

บทที่ 2

แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

โครงการเรื่อง การสร้างโมเดลวิเคราะห์ข้อมูลแนวโน้มที่นำไปสู่การฆ่าตัวตายของผู้ป่วยจิตเวชในจังหวัดเชียงใหม่ ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของ การสร้างโมเดลวิเคราะห์ข้อมูลแนวโน้มที่นำไปสู่การฆ่าตัวตายของผู้ป่วยจิตเวชในจังหวัดเชียงใหม่ และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับดังนี้

2.1 แนวคิด

- 2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)
- 2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ
- 2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

2.2 ทฤษฎี

- 2.2.1 การทำเหมืองข้อมูล (Data Mining)
- 2.2.2 หลักในการทำเหมืองข้อมูล
- 2.2.3 การจำแนกประเภทของข้อมูล (Classification)
- 2.2.4 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์
- 2.2.5 ทฤษฎีเกี่ยวกับภาวะซึมเศร้า

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

- 2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)
- 2.3.2 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการสร้างโมเดลวิเคราะห์ข้อมูลแนวโน้มที่นำไปสู่การฆ่าตัวตายของผู้ป่วยจิตเวชในจังหวัดเชียงใหม่ ทางผู้วิเคราะห์ข้อมูลได้ศึกษาหลักการ และทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

Data Analytics คือ การวิเคราะห์ข้อมูลที่มีอยู่ตั้งแต่ อดีต จนถึงปัจจุบัน เพื่อทำนายอนาคต ที่เป็นประโยชน์ในการพัฒนาการตลาด ให้ตรงใจลูกค้ามากยิ่งขึ้น Data Analytics เป็นเครื่องมือสำหรับธุรกิจ (Business Intelligence) เพราะว่าการที่บริษัทคุณไม่รู้ข้อมูล ก็เหมือนบริษัทคุณกำลังหาทาง โดยไม่มีจุดหมายปลายทาง ดังนั้นการทำ Data Analytics นี้ไม่จำเป็นต้องเป็นธุรกิจขนาดใหญ่เท่านั้น แต่ธุรกิจขนาดกลางและเล็กก็สามารถทำได้เช่นกัน สำหรับรูปแบบของการวิเคราะห์ข้อมูล (Data Analytics) สามารถแบ่งได้ดังนี้

1) การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive analytics) เป็นการวิเคราะห์เพื่อแสดงผลของรายการทางธุรกิจ เหตุการณ์ หรือกิจกรรมต่างๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้นในลักษณะที่ง่ายต่อการเข้าใจ หรือต่อการตัดสินใจ ตัวอย่างเช่น รายงานการขาย รายงานผล การดำเนินงาน

2) การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้น ปัจจัยต่างๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่างๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น ตัวอย่างเช่น ความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่จะช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

3) การวิเคราะห์แบบพยากรณ์ (Predictive analytics) เป็นการวิเคราะห์เพื่อพยากรณ์สิ่งที่กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้น โดยใช้ข้อมูลที่ได้เกิดขึ้นแล้วกับแบบจำลองทางสถิติ หรือ เทคโนโลยีปัญญาประดิษฐ์ต่างๆ (Artificial intelligence) ตัวอย่างเช่น การพยากรณ์ยอดขาย การพยากรณ์ผลประชามติ

4) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive analytics) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด เป็นทั้งการพยากรณ์สิ่งต่างๆ ที่จะเกิดขึ้น ข้อดี ข้อเสีย สาเหตุ และระยะเวลาของสิ่งที่เกิดขึ้น รวมถึงการให้คำแนะนำทางเลือกต่างๆ ที่มีอยู่ และผลของแต่ละทางเลือก (แอฟฟินีตี้ โชลูชั่น, 2565)

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

การประยุกต์ใช้ข้อมูลทางสถิติมีความสำคัญในหลายด้านของชีวิตและอุตสาหกรรมต่าง ๆ เพื่อให้เกิดประโยชน์ต่อการตัดสินใจ การวิเคราะห์ปัญหา และการพัฒนานโยบาย นี่คือบางแนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

1) การวิเคราะห์ข้อมูล การใช้สถิติเพื่อวิเคราะห์ข้อมูลชุดใหญ่เป็นเครื่องมือสำคัญในการหาความสัมพันธ์ แนวโน้ม และแบบแผนที่ซ่อนอยู่ในข้อมูล นี่เป็นสิ่งสำคัญในการทำความเข้าใจเกี่ยวกับปัญหาและตัดสินใจทางธุรกิจ วิทยาศาสตร์ และสาขาอื่น ๆ

2) การวางแผนและการตัดสินใจ การใช้ข้อมูลสถิติช่วยในการวางแผนกิจกรรมและการตัดสินใจ โดยมีข้อมูลที่เป็นหลักในการตัดสินใจ เช่น การตลาด การจัดการโครงการ และการวางแผนการเงิน

3) การควบคุมคุณภาพ: ในสาขาอุตสาหกรรมและการผลิต การใช้สถิติช่วยในการควบคุมคุณภาพของสินค้าหรือบริการเป็นสิ่งสำคัญ เช่น การตรวจสอบคุณภาพและการควบคุมกระบวนการผลิต

4) การวิเคราะห์การทดลอง: ในวิทยาศาสตร์และวิจัยการทดลอง สถิติมีบทบาทสำคัญในการวิเคราะห์ผลลัพธ์การทดลองเพื่อหาความหมายทางสถิติและการตรวจสอบสมมติฐาน

5) การวิเคราะห์การเงิน: สถิติมีความสำคัญในการวิเคราะห์ข้อมูลการเงิน เช่น การวิเคราะห์รายได้และรายจ่าย การวิเคราะห์ความเสี่ยงการลงทุน และการวิเคราะห์การเงินส่วนบุคคล

6) การคาดการณ์และการวางแผนอนาคต: การใช้สถิติเพื่อคาดการณ์และวางแผนในอนาคต ในการทำนายข้อมูลเช่นยอดขายในอนาคต การเจริญเติบโตของประชากร หรือความเสี่ยงทางการเงิน

7) การปรับปรุงกระบวนการ: การใช้ข้อมูลสถิติในการตรวจสอบและปรับปรุงกระบวนการทางธุรกิจ เพื่อเพิ่มประสิทธิภาพและลดความผิดพลาด

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

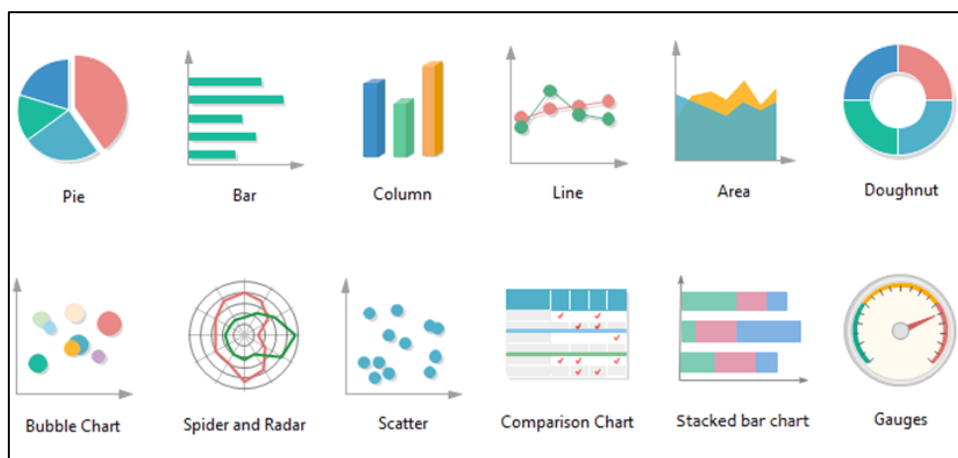
ความหมายของการแสดงผลข้อมูล

Data Visualization คือ การนำข้อมูลหรือ Data ที่ได้มาจากแหล่งข้อมูลต่างๆ มาวิเคราะห์ประมวลผลแล้วนำเสนอออกมาในรูปแบบที่มองเห็นและทำความเข้าใจได้ด้วยตา เช่น แผนภูมิ รูปภาพ แผนที่ กราฟแสดงเทรนด์ ตาราง วิดีโอ อินโฟกราฟิก (Infographic) แดชบอร์ด (dashboard) จุดประสงค์สำคัญของการทำ Data Visualization คือ การนำเสนอข้อมูลให้เข้าใจง่าย ผู้อ่านข้อมูลสามารถเข้าใจได้ทันทีว่าตัวชี้งาน (media) ต้องการสื่อสารอะไร ซึ่งจุดสำคัญของเนื้อหา และชี้ Insight ข้อเปรียบเทียบให้เห็นอย่างชัดเจน ช่วยให้สังเกตเห็นจุดที่น่าสนใจของข้อมูลได้ง่ายขึ้น

Data Visualization มีหลากหลายรูปแบบและไม่จำกัดว่าต้องใช้รูปแบบต่อไปนี้ในการนำเสนอข้อมูลเท่านั้น เพราะแต่ละรูปแบบก็มีฟังก์ชันเฉพาะของการนำเสนอข้อมูล บางรูปแบบใช้เปรียบเทียบข้อมูลแต่ละชุดได้ดี บางรูปแบบช่วยให้มองเห็นเทรนด์ได้ง่าย บางรูปแบบ

ช่วยให้ข้อมูลที่โกลดตัวให้เข้าใจได้ง่ายโดยการเปรียบเทียบให้สอดคล้องกับสิ่งที่คุ้นเคยในชีวิตประจำวัน (เพิร์สคราฟท์ ดิจิทัล โซลูชัน, 2563)

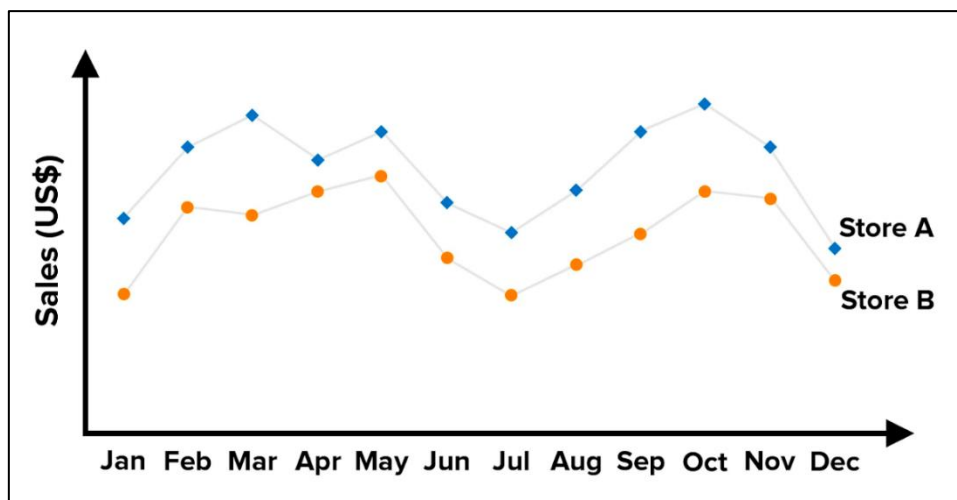
1) แผนภูมิ (Charts) ซึ่งเป็นรูปแบบที่น่าจะคุ้นเคยกันมากที่สุด และเป็นรูปแบบที่มีหลากหลายชนิดที่เหมาะสมกับการนำเสนอข้อมูลที่แตกต่างกันไปตามวัตถุประสงค์ เช่น Pie chart จะช่วยให้เราเห็นปริมาณความแตกต่างได้ชัดเจน, Comparison chart เหมาะสำหรับการเปรียบเทียบคุณสมบัติหลายๆ ข้อ, มาตรวัด (Gauges) จะช่วยให้เห็นความเข้มข้น ความรุนแรง หรือน้ำหนัก



ภาพที่ 2.1 แสดงตัวอย่างแผนภูมิรูปแบบต่าง ๆ

(ที่มา: medium.com/@Lynia_Li/)

2) กราฟ (Graphs) คือ subset หรือประเภทหนึ่งของแผนภูมิ โดยกราฟจะทำหน้าที่แสดงความสัมพันธ์ระหว่างข้อมูล 2 ตัวแปร ผ่านแกนแนวนอน (แกน X) และแกนแนวตั้ง (แกน Y) ช่วยให้เห็นเทรนด์สถานการณ์ประกอบกับบริบทได้เป็นอย่างดี



ภาพที่ 2.2 แสดงกราฟตัวอย่างยอดขายแต่ละเดือน

(ที่มา: mindtools.com)

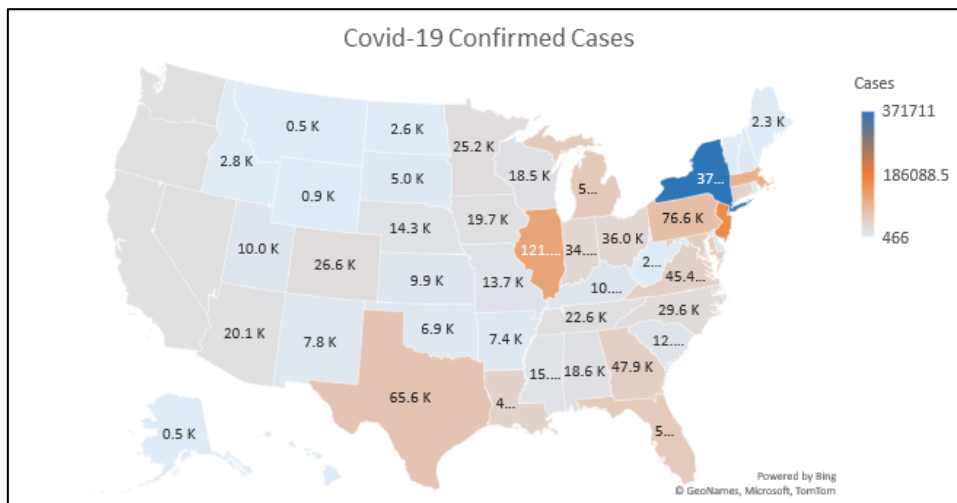
3) ตาราง (Tables) ก็เป็นอีกรูปแบบที่ใช้กันมากเพื่อนำเสนอข้อมูลให้ออกมาได้ง่าย ตารางประกอบไปด้วย 2 ส่วน ได้แก่ คอลัมน์และแถว ซึ่งช่วยจัดการข้อมูลให้เรียบร้อย ช่วยให้มองเห็นบริบทและความสัมพันธ์ของข้อมูลหลายๆ ชุดได้อย่างง่ายดาย

Marks	Number of Students		Total
	Males	Females	
30 – 40	8	6	14
40 – 50	16	10	26
50 – 60	14	16	30
60 – 70	12	8	20
70 – 80	6	4	10
Total	56	44	100

ภาพที่ 2.3 แสดงตัวอย่างตารางช่วงอายุ

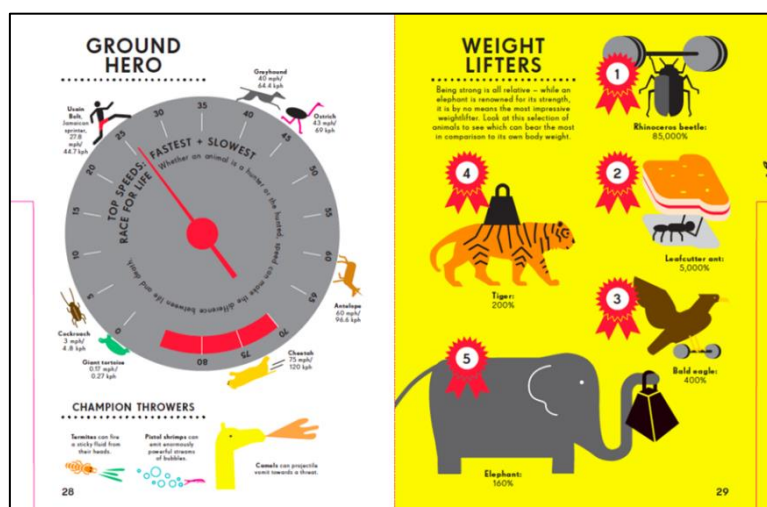
(ที่มา: embibe.com)

4) แผนที่ (Maps) เป็นการนำเสนอข้อมูลบนแผนที่เพื่อแสดงข้อมูลเกี่ยวกับพื้นที่ต่างๆ ยกตัวอย่างเช่น การนำเสนอข้อมูลยอดผู้ติดเชื้อ Covid-19 ในแต่ละรัฐของประเทศสหรัฐอเมริกา ซึ่งนอกจากการใส่ข้อมูลลงไปยังพื้นที่ต่างๆ แล้ว ยังสามารถใช้สีส้มเพื่อบอกระดับปริมาณหรือความหนาแน่นของผู้ติดเชื้ออีกด้วย



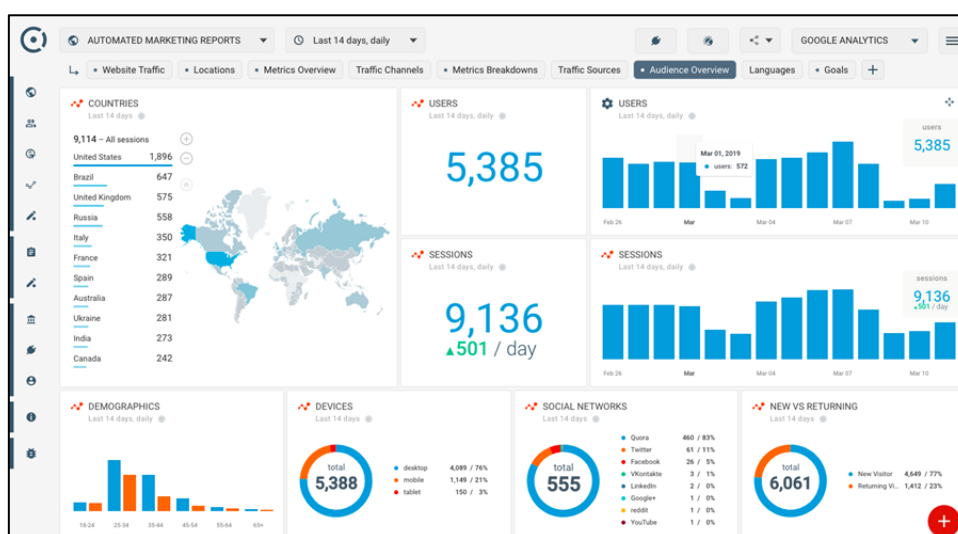
ภาพที่ 2.4 แสดงแผนที่จำนวนผู้ป่วย Covid-19
(ที่มา: spreadsheetweb.com)

5) อินโฟกราฟิก (Infographic) คือ การนำเสนอสารสนเทศ (Info: information) ด้วยภาพกราฟิก (Graphic) เป็นรูปแบบการนำเสนอข้อมูลที่ใช้ภาพสื่อแทน ทำให้ผู้อ่านข้อมูลเข้าใจข้อมูลได้ง่ายหรือสามารถทำความเข้าใจผ่านภาพแทนที่คุ้นเคย นอกจากนี้ อินโฟกราฟิกยังเป็นรูปแบบการนำเสนอข้อมูลที่น่าสนใจ มีการนำเทคนิคการเล่าเรื่อง (Storytelling) มาใช้ทำให้ข้อมูลน่าสนใจ น่าดึงดูด จึงมักจะใช้เพื่อนำเสนอเนื้อหา ความรู้ หรือเป็นสื่อการเรียนการสอน



ภาพที่ 2.5 แสดงภาพอินโฟกราฟิก
(ที่มา: simonrogers.net)

6) แดชบอร์ด (Dashboards) คือ การนำข้อมูลต่างๆ มาเรียบเรียงและสรุปเป็นภาพ โดยใช้แผนภูมิและกราฟต่างๆ มาใช้นำเสนอ ปัจจุบันแดชบอร์ดเป็น Data Visualization ที่นิยมใช้กับการนำเสนอข้อมูลแบบ Real-time ผ่านซอฟต์แวร์หรือเครื่องมือจัดการและวิเคราะห์ข้อมูลต่างๆ เช่น เครื่องมือการตลาด เครื่องมือบริหารจัดการข้อมูล เครื่องมือติดตามและดูแลเว็บไซต์



ภาพที่ 2.6 แสดงตัวอย่างหน้าแดชบอร์ด

(ที่มา: octoboard.com)

2.2 ทฤษฎี

2.2.1 การทำเหมืองข้อมูล (Data Mining)

(เกียรติพงษ์ อุดมชนะธีระ, 2563) การทำเหมืองข้อมูล (Data Mining) หรือ การค้นหาค้นหาองค์ความรู้ในฐานข้อมูล (Knowledge Discovery In Databases หรือ KDD) เป็นเทคนิคการค้นหาค้นหาองค์ความรู้ที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่เพื่อค้นหาแนวโน้ม รูปแบบรวม ความสัมพันธ์หรือความรู้ใหม่อื่น ๆ โดยอาศัยข้อมูลในอดีตความรู้ที่ได้ทำให้เข้าใจและทราบปัจจัยที่ทำให้เกิดลักษณะบางอย่างของข้อมูล ซึ่งจะช่วยให้สามารถทำนายแนวโน้มของข้อมูลใหม่ที่จะเกิดขึ้นในอนาคตได้ผลที่ได้จะมีลักษณะของข้อมูลอยู่ 3 แบบคือข้อมูลแบบที่ไม่ทราบมาก่อน (Unknown) ไม่มีความชัดเจนและไม่สามารถตั้งสมมติฐานก่อนได้ว่าควรเป็นแบบใด ข้อมูลแบบที่มีเหตุผล (Valid) และข้อมูลแบบที่สามารถนำไปใช้ได้ (Actionable)

กระบวนการทำงานของการทำเหมืองข้อมูลประกอบด้วยขั้นตอนต่าง ๆ ดังนี้

1) การทำความสะอาดข้อมูล (Data Cleaning) เป็นขั้นตอนในการกำจัดข้อมูลที่ผิดปกติ หรือตัดรายละเอียดที่ไม่จำเป็นออกไป เพื่อแก้ไขข้อมูลที่บกพร่อง

2) การรวมข้อมูล (Data Integration) เป็นขั้นตอนการรวบรวมข้อมูลที่มาจากรายแหล่งให้รวมกันเป็นชุดข้อมูลเดียว และมีการตรวจสอบรูปแบบของการจัดเก็บข้อมูลว่าเหมือนหรือต่างกัน เพื่อลดความซ้ำซ้อนของข้อมูลด้วย

3) การคัดเลือกข้อมูล (Data Selection) เป็นขั้นตอนในการคัดเลือกข้อมูลที่เกี่ยวข้องกับวัตถุประสงค์ที่ตั้งไว้เพียงบางส่วนจากแหล่งข้อมูลที่บ้านที่กไว้เพื่อให้ผลการวิเคราะห์ข้อมูลสามารถนำไปประกอบการตัดสินใจได้ตรงตามความต้องการ

4) การแปลงข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลเพื่อใช้กับแบบจำลองต่าง ๆ ของกระบวนการทำเหมืองข้อมูล ซึ่งมีวิธีแปลงข้อมูลในรูปแบบต่าง ๆ เช่น การนอร์มัลไลซ์ (Normalization) การดิสครีต (Discretization) การเข้ารหัส (1 Of N Coding) และการเจเนอรัลไลซ์ (Generalization)

5) การทำเหมืองข้อมูล (Data mining) เป็นขั้นตอนในการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่โดยมีเทคนิคหลายรูปแบบ เช่น การแบ่งกลุ่มข้อมูล (Clustering), การจำแนกประเภทข้อมูล (Classification) และ การค้นหากฎความสัมพันธ์ (Association)

6) การประเมินผล (Evaluation) เป็นขั้นตอนสำหรับการวิเคราะห์ข้อมูลและประเมินผลที่ได้จากการทำเหมืองข้อมูล ซึ่งจะต้องเก็บผลลัพธ์ของกระบวนการทำเหมืองข้อมูล มาสรุปความหมายและตีความของผลลัพธ์ให้ออกมาเป็นข้อมูลความรู้ใหม่

7) อภิปรายผล (Knowledge Presentation) เป็นขั้นตอนการนำเสนอความรู้ที่ได้

2.2.2 หลักในการทำเหมืองข้อมูล

จรัสศรี รุ่งรัตนอุบล, สำนักพิมพ์มหาวิทยาลัยนเรศวร, (2566)การทำเหมืองข้อมูลจัดเป็นกระบวนการอัตโนมัติ เพื่อค้นพบข้อสนเทศหรือองค์ความรู้ รูปแบบ หรือแม้แต่ตัวแบบเพื่อการพยากรณ์จากฐานข้อมูลขนาดใหญ่ การทำเหมืองข้อมูล มีหลักการการทำเหมืองข้อมูล 3 หลักการทำงานหลัก ๆ ดังนี้

1) การแบ่งกลุ่มข้อมูล (Clustering) เป็นการรวมข้อมูลที่มีลักษณะ รูปแบบ และแนวโน้มที่เหมือนกันไว้ในกลุ่มเดียวกันโดยเริ่มต้นจากการาตัวแทนของกลุ่ม และทำการเปรียบเทียบข้อมูลตัวที่เหลือกับตัวแทนของแต่ละกลุ่ม ถ้าข้อมูลของตัวที่นำมาเปรียบเทียบเหมือน คล้ายคลึง หรือสอดคล้องกับตัวแทนของกลุ่มไหนก็จะถูกจัดประเภทให้อยู่ในกลุ่มนั้น

2) การจำแนกประเภท (Classification) เป็นการจำแนกประเภทของข้อมูลโดยการเรียนรู้ข้อมูลเพื่อสร้างเป็นฐานความรู้ซึ่งแต่ละกลุ่มจะมีคุณสมบัติเฉพาะของตัวเองที่เรียกว่าตัวจำแนกข้อมูล (Classifier) จะมีหลักในการพัฒนาและอัลกอริทึมที่เกี่ยวข้องหลายตัวที่นิยมใช้ในปัจจุบัน โดยในที่นี้จะกล่าวถึง ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว (K-nearest Neighbor Algorithm) วิธีต้นไม้ตัดสินใจ (Decision Tree) การสร้างกฎ (Rule-based Classifier) วิธีเบย์อย่างง่าย (Naive Bayes Classifier) และโครงข่ายประสาทเทียม (Artificial Neural Network) ซึ่งสามารถใช้ในการวิเคราะห์ และแยกแยะข้อมูลที่เข้ามาใหม่ได้

3) การพยากรณ์ (Prediction) เป็นการนำข้อมูลมาทำนายค่าตอบ แต่ค่าของการพยากรณ์หรือการทำนายจะเป็นค่าแบบต่อเนื่อง (Continuous Value) ซึ่งแตกต่างจากเทคนิคการจำแนกที่ค่าตอบของการทำนายจะเป็นค่าเต็มหน่วย (Discrete Value) หรือที่เรียกว่าคลาส (Class) ที่เป็นการสื่อถึงค่าคำตอบแบบเต็มหน่วย ขั้นตอนการพัฒนาตัวพยากรณ์จะมีความคล้ายคลึงกับการพัฒนาตัวจำแนก โดยจะมีการแบ่งข้อมูล เป็นข้อมูลฝึกสอน และข้อมูลทดสอบเหมือนกัน แต่สิ่งที่แตกต่างกันคือการวัดประสิทธิภาพ ของการพยากรณ์หรือความแม่นยำในการพยากรณ์ (Predicted Accuracy) ซึ่งจะใช้เกณฑ์การวัดค่าความแม่นยำอีกลักษณะหนึ่งที่ไม่ใช่การวัดร้อยละการจำแนกที่ถูกต้อง และเมตริกซ์สับสนเหมือนเทคนิคการจำแนก โดยเกณฑ์การวัดประสิทธิภาพที่นิยมใช้กัน เช่น รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) เป็นต้น

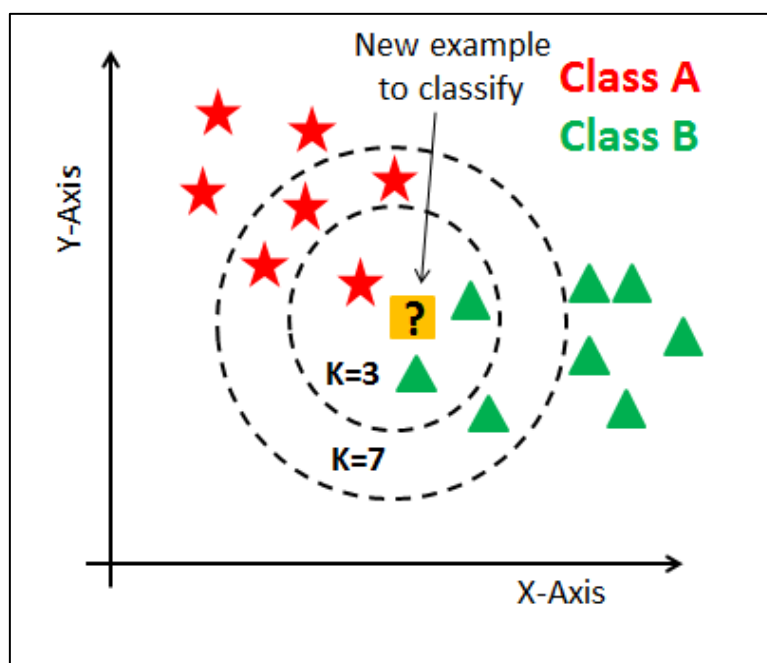
2.2.3 การจำแนกประเภทของข้อมูล(Classification)

เทคนิคการจำแนก เป็นเทคนิคหนึ่งที่ใช้ในงานด้านการทำเหมืองข้อมูล (Data Mining) ในการสร้างแบบจำลองเพื่อทำนายค่าตอบที่มีลักษณะค่าข้อมูลเป็นแบบเชิงคุณภาพ (Qualitative value) หรือเชิงกลุ่ม(Category data) เช่น จัดอยู่ในกลุ่ม เสี่ยง/ไม่เสี่ยง ระดับความเสี่ยงต่อการเป็นโรค เสี่ยงมาก/เสี่ยงน้อย/เสี่ยงปานกลาง หรือวิธีการรักษา 1/2/3 เป็นต้น โดยหลักการทำงานของเทคนิค คือ จะเป็นการสร้างแบบจำลองจากชุดข้อมูลที่มีอยู่และนำมาประยุกต์ใช้ทำนายค่าตอบของชุดข้อมูลใหม่ (Unseen Object) ซึ่งเป็นเทคนิคที่ได้รับความนิยมและถูกนำไปประยุกต์ใช้ในงานหลากหลายด้าน เช่น การจำแนกประเภทกลุ่มของลูกค้า การตรวจสอบและพิจารณาอนุมัติสินเชื่อแก่ลูกค้า และการตรวจสอบความผิดปกติหรือภาวะเสี่ยงของการเป็นโรคเป็นต้นในขั้นตอนของการสร้างแบบจำลองเทคนิคการจำแนกนั้นมีอัลกอริทึมที่สามารถใช้ได้หลาย ตัวอย่างอัลกอริทึมที่นิยมใช้ เช่น ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว (K-nearest neighbors) ต้นไม้ตัดสินใจ (Decision Tree) ซัพพอร์ตเวกเตอร์แมชชีน (Support

Vector Machine) นาอิวเบย์ (Naïve Bayes) และโครงข่ายประสาทเทียม (Artificial neural network) เป็นต้น

1) ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว (K-nearest neighbors)

อัลกอริทึม K-nearest neighbors (KNN) นับเป็นเทคนิคที่มีวิธีการไม่ซับซ้อนและเข้าใจได้ง่ายที่สุดที่ใช้ในการจำแนกประเภทข้อมูล โดยหลักการทำงาน คือ จะใช้หลักการเปรียบเทียบความคล้ายคลึงกันของข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงหรืออยู่ใกล้กับข้อมูลใดมากที่สุด k ตัว จากนั้นจะทำการตัดสินใจว่าคำตอบของข้อมูลที่สนใจนั้นควรเป็นคำตอบเดียวกับข้อมูลที่อยู่ใกล้ที่สุด k ตัวนั้น ทั้งนี้ k คือความถี่ของข้อมูลที่อยู่ใกล้กับข้อมูลที่สนใจ



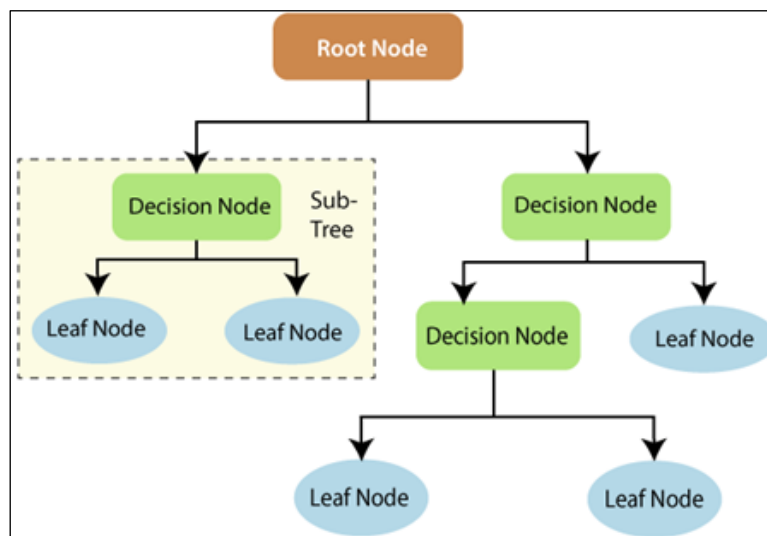
ภาพที่ 2.7 แสดงอัลกอริทึม K-nearest neighbors (KNN)

(ที่มา: <https://shorturl.asia/gKbul>)

2) ต้นไม้ตัดสินใจ (Decision Tree)

ลักษณะของต้นไม้การตัดสินใจในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ โดยมีลักษณะการตัดสินใจเป็นแบบแผนผังต้นไม้ประกอบด้วย ส่วนที่เป็นบัพภายใน (inner node) หรือที่เรียกว่าโหนดของต้นไม้ใช้แทนตัวแปรต่างๆ ที่ใช้ในการตัดสินใจ ส่วนที่กิ่ง (Branch) ใช้แทนค่าที่เป็นได้ของตัวแปร และส่วนที่เป็นบัพใบ (leaf node) หรือที่เรียกว่า

โหนดใบ ใช้แทนค่าคำตอบของการตัดสินใจหรือการจำแนก ต้นไม้ตัดสินใจ เป็นเทคนิคที่ใช้ได้กับข้อมูลที่มีค่าต่อเนื่อง (Continuous values) และไม่ต่อเนื่อง (Discrete values) โดยต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง จะเรียกว่าต้นไม้ตัดสินใจแบบจำแนก (Classification trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (Continuous values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (regression trees)

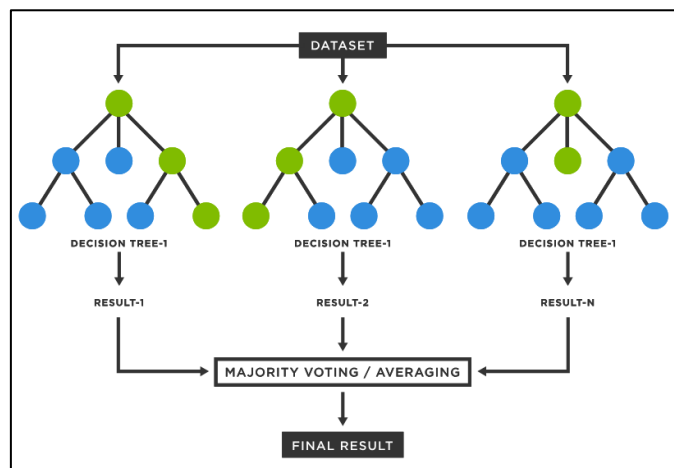


ภาพที่ 2.8 แสดงต้นไม้ตัดสินใจ (Decision Tree)

(ที่มา: <https://medium.com/@favourphilic/decision-tree-5c1c7b6db59>)

3) เทคนิค Random Forest

Random Forest คือขั้นตอนวิธีหนึ่งของการเรียนรู้ของเครื่องที่นิยมใช้มาก ใช้ได้ดีในหลายๆ ปัญหา ทั้งกับปัญหาแบบ Regression และ Classification โดย Random Forest เป็นขั้นตอนวิธีพัฒนาต่อยอดมาจาก Decision Tree ต่างกันที่ Random Forest เป็นการเพิ่มจำนวนต้นไม้ (Tree) เป็นหลายๆ ต้น ทำให้ประสิทธิภาพการทำงานและพยากรณ์สูงขึ้น Random Forest มีหลักการการทำงาน คือ จะแบ่งข้อมูลออกเป็นต้นไม้ตัดสินใจ (Decision Tree) หลายๆ ต้น โดยแต่ละต้นจะได้รับคุณลักษณะ (Feature) และข้อมูล (Data) ที่ไม่เหมือนกันทั้งหมด เพื่อให้ได้ต้นไม้ที่มีความหลากหลายและมีความอิสระต่อกันมากขึ้น ตัวอย่างการแบ่งข้อมูล และการสุ่มเลือก feature



ภาพที่ 2.9 แสดงเทคนิค Random Forest

(ที่มา: <https://shorturl.asia/67cM9>)

2.2.4 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.2.4.1 การสร้างเว็บไซต์สิ่งสำคัญอยู่ที่การ ออกแบบเว็บ เพราะเว็บไซต์ที่มีรูปแบบสวยงาม จะสามารถดึงดูดความสนใจจากผู้คนได้ดีกว่า ทำให้ผู้คนเกิดความรู้สึกประทับใจ อยากกลับมาใช้งานเว็บไซต์อีกครั้งในอนาคต ดังนั้นเริ่มแรกก่อนทำเว็บไซต์ จึงจำเป็นต้องทำความเข้าใจ กับหลักการออกแบบ และรูปแบบโครงสร้างของเว็บก่อน การออกแบบเว็บไซต์เพื่อให้มีประสิทธิภาพ และสามารถดึงดูดความสนใจของผู้คนได้ดี จะต้องมียอดประกอบของเว็บไซต์อย่างครบถ้วน(วันปีลีฟ, 2560)

1) ความเรียบง่าย เข้าใจง่าย การออกแบบเว็บไซต์ที่ดี จะต้องเน้นที่ความเรียบง่ายเป็นหลัก โดยเลือกนำเสนอเฉพาะสิ่งที่ต้องการนำเสนอจริงๆ ในรูปแบบที่หลากหลาย โดยอาจจะเป็นสี สัน กราฟิก ภาพเคลื่อนไหวหรือตัวอักษร ที่สำคัญจะต้องมีการนำเสนอที่ไม่ดูรกหน้าเว็บจนเกินไป เพื่อไม่ให้เกิดความรู้สึกรกสายตา หรือสร้างความเบื่อหน่าย นำรำคาญให้กับผู้ที่เข้าชมเว็บไซต์ มีตัวอย่างเว็บไซต์ที่มีการออกแบบโดยเน้นความเรียบง่ายได้ดี

2) ความสม่ำเสมอ ไม่สับสนควรออกแบบเว็บไซต์ด้วยความสม่ำเสมอ คือจะต้องมีรูปแบบ กราฟิก โทนมสีและการตกแต่งต่างๆ ให้แต่ละหน้าบนเว็บไซต์มีความคล้ายคลึงกัน และเป็นแนวเดียวกันไปตลอดทั้งเว็บไซต์ ดังตัวอย่างเว็บไซต์ทั่ว ๆ ไปที่จะสังเกตเห็นได้ว่า

ทุกหน้าของเว็บไซต์นั้น จะเน้นการตกแต่งในรูปแบบเดียวกันทั้งหมด ต่างก็แค่การนำเสนอของแต่ละหน้าเท่านั้น

3) สร้างความโดดเด่น เป็นเอกลักษณ์ การออกแบบเว็บไซต์เพื่อให้สามารถสื่อถึงจุดประสงค์ในการนำเสนอเว็บได้ดี จะต้องมีการสร้างความเป็นเอกลักษณ์และจุดเด่นให้กับเว็บไซต์ เพื่อให้สามารถสะท้อนถึงลักษณะขององค์กรได้มากที่สุด โดยการสร้างเอกลักษณ์ดังกล่าวนั้น อาจใช้ชุดสี รูปภาพ ตัวอักษรหรือกราฟิก นอกจากนี้ก็ต้องขึ้นอยู่กับว่า เป็นเว็บไซต์แบบทางการหรือไม่ เพื่อจะได้ออกแบบได้อย่างเหมาะสมที่สุด

4) เนื้อหาต้องดี ครบถ้วน เนื้อหาเป็นสิ่งที่สำคัญที่สุดของการสร้างเว็บไซต์ เพราะสิ่งที่ทำให้ผู้คนเกิดความสนใจ และหมั่นติดตามเว็บไซต์เหล่านั้นอยู่เสมอ ก็คือเนื้อหาที่มีความสมบูรณ์และน่าสนใจ นอกจากนี้จะต้องมีการปรับปรุง พัฒนาเนื้อหาบนเว็บให้มีความทันสมัยอยู่เสมอ รวมถึงข้อมูลต้องมีความถูกต้องที่สุด

5) ระบบเนวิเกชัน เป็นเสมือนป้ายบอกทางเพื่อให้ผู้ใช้งาน ไม่เกิดความสับสนในขณะที่ใช้งานเว็บไซต์ ซึ่งการออกแบบเนวิเกชันก็จะต้องเน้นที่ความเรียบง่าย ใช้งานสะดวก และมีความเข้าใจได้ง่าย ที่สำคัญจะต้องมีตำแหน่งการวางที่สม่ำเสมอเพื่อให้ดูเป็นแนวทางเดียวกัน ทำให้ผู้ใช้งานหรือผู้ชมรู้สึกประทับใจ และจดจำเว็บไซต์ได้ง่ายขึ้นส่วนใครที่มีการนำกราฟิกมาใช้ในระบบเนวิเกชัน ก็จะต้องเลือกกราฟิกที่สามารถสื่อความหมายได้ดีเช่นกัน

6) คุณภาพของเว็บไซต์ที่ดีจะต้องมีคุณภาพ ทั้งสิ่งที่ปรากฏให้เห็นบนเว็บไซต์ ไม่ว่าจะเป็นกราฟิก ชนิดตัวอักษร รูปภาพหรือสีสันทันทีใช้ เนื้อหาที่นำมาแสดงผล ซึ่งหากเว็บไซต์มีคุณภาพก็จะสร้างความน่าเชื่อถือ และเป็นจุดเด่นที่ทำให้ผู้คนส่วนใหญ่เกิดความสนใจได้ดี เพราะฉะนั้นห้ามละเลยในส่วนของคุณภาพเด็ดขาด

7) ความสะดวกในการเข้าใช้งานเว็บไซต์ควรให้ความสะดวกสบายแก่ผู้ใช้งานได้ดี คือจะต้องมีการแสดงผลได้ในทุกระบบปฏิบัติการ ไม่ว่าจะเป็นเว็บเบราว์เซอร์ คอมพิวเตอร์ โน้ตบุ๊ก หรือบนโทรศัพท์มือถือ ที่สำคัญจะต้องมีความละเอียดของการแสดงผล และสามารถใช้งานได้โดยไม่มีปัญหาด้วย

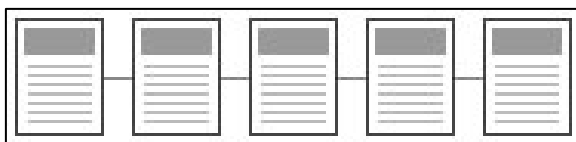
8) ความคงที่ของการออกแบบ การออกแบบเว็บไซต์ควรจะมี ความคงที่ในการออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกัน และมีการเรียบเรียงเนื้อหาอย่างรอบคอบ ทำให้เว็บมีความน่าเชื่อถือ และดูมีคุณภาพ ช่วยสร้างความประทับใจให้กับผู้ใช้งานได้เป็นอย่างดี

9) ความคงที่ของการทำงานระบบการทำงานบนเว็บไซต์จะต้องมีความคงที่ และสามารถใช้งานได้ดี ซึ่งนอกจากการออกแบบระบบการทำงานให้มีความทันสมัยและ

สร้งสรรค์แล้ว ก็จะต้องหมั่นตรวจสอบอยู่เสมอ เพราะหากระบบการใช้งานมีความผิดปกติก็จะได้แก้ปัญหาได้ทัน นอกจากนี้อาจมีการอัปเดตดีไซน์ให้ทันสมัยขึ้นบ่อยๆ เพื่อให้ผู้ใช้งานรู้สึกสนุกไปกับการใช้งานเว็บไซต์

2.2.4.2 รูปแบบโครงสร้างของเว็บไซต์ การออกแบบโครงสร้างของเว็บไซต์สามารถทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะสมกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่มักจะประกอบไปด้วย 4 รูปแบบดังนี้

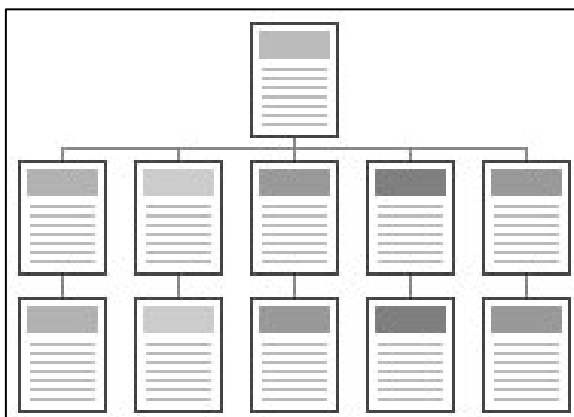
1) โครงสร้างเว็บไซต์แบบเรียงลำดับ จะเป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิงค์เนื้อหา ก็จะลิงค์ไปที่หน้า มีทิศทางการเข้าสู่เนื้อหาต่างๆ ในแบบเส้นตรง ใช้ปุ่มเดินทาง-ถอยหลังในการกำหนดทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับก็มีข้อเสีย คือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหาเพราะไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาด้วยตัวเองได้



ภาพที่ 2.10 แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ

(ที่มา: <https://www.1belief.com/article/website-design/>)

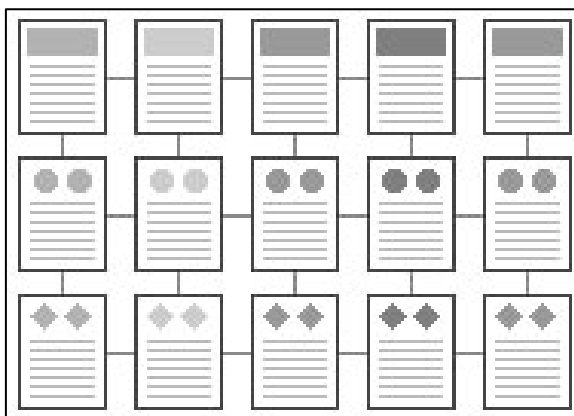
2) โครงสร้างแบบลำดับขั้น นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่างๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆ และมีการนำเสนอรายละเอียดย่อยๆ ที่ลัดหลั่นกันมา ทำให้สามารถทำความเข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีไฮมเพจเป็นจุดเริ่มต้น และจุดรวมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง



ภาพที่ 2.11 แสดงโครงสร้างแบบลำดับชั้น

(ที่มา: <https://www.1belief.com/article/website-design/>)

3) โครงสร้างแบบตาราง เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีคามยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่างๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งานสามารถเปลี่ยนทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลา แถมยังทำให้เว็บไซต์มีความทันสมัยขึ้น

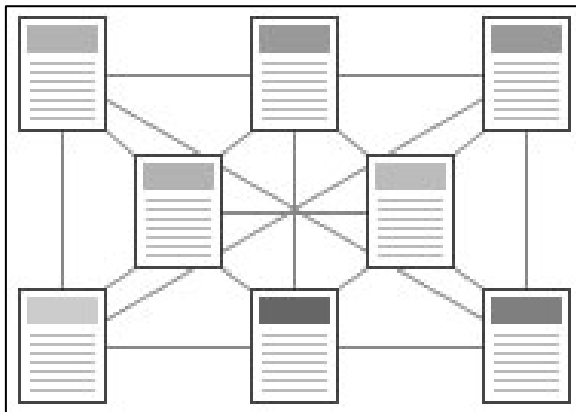


ภาพที่ 2.12 แสดงโครงสร้างแบบตาราง

(ที่มา: <https://www.1belief.com/article/website-design/>)

4) โครงสร้างแบบใยแมงมุม เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถ

เข้าถึงหน้าเว็บเพจต่างๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2.13 แสดงโครงสร้างแบบใยแมงมุม

(ที่มา: <https://www.1belief.com/article/website-design/>)

2.2.4.3 ส่วนประกอบสำคัญของหน้าเว็บเพจ หน้าเว็บเพจ จะมีส่วนประกอบสำคัญที่จำเป็นต้องมีอยู่ 3 ส่วน

1) ส่วนหัวของหน้า (Header) อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซต์ต่อไป ซึ่งส่วนใหญ่มักจะมีการใส่ภาพกราฟิกให้ดูสวยงาม สิ่งสำคัญหลักๆ เลย ก็คือ โลโก้ ชื่อเว็บไซต์และเมนูหลักที่สามารถคลิกไปยังเนื้อหาในหน้าเว็บเพจต่างๆ ได้

2) ส่วนของเนื้อหา (Body) อยู่บริเวณตอนกลางของหน้าเว็บ โดยจะแสดงข้อมูลเกี่ยวกับเนื้อหาบนเว็บแบบคร่าวๆ ซึ่งก็จะมีข้อความ กราฟิก ตารางข้อมูลหรือวิดีโอประกอบอยู่ และหากมีเมนูแบบเฉพาะกลุ่มก็จะถูกจัดไว้ในหน้านี้เช่นกัน และที่สำคัญเนื้อหาในส่วนนี้ควรจะต้องมีความกระชับ เข้าใจง่าย มีการใช้รูปแบบตัวอักษรแบบเรียบง่ายและเป็นระเบียบ

3) ส่วนท้ายของหน้า (Footer) อยู่ล่างสุดของหน้าเว็บ ซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่างๆ เพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูลเจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่างๆ เกี่ยวกับการใช้งานเว็บไซต์อย่างถูกต้อง

2.2.4.4 วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์ การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บ ว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไรต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่ม

1) สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอดภัยและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2) สีโทนเย็น (Cool Colors) เป็นสีแห่งความสุภาพและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้น้ำหนักจากในระยะไกลได้อีกด้วย

3) สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

2.2.5 ทฤษฎีเกี่ยวข้องกับภาวะซึมเศร้า

องค์การอนามัยโลก ได้ให้คำนิยามของภาวะซึมเศร้าว่า หมายถึงภาวะที่เกิดจากความผิดปกติของสารสื่อประสาทในสมอง ซึ่งเกิดขึ้นได้กับทุกคน เป็นภัยเงียบด้านสุขภาพ ไม่ใช่เพียงแค่อารมณ์เบื่อเซ็ง ท้อแท้ หดหู่ เศร้า เหงาโดยทั่วไป อาการที่สำคัญของภาวะซึมเศร้า คือมีอาการมึนเศร้า ท้อแท้ หดหู่ สิ้นหวังอย่างรุนแรง เกิดขึ้นเกือบตลอดเวลาทั้งวันติดต่อกันอย่างน้อย 2 สัปดาห์ ร่วมกับเบื่อหน่าย หมดความสนใจในการทำงาน หรือกิจกรรมที่เคยชอบทำ หากไม่ได้รับการบำบัดอย่างถูกต้องจะเป็นเรื้อรัง และกลับไปเป็นซ้ำได้ หากอาการซึมเศร้ารุนแรง อาจจบชีวิตด้วยการฆ่าตัวตายได้มากกว่าคนทั่วไปถึง 20 เท่า สำหรับภาวะซึมเศร้าในผู้สูงอายุ นั้น เป็นความเจ็บป่วย ไม่ใช่ความอ่อนแอทางจิตใจ ผู้สูงอายุจะมีความรู้สึกไม่สบายใจ เซ็ง ทุกข์ใจ เศร้า ท้อแท้ ซึม หงอย เบื่อ ไม่อยากพูด ไม่อยากทำอะไร หรือทำอะไรก็ไม่เพลิดเพลินเหมือนเดิม ภาวะซึมเศร้าในผู้สูงอายุ มักเกี่ยวข้องกับโรคทางกายต่าง ๆ เช่น โรคหัวใจ ความดันโลหิตสูง เบาหวาน โรคหลอดเลือดสมอง โรคมะเร็ง โรคไตวาย ปวดเรื้อรัง และมีการสูญเสีย (คนที่รักหรือทรัพย์สินจำนวนมาก) การเผชิญกับความลำบากในชีวิต เป็นต้น

จากการทบทวนวรรณกรรม สรุปได้ว่า ภาวะซึมเศร้า เป็นอาการเจ็บป่วยทางจิตใจชนิดหนึ่ง ส่งผลให้มีพฤติกรรมเปลี่ยนแปลงไป เช่น มีความรู้สึกไม่สบายใจ เศร้า ท้อแท้ ไม่อยากทำอะไร มีความรู้สึกสิ้นหวัง มองตนเองไร้ค่า เป็นภาระต่อคนอื่น ถ้ามีอาการมาก จะมีความรู้สึกเบื่อชีวิตคิดฆ่าตัวตาย โดยภาวะซึมเศร้าในผู้สูงอายุ เกี่ยวข้องกับโรคทางกายต่าง ๆ เช่น โรคหัวใจ ความดันโลหิตสูง เบาหวาน โรคหลอดเลือดสมอง โรคมะเร็ง โรคไตวาย ปวดเรื้อรัง และมีการสูญเสีย การเผชิญกับความลำบากในชีวิต เป็นต้น

2.2.5.1 ทฤษฎีเกี่ยวกับภาวะซึมเศร้า สามารถแบ่งได้เป็น 2 ประเภท

1. ด้านร่างกาย

ทฤษฎีที่ใช้อธิบายการเกิดภาวะซึมเศร้า ได้แก่ ทฤษฎีเกี่ยวกับพันธุกรรม (Genetic) และทฤษฎีชีวเคมี (Biochemical Theory of Depression)

1) ทฤษฎีเกี่ยวกับพันธุกรรม (Genetic Transmission) เป็นทฤษฎีที่เชื่อว่าภาวะซึมเศร้าเป็นโรคที่ถ่ายทอดทางพันธุกรรม โดยผลการศึกษา พบว่าหากพ่อหรือแม่คนใดคนหนึ่งเป็นโรคซึมเศร้า ลูกมีโอกาสเป็นโรคดังกล่าวร้อยละ 27 แต่ถ้าทั้งพ่อและแม่เป็นโรคซึมเศร้า ลูกมีโอกาสเป็นโรคดังกล่าวเพิ่มขึ้นเป็นร้อยละ 54 นอกจากนี้ถ้าลูกเป็นคู่แฝดที่เกิดจากไข่ใบเดียวกัน จะมีโอกาสเป็นโรคซึมเศร้าสูงถึงร้อยละ 70

2) ทฤษฎีชีวเคมี (Biochemical Theory of Depression) มีแนวคิดว่าภาวะซึมเศร้าเกิดจากสารสื่อประสาทในสมองขาดความสมดุล (neurotransmitter) กลุ่มไบโอเจนิคเอมีนส์ (biogenic amines) ได้แก่ อินโดเลมีนส์ (indoleamines) หรือ ซีโรโทนิน (serotonin) ที่มีความสำคัญหลักต่อการเกิดภาวะซึมเศร้า และกลุ่มแคทีโคลามีนส์ (catecholamines) คือ นอร์เอพิเนฟริน (norepinephrine) และโดพามีน (dopamine) โดยมีความผิดปกติของกระบวนการดูดกลับสารเข้าสู่เซลล์ประสาทส่งต่อ ทำให้สารสื่อประสาทกลุ่มดังกล่าวมีปริมาณลดลง ผลคือทำให้บุคคลรู้สึกท้อแท้ หดหู่ หงอยเหงา เบื่อหน่าย สิ้นหวัง นอนไม่หลับ ฯลฯ ทำให้ประสิทธิภาพและความสามารถในการทำงานลดลง ในระยะยาวอาจส่งผลให้เป็นโรคซึมเศร้าได้ (โรงพยาบาลบำรุงราษฎร์, 2566)

2. ด้านจิตใจ

ทฤษฎีที่นิยมใช้ในการอธิบายการเกิดภาวะซึมเศร้า ได้แก่ ทฤษฎีปัญญานิยม (Cognitive Theory) ทฤษฎีจิตวิเคราะห์ (Psychoanalytic Theory) และทฤษฎีการสูญเสียและความเศร้าโศก (Grief and Loss Theory) มีรายละเอียดดังนี้

1) ทฤษฎีปัญญานิยม (Cognitive Theory: CT) (Beck, 1967) ประกอบด้วยองค์ประกอบพื้นฐาน 3 ประการ ได้แก่ ความคิดหลัก 3 ประการ (Cognitive triad) โครงสร้างทางความคิด (Schema) และเนื้อหาความคิดที่บิดเบือน (Cognitive distortion) มีรายละเอียดดังนี้

ความคิดหลัก 3 ประการ (Cognitive triad) เป็นความคิดเกี่ยวกับตนเอง โลก และอนาคตที่บิดเบือนไปในทางลบ ได้แก่ 1. รูปแบบความคิดด้านลบต่อตนเอง (Negative view of self) โดยผู้ที่มีภาวะซึมเศร้าจะมองตนเองในทางลบว่า ตนเองไม่มีคุณค่า ไร้ประสิทธิภาพ ทำให้เกิดข้อบกพร่องทางด้านร่างกาย และจิตใจ นอกจากนี้ยังมองว่าตนเองไม่เป็นที่ต้องการ ไม่มีคุณค่าอันเนื่องมาจากข้อบกพร่องที่ตนเองคิดขึ้นเอง จึงมีแนวโน้มที่จะไม่ยอมรับตนเอง เมื่อเกิดเหตุการณ์ต่าง ๆ ขึ้นก็จะคิดว่าเป็นความผิดของตนเอง ทำให้มีความคิดด้านลบต่อตนเอง ส่งผลให้เกิดภาวะซึมเศร้าตามมาได้ 2. รูปแบบความคิดด้านลบต่อโลก (Negative view of world) ผู้ที่มีภาวะซึมเศร้าจะแปลการปฏิสัมพันธ์ของตนเองกับสิ่งแวดล้อม

ในเชิงความสูญเสีย รู้สึกถูกบีบบังคับ มองชีวิตเต็มไปด้วยภาระอุปสรรค รู้สึกเจ็บปวด ทำให้รู้สึกว่าคุณค่าในตนเองลดลง หรือแปลสถานการณ์ต่าง ๆ ในทางลบ ส่งผลให้เกิดความคิดด้านลบโดยอัตโนมัติและเกิดภาวะซึมเศร้าตามมา 3. รูปแบบความคิดด้านลบต่ออนาคต (Negative view of future) ผู้ที่มีภาวะซึมเศร้าจะคิดว่าความลำบากในปัจจุบันจะดำเนินต่อไปไม่มีที่สิ้นสุด ไม่มีความหวังในอนาคต ทำให้เกิดภาวะซึมเศร้าตามมาโครงสร้างทางความคิด (Schema) ได้แก่การแปลความ การประเมินค่าและการอธิบายความหมายของเหตุการณ์ต่าง ๆ โดยผู้ที่มีภาวะซึมเศร้าจะมีความคิดประเภทบิดเบือนไปในทางลบเนื้อหาความคิดที่บิดเบือน (Cognitive distortion) เป็นการแปลความจริงที่บิดเบือนไปในแง่ลบ โดยแบ่งความคิดที่บิดเบือนไปดังนี้การเลือกสนใจเฉพาะประสบการณ์ในทางลบไม่สนใจเรื่องที่ดีหรือเป็นกลาง (Selective abstraction) การด่วนสรุปโดยไม่มีข้อมูลที่เพียงพอ (Arbitrary inference) การแปลความเกินกว่าพื้นฐานความเป็นจริง มองทุกอย่างโดยรวม เหมือนกันไปหมดมากเกินไป (Overgeneralization) การขยายต่อเติมหรือการตัดทอนเรื่องราว (Magnification and minimization) มีความคิดเอนเอียงไปกับเหตุการณ์ที่ไม่ได้เกี่ยวข้องกับตนเอง แปลพฤติกรรมหรือเหตุการณ์ภายนอกว่าเป็นตัวแสดงหรือยืนยันให้เห็นว่าตนนั้นไม่ดี (Personalization) มีความคิดเฉพาะด้านใดด้านหนึ่งมากกว่าที่จะมองข้อมูลทั้งหมด รับรู้เฉพาะด้านที่ลุ่มเหลว

ระบบความคิดดังกล่าว ส่งผลให้บุคคลมองตนเองว่าไร้ค่า ไร้ความสามารถไร้ทางออก ไม่ควรมีชีวิตอยู่ต่อไป และมองสิ่งแวดล้อมบิดเบือนจากความเป็นจริง มองว่าทุกคนรอบข้างเปื้อนหนาย และจับผิดตนเอง ไม่มีใครต้องการ และต้องการแยกตัวจากสังคม หมกมุ่นอยู่กับตนเองในทางลบ และพฤติกรรมดังกล่าวจะวนเวียนซ้ำไปมาตลอด จนส่งผลให้บุคคลเกิดความผิดปกติทางจิตได้ในที่สุด

2) ทฤษฎีจิตวิเคราะห์ (Psychoanalytic Theory) เชื่อว่าพฤติกรรมของบุคคลเกิดจากจิตใต้สำนึก (Subconscious) หรือสัญชาตญาณดั้งเดิม (basic instinct) การแก้ไขพฤติกรรมสามารถกระทำได้โดยการเจาะลึกถึงจิตใต้สำนึก ซึ่งประกอบด้วย 3 ส่วน ได้แก่ Id (สัญชาตญาณที่คอยผลักดันให้บุคคลแสดงพฤติกรรมเพื่อตอบสนองความสุข และความต้องการของตนเอง), Ego (ทำหน้าที่ในการตระหนักถึงความเป็นเหตุเป็นผลตามสภาพความเป็นจริง) และ Superego (ทำหน้าที่คอยควบคุมพฤติกรรม) โดยปกติการทำงานของ 3 ส่วนจะมีความสมดุลระหว่างกัน ทฤษฎีนี้สามารถอธิบายปรากฏการณ์การเกิดภาวะซึมเศร้าว่าเกิดจาก 2 สาเหตุคือ 1. เกิดจากการที่บุคคลขาดความสามารถในการควบคุมสัญชาตญาณของมนุษย์ ขาดความรักความอบอุ่นในครอบครัว ขาดการเลี้ยงดูที่เหมาะสมตามพัฒนาการ 2. เกิดจากการที่บุคคลไม่พัฒนา Superego มีการใช้กลไกการเก็บกด การโทษตนเอง เมื่อต้องเผชิญ

เหตุการณ์ที่เป็นความสูญเสีย ทำให้เกิดความรู้สึกเครียด กังวล 3) ทฤษฎีการสูญเสียและความเศร้าโศก (Grief and Loss Theory) อธิบายการสูญเสีย (loss) ว่าเป็นสถานการณ์ที่บุคคลประสบกับความสูญเสียสิ่งที่มีค่า มีความหมาย และความสำคัญต่อชีวิต ไม่ว่าจะเป็นการสูญเสียบุคคลสำคัญในชีวิต สูญเสียอวัยวะ สูญเสียทรัพย์สินความภาคภูมิใจในตนเอง รวมทั้งสูญเสียความมั่นคงปลอดภัยในชีวิต

3. ด้านสิ่งแวดล้อม/ สิ่งเร้าที่เข้ามากระตุ้น โดยทฤษฎีที่นิยมมาอธิบายการเกิดภาวะซึมเศร้าของบุคคล ได้แก่ ทฤษฎีด้านสังคม (Social Theory) โดยสถานการณ์ปัญหาที่ส่งผลต่อบุคคล ได้แก่ ปัญหาความรัก การปรับตัวกับเพื่อน/ ครอบครัว/ เพื่อนร่วมงาน การเงิน ความเจ็บป่วย หรือเหตุการณ์ที่เกิดขึ้นในแต่ละช่วงวัยของชีวิต เมื่อหลายสถานการณ์รวมกันเป็นสิ่งเร้าที่เข้ามาคุกคามภาวะสมดุลของบุคคล

2.3 เครื่องมือในการออกแบบ และวิเคราะห์ข้อมูล

2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

2.3.1.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

ชัยภพ แจ่มจรัส (2565) กล่าวว่าเทคนิคต้นไม้ตัดสินใจเป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูล ในการเรียนรู้ของเครื่อง (Machine Learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (Inner Node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (Leaf Node) จะแสดงประเภทของวัตถุ

ต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่าต้นไม้ตัดสินใจแบบจำแนก (Classification Trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (Continuous Values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (Regression Trees)

ต้นไม้การตัดสินใจในการบริหารธุรกิจ เป็นแผนผังต้นไม้ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (Risk Management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (Decision Theory) และ ทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่ง

สำหรับการทำเหมืองข้อมูล

2.3.1.2 ปัญหาที่เหมาะสมสำหรับต้นไม้การตัดสินใจ

เนื่องจากต้นไม้การตัดสินใจเป็นต้นไม้ที่แต่ละกิ่งที่ออกมาจากบัพแทนค่าของข้อมูลที่เป็นไปได้ในบัพนั้น เนื่องจากต้นไม้มีจำนวนกิ่งที่จำกัด ดังนั้นค่าของตัวแปรที่เป็นไปได้จึงต้องจำกัดด้วย จึงต้องมีจำนวนตัวแปรที่จำกัด และนอกจากนั้นยังบังคับว่าค่าของตัวแปรนั้นต้องไม่ต่อเนื่องด้วย โดยข้อมูลที่เข้ามานั้นอาจมีความผิดพลาดได้บ้าง โดยต้นไม้การตัดสินใจจะมีกระบวนการที่จะไม่นำความผิดพลาดนั้นมาพิจารณาเรียกว่าการตัดแต่งกิ่ง (Post-Pruning)

2.3.1.3 ขั้นตอนวิธีการสร้างต้นไม้การตัดสินใจ เพื่อจำแนกข้อมูล มีดังนี้

- 1) เลือก Attribute ที่ทำหน้าที่เป็น Root Node
- 2) จาก Root Node สร้างเส้นเชื่อมโยงไปยังโหนดลูก จำนวนเส้นเชื่อมโยงจะเท่ากับจำนวนค่าที่เป็นไปได้ทั้งหมดของ Attribute ที่เป็น Root Node
- 3) ถ้าโหนดลูกเป็นกลุ่มของข้อมูลที่อยู่ในคลาสเดียวกันทั้งหมด ให้หยุดสร้างต้นไม้ แต่ถ้าโหนดมีข้อมูลของหลายคลาสปะปนกันอยู่ ต้องสร้าง Subtree เพื่อจำแนกข้อมูลต่อไป โดยเลือก subtree มาทำหน้าที่เป็น Root Node ของ Subtree มาทำซ้ำในขั้นตอนที่ 2.) , 3.) ซึ่งการคำนวณมีดังนี้

เอนโทรปี (Entropy) การสร้างต้นไม้ตัดสินใจจากบนลงล่างด้วยการถามว่าลักษณะใด (ขอใช้คำว่าลักษณะแทนตัวแปรต้น) ควรจะเป็นรากของต้นไม้การตัดสินใจต้นนี้ และถามซ้ำ ๆ ไปเรื่อย ๆ เพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด (อังกฤษ: recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดที่สุ่นั้นดูจากค่าของลักษณะเรียกว่าเกนความรู้ (Information gain) ก่อนที่จะรู้จักเกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ $E(S)$ ดังนี้

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

เมื่อพิจารณาเอนโทรปีแล้วจะเห็นว่าเอนโทรปีจะมีค่าอยู่ระหว่าง 0 กับ 1 โดยจะมีค่าเป็นศูนย์เมื่อทุก ๆ กรณีมีผลลัพธ์เพียงแบบเดียว เช่น ใช่ทั้งหมด หรือ ไม่ใช่ทั้งหมด และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น หรือจะพูดอีกนัยหนึ่งก็คือเอนโทรปีจะมีค่ามากขึ้นหากข้อมูลไม่บริสุทธิ์ และจะตัดสินใจได้ว่าผลลัพธ์จะเป็นอะไรเมื่อเอนโทรปีเป็น 0

เท่านั้น

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

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

เมื่อแทนค่า S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

E คือ เอนโทรปีของตัวอย่าง

A คือ ตัวแปรต้นที่พิจารณา

value (A) คือ เซตของค่าของ A ที่เป็นไปได้

S_v คือ ตัวอย่างที่ A มีค่า v ทั้งหมด

จะเห็นว่าหากเกนความรู้ของ A ยิ่งมากแสดงว่าหลังจากแบ่งตัวอย่าง S ด้วย A แล้วในแต่ละชุดที่แบ่งได้จะมี Entropy เข้าใกล้ศูนย์มากยิ่งขึ้น ทำให้เกลที่จะตัดสินใจได้มากขึ้น เกนความรู้จึงเป็นค่าที่ดีที่จะบอกความดีของตัวแปรต้นที่นำมาพิจารณา

2.3.2 การประเมินผลโมเดล (Decision Tree)

สถิติในการทดสอบหาความแม่นยำของอัลกอริทึม และทดสอบความถูกต้องและความน่าเชื่อถือของโมเดลโดยใช้การประเมินค่าความแม่นยำ Confusion Matrix

Confusion Matrix คือ การประเมินผลลัพธ์การทำนาย (หรือผลลัพธ์จากโปรแกรม) เปรียบเทียบกับผลลัพธ์จริงโดยที่

True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่าจริง และคนบอกว่ามันจริง

True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง และคนบอกว่ามันไม่

จริง

False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่าจริง แต่คนบอกว่าไม่จริง

False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง แต่คนบอกว่าจริง

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

ภาพที่ 2.14 แสดงการประเมินค่าความแม่นยำ Confusion Matrix
(ที่มา: <https://shorturl.asia/4dXLQ>)

แสดงเป็นสมการได้ ดังนี้

Accuracy คือ ค่าที่บอกว่าโปรแกรมสามารถทำนายได้แม่นยำขนาดไหน

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}$$

สมการ accuracy = total correct predictions / total predictions made * 100

$$\text{accuracy} = 7 / 10 * 100$$

Expected,	Predicted
man,	woman
man,	man
woman,	woman
man,	man
woman,	man
woman,	woman
woman,	woman
man,	man
man,	woman
woman,	woman

จากข้อมูลที่ติด label ถูกต้อง คือ Expected และข้อมูลที่ทำนายออกมา คือ Prediction Precision คือ ค่าที่บอกว่าโปรแกรมทำนายว่าจริง ถูกต้องเท่าไร

$$precision = \frac{TP}{TP + FP}$$

สมการ Precision = TP/(TP+FP) ใช้ในการทำนายว่าถ้าทำนายผิดว่าเป็นแล้วเสียหาย (ถ้าทำนายผิดแล้วเกิดความเสียหาย) ถ้าค่า Precision มีค่าน้อยแสดงว่าโมเดลไม่เหมาะสม เช่น อีเมลไม่ได้เป็น spam แต่ทำนายว่าเป็น spam

Recall (True Positive Rate) คือ ค่าที่บอกว่าโปรแกรมทำนายได้ว่าจริง เป็นอัตราส่วนเท่าไรของจริงทั้งหมด

$$recall = \frac{TP}{TP + FN}$$

สมการ Recall = TP/(TP+FN) ใช้ในการทำนายที่ผิดแบบ negative แล้วเสียหาย เช่น เป็น HIV แล้วบอกว่าไม่เป็น

F1-score คือ การนำค่า Precision กับ Recall มาคำนวณค่าเฉลี่ยกันสำหรับพิจารณา รวมกันเพื่อความแม่นยำและความครบถ้วน

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

สมการ F1-score = 2 x (Precision x Recall) / (Precision + Recall) มีบางโมเดลที่จำเป็นต้องใช้ทั้ง 2 สูตร จึงต้องใช้หาค่าเฉลี่ยระหว่าง Precision * Recall แต่หน่วยหารมันไม่เท่ากัน จึงต้องใช้สูตร F1-score

2.3.2 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM) หรือ Cross Industry Standard Process for Data Mining หมายถึง กระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูลเพื่อการวิเคราะห์และนำไปใช้ประโยชน์ในทางธุรกิจ พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ DaimlerChrysler, SPSS และ NCR ในกระบวนการ CRISP-DM ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่

- 1) การทำความเข้าใจในธุรกิจ (Business understanding) เป็นการกำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์เพื่อหาความได้เปรียบทางการตลาด เพื่อนำมาแก้ไขปัญหาองค์กรและต้องสามารถระบุผลลัพธ์ที่มีได้
- 2) การทำความเข้าใจในข้อมูล (Data Understanding) โดยการรวบรวมข้อมูลที่เกี่ยวข้องให้อยู่รวมกลุ่มกัน คัดเลือกให้เหลือเพียงข้อมูลที่มีความสำคัญต่อการทำงาน
- 3) การเตรียมข้อมูล (Data Preparation) การจัดเตรียมข้อมูลประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล
- 4) การสร้างแบบจำลอง (Modeling) นำผลที่ได้มาทดลองทำแบบจำลอง เพื่อใช้ถ่ายทอดข้อมูลหรือทำการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปสารสนเทศเป็นข้อมูลที่น่าไปใช้ประโยชน์ได้จริงในทางธุรกิจ
- 5) การประเมินผล (Evaluation) การวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ตั้งไว้ในตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใดเพื่อเปลี่ยนแปลงแก้ไขให้ได้ตามผลลัพธ์ที่ต้องการ
- 6) เผยแพร่ผลวิเคราะห์ (Deployment) การนำเอาข้อมูลที่เป็นผลลัพธ์มาลองปฏิบัติจริงกับธุรกิจ โดยแปลงแนวคิดที่มีให้เกิดเป็นการนำไปใช้ประโยชน์ และติดตามรวบรวมผลที่ได้เพื่อปรับปรุง

2.4 วรรณกรรมที่เกี่ยวข้อง

อุไรวรรณ อินทร์แหยม (2563) ได้ศึกษาวิธีการจำแนกประเภทอาการภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อมูล โดยการทบทวนงานวิจัยที่เน้นในการจำแนกประเภทข้อมูลด้วยเทคนิคการทำเหมืองข้อมูลในโดเมนที่แตกต่างกัน ผู้วิจัยทำการเปรียบเทียบเทคนิคต่าง ๆ ที่ใช้ในงานวิจัยเพื่อเลือกเทคนิคที่เหมาะสมสำหรับการจำแนกประเภทอาการภาวะซึมเศร้า ผลการวิจัยแสดงว่า การใช้เทคนิคการทำเหมืองข้อมูลหลากหลายแบบ เช่น ต้นไม้ตัดสินใจ, ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines), นาอิวเบย์ (Naive Bayes), และโครงข่ายประสาทเทียม (Artificial neural networks) เป็นทางเลือกที่ดี ผู้วิจัยเลือกอัลกอริทึมที่ดีที่สุดจากทั้งหมดเพื่อพัฒนาเป็นแอปพลิเคชันที่สามารถใช้งานได้ง่ายผลลัพธ์แสดงว่าโมเดลที่พัฒนาขึ้นมีความถูกต้องและมีประสิทธิภาพสูงในการนำไปใช้กับงานจริง

อุกฤษฏ์ ศรีสุข และ จารีย์ ทองคำ (2564) ได้ทำการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเกิดโรคได้ผลดีจากการใช้เทคนิค Machine Learning 5 แบบ (Decision Tree C4.5, Naive Bayes, Neural Networks, Random Forest, Deep Learning)

บนข้อมูลผู้ป่วยโรคมะเร็งเต้านม, เบาหวาน, และไฮเปอร์ไทรอยด์จาก UCI Database ทั้งหมด 3 ชุด. ผลการทดลองด้วย 10-fold cross validation แสดงให้เห็นว่า Decision Tree C4.5 มีประสิทธิภาพดีที่สุดในการพยากรณ์โรคไฮโปไทรอยด์, มีค่าความถูกต้อง 99.86%, ค่าความไว 99.85%, และค่าจำเพาะ 100%

อัชฌาพร กว่างสวัสดิ์ และคณะ (2562) ได้ทำการสร้างระบบทำนายระดับความเครียดโดยใช้เทคนิคต้นไม้การตัดสินใจ (Decision tree) ซึ่งทำนายระดับความเครียดในกลุ่มผู้ตอบแบบสอบถามจำนวน 2,000 ตัวอย่าง โดยใช้ภาษา PHP เป็นเครื่องมือในการพัฒนา ระบบ ผลลัพธ์ทำนายมีการแบ่งข้อมูลด้วย Cross-validation Test และได้กฎการทำนายจำนวน 120 กฎ ความถูกต้องของโมเดลร้อยละ 91.10 ระบบทำนายระดับความเครียดถูกพัฒนาขึ้นจากโมเดลที่ได้ และสามารถทำนายระดับความเครียดใน 4 ระดับ เครียดน้อย, เครียดปานกลาง, เครียดมาก, และเครียดมากที่สุด ระบบยังมีสาระความรู้ที่ช่วยในการเข้าใจและนำไปใช้ในชีวิตประจำวัน

นรินทร์ จิวิตัน (2565) ได้ทำการวิเคราะห์ปัจจัยที่เกี่ยวข้องกับการเสียชีวิตจากอุบัติเหตุบนโครงข่ายถนนและเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ที่ใช้เทคนิคต่าง ๆ เช่น เกรเดียนท์ บูตทริ่ง (Gradient Boosting), ต้นไม้ตัดสินใจ (Decision tree), นาอิวเบย์ (Naive Bayes), และแรนดอมฟอเรสต์ (Random Forest) การทบทวนข้อมูลและวิเคราะห์ในกระบวนการทั้งหมดถูกดำเนินการตามขั้นตอน CRISP-DM ผลการวิเคราะห์ข้อมูลแสดงให้เห็นถึงปัจจัยที่เกี่ยวข้องมากที่สุดคือประเภทผู้ใช้ถนน ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ โดยพบว่าเทคนิคนาอิวเบย์มีประสิทธิภาพการจำแนกข้อมูลดีที่สุดในแง่ความถูกต้องร้อยละ 72.23 ค่าประสิทธิภาพโดยรวมร้อยละ 73.35 และค่าเฉลี่ยเรขาคณิตร้อยละ 72.10 ผลลัพธ์นี้สามารถนำไปใช้เพื่อป้องกันอุบัติเหตุทางถนนได้อย่างมีประสิทธิภาพ

จิรโรจน์ ตอสะสุกุล (2564) ได้พยากรณ์การระบาดของโรคไข้เลือดออกในเขตพื้นที่ของสำนักงานป้องกันควบคุมโรคที่ 2 ในจังหวัดพิษณุโลก โดยใช้เทคนิคการทำเหมืองข้อมูลและเปรียบเทียบประสิทธิภาพของแบบจำลองต่าง ๆ โดยใช้ข้อมูลผู้ป่วยไข้เลือดออกและข้อมูลสภาพอากาศจาก 5 จังหวัด ระหว่าง 1 มกราคม 2552 ถึง 31 พฤษภาคม 2564 โดยใช้ 4 เทคนิคการทำเหมืองข้อมูล ได้แก่ ต้นไม้ตัดสินใจ (Decision tree), แบบเบย์ (Bayes), โครงข่ายประสาท (Neural Network), และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ผลลัพธ์แสดงว่าต้นไม้ตัดสินใจเป็นเทคนิคที่เหมาะสมที่สุดในการพยากรณ์การระบาดของโรคไข้เลือดออกมีความแม่นยำ และประสิทธิภาพโดยรวมสูงที่สุดการนำแบบจำลองที่ได้มาใช้จะมี

ประโยชน์ในการวางแผนนโยบายและการรณรงค์เพื่อป้องกันการระบาดของโรคไข้เลือดออก โดยเน้นที่กลุ่มเสี่ยงที่มีการจัดสรรทรัพยากรและงบประมาณจำกัด

รัตนาวดี พานทอง (2565) ได้ทำการสำรวจและจำแนกประเภทข้อมูลการท้องเที่ยวของนักท่องเที่ยวในจังหวัดพะเยา โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เพื่อวิเคราะห์และจำแนกข้อมูลในการสร้างโมเดลต้นแบบ (Model) ที่ใช้ในการจำแนกประเภทข้อมูลของนักท่องเที่ยวในจังหวัดพะเยา โดยใช้โปรแกรม WEKA และอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree Algorithm) เพื่อสร้างโมเดลที่สามารถจำแนกประเภทข้อมูลได้อย่างแม่นยำ ผลลัพธ์ที่ได้จากการวิจัยคือได้รับค่าความถูกต้องที่ดีที่สุดคือ 81.70%

พุทธิพร ธนธรรมเมธิ และเยาวเรศ ศิริสถิตกุล (2561) ได้ทำการเพื่อพัฒนาตัวแบบการพยากรณ์ภาวะข้อเข่าเสื่อมในผู้สูงอายุจากการนำข้อมูลมาปรับสมดุลด้วยวิธี SMOTE และ ADASTN และนำเข้าสู่กระบวนการจำแนกข้อมูลด้วยวิธี one-vs-one และ one-vs-all และ Gentle boost เพื่อพยากรณ์ภาวะโรคข้อเข่าเสื่อม ผลการทดสอบพบว่าการใช้วิธี ADASYN และ one-vs-one ให้ค่าความถูกต้องที่ร้อยละ 97.31% นอกจากนี้การทดสอบโดยใช้ชุดข้อมูลจริงที่ไม่สมดุลจำนวน 232 เรคอร์ด พบว่าการจำแนกถูกต้องทั้งหมดอยู่ที่ 85.78%

สำราญ วานนท์ และรจนา เมืองแสน (2564) ได้ศึกษาโมเดลพยากรณ์ของอาชีพนักศึกษาทางธุรกิจด้วยเทคนิคการสุ่มตัวอย่างเรียนรู้บนชุดข้อมูลที่ไม่สมดุลแบบหลายคลาส ด้วย 1) เทคนิค Over sampling 2) เทคนิค Under sampling 3) เทคนิค SMOTE ผลการทดสอบพบว่า ชุดข้อมูลที่ปรับความไม่สมดุลด้วยเทคนิค Over Sampling และเทคนิคแรนดอมฟอเรส ให้ประสิทธิภาพในการพยากรณ์สูงสุด โดยมีค่าความถูกต้องร้อยละ 67.17 ค่าความแม่นยำ 0.66 ค่าความระสีก 0.67 และค่าเอฟเมเชอร์ 0.66

จากการศึกษาวรรณกรรมที่เกี่ยวข้อง โรคจิตเวชที่มีแนวโน้มนำไปสู่การฆ่าตัวตาย พบว่าปัญหาการฆ่าตัวตายมาจากหลายปัจจัย ปัญหาทางเศรษฐกิจ ปัญหาโรคจิตหรือโรคประสาท ขาดความสมดุลในการรับมือกับความเครียดและปัญหาทางชีวิต รวมถึงความคิดเกี่ยวกับการตายและความรู้สึกที่ว่าตนไม่มีคุณค่าหรือไม่คุ้มค่า การวิเคราะห์ข้อมูลโรคจิตเวชที่มีแนวโน้มนำไปสู่การฆ่าตัวตายด้วย Decision Tree Model เป็นกระบวนการที่ต้องพิจารณาความซับซ้อนของโรคจิตเวชและปัจจัยต่างๆ ซึ่งควรได้รับการดำเนินการโดยผู้เชี่ยวชาญทางด้านจิตเวชและการวิเคราะห์ข้อมูลเพื่อให้ได้ผลลัพธ์ที่ถูกต้องและมีคุณค่าในการตัดสินใจการรักษาและการป้องกัน

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือ และวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นที่เกี่ยวข้องกับการสร้างโมเดลวิเคราะห์ข้อมูลแนวโน้มที่นำไปสู่การฆ่าตัวตายของผู้ป่วยจิตเวชในจังหวัดเชียงใหม่ ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม RapidMiner Studio ในการสร้างโมเดล Decision Tree เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาซึ่งการนำข้อมูลโรคจิตเวช เพื่อหาปัจจัยที่มีผลต่อผลลัพธ์ จากนั้นนำข้อมูลมาทำการแสดงผลแบบ Visualization ในรูปแบบของแดชบอร์ด และสร้างส่วนติดต่อประสานระหว่างผู้ใช้กับระบบการทำเหมืองข้อมูลเผยแพร่ผ่านเว็บไซต์