17 กลุ่มอาศัยความคล้ายคลึงกันของข้อมูล Association rules อาศัยความสัมพันธ์ของข้อมูลที่เกิดร่วมกัน สร้างเป็นกฎความสัมพันธ์ เช่น “ซื้อเบียร์แล้วจะ ซื้อผ้าอ้อมไปด้วย”

ขั้นตอนในการวิเคราะห์ข้อมูลได้ใช้หลักการของกระบวนการหาความรู้แบบ Cross Industry Standard Process for Data Mining (CRISP-DM) ซึ่งเป็นแนวทางในการดำเนินงาน CRISP-DM ประกอบไปด้วย 6 ขั้นตอน

1) ความเข้าใจทางธุรกิจ (Business Understanding) เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ ข้อมูลทางดาต้าไมน์นิ่งพร้อมทั้งวางแผนในการดำเนินการคร่าว ๆ

2) ความเข้าใจเกี่ยวกับข้อมูล (Data Understanding) ในตอนนี้เริ่มจากการเก็บรวบรวม ข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของ ข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

3) การเตรียมข้อมูล (Data Preparation) ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) เช่น การแปลงข้อมูล ให้อยู่ ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น โดยขั้นตอนนี้จะ เป็น ขั้นตอนที่ใช้ เวลา มาก ที่สุด ของ กระบวนการ CRISP-DM

4) การพัฒนาแบบจำลอง (Modeling) ในขั้นตอนนี้จะ เป็น ขั้นตอนการวิเคราะห์ ข้อมูล ด้วยเทคนิคทางดาต้าไมน์นิ่ง ที่ได้แนะนำไปแล้ว เช่น การจำแนกประเภทข้อมูล หรือ การแบ่งกลุ่ม ข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปขั้นตอน Data Preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย เทคนิคในการวิเคราะห์ข้อมูลต่าง ๆ

5) การทดสอบแบบจำลอง (Evaluation) ในขั้นตอนนี้เราจะ ได้ ผลการวิเคราะห์ ข้อมูล ด้วยเทคนิคทางดาต้าไมน์นิ่งแล้วแต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปก็จะต้องมีการวัด ประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความ น่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไข เพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้

6) การนำแบบจำลองไปใช้ (Deployment) ในกระบวนการทำงานของ CRISP-DM นั้นไม่ได้หยุดเพียงแค่ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิ่งเท่านั้น แม้ว่าผลลัพธ์ที่ได้จะแสดงถึงองค์ความรู้ที่มีประโยชน์ แต่จะต้องนำองค์ความรู้ที่ได้เหล่านี้ไป ใช้ได้จริง ในองค์กรหรือบริษัท ตัวอย่างเช่น การสร้างรายงานเพื่อให้ผู้บริหารหรือนักการตลาด เข้าใจได้ง่าย และสามารถนำไปออกโปรโมชั่นได้ เป็นต้น

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

เว็บไซต์เป็นสื่อที่ได้รับความนิยมอย่างมากบนอินเทอร์เน็ต ซึ่งเว็บไซต์เป็นสื่อที่อยู่ในความควบคุมของผู้ใช้โดยสมบูรณ์ กล่าวคือ ผู้ใช้สามารถตัดสินใจเลือกได้ว่าจะดูเว็บไซต์ใดและจะไม่เลือกดูเว็บไซต์ใด ได้ตามต้องการ จึงทำให้ผู้ใช้ไม่มีความอดทนต่ออุปสรรคและปัญหาที่เกิดขึ้นจากการออกแบบเว็บไซต์ผิดพลาดถ้าผู้ใช้เห็นว่าเว็บที่กำลังดูอยู่นั้นไม่มีประโยชน์ต่อตัวเขาหรือไม่เข้าใจว่าเว็บไซต์นี้จะใช้งานอย่างไร เขาก็สามารถที่จะเปลี่ยนไปดูเว็บไซต์อื่น ๆ ได้อย่างรวดเร็ว เนื่องจากในปัจจุบันมีเว็บไซต์อยู่มากมาย และยังมีเว็บไซต์ที่เกิดขึ้นใหม่ ๆ ทุกวัน ผู้ใช้จึงมีทางเลือกมากขึ้น และสามารถเปรียบเทียบคุณภาพของเว็บไซต์ต่าง ๆ ได้เอง

เว็บไซต์ที่ได้รับการออกแบบอย่างสวยงาม มีการใช้งานที่สะดวก ย่อมได้รับความนิยมสนใจจากผู้ใช้งานมากกว่าเว็บไซต์ที่ดูสับสนวุ่นวาย มีข้อมูลมากมายแต่หาอะไรไม่เจอ นอกจากนี้ยังใช้เวลาในการแสดงผลแต่ละหน้านานเกินไป ซึ่งปัญหาเหล่านี้ล้วนเป็นผลมาจากการออกแบบเว็บไซต์ไม่ดีทั้งสิ้น

ดังนั้น การออกแบบเว็บไซต์จึงเป็นกระบวนการสำคัญในการสร้างเว็บไซต์ ให้ประทับใจผู้ใช้ ทำให้เขาอยากกลับเข้ามาเว็บไซต์เดิมอีกในอนาคต ซึ่งนอกจากต้องพัฒนาเว็บไซต์ที่ดีมีประโยชน์แล้ว ยังต้องคำนึงถึงการแข่งขันกับเว็บไซต์อื่น ๆ อีกด้วย

2.2.3.1 องค์ประกอบของการออกแบบเว็บไซต์

1) ความเรียบง่าย (Simplicity) หมายถึง การจำกัดองค์ประกอบเสริมให้ เหลือเฉพาะองค์ประกอบหลัก กล่าวคือในการสื่อสารเนื้อหากับผู้ใช้นั้น เราต้องเลือกเสนอลิ่งที่ เราต้องการนำเสนอจริง ๆ ออกมาในส่วนของกราฟิก สี สัน ตัวอักษรและภาพเคลื่อนไหว ต้อง เลือกให้พอเหมาะ ถ้าหากมีมากเกินไปจะรบกวนสายตาและสร้างความรำคาญต่อผู้ใช้ ตัวอย่าง เว็บไซต์ที่ได้รับการออกแบบที่ดี ได้แก่ เว็บไซต์ของบริษัทใหญ่ ๆ อย่างเช่น Apple Adobe Microsoft หรือ Kokia ที่มีการออกแบบเว็บไซต์ในรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งาน อย่างสะดวก

2) ความสม่ำเสมอ (Consistency) หมายถึง การสร้างความสม่ำเสมอ ให้ เกิดขึ้นตลอดทั้งเว็บไซต์ โดยอาจเลือกใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ก็ได้ เพราะถ้าหากว่า แต่ละหน้าในเว็บไซต์นั้นมีความแตกต่างกันมากจนเกินไป อาจทำให้ผู้ใช้เกิดความสับสนและไม่ แน่ใจว่ากำลังอยู่ในเว็บไซต์เดิมหรือไม่ เพราะฉะนั้นการออกแบบเว็บไซต์ในแต่ละหน้า

ควรที่จะมีรูปแบบ สไตล์ของกราฟิก ระบบเนวิเกชัน (Navigation) และโทนสีที่มีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ (Identity) ในการออกแบบเว็บไซต์ต้องคำนึงถึง ลักษณะขององค์กรเป็นหลัก เนื่องจากเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กร การเลือกใช้ตัวอักษร ชุดสี รูปภาพหรือกราฟิก จะมีผลต่อรูปแบบของเว็บไซต์เป็นอย่างมาก ตัวอย่างเช่น ถ้าเราต้องออกแบบเว็บไซต์ของธนาคารแต่เรากลับเลือกสีส้มและกราฟิกมากมาย อาจทำให้ผู้ใช้คิดว่าเป็นเว็บไซต์ของสวนสนุกซึ่งส่งผลต่อความเชื่อถือขององค์กรได้

4) เนื้อหา (Useful Content) ถือเป็นสิ่งสำคัญที่สุดในเว็บไซต์ เนื้อหาในเว็บไซต์ต้องสมบูรณ์และได้รับการปรับปรุงพัฒนาให้ทันสมัยอยู่เสมอ ผู้พัฒนาต้องเตรียมข้อมูลและเนื้อหาที่ผู้ใช้ต้องการให้ถูกต้องและสมบูรณ์ เนื้อหาที่สำคัญที่สุดคือเนื้อหาที่ทีมผู้พัฒนาสร้างสรรค์ขึ้นมาเอง และไม่ไปซ้ำกับเว็บอื่น เพราะจะถือเป็นสิ่งที่ดึงดูดผู้ใช้ให้เข้ามาเว็บไซต์ได้เสมอ แต่ถ้าเป็นเว็บที่ดึงข้อมูลจากเว็บอื่น ๆ มาเมื่อใดก็ตามที่ผู้ใช้ทราบว่ามีข้อมูลนั้นมาจากเว็บใด ผู้ใช้ก็ไม่จำเป็นต้องกลับมาใช้งานลิงค์เหล่านั้นอีก

5) ระบบเนวิเกชัน (User-Friendly Navigation) เป็นส่วนประกอบที่มีความสำคัญต่อเว็บไซต์มาก เพราะจะช่วยไม่ให้เกิดความสับสนระหว่างดูเว็บไซต์ ระบบเนวิเกชันจึงเปรียบเสมือนป้ายบอกทาง ดังนั้นการออกแบบเนวิเกชัน จึงควรให้เข้าใจง่าย ใช้งานได้สะดวก ถ้ามีการใช้กราฟิกก็ควรสื่อความหมาย ตำแหน่งของการวางเนวิเกชันก็ควรวางให้สม่ำเสมอ เช่น อยู่ตำแหน่งบนสุดของทุกหน้า เป็นต้น ซึ่งถ้าจะให้ดีเมื่อมีเนวิเกชันที่เป็นกราฟิกก็ควรเพิ่มระบบเนวิเกชันที่เป็นตัวอักษรไว้ส่วนล่างด้วย เพื่อช่วยอำนวยความสะดวกให้กับผู้ใช้ที่ยกเลิกการแสดงผลภาพกราฟิกบนเว็บเบราว์เซอร์

6) คุณภาพของสิ่งที่ปรากฏให้เห็นในเว็บไซต์ (Visual Appeal) ลักษณะที่น่าสนใจของเว็บไซต์นั้น ขึ้นอยู่กับความชอบส่วนบุคคลเป็นสำคัญ แต่โดยรวมแล้วก็สามารถสรุปได้ว่าเว็บไซต์ที่น่าสนใจนั้นส่วนประกอบต่าง ๆ ควรมีคุณภาพ เช่น กราฟิกควรสมบูรณ์ไม่มีรอยหรือขอบขั้วกันให้เห็น ชนิดตัวอักษรอ่านง่ายสบายตา มีการเลือกใช้โทนสีที่เข้ากันอย่างสวยงาม เป็นต้น

8) ความคงที่ในการออกแบบ (Design Stability) ถ้าต้องการให้ผู้ใช้ใช้งานรู้สึก ว่าเว็บไซต์มีคุณภาพ ถูกต้อง และเชื่อถือได้ ควรให้ความสำคัญกับการออกแบบเว็บไซต์เป็น อย่างมาก ต้องออกแบบวางแผนและเรียบเรียงเนื้อหาอย่างรอบคอบ ถ้าเว็บที่จัดทำขึ้น

อย่าง ลวก ๆ ไม่มีมาตรฐานการออกแบบและระบบการจัดการข้อมูล ถ้ามีปัญหามากขึ้นอาจส่งผลให้ เกิดปัญหาและทำให้ผู้ใช้หมดความเชื่อถือ

9) ความคงที่ของการทำงาน (Function Stability) ระบบการทำงานต่างๆ ในเว็บไซต์ควรมีความถูกต้องแน่นอน ซึ่งต้องได้รับการออกแบบสร้างสรรค์และตรวจสอบอยู่เสมอ ตัวอย่างเช่น ลิงค์ต่าง ๆ ในเว็บไซต์ ต้องตรวจสอบว่ายังสามารถลิงค์ข้อมูลได้ถูกต้องหรือไม่ เพราะเว็บไซต์อื่นอาจมีการเปลี่ยนแปลงได้ตลอดเวลา ปัญหาที่เกิดจากลิงค์ ก็คือ ลิงค์ขาด ซึ่งพบได้บ่อยเป็นปัญหาที่สร้างความรำคาญกับผู้ใช้เป็นอย่างมาก

2.2.3.2 รูปแบบโครงสร้างของเว็บไซต์

การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบ ซึ่งก็ ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างแบบเรียงลำดับ จะเป็นโครงสร้างแบบธรรมดาที่นิยมใช้งาน กันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิงค์เนื้อหา ก็จะลิงค์ไปที่ละหน้า มีทิศทางการเข้าสู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ปุ่มเดินหน้า-ถอยหลังในการกำหนดทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับก็มีข้อเสีย คือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหาเพราะไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาด้วยตัวเองได้



ภาพที่ 2.10 แสดงโครงสร้างแบบเรียงลำดับ

(ที่มา : <https://sites.google.com/site/kruousaratwebsite/hlak-kar-xxkbaeb-websit/khorngsrang-khxng-websit>)

2) โครงสร้างแบบลำดับชั้น นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่าง ๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆ และมีการนำเสนอรายละเอียดย่อย ๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความเข้าใจกับโครงสร้างเนื้อหา

ได้ง่ายขึ้น โดยจะมีโฮมเพจเป็นจุดเริ่มต้น และจุดร่วมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง



ภาพที่ 2.11 แสดงโครงสร้างแบบลำดับชั้น

(ที่มา : <https://sites.google.com/site/kruousaratwebsite/hlak-kar-xxkbaeb-website/khorngsrang-khxng-website>)

3) โครงสร้างแบบตาราง เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีความยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่าง ๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งาน สามารถเปลี่ยนทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลา แถมยังทำให้เว็บไซต์มีความทันสมัยขึ้น



ภาพที่ 2.12 แสดงโครงสร้างแบบตาราง

(ที่มา : <https://sites.google.com/site/kruousaratwebsite/hlak-kar-xxkbaeb-website/khorngsrang-khxng-website>)

4) โครงสร้างแบบใยแมงมุม เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถ

เข้าถึงหน้าเว็บเพจต่าง ๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2.13 แสดงโครงสร้างแบบใยแมงมุม

(ที่มา : <https://sites.google.com/site/kruousaratwebsite/hlak-kar-xxkbaeb-website/khorngsrang-khxng-website>)

2.2.3.3 ส่วนประกอบสำคัญของหน้าเว็บเพจ บนหน้าเว็บเพจ จะมีส่วนประกอบสำคัญที่จำเป็นต้องมีอยู่ 3 ส่วน ได้แก่

1) ส่วนหัวของหน้า (Header) อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซด์ต่อไป ซึ่งส่วนใหญ่มักจะมีการใส่ภาพกราฟฟิกให้ดูสวยงาม สิ่งสำคัญหลักๆ เลย ก็คือ โลโก้ ชื่อเว็บไซด์และ เมนูหลักที่สามารถลิงค์ไปยังเนื้อหาในหน้าเว็บเพจต่าง ๆ ได้

2) ส่วนของเนื้อหา (Body) อยู่บริเวณตอนกลางของหน้าเว็บ โดยจะแสดง ข้อมูลเกี่ยวกับเนื้อหาบนเว็บแบบคร่าวๆ ซึ่งก็จะมีข้อความ กราฟฟิก ตารางข้อมูลหรือวิดีโอ ประกอบอยู่ และหากมีเมนูแบบเฉพาะกลุ่มก็จะถูกจัดไว้ในหน้านี้เช่นกัน และที่สำคัญเนื้อหาในส่วนนี้ควรจะมีคามกระชับ เข้าใจง่าย มีการใช้รูปแบบตัวอักษรแบบเรียบง่ายและเป็นระเบียบ

3) ส่วนท้ายของหน้า (Footer) อยู่ล่างสุดของหน้าเว็บ ซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่าง ๆ เพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูล เจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่าง ๆ เกี่ยวกับการใช้งานเว็บไซด์อย่างถูกต้อง เป็นต้น

2.2.3.4 วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์

การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสี สามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ ใช้จึงต้องมีความ

สอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บ ว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไรต่อเนื้อหาที่ได้ อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่ง ออกเป็น 3 กลุ่มดังต่อไปนี้

1) สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอดภัยและกระตุ้น ความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2) สีโทนเย็น (Cool Colors) เป็นสีแห่งความสุภาพและความอ่อนโยน ทำให้ ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้โน้มน้าวจากในระยะไกลได้ อีกด้วย

3) สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

สีสามารถสื่อถึงอารมณ์ความรู้สึกและสื่อความหมายของเว็บไซต์นั้น ๆ ได้ อย่างชัดเจน ดังนั้นหากเลือกใช้สีไม่เหมาะสมกับเนื้อหาหรือจุดประสงค์ของเว็บ ก็จะทำให้เว็บดู ไม่น่าสนใจ ขาดความน่าเชื่อถือและทำให้ผู้ที่เคยเข้ามาใช้บริการไม่คิดจะกลับมาใช้บริการอีก

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

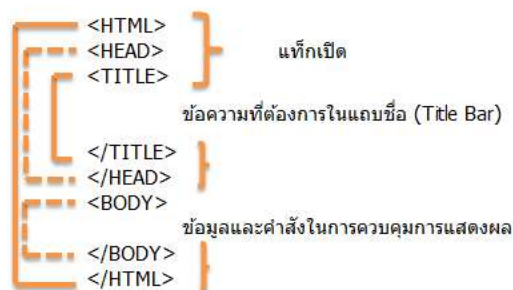
ภาษา HTML (Hyper Text Markup Language) เป็นภาษาที่ใช้ในการเขียนโปรแกรม ข้อมูลที่ใช้แสดงผลบนเครือข่ายอินเทอร์เน็ตในลักษณะของข้อความ รูปภาพ เสียง และ ภาพเคลื่อนไหวต่าง ๆ ภาษา HTML เป็นภาษาที่ง่ายต่อการเรียนรู้ สามารถกำหนดรูปแบบและ โครงสร้างได้ง่าย ทำให้ได้รับความนิยม และมีการพัฒนาอย่างต่อเนื่องเพื่อให้ใช้งานง่ายขึ้น และตอบสนองต่องานด้านกราฟิกมากยิ่งขึ้น และสนับสนุนการแสดงผลในเว็บเบราว์เซอร์ มากมาย และบันทึกในรูปแบบของไฟล์นามสกุล html หรือ htm

(Tag) เป็น คำสั่งหลักของ HTML แทบจะพูดได้ว่า ทุกสิ่งทุกอย่างของ HTML จะ ขึ้นอยู่กับ tag ทั้งนี้ ไม่ว่าจะเป็น การเน้นข้อความ การแสดงภาพประกอบ หรือการสร้างจุด เชื่อมโยง (link) โปรแกรมเบราว์เซอร์จะตีพิมพ์เอกสารออกมาในรูปแบบใด ก็โดยคำสั่ง tag ทั้งสิ้น

2.2.4.1 ไวยากรณ์ของ HTML จะแบ่งออกเป็น 2 ส่วน คือ

1) ส่วนของคำสั่ง (Tag) เป็นส่วนที่กำหนดรูปแบบของข้อความที่แสดง ซึ่ง เราเรียกว่า Tag โดยจะอยู่ในเครื่องหมาย < ... >

2) ส่วนของบทความทั่ว ๆ ไป เป็นส่วนของข้อความที่เราต้องการแสดงผล ตัวอย่างการใช้งานภาษา HTML



ภาพที่ 2.14 ไวยากรณ์ของ HTML

(ที่มา : <https://sites.google.com/site/kanpattanawebdouypasapeaspe/bth-reiyn/bth-thi-1-thakhwam-rucak-kab-php/xngkh-prakxb-khxng-kar-kheiyn-skhipt>)

<HTML>.....</HTML>

คำสั่ง <HTML> เป็นคำสั่งเริ่มต้นในการเขียนโปรแกรมและคำสั่ง </HTML> เป็นการสิ้นสุดโปรแกรม HTML คำสั่งนี้จะไม่แสดงผลในโปรแกรมเว็บเบราว์เซอร์ แต่ต้องเขียนเพื่อให้เกิดความเป็นระบบของงาน และเพื่อจะให้เราทราบว่าเอกสารนี้เป็นเอกสารของภาษา HTML ส่วนหัวเรื่องเอกสารเว็บ (Head Section)

<HEAD>.....</HEAD>

Head Section เป็นส่วนที่ใช้อธิบายเกี่ยวกับข้อมูลเฉพาะของหน้านั้น ๆ เช่น ชื่อเรื่องของหน้าเว็บ (Title), ชื่อผู้จัดทำเว็บ(Author), คีย์เวิร์ดสำหรับการค้นหา (Keyword)

<TITLE>.....</TITLE >

ข้อความที่ใช้เป็น TITLE ไม่ควรพิมพ์เกิน 64 ตัวอักษร, ไม่ต้องใส่ลักษณะพิเศษ เช่น ตัวหนา เอียง หรือสี และควรใช้ภาษาที่มีความหมายครอบคลุมถึงเนื้อหาของเว็บเพจ นั้น หรือเป็นคำสำคัญในการค้นหา (Keyword)

<BODY>.....</BODY>

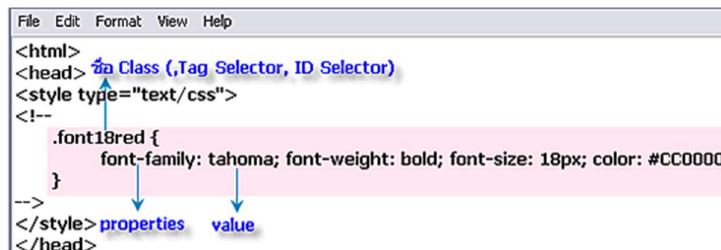
Body Section เป็นส่วนเนื้อหาหลักของหน้าเว็บ ซึ่งการแสดงผลจะต้องใช้ Tag จำนวนมาก ขึ้นอยู่กับลักษณะของข้อมูล เช่น ข้อความ, รูปภาพ, เสียง, วิดีโอ หรือไฟล์ต่างๆ ส่วนเนื้อหาเอกสารเว็บ เป็นส่วนการทำงานหลักของหน้าเว็บ ประกอบด้วย Tag มากมายตามลักษณะของข้อมูล ที่ต้องการนำเสนอ การป้อนคำสั่งในส่วนนี้ ไม่มีข้อจำกัดสามารถป้อนติดกันหรือ 1 บรรทัดต่อ 1 คำสั่งก็ได้ แต่ส่วนใหญ่จะยึดรูปแบบที่อ่านง่าย คือ การทำย่อหน้าในชุดคำสั่งที่เกี่ยวข้องกัน ทั้งนี้ให้ป้อนคำสั่งทั้งหมดภายใต้ Tag <BODY> ... </BODY>

2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

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

2.2.5.1 ไวยากรณ์ของ CSS

- 1) CSS จะประกอบไปด้วย ชื่อ Class หรือชื่อ ID หรือชื่อ Tag HTML ใน CSS จะเรียกว่า Selector ที่ใช้สำหรับการอ้างถึง
- 2) มีชื่อ Properties หรือชื่อคุณสมบัติ และตามด้วย Value เพื่อกำหนดค่า ให้แก่คุณสมบัตินั้น และระหว่างชื่อ Properties กับ Value จะถูกคั่นด้วยเครื่องหมาย colon (properties: value)
- 3) คุณสมบัติภายในจะมีได้มากกว่า 1 คุณสมบัติตามที่เรากำหนด และ properties ในแต่ละ properties จะถูกคั่นด้วยเครื่องหมาย semi-colon (;) 4) ต้องเขียนคุณสมบัติของ CSS นั้นไว้ในเครื่องหมายปีกกา {...} เท่านั้น



```
File Edit Format View Help
<html>
<head>
<style type="text/css">
<!--
    .font18red {
        font-family: tahoma; font-weight: bold; font-size: 18px; color: #CC0000;
    }
-->
</style>
</head>
```

ภาพที่ 2.15 ไวยากรณ์ของ CSS

(ที่มา : <https://www.dwthai.com/dwarticle/?t=6&aid=196&atitle=รูปแบบไวยากรณ์ของ+CSS+เบื้องต้น>)

2.2.6 ทฤษฎีเกี่ยวกับการ Visualization

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้งานสามารถเรียนรู้และจดจำข้อมูลผ่านการมองเห็นได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวว่า Visualization ก็คือการสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการ นำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

Visualization เป็นมากกว่าวิธีการทาง Computer Visualization เป็นการนำเสนอ ข้อมูลแบบหนึ่งที่ทำให้การแสดงข้อมูลให้อยู่ในรูปแบบของ Visual Form ซึ่งอาจจะเป็นการแสดง ในรูปแบบของรูปภาพ, กราฟ หรือแผนภาพ ซึ่งผลลัพธ์ของการทำ Visualization คือ การนำข้อมูลต่าง ๆ ที่ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้งานสามารถสัมผัสได้ซึ่งกระบวนการต่าง ๆ ในการแสดงหรือนำข้อมูลต่าง ๆ ออกมา อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้งานไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือใช้ในการวิเคราะห์ข้อมูล โดยจุดมุ่งหมายของ Visualization ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ เพื่อช่วยในการลดช่องว่างระหว่างผู้ใช้งานและตัวข้อมูล และยังช่วยผู้ใช้งานสามารถเรียนรู้ ได้ อย่างมีประสิทธิภาพ

2.2.6.1 วิธีการ Visualization

การใช้ Visual เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้งานจะทำขั้นตอนหลักๆอยู่ 3 ขั้นตอน คือ Overview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้งานต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้งานก็จะทำการตัดสินใจ

เลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้ก็จะทำการเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็อาจอ้างหรือพัฒนาจากขั้นตอนเหล่านี้ซึ่ง Visualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ใช้งานต้องการโดยอาจจะใช้หลายๆวิธีการรวมกันเพื่อให้ได้ข้อมูลที่ผู้ใช้งานต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามารถนำมาผ่านกระบวนการของ Visualization มีลักษณะต่าง ๆ มากมายดังนี้ ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง, ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi Dimensional Data ได้แก่ Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อข่าวต่าง ๆ และ Web Document, Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document ,Algorithms และ Software ได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลาย

2.2.7 ทฤษฎีเกี่ยวกับกฎความสัมพันธ์ ด้วย Apriori Algorithm

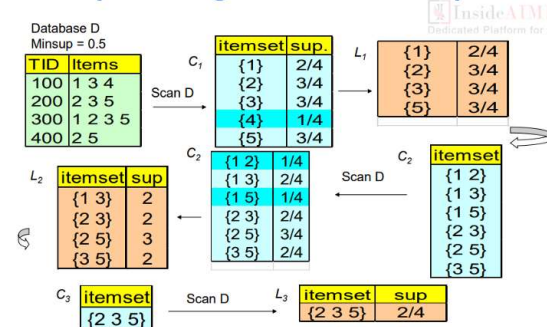
ความสัมพันธ์ (Association) โดยหลักการทำงาน คือการหาความสัมพันธ์ของข้อมูลภายในกลุ่มข้อมูล เพื่อใช้ลักษณะของข้อมูลหนึ่งในการบอกถึงลักษณะที่จะเกิดขึ้นกับอีกตัวหนึ่ง ซึ่งอาจจะเป็นการหาความสัมพันธ์ของข้อมูล ในกลุ่มเดียวกัน ตัวอย่างเช่น การระบุว่าในกลุ่มของนักเรียน ที่สมัครเรียนสาขาวิชาพยาบาลศาสตร์แล้วจะเป็นเพศหญิงนั้น มีโอกาสเกิดขึ้นร่วมกัน ผลลัพธ์ที่ได้จากการค้นหาความสัมพันธ์นั้นสามารถนำไปใช้ในการวิเคราะห์สนับสนุนหรือ พัฒนางานที่เกี่ยวข้องกับฐานข้อมูลนั้น ได้อย่างมีประสิทธิภาพมากยิ่งขึ้นซึ่งวิธีที่ได้รับค่านิยมและเป็นที่ยอมรับ คือการค้นหาหาความสัมพันธ์ที่ปรากฏข้อมูลร่วมกันบ่อย

การค้นหาหาความสัมพันธ์ของข้อมูลในฐาน ข้อมูลได้พัฒนาขึ้นครั้งแรกโดยนักวิจัยจากศูนย์วิจัย IBM (International Business Machines Corporation) ประเทศสหรัฐอเมริกา มีจุดประสงค์เพื่อค้นหาความสัมพันธ์ที่น่าสนใจ ซึ่งซ่อนอยู่ในข้อมูลการซื้อสินค้าของลูกค้าว่าจะซื้อสินค้าใดบ้างร่วมกันในตะกร้ารถเข็นในห้างสรรพสินค้า (Market Basket Analysis) เพื่อให้ทราบถึงพฤติกรรมการซื้อสินค้าของลูกค้า เช่นเมื่อลูกค้าซื้อนมแล้วจะซื้อขนมปังด้วย การค้นหา กฎความสัมพันธ์มีขั้นตอนที่สำคัญที่สุดคือการค้นหาหา กลุ่ม ข้อมูลที่ปรากฏร่วมกันบ่อยซึ่งจะมีขั้นตอนกระบวนการการทำงานนานที่สุด ดังนั้นในการเลือกขั้นตอนวิธีในการค้นหาหาความสัมพันธ์จึงเป็นสิ่งสำคัญเพื่อให้เหมาะสมกับลักษณะของ ข้อมูล เพื่อให้ลดระยะเวลา และเนื้อที่หน่วยความจำในการประมวลผลการทำงาน กระบวนการค้นหาหาความสัมพันธ์ มีขั้นตอนการค้นหาหาข้อมูลปรากฏร่วมกันบ่อยทั้งหมด กลุ่มข้อมูลเหล่านั้นจะต้องมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่ผู้ใช้กำหนดจึงจะถือว่าเป็นกลุ่มข้อมูลที่

ปรากฏร่วมกันบ่อย และขั้นตอนการนำกลุ่มข้อมูลที่ปรากฏ บ่อยมาสร้างกฎความสัมพันธ์ซึ่งกฎความสัมพันธ์จะเป็นที่ ยอมรับได้หากตรวจสอบค่าความเชื่อมั่นของกฎนั้นมีค่า มากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่ผู้ใช้กำหนด

ขั้นตอนวิธี Apriori (Apriori Algorithm) เป็นขั้นตอนวิธีที่ได้รับการยอมรับและเป็นที่ยอมรับ ในการค้นหากฎ ความสัมพันธ์โดยเซตที่มีความถี่มากกว่าค่าสนับสนุนขั้นต่ำ เซตย่อยจะมีความถี่ที่มากกว่าด้วยและสามารถตัดเซตรายการ ที่มีความถี่ต่ำออก ขั้นตอนวิธีในการค้นหาเซต รายการที่ปรากฏร่วมกันบ่อยและนำ มาสร้างกฎความสัมพันธ์ข้อดีของขั้นตอนวิธี Apriori คือสามารถทำงานได้ดีหากกำหนดค่า สนับสนุนขั้นต่ำมีค่ามาก มีขนาดของฐานข้อมูลขนาดเล็ก และมีจำนวนของเซตรายการน้อย ส่วนข้อเสียคือต้องอ่านข้อมูลจากฐานข้อมูลหลายครั้ง

The Apriori Algorithm — Example



ภาพที่ 2.16 ขั้นตอนวิธี Apriori

(ที่มา : <https://ichi.pro/th/apriori-method-in-machine-learning-khux-xari-6468477945009%206>)

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม Rapid Miner Studio

Rapidminer คือซอฟต์แวร์ Data Science ใช้สำหรับการเตรียมข้อมูล การเรียนรู้เครื่อง การเรียนรู้ลึก การทำเหมืองข้อความ และการวิเคราะห์การทำนาย (Predictive analysis) เป็นซอฟต์แวร์ที่ช่วยในการจัดส่งข้อมูล และลดข้อผิดพลาดจนแทบจะไม่จำเป็นต้องเขียนโค้ดเพิ่ม แต่ที่ทำให้เป็นเครื่องมือที่เหล่า Data Scientist นิยมเลือกใช้เป็นเพราะว่าตัว Rapidminer มีขั้นตอนพร้อมสำหรับการทำ Data mining (ขุดข้อมูล) และ Machine learning ซึ่งรวมไปถึงการโหลดและการแปลงข้อมูล(ETL) การประมวลผลล่วงหน้าและการวาดภาพจาก

ข้อมูล การวิเคราะห์เชิงพยากรณ์และการสร้างแบบจำลองทางสถิติ การประเมินผลและการปรับใช้ ต่างๆ ล้วนเป็นสิ่งที่เหล่า Data Scientist จำเป็นต้องทำในการเข้าใจข้อมูลมากขึ้น

แต่ที่บอกว่า Rapidminer ใช้งานง่ายสำหรับผู้ใช้งานที่ไม่มีพื้นฐานเลยนั้นเป็นเพราะว่า หน้าตาของซอฟต์แวร์เข้าใจง่าย และการทำงานนั้นเป็นแบบ Drag and drop สำหรับการวิเคราะห์ในชั้นเบสิค Rapidminer ให้ยูสเซอร์ได้ออกแบบ Workflow ในการวิเคราะห์ข้อมูลในรูปแบบของรูปภาพ ซึ่งทุกอย่างจะเกิดขึ้นในหน้า Design View

เพื่อให้ผู้อ่านเข้าใจการใช้งานซอฟต์แวร์ตัวนี้ง่ายขึ้น Achieve Plus จะอธิบายคำศัพท์ที่จะเห็นบ่อยในซอฟต์แวร์ เพื่อว่าผู้อ่านจะมีโอกาสได้นำความรู้ตรงนี้ไปต่อยอดได้

1. Repository

อธิบายอย่างเข้าใจง่ายที่สุดก็คือ เปรียบเสมือนกับ Folder ในคอมพิวเตอร์ของเรา เป็นพื้นที่จัดเก็บข้อมูล Process และผลลัพธ์

2. Operators

เป็นองค์ประกอบสำคัญ โดยจะเป็นตัวการในการทำงานต่อจาก input ที่ได้รับ ซึ่งเหล่า Operators มีการทำงานที่มากกว่าหลายร้อยเลยทีเดียว ตั้งแต่การทำความสะอาดข้อมูล Cleansing การทำโมเดล Modeling รวมไปถึงการ Blend ข้อมูลเข้าด้วยกัน

3. Ports

ตามความหมายของคำศัพท์เลย เป็นพอร์ตที่เป็นส่วนของ output ของ Operators การเชื่อมต่อ Port เหล่านี้ต้องมั่นใจก่อนว่า output กับ input ของ Operator อีกตัวหนึ่งเข้ากันได้

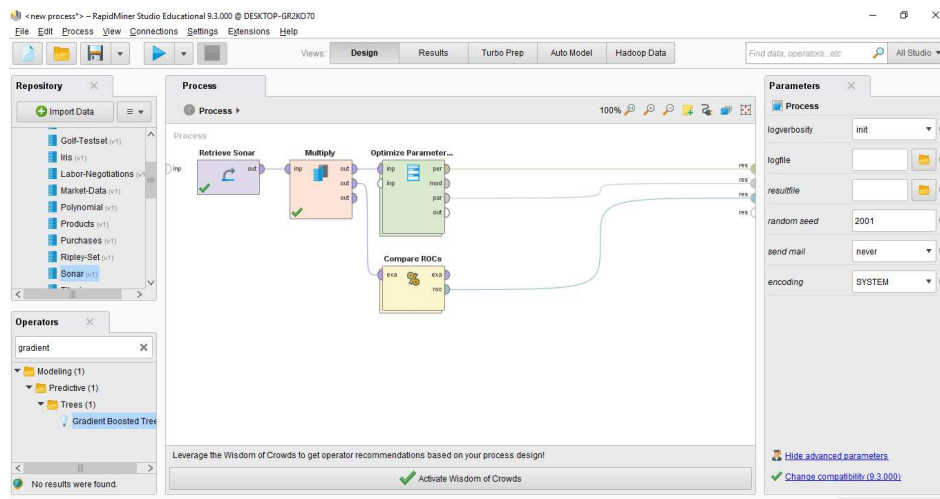
4. Process

เข้าใจอีกอย่างหนึ่งว่า Flow หรือ Pipeline ก็ได้ Process คือ workflow ที่เชื่อมต่อ Operators ต่างๆครบ และสามารถนำไปวิเคราะห์ข้อมูลได้ ตัวอย่าง Process ที่เข้าใจง่ายคือ

5. Parameters

พารามิเตอร์นี้เปรียบเสมือน การตั้งค่าการทำงานของแต่ละ Operators ซึ่งยูสเซอร์จะสามารถกำหนดค่าของแต่ละ Operators ที่แตกต่างกัน

เมื่อเข้าใจคำศัพท์คร่าวๆของซอฟต์แวร์ Rapidminer แล้ว มั่นใจว่าการจ่อยอดหลังจากนี้จะง่ายขึ้นแน่นอน หากว่าสนใจลงลึกการใช้งานของตัวซอฟต์แวร์ สามารถศึกษาด้วยตนเองผ่านเว็บไซต์ของซอฟต์แวร์เลย แต่หากว่าต้องการพัฒนาความรู้ด้าน Data Science โดยได้เรียนรู้จากการใช้งานจริง ๆ Data Science Pathway จาก Achieve Plus ได้รวบรวมเนื้อหาที่เน้นการใช้งาน เข้าใจ และที่สำคัญ สอนการใช้งานแต่ละเครื่องมือจริง ๆ



ภาพที่ 2.17 ภาพโปรแกรม Rapidminer

(ที่มา : <https://medium.com/@jamesjetana/การทำ-automated-optimization-models-ใน-rapidminer-studio-9-3-กัน-d3840977412>)

2.4 วรรณกรรมที่เกี่ยวข้อง

วรุฑ์ ทัพพงษ์ (2559) งานวิจัยเรื่อง “การจัดกิจกรรมการส่งเสริมการขายสำหรับร้านอาหารด้วยขั้นตอนวิธีอะเพียวรี” แนวคิดในการพัฒนาซอฟต์แวร์ที่จะช่วยทำการวิเคราะห์การจัดชุดอาหารโดยใช้หลักการของการหาความสัมพันธ์มาช่วยในการสร้างความสัมพันธ์เพื่อช่วยในการสนับสนุนการตัดสินใจของผู้บริหารให้มีทางเลือกมากขึ้นในการจัดกิจกรรมส่งเสริมการขาย ได้นำเทคนิคอะเพียวรีมาใช้ในการจัดชุดอาหารและนำการคำนวณหาค่าความสัมพันธ์ เทคนิคอะเพียวรีมีความเป็นไปได้ที่จะนำมาวิเคราะห์ข้อมูลการขายอาหารเพื่อจัดชุดอาหารโดยใช้ข้อมูลยอดขายแต่ละรายการอาหารที่ได้ในแต่ละวันเป็นข้อมูลในการวิเคราะห์และหาความสัมพันธ์ร่วมของกลุ่มข้อมูลการขายอาหารเพื่อแสดงทางเลือกที่จะนำมาจัดกิจกรรมทำให้ร้านอาหารมีทางเลือกที่หลากหลายมากยิ่งขึ้นเพิ่มความสะดวกรวดเร็วในการจัดเก็บค้นหาข้อมูลอาหาร ยอดขาย และลดปัญหาด้านการสูญหายของเอกสาร

สุทธิดา ร่มสุข (2560) งานวิจัยเรื่อง “ปัจจัยที่มีผลในการเลือกใช้บริการร้านกาแฟ” กาแฟเป็นเครื่องดื่มที่ได้รับความนิยมอย่างมากในหมู่ผู้บริโภคชาวไทย สะท้อนให้เห็นพฤติกรรมการดื่มกาแฟของผู้บริโภคสมัยใหม่ที่มีความหลากหลายมากยิ่งขึ้น ไม่ได้เพียงแค่การเป็นเครื่องดื่มเพื่อคลายความง่วงเหมือนในอดีตเท่านั้น แต่ร้านกาแฟสดทั้งขนาดใหญ่และขนาด

ใหญ่รวมไปถึงร้านกาแฟ พรีเมียมทั้งในและต่างประเทศที่มีเพิ่มมากขึ้นอย่างต่อเนื่องในช่วงระยะเวลา 4-5 ปีที่ผ่านมา ทำให้ร้านกาแฟเข้ามาช่วงชิงส่วนแบ่งทางการตลาด ส่งผลให้ตลาดกาแฟเริ่มมีการเปลี่ยนแปลงมากขึ้น กล่าวคือ มีการเข้ามาของผู้ประกอบการรายใหม่และนำเสนอผลิตภัณฑ์ใหม่ๆร้านกาแฟเป็นหนึ่งในธุรกิจที่ได้รับความสนใจจากผู้ประกอบการตั้งแต่รายเล็กไปจนถึงรายใหญ่ เห็นได้จากรายงานของ Euromonitor ที่แสดงตัวเลขของตลาดร้านกาแฟในประเทศไทยโดยมีมูลค่าสูงถึง 21,220 ล้านบาท ในปี 2560 นอกเหนือจากนี้ ปริมาณการบริโภคกาแฟในประเทศที่เพิ่ม ขึ้นอย่างต่อเนื่องรวมถึงปริมาณเมล็ดกาแฟชนิดที่ยังไม่ได้คั่ว ที่ถูกนำเข้ามาจากต่างประเทศในปี 2560 นอกเหนือจากปริมาณการบริโภคกาแฟของชาวไทยที่เพิ่มขึ้นแล้วยังมีปัจจัยบางอย่างที่ส่งผลต่อตลาด ร้านกาแฟขยายตัวขึ้น ได้แก่ รายได้และจำนวนชนชั้นกลางที่เพิ่มขึ้น โดยอาจมีผลให้ผู้บริโภคมีอำนาจการใช้จ่ายเพิ่ม ขึ้น ในส่วนของผู้ประกอบการก็ได้มีการตอบสนองกับจำนวนอุปสงค์ที่มีแนวโน้มเพิ่มขึ้น โดยมีการเพิ่มจำนวนสาขา ร้านกาแฟที่จะเปิดตัวจากแผนธุรกิจในหลาย ๆ แพรนไชส์โดยปัจจัยเหล่านี้สะท้อนให้เห็นถึงการคาดการณ์ในเชิงบวกของผู้ประกอบการทั้งชาวไทยและชาวต่างชาติที่ลงทุนในธุรกิจกาแฟในประเทศไทย อย่างไรก็ตาม แม้ว่าธุรกิจร้านกาแฟจะมีแรงสนับสนุนจากทั้งฝั่งอุปสงค์และอุปทานแต่การลงทุนในธุรกิจชนิดนี้ก็ ยังมีความท้าทายทั้งทางด้านการเงินและทางด้านธุรกิจอยู่ไม่น้อยโดยเฉพาะอย่างยิ่งการแข่งขันที่สูงมากในร้านกาแฟ

วรัญญิววิญญ์ แจ้งพลอย (2559) งานวิจัยเรื่อง“กลยุทธ์การสร้างรายได้เปรียบทางการแข่งขันในการดำเนินธุรกิจร้านกาแฟ” กาแฟ เป็นเครื่องดื่มที่มีกลิ่นและรสเป็นเอกลักษณ์ เป็นที่ชื่นชอบของคนทั่วโลกมาอย่างยาวนาน และกาแฟถือเป็นหนึ่งในเครื่องดื่มที่ได้รับความนิยมอย่างมากในประเทศไทย โดยเฉพาะในกลุ่มคนรุ่นใหม่วัยทำงานและผู้คนที่ใช้ชีวิตอยู่ในสังคมเมืองที่หันมาบริโภคกาแฟสดกันมากขึ้นแต่อย่างไรก็ตาม อัตราการบริโภคกาแฟของคนไทยยังอยู่ในเกณฑ์ต่ำ คือ น้อยกว่า 0.5 กิโลกรัมต่อคนต่อปีหรือคิดเป็นเพียง 130-150 ถ้วยต่อคนต่อปี ซึ่งคาดว่าคนไทยเพียงร้อยละ 30 ของประชากรทั้งหมดที่บริโภคกาแฟเท่านั้น ซึ่งยังมีปริมาณน้อยเมื่อเทียบกับประเทศในแถบเอเชีย อย่างเช่น ญี่ปุ่นดื่มกาแฟเฉลี่ย 500 แก้วต่อคนต่อปี หรืออเมริกาที่ดื่มกาแฟเฉลี่ย 700 แก้วต่อคนต่อปี ส่งผลให้อัตราการเติบโตของธุรกิจกาแฟภายในประเทศยังมีแนวโน้มสูงขึ้นไปอย่างต่อเนื่องทำให้ธุรกิจร้านกาแฟรายใหญ่ๆ จากต่างประเทศเข้ามาลงทุนและกระตุ้นตลาดกาแฟให้เกิดความคึกคักและตื่นตัวเป็นอย่างมาก ส่งผลให้อัตราการเติบโตของธุรกิจร้านกาแฟทั้งขนาดเล็กและขนาดใหญ่ในประเทศเพิ่มขึ้นอย่างต่อเนื่อง

เอกสิทธิ์ พชรวงศ์ศักดิ์ดา (2557) งานวิจัยเรื่อง “การวิเคราะห์ข้อมูลด้วย Data Mining” ในยุคปัจจุบันได้ก้าวเข้าไปสู่ยุคที่เรียกว่า Big Data หรือ ข้อมูลมหาศาล เนื่องจากในแต่ละวันมีข้อมูลเกิดขึ้นมากมาย เช่น ข้อมูลสมาชิกของ Facebook ข้อมูลการซื้อขายสินค้าในซูเปอร์มาร์เกตต่าง ๆ และเพื่อให้เกิดประโยชน์มากที่สุดจำเป็นต้องนำข้อมูลมหาศาลเหล่านี้มาทำการวิเคราะห์ (Analyze) ซึ่งเทคนิคหนึ่งที่ได้รับการนิยมนอย่างสูงในปัจจุบันคือ เทคนิค Data Mining ซึ่งเป็นเทคนิคที่ค้นหาความสัมพันธ์ในข้อมูล เช่น ถ้าลูกค้าซื้อเบียร์แล้วลูกค้าจะซื้อผ้าอ้อมรวมไปด้วย หรือถ้าเรากด Like หน้า Facebook Page เราจะเห็นว่า Facebook มีระบบแนะนำ Page อื่น ๆ ที่เกี่ยวข้องมาให้ด้วย หรือ การสร้างโมเดลเพื่อทำนายสิ่งที่จะเกิดขึ้นในอนาคต เช่น ทำนายยอดขายในไตรมาสถัดไป หรือ การทำนายว่าพนักงานคนไหนที่จะลาออกจากบริษัทในช่วง 3 เดือนข้างหน้า ตัวอย่างเหล่านี้ล้วนเป็นผลมาจากการวิเคราะห์ข้อมูลทางด้าน Data Mining โดยใช้ซอฟต์แวร์ที่ช่วยให้ทำการวิเคราะห์ได้ง่ายขึ้นแต่ซอฟต์แวร์ส่วนใหญ่จะเป็นซอฟต์แวร์เชิงพาณิชย์ (Commercial Software) เช่น SAS Enterprise Miner หรือ IBM Intelligent Miner การลงทุนซื้อซอฟต์แวร์เชิงธุรกิจเหล่านี้มาใช้งานอาจจะไม่คุ้มค่าในการลงทุนสำหรับผู้ประกอบการวิสาหกิจขนาดกลางและขนาดย่อม (SMEs) หรืออาจารย์ นักวิจัยและนักศึกษาในมหาวิทยาลัยต่าง ๆ ดังนั้นวิธีการหนึ่งที่จะทำให้สามารถวิเคราะห์ข้อมูลเหล่านี้ได้คือการใช้ซอฟต์แวร์เวอร์ชันฟรี (Free Version) ที่สามารถดาวน์โหลด (Download) มาใช้งานได้โดยไม่เสียค่าใช้จ่าย เช่น RapidMiner Studio Educational RapidMiner Studio แรกเริ่มพัฒนาขึ้นจากบริษัทที่ชื่อว่า Rapid-I ในประเทศเยอรมนี และเมื่อช่วงปลายปีพ.ศ. 2556 ที่ผ่านมามีได้รับทุนจากนักลงทุนในประเทศสหรัฐอเมริกาจึงเปลี่ยนชื่อบริษัทจาก Rapid-I เป็น RapidMiner แทนและย้ายสำนักงานใหญ่มาอยู่ประเทศสหรัฐอเมริกา สามารถดาวน์โหลดซอฟต์แวร์ RapidMiner Studio ซึ่งเป็นเวอร์ชันปัจจุบันได้จากเว็บไซต์

ธิดา สมณะ (2563) งานวิจัยเรื่อง “การวิเคราะห์ข้อมูลทางธุรกิจ และการใช้โปรแกรมประยุกต์ทางธุรกิจ” การวิเคราะห์ทางธุรกิจ (Business Analytics: BA) หมายถึง ทักษะทางด้านเทคโนโลยีเพื่อการปฏิบัติสำหรับการสำรวจ การวิเคราะห์ การออกแบบ และการตรวจตราอย่างต่อเนื่องนอกจากนั้น การวิเคราะห์ทางธุรกิจ ยังเป็นกระบวนการตรวจสอบประสิทธิภาพทางธุรกิจ ในช่วงเวลาที่ผ่านมา เพื่อให้ได้ข้อมูลเชิงลึก และผลักดันการวางแผนธุรกิจให้ประสบความสำเร็จการวิเคราะห์ธุรกิจมุ่งเน้นไปที่การพัฒนาความเข้าใจเดิม และความเข้าใจใหม่เกี่ยวกับประสิทธิภาพทางธุรกิจ โดยการใช้ข้อมูลและวิธีการทางสถิติ เพื่อเป็นการพัฒนาระบบธุรกิจอัจฉริยะแบบดั้งเดิมที่มุ่งเน้นไปที่การใช้ชุดของตัวชี้วัดที่สอดคล้องกัน เพื่อใช้วัดประสิทธิภาพการดำเนินงานธุรกิจที่ผ่านมา และเป็นแนวทางการวางแผนธุรกิจเดิม ซึ่งยังขึ้นอยู่กับ

ข้อมูลและวิธีการทางสถิติพื้นฐาน การวิเคราะห์ธุรกิจขึ้นอยู่กับปริมาณข้อมูลที่เพียงพอ โดยที่ความยากลำบากในการประกันคุณภาพข้อมูล คือ การรวบรวม และคัดเลือกข้อมูลในระบบ และตัดสินใจเลือกใช้ข้อมูลก่อนหน้าการวิเคราะห์ที่ใช้การพิจารณาตามแบบแผน และความเป็นจริง ซึ่งในการพยากรณ์พฤติกรรมผู้บริโภค ตัวอย่างเช่น การตรวจสอบหน่วยขายรายไตรมาส รายเดือน รายปี การคาดคะเนแนวโน้มที่ต้องการจำเป็นต้องใช้ข้อมูลที่เก็บไว้ในปริมาณมาก คลังข้อมูลประเภทนี้ต้องการพื้นที่เก็บข้อมูลมากกว่าความเร็ว ขณะที่การวิเคราะห์ข้อมูลธุรกิจกลายเป็นเครื่องมือที่มีอิทธิพลต่อผลลัพธ์และการมีปฏิสัมพันธ์กับลูกค้า ตัวอย่างเช่น เมื่อลูกค้าบางประเภทกำลังพิจารณาการสั่งซื้อสินค้า องค์กรจำเป็นต้องเรียกใช้งานการวิเคราะห์เพื่อให้สามารถปรับเปลี่ยนการขาย เพื่อดึงดูดลูกค้ารายนั้นให้ได้ หมายความว่า ข้อมูลทั้งหมดจะต้องตอบสนองอย่างรวดเร็ว เพื่อให้ได้ข้อมูลที่จำเป็นในแบบเรียลไทม์ (Real Time) อาจกล่าวได้ว่าความท้าทายที่แท้จริง คือ ทำอย่างไรให้ได้ผลลัพธ์ตามที่ต้องการ

2.5 บทสรุป

ข้อมูลได้เลือกใช้การทำเหมืองข้อมูลแบบ Association rule ด้วยเทคนิค Apriori เพื่อสร้างชุดข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม RapidMiner และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผล แบบ Visualization ในรูปแบบของภาพโดยใช้ โปรแกรม Tableau Public เผยแพร่บน Web Browser ที่เป็นที่ยอมรับในยุคอินเทอร์เน็ตคือการ เผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา