บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการศึกษาปัจจัยที่ส่งผลต่อการเกิดอุบัติเหตุบนท้องถนน และการเปรียบเทียบประสิทธิภาพตัวแบบการจำแนกข้อมูล บนชุดข้อมูลอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม ปี 2562 – 2567 คณะผู้จัดทำได้ศึกษาหลักการและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

ความหมายของการวิเคราะห์ข้อมูล

หัสพร ทองแดง, พิศุทธิภา เมธีกุล, สมคิด ทุมวงศ์ และแพรว สมบัติใหม่(2559) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเป็นการจำแนกข้อมูลให้เป็นระบบ เพื่อทำให้เกิดความเข้าใจในความหลากหลาย ความหมายและความสัมพันธ์ของ ข้อมูลในบริบทของสังคมและวัฒนธรรมหนึ่ง ๆ ในการวิเคราะห์ข้อมูลเชิงคุณภาพนี้หลังจากที่มีการแยกข้อมูลออกได้เป็นหมวดหมู่แล้ว นักวิจัยต้องใช้วิธีการนำเสนอโดยการพรรณนาให้เห็นสภาพเงื่อนไข กระบวนการ ขั้นตอนการสัมพันธ์ต่าง ๆ รวมทั้งพยายามหาความหมายทาง วัฒนธรรมของปรากฏการณ์ในทัศนะของบุคคลหรือกลุ่มตัวอย่างซึ่งเป็นผู้ให้ข้อมูล และอาจมีความเข้าใจและดีความแตกต่างกันออกไปหรือแตกต่างไปจากของผู้วิจัยเอง การวิเคราะห์ข้อมูลสำหรับการวิจัยเชิงคุณภาพ จึงเป็นเสมือนการทำความเข้าใจในแบบแผนและความหมายของพฤติกรรมมากกว่าที่จะมุ่งหาระดับความมากน้อยหรือระดับความเข้มข้นของความสัมพันธ์ของปรากฏการณ์จึงเป็นการวิเคราะห์ข้อมูลที่อาศัยเครื่องมือทางความคิด (Conceptual tools)มากกว่าที่จะใช้เครื่องมือทางสถิติ และที่สำคัญต้องอาศัยความสามารถของนักวิจัยที่จะอธิบายให้เห็นความสัมพันธ์และความหมายของปรากฏการณ์บนพื้นฐานความเข้าใจในวัฒนธรรมและ สังคมที่เป็นบริบทของปรากฏการณ์ที่ศึกษา

จุมพล หนิมพานิช และวรวลัญช์ โรจนพล (2561) การวิเคราะห์ข้อมูลการวิจัยเชิงคุณภาพในทางรัฐศาสตร์ในภาพรวมมีลักษณะเหมือนกับการวิเคราะห์ข้อมูลเชิงคุณภาพโดยทั่วไป คือ เป็นเรื่องของการนำข้อมูลที่ได้จากการค้นคว้าวิจัยมาจัดกระทำให้เป็นระบบและหาความหมาย แยกแยะองค์ประกอบ รวมทั้งเชื่อมโยงและหาความสัมพันธ์ของข้อมูล

เอื้อมพร หลินเจริญ (2555) การวิเคราะห์ข้อมูลเชิงคุณภาพ เป็นวิธีการสร้างข้อสรุปจากข้อมูลจำนวนหนึ่งซึ่งมักไม่ใช้สถิติในการวิเคราะห์ ทั้งนี้การวิเคราะห์ข้อมูลเชิงคุณภาพ อาจใช้กับการวิจัยเชิงปริมาณที่ผู้วิจัยมีการเก็บรวบรวมข้อมูลเชิงคุณภาพ เช่น แบบสอบถามปลายเปิด การสัมภาษณ์ การสังเกต จดบันทึก สำหรับสาระในบทความนี้ผู้เขียน มุ่งนำ เสนอสาระ เกี่ยวกับเทคนิควิเคราะห์ข้อมูลเชิงคุณภาพที่นักวิจัยนิยมใช้ ได้แก่ การจำแนกหรือ การจัดกลุ่มข้อมูล การเปรียบเทียบเหตุการณ์ การวิเคราะห์ส่วนประกอบการ วิเคราะห์แบบอุปนัย และการวิเคราะห์ข้อมูลที่เป็นเอกสาร

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

ความหมายของการประยุกต์ใช้ข้อมูลทางสถิติ

จากเอกสารประกอบการสอนการ วิจัยสำหรับครู ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า คำว่าสถิติ (Statistics) มาจากภาษาเยอรมัน ว่า Statistics มีรากศัพท์มาจาก Stat หมายถึง ข้อมูลหรือสารสนเทศ ซึ่งจะอำนวยประโยชน์ต่อการบริหารประเทศในด้านต่าง ๆ เช่น การทำ สำมะโนครัวเพื่อจะทราบจำนวนพลเมืองในประเทศทั้งหมดในสมัยต่อมา คำว่า สถิติ ได้หมายถึงตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวมเช่น จำนวนผู้ประสบอุบัติเหตุบนท้องถนน อัตราการเกิดของเด็กทารก ปริมาณน้ำฝนในแต่ละปีเป็นต้น สถิติในความหมายที่กล่าวมานี้ เรียกอีกอย่างหนึ่งว่า ข้อมูลทางสถิติ (Statistica data)

ค่ากึ่งกลาง (Median) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางวิธีหนึ่งที่ใช้การเรียงค่าการสะท้อนของจุดภาพจากค่าน้อยที่สุดไปหาค่ามากที่สุด โดยค่ากึ่งกลางเป็นค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งชุด ค่ากึ่งกลางจึงเป็นตัวแทนค่าการสะท้อนของจำนวนจุดภาพ ทั้งหมดในช่วงคลื่นหนึ่งๆ ที่แสดงให้ทราบว่ามีจำนวนจุดภาพที่มีค่าการสะท้อนมากกว่าและ น้อยกว่าค่ากึ่งกลางอยู่ประมาณร้อยละ 50

ค่าฐานนิยม (Mode) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางอีกวิธีหนึ่ง โดยดูจากจำนวนความถี่ของค่าการสะท้อนซึ่งมีความถี่สูงที่สุด นิยมนำมาใช้กับข้อมูลที่เป็นนาม

บัญญัติเช่น ค่าของประเภทข้อมูลหลังจากการจำแนกประเภทแล้วถือเป็นค่าการสะท้อนแสดงการใช้ที่ดินประเภทต่าง ๆ ไม่ใช่ค่าการสะท้อนของวัตถุอีกต่อไป

ความเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D.) เป็นการวัดการกระจาย ที่นิยมใช้มากที่สุดการคำนวณใช้วิธียกกำลังสองของผลต่างระหว่างค่าการสะท้อนของทุก จุดภาพในแต่ละช่วงคลื่นกับค่าเฉลี่ยเลขคณิตของช่วงคลื่นนั้น

ภัทรธิดา ผลงาม (2558) ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า สถิติ คือ ตัวเลขต่าง ๆ ที่ได้มีการรวบรวมขึ้นเพื่อบอกข้อเท็จจริงเกี่ยวกับคุณสมบัติหรือลักษณะบางสิ่งบางอย่างที่สามารถแสดงออกเป็นตัวเลขได้สถิติแบ่งเป็นประเภทใหญ่ๆ ได้ 2 ประเภท ดังนี้ คือ

1) **สถิติพรรณนา** (Descriptive statistics) คือสถิติที่บรรยายถึงลักษณะของข้อมูลเฉพาะกลุ่มนั้นๆ โดยไม่สรุปอ้างอิงไปยังประชากรกลุ่มอื่นๆ สถิติประเภทนี้นิยมศึกษาในกลุ่มเล็กหรือกลุ่มใหญ่ก็ได้สถิติประเภทนี้เป็นสถิติที่บรรยายลักษณะของข้อมูล เช่น ค่าเฉลี่ย ร้อยละ มัธยฐาน พิสัย ส่วนเบี่ยงเบนมาตรฐาน และการวัดความสัมพันธ์ต่าง ๆ เช่น สหสัมพันธ์

2) **สถิติอ้างอิงหรือสถิติอนุมาน** (Inferential or inductive statistics) คือ สถิติที่นำ ค่าสถิติพรรณนามาสรุปอ้างอิงไปยังประชากรหรือเป็นสถิติที่ศึกษาจากกลุ่มตัวอย่างแล้วนำผล อ้างอิงไปยังกลุ่มประชากร เช่น การศึกษาความคิดเห็นต่อการเลือกตั้งของประชาชนไม่จำเป็นต้องศึกษาจากประชาชนทุกคน แต่สามารถเลือกศึกษาจากประชาชนบางกลุ่มซึ่งจะเป็น ตัวแทนของประชาชนทั้งหมด แล้วจึงสรุปว่าประชาชนมีความคิดเห็นอย่างไรต่อการเลือกตั้งได้ดังนั้นในการใช้สถิติอ้างอิงนี้จึงจำเป็นต้องเลือกลุ่มตัวอย่างที่เหมาะสม สถิติอ้างอิงนี้ก่อนนำไป อ้างอิงกลุ่มประชากรต้องมีการทดสอบทางสถิติก่อนทุกครั้งจึงสามารถอ้างอิงประชากรได้อย่างมีประสิทธิภาพ

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

ความหมายการแสดงผลข้อมูล

กานต์ ยงศิริวิทย์ และภาคภูมิ ชัยศิริประเสริฐ (2560) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลควรนำเสนอในรูปแบบที่น่าสนใจ และเข้าใจง่าย ดังนั้นเครื่องมือควรที่จะ

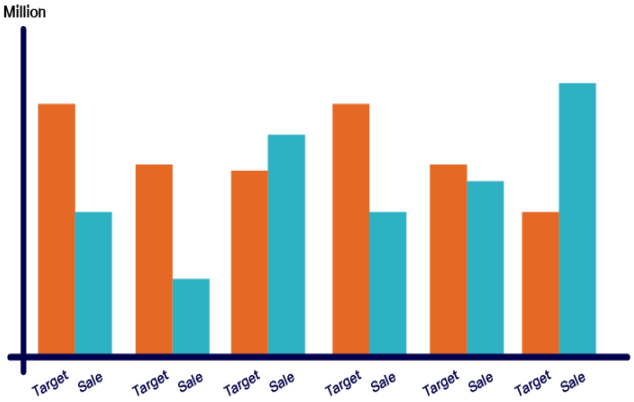
สามารถนำเสนอข้อมูลในรูปแบบของกราฟและแสดงผลลัพธ์บนแผนที่ในขั้นพื้นฐานได้ โดยไม่ควรที่จะต้องมีการติดตั้งซอฟต์แวร์หรือตัวช่วยเพิ่มเติม

ชนาธิป ชื่นมนัส (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) คือวิธีการที่ใช้ในการนำข้อมูลที่เป็นนามธรรมมาแสดงให้เป็นภาพเพื่อเพิ่มประสิทธิผลในการติดต่อสื่อสารหรือทำให้เข้าใจข้อมูลได้ง่ายขึ้น การใช้เทคนิค Visualzation มีหลายวิธีซึ่งแต่ละวิธีเหมาะสมที่จะใช้ในการแสดงข้อมูลที่แตกต่างกัน การเลือกใช้เทคนิคใดนั้นขึ้นกับองค์ประกอบหลายอย่าง เช่น ชนิดของข้อมูลที่จะนำมาแสดง วิธี ที่ต้องการแสดงข้อมูล เป็นต้น ในปัจจุบันมีงานวิจัยที่นำเอาเทคนิค Visualization มาใช้โดยมีจุดประสงค์เพื่อเพิ่มประสิทธิผลในการแสดงข้อมูลการติดต่อสื่อสารภายในเครือข่ายคอมพิวเตอร์ และช่วยในการวิเคราะห์ความผิดปกติที่เกิดขึ้นในเครือข่าย

อาทิตย์ สิทธิบรรเจิด (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพโดยผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่นๆ หรือจะกล่าวว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนามาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ การเลือกรูปแบบ Visudization ให้เหมาะสมกับข้อมูลในปัจจุบันเป็นยุคเทคโนโลยีเข้าถึงทุกคนทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่ายและรวดเร็วมากขึ้น คนที่นำเสนอ ข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visudization ขึ้นมา Data Visuaization เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ซึ่งอาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่นๆ อีกมากมาย เพื่อให้เข้าใจได้

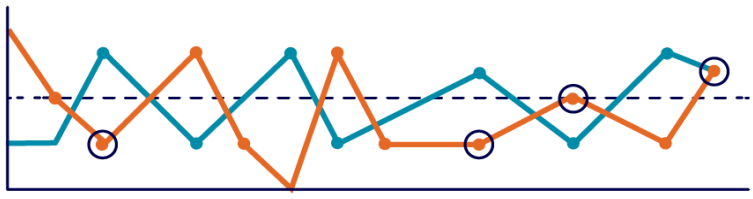
1) แผนภูมิแท่ง (Bar chart) คือกราฟที่ประกอบด้วยรูปสี่เหลี่ยมผืนผ้า (เรียกว่าแท่ง) จำนวนหนึ่ง โดยมีลักษณะเป็นแท่งสูงหรือยาวที่เปลี่ยนแปลงตามขนาด แต่มีความกว้างเท่ากันหมด เราอาจเรียงแท่งเหล่านี้ในทางตั้งหรือทางนอนก็ได้ โดยเว้นระยะช่องว่างตามสมควรและจะต้องเขียนโครงเรื่องจำแนกแต่ละแท่งให้ชัดเจนด้วย แผนภูมิแท่งอาจจะมีการระบายสีหรือแรเงาเพื่อให้ดูเด่น และในกรณีที่มีการเปรียบเทียบกันหลายแท่ง เช่น แผนภูมิแท่ง ซับซ้อนหรือเชิงประกอบ จำเป็นจะต้องระบายสีหรือแรเงา เพื่อจำแนกความ

แตกต่างของแผนภูมิแต่ละชุดที่นำมาเปรียบเทียบกันนั้น วัตถุประสงค์ของการนำเสนอข้อมูลในรูปแผนภูมิแห่งเพื่อแสดงความเปลี่ยนแปลงและการเปรียบเทียบข้อมูล



ภาพที่ 2.1 แผนภูมิแท่ง (Bar chart)

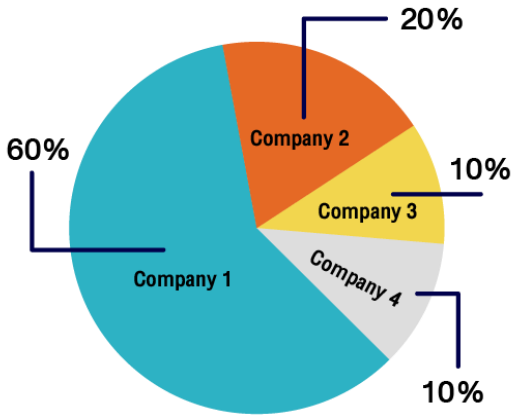
2) **แผนภูมิเส้น (Line Charts)** แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแห่งข้อมูลเป็นจุดบนแผนภูมินั่นเอง แผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้ เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดูแนวโน้มเป็นต้น



ภาพที่ 2.2 แผนภูมิเส้น (Line Charts)

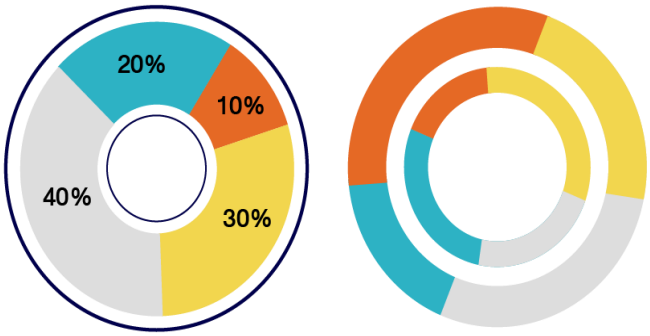
3) **แผนภูมิวงกลม (Ple Chart)** แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่ายและสวยงามแต่ในทางกลับกันอาจจะดูมากในเรื่องของการประมาณขนาดของแต่ละชิ้นยิ่งถ้ามีจำนวนชิ้นมากจะ

ยิ่งแยกยากเพราะต้องใช้หลายสีในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share) ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น



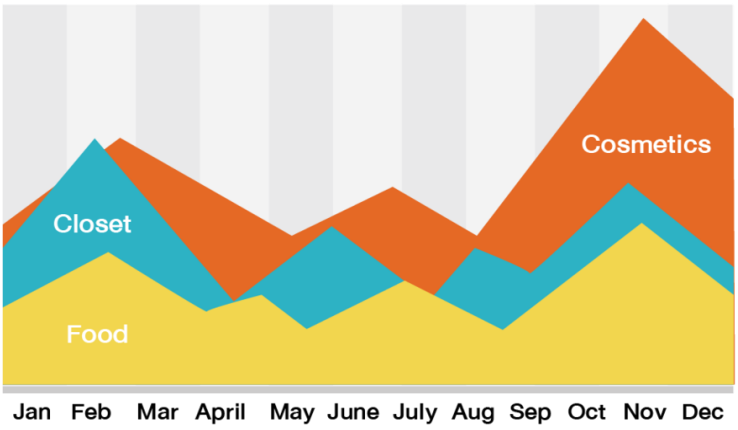
ภาพที่ 2.3 แผนภูมิวงกลม (Ple Chart)

4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลายๆ ชั้น



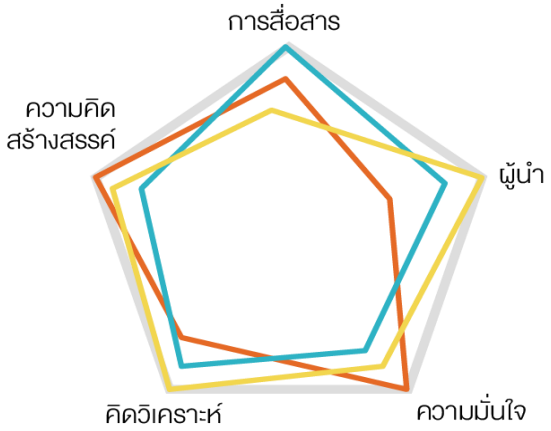
ภาพที่ 2.4 แผนภูมิโดนัท (Doughnut Charts)

5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้นเหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่เดือนตามหมวดหมู่ต่าง ๆไล่ไป เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



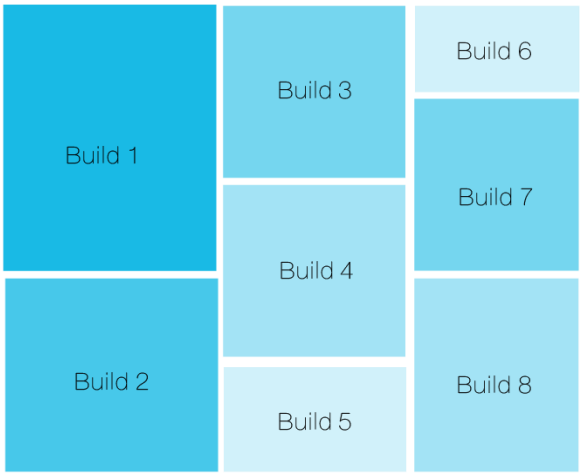
ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

6) **แผนภูมิเรดาร์ (Rodar Charts)** มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความ ต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อน จุดแข็งของข้อมูล เช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อน จุดแข็งของแต่ละคน เป็นต้น



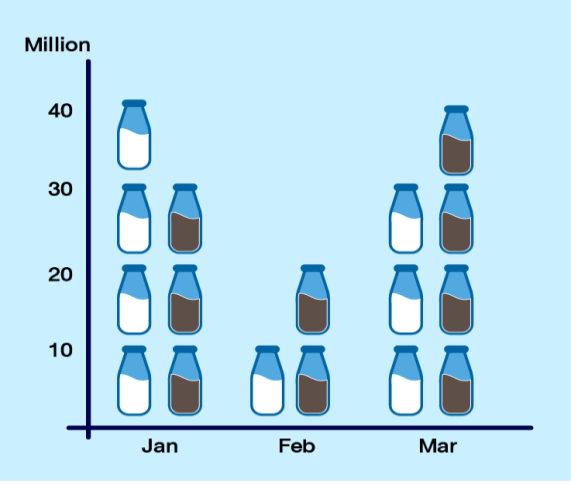
ภาพที่ 2.6 แผนภูมิเรดาร์ (Rodar Charts)

7) **แผนภูมิต้นไม้ (Tree Maps)** คือ การนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่แสดงผลได้ในแบบลำดับชั้นเหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้



ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอนและแทนตั้ง แต่เลือกใช้แคนตั้ง แต่เลือกใช้: ภาพหรือโอคอนแทนจำนวนขุองสิ่งของนั้นๆ เช่น การแสดงผลจำนวนของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละเดือนซึ่งมีการนำเสนอลักษณะคล้ายกับคราฟแห่ง แต่เปลี่ยนจากแห่งเป็นรูปกราฟิกของนม 2รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะ นี่ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคนอาจจะแทนด้วยภาพไอคอนคนหรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงิน หรือแบงค์ ก็จะทำให้คนดูเข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคย



ภาพที่ 2.8 แผนภูมิรูปภาพ (Picture Graph)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับโมเดลวิเคราะห์ปัจจัยการเกิดอุบัติเหตุ

2.2.1.1 เทคนิค Chi-Square (วรการ ใจดี และ นพณัฐ วรรณภีร์,2020) เป็นเทคนิคการประเมินค่าของคุณลักษณะโดยการใช้การคำนวนค่า Chi-Square ทางสถิติ เพื่อศึกษาว่าการแจกแจงความถี่ของตัวแปรคุณลักษณะเป็นไปตามรูปแบบที่กำหนดไว้หรือไม่ สมการของ Chi-Square คือ

$$x^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

โดยที่ O_1, O_2 ..... O_n เป็นความถี่ของตัวแปรที่ได้จากการศึกษา

E_1, E_2 .... E_n เป็นความถี่ที่คาดหวัง(หรือความถี่ที่ควรจะเป็น)

หลักการของการทดสอบไคสแควร์ (มหาวิทยาลัยราชภัฏนครปฐม) คือ เปรียบเทียบความถี่ที่ได้จากตัวอย่าง หรือเรียกว่าความถี่ที่สังเกตได้ (observed frequency) แทนด้วยสัญลักษณ์ O_i กับความถี่ตามทฤษฎี หรือเรียกว่าความถี่คาดหวัง (expected frequency) แทนด้วยสัญลักษณ์ E_i ว่ามีความแตกต่างกันมากเกินไปหรือไม่ สมมติว่าข้อมูลที่ศึกษามี k กลุ่ม เลือกตัวอย่างมาจำนวน n จะได้ความถี่ที่สังเกตได้ O_i และ ความถี่คาดหวัง E_i ตามตารางที่ 2.1

ตารางที่ 2.1 ตารางแสดงหลักการของการทดสอบไคสแควร์ (มหาวิทยาลัยราชภัฏนครปฐม)

ตัวแปร	1	2	3	....	k	รวม
ความถี่ที่สังเกตได้	O_1	O_2	O_3	....	O_k	n
ความถี่คาดหวัง	E_1	E_2	E_3	....	E_k	n

หมายเหตุ 1. การหาความถี่คาดหวัง E_i ต้องทราบสัดส่วนที่ตั้งไว้ในสมมติฐานแล้วคูณกับผลรวมความถี่ทั้งหมด นั้นคือ $E_i = np_i$ หรือ $E_i = np_{io}$

2. $\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i = n$

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

โดยที่ X^2 แทนตัวสถิติทดสอบไคสแควร์ที่องศาแห่งความเป็นอิสระ k–1

k แทนจำนวนกลุ่มข้อมูล

ค่าวิกฤต X^2 ที่องศาแห่งความเป็นอิสระ k–1

หมายเหตุ 1.ถ้าความถี่ที่ได้จากการสังเกตใกล้เคียงกับความที่คาดหวัง X^2 จะมีค่าน้อย

2.ถ้าความถี่ที่ได้จากการสังเกตแตกต่างกับความที่คาดหวัง X^2 จะมีค่ามาก

3.การทดสอบ X^2 จะมีประสิทธิภาพ n ควรมีขนาดมากพอที่ทำให้ค่าคาดหวังในแต่ละช่องมีค่าไม่ต่ำกว่า 5

4.เพื่อสะดวกในการคำนวณสามารถใช้ $\chi^2 = \Sigma \frac{O^2}{E} - n$ ได้

ในการทดสอบไคสแควร์มีข้อตกลงเบื้องต้น ดังนี้

1. ตัวอย่างแต่ละชุดเป็นตัวอย่างที่ได้จากการสุ่ม

2. ผลที่เกิด (outcomes) ของตัวอย่างเป็นอิสระต่อกัน (mutually independence)

3. ค่าสังเกตแต่ละต้องถูกจัดให้อยู่ในชั้นใดชั้นหนึ่งเท่านั้น

การทดสอบไคสแควร์แบ่งได้ 2 ประเภท คือ การทดสอบไคสแควร์กับข้อมูลจำแนกทาง เดียวและการทดสอบไคสแควร์กับข้อมูลจำแนกสองทาง

1. การทดสอบไคสแควร์กับข้อมูลจำแนกทางเดียว การทดสอบไคสแควร์กับข้อมูลจาแนกทางเดียว เรียกว่าการทดสอบภาวะสารูปสนิทดี (Goodness of fit test) แบ่งออกเป็น 2 ลักษณะ ดังนี้

1.1 การทดสอบสัดส่วนของประชากร k กลุ่ม จุดประสงค์ต้องการทดสอบสัดส่วน ของความถี่ในแต่ละกลุ่มว่าเท่ากันหรือไม่ หรือทดสอบสัดส่วนของความถี่ในแต่ละกลุ่มว่าเท่ากับ ค่าคงที่หรือไม่ เช่น สัดส่วนของนักเรียนที่ขาดเรียนในแต่ละวันเท่ากันหรือไม่ สัดส่วนของนักเรียน ที่ชอบเรียนในแต่ละรายวิชาเท่ากันหรือไม่ สัดส่วนของลูกค้าที่ใช้โทรศัพท์ยี่ห้อ Samsung:i-phone: Nokia เท่ากับ 0.6:0.3:0.2 หรือไม่ ซึ่งถ้าจะทดสอบความเท่ากันของสัดส่วนของประชากร 2 กลุ่ม ใช้ตัวสถิติทดสอบ Z แต่การทดสอบความเท่ากันของสัดส่วนของประชากรมากกว่า 2 กลุ่ม ใช้ตัว สถิติทดสอบ X^2 ในการทดสอบ

กำหนด p_i แทนสัดส่วนของประชากรกลุ่มที่ i เมื่อ i = 1,2,3,..,k

สมมติฐานเชิงสถิติ

$$H_0 : p_1 = p_2 = p_3 = \ldots = p_k = \frac{1}{k}$$

$$H_1 : p_I \neq p_j \text{ อย่าง 1 คู่ เมื่อ } i \neq j$$

หรือถ้ามีค่าสัดส่วนที่คาดหวัง (p_0) สมมติฐานเชิงสถิติเป็น ดังนี้

$$H_0 : p_1 = p_{10}, p_2 = p_{20}, p_3 = p_{30}, \ldots p_k = p_{k0} \text{ หรือ}$$

$$H_0 : p_1 : p_2 : p_3 : \ldots p_k = p_{10} : p_{20} : p_{30} : \cdots p_{k0}$$

$$H_1 : p_I \neq p_{i0} \text{ อย่างน้อย 1 คู่ เมื่อ } i = 1, 2 \ldots, k$$

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

เมื่อ O_i แทนความถี่ที่ได้จากการสังเกต

E_i แทนความถี่คาดหวังหรือตามทฤษฎี และ $E_i = np_i$

k แทนจำนวนกลุ่ม

n แทนจำนวนตัวอย่าง

ค่าวิกฤต $\chi^2_{1-\alpha,k-1}$

ตัวอย่าง ห้างสรรพสินค้า B ได้ผลิตขนมเค้กยี่ห้อ B ออกจำหน่ายตามสาขาต่าง ๆ ของตน โดย จะขายในราคาที่ถูกกว่ายี่ห้ออื่น ๆ ที่มีชื่อเสียงอีก 4 ยี่ห้อ คือ C, D, E และ F ที่ทางห้างรับมาขายทางห้างต้องการทราบว่าสัดส่วนของลูกค้าที่ชอบขนมเค้กยี่ห้อ B เท่ากับสัดส่วนของลูกค้าที่ชอบขนม เค้กยี่ ห้อ C, D, E และ F หรือไม่ จึงสุ่มลูกค้ามา 100 คน ให้ชิมขนมเค้กทั้ง 5 ยี่ห้อ แล้วให้ลูกค้าบอก ว่าชอบยี่ห้อใดมากที่สุด ได้ผลดังนี้

ค่าสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$
$$= 4.4$$

ค่าวิกฤติ $\chi^2_{1-\alpha,k-1} = \chi^2_{0.95,4} = 9.49$

เนื่องจากค่าสถิติทดสอบ $X^2 = 4.4$ มากกว่า $X^2_{0.95,4} = 9.49$ อยู่ในบริเวณยอมรับ H_0 หมายความว่าสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้อ B เท่ากับสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้ออื่นๆอีก 4 ยี่ห้อ ที่ระดับนัยสำคัญ 0.05

1.2 การทดสอบการแจกแจงของประชากร ในเรื่องการทดสอบสมมติฐานนั้นตัว สถิติทดสอบที่ใช้ขึ้นอยู่กับการแจกแจงของข้อมูล หากเราทราบว่าข้อมูลมีการแจกแจงลักษณะใด จะทำให้สามารถเลือกใช้ตัวสถิติทดสอบได้ถูกต้องตามข้อจำกัดของวิธีการทดสอบนั้น ๆ ซึ่งการ X^2 ทดสอบ เป็นวิธีการหนึ่งที่จะช่วยทำให้ทราบว่าข้อมูลชุดนั้นมีการแจกแจงแบบใด โดยที่

สมมติฐานเชิงสถิติ

H_0 : ข้อมูลมีการแจกแจงแบบที่ต้องการ

H_1 : ข้อมูลไม่มีการแจกแจงแบบที่ต้องการ

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

เมื่อ o_i แทนความถี่ที่ได้จากการสังเกต

E_i แทนความถี่คาดหวังหรือตามทฤษฎี และ $E_i = np_i$

k แทนจำนวนกลุ่ม

ค่าวิกฤต $\chi^2_{1-\alpha,k-1-m}$

เมื่อ m แทนจำนวนพารามิเตอร์ที่ประมาณ (จาก H_0)

ตัวอย่าง ในการทดสอบการกระจายของกรดยูริคของคนไข้ 250 คน ว่ามีการแจกแจงแบบปกติด้วยค่าเฉลี่ย 5.74 และ ส่วนเบี่ยงเบนมาตรฐาน 2.01 หรือไม่ที่ระดับนัยสำคัญ 0.05 ตารางข้อมูลเป็นดังนี้

กรดยูริค	จำนวนคนไข้
< 1	1
1 - 1.99	5
2 - 2.99	15
3 - 3.99	24
4 - 4.99	43
5 - 5.99	50
6 - 6.99	45
7 - 7.99	30
8 - 8.99	29
9 - 9.99	10
≥ 10	5

วิธีทำ กำหนด X เป็นตัวแปรสุ่มแทนปริมาณกรดยูริค

สมมติฐานเชิงสถิติ

H_0 : X มีการแจกแจง N(5.74 , 4.04)

H_1 : X ไม่มีการแจกแจง N(5.74,4.04)

กรดยูริค	ขอบเขตที่แท้จริง	$Z=\frac{(X-\mu)}{\sigma}$	P_i	O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
< 1	< 0.995	< - 2.36	0.0091	1	2.275	
1 - 1.99	0.995 - 1.995	(-2.36) - (-1.86)	0.0223	5	5.575	0.44
2 - 2.99	1.995 - 2.995	(-1.86) - (-1.37)	0.0539	15	13.475	0.17
3 - 3.99	2.995 - 3.995	(-1.37) - (-0.87)	0.1069	24	26.725	0.28
4 - 4.99	3.995 - 4.995	(-0.87) - (-0.37)	0.1635	43	40.875	0.11
5 - 5.99	4.995 - 5.995	(-0.37) - (0.13)	0.1960	50	49.000	0.02
6 - 6.99	5.995 - 6.995	(0.13) - (0.62)	0.1807	45	45.175	0.006
7 - 7.99	6.995 - 7.995	(0.62) - (1.12)	0.1362	30	35.050	0.48
8 - 8.99	7.995 - 8.995	(1.12) - (1.62)	0.788	22	19.700	0.29
9 - 9.99	8.995 - 9.995	(1.62) - (2.12)	0.356	10	8.900	0.26
≥ 10	> 9.995	> 2.12	0.170	5	4.250	
รวม			1000	250	250	2.03

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$
$$= 2.03$$

ค่าวิกฤต $\chi^2_{1-\alpha,k-1-m} = \chi^2_{0.95,8} = 15.5$ (k=9 และ m=0)

เนื่องจากค่าสถิติทดสอบ X^2 = 2.03 มีค่าน้อยกว่า $X^2_{0.95,8}$=15.5 จึงไม่สามารถ ปฏิเสธสมมติฐาน H_0 หมายความว่าการกระจายของกรดยูริคนี้มีการแจกแจงแบบปกติด้วยค่าเฉลี่ย 5.44 และส่วนเบี่ยงเบนมาตรฐาน 2.01 ที่ระดับนัยสำคัญ 0.5

หมายเหตุ 1. ในกรณีที่มีความถี่คาดหวัง E_i บางค่าต่ำกว่า 5 อาจจะรวม ความถี่ของกลุ่มที่มีค่า E น้อยเข้ากับความถี่ของกลุ่มที่อยู่ติดกัน หรือรวมกลุ่มที่มี ลักษณะใกล้เคียงกันเป็นกลุ่มเดียวกัน เพื่อให้ E มีค่ามากกว่า 5 ซึ่งจะทำให้จำนวนกลุ่มลดลง (k ลดลง) จึงมีผลทำให้ df ลดลงด้วย

2. ในกรณีที่ระดับของข้อมูลมี 2 ระดับ (k = 2) องศาเท่าความเป็นอิสระของการทดสอบ จะเหลือเพียง 1 จะมีผลทำให้ค่า X^2 มีจำนวนสูงกว่าที่ควรจะเป็น Frank Yate เสนอให้ปรับค่า X^2 เป็นดังนี้

$$\chi^2 = \sum_{i=1}^{2} \frac{(|O_i - E_i| - .05)^2}{Ei}$$

แต่ถ้าขนาดตัวอย่าง $n \geq 50$ ก็ไม่จำเป็นตัวปรับค่า χ^2

2. การทดสอบไคสแควร์กับข้อมูลจำแนกสองทาง เรียกว่าการทดสอบความเป็นอิสระ (testing of independence) เป็นวิธีการที่ใช้ทดสอบสมมติฐานการวิจัยที่เกี่ยวกับความเป็นอิสระของตัวแปรสองตัวหรือความสัมพันธ์ของตัวแปรสองตัวว่าเกี่ยวข้องกันหรือไม่ เช่น สนใจศึกษาความสัมพันธ์ของรายได้กับอายุ ความสัมพันธ์ของประเภทสินค้าที่ขายได้กับรายรับ ความเป็นอิสระของน้ำหนักเด็กแรกเกิดกับปริมาณการสูบบุหรี่ของแม่ขณะตั้งครรภ์ หรือความสัมพันธ์ ของยี่ห้อรถยนต์ที่กับอาชีพ เป็นต้น ซึ่งข้อมูลจำแนกสองทางมีลักษณะเป็น ดังนี้

2.1 ลักษณะข้อมูลจำแนกสองทาง ในกรณีที่ข้อมูลเรื่องใดเรื่องหนึ่งถูกจำแนกโดยตัวแปร 2 ตัวแปร ซึ่งโดยทั่วไปข้อมูลที่นำมาทดสอบจะเป็นข้อมูลจำแนก 2 ทาง โดยตัวแปรที่ 1 จะ แบ่งเป็น r กลุ่ม (row) และ ตัวแปรที่ 2 จะแบ่งเป็น c กลุ่ม (column) เราจะเรียกตารางนี้ว่า ตารางการณ์จรขนาด rxc (rxc contingency table) โดยข้อมูลที่วิเคราะห์อยู่ในรูปแบบตารางดังนี้

สมมติฐานเชิงสถิติ

H_0 : ตัวแปรที่ 1 เป็นอิสระกับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 ไม่เป็นอิสระกับตัวแปรที่ 2

หรือ

H_0 : ตัวแปรที่ 1 ไม่มีความสัมพันธ์กับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 มีความสัมพันธ์กับตัวแปรที่ 2

หรือ

H_0 : ตัวแปรที่ 1 ไม่มีผลต่อตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 มีผลต่อตัวแปรที่ 2

หรือ

H_0 : ตัวแปรที่ 1 ไม่ขึ้นกับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 ขึ้นกับตัวแปรที่ 2

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

โดยที่ χ^2 ~ไคสแควร์ที่องศาแห่งความเป็นอิสระ $(r-1)(c-1)$

O_{ij} แทนความถี่ที่ได้จากการสังเกตใน cell(i,j)

E_{ij} แทนความถี่คาดหวังใน cell(i,j)

หมายเหตุ การหา E_{ij} นั้น ต้องหาค่าความน่าจะเป็นที่จะเกิดเหตุการณ์ในแต่ละแถวในสดมภ์ที่ j นั้น คือถ้า H_o เป็นจริงแล้วจะมีค่าเท่ากับ $\frac{c_i}{n}$ และ จำนวนความถี่ที่ได้จากการสังเกตใน cell(i,j) ควรจะมีค่าใกล้เคียงกับขนาดของตัวอย่างชุดที่ i (r_1) คูณกับ $\frac{c_i}{n}$ ดังนั้น

$$E_{ij} = \frac{r_i c_i}{n}$$

เมื่อ r_i แทนความถี่รวมของตัวแปรที่ 1 กลุ่ม i ; i = 1,... ,x

c_j แทนความถี่รวมของตัวแปรที่ 2 กลุ่ม j : j = 1,.... ,c

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$
$$= \frac{(328-233)^2}{233} + \frac{(172-267)^2}{267} + \frac{(138-233)^2}{233} + \frac{(362-267)^2}{267}$$
$$= 145.07$$

ค่าวิกฤต $\chi^2_{1-\alpha,(r-1)(c-1)} = \chi^2_{0.95,1} = 3.84$

เนื่องจากค่าสถิติทดสอบ X^2 = 145.07 อยู่ในบริเวณปฏิเสธ H_0 หมายความว่า ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้ ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง บริษัทแห่งหนึ่งต้องการเปิดทำงานในช่วงวันเสาร์และอาทิตย์ โดยคิดว่าพนักงาน ต้องการมีรายได้เสริม เพื่อการตัดสินใจที่ถูกต้องจึงทำการเก็บรวบรวมข้อมูลจากคนงานในบริษัท จำนวน 250 คน โดยแบ่งรายได้ออกเป็น 2 ช่วงคือ รายได้ต่า กว่า 10,000 บาทและตั้งแต่ 10,000 บาทขึ้นไป และความคิดเห็นเป็น 3 ระดับ ได้ผลดังนี้

รายได้	ความคิดเห็น			รวม
	เห็นด้วย	ไม่มีความเห็น	ไม่เห็นด้วย	
ต่ำกว่า 10,000	68	12	110	190
ตั้งแต่ 10,000 ขึ้นไป	48	2	10	60
รวม	116	14	120	250

อยากทราบว่าความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นอยู่กับรายได้หรือไม่ที่ระดับนัยสำคัญ 0.05

วิธีทำ สมมติฐานเชิงสถิติ

H_0 : ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ไม่ขึ้นกับรายได้

H_i : ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นกับรายได้

ความถี่คาดหวัง

$$E_{11}=\frac{190\times166}{250}=88.16 \qquad E_{12}=\frac{190\times14}{250}=10.64 \qquad E_{13}=\frac{190\times120}{250}=91.20$$

$$E_{21}=\frac{60\times166}{250}=27.84 \qquad E_{22}=\frac{60\times14}{250}=3.36 \qquad E_{23}=\frac{60\times120}{250}=28.80$$

ตัวสถิติทดสอบ

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$=\frac{(68-88.16)^2}{88.16}+\frac{(12-10.64)^2}{10.64}+\frac{(110-91.20)^2}{91.20}+\ldots+\frac{(10-28.80)^2}{28.80}$$

$$=36.08$$

ค่าวิกฤต $\chi^2_{1-\alpha,(r-1)(c-1)}=\chi^2_{0.95,2}=5.99$

เนื่องจากค่าสถิติทดสอบ X^2 = 36.08 อยู่ในบริเวณปฏิเสธ H_0 หมายความว่าความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นอยู่กับรายได้ ที่ระดับนัยสำคัญ 0.05

2.2 การทดสอบความเป็นอิสระกรณีตารางการณ์จรขนาด 2x2 ในกรณีที่ตัวแปรที่ 1 และ 2 แบ่งออกเป็นตัวแปรละ 2 กลุ่ม ความถี่นั้นจะอยู่ในรูปตาราง ตามตารางที่ 2.2

ตารางที่ 2.2 ตารางการณ์จรขนาด 2x2 (มหาวิทยาลัยราชภัฏนครปฐม)

ตัวแปรที่ 1	ตัวแปรที่ 2		รวม
	1	2	
1	0_{11}	0_{12}	r_1
2	0_{21}	0_{22}	r_2
รวม	c_1	c_2	n

เมื่อตารางการณ์จร มีขนาด 2X2 ตัวสถิติทดสอบ 2 ใช้หลักการเช่นเดิม หรืออาจคำนวณหาค่าตัวสถิติทดสอบ 2 ได้สะดวก และง่ายขึ้นโดยใช้สูตร ดังนี้

$$\chi^2 = \frac{n(O_{11}O_{22} - O_{12}O_{21})^2}{r_1 r_2 c_1 c_2}$$

ค่าวิกฤต $\chi^2_{1-\alpha,(r-1)(c-1)}$

หมายเหตุ ข้อจำกัดของการทดสอบ X^2 สำหรับตารางการณ์จร 2x2

1. $E_{ij} \geq 5$ ทุก ๆ ค่า I,j เมื่อ i = 1,2,.....r และ j = 1,...,c

2. เนื่องจากการแจกแจง X^2 เป็นการแจกแจงของตัวแปรเชิงสุ่มแบบต่อเนื่อง แต่ความถี่ที่ได้มักจะเป็นตัวแปรเชิงสุ่มแบบไม่ต่อเนื่อง อย่างไรก็ดีการแจกแจงแบบ X^2 เป็นการแจกแจงโดยประมาณที่ดีของค่า x^2 ที่คำนวณได้ ถ้า df > 1 ในกรณีที่ df = 1 เช่น ตารางขนาด 2x2 เราจำเป็นต้องแก้ไขโดยใช้ Yates correction for continuity ดังนี้

$$\chi^2_{\text{corrected}} = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}$$

หรือ อาจคำนวณหาโดยใช้สูตรอย่างง่าย ได้ดังนี้

$$\chi^2 = \frac{n(|O_{11}O_{22} - O_{12}O_{21}| - \frac{1}{2}n)^2}{r_1 r_2 \; c_1 c_2}$$

โดยที่ $\chi^2 \sim \chi^2$ df = (r-1) (c-1) ถ้า $n \geq 50$ ไม่ต้องปรับ χ^2

ตัวอย่าง บริษัทผลิตมือถือแห่งหนึ่งต้องการทราบว่าความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้หรือไม่ จึงทำการเก็บรวบรวมข้อมูลจากผู้ใช้โทรศัพท์มือถือจำนวน 1,000 คน เป็นเพศชาย 500 คน และ หญิง 500 คน ได้ผลดังนี้

เพศ	สีของโทรศัพท์มือถือ		รวม
	สีดำ	สีอื่น ๆ	
ชาย	328	172	500
หญิง	138	362	500
รวม	466	534	1,000

จงคำนวณค่าสถิติทดสอบสำหรับตารางการณ์จรขนาด 2x2

วิธีทำ ตัวสถิติทดสอบ

$$\chi^2 = \frac{n(O_{11}O_{22} - O_{12}O_{21})^2}{r_1 r_2 c_1 c_2}$$

$$= \frac{1000((328)(362) - (172)(138))^2}{(500)(500)(466)(534)}$$

$$= \frac{1000(118736 - 23736)^2}{6.2211 \times 10^{10}}$$

$$= 145.07$$

2.2.1.2 ทฤษฎีเกี่ยวกับเทคนิคการหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Based Feature Selection : CFS) (วรการ ใจดี และ นพณัฐ วรรณภีร์,2020) เป็นเทคนิคในการเลือกคุณสมบัติของคุณลักษณะโดยใช้การพิจารณาบนพื้นฐานความสัมพันธ์ของกลุ่มคุณลักษณะที่ได้จากการประเมินค่าความสามารถในการคาดการณ์ และยังสามารถจัดการกับคุณลักษณะที่ไม่เกี่ยวข้อง โดย CFS จะจัดอันดับกลุ่มย่อยของมิติข้อมูลและทำการคัดเลือกกลุ่มย่อยของมิติข้อมูลที่มีความสัมพันธ์กันสูงกับคลาสและไม่มีความสัมพันธ์กับคลาสอื่นๆ สมการประเมินกลุ่มย่อยของมิติข้อมูลแบบ CFS คือ

$$M_s = \frac{kr_{cf}}{\sqrt{k+k(k-1)\overline{r_{cf}}}}$$

โดยที่ M_s คือค่าที่ค้นหาได้ของมิติข้อมูลกลุ่มย่อย S ซึ่งประกอบด้วยมิติข้อมูล K

$\overline{r_{cf}}$ คือค่าเฉลี่ยความสัมพันธ์ของตัวแปรกับคลาส $(f \in S)$

$\overline{r_{ff}}$ คือค่าเฉลี่ยความสัมพันธ์ระหว่างมิติข้อมูล

2.2.1.3 ทฤษฎีเกี่ยวกับเทคนิคการหาค่าเกนความรู้ (Information Gain) (วรการ ใจดี และ นรินทร์ จิวิตัน,2023) เป็นเทคนิคการเลือกคุณลักษณะเพื่อประเมินค่าในการแบ่งข้อมูลด้วยการคำนวณค่า Gain สำหรับแต่ละมิติข้อมูล ถ้ามิติข้อมูลใดมีค่า Gain สูงสุด จะถูกเลือกให้เป็นกลุ่มย่อยที่มีอำนาจในการจำแนก สมการของค่าเกน คือ

$$\text{Gain} = \text{Entropy}(p) - \sum_{i=1}^{k} \frac{n_i}{n} \text{Entropy}(i)$$

โดยที่ $\text{Entropy}(p)$ คือค่า Entropy ของตัว Root

$\sum_{i=1}^{k} \frac{n_i}{n} \text{Entropy}(i)$ คือค่า Entropy ในแต่ละโหนดย่อย

คำนวณหาค่า Entropy ได้จากสมการที่ 5

$$\text{Entropy}\ (p) = \sum_{i=0}^{c-1} p(j|t) \log_2 p(j|t)$$

โดยที่ $\sum_j$ คือผลรวมของความน่าจะเป็นของค่า j ที่เกิดในคลาส t

$p(j|t)$ คือค่าความถี่ที่มีความสัมพันธ์ของกลุ่ม j กับโหนด t

การคัดเลือกฟีเจอร์แบบ Gain Ratio Feature Selection (วรการ ใจดี,2562) เป็นวิธีที่ง่ายในการให้ค่าน้ำหนักของแต่ละฟีเจอร์ ซึ่งใช้หลักการเดียวกับการเลือกฟีเจอร์ของต้นไม้ตัดสินใจ (Decision Tree) ในเทคนิคนี้แต่ละฟีเจอร์จะได้รับค่าคะแนนที่เรียกว่า Entropy ซึ่งเป็นค่าที่วัดความแตกต่างหรือการกระจัดกระจายของข้อมูลกับคลาสทานาย (Class) ว่ามีความแตกต่างกันมากเพียงใดต่อคลาสคุณลักษณะหรือฟีเจอร์ (Attribute) ถ้ามีความแตกต่างกันมาก ค่า Entropy ของฟีเจอร์นั้นๆ จะมีค่าสูงแต่ถ้าข้อมูลมีความเหมือนกันมากค่า Entropy จะมีค่าต่ำ การคำนวณหาค่า Information Gain และ Entropy มีสมการดังต่อไปนี้

$$\text{Information Gain} = Entropy(parent) - \sum_{n=1}^{k} (Prob(Cn)x\ Entropy(Cn))$$

$$\text{Entropy} = -\sum_{n=1}^{k} (Prob(Cn)x\ log_2\ (Prob(Cn)))$$

ภาพที่ 2.9 แสดงตัวอย่างคุณลักษณะของ Entropy

(ที่มา: http://dataminingtrend.com/2014/data-mining-techniques/feature-selection-information-gain/)

จากภาพแต่ละจุดคือข้อมูลแต่ละตัว จะเห็นว่าถ้าข้อมูลมีคาตอบหรือคลาสเดียวกัน เช่น เป็นคลาสสีฟ้าหรือสีส้มทั้งหมดจะมีค่า Entropy ที่ต่ำที่สุด คือ Entropy เท่ากับ 0 แต่ถ้ามีความแตกต่างกันมาก เช่น เป็นคลาสสีฟ้าครึ่งหนึ่งและคลาสสีส้มอีกครึ่งหนึ่งจะมีค่าEntropy สูงสุด คือ Entropy เท่ากับ 1

2.2.2 ทฤษฎีเกี่ยวกับโมเดลวัดประสิทธิภาพ

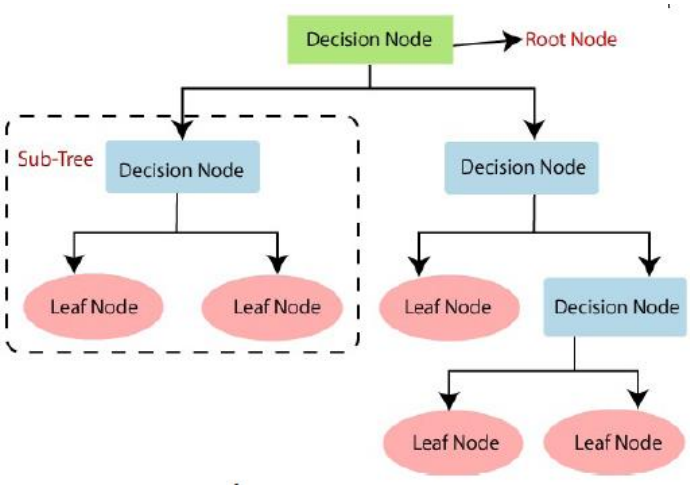
2.2.2.1 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ(Decision Tree) (มนิศรา เทพฝั้นและศิริขวัญ ขันนาแล,2562) ต้นไม้การตัดสินใจจะทำการจัดกลุ่ม (classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (node) ของต้นไม้การตัดสินใจคือตัวแปร (attribute) ต่าง ๆ ของชุด เช่น หากต้องการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณาคือ ข้อมูลทัศนียภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์จากต้นไม้คือการ ตัดสินใจว่าจะไปเล่นกีฬารึเปล่า ซึ่งแต่ละตัวแปรนั้นก็จะมีค่าของตัวเอง (value) เกิดเป็นชุดของ ตัวแปร-ค่าของตัวแปร (attribute-value pair) เช่น ทัศนียภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่านั้นก็อาจมีค่าได้เป็นใช่ กับ ไม่ใช่ เป็นต้น การทำนายประเภทด้วยต้นไม้ตัดสินใจจะเริ่มจากบัพราก โดยทดสอบค่าตัวแปรของบัพ แล้ว จึงตามกิ่งของต้นไม้ที่กำหนดค่าเพื่อไปยังบัพลูกถัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอ บัพใบซึ่งจะแสดงผลการทำนาย

ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย

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

2) กิ่ง คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด

3) ใบ คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ



ภาพที่ 2.10 เทคนิค Decision Tree

(ที่มา : https://onlinelibrary.wiley.com/doi/10.1155/2021/6053824)

ขั้นตอนในการสร้าง Decision Tree เพื่อใช้จำแนกข้อมูล มีดังนี้

1) เลือก Attribute ที่ทำหน้าที่เป็น Root Node

2) จาก Root Node สร้างเส้นเชื่อมโยงไปยังโหนดลูก จำนวนเส้นเชื่อมโยงจะเท่ากับจำนวนค่าที่เป็นไปได้ทั้งหมดของ Attribute ที่เป็น root node

3) ถ้าโหนดลูกเป็นกลุ่มของข้อมูลที่อยู่ในคลาสเดียวกันทั้งหมด ให้หยุดสร้างต้นไม้ แต่ถ้าโหนดลูกมีข้อมูลของหลายคลาสปะปนกันอยู่ ต้อง สร้าง subtree เพื่อจำแนก ข้อมูลต่อไป โดยเลือก subtree มาทำหน้าที่ เป็น root node ของ subtree มาทำซ้ำในขั้นตอนที่ 2.), 3.) ซึ่งการคำนวณมีดังนี้

เอนโทรปี (Entropy) การสร้างต้นไม้การตัดสินใจจากบนลงล่างด้วย การถามว่าลักษณะใด ควรจะเป็นรากของต้นไม้การตัดสินใจต้นนี้ และถามซ้ำ ๆ ไปเรื่อย ๆ เพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด (อังกฤษ: recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดนั้นดูจากค่าของลักษณะเรียกว่าเป็นความรู้ (Information Gain) ก่อนที่จะรู้จักเกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ E (S) ดังนี้

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

เมื่อแทนค่า S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

$p_S(j)$ คือ อัตราส่วนของกรณีใน S ที่ตัวแปรตามหรือผลลัพธ์มีค่า j

โดยสำหรับต้นไม้การตัดสินใจที่มีผลลัพธ์เป็นแค่เพียงค่าตรรกะ (Boolean) ใช่กับไม่ใช่เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปี คือ

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

เมื่อพิจารณาเอนโทรปีแล้วจะเห็นว่าเอนโทรปีจะมีค่าอยู่ระหว่าง 0 กับ 1 โดยจะมีค่าเป็นศูนย์เมื่อทุก ๆ กรณีมีผลลัพธ์เพียงแบบเดียว เช่น ใช่ทั้งหมด หรือไม่ใช่ ทั้งหมด และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น หรือจะพูดอีกนัยหนึ่งก็คือเอนโทร ปีจะมีค่ามากขึ้นหากข้อมูลไม่บริสุทธิ์ และจะตัดสินใจได้ว่าผลลัพธ์จะเป็นอะไรเมื่อเอนโทรปี เป็น 0 เท่านั้น

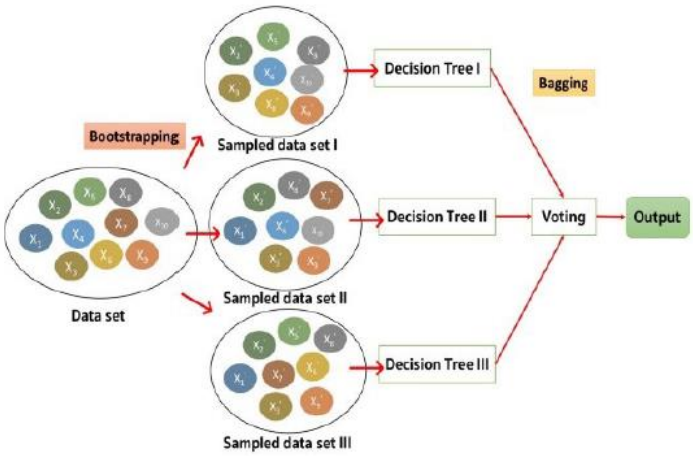
ตารางที่ 2.3 ตารางเปรียบเทียบข้อดีและข้อเสียของเทคนิค Decision Tree (นรินทร์ จิวิตัน,2024)

ข้อดี	**ข้อเสีย**
• เข้าใจง่าย สามารถมองเห็นกฎการตัดสินใจได้ชัดเจน • สามารถจัดการกับข้อมูลทั้งตัวเลขและ categorical ได้ ไม่จำเป็นต้องแปลงข้อมูล • ไม่จำเป็นต้องทำ normalization ข้อมูล • สามารถใช้เพื่อหาความสำคัญของแต่ละ attribute ได้	• อาจเกิด Overfitting ได้ง่าย ต้นไม้ที่ซับซ้อนเกินไปอาจจำข้อมูลฝึกสอนได้ดีเกินไป แต่ไม่สามารถทำนายข้อมูลใหม่ได้ดี • Sensitive ต่อการเปลี่ยนแปลงของข้อมูลเล็กน้อย การเปลี่ยนแปลงข้อมูลเพียงเล็กน้อยอาจทำให้โครงสร้างของต้นไม้เปลี่ยนแปลงไปมาก • อาจสร้างต้นไม้ที่ซับซ้อนเกินไป ยากต่อการตีความ

ความเหมาะสมการใช้ เทคนิค Decision Tree

- เหมาะสำหรับปัญหาการจำแนกประเภทที่มีทั้งข้อมูลตัวเลขและ categorical
- เหมาะสำหรับปัญหาที่ต้องการทราบเหตุผลว่าทำไมจึงจัดข้อมูลให้อยู่ใน class นั้นๆ
- เหมาะสำหรับข้อมูลที่มีปริมาณปานกลาง

2.2.2.2 ทฤษฎีเกี่ยวกับป่าสุ่ม (Random Forest) (นรินทร์ จิวิตัน, 2022) เป็นการสร้างแบบจำลองด้วยวิธีการต้นไม้ตัดสินใจขึ้นมาหลายๆแบบจำลองโดยวิธีการสุ่มตัวแปร แล้วนำผลที่ได้แต่ละแบบจำลองมารวมกันพร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย

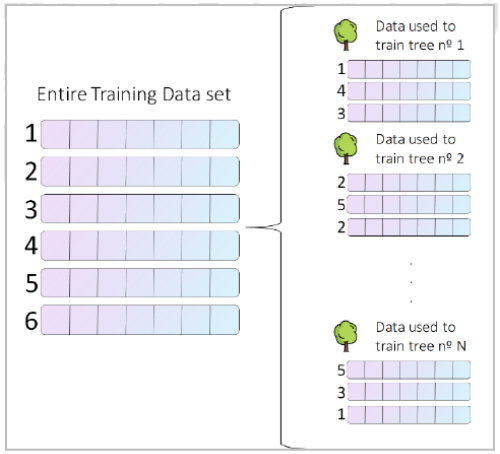


ภาพที่ 2.11 เทคนิค Random Forest

(ที่มา : https://medium.com/@witchapongdaroontham/เจาะลึก-random-forest-part-2-of-รู้จัก-tree-random-forest-และ-xgboost-79b9f41a1c1c)

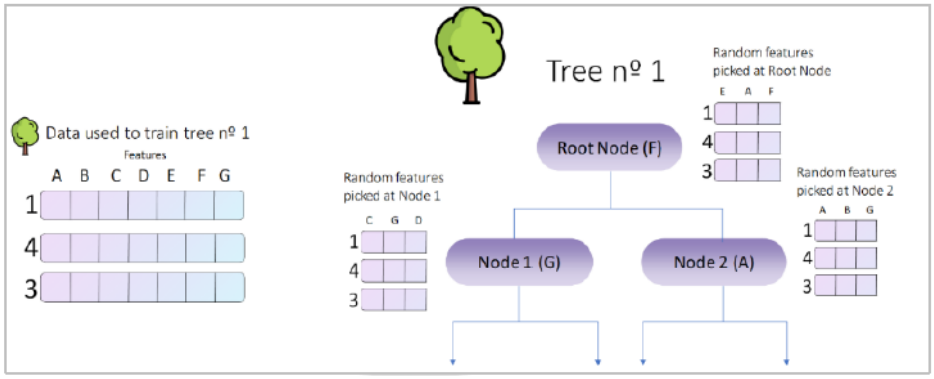
หลักการทำงานของ Random Forest (ปริยานุช ประเสริฐสิริกุล,2564) มี 2 วิธีดังนี้

1. Bootstrapping เป็นเทคนิคที่เพิ่มปริมาณข้อมูล และไม่ได้ทำให้ข้อมูลมีความหลากหลายเพิ่ม โดยจัดการข้อมูลตามแนวตั้ง จะได้ชุดข้อมูลที่ต่างกับชุดข้อมูลดั้งเดิมหลาย ๆ ชุดดังภาพที่ 2.12



ภาพที่ 2.12 การทำงานของ Random Forest แบบ Bootstrapping
(ที่มา : (Z_ai,2020b))

2. Random Feature Projection เป็นเทคนิคที่สุ่มเลือกคุณลักษณะแต่ละแบบจำลอง โดยจัดการข้อมูลตามแนวนอน ทำให้ได้ชุดข้อมูลที่ต่างกันหลายชุด ดังภาพที่ 2.13



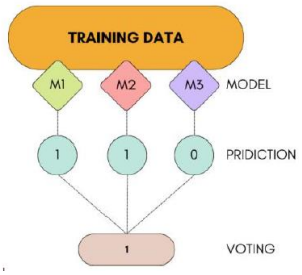
ภาพที่ 2.13 การทำงานของ Random Forest แบบ Random Feature Projection
(ที่มา : (Z_ai,2020b))

ตารางที่ 2.4 ตารางเปรียบเทียบข้อดีและข้อเสียของเทคนิค Random Forest (นรินทร์ จิวิตัน,2024)

ข้อดี	ข้อเสีย
• มีความแม่นยำสูง สามารถลดปัญหา Overfitting ได้ดี • สามารถจัดการกับข้อมูลที่มีมิติสูงได้ • สามารถจัดการกับข้อมูลที่ขาดหายไปได้ • สามารถบอกความสำคัญของแต่ละคุณสมบัติ (Attribute) ได้	• โมเดลมีความซับซ้อน อาจยากต่อการตีความผลลัพธ์ • ใช้ทรัพยากรในการคำนวณค่อนข้างสูง เนื่องจากต้องสร้างต้นไม้ตัดสินใจจำนวนมาก

ความเหมาะสมในการใช้ Random Forest เหมาะสำหรับข้อมูลที่มีความซับซ้อนสูงและต้องการการทำนายที่แม่นยำ เช่น การจัดหมวดหมู่ข้อความ

2.2.2.3 ทฤษฎีเกี่ยวกับการโหวต (Voting) (วรการ ใจดี และ นรินทร์ จิวิตัน ,2023) เป็นเทคนิคที่จัดอยู่ในการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning–based Model) เป็นการสร้างแบบจำลองการเรียนรู้โดยการใช้โมเดลหรือตัวจำแนก (Classifier) มากกว่าหนึ่งตัวในการเรียนรู้ เทคนิคการโหวตเป็นการนำข้อมูลการเรียนรู้ชุดเดียวกันให้โมเดลแต่ละโมเดลทำการเรียนรู้ เมื่อได้ผลการทำนายของแต่ละโมเดลเรียบร้อยแล้ว จะนำผลการทำนายมาใช้หลักการโหวตว่าคลาสคำตอบใดที่แต่ละโมเดลทำนายผลมากที่สุด ก็จะนำคลาสคำตอบนั้นเป็นผลการทำนายของคลาสที่กำลังพิจารณาอยู่ ดังภาพที่ 2.14



ภาพที่ 2.14 เทคนิคการโหวต

(ที่มา : ปัจจัยที่ส่งผลต่อการตัดสินใจเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและศิลปศาสตร์ โดยใช้เทคนิคการคัดเลือกคุณลักษณะและเทคนิคการทำเหมืองข้อมูล,วรการ ใจดี และนรินทร์ จิวิตัน,2023)

2.2.2.4 ทฤษฎีเกี่ยวกับนาอีฟเบย์ (Naïve Bayes) (วรการ ใจดี และ นรินทร์ จิวิตัน,2023) เป็นเทคนิคที่ใช้ความน่าจะเป็น (Probability) ในการจำแนกข้อมูลตามกฎของเบย์ (Bayes' Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่งๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่งๆ เพื่อหาสมมติฐานที่ดีที่สุด เป็นเทคนิคที่ได้รับความนิยมอย่างแพร่หลายในปัจจุบัน สมการของนาอีฟเบย์ คือ

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

โดยที่ $P(A|B)$ คือค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา (Conditional Probability)
$P(A \cap B)$ คือค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B จะเกิดร่วมกัน (Joint Probability)
$P(B)$ คือค่าความน่าจะเป็นที่เหตุการณ์ B จะเกิดขึ้น

วิธีการเรียนรู้เบย์อย่างง่าย (Naive Bayesian Learning) (วุฒิชัย กำจรกิตติคุณ,2561)เป็นวิธีจำแนกข้อมูลที่มีประสิทธิภาพวิธีหนึ่งโดยที่ใช้งานได้ดี เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน มีการนำการจำแนกประเภทของข้อมูลแบบเบย์อย่างง่ายไป ประยุกต์ใช้ในการแยกประเภทข้อความ (Text Classification) การวินิจฉัย (Diagnosis) และพบว่าใช้ งานได้ดีไม่ต่างจากการจำแนกประเภทด้วยวิธีการอื่นทำให้ผู้วิจัยเลือกวิธีการนี้มาใช้ในงานวิจัย เนื่องจากเป็นวิธีการจำแนกข้อมูลที่มีประสิทธิภาพและมีอัลกอริทึมในการทำงานที่ไม่ซับซ้อน กำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็นกลุ่ม V_j สำหรับข้อมูลที่มีคุณสมบัติ n ตัว

$$X = \{A_1, A_2, \dots, A_n\} \text{ หรือใช้สัญลักษณ์ว่า } P(A_1, A_2, \dots, A_n|V_j)$$
$$P(A_1, A_2, \dots, A_n|V_j) = \prod_{i=1}^{n} P(A_i|V_j)$$

โดยที่ $\prod$ หมายถึงผลคูณของค่า $P(A_i|V_j)$ ทั้งหมด i = 1,2,3,...,n และ j = 1,2,3,... การนำวิธีการเรียนรู้เบย์อย่างง่ายไปใช้ มีวิธีการดังต่อไปนี้

1. หาความน่าจะเป็นของสิ่งที่พบในในแต่ละกลุ่ม โดยนำค่า $P(A_1, A_2, \ldots, A_n|V_j) = \prod_{i=1}^{n} P(A_i|V_j)$ คูณกับค่าความน่าจะเป็นของกลุ่มคือ $P(V_j)$ จะได้เท่ากับ V_{NB}

2. นำค่าที่ได้มาเปรียบเทียบกัน กลุ่มที่มีความน่าจะเป็นสูงสุดคือคำตอบ ดังนั้นจะได้วิธีการจำแนกประเภทแบบนาอีฟเบย์ดังสมการ

$$V_{NB} = arg\,max P(V_j) \times \prod_{i=1}^{n} P(A_i|V_j)$$

นาอีฟเบย์จัดเป็นเทคนิคในการแก้ปัญหาแบบมีผู้สอน (Supervised Learning) ที่สามารถคาดการณ์ผลลัพธ์และสามารถอธิบายได้ด้วยการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับ แต่ละความสัมพันธ์ดังสมการ

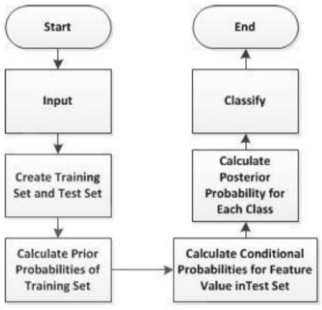
$$P(X_1, X_2, \cdots, X_n \mid C) = P(X_1 \mid C) \cdot P(X_2 \mid C) \cdots P(X_n \mid C)$$

และเมื่อผลคูณของความน่าจะเป็นจากเหตุการณ์ใดมีค่ามากกว่า จะถือว่าเหตุการณ์นั้นคือผลลัพธ์ของนาอีฟเบย์ดังอสมการ

$$[P(x_1 \mid c^*) \cdots P(x_n \mid c^*)]P(c^*) > [P(x_1 \mid c) \cdots P(x_n \mid c)]P(c)$$

เมื่อ $c \neq c^*, c = c_1, \cdots, c_L$

จะสรุปได้ว่าผลลัพธ์ของนาอีฟเบย์นั้นคือ c^* สามารถอธิบายหลักการทำงานอย่างง่ายได้ดังภาพที่ 2.15



ภาพที่ 2.15 Naïve Bayes Flow Chart

(ที่มา : การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก,

วุฒิชัย กำจรกิตติคุณ,2561)

เพื่อให้เข้าในการสร้างนาอีฟเบย์โมเดลมากยิ่งขึ้นนั้นจึงยกตัวอย่างของ Data set ของสภาพอากาศ ซึ่งประกอบไปด้วย outlook, temperature, humidity และ windy สำหรับการทำนายเงื่อนไขของการ เล่นเทนนิสดังดังตารางที่ 1 จากนั้นสร้างนาอีฟเบย์โมเดลจากข้อมูลที่ให้มาซึ่งจะได้ผลลัพธ์ดัง ตารางที่ 2

ตารางที่ 1 แสดงชุดข้อมูลของสภาพอากาศ

No	outlook	temperature	humidity	windy	play
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	mild	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

ตารางที่ 2 แสดงผลการสร้างนาอีฟเบย์โมเดล

outlook			temperature			humidity			windy			play	
	yes	no		yes	no		yes	no		yes	no	yes	no
Sunny	2	3	Hot	2	2	High	3	4	FALSE	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	TRUE	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/9	Hot	2/9	2/5	High	3/9	4/5	FALSE	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	TRUE	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

จากตารางแสดงให้เห็นความถี่ที่เกิดขึ้นของเหตุการณ์แต่ละส่วนที่ใช้ในการทำนายว่าจะเล่นเทนนิส หรือไม่ ยกตัวอย่างเช่นจาก Data set ทั้งหมดแสดง outlook=sunny เมื่อ play=yes ดังนี้

$$P(outlook = sunny|play = yes) = \frac{2}{9}$$

$$P(play = yes) = \frac{9}{14}$$

เมื่อได้นาอีฟเบย์โมเดลมาแล้ว จะสามารถทำนายเหตุการณ์ของการเล่น “play” ได้จากหลักฐาน ต่างๆที่ได้เตรียมไว้แล้ว จากตัวอย่างกำหนดให้เหตุการณ์ E มีสภาพอากาศดังนี้ outlook = sunny, temperature = cool, humidity = high และ windy = TRUE แล้วสามารถหาค่าความน่าจะเป็นได้ดังนี้

$$P[yes|E] = P[outlook = sunny|yes] \times P[temperature = cool|yes] \times P[humidity = high|yes] \times P[windy = TRUE|yes] \times \frac{P[yes]}{P[E]}$$

$$= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{P[E]}$$

ในที่นี้สามารถมองข้าม P[E] ได้เนื่องจากเราเพียงต้องการหาค่าความสัมพันธ์ (relatively) เพื่อเปรียบเทียบค่าระหว่าง 2 Class ซึ่งได้ผลดังนี้

Likelihood of two class

For “yes” = 2/9 * 3/9 * 3/9 * 3/9 * 9/14 = 0.0053

For "no" = 3/5 * 1/5 * 4/5 * 3/5 * 5/14 = 0.0206

Conversion into a probability by normalization

P(“yes”) = 0.0053/ (0.0053+0.0206) = 0.205

P(“no”) = 0.0206/ (0.0053+0.0206) = 0.795

สังเกตได้จากนาอีฟเบย์โมเดลข้างต้น จะเห็นได้ว่า $P(outlook = overcast|play = no) = 0/5$ ซึ่งจะทำให้เกิดปัญหาเมื่อหาค่า P(“no”) เนื่องจากผลลัพธ์ที่ได้จะเท่ากับ 0 สามารถแก้ปัญหาโดยใช้เทคนิคที่ เรียกว่า “smoothing technique” ซึ่งหนึ่งในวิธีการ smoothing technique นั้นคือ Laplace estimation

$$P(outlook = sunny|play = no) = \frac{3 + \mu p_1}{5 + \mu}$$

$$P(outlook = overcast|play = no) = \frac{0 + \mu p_2}{5 + \mu}$$

$$P(outlook = rainy|play = no) = \frac{2 + \mu p_3}{5 + \mu}$$

โดยที่ $p_1+p_2+p_3 = 1.0$ และตามที่ได้มีการกำหนดค่าให้ข้อมูลทั้งหมดมีการกระจายแบบเท่ากันจะได้ $p_1 = p_2 = p_3 = 1/3$

$$P(outlook = sunny|play = no) = \frac{3 + \mu/3}{5 + \mu} = \frac{3 + 3/3}{5 + 3} = \frac{4}{8}$$

$$P(outlook = overcast|play = no) = \frac{0 + \mu p_2}{5 + \mu} = \frac{0 + 3/3}{5 + 3} = \frac{1}{8}$$

$$P(outlook = rainy|play = no) = \frac{2 + \mu p_3}{5 + \mu} = \frac{2 + 3/3}{5 + 3} = \frac{3}{8}$$

นาอีฟเบย์จึงเป็นการจำแนกข้อมูลวิธีหนึ่งที่มีประสิทธิภาพและมีการทำงานที่ไม่ซับซ้อนเหมาะกับกรณีของเซ็ตตัวอย่างที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างที่ไม่ขึ้นต่อกัน ที่ผ่านมานั้นมีการนำนาอีฟเบย์ไปประยุกต์ใช้งานกันอย่างแพร่หลาย เช่น การกรองข้อความภาษาไทยและภาษาอังกฤษของบริการส่งข้อความสั้นบนเครือข่ายโทรศัพท์เคลื่อนที่ การจำแนกสายพันธุ์ยางพารา เป็นต้น

ตารางที่ 2.5 ตารางเปรียบเทียบข้อดีและข้อเสียของเทคนิค Naïve Bayes (นรินทร์ จิวิตัน ,2024)

ข้อดี	ข้อเสีย
• ง่ายต่อการใช้งาน มีสูตรการคำนวณที่ค่อนข้างตรงไปตรงมา • เหมาะสำหรับข้อมูลขนาดใหญ่ สามารถเรียนรู้ได้เร็ว	• สมมติฐานความเป็นอิสระ สมมติฐานที่ว่าคุณสมบัติ (Attribute) เป็นอิสระต่อกัน อาจไม่เป็นจริงในโลกแห่งความเป็นจริงเสมอไป

ตารางที่ 2.5 ตารางเปรียบเทียบข้อดีและข้อเสียของเทคนิค Naïve Bayes (นรินทร์ จิวิตัน ,2024) (ต่อ)

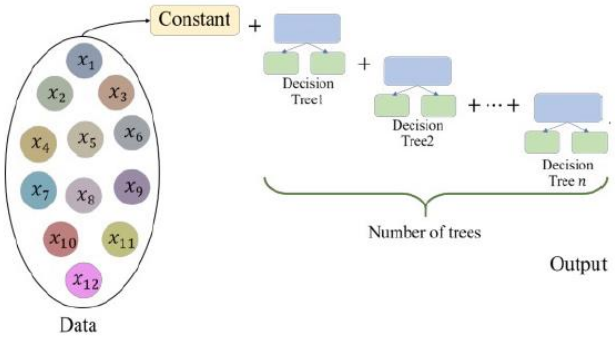
ข้อดี	ข้อเสีย
• ง่ายต่อการใช้งาน มีสูตรการคำนวณที่ค่อนข้างตรงไปตรงมา • เหมาะสำหรับข้อมูลขนาดใหญ่ สามารถเรียนรู้ได้เร็ว • ทำงานได้ดีกับข้อมูลที่มีมิติสูง เนื่องจากสมมติฐานที่ว่าคุณสมบัติ (Attribute) เป็นอิสระต่อกัน • เหมาะสำหรับปัญหา text classification เช่น การจัดประเภทอีเมลเป็นสแปมหรือไม่	• อาจให้ผลลัพธ์ที่ไม่ดีเมื่อข้อมูลมีการกระจายตัวไม่สม่ำเสมอ เช่น ข้อมูลที่มี class imbalance • สมมติฐานความเป็นอิสระ สมมติฐานที่ว่าคุณสมบัติ (Attribute) เป็นอิสระต่อกันอาจไม่เป็นจริงในโลกแห่งความเป็นจริงเสมอไป • อาจให้ผลลัพธ์ที่ไม่ดีเมื่อข้อมูลมีการกระจายตัวไม่สม่ำเสมอ เช่น ข้อมูลที่มี class imbalance • ไม่สามารถจับรูปแบบที่ซับซ้อนได้ดี เช่น ข้อมูลที่มีการโต้ตอบระหว่างคุณสมบัติ (Attribute)

ความเหมาะสมในการใช้เทคนิค Naïve Bayes

- เหมาะสำหรับปัญหาการจำแนกประเภทที่มีข้อมูลจำนวนมากและมีมิติสูง
- เหมาะสำหรับปัญหา text classification
- เหมาะสำหรับปัญหาที่ต้องการผลลัพธ์ที่รวดเร็ว

2.2.2.5 ทฤษฎีเกี่ยวกับเกรเดียนท์บูตทรีส์ (Gradient Boosting) (นรินทร์ จิวิตัน, 2022) เป็นเทคนิคการเรียนรู้ของเครื่อง ใช้แก้ปัญหาการถดถอยและปัญหาการจำแนกประเภท โดยการสุ่มสร้างต้นไม้การตัดสินใจหลายร้อยแบบจำลองและประเมินผลแต่ละแบบจำลองจนกว่าจะได้แบบจำลองที่เหมาะสมที่สุด ลดความคลาดเคลื่อนจากการเรียนรู้ก่อนหน้า สามารถจัดการข้อมูลสูญหาย(Missing Value) มีพารามิเตอร์ที่หลากหลาย

ให้ปรับเพื่อให้เหมาะสมกับแบบจำลอง เช่น จำนวนต้นไม้(Number Of Tree) จำนวนชั้นสูงสุดของต้นไม้(Maximal Depth) อัตราการเรียนรู้(Learning Rate) เป็นต้น



ภาพที่ 2.16 เทคนิค Gradient Boosting

(ที่มา : การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของข้อมูลการเสียชีวิต จากอุบัติเหตุบนโครงข่ายถนน โดยใช้เทคนิคการทำเหมืองข้อมูล, นรินทร์ จิวิตัน, 2022)

ขั้นตอนของอัลกอริธึม Gradient Boosting (อภิชา กวินพลอาสา และ ธฤตพน อู่สวัสดิ์,2565) สามารถเขียนได้ ดังนี้

1.เริ่มต้นด้วย $\hat{F}_0(x_{t-h}) = 0$ และตั้งค่า learning rate $\varepsilon > 0$ ซึ่งค่านี้สามารถควบคุมอัตราการบูตติ่งให้พอดีหรือมากเกินไปได้

2. สำหรับ 1= 1,...,L

2.1 คำนวณค่าความผิดพลาดเทียม (pseudo–residuals) เช่น negative gradient ของฟังก์ชั่นสูญเสีย (Loss Function) ถูกมินิไมเซอร์ (minimized) ที่ความพอดีปัจจุบันในแต่ละจุดของข้อมูลที่ใช้ฝึก สามารถเขียนเป็นสมการ ดังนี้

$$R_t = -\frac{\partial L(Y_t, F(X_{t-h}))}{\partial F(X_{t-h})} \Big| F(X_{t-h}) = \hat{F}_{l-1}(X_{t-h})$$

2.2 ประมาณค่า negative gradient โดยใช้ความลึกของต้นไม้ตัดสินใจ (d) และแก้ไขปัญหาตามวิธีออฟติไมเซชั่น (Optimization) สามาถเขียนสมการได้ดังนี้

$$minimize_{\gamma} = \sum_{t=1}^{T} \left(R_t - g(X_{t-h};\, \gamma)\right)^2$$

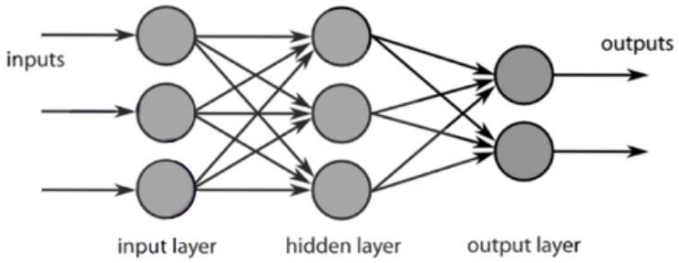
2.3 ปรับปรุง $\hat{F}_l(X_{t-h})$ ให้เป็นไปตามสมการนี้

$$\hat{F}_l(X_{t-h}) = \hat{F}_{l-1}(X_{t-h}) + \hat{g}_l(X_{t-h})$$

เมื่อ $\hat{g}_l(X_{t-h}) = \epsilon g(x_{t-h}; \hat{\eta})$

2.4 ส่งกลับค่าในลำดับถัดไปที่ $\hat{F}_l(X_{t-h}), l = 1, \ldots, L$

2.2.2.6 ทฤษฎีเกี่ยวกับการเรียนรู้เชิงลึก (Deep learning) (นรินทร์ จิวิตัน, 2022) เป็นวิธีการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์(Neurons) โดยนำระบบโครงข่ายประสาท(Neural Network) มาซ้อนกันหลายชั้น (Layer) และทำการเรียนรู้ข้อมูลตัวอย่าง ข้อมูลดังกล่าวจะถูกนำไปใช้ในการจัดหมวดหมู่ข้อมูล (Classify The Data)



ภาพที่ 2.17 เทคนิค Deep learning

(ที่มา : โครงงานการพยากรณ์ความต้องการสินค้าด้วยเทคนิคการเรียนรู้ของเครื่องในธุรกิจค้าปลีก,ชุติมณฑน์ ภักดีสิโรตม์,2021)

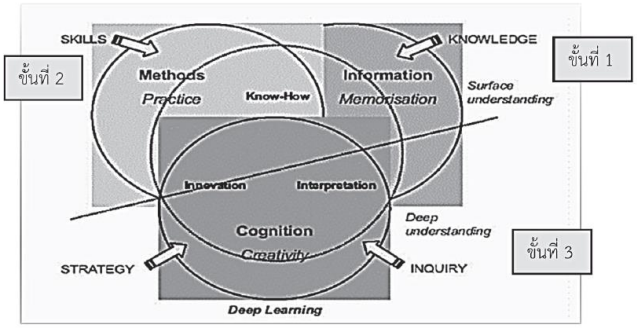
กิจกรรมการเรียนรู้เชิงลึก (สุมนา โสตถิผลอนันต์,2564)

การจัดกิจกรรมการเรียนรู้เชิงลึก ประกอบด้วย กิจกรรม 3 ขั้น ได้แก่ (Havard, et al., 2005: 126)

ขั้นที่ 1 การบูรณาการความรู้ในระดับความเข้าใจ หรือความรู้ผิวเผิน (Surface understand) เพื่อใช้เป็นความรู้พื้นฐานสำหรับพัฒนาความคิด

ขั้นที่ 2 การนำความรู้เฉพาะส่วนไปใช้ฝึกปฏิบัติ (Practice) โดยมาจากความรู้เดิมในขั้นที่ 1และขั้นที่ 2 เป็นการเรียนรู้ว่าจะทำอย่างไรจากสิ่งที่มีอยู่หรือความรู้เดิม (Know–how of the material)

ขั้นที่ 3 การใช้ความคิดรวบยอดที่เป็นความรู้เชิงลึก (Deep understanding) และนำไปประยุกต์ใช้แก้ปัญหา ซึ่งเป็นการผลิตนวัตกรรมที่มาจากการใช้ความคิดในขั้นสูง (Higher–order thinking) หรือความคิดสร้างสรรค์ โดยอาศัยการมอบหมายงานเพื่อนำไปสู่กระบวนการคิด ดังภาพที่ 2.18



ภาพที่ 2.18 การพัฒนากระบวนการคิดเชิงสร้างสรรค์โดยผ่านการเรียนรู้เชิงลึก (ที่มา: Havard et al., 2005: 126)

การออกแบบกิจกรรมการเรียนรู้เชิงลึกเพื่อพัฒนาความคิดสร้างสรรค์

การใช้การเรียนรู้เชิงลึก เป็นการใช้ความรู้พื้นฐานที่รับรู้มา ซึ่งอาจเป็นความรู้ที่ผิวเผิน แล้วมาทำความเข้าใจให้ลึกซึ้ง หรือขยายความรู้ให้ชัดเจนขึ้นด้วยการสืบค้น (Inquiry) ร่วมกับการใช้ทักษะต่าง ๆ(Skills) ที่เกี่ยวข้องกับการปฏิบัติที่จะต้องนำไปใช้จากการมอบหมายงาน ผู้เรียนจะเรียนรู้ว่า จะต้องทำอย่างไร (Know–how) เพื่อให้เกิดผลสัมฤทธิ์ของงานที่ได้รับมอบหมาย โดยอาศัยกลยุทธ์(Strategy) ที่เป็นการใช้ความสามารถจากการเรียนรู้วิธีการที่ใช้ร่วมกับการคิดสร้างสรรค์ ซึ่งในทางปฏิบัติ การจัดกิจกรรมการเรียนรู้สามารถใช้วิธีการต่าง ๆ เช่น การใช้สื่อมัลติมีเดีย การใช้สื่อวิดีโอเพื่อให้เกิดความรู้พื้นฐานที่เข้าใจง่าย และฝึกการใช้กระบวนการคิดจากงานที่ได้รับมอบหมายผู้เรียนสามารถใช้วิธีการเรียนรู้เชิงลึกเพื่อพัฒนาความคิดสร้างสรรค์ได้การออกแบบกิจกรรมการเรียนรู้เชิงลึกจึงประกอบด้วย 2 ส่วน คือ

1. กิจกรรมที่ทำให้เกิดความรู้เบื้องต้นเพื่อให้เข้าใจในเนื้อหาความรู้

2. กิจกรรมที่เป็นการขยายความรู้ เพื่อให้ผู้เรียนสามารถใช้ทักษะต่าง ๆ และกระบวนการคิดในขั้นสูง

แนวทางพัฒนาความคิดสร้างสรรค์จากการใช้วิธีเรียนรู้เชิงลึก

เมื่อต้องการให้ผู้เรียนได้พัฒนาความคิดสร้างสรรค์ในรายวิชาใดก็ตาม สิ่งที่จะใช้ในการพัฒนาผู้เรียนประกอบด้วย

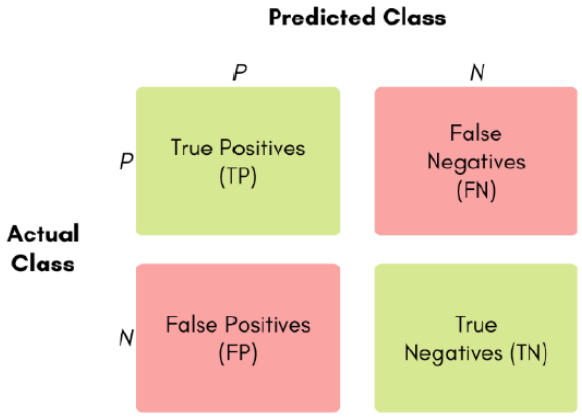
1. การกำหนดเป้าหมายการเรียนรู้ หรือจุดประสงค์การเรียนรู้โดยใช้มิติความรู้ที่สอดคล้องกับงานที่ต้องการให้ทำกับมิติกระบวนการคิดในขั้นสูง

2. การออกแบบกิจกรรมการเรียนรู้ แบ่งออกเป็น 2 ส่วน คือ 1) ส่วนที่เป็นการให้ความรู้เบื้องต้น ซึ่งเกี่ยวข้องกับงานที่จะมอบหมายให้ผู้เรียนทำ จำเป็นต้องให้ผู้เรียนได้เรียนรู้อย่างรวดเร็ว และ 2) ส่วนที่เป็นการมอบหมายงานให้ผู้เรียนรับผิดชอบ ซึ่งเป็นการให้ผู้เรียนใช้วิธีการเรียนรู้เชิงลึก ผู้เรียนจะเป็นผู้ค้นคว้าหาความรู้เฉพาะเจาะจงเพื่อให้เกิดความเข้าใจเชิงลึกสำหรับนำมาใช้ในการแก้โจทย์ปัญหาหรือผลิตผลงาน/สร้างนวัตกรรมตามลักษณะงานที่ได้รับมอบหมาย

3. การนำแผนการจัดการเรียนรู้ไปใช้บทบาทผู้สอนจะเป็นเพียงผู้อำนวยความสะดวกเป็นที่ปรึกษา และช่วยเหลือบางอย่างที่ผู้เรียนต้องการ ส่วนบทบาทผู้เรียนจะเป็นผู้ดำเนินการมีการวางแผนการเรียนรู้ด้วยตนเอง มีการระดมสมอง มีการใช้กลยุทธ์ในการปฏิบัติงานเพื่อให้งานสัมฤทธิ์ผลตามเป้าหมายการเรียนรู้

2.2.3 ทฤษฎีการประเมินผลโมเดล

คอมฟิวชันเมทริกซ์ (Confusion Matrix) (วรการ ใจดี และ นรินทร์ จิวิตัน, 2023) เป็นตารางประเมินผลลัพธ์การทำนายของแบบจำลอง สามารถวัดจากผลลัพธ์การจำแนกข้อมูล (classification) รูปแบบของคอมฟิวชันเมทริกซ์เป็น ดังภาพที่ 2.19



ภาพที่ 2.19 Confusion Matrix

(ที่มา : ปัจจัยที่ส่งผลต่อการตัดสินใจเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและศิลปศาสตร์ โดยใช้เทคนิคการคัดเลือกคุณลักษณะและเทคนิคการทำเหมืองข้อมูล ,วรการ ใจดี และ นรินทร์ จิวิตัน, 2023)

โดยที่ TP (True Positive) คือ ผลการทำนายตรงกับสิ่งที่เกิดขึ้นจริง ในกรณีที่ทำนายว่าจริงและสิ่งนั้นเกิดขึ้นจริง

FP (False Positive) คือ ผลการทำนายตรงกับสิ่งที่เกิดขึ้น ในกรณีที่ทำนายว่าจริงและสิ่งที่เกิดขึ้นไม่จริง

TN (True Negative) คือ ผลการทำนายตรงกับสิ่งที่เกิดขึ้น ในกรณีที่ทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นไม่จริง

FN (False Negative) คือ ผลการทำนายไม่ตรงกับสิ่งที่เกิดขึ้นจริง ในกรณีที่ทำนายว่าไม่จริง แต่สิ่งนั้นเกิดขึ้นจริง

โดยบริบทของงานวิจัยนี้

TP คือ กรณีที่ทำนายว่านักศึกษามารายงานตัว และนักศึกษามารายงานตัวจริง

FP คือ กรณีที่ทำนายว่านักศึกษามารายงานตัว แต่นักศึกษาไม่มารายงานตัว

TN คือ กรณีที่ทำนายว่านักศึกษาไม่มารายงานตัว และนักศึกษาไม่มารายงานตัวจริง

FN คือ กรณีที่ทำนายว่านักศึกษาไม่มารายงานตัว แต่นักศึกษามารายงานตัว

จากตาราง Confusion Matrix สามารถนำมาคำนวณเพื่อหาค่าการวัดประสิทธิภาพของโมเดลดังนี้

1.ค่าความถูกต้อง (Accuracy) วิธีวัดความถูกต้องจากจำนวนคำตอบที่ถูกต้องเทียบกับจำนวนคำตอบทั้งหมดที่นำไปให้โมเดลทำการตอบ โดยใช้สูตรคำนวณ Accuracy = (TP + TN) / (TP + FP + TN + FN)

2.ค่าความถูกต้อง (Recall) วิธีวัดค่าสัดส่วนของข้อมูลที่ตรงตามความต้องการที่ถูกค้นคืนกับข้อมูลที่ต้องการทั้งหมด โดยใช้สูตรคำนวณ Recall = TP / (TP + FN)

3.ค่าความแม่นยำ (Precision) วิธีวัดค่าจำนวนที่ทายถูก จากข้อมูลที่ทำนายว่าเป็นคลาสที่พิจารณาอยู่ โดยใช้สูตรคำนวณ Precision = TP / (TP + FP)

4.ค่าความถ่วงดุล (F–Measure) วิธีการหาค่าน้ำหนักหรือค่าเฉลี่ยของ ค่าความแม่นยำ (Precision) และ ค่าเรียกกลับ (Recall) หรือเป็นการหาประสิทธิภาพโดยรวมของคลาสคำตอบ โดยใช้สูตรคำนวณ F–Measure = 2 x (Precision x recall) / (Precision + Recall)

2.2.4 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.4.1. ข้อมูลขนาดใหญ่

ข้อมูลขนาดใหญ่ หมายถึง ข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยากหรือเป็นไปไม่ได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมากเพื่อทำการวิเคราะห์มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วง ต้นปีค.ศ. 2000 เมื่อดั๊ก ลานีย์ นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันในขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วย3Vs ดังนี้

1) Volume (ปริมาณ) : องค์กรต่าง ๆ รวบรวมข้อมูลจากหลากหลายแหล่งซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IOT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดียและอื่นๆ ในอดีต การจัดเก็บข้อมูลถือเป็นปัญหาใหญ่ – แต่เมื่อค่าใช้จ่ายในการจัดเก็บบนแพลตฟอร์มต่าง ๆ เช่น พื้นที่จัดเก็บข้อมูลส่วนกลาง (Data Lake) และ Hadoop ลดลง ภาระนี้จึงบรรเทาลง

2) Velocity (ความเร็ว) : ด้วยการเติบโตของ Internet of Things ข้อมูลจะถูกส่งไปยังธุรกิจต่าง ๆ ด้วยความเร็วที่ไม่เคยมีมาก่อนและต้องได้รับการจัดการในเวลาที่

เหมาะสม แท็ก RFID, เซ็นเซอร์ และสมาร์ทมิเตอร์ช่วยผลักดันความต้องการในการจัดการกับกระแสข้อมูลเหล่านี้ในแบบเรียลไทม์

3) Variety (ความหลากหลาย) : ข้อมูลมีในทุกรูปแบบ นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ อีเมล วิดีโอ เสียง ข้อมูล หุ้น และธุรกรรมทางการเงิน

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

1) ลดต้นทุน : เทคโนโลยีข้อมูลขนาดใหญ่เช่น Hadoop และการวิเคราะห์บนคลาวด์นำมาซึ่งความได้เปรียบด้านต้นทุนอย่างมีนัยสำคัญเมื่อพูดถึงการจัดเก็บข้อมูลจำนวนมากรวมทั้งสามารถระบุวิธีการทำธุรกิจที่มีประสิทธิภาพมากขึ้น

2) เร็วกว่าและตัดสินใจดีกว่า : ด้วยความเร็วของ Hadoop และการวิเคราะห์ในหน่วยความจำรวมกับความสามารถในการวิเคราะห์แหล่งข้อมูลใหม่ๆ ของธุรกิจจะสามารถสร้างข้อมูลได้ทันทีและสามารถดำเนินการต่อได้ทันทีจากการวิเคราะห์นั้น ๆ

3) ผลิตภัณฑ์และบริการใหม่ : ด้วยความสามารถในการวัดความต้องการและความพึงพอใจของลูกค้าผ่านการวิเคราะห์นำมาซึ่งสิ่งที่ลูกค้าต้องการ ซึ่งยกตัวอย่างโดยดาเวนพอร์ท สามารถชี้ให้เห็นว่าด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ของบริษัทต่าง ๆ จะสามารถสร้างผลิตภัณฑ์ใหม่ ๆ เพื่อตอบสนองความต้องการของลูกค้าได้เป็นอย่างดี

2.2.4.2 การจัดการข้อมูลขนาดใหญ่

1) การรวบรวมข้อมูล Big Data เป็นการรวบรวมข้อมูลของจากหลากหลายทั้งที่มาและการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบ ดั้งเดิม ETL (extract, transform, and load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิค วิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาด เทราไบต์ และอาจจะเป็นระดับเพธาไบต์เลยก็มี ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล

จัดรูปแบบ ให้เหมาะสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้น ๆ

2) การจัดการข้อมูล ข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ on premises หรือแบบ cloud ขึ้นกับความต้องการหรือความสะดวกในการใช้ ซึ่งเราสามารถใช้และประเมินผลได้เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูล หรือข้อมูลบางอัน ต้องการความยืดหยุ่นสูงและไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยม กันเป็นอย่างมาก

3) การวิเคราะห์ การลงทุนสร้างข้อมูลขนาดใหญ่ หรือ Big data จะมีประโยชน์หรือคุ้มค่าก็ต่อเมื่อคุณใช้และวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่างและชัดเจนในชุด ข้อมูลที่คุณมีอยู่ การสำรวจข้อมูลยังทำให้เราค้นพบสิ่งใหม่ แชร์สิ่งที่ค้นพบใหม่ๆต่อคนอื่นสร้างรูปแบบจำลองข้อมูล ด้วยการเรียนรู้ของเครื่องจักรและปัญญาประดิษฐ์ AI และนำข้อมูลเหล่านั้นไปใช้งาน

2.2.4.3 เทคโนโลยีการประมวลผลข้อมูล

การประมวลผลข้อมูลที่เป็น Big Data จะมีทั้งการวิเคราะห์ข้อมูลที่เป็นbusiness inteligence (BI) เพื่อที่จะดึงข้อมูลมานำเสนอ หรือการทำ Predictive Analytis โดยใช้หลักการของ Data Science ความยากของการประมวลผลคือต้องการความเร็วในการประมวลผลข้อมูลที่นอกจากมีขนาดใหญ่แล้วบางครั้งยังเป็นข้อมูลที่ไม่มีโครงสร้าง ดังนั้นต้อง จึงมีการนำเทคโนโลยีหรือภาษาต่าง ๆ มาเพื่อให้สามารถประมวลผลข้อมูลได้ ซึ่งในบางครั้งหน่วยงานอาจต้องพิจารณาต้องเลือกใช้

2.2.5 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

2.2.5.1 การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการ จัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การ

จัดเก็บใน รูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถ ค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

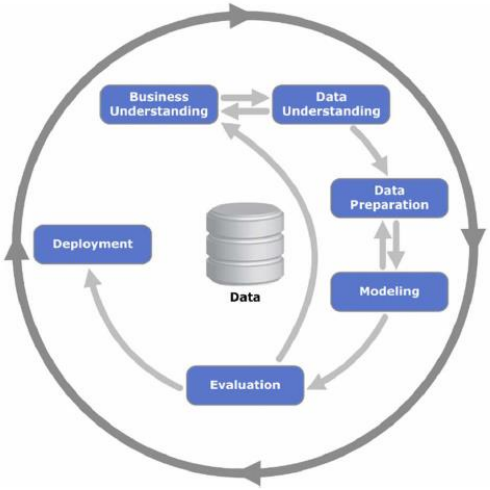
1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง

2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย

3) การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล

4) การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

2.2.5.2 ขั้นตอนการทำเหมืองข้อมูล ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้



ภาพที่ 2.20 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM

(ที่มา : http://dataminingtrend.com/2014/data-mining-techniques/crisp-dm-example/)

1) Business Understanding เน้นไปที่การทำความเข้าใจในงาน ระบุโอกาสและหาปัญหาที่จะเกิดขึ้น กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์ ซึ่งต้องสามารถระบุผลลัพธ์ที่มีได้

2) Data Understanding ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์

3) Data Preparation ทำการแปลงข้อมูล (Raw Data) ให้กลายเป็นข้อมูลที่สามารถนำมาช่วยในการวิเคราะห์ต่อไปได้ ขั้นตอนนี้จะใช้เวลามากที่สุดในทุกขั้นตอน เพราะ คุณภาพของงานที่ได้จะดีเพียงใดขึ้นอยู่กับคุณภาพข้อมูลที่จัดเตรียมในขั้นนี้ การเตรียมข้อมูลประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4) Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากขั้นตอนที่ 3 พร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้คำตอบที่ดีที่สุด บางครั้งอาจมีการย้อนกลับไปปรับการเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

5) Evaluation การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับวัตถุประสงค์หรือเป้าหมายที่ได้ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด หากไม่ได้ผลลัพธ์ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า

6) Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์จากทั้งหมด มาใช้ประโยชน์ตามวัตถุประสงค์ และทำการประเมินผลลัพธ์ที่ได้ว่ามีประสิทธิภาพตรงตาม วัตถุประสงค์ที่ตั้งไว้หรือไม่

2.2.6 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.6.1 หลักในการออกแบบเว็บไซต์

หน้าเว็บเป็นสิ่งแรกที่ผู้ใช้จะได้เห็นขณะที่เปิดเข้าสู่เว็บไซต์ และยังเป็นสิ่งแรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วย หน้าเว็บจึงเป็นสิ่งสำคัญมากเพราะเป็นสื่อกลางให้ผู้ชมสามารถใช้ประโยชน์จากข้อมูลของระบบงานของเว็บไซต์นั้นได้ โดยปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้น ระบบเนวิเกชั่น และองค์ประกอบอื่น ๆ ที่ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการใช้งานหลักสำคัญในการออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อความหมายเกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้ การออกแบบเว็บไซต์ ต้องคำนึงถึง

1) **ความเรียบง่าย** ได้แก่ มีรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานได้สะดวก ไม่มีกราฟิกหรือตัวอักษรที่เคลื่อนไหวอยู่ตลอดเวลา ชนิดและสีของตัวอักษรไม่มากจนเกินไป ทำให้วุ่นวาย

2) **ความสม่ำเสมอ** ได้แก่ ใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ เช่นรูปแบบของหน้า สไตล์ของกราฟิก ระบบเนวิเกชันและโทนสี ควรมีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) **ความเป็นเอกลักษณ์** การออกแบบเว็บไซต์ควรคำนึงถึงลักษณะขององค์กร เพราะรูปแบบของเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กรนั้น ๆ เช่น ถ้าเป็นเว็บไซต์ของทาง ราชการ จะต้องดูน่าเชื่อถือไม่เหมือนสวนสนุก ฯลฯ

4) **เนื้อหาที่มีประโยชน์** เนื้อหาเป็นสิ่งที่สำคัญที่สุดในเว็บไซต์ ดังนั้นควรจัดเตรียมเนื้อหาและข้อมูลที่ผู้ใช้ต้องการให้ถูกต้อง และสมบูรณ์ มีการปรับปรุงและเพิ่มเติมให้ทันเหตุการณ์อยู่เสมอ เนื้อหาไม่ควรซ้ำกับเว็บไซต์อื่น จึงจะดึงดูดความสนใจ

5) **ระบบเนวิเกชันที่ใช้งานง่าย** ต้องออกแบบให้ผู้ใช้เข้าใจง่ายและใช้งานสะดวก ใช้กราฟิกที่สื่อความหมายร่วมกับคำอธิบายที่ชัดเจน มีรูปแบบและลำดับขอรายการที่สม่ำเสมอ เช่น วางไว้ ตำแหน่งเดียวกันของทุกหน้า

6) **ลักษณะที่น่าสนใจ** หน้าตาของเว็บไซต์จะต้องมีความสัมพันธ์กับคุณภาพขององค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่จะต้องสมบูรณ์ การใช้สี การใช้ตัวอักษรที่อ่านง่าย สบายตา การใช้โทนสีที่เข้ากันลักษณะหน้าตาที่น่าสนใจนั้นขึ้นอยู่กับความชอบของแต่ละบุคคล

7) **การใช้งานอย่างไม่จำกัด** ผู้ใช้ส่วนใหญ่สามารถเข้าถึงได้มากที่สุดเลือกใช้บราวเซอร์ชนิดใดก็ได้ในการเข้าถึงเนื้อหาสามารถแสดงผลได้ทุกระบบปฏิบัติการและความละเอียดหน้าจอต่าง ๆ กันอย่างไม่มีปัญหาเป็นลักษณะสำคัญสำหรับผู้ใช้ที่มีจำนวนมาก

8) **คุณภาพในการออกแบบ** การออกแบบและเรียบเรียงเนื้อหาอย่างรอบคอบสร้างความรู้สึกว่าเว็บไซต์มีคุณภาพ ถูกต้อง และเชื่อถือได้

9) **ลิงค์ต่าง ๆ** จะต้องเชื่อมโยงไปหน้าที่มีอยู่จริงและถูกต้อง ระบบการทำงานต่าง ๆ ในเว็บไซต์จะต้องมีความแน่นอนและทำหน้าที่ได้อย่างถูก

2.2.6.2 องค์ประกอบของการออกแบบเว็บไซต์

1) โครงสร้างที่ชัดเจน ผู้ออกแบบเว็บไซต์ควรจัดโครงสร้างหรือจัดระเบียบของข้อมูลที่ชัดเจน แยกย่อยเนื้อหาออกเป็นส่วนต่าง ๆ ที่สัมพันธ์กันและให้อยู่ในมาตรฐานเดียวกัน จะช่วยให้น่าใช้งานและง่าย ต่อการอ่านเนื้อหาของผู้ใช้เสร็จสิ้น

2) การใช้งานที่ง่าย ลักษณะของเว็บที่มีการใช้งานง่ายจะช่วยให้ผู้ใช้รู้สึกสบายใจต่อการอ่านและสามารถทำความเข้าใจกับเนื้อหาได้อย่างเต็มที่ โดยไม่ต้องมาเสียเวลาอยู่กับการทำความเข้าใจ การใช้งานที่สับสนด้วยเหตุนี้ผู้ออกแบบบจึงควรกำหนดปุ่มการใช้งานที่ชัดเจน เหมาะสม โดยเฉพาะปุ่มควบคุมเส้นทางการเข้าสู่เนื้อหา (Navigation) ไม่ว่าจะเป็นเดินหน้า ถอยหลัง หากเป็นเว็บไซ์ที่มีเว็บเพจจำนวนมาก ควรจะจัดทำแผนผังของเว็บไซต์ (Site Map) ที่ช่วยให้ผู้ใช้ทราบว่า ตอนนี้อยู่ ณ จุดใด หรือเครื่องมือสืบค้น (Search Engine) ที่ช่วยในการค้นหาหน้าที่ที่ต้องการ

3) การเชื่อมโยงที่ดี ลักษณะไฮเปอร์เท็กซ์ที่ใช้ในการเชื่อมโยง ควรอยู่ในรูปแบบที่เป็นมาตรฐาน ทั่วไปและต้องระวังเรื่องของตำแหน่งในการเชื่อมโยง การที่จำนวนการเชื่อมโยงมากและกระจัดกระจายอยู่ทั่วไปในหน้าอาจก่อให้เกิดความสับสน นอกจากนี้คำที่ใช้สำหรับการเชื่อมโยงจะต้องเข้าใจง่ายมีความชัดเจนและไม่สั้นจนเกินไป นอกจากนี้ในแต่ละเว็บเพจที่สร้างขึ้นมาควรมี จุดเชื่อมโยงกลับมายังหน้าแรกของเว็บไซต์ที่กำลังใช้งานอยู่ด้วย ทั้งนี้เผื่อว่าผู้ใช้เกิดหลงทาง และไม่ทราบว่าจะทำอย่างต่อไปดีจะได้มีหนทางกลับมาสู่จุดเริ่มต้นใหม่ระวังอย่าให้มีหน้าที่ไม่มีการเชื่อมโยง (orphan Page) เพราะจะทำให้ผู้ใช้ไม่รู้จะทำอย่างไรต่อไป

4) ความเหมาะสมในหน้าจอ เนื้อหาที่นำเสนอในแต่ละหน้าจอควรสั้นกระชับและทันสมัย หลีกเลี่ยงการใช้หน้าจอที่มีลักษณะการเลื่อนขึ้นลง (Scrolling) แต่ถ้าจำเป็นต้องมี ควรจะให้ข้อมูลที่มี ความสำคัญอยู่บริเวณด้านบนสุดของหน้าจอ หลีกเลี่ยงการใช้กราฟิกด้านบนของหน้าจอ เพราะถึงแม้จะดูสวยงาม แต่จะทำให้ผู้ใช้เสียเวลาในการได้รับ ข้อมูลที่ต้องการ แต่หากต้องมีการใช้ภาพประกอบก็ควรใช้เฉพาะที่มีความสัมพันธ์กับเนื้อหาเท่านั้น นอกจากนี้การใช้รูปภาพเพื่อเป็นพื้นหลัง (Background) ไม่ควรเน้นสีสันที่ฉูดฉาดมาก นัก เพราะอาจจะไปลดความเด่นชัดของเนื้อหาลง ควรใช้ภาพที่มีสีอ่อน ๆ ไม่สว่างจนเกินไปรวมไปถึงการใช้เทคนิคต่าง ๆ เช่น ภาพเคลื่อนไหว หรือตัวอักษรวิ่ง (Marquees) ซึ่งอาจจะเกิดการรบกวนการอ่านได้ ควรใช้เฉพาะที่จำเป็นจริง ๆ เท่านั้น

ตัวอักษรที่นำมาแสดงบนจอภาพควรเลือกขนาดที่อ่านง่าย ไม่มีสีสันและลวดลายมากเกินไป

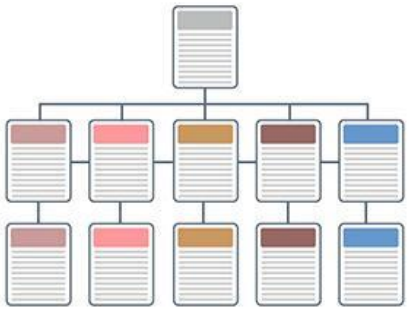
5) ความรวดเร็ว ความรวดเร็วเป็นสิ่งสำคัญประการหนึ่งที่ส่งผลต่อการเรียนรู้ ผู้ใช้จะเกิดอาการเบื่อหน่ายและหมดความสนใจกับเว็บที่ใช้เวลาในการแสดงผลนานสาเหตุสำคัญที่จะทำให้การแสดงผลนานคือการใช้ภาพกราฟิกหรือภาพเคลื่อนไหว ซึ่งแม้ว่าจะช่วยดึงดูดความสนใจได้ดี ฉะนั้นในการออกแบบจึงควรหลีกเลี่ยงการใช้ภาพขนาดใหญ่หรือภาพเคลื่อนไหวที่ไม่จำเป็น และพยายามใช้กราฟิกแทนตัวอักษรธรรมดาให้น้อยที่สุด โดยไม่ควรใช้มากเกินกว่า 2 – 3 บรรทัดในแต่ละหน้าจอ

2.2.6.3 โครงสร้างเว็บไซต์

1) เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure) เป็นโครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยม จัด ด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การเรียงลำดับตามตัวอักษร ดรรชนี สารานุกรม หรืออภิธานศัพท์ โครงสร้างแบบนี้ เหมาะกับเว็บไซต์ที่มีขนาดเล็ก เนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปทีละหน้า ทิศทางของการเข้าสู่เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรง โดยมี ปุ่มเดินหน้า–ถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาของตนเองได้ ทำให้เสียเวลาเข้าสู่เนื้อ

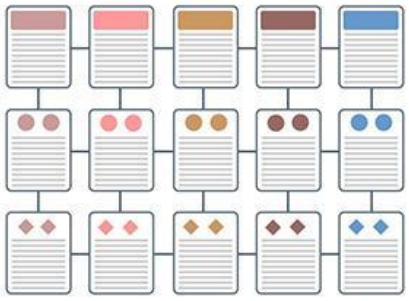
ภาพที่ 2.21 เว็บที่มีโครงสร้างแบบเรียงลำดับ

2) เว็บที่มีโครงสร้างแบบลำดับขั้น (Hierarchical Structure) เป็นวิธีที่ดีที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วนต่างๆ และมีรายละเอียดย่อยๆ ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับแผนภูมิองค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับโครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของ เว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดร่วมจุดเดียว นั่นคือ โฮมเพจ (Homepage)และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง



ภาพที่ 2.22 เว็บที่มีโครงสร้างแบบลำดับขั้น

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

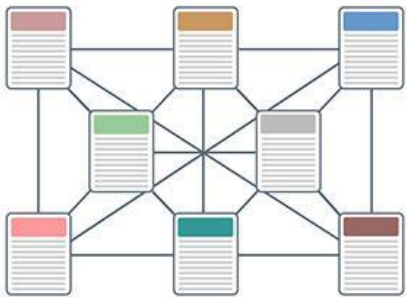


ภาพที่ 2.23 เว็บที่มีโครงสร้างแบบตาราง

ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมีลักษณะที่เหมือนกัน และ สามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมาบรรจุลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับโครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือก หัวข้อใด ก็จะเข้าไปสู่หน้าเนื้อหา (Topic Page) ที่แสดงรายละเอียดของหัวข้อนั้นๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้ารายละเอียดของ หัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำ โครงสร้างแบบเรียงลำดับและแบบลำดับ ขั้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจจะสร้างความยุ่งยากในการเข้าใจได้ และอาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive

Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ใช้ได้เข้าใจถึงความสัมพันธ์ ระหว่างเนื้อหา ในส่วนของการออกแบบจำเป็นจะต้องมีการวางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้น ได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจเกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4) เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure) โครงสร้างประเภทนี้จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกัน ได้หมด เป็นการสร้างรูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วย ตนเอง การเชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็นรูปแบบที่ ไม่มีโครงสร้างที่แน่นนอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัดเฉพาะเนื้อหาภายในเว็บนั้น ๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้



ภาพที่ 2.24 เว็บที่มีโครงสร้างแบบใยแมงมุม

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย กับข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้จะปรากฏอยู่บริเวณใด บริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการเพื่อเลือกที่จะเข้าไปสู่หน้าใด ๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการ ท่องเที่ยวบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือ ถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอจะเป็นการยากในการปรับปรุง นอกจากนี้การเชื่อมโยงระหว่างข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสน และ เกิดปัญหาการคงค้างของหัวข้อ(Cognitive Overhead)

2.2.6.4 การใช้สีในการออกแบบเว็บไซต์

การสร้างสีสันบนหน้าเว็บเป็นสิ่งที่สื่อความหมายของเว็บไซต์ได้อย่างชัดเจน การเลือกใช้สีให้เหมาะสม กลมกลืน ไม่เพียงแต่จะสร้างความพึงพอใจให้กับผู้ใช้ แต่ยังสามารถทำให้เห็นถึงความแตกต่างระหว่างเว็บไซต์ได้ สีเป็นองค์ประกอบหลักสำหรับการตกแต่งเว็บ จึงจำเป็นอย่างยิ่งที่จะต้องทำความเข้าใจเกี่ยวกับการใช้สีระบบสีที่แสดงบนจอคอมพิวเตอร์ มีระบบการแสดงผลผ่านหลอดลำแสงที่เรียกว่า CRT (Cathode ray tube) โดย มีลักษณะระบบสีแบบบวก อาศัยการผสมของของแสงสีแดง สีเขียว และสีน้ำเงิน หรือระบบสี RGB สามารถกำหนดค่าสีจาก 0 ถึง 255 ได้ จากการรวมสีของแม่สีหลักจะทำให้เกิดแสงสีขาว มีลักษณะเป็นจุดเล็ก ๆ บนหน้าจอไม่สามารถมองเห็นด้วยตาเปล่าได้ จะมองเห็นเป็นสีที่ถูกผสมเป็นเนื้อสีเดียวกันแล้ว จุดแต่ละจุดหรือพิกเซล (Pixel) เป็นส่วนประกอบของภาพบนหน้าจอคอมพิวเตอร์ โดยจำนวนบิตที่ใช้ในการกำหนดความสามารถของการแสดงสีต่าง ๆ เพื่อสร้างภาพบนจอนั้นเรียกว่า บิตเด็ป (Bit–depth) ในภาษา HTML มีการกำหนดสีด้วยระบบเลขฐานสิบหก ซึ่งมีเครื่องหมาย (#) อยู่ด้านหน้าและตามด้วยเลขฐานสิบหกจำนวนอักษรอีก 6หลัก โดยแต่ละไบต์ (byte) จะมีตัวอักษรสองตัว แบ่งออกเป็น 3 กลุ่ม เช่น #FF12AC การใช้ตัวอักษรแต่ละไบต์นี้เพื่อกำหนดระดับความเข้มของแม่สีแต่ละสีของชุดสี RGB โดย 2 หลักแรกแสดงถึงความเข้มของสีแดง 2 หลักต่อมา แสดงถึงความเข้มของสีเขียว 2 หลักสุดท้ายแสดงถึงความเข้มของสีน้ำเงินสีมีอิทธิพลในเรื่องของอารมณ์การสื่อความหมายที่เด่นชัด กระตุ้นการรับรู้ทางด้านจิตใจมนุษย์ สีแต่ละสีให้ความรู้สึก อารมณ์ที่ไม่เหมือนกัน สีบางสีให้ความรู้สึกสงบ บางสีให้ความรู้สึกตื่นเต้นรุนแรง สีจึงเป็นปัจจัยสำคัญอย่างยิ่งต่อการออกแบบเว็บไซต์ดังนั้นการเลือกใช้โทนสีภายในเว็บไซต์เป็นการแสดงถึงความแตกต่างของสีที่แสดงออกทางอารมณ์มีชีวิตชีวาหรือเศร้าโศก รูปแบบของสีที่สายตาของมนุษย์มองเห็น สามารถแบ่งออกเป็น 3 กลุ่ม คือ

1) **สีโทนร้อน (Warm Colors)** เป็นกลุ่มสีที่แสดงถึงความสุข ความปลอบโยน ความอบอุ่น และดึงดูดใจ สีกลุ่มนี้เป็นกลุ่มสีที่ช่วยให้หายจากความเฉื่อยชา มีชีวิตชีวามากยิ่งขึ้น
2) **สีโทนเย็น (Cool Colors)** แสดงถึงความที่ดูสุภาพ อ่อนโยน เรียบร้อยเป็นกลุ่มสีที่มีคนชอบมากที่สุด สามารถโน้มนาวในระยะไกลได้

3) **สีโทนกลาง (Neutral Colors)** สีที่เป็นกลาง ประกอบด้วย สีดำ สีขาว สีเทา และสีน้ำตาล กลุ่มสีเหล่านี้คือ สีกลางที่สามารถนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีกลางขึ้นมา

2.2.7 ทฤษฎีเกี่ยวกับการ visualization

Data Visualization หรือ Information Visualization คือ การนำข้อมูลในเชิงปริมาณ ทั้งที่จัดเก็บไว้ในรูปแบบของข้อมูลจำนวนน้อย และข้อมูลจำนวนมหาศาล (Big Data) มาประมวลผล จากนั้นจึงนำมาแสดงผลในรูปแบบของกราฟ แผนภูมิอินฟอร์เมชันกราฟิก หรือแม้กระทั่งอินเทอร์แอคทีฟกราฟิก ที่ผู้บริโภคสามารถคลิกหรือมีปฏิสัมพันธ์กับกราฟิกนั้น ๆ ได้ซึ่งปัจจุบันสื่อหลายสำนักในประเทศไทย เริ่มมีการนำเสนอข่าวโดยการนำข้อมูลแบบประยุกต์ด้วยภาพ (Data Visualization) การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมีประสิทธิผล ซึ่งคำว่า "ประสิทธิผล" ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้ว อาจทำให้เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์เชื่อมโยงได้

Visualization คือ การจินตนาการ หรือสร้างภาพขึ้นในความคิด ซึ่งเป็นกระบวนการทำให้เกิดความรู้ความเข้าใจสำหรับเรื่องที่จินตนาการยาก เข้าใจยาก วิธีการที่เป็นทางลัดก็คือการสร้าง ภาพ ให้เป็น บันไดความคิด ไปสู่ การใช้ความคิดอีกระดับ บันไดนี้จะช่วยตัดปริมาณ ข้อมูล ช่วยลดภาระการคำนวณหรือการนำไปผ่านหลากหลายกระบวนการความคิด เพื่อนำไปสู่คำตอบที่ต้องการได้โดยเร็วและถูกต้อง

Visualization System คือระบบ ซึ่งประกอบด้วยกระบวนการต่าง ๆ ที่ออกแบบมาเพื่อสร้าง รักษา นำไปใช้ และปรับปรุงทัศนสนเทศ เพื่อทำให้เกิดความรู้ความเข้าใจอย่าง ถูกต้อง รวดเร็ว และได้ผลเป็นอย่างสูง

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพในความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณ ให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

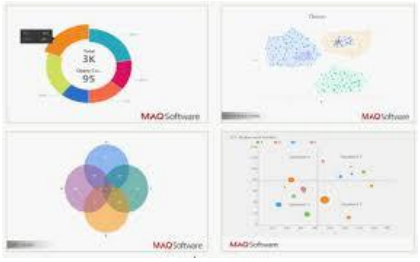
2.2.7.1 โปรแกรมสำหรับการสร้างแดชบอร์ด (Data Visualization) Toos ที่นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qik View, Google Charts, Fusion Charts,Data wrapper และอื่น ๆ อีกมากมาย

ภาพที่ 2.25 Tools Data Visualization

2.2.7.2 รูปแบบในการใช้ Data Visualization การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending) เราใช้กราฟที่แสดงผลแบบทิศทางหรือแนวโน้ม เพื่อนำเสนอข้อมูลให้เห็นจำนวนข้อมูลที่เกิดขึ้นในแต่ละช่วงเวลา (period) รวมถึงเน้นข้อมูลที่ต้องการนำเสนอ เช่น Line Chart, Bar Chart,Radar Chart, Area Chart เป็นต้น

ภาพที่ 2.26 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

2.2.7.3 การนำเสนอแบบกลุ่มข้อมูล (Classification) เป็นการนำเสนอโดยนำข้อมูลมาจัดเป็นกลุ่มๆ เช่น Donut Chart, Ring Chart, Pie Chart



ภาพที่ 2.27 การนำเสนอแบบกลุ่มข้อมูล (Classfication)

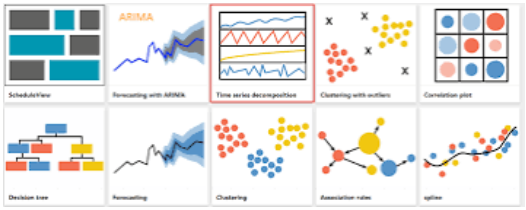
2.2.7.4 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison) เหมาะสำหรับการนำเสนอที่ต้องการเปรียบเทียบผลการดำเนินงาน เช่นเทียบกับปีที่แล้ว(YoY) เปรียบเทียบกับเป้า ที่ตั้งไว้ (Target) ซึ่งกราฟที่เหมาะสมและมักนำมาใช้ เช่น KPI Indicator, Bullet Chart, Power BI Card with state เป็นต้น

ภาพที่ 2.28 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)

2.2.7.5 การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับนำเสอข้อมูลบนแผนที่ โดยสามารถที่จะนำยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อ Focus กลุ่ม ลูกค้าในแต่ละพื้นที่ที่เราสนใจ เช่น Globe Map, Google Map, Flow Map เป็นต้น

ภาพที่ 2.29 การนำเสนอรูปแบบแผนที่ (Geographical)

2.2.7.6 กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Andlytics) เราสามารถที่จะใช้ภาษา R หรือ Python ดึงข้อมูลในอดีตมาเพื่อวิเคราะห์เพื่อพยากรณ์อนาคตและนำเสนอข้อมูลที่ได้ในรูปของกราฟ เช่น Association Rules, Clustering, Forecasting Time series, Calculation plot เป็นต้น



ภาพที่ 2.30 กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Andlytics)

ข้อดีของ Data Visualization

- ข้อมูลที่มีปริมาณมากมายมหาศาลยากต่อการเข้าใจ เราสามารถทำให้เข้าใจง่ายได้ด้วยรูปภาพ
- ช่วยจัดระเบียบความคิดวิเคราะห์ข้อมูลให้มีความน่าเชื่อถือ
- ประหยัดเวลาในการนำเสนอ

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

ภาพที่ 2.31 โปรแกรม Rapid Miner

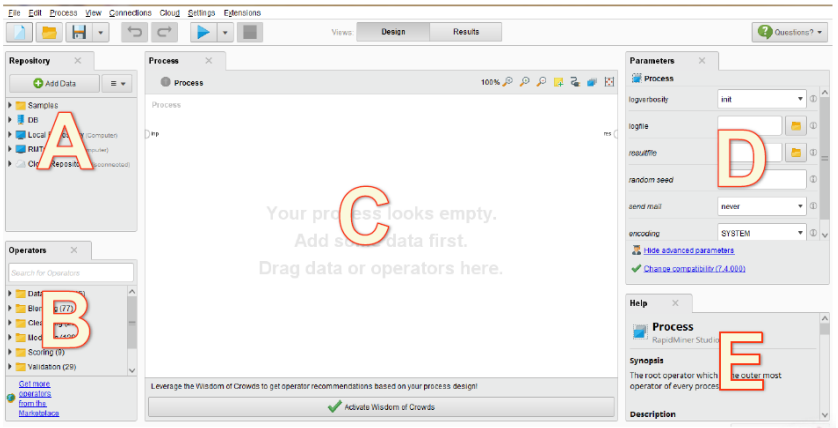
(ที่มา : https://www.dataversity.net/rapidminer-informatica-bring-ai-powered-data-analytics-enterprise/rapidminer-logo/)

โปรแกรม Rapid Miner (นฤชล โรจนบุรานนท์) ซอฟต์แวร์ RapidMiner Studio 7 แรกเริ่มพัฒนาขึ้นจากบริษัทที่ชื่อว่า Rapid-I ในประเทศเยอรมนีและเมื่อช่วงปลายปี 2013 ที่ผ่านมาได้รับทุนก้อนโตจากนักลงทุนในประเทศสหรัฐอเมริกาจึงเปลี่ยนชื่อบริษัทจาก Rapid-I เป็น RapidMiner แทน และย้ายสำนักงานใหญ่มาอยู่ประเทศสหรัฐอเมริกา สามารถดาวน์โหลดซอฟต์แวร์ RapidMiner Studio 9 ซึ่งเป็นเวอร์ชันปัจจุบันได้จากเว็บไซต์ https://rapidminer.com

2.3.1.1 ข้อดีของซอฟต์แวร์ RapidMiner Studio 9 คือ

- รองรับการใช้งานไฟล์ได้หลายประเภท เช่น ไฟล์ Excel 2007
- สามารถแสดงข้อมูลได้หลายรูปแบบ เช่น scatter plot 3D
- สามารถแสดงผลโมเดลที่สวยงามและแก้ไขการแสดงผลให้สามารถอ่านได้ง่ายขึ้น
- สามารถบันทึกไฟล์โมเดลออกเป็นไฟล์ภาพประเภทต่างๆ เช่น PNG, JPG หรือ PDF
- มีวิธีการเตรียมข้อมูล (preprocess) และการวิเคราะห์ได้หลากหลายรูปแบบ

2.3.1.2 องค์ประกอบของ RapidMiner Studio 9 (ผศ.กมลรัตน์ สมใจ)

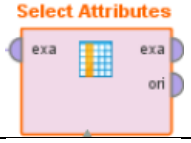
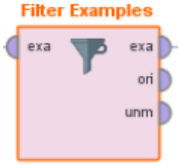


ภาพที่ 2.32 องค์ประกอบของ RapidMiner Studio 9

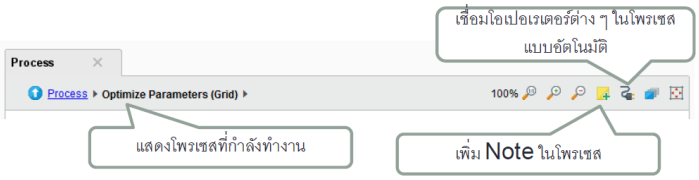
1) **A : Repositories** ส่วนนี้จะใช้ในการจัดการไฟล์ต่าง ๆ หลักการของ RapidMiner จะเก็บไฟล์ข้อมูลหรือโพรเซสต่าง ๆ ไว้ใน โฟลเดอร์เพื่อความสะดวกในการเรียกใช้งานครั้งถัดไป

2) **B : Operators** ส่วนนี้จะเก็บโอเปอเรเตอร์ในการใช้งานต่าง ๆ ไว้เป็นกลุ่มตามหน้าที่ที่คล้ายคลึงกัน และในส่วนของโอเปอเรเตอร์นี้ยังมีที่สาหรับค้นหาโอเปอเรเตอร์อีกด้วย

ตารางที่ 2.6 แสดงตัวอย่างโอเปอเรเตอร์

โอเปอเรเตอร์	หน้าที่
	อ่านไฟล์ CSV
	เลือกแอตทริบิวต์
	เลือกข้อมูล(Example)

3) C : Process ส่วนนี้เป็นอีกส่วนที่สาคัญของRapidMiner เพราะหลักการทางานของซอฟต์แวร์นี้คือ การนาเอาโอเปอเรเตอร์ ต่าง ๆ มาประกอบกันให้เป็นโพรเซสขึ้นมา

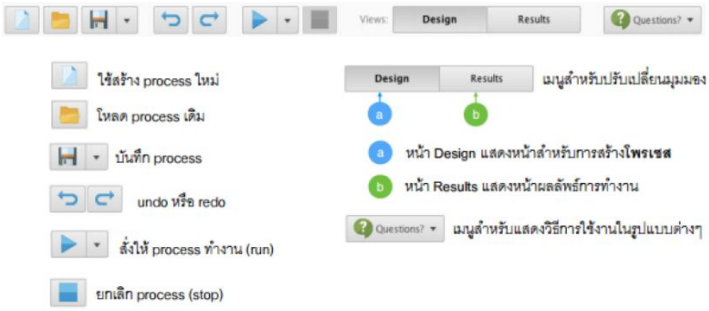


ภาพที่ 2.33 Process

4) D : Parameter ส่วนนี้จะเป็นส่วนที่แสดงพารามิเตอ์(ที่เกี่ยวข้องกับแต่ละ โอเปอเรเตอร์ เช่นโอเปอเรเตอร์Read CSV CSVสาหรับอ่านไฟล์CSVCSVจะมีพารามิเตอร์ที่เกี่ยวข้อง เช่น ชื่อและที่อยู่ของไฟล์CSVCSVเป็นต้น

5) E : Help ส่วนนี้จะเป็นส่วนที่แสดงข้อความช่วยเหลือหรือรายละเอียดของโอเปอเรเตอร์ที่ เลือกใช้งานอยู่ซึ่งประกอบด้วยรายละเอียดเบื้องต้น ความหมายของแต่ละ พารามิเตอร์ และตัวอย่างการใช้งานในส่วนท้ายสุด

6) เมนูใน RapidMiner Studio



ภาพที่ 2.34 เมนูใน RapidMiner Studio

2.3.2 โปรแกรม Xampp

ภาพที่ 2.35 โปรแกรม Xampp

(ที่มา: https://arit.rmutsv.ac.th/th/blogs/801-การติดตั้งเว็บเซิร์ฟเวอร์สำหรับทดสอบ-xampp-933)

โปรแกรม Xampp (สสวท.) คือ โปรแกรมสำหรับจำลองเครื่องคอมพิวเตอร์ส่วนบุคคลของเรา ให้ทำงานในลักษณะของ WebServer นั่นคือ เครื่องคอมพิวเตอร์ของเราจะเป็นทั้งเครื่องแม่ และเครื่องลูกในเครื่องเดียวกัน ทำให้ไม่ต้องเชื่อมต่อกับ Internet คุณก็สามารถทดสอบเว็บไซต์ที่คุณสร้างขึ้น ได้ทุกที่ทุกเวลา ปัจจุบันได้รับความนิยมจากผู้ใช้ CMS ในการสร้างเว็บไซต์ โปรแกรม XAMPP จะมาพร้อมกับ

- PHP ภาษาสำหรับพัฒนาเว็บแอพลิเคชั่นที่เป็นที่นิยม
- MySQL ฐานข้อมูล
- Apache จะทำหน้าที่เป็นเว็บเซิร์ฟเวอร์
- Perl อีกทั้งยังมาพร้อมกับ OpenSSL
- phpMyadmin ระบบบริหารฐานข้อมูลที่พัฒนาโดย PHP เพื่อใช้เชื่อมต่อไปยังฐานข้อมูล สนับสนุนฐานข้อมูล MySQL และ SQLite

โปรแกรม XAMPP จะอยู่ในรูปแบบของไฟล์ Zip, tar, 7z หรือ exe โปรแกรม XAMPP อยู่ภายใต้ใบอนุญาตของ GNU General Public License แต่บางครั้งอาจจะมีการเปลี่ยนแปลงเรื่องของลิขสิทธิ์ในการใช้งาน จึงควรติดตามและตรวจสอบโปรแกรมด้วย

โปรแกรม XAMPP พัฒนาโดยโครงการ Apache Friends ที่เป็นโครงการไม่แสวงหาผลกำไร ที่จัดตั้งในปี ค.ศ. 2002 โดย Kai 'Oswald' Seidler และ Kay Vogelgesang ทั้งนี้ XAMPP ประกอบด้วยโปรแกรมย่อยได้แก่โปรแกรม Apache โปรแกรมฐานข้อมูล MySQL โปรแกรมภาษา PHP และภาษา Perl

โปรแกรม XAMPP สามารถใช้งานได้ 4 OS ได้แก่

1. Windows สามารถใช้งานได้กับ windows รุ่น 2000, 2003, xp, vista, windows 7

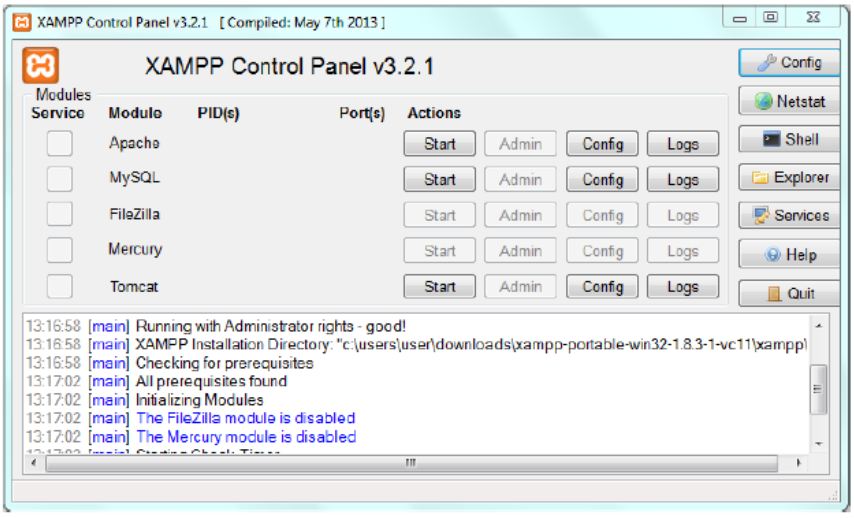
2. Linux สาหรับ SuSE, RedHat, Mandrake, Debian และ Ubuntu

3. Mac OS X

4. Solaris สำหรับ Solaris 8 และ Solaris 9

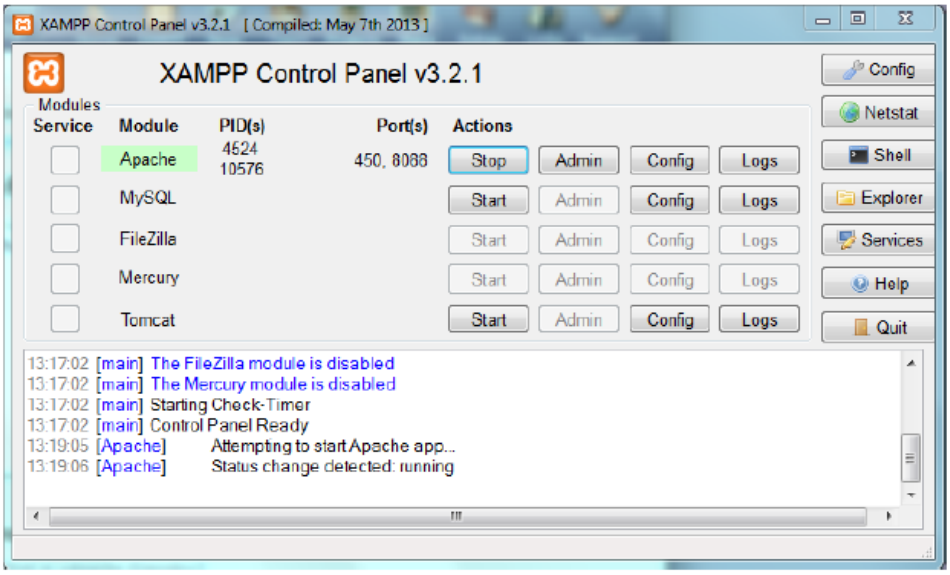
2.3.2.1 การเรียกใช้งาน XAMPP

การเรียกใช้งาน XAMPP เริ่มด้วยการดับเบิลคลิกโปรแกรม xampp-control.exe ซึ่งจะปรากฏหน้าต่างโปรแกรม XAMPP Control Panel ดังภาพที่ 2.36



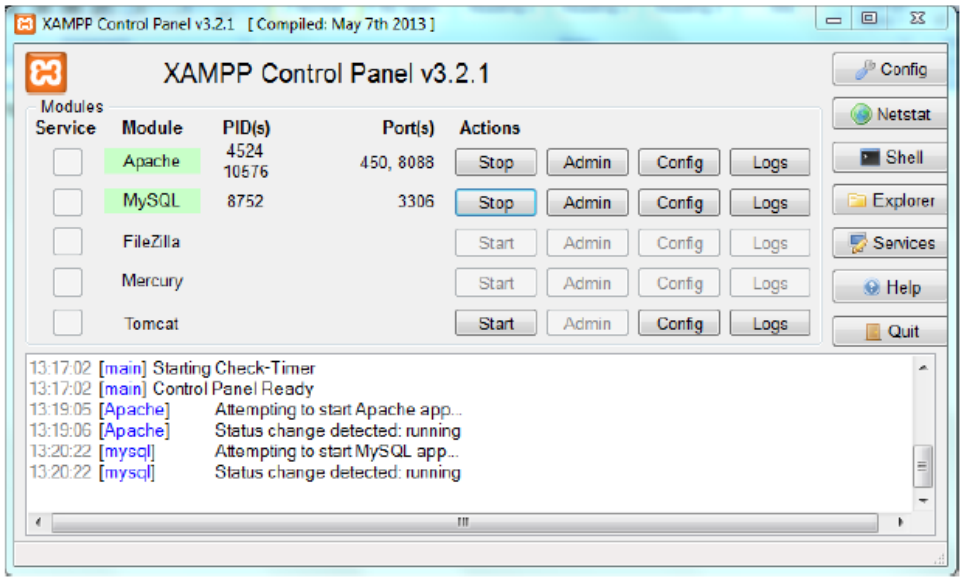
ภาพที่ 2.36 หน้าต่างโปรแกรม XAMPP Control Panel

จากหน้าต่างโปรแกรม XAMPP Control Panel ให้คลิกปุ่ม Start ของโปรแกรม Apache เพื่อรันโปรแกรม Apache ซึ่งจะแสดงสถานภาพการทำงาน ดังภาพที่ 2.37



ภาพที่ 2.37 ปุ่ม Start ของโปรแกรม Apache เพื่อรันโปรแกรม Apache

เมื่อสั่งให้โปรแกรม Apache ทำงานแล้วก็ต่อด้วยการสั่งให้โปรแกรมฐานข้อมูลทำงาน โดยคลิกปุ่ม Start ของ MySQL ดังภาพที่ 2.38



ภาพที่ 2.38 คลิกปุ่ม Start ของ MySQL เพื่อสั่งให้โปรแกรมฐานข้อมูลทำงาน

การเรียกใช้งาน Apache และ MySQL หากพบปัญหาในการเรียกใช้งาน ให้ Stop โปรแกรมที่เปิดใช้งาน และปิด XAMPP ก่อน จากนั้นให้เข้าไปยังโฟลเดอร์ของ XAMPP เพื่อปรับแก้ไข ดังนี้

1) เปิดแฟ้ม httpd.conf จาก \xampp-portable-win32-1.8.3-1-VC11\xampp\apache\conf แล้วแก้ไขรายการ

- Listen ปรับตัวเลขให้เป็น 80
- ServerName localhost ปรับตัวเลขให้เป็น 80

2) เปิดแฟ้ม php.ini จาก \xampp-portable-win32-1.8.3-1-VC11\xampp\php

- upload_max_filesize ปรับตัวเลขให้เป็น 200M
- date.timezone แก้ไขเป็น Asia/Bangkok

เมื่อปรับแก้ไขข้อมูลข้างต้น ให้รันโปรแกรมอีกครั้ง ซึ่งจะเข้าสู่หน้าเว็บ ดังภาพที่ 2.39

ภาพที่ 2.39 หน้าเว็บ XAMPP

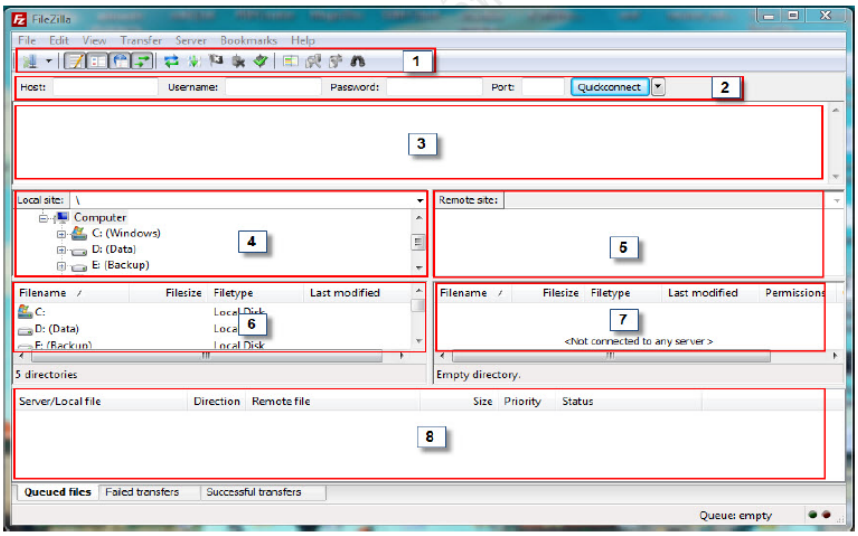
2.3.3 โปรแกรม FileZilla (สำนักบริการคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์)

ภาพที่ 2.40 โปรแกรม FileZilla

(ที่มา : https://software.thaiware.com/3723-FileZilla-Download.html)

โปรแกรม FileZilla คือ โปรแกรม FTP Client คือโปรแกรมสำหรับรับส่งข้อมูลไปยัง Server ซึ่ง FileZilla เป็นโปรแกรม OpenSource ที่สามารถนำมาใช้งานได้ฟรี

องค์ประกอบของ FileZilla แสดงดังภาพที่ 2.41



ภาพที่ 2.41 องค์ประกอบของ FileZilla

1. Genaral Toolbar -- เมนูทั่วไป
2. Quick Connect -- ล็อกอินแบบรวดเร็ว
3. Server Information -- แสดงข้อมูลจาก Server
4. Local Site Folder Tree – แสดงโฟลเดอร์ในเครื่อง
5. Remote Site Folder Tree -- แสดงโฟลเดอร์ที่อยู่ใน Server

6. Local Site Files -- แสดงไฟล์ในโฟลเดอร์ที่เลือกไว้
7. Remote Site Files -- แสดงไฟล์ในโฟลเดอร์ที่อยู่ใน Server ที่เลือกไว้
8. Queue Files -- แสดงรายชื่อไฟล์ที่จะ Upload / Download

2.3.4 โปรแกรม Power BI

ภาพที่ 2.42 โปรแกรม Power BI
(ที่มา : https://www.datacamp.com/blog/all-about-power-bi)

โปรแกรม Power BI คือ เครื่องมือในการวิเคราะห์และแสดงข้อมูลที่สร้างขึ้นโดย Microsoft เพื่อช่วยให้ผู้ใช้สามารถเชื่อมต่อและแสดงผลข้อมูลจากแหล่งต่าง ๆ ในรูปแบบที่สวยงามและออกแบบได้อย่างมีประสิทธิภาพ ซึ่งรวมถึงกราฟและผังข้อมูลที่เป็นเอกลักษณ์ การใช้งาน Power BI ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลในแบบที่น่าสนใจและให้ความหมาย โดยใช้เครื่องมือต่าง ๆ เช่น ภาพวาดข้อมูล กราฟ แผนภูมิ ตาราง แผนภูมิแท่ง แผนภูมิวงกลม และอื่น ๆ

Power BI สามารถเชื่อมต่อกับแหล่งข้อมูลต่าง ๆ ได้ รวมถึงฐานข้อมูลในองค์กร แฟ้มข้อมูล Excel หรือไฟล์ CSV บริการคลาวด์เช่น Azure SQL Database และหลายบริการอื่น ๆ โดยผู้ใช้สามารถสร้างแผนภูมิ รายงาน และแดชบอร์ดเพื่อให้ข้อมูลที่สำคัญและสื่อความหมายได้เป็นอย่างดี โดยสามารถเข้าถึงและแชร์ข้อมูลได้ทั้งในองค์กรและต่างประเทศผ่านเว็บบราวเซอร์หรืออุปกรณ์พกพา นอกจากนี้ Power BI ยังสามารถปรับแต่งและตั้งค่าเพื่อตอบสนองต่อความต้องการและรูปแบบการแสดงผลข้อมูลที่แตกต่างกันได้ อีกทั้งยังมีฟีเจอร์การเขียนสูตร การใช้งาน AI และการสร้างแผนภูมิแบบสดในเวลาจริงที่ช่วยให้ผู้ใช้ติดตามและวิเคราะห์ข้อมูลได้ในขณะเดียวกัน

ประโยชน์ของ Power BI

- Power BI ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลจากแหล่งต่าง ๆ ในรูปแบบที่มีประสิทธิภาพ ด้วยการสร้างกราฟและแผนภูมิที่สวยงาม ผู้ใช้สามารถเห็นแนวโน้มและข้อมูลที่สำคัญได้ในทันที และเข้าใจข้อมูลอย่างรวดเร็ว
- Power BI ช่วยให้ผู้ใช้สามารถเชื่อมต่อและรวมข้อมูลจากแหล่งต่าง ๆ เข้าด้วยกัน เช่น ฐานข้อมูล, แฟ้ม Excel, บริการคลาวด์ เป็นต้น การรวมข้อมูลเหล่านี้ช่วยให้ผู้ใช้สามารถมีภาพรวมและวิเคราะห์ข้อมูลอย่างครอบคลุมมากขึ้น
- Power BI มีชุดเครื่องมือที่ใช้ในการสร้างและแสดงผลรายงานและแผนภูมิที่สวยงามและมีประสิทธิภาพ ผู้ใช้สามารถสร้างแผนภูมิและกราฟที่ตอบสนองต่อความต้องการและรูปแบบการแสดงผลข้อมูลได้ตามที่ต้องการ
- ช่วยให้ผู้ใช้สามารถแสดงผลข้อมูลในเวลาจริงด้วยแผนภูมิและกราฟที่อัพเดทอัตโนมัติ ผู้ใช้สามารถติดตามข้อมูลและการเปลี่ยนแปลงได้ในขณะเกิดขึ้น ซึ่งช่วยให้ผู้ใช้ตอบสนองกับสถานการณ์ในเวลาที่เป็นปัจจุบัน
- ช่วยให้ผู้ใช้สามารถแบ่งปันรายงานและแดชบอร์ดกับผู้อื่นในองค์กร หรือแม้แต่นำเสนอข้อมูลสู่ผู้ใช้งานภายนอก ผู้ใช้สามารถเข้าถึงข้อมูลผ่านเว็บบราวเซอร์หรือแอปพลิเคชัน Power BI และแชร์ลิงก์หรือฝังรายงานในเว็บไซต์หรือแอพพลิเคชันอื่น
- Power BI ช่วยให้ผู้ใช้ตั้งค่าการติดตามและการแจ้งเตือนเพื่อเตือนให้ทราบเมื่อข้อมูลเปลี่ยนแปลงหรือเกิดเหตุการณ์ที่สำคัญ ผู้ใช้สามารถตั้งค่าการติดตามเพื่อรับข้อมูลทันทีหรือตามกำหนดเวลาที่ต้องการ
- การทำนายและวิเคราะห์ข้อมูล Power BI มีความสามารถในการใช้งาน AI (Artificial Intelligence) เพื่อทำนายและวิเคราะห์ข้อมูล ซึ่งช่วยให้ผู้ใช้สามารถระบุแนวโน้มและคาดการณ์การเกิดเหตุการณ์ในอนาคตได้อย่างมีประสิทธิภาพ
- การสร้างแผนภูมิแบบ Real Time Power BI ช่วยให้ผู้ใช้สามารถสร้างแผนภูมิและกราฟที่อัพเดทแบบสดในเวลาจริง ซึ่งช่วยในการติดตามและวิเคราะห์ข้อมูลที่กำลัง

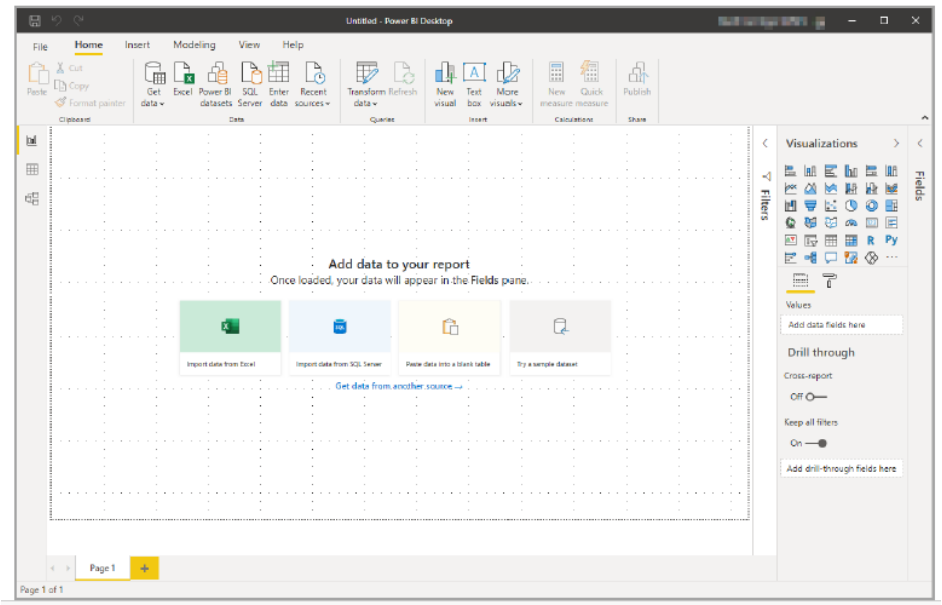
เปลี่ยนแปลงอยู่ในขณะนั้น ผู้ใช้สามารถเห็นข้อมูลในรูปแบบสดในทันทีเพื่อตอบสนองต่อสถานการณ์ปัจจุบัน

- Power BI มีเครื่องมือที่ช่วยให้ผู้ใช้สามารถเขียนสูตรหรือปรับแต่งข้อมูลได้ตามความต้องการ โดยสามารถทำการคำนวณและแปลงข้อมูลให้เหมาะสมกับการวิเคราะห์และแสดงผลได้อย่างแม่นยำ
- Power BI สามารถเชื่อมต่อกับแหล่งข้อมูลแบบเรียลไทม์ ซึ่งช่วยให้ผู้ใช้ได้รับข้อมูลที่อัพเดทอยู่ในขณะนั้นและมีการติดตามเหตุการณ์แบบเรียลไทม์ เช่น การติดตามการขายออนไลน์หรือการติดตามและวิเคราะห์การทำงานของอุปกรณ์ IOT
- Power BI มีความยืดหยุ่นสูงที่ช่วยให้ผู้ใช้งานสามารถปรับแต่งและตั้งค่าตามความต้องการขององค์กร และสามารถปรับปรุงรายงานและแผนภูมิตามความเปลี่ยนแปลงในองค์กรได้โดยง่าย
- Power BI สนับสนุนการร่วมมือและการทำงานเป็นทีม ผู้ใช้สามารถแชร์และร่วมงานกันในการสร้างรายงานและแดชบอร์ด และสามารถแสดงความคิดเห็นและประเมินผลร่วมกันได้

Power BI เป็นเครื่องมือที่มีประสิทธิภาพสำหรับการวิเคราะห์และแสดงผลข้อมูล ช่วยให้ผู้ใช้งานสร้างรายงานที่สวยงามและมีคุณภาพสูง และช่วยให้องค์กรตัดสินใจที่มีมูลค่ามากขึ้นได้

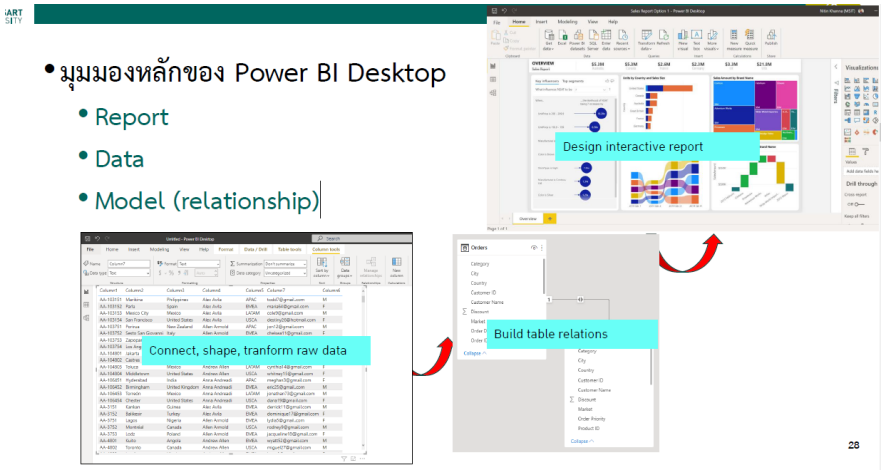
ส่วนประกอบต่าง ๆ ใน Power BI Desktop (สนผ.สพฐ,2022)

1) หน้าจอแรกของ Power BI Desktop มี Link ช่วยให้เพิ่มข้อมูลไปยังรายงาน



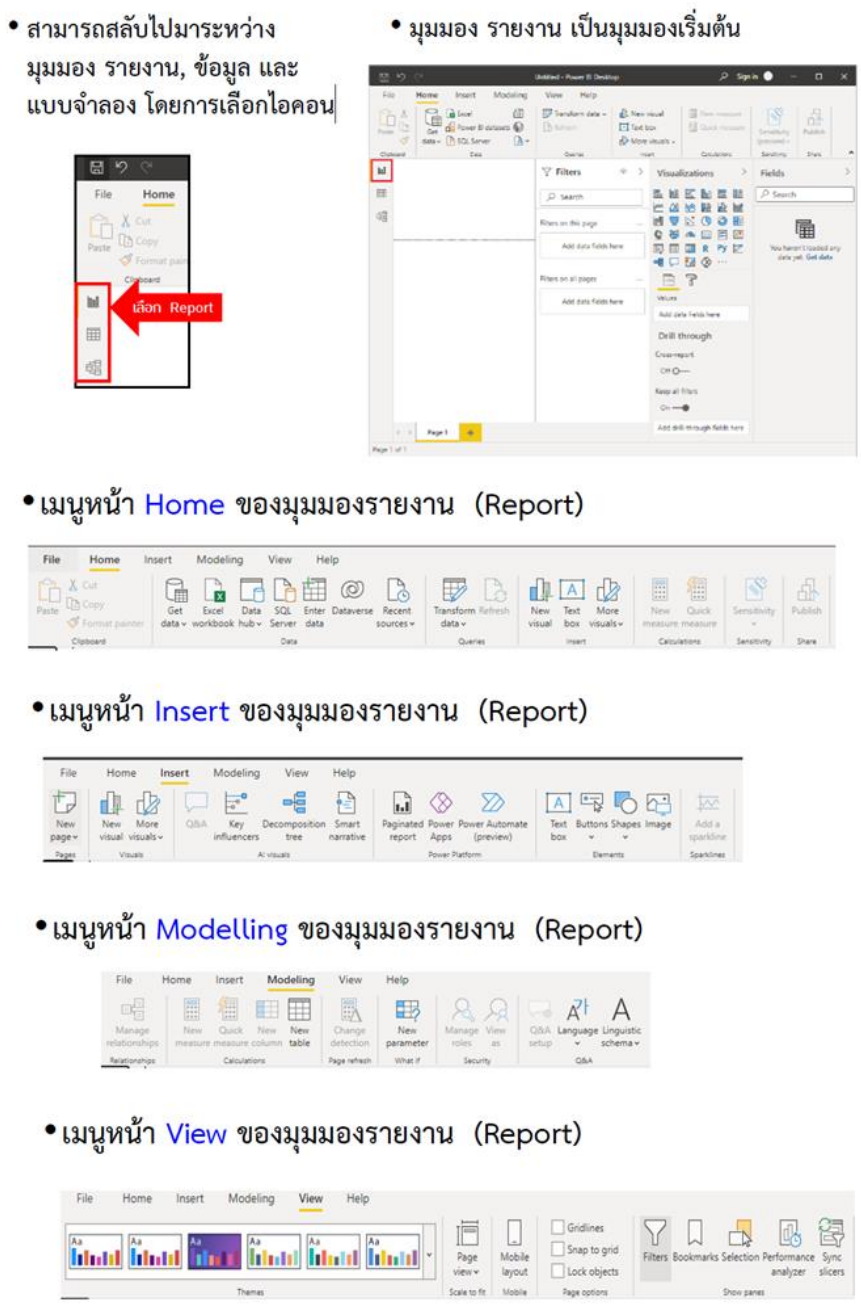
ภาพที่ 2.43 หน้าจอแรกของ Power BI

2) มุมมองหลักของ Power BI Desktop



ภาพที่ 2.44 มุมมองหลักของ Power BI Desktop

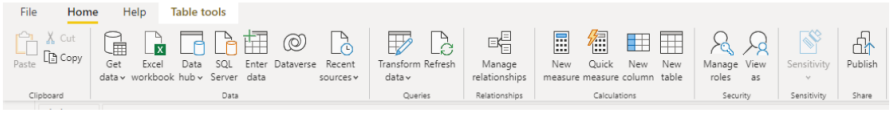
3) มุมมองรายงานของ Power BI Desktop



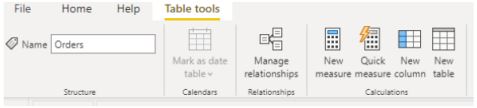
ภาพที่ 2.45 มุมมองรายงานของ Power BI Desktop

4) **มุมมองข้อมูลของ Power BI Desktop**

- เมนูหน้า Home ของมุมมองข้อมูล (Data)

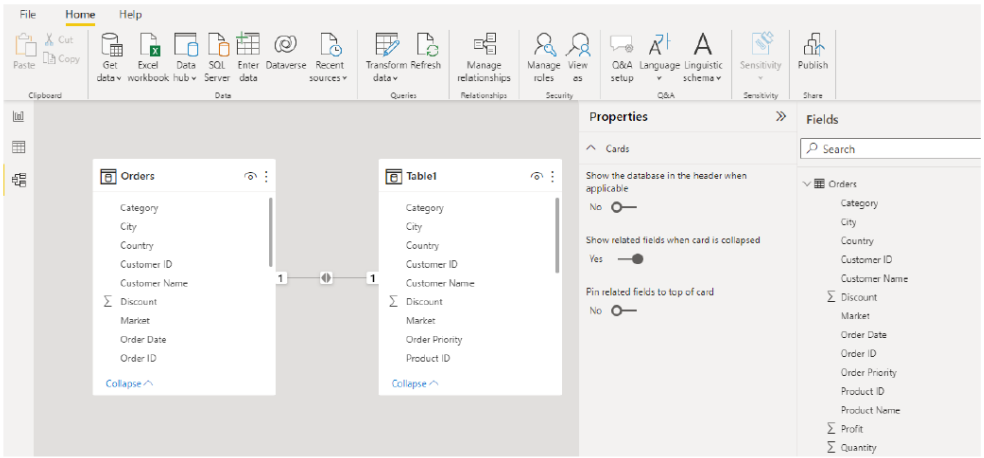


- เมนูหน้า Table tools ของมุมมองข้อมูล (Data)



ภาพที่ 2.46 มุมมองข้อมูลของ Power BI Desktop

5) **มุมมองความสัมพันธ์** (Model)



ภาพที่ 2.47 มุมมองความสัมพันธ์ (Model)

2.3.5 โปรแกรม Sublime Text 3

ภาพที่ 2.48 โปรแกรม Sublime Text 3

(ที่มา : https://software.thaiware.com/13452-Sublime-Text-Download.html)

Sublime Text 3 (Naruebet Chanthong , 2017) เป็นโปรแกรมเขียนโค้ดซึ่งสนับสนุนภาษาที่หลากหลาย C, C++, C#, CSS, D, Erlang, HTML, Groovy, Haskell, HTML, Java, JavaScript, LaTeX, Lisp, Lua, Markdown, Matlab, OCaml, Perl, PHP, Python, R, Ruby, SQL, TCL, Textile และ XML

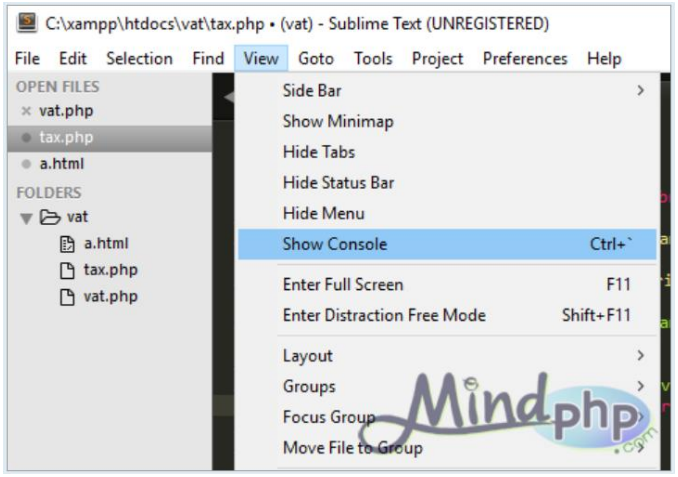
ข้อดีของ Sublime Text 3

- เร็วมาก ทั้งตอนเปิดโปรแกรม เรียกไฟล์ หรือฟังก์ชั่นต่างๆ
- Multiple Cursors สามารถ แก้ไขหลายๆ ที่ในทีเดียว ช่วยประหยัดเวลาได้มากๆ
- แต่ง Theme ได้เอง และมีแบบที่คนทำไว้ให้เยอะมาก
- Split Screen สามารถแบ่งหน้าจอการทำงานได้ แบบเป็นคอลัมป์ เป็นแถว หรือเป็น grid
- Command Palette > ทำหน้าที่คล้ายๆ spotlight ใน Mac ที่หาอะไรไม่เจอ ก็พิมพ์เข้าไป เดี๋ยวมันจะหาคำสั่งนั้นมาให้เอง สะดวกมากๆ
- Minimap สำหรับดูว่าเราแก้โค้ดส่วนไหนของไฟล์อยู่
- Sublime Package Control เป็น plugin ที่ช่วยให้เราควบคุม package ต่างๆ ที่เราจะลงเพิ่มใน sublime text ได้

เขียนโปรแกรมใน Sublime text (moomai,2017)

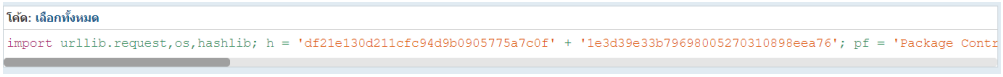
1.ติดตั้ง Package ดังนี้

1.1 เปิดตัวโปรแกรม Sublime Text 3 ขึ้นมา จากนั้นให้ไปคลิกที่ View บนแถบของโปรแกรม แล้วเลือก Show Console



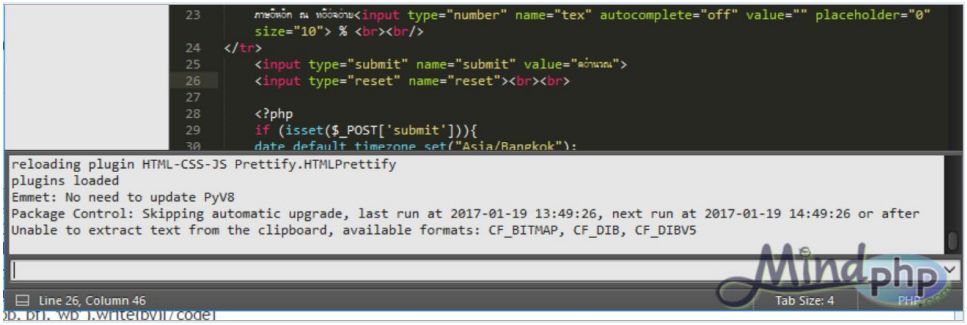
ภาพที่ 2.49 ตัวโปรแกรม Sublime Text 3

1.2 เมื่อเราคลิก Show Console แล้วให้เราคัดลอกโค้ดตามภาพที่ 2.50



ภาพที่ 2.50 โค้ดในShow Console

1.3 เมื่อเราคัดลอกโค้ดมาแล้วให้เรามาวาง ในคอนโซลที่โชว์ แล้วกด Enter

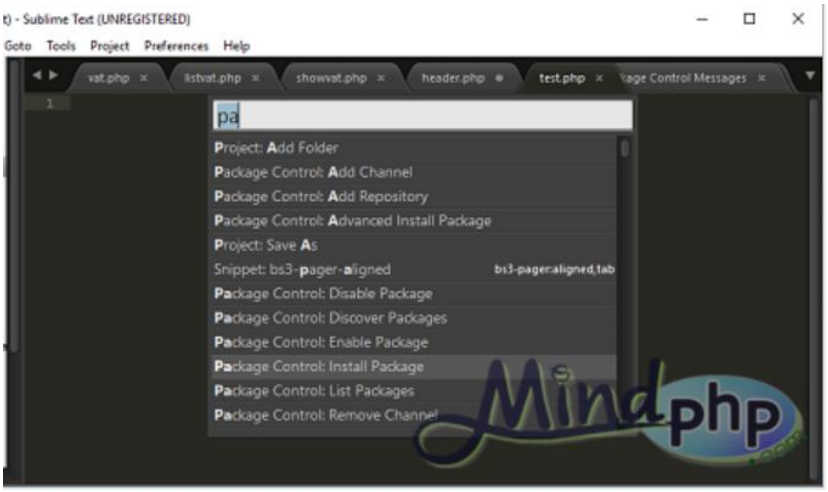


ภาพที่ 2.51 คัดลอกโค้ดวางในShow Console

1.4 เมื่อกด Enter แล้วให้รอทำการติดตั้งสักครู่ เมื่อติดตั้งเสร็จแล้วให้เราปิดโปรแกรมแล้วเปิดใหม่ก็เสร็จเรียบร้อย

ภาพที่ 2.52 ติดตั้งโค้ดในShow Console เสร็จ

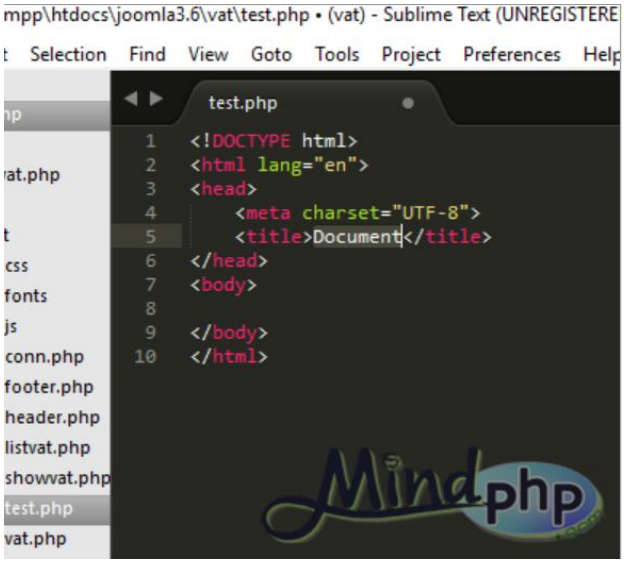
1.5 ให้เปิดโปรแกรม Sublimte Text 3 ขึ้นมาแล้วเปิดไฟล์ จากนั้นเราต้องทำการติดตั้ง package ก่อน โดยกดคีย์บรอดแป้นพิมพ์ Shift+Ctrl+P (ชิป+คอนโทล+พี) พร้อมกันแล้วพิมพ์ Package Control: Install Package แล้วกด Enter แล้วรอโปรแกรมทำการติดตั้งสักครู่ เมื่อติดตั้ง Package เสร็จเริ่มลองใช้ Tab ในการเพิ่มความเร็ว



ภาพที่ 2.53 ติดตั้ง Package

2.การใช้ Tab

2.1 เริ่มจากเปิดโปรแกรมแล้วก็เซฟไฟล์จากนั้น พิมพ์เครื่องหมาย ! แล้วก็กด Tab โค้ด HTML ก็จะปรากฏขึ้น ดังภาพที่ 2.54



ภาพที่ 2.54 โค้ด HTML

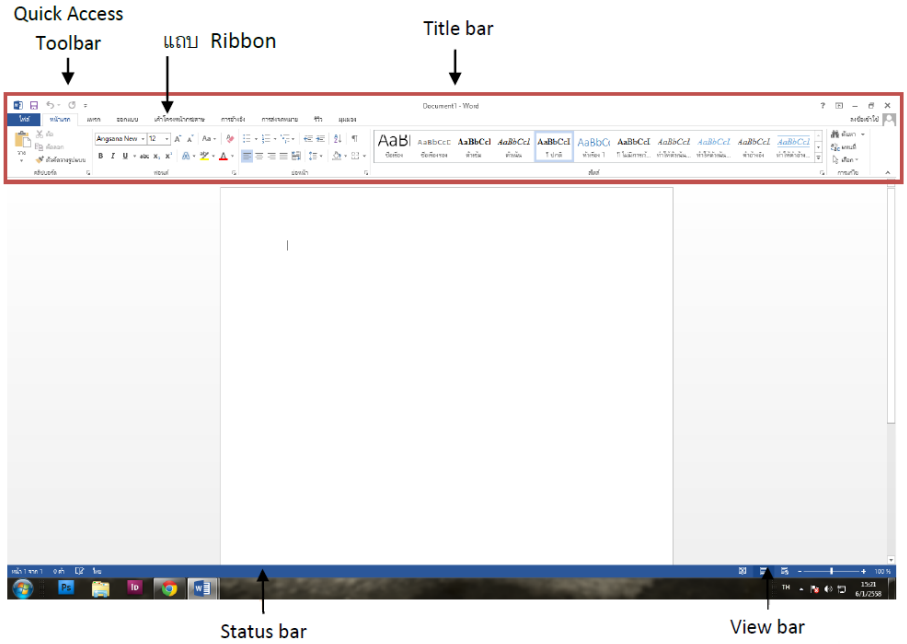
2.3.6 โปรแกรม Microsoft word

ภาพที่ 2.55 โปรแกรม Microsoft word

(ที่มา : https://race.nstru.ac.th/home_ex/blog/topic/show/5480)

โปรแกรม Microsoft Word (กรมส่งเสริมการปกครองท้องถิ่น) เป็นโปรแกรมสาหรับสร้างและจัดการกับเอกสารทุกรูปแบบ ไม่ว่าจะเป็นรายงาน จดหมาย ใบปะหน้าแฟกซ์ ไปจนถึงการจัดทาคู่มือและหนังสือ โดยมีระบบอัตโนมัติที่ช่วยในการจัดทำเอกสาร เช่น การตรวจคาสะกด การตรวจสอบไวยกรณ์ การใส่ข้อความอัตโนมัติ เป็นต้นคุณจะมีเวลาในการเขียนมากขึ้น และใช้เวลาน้อยลงในการจัดรูปแบบ

ส่วนประกอบของหน้าจอโปรแกรม (ม.ราชภัฏนครสวรรค์)



ภาพที่ 2.56 ส่วนประกอบของหน้าจอโปรแกรม Microsoft Word

– **Quick Access Toolbar** เป็นแถบเครื่องมือให้คุณเรียกใช้งานได้อย่างรวดเร็ว ผู้ใช้สามารถเพิ่มปุ่มคาสั่งที่ใช้งานบ่อยๆ ไว้ในแถบเครื่องมือนี้ได้

– **Title bar** แถบแสดงชื่อโปรแกรมและชื่อไฟล์ปัจจุบันที่เปิดใช้งานอยู่

– **แถบ Ribbon** เป็นแถบที่รวบรวมคาสั่งต่างๆ ของเมนูหรือทูลบาร์ เพื่อให้ผู้ใช้เลือกใช้งานง่ายขึ้น

– **Status bar** แถบแสดงสถานะการทางานปัจจุบันบนหน้าจอ

– **View bar** แถบแสดงมุมมองเอกสารในแบบต่างๆ

2.3.7 โปรแกรม Microsoft excel

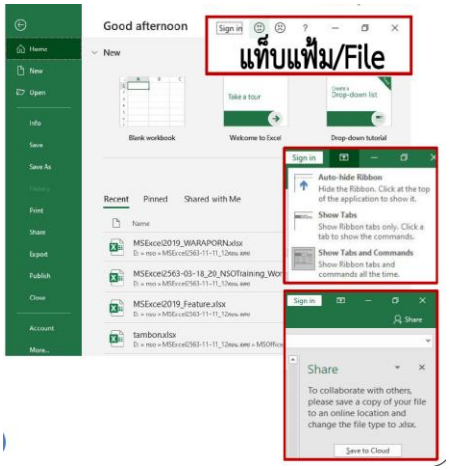
ภาพที่ 2.57 โปรแกรม Microsoft excel

(ที่มา https://en.m.wikipedia.org/wiki/File:Microsoft_Excel_2013–2019_logo.svg)

โปรแกรม Microsoft excel (วราภรณ์ ไตรศักดิ์ศรี) คือ เป็นโปรแกรมประเภท Spreadsheet ซึ่งออกแบบมาสำหรับบันทึกวิเคราะห์ และแสดงข้อมูลเกี่ยวกับตัวเลขได้อย่างมีประสิทธิภาพ ทั้งในรูปแบบของแผนภาพหรือรายงาน ซึ่งโปรแกรม Microsoft Excel ยังมีความสามารถในการจัดรูปแบบเอกสารได้สวยงาม และง่ายดายไม่แพ้โปรแกรมอื่นๆ

แท็บของโปรแกรม MS Excel

1.แท็บ File



ภาพที่ 2.58 แท็บ File

2.แท็บ Home : กลุ่มของคำสั่งที่เกี่ยวของกับการจัดรูปแบบข้อความ Font ตำแหน่งข้อความภายใน Cell เป็นต้น

ภาพที่ 2.59 แท็บ Home

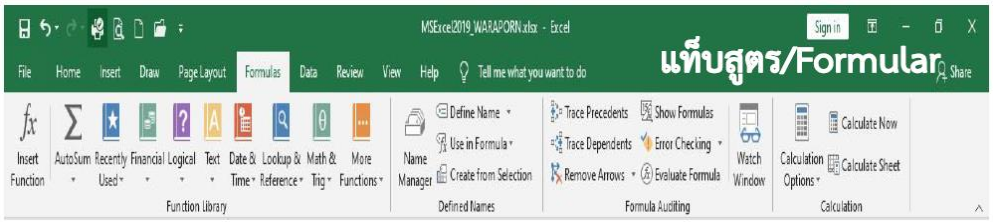
3.แท็บ Insert : กลุ่มของคำสั่งเกี่ยวกับการเพิ่มตาราง รูปภาพ ตัวอักษรประดิษฐ์ กราฟ สัญลักษณ์พิเศษ เป็นต้น

ภาพที่ 2.60 แท็บ Insert

4.แท็บ Page Layout : กลุ่มของคำสั่งเกี่ยวกับการจัดขนาดของกระดาษ การวางแนวเอกสาร ใส่สีพื้นหลัง เป็นต้น

ภาพที่ 2.61 แท็บ Page Layout

5. แท็บ Formulas : กลุ่มของคำสั่งเกี่ยวกับการใส่สูตรคำนวณ และเรียกใช้ฟังก์ชั่นต่างๆ ที่โปรแกรมจัดเตรียมไว้ให้



ภาพที่ 2.62 แท็บ Formulas

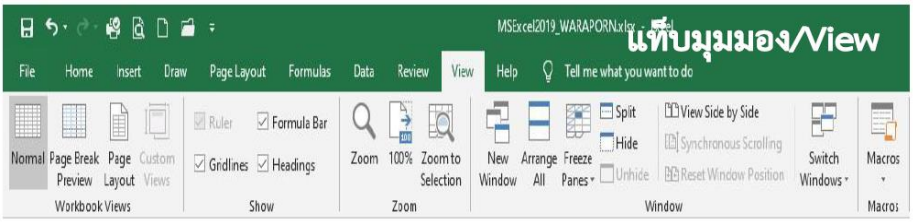
6.แท็บ Data : กลุ่มของคำสั่งเกี่ยวกับการจัดการด้านข้อมูล เช่น การนำเข้าข้อมูล การเรียงลำดับ การกรองข้อมูล เป็นต้น

ภาพที่ 2.63 แท็บ Data

7.แท็บ Review : กลุ่มของคำสั่งเกี่ยวกับการตรวจสอบการสะกด ไวยกรณ์ สร้างหมายเหตุ การป้องกันข้อมูลของแผ่นงาน การทำงานร่วมกัน เป็นต้น

ภาพที่ 2.64 แท็บ Review

8.แท็บ View : กลุ่มของคำสั่งเกี่ยวกับการจัดมุมมองการทำงานของเอกสาร เช่น การย่อ–ขยาย การแสดงGridline) เป็นต้น

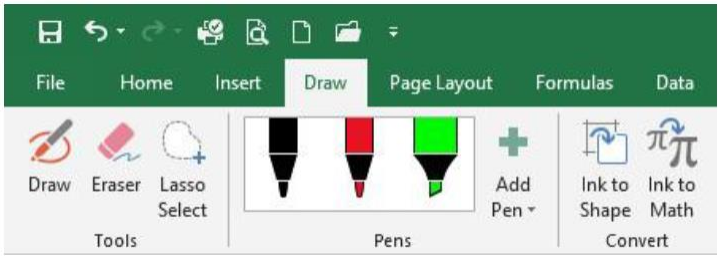


ภาพที่ 2.65 แท็บ View

9.แท็บ Developer : กลุ่มของคำสั่งในการจัดทำ Marco เพื่อใช้งานในขั้นที่สูงขึ้น

ภาพที่ 2.66 แท็บ Developer

10.แท็บ Draw : กลุ่มของชุดคำสั่งใหม่ที่เพิ่มความสะดวกในการจัดทำเอกสารที่เน้นการจัดงาน Graphic



ภาพที่ 2.67 แท็บ Draw

2.3.8 ชุดคำสั่ง HTML

ภาษา HTML (Hyper Text Markup Language) (กังวาน อัศวไชยวศิน, 2556) เป็นภาษามาตรฐานสากลที่ใช้นำเสนอ ข้อมูลแบบผสมผสานในการสื่อสารบนอินเทอร์เน็ตแบบ World – Wide – Web : WWW (Web) ซึ่งเป็นการเชื่อมต่อเครือข่ายของเครื่องคอมพิวเตอร์ทั่วโลก (Internet) รูปแบบหนึ่ง ข้อมูลในรูปแบบ ต่างๆ ไม่ว่าจะเป็นข้อความ รูปภาพ เสียง ภาพเคลื่อนไหว หรืออื่น ๆ จะถูกเชื่อมโยงเข้าหากันด้วยชุดคำสั่ง ต่าง ๆ เพื่อให้แสดงผลออกมาคล้ายกับสิ่งพิมพ์ สไลด์ หรือ แบบมัลติมีเดีย HTML มีโครงสร้างการเขียนโดยอาศัยตัวกำกับ (Tag) ควบคุมการแสดงผลข้อความ รูปภาพ หรือวัตถุอื่น ๆ ผ่านโปรแกรมเบราเซอร์ แต่ละ Tag อาจจะมีส่วนขยายที่เรียกว่า Attribute สำหรับระบุ

หรือควบคุมการแสดงผลของเว็บได้ด้วย HTML เป็น ภาษาที่ถูกพัฒนาโดย World Wide Consortium (W3C) จากแม่แบบของภาษา SGML (Standard Generalized Markup Language) โดยตัดความสามารถบางส่วนออกเพื่อให้สามารถทำความเข้าใจและเรียนรู้ได้ง่ายและด้วยประเด็นดังกล่าว ทำให้บริการ WWW เติบโตขยายตัวอย่างกว้างขวางตามไปด้วย Tag

ส่วนประกอบของไฟล์ HTML ไฟล์ HTML ประกอบด้วยส่วนสำคัญ 2 ส่วน คือ ข้อความของเอกสาร (Text) และ คำสั่งของภาษา HTML (Tag) ซึ่งบ่งบอกถึงส่วนประกอบของเอกสารโครงสร้าง รูปแบบและการลิงค์ (Link) ไปยังเอกสารอื่น ๆ หรืออาจจะรวมถึงระบบมัลติมีเดียคำสั่ง (Tag) เป็นคำสั่งหลักของ HTML แทบจะพูดได้ว่า ทุกสิ่งทุกอย่างของ HTML จะขึ้นอยู่กับ Tag ทั้งนั้นไม่ว่าจะเป็น การเน้นข้อความ การแสดงภาพประกอบ หรือการสร้างจุดเชื่อมโยง (Link) โปรแกรมบราวเซอร์ จะตีพิมพ์เอกสารออกมาในรูปแบบใด ก็โดยคำสั่ง Tag ทั้งสิ้นสามารถเขียนคำสัง (Tag) ได้ด้วยตัวอักษรใหญ่หรือเล็กจะมีความหมายเหมือนกัน ข้อบังคับข้อหนึ่งของ Tag ก็คือ จะต้องอยู่ภายในเครื่องหมาย "<" และ "y" Tag มี 2 ชนิด คือ คอนแทนเนอร์แท็ก (Container Tag) และแท็กเปล่า (Empty Tag) คอนแทนเนอร์แท็ก (Container Tag) เป็นคำสั่งที่บรรจุบางสิ่งบางอย่างอยู่ภายใน ซึ่งอาจจะเป็นข้อความ และ / หรือคำสั่ง (Tag) อื่น ๆ คอนแทนเนอร์แท็ก จะประกอบด้วย Tag เปิดและ Tag ปิด ใช้คู่กัน เรียกว่า Tag คู่ Tag เปิด จะอยู่ในรูปแบบ <Tag> โดยที่ "Tag" คือส่วนที่เป็นคำสั่งของภาษา HTML ที่ต้องการจะใช้ Tag ปิด จะอยู่ในรูปแบบ </Tag> คล้ายกับ Tag เปิด มีเพียงเครื่องหมาย "/" ที่เพิ่มเข้ามาเพื่อบอกให้รู้ว่าเป็น Tag ปิด Tag จะต้องเหมือนกันทั้งสองส่วน ข้อมูลจะบรรจุอยู่ระหว่าง Tag เปิด และ Tag ปิด เช่น Tag <body> เป็น Tag เปิด และ </body> เป็น Tag ปิด แท็กเปล่า (Empty Tag) Tag ชนิดนี้ จะต่างจากคอนแทนเนอร์แท็ก เนื่องจากไม่มีข้อมูลใด 7 บรรจุอยู่ภายใน Tag เปล่าจะมีเพียงส่วนของ Tag เปิดเท่านั้น ไม่มี Tag ปิดเรียกว่า Tag เดียว เช่น Tag <hr> Tag
 เป็นต้น ไฟล์เอกสาร HTML เป็นไฟล์ข้อความรูปแบบหนึ่ง (Text File) ที่เก็บ ชุดคำสั่ง HTML

ตารางที่ 2.7 ตารางแสดงคำสั่งภาษา HTML

คำสั่ง	ความหมาย	รูปแบบ
<title>	กำหนดข้อความบนไตเติ้ลบาร์ เป็นคำสั่งในส่วน chead>	<head> <title> ข้อความ </title> </head>
< font color size face>	–แสดงข้อความ –กำหนดสีของข้อความ –กำหนดขนาดของข้อความ –กำหนดชนิดของข้อความ	 ข้อความ
<bgcolor>	ใช้ร่วมกับ <body> หรือ <table>เพื่อกำหนดสีของพื้นหลัง	<body bgcolor = ชื่อสี>
<background>	ใช้ร่วมกับ <body> หรือ <table>เพื่อกำหนดพื้นหลังเป็นรูปภาพ	<body background = "ชื่อนามสกุล รูปภาพ>
<u>	– ข้อความหนา – ข้อความเอียง – ข้อความขีดเส้นใต้	ข้อความ ข้อความ <u> ข้อความ </u>
<marquee>	– ทำให้ข้อความเลื่อนจากซ้ายไปขวา	<marquee>ข้อความ </marquee>

ตารางที่ 2.7 ตารางแสดงคำสั่งภาษา HTML (ต่อ)

คำสั่ง	ความหมาย	รูปแบบ
<marquee Behavior= alternate>	–ทำให้ข้อความเลื่อนจากซ้ายไปขวาแล้วย้อนกลับ	‹marquee behavior=alternate> ข้อความ</marquee>
 	– ขึ้นบรรทัดใหม่	
<hr>	– ขีดเส้น	<hr>
<img src >	– แสดงรูปภาพ – กำหนดรูปภาพ	<img src– "ชื่อนามสกุลของภาพ" 0 width="ความกว้าง"0
alt width height>	–กำหนดข้อความเมื่อเอาเมาส์วางบนภาพ –กำหนดความกว้างของภาพ – กำหนดความยาวของภาพ	height="ความยาว" 0 alt="hv8;k,">เซน0<img src="pic1.jpg width= "80"height = "250" alt="ภาพท้องฟ้า">
<a href target>	– กำหนดการเชื่อมโยง – กำหนดจุดเชื่อมโยง – กำหนดการเปิดหน้าต่างใหม่	–ข้อความ เช่น <a href= "http://www.holy.ac.th" target="_blank'"›ภาควิชาวิศวกรรมไฟฟ้า
<table width bgcolor border bordercolor> <tr> <td>	– สร้างตาราง – กำหนดความกว้างของตาราง – กำหนดสีพื้นของตาราง	เช่น <table width="80%" gcolor=blue border="2" bordercolor=red> <tr><td>เกียร์</td> <td>ไฟฟ้า</td>

ตารางที่ 2.7 ตารางแสดงคำสั่งภาษา HTML (ต่อ)

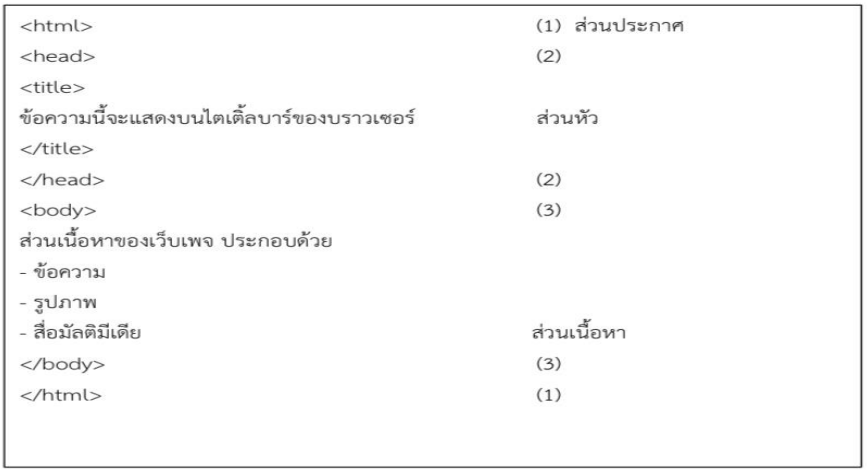
คำสั่ง	ความหมาย	รูปแบบ
	–กำหนดความหนาของเส้นขอบ – กำหนดสีของเส้นขอบ	</tr>
	– กำหนดจำนวนแถว (แนวนอน) – กำหนดจำนวนคอลัมน์ (แนวตั้ง)	จะได้ ตาราง 1 แถว 2 คอลัมน์ ความยาว 8 ของหน้าจอ สีน้ำเงิน และขอบ 2 pixel สีแดง คอลัมน์ที่ 1 มีข้อความคอลัมน์ที่ 2 มีข้อความว่า ไฟฟ้า

โครงสร้างหลักของภาษา HTML (Hyper Text Markup Language) ในการเขียนภาษา HTML นั้นจะมีรูปแบบโครงสร้างการเขียนแบ่งออกเป็น 3 ส่วนดังนี้

1) ส่วนประกาศ เป็นส่วนที่กำหนดให้บราวเซอร์ทราบว่า นี่คือภาษา HTML และจะต้องทำการแปรผลอย่างไรมีคำสั่งคู่เดียวคือ <html> และ </html> ปรากฏที่หัวและท้ายไฟล์

2) ส่วนหัวเรื่อง (head) เป็นส่วนที่แสดงผลข้อความบนไตเติ้ลบาร์ของบราวเซอร์ และอาจมีคำสั่งสำหรับกำหนดรายละเอียดด้านเทคนิคอื่น ๆ แทรกอยู่ระหว่างคำสั่ง ‹head> และ </head>เป็นคำสั่งที่ใช้กำหนดข้อความที่ต้องการนำมาแสดงผลบนแถบ Title Bar คำสั่งนี้จะอยู่ภายในคำสั่ง ส่วน <head>...../head> โดยกำหนดความยาวของตัวอักษรไม่เกิน 64 ตัวอักษร

3) ส่วนเนื้อหา (body) เป็นส่วนที่มีความซับซ้อนมากที่สุด และสามารถใส่เทคนิคลูกเล่นเพื่อ ดึงดูดความสนใจได้มากความแตกต่างระหว่างเว็บไซต์ต่าง ๆ แสดงความมีฝีมือของผู้จัดทำศิลปะในการออกแบบจะอยู่ในส่วนนี้ทั้งหมด ซึ่งจะแทรกอยู่ระหว่างคำสั่ง <body> และ </body> เป็นคำสั่ง ที่กำหนดข้อความและรูปแบบของคำสั่งต่าง ๆ ที่ใช้สำหรับปรับแต่งเอกสารที่จะนำเสนอออกทางส่วนแสดงผลหลักของ Web Browser



ภาพที่ 2.68 โครงสร้างหลักของภาษา HTML (Hyper Text Markup Language)
(ที่มา : กังวาน อัศวไชยวศิน, 2556)

2.3.9 ชุดคำสั่ง CSS

ภาษา CSS (Cascading Style Sheet) (เนื่องวงศ์ ทวยเจริญ, 2560)หรือ สไตล์ชีส เป็นภาษาที่ใช้เป็นส่วนของการ จัดรูปแบบการแสดงผลของภาษา HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบของเนื้อหาในเอกสารซึ่งสามารถกำหนดรูปแบบของสีข้อความสีพื้นหลัง ประเภทตัวอักษร เป็นต้น ซึ่งการกำหนด รูปแบบ หรือ Stle นี้ใช้หลักการของการแยกเนื้อหาของ ภาษา HTML ออกจากคำสั่งที่ใช้ในรูปแบบการแสดงผล กำหนดให้รูปแบบของการแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลลัพธ์ของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือต้องการควบคุมให้รูปแบบการแสดงผลเอกสาร HTML

คำสั่งพื้นฐานของภาษา CSS (Cascading Style Sheet) คำสั่งพื้นฐานที่ใช้ในการกำหนดรูปแบบของเนื้อหา มีตัวอย่างดังนี้

ตารางที่ 2.8 คำสั่งพื้นฐานของภาษา CSS (Cascading Style Sheet)

คำสั่ง	อธิบาย
width : 100px;	กำหนดความกว้าง
min–width : 100px;	กำหนดความกว้าง ขั้นต่ำ

ตารางที่ 2.8 คำสั่งพื้นฐานของภาษา CSS (Cascading Style Sheet) (ต่อ)

คำสั่ง	อธิบาย
max-width : 100px;	กำหนดความกว้าง กว้างสุด
height: 100px;	กำหนดความสูง
min-height: 100px;	กำหนดความสูง ต่ำสุด
max-height : 100px;	กำหนดความสูง สูงสุด
font-size: 14px;	กำหนดขนาดตัวอักษร
font-weight: normal;	กำหนดความหนาตัวอักษร
font-family : 'sans-serif;	กำหนดรูปแบบตัวอักษ
text-align: center;	จำตำแหน่งตัวอักษร
line-height: 1.5;	กำหนดส่วนสูงของบรรทัด
color: #286F65;	กำหนดสีตัวอักษร
white-space: nowrap;	กำหนดให้ตัวอักษรไม่ให้ขึ้นบรรทัดใหม่ใส่ช่องว่างภายใน
padding : 10px 15px;	ใส่ช่องว่างภายใน
margin: 10px 15px;	ใส่ช่องว่างภายนอก
border: 5px solid #000;	ใส่เส้นขอบใส่เงา
box-shadow: 0 0 8px rgba(0, 0, 0, 0.5);	ใส่เงา
float: left;	จัดชิดซ้าย

ตารางที่ 2.8 คำสั่งพื้นฐานของภาษา CSS (Cascading Style Sheet) (ต่อ)

คำสั่ง	อธิบาย
float: right;	จัดชิดขวา
clear: foth;	เคลียร์ float
overflow: hidden;	การสร้างขอบเขตถ้าอยู่นอกขอบเขตจะไม่แสดงผล
opacity : 0.5;	ใส่ค่าความโปร่งทึบ
display: block;	แสดงผลแบบเต็มช่อง
display: inline;	แสดงผลต่อกันในบรรทัด
Display:none;	ปิดการแสดงผล
Background – color: #00ff00;	ใส่สีพื้นหลัง
Background – image url"paper. jpg");	ใส่รูปพื้นหลัง
Position: relative;	การจัดการเรื่องตำแหน่งการแสดงผล
Content : '....';	ใส่ข้อความ
Cursor : pointer;	กำหนดรูปแบบเคอร์เซอร์
Transition : all 1s ease–in–out;	การทำอนิเมชันด้วย CSS

2.3.9 แนวคิดแผนภาพกระแสข้อมูล (Concept of Data Flow Diagram)

2.3.9.1 แผนภาพกระแสข้อมูล (Data Flow Diagram: DFD) (เถกิง วงศ์ศิริโชติ,2015)

แผนภาพกระแสข้อมูล (Data Flow Diagram: DFD) คือ แบบจำลองกระบวนการ (Process Model)ประเภทหนึ่ง มักนำมาใช้ในการวิเคราะห์และออกแบบระบบเชิงโครงสร้าง โดยแสดงความสัมพันธ์ระหว่างกระบวนการกับข้อมูล ซึ่งเป็นการแสดงการเคลื่อนที่ของข้อมูลจากกระบวนการหนึ่งไปยังกระบวนการหนึ่งโดยไม่อิงเทคโนโลยีการจัดเก็บข้อมูล เช่น สื่อบันทึกข้อมูล ฯลฯ

2.3.9.2 คำศัพท์ที่เกี่ยวข้อง

1) กระบวนการ (Process) แสดงการทำงาน หรือการประมวลผลของระบบ โดยที่ตั้งชื่อเพื่อให้สื่อถึงว่า ณ.ตำแหน่งหนึ่ง ๆ ระบบต้องมีกระบวนการทำงานอย่างไรตามลำดับ

2) การเคลื่อนที่ของข้อมูล (Data Flow) แสดงการเคลื่อนที่ของข้อมูลจากตำแหน่งหนึ่งไปยังอีกตำแหน่งหนึ่งของแผนภาพ โดยใช้สัญลักษณ์เส้นและมีลูกศรแสดงทิศทางการเคลื่อนที่ของข้อมูล การเคลื่อนที่ของข้อมูลแสดงเฉพาะการเคลื่อนที่ของข้อมูลเท่านั้น เอกสารบางอย่างถูกจัดอยู่ในกลุ่มของข้อมูล เช่น รายงานต่าง ๆเป็นต้น

3) หน่วยภายนอก (External Entity) แสดงหน่วยภายนอกที่ติดต่อกับระบบโดยอ้างอิงจากขอบเขตของระบบ (System Boundary) หน่วยภายนอกของระบบ ได้แก่ บุคคล กลุ่มคน หน่วยงาน ระบบสารสนเทศ เป็นต้น ซึ่ง มีหน้าที่หลัก คือ ส่งข้อมูลเข้าและ/หรือ รับข้อมูลที่ออกจากระบบที่กำลังศึกษา

4) แฟ้มข้อมูล (Data Store / File) แสดงที่เก็บข้อมูลภายในระบบที่กำลังศึกษาที่เก็บข้อมูล ได้แก่ แฟ้มข้อมูลแฟ้มเอกสาร ตู้เอกสาร เป็นต้น

	ภาพปัจจุบัน	ภาพใหม่
กายภาพ	1. ศึกษาระบบปัจจุบันเชิงกายภาพ	3. ศึกษาระบบใหม่ (ที่กำลังพัฒนา) เชิงกายภาพ
ตรรกะ	2. ศึกษาระบบปัจจุบันเชิงตรรกะ	4. ศึกษาระบบใหม่ (ที่กำลังพัฒนา) เชิงตรรกะ

ภาพที่ 2.69 ขั้นตอนการวิเคราะห์และออกแบบระบบ

2.3.9.3 สัญลักษณ์ที่ใช้ในการวาดแผนภาพกระแสข้อมูล

สัญลักษณ์ที่นิยมใช้ในการวาดแผนภาพกระแสข้อมูลมี 2 แบบ คือ

1. สัญลักษณ์แบบ Gane and Sarson Symbol

ตารางที่ 2.9 ตารางแสดงสัญลักษณ์แบบ Gane and Sarson Symbol

สัญลักษณ์	คำอธิบาย
	กระบวนการ (Process)
	การเคลื่อนที่ของข้อมูล (Data Flow)
	หน่วยภายนอก (External Entity)
	แฟ้มข้อมูล (Data Store/File)

2. สัญลักษณ์แบบ Yourdons

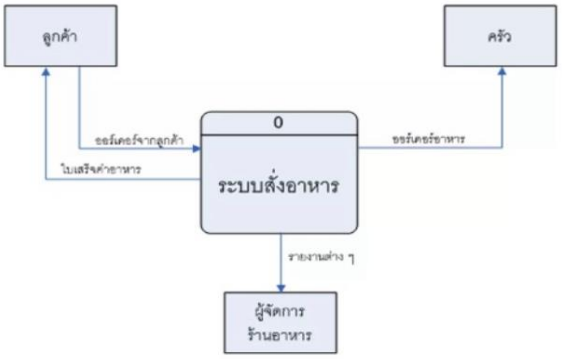
ตารางที่ 2.10 ตารางแสดงสัญลักษณ์แบบ Yourdons

สัญลักษณ์	คำอธิบาย
	กระบวนการ (Process)
	การเคลื่อนที่ของข้อมูล (Data Flow)
	หน่วยภายนอก (External Entity)
	แฟ้มข้อมูล (Data Store/File)

2.3.9.4 ลำดับแผนภาพกระแสข้อมูล (Data Flow Diagram Hierarchy)

ลำดับแผนภาพกระแสข้อมูล คือ ลำดับขั้นของการแสดงแผนภาพ โดยเริ่มจากภาพรวม (ภาพบนสุด)ของระบบและทำการแยกออกเป็นส่วนหรือการทำให้ภาพปัจจุบันละเอียดขึ้น(Functional Decomposition) ลำดับแผนภาพกระแสข้อมูลแบ่งออกเป็น 3 ลำดับ คือ

1. แผนภาพบริบท (Context Diagram) คือ แผนภาพที่อยู่บนสุด แสดงภาพรวมของระบบทั้งหมด ซึ่งประกอบด้วยกระบวนการเพียงกระบวนการเดียว (อยู่ตรงกลางของภาพ) นั่นคือ ระบบที่กำลังศึกษา ณ. ปัจจุบัน บุคคล/ระบบภายนอก และการเคลื่อนที่ของข้อมูลจากภายนอกระบบสู่ระบบ ดังตัวอย่าง

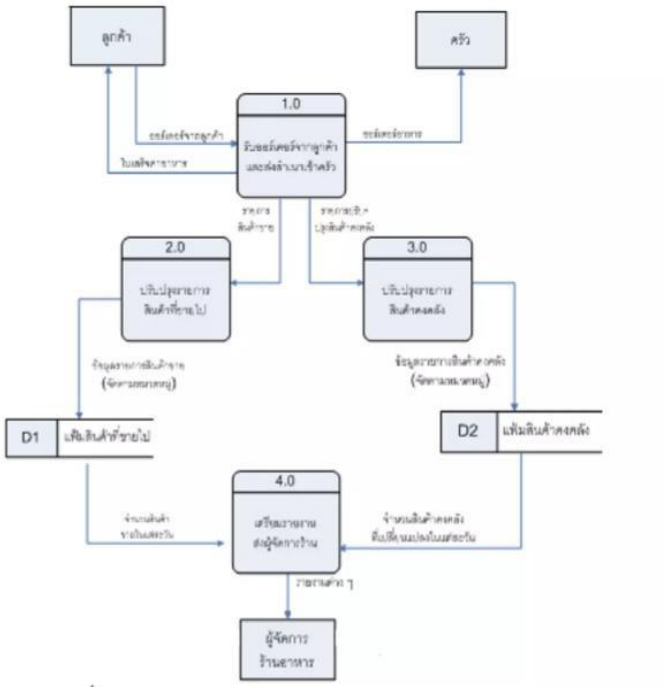


ภาพที่ 2.70 แผนภาพบริบท (Context Diagram)

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

2. แผนภาพกระแสข้อมูลระดับที่ 1 (DFD-1)

แผนภาพกระแสข้อมูลระดับที่ 1 คือ แผนภาพบริบทซึ่งผ่านการแยกออกเป็นส่วน ๆ โดยที่นักวิเคราะห์ระบบยังคงวาด บุคคล/ระบบภายนอก และการเคลื่อนที่ของข้อมูลจากภายนอกระบบสู่ระบบไว้ แต่จะทำการแยกกระบวนการหลัก (ซึ่งอยู่ตรงกลางภาพในแผนภาพบริบท) ให้ออกเป็นกระบวนการย่อย ๆ (Sub Processes) ในแผนภาพกระแสข้อมูลระดับที่ 1 นักวิเคราะห์ระบบใช้ตัวเลขเพื่อเรียกกระบวนการ โดยเริ่มจาก 1.0 ไปเรื่อย ๆ จนกว่าจะครบทุกกระบวนการ หลักการให้ตัวเลขกับกระบวนการนั้น ไม่มีกฎเกณฑ์แน่นอนแต่นิยมให้ตัวเลขเพิ่มขึ้นตามการเคลื่อนที่ของข้อมูลจากกระบวนการหนึ่งสู่กระบวนการหนึ่งหากมีการจัดเก็บหรือค้นคืนข้อมูลจากแฟ้มข้อมูล ให้วาดเส้นเชื่อมต่อระหว่างแฟ้มข้อมูลกับกระบวนการที่เกี่ยวข้อง

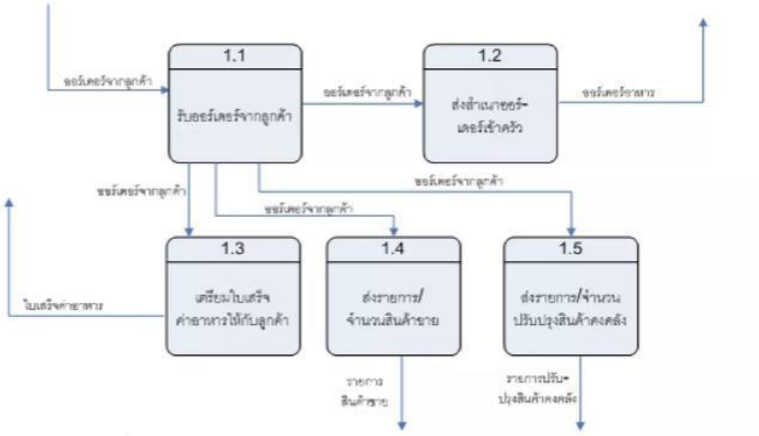


ภาพที่ 2.71 แผนภาพกระแสข้อมูลระดับที่ 1 (DFD–1)

(ที่มา : https://www.slideshare.net/slideshow/ss–47598456/47598456)

3. แผนภาพกระแสข้อมูลระดับที่ 2 (DFD–2)

แผนภาพกระแสข้อมูลระดับที่ 2 คือ แผนภาพกระแสข้อมูลระดับที่ 1 ซึ่งผ่านการแยกออกเป็นส่วน ๆ โดยนักวิเคราะห์ระบบจะเริ่มจากการเลือกกระบวนการที่อยู่ในแผนภาพกระแส ข้อมูลระดับที่ 1 ทีละ 1 กระบวนการ เช่น เลือกกระบวนการหมายเลข 1.0 หลังจากเลือกกระบวนการที่ต้องการศึกษา (1.0) แล้ว นักวิเคราะห์ระบบยังคงวาด การเคลื่อนที่ของข้อมูลเข้า และการเคลื่อนที่ของข้อมูลออก ที่มีผลต่อกระบวนการที่เลือก (1.0) และทำการแยกกระบวนการที่เลือก(1.0) ให้ออกเป็นกระบวนการย่อย ๆ ในแผนภาพกระแสข้อมูลระดับที่ 2 นักวิเคราะห์ระบบใช้ตัวเลขเพื่อเรียกกระบวนการ ซึ่งเลขตัวแรกเป็นหมายเลขของกระบวนการที่เลือก และเมื่อได้กระบวนการที่ได้รับการที่ผ่านการแยกกระบวนการ เช่น เลือกกระบวนการหมายเลข 1.0 (ในระดับ DFD–I) ดังนั้น ในแผนภาพกระแส ข้อมูลระดับที่ 2 กระบวนการย่อยที่อยู่ภายในกระบวนการหมายเลข 1.0 ได้แก่ 1.1 1.2 1.3 ...ตามลำดับ

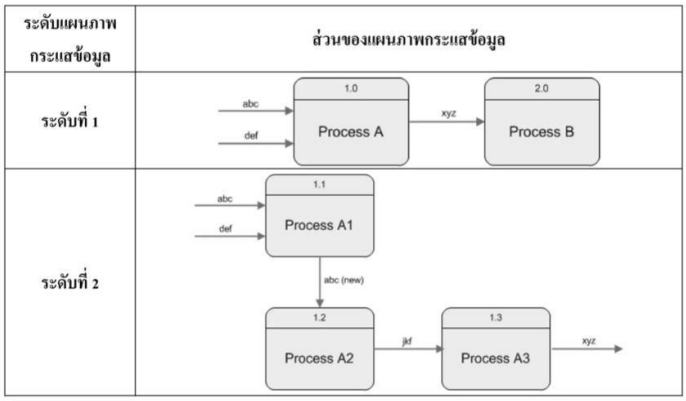


ภาพที่ 2.72 แผนภาพกระแสข้อมูลระดับที่ 2 (DFD–2)

(ที่มา : https://www.slideshare.net/slideshow/ss–47598456/47598456)

4. แผนภาพกระแสข้อมูลระดับอื่น ๆ

แผนภาพกระแสข้อมูลระดับอื่น ๆ ที่ลงลึกกว่าระดับที่ 2 ทำให้นักวิเคราะห์ระบบมองภาพของระบบนั้น ๆ ละเอียดขึ้นความสมดุลของแผนภาพกระแสข้อมูล (Balancing Data Flow Diagram) การเขียนแผนภาพกระแสข้อมูลนั้น สิ่งที่ต้องคำนึงถึงมากที่สุดอย่างหนึ่ง คือ ความสมดุลของแผนภาพกระแสข้อมูลในแต่ละระดับ เช่น กระบวนการ 1.0 ใน DFD ระดับที่ 1 มีข้อมูลเคลื่อนที่จาก 2แหล่งข้อมูล และมีการเคลื่อนที่ออกของข้อมูล 1 อย่างไปยังกระบวนการ 2.0 หากนักวิเคราะห์ระบบต้องการแยกกระบวนการ 1.0 ออกเป็นส่วน ๆ ลงไปอีกหนึ่งระดับ คือ DFD ระดับที่ 2 ต้องคำนึงถึงข้อมูลเข้าจำนวน 2 เส้น และข้อมูลออกจำนวน 1 เส้น ดังตัวอย่าง



ภาพที่ 2.73 แผนภาพกระแสข้อมูลระดับอื่น ๆ

(ที่มา : https://www.slideshare.net/slideshow/ss–47598456/47598456)

2.2.9.5 ข้อบังคับการวาดแผนภาพกระแสข้อมูล

1.กระบวนการ

1.1. หลักการตั้งชื่อกระบวนการ มักใช้คำกริยาขึ้นต้นเพื่อแสดงถึงการทำงานของกระบวนการนั้น ๆ

1.2. ตัวเลขที่ใช้เรียกกระบวนการ ในแผนภาพบริบทมีเพียงตัวเลขเดียว คือ 0 ศูนย์) ซึ่งใช้แทนระบบทั้งระบบ ในแผนภาพกระแสข้อมูลระดับ 1 (DFD–I) ใช้ตัวเลข 1.0 2.0 3.04.0 .. ตามลำดับ เท่ากับจำนวนของกระบวนการที่มีอยู่ในระดับนี้ทั้งหมด ในแผนภาพกระแสข้อมูลระดับ 2 (DFD–2) ใช้ตัวเลข 1.11.2 1.3 1.4... ตามลำดับ เท่ากับจำนวนของกระบวนการที่มีอยู่ในระดับนี้ทั้งหมด

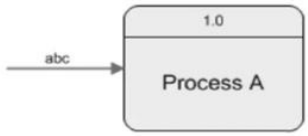
1.3. กระบวนการต้องมีการเคลื่อนที่ของข้อมูลเข้า (Data Input Flow) และ การเคลื่อนที่ของข้อมูลออก (Data Output Flow) อย่างน้อยอย่างละ 1 เส้นเสมอ อย่างไรก็ตาม ไม่มีการกำหนดจำนวนสูงสุดเส้นของการเคลื่อนที่ของข้อมูลเข้าและออก นักวิเคราะห์ระบบต้องคำนึงถึงความเหมาะสมความเพียงพอของข้อมูลต่อการประมวลผล และความสอดคล้องของข้อมูล

1.4. โดยทั่วไป ชื่อของข้อมูลที่เคลื่อนที่เข้า และชื่อของข้อมูลที่เคลื่อนที่ออก จะไม่ซ้ำกัน หากมีความจำเป็นต้องใช้ชื่อเดียวกัน ให้เขียนวงเล็บต่อท้ายชื่อของข้อมูลที่เคลื่อนที่ออกว่า new หรือ ใหม่

ภาพที่ 2.74 ลักษณะแผนภาพแผนภาพกระแสข้อมูลโดยชื่อของข้อมูลที่เคลื่อนที่เข้าและชื่อของข้อมูลที่เคลื่อนที่ออก ใช้ชื่อเดียวกัน

(ที่มา : https://www.slideshare.net/slideshow/ss–47598456/47598456)

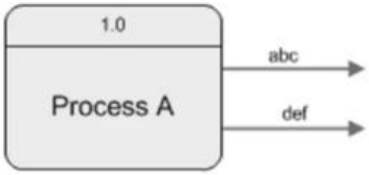
1.5. ในแผนภาพกระแสข้อมูลทุกระดับ ต้องไม่มีกระบวนการใด ที่มีการเคลื่อนที่ของข้อมูลเข้า และไม่มีการเคลื่อนที่ของข้อมูลออก ลักษณะแบบนี้ นักวิเคราะห์ระบบเรียกว่า หลุมดำ (Black Hole) ซึ่ง ขัดแย้งกับแนวคิดกิจกรรมทางสารสนเทศ (Information System Activities) ซึ่งอธิบายว่า ทุกกระบวนการต้องมีการเคลื่อนที่ของข้อมูลเข้า และการเคลื่อนที่ของข้อมูลออก อย่างน้อย 1 เส้นเสมอ



ภาพที่ 2.75 ลักษณะแผนภาพแผนภาพกระแสข้อมูลแบบมีการเคลื่อนที่ของข้อมูลเข้า แต่ไม่มีการเคลื่อนที่ของข้อมูลออก

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

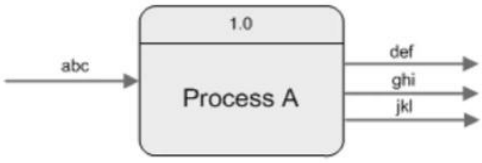
1.6. จากข้อ 1.5 ในทางกลับกัน ต้องไม่มีกระบวนการใด ที่มีการเคลื่อนที่ของข้อมูลออกเพียงอย่างเดียวโดยที่ไม่มีการเคลื่อนที่ของข้อมูลเข้า



ภาพที่ 2.76 ลักษณะแผนภาพแผนภาพกระแสข้อมูลแบบไม่มีการเคลื่อนที่ของข้อมูลเข้า แต่มีการเคลื่อนที่ของข้อมูลออก

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

1.7. จำนวนของข้อมูลที่เคลื่อนที่เข้าสู่กระบวนการต้องเพียงพอต่อการทำงาน ซึ่งทำให้เกิดการเคลื่อนที่ของข้อมูลออก หากข้อมูลที่เคลื่อนที่เข้าไม่เพียงพอ ลักษณะแบบนี้นักวิเคราะห์ระบบเรียกว่า หลุมเทา (Gray Hole) เช่น กระบวนการ 1.0 ชื่อ คำนวณเงินเดือน ซึ่งรับข้อมูลเฉพาะ ชื่อ-สกุล ของพนักงาน โดยไม่ได้รับข้อมูลเกี่ยวกับ จำนวนวันทำงาน รายได้ต่อวัน เป็นต้น แต่สามารถส่งข้อมูลเงินเดือนของพนักงานคนนั้น ๆ ได้ ซึ่งเมื่อพิจารณาแล้ว ลักษณะแบบนี้ถือว่าเป็นแบบ หลุมเทา



ภาพที่ 2.77 ลักษณะแผนภาพแผนภาพกระแสข้อมูลแบบข้อมูลที่เคลื่อนที่เข้าไม่เพียงพอ

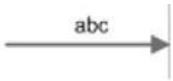
(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

2. การเคลื่อนที่ของข้อมูล

2.1. การเคลื่อนที่ของข้อมูล แสดงเฉพาะการเคลื่อนที่ของข้อมูล จะไม่แสดงการเคลื่อนที่ของสิ่งอื่น ๆ ที่ไม่ใช่ข้อมูล

2.2. หลักการตั้งชื่อการเคลื่อนที่ของข้อมูล มักใช้กลุ่มของคำ หรือคำนามในการตั้งชื่อ โดยทั่วไป มักจะ มีคำว่า ข้อมูล ประกอบอยู่ในกลุ่มของคำนั้น ๆ

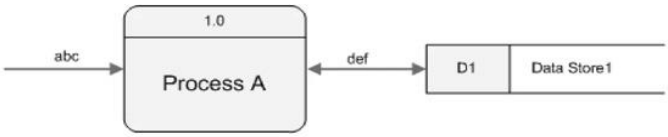
2.3. การเคลื่อนที่ของข้อมูลใช้สัญลักษณ์เป็นเส้นและมีลูกศรฝั่งเดียวแทนทิศทางการเคลื่อนที่ของข้อมูล



ภาพที่ 2.78 การเคลื่อนที่ของข้อมูลโดยใช้สัญลักษณ์เป็นเส้นและมีลูกศรฝั่งเดียว

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

2.4. การเคลื่อนที่ของข้อมูลใช้สัญลักษณ์เป็นเส้นและมีลูกศรสองฝั่ง ข้อมูลระหว่างกระบวนการกับแฟ้มข้อมูลเท่านั้น



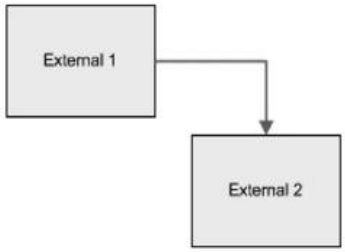
ภาพที่ 2.79 การเคลื่อนที่ของข้อมูลใช้สัญลักษณ์เป็นเส้นและมีลูกศรสองฝั่ง

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

3. หน่วยภายนอก

3.1. หลักการตั้งชื่อหน่วยภายนอก ต้องใช้คำนามเสมอ หน่วยภายนอก สามารถเป็นได้ทั้ง บุคคล กลุ่มคน หน่วยงาน ระบบสารสนเทศเป็นต้น

3.2. นักวิเคราะห์ระบบสนใจเฉพาะข้อมูลเข้าและออกจากหน่วยภายนอกที่มีผลต่อระบบเท่านั้น การสื่อสารระหว่างหน่วยภายนอก นักวิเคราะห์ระบบจะไม่สนใจเนื่องจากอยู่นอกเหนือขอบเขตของการศึกษา



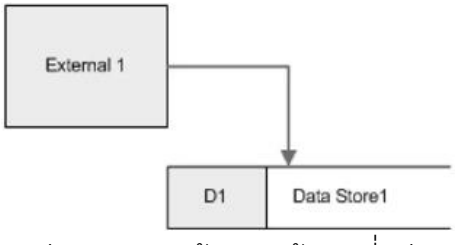
ภาพที่ 2.80 หน่วยภายนอก

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

4. แฟ้มข้อมูล

4.1. หลักการตั้งชื่อแฟ้มข้อมูล ต้องใช้คำนามเสมอ มักขึ้นต้นด้วยคำว่า แฟ้มข้อมูล

4.2. หน่วยภายนอก ต้องไม่มีการติดต่อกับแฟ้มข้อมูลโดยตรง เนื่องจากแฟ้มข้อมูลถือเป็นข้อมูลที่อยู่ภายในระบบ หากหน่วยภายนอกต้องการข้อมูลที่อยู่ภายในแฟ้มข้อมูล ต้องทำการร้องขอผ่านกระบวนการภายในระบบ



ภาพที่ 2.81 หน่วยภายนอกต้องการข้อมูลที่อยู่ภายในแฟ้มข้อมูล

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

4.3. ต้องไม่มีการสื่อสารระหว่างแฟ้มข้อมูลโดยตรง หากต้องการถ่ายโอนข้อมูลจากแฟ้มข้อมูลหนึ่ง ไป ยังอีกแฟ้มข้อมูลหนึ่ง ต้องแทรกกระบวนการอย่างน้อยหนึ่งกระบวนการเพื่อถ่ายโอนข้อมูล

ภาพที่ 2.82 ต้องการถ่ายโอนข้อมูลจากแฟ้มข้อมูลหนึ่งไปยังอีกแฟ้มข้อมูลหนึ่ง

(ที่มา : https://www.slideshare.net/slideshow/ss-47598456/47598456)

ข้อสังเกตการแยกออกเป็นส่วน ๆ หรือการทำให้ภาพปัจจุบันละเอียดขึ้น (Functional Decomposition)ไม่มีกฎเกณฑ์แน่นอนว่า นักวิเคราะห์ระบบควรที่จะหยุดการแขกออกเป็นส่วน ๆ เมื่อใด แต่ให้คำนึงถึงความเหมาะสม อย่างไรก็ตาม ข้อสังเกตต่อไปนี้สามารถใช้ประกอบการตัดสินใจว่าจะหยุดการแยกออกเป็นส่วนๆ

1. ชื่อของข้อมูลที่เคลื่อนที่เข้า และข้อมูลที่เคลื่อนที่ออก เหมือนกัน หรือคล้ายกันมาก

2. เมื่อทำการแยกออกเป็นส่วน ๆ อีกครั้งหนึ่งแล้ว ปรากฏว่า จะเหลือกระบวนการเพียง2กระบวนการ หรือน้อยกว่า

3.เมื่อทำการแยกออกเป็นส่วน ๆ อีกครั้งหนึ่งแล้ว ปรากฏว่า กระบวนการที่ได้เป็นกระบวนการที่เกี่ยวข้องกับการบำรุงรักษาข้อมูลภายในแฟ้มข้อมูล เช่น เพิ่ม ลด แก้ไข ปรับปรุง เป็นต้น

4. เมื่อได้แผนภาพกระแสข้อมูลที่ไม่สามารถแยกออกเป็นส่วน ๆ ได้อีก แผนภาพนั้นถูกเรียกว่า Primitive DFD

2.3.10 E–R โมเดล

2.3.10.1 เอนทิตี้ (Entity)

เอนทิตี้ (Entity) บางครั้งเรียกว่า Entity type หมายถึง อ๊อบเจ็กท์หนึ่งๆ อาจเป็นคน สถานที่ สิ่งของ การกระทำ หรือกิจกรรมต่างๆ ฯลฯ ที่ผู้ใช้ต้องการเก็บข้อมูลไว้ เช่น ผู้ผลิต ลูกค้า แผนก สินค้า พนักงาน การสั่งซื้อเอนทิตี้สามารถใช้แอททริบิวต์ที่มีอยู่ในเอนทิตี้เพื่อกำหนดเป็นคีย์ได้โดยทั่วไปเอนทิตี้จะเป็นเอนทิตี้ปกติ (Regular Entity หรือ Strong Entity) เอนทิตี้อีกประเภทหนึ่งเรียกว่า เอนทิตี้อ่อนแอ (Weak Entity) เอนทิตี้ แบบนี้จะไม่คงอยู่หากไม่มีอีกเอนทิตี้อยู่ กล่าวคือ เอนทิตี้ลักษณะนี้จะไม่มีข้อมูลใน ฐานข้อมูล (Existence Dependency) หากไม่มีอีกเอนทิตี้หนึ่ง (Owner Entity) ลักษณะ ที่สำคัญอีกอย่างหนึ่งของเอนทิตี้อ่อนแอคือ เอนทิตี้นี้ไม่สามารถกำหนดคีย์โดยใช้ แอททริบิวต์ในเอนทิตี้เพียงลำพัง แต่ต้องใช้แอททริบิวต์จากอีกเอนทิตี้หนึ่ง (Owner Entity) ประกอบกันเป็นคีย์ (Partial Key) ตัวอย่างเช่น เอนทิตี้ครอบครัวของพนักงาน (Dependent) จะไม่มีในฐานข้อมูลหากไม่มีเอนทิตี้พนักงาน (Employee) การกำหนด แอททริบิวต์ที่เป็นคีย์หลักประกอบด้วยแอททริบิวต์บางตัวจากเอนทิตี้ครอบครัวของพนักงาน และแอททริบิวต์บางตัวจากเอนทิตี้พนักงาน (Owner Entity) เช่น แอททริบิวต์ ชื่อสมาชิกในครอบครัวประกอบกับชื่อพนักงานรวมกัน

เป็นคีย์ของเอนทิตี้ครอบครัวของพนักงาน ดังนั้น เอนทิตี้นี้เป็นเอนทิตี้อ่อนแอ เพราะขึ้นอยู่กับการมีอยู่ของเอนทิตี้หนึ่งและการกำหนดคีย์หลักเป็นลักษณะ Partial Key

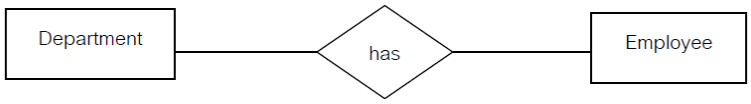
2.3.10.2 คุณลักษณะของเอนทิตี้ (Property)

การกำหนดรายละเอียดข้อมูลของเอนทิตี้ คือการกำหนดแอททริบิวต์ของเอนทิตี้ว่า ประกอบด้วยแอททริบิวต์อะไร รายละเอียดของเอนทิตี้บางครั้งอาจเป็นแอททริบิวต์ผสม (Composite Attribute) ซึ่งสามารถแยกแอททริบิวต์นั้น ๆ เป็นแอททริบิวต์ย่อยได้ เช่น แอททริบิวต์ Address สามารถแยกเป็นแอททริบิวต์ถนนหรือแอททริบิวต์จังหวัด กรณีของแอททริบิวต์ที่ถูกแปลค่ามาจากแอททริบิวต์อื่น (Derived Attribute) เช่น แอททริบิวต์อายุ (Age) เกิดจากการนำแอททริบิวต์วันที่ปัจจุบันลบด้วย แอททริบิวต์วันเกิด (บางครั้งเรียกว่า Stored Attribute) แอททริบิวต์ที่แปลค่ามาอาจจะ แปลค่ามาจากแอททริบิวต์ของเอนทิตี้อื่นที่สัมพันธ์กัน (Related Entity) เช่น แอททริบิวต์ จำนวนพนักงานในแต่ละแผนก (No_of_Emp) ในเอนทิตี้ Department สามารถคำนวณ จากเอนทิตี้พนักงานด้วยการคำนวณจำนวนพนักงานที่ทำงานในแต่ละแผนก นอกจากนี้ แอททริบิวต์บางแอททริบิวต์มีค่าของข้อมูลหลายค่า เช่น แอททริบิวต์ประวัติการศึกษา ซึ่งอาจจะมีตั้งแต่ปริญญาตรี โท และเอก แอททริบิวต์ ลักษณะนี้เรียกว่า แอททริบิวต์ที่มีหลายค่า (Multivalued Attribute)

2.3.10.3 ความสัมพันธ์ระหว่างเอนทิตี้ (Relationship)

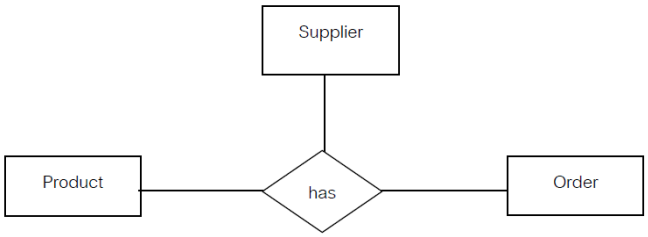
ในการออกแบบฐานข้อมูลด้วยโมเดล จะต้องวิเคราะห์ถึงความสัมพันธ์ระหว่างเอนทิตี้ต่างๆ ของฐานข้อมูล ความสัมพันธ์เหล่านี้สามารถนำไป ออกแบบเค้าร่างของฐานข้อมูลเพื่อกำหนดคีย์นอกที่ใช้อ้างอิงข้อมูลระหว่างเอนทิตี้ที่สัมพันธ์กันลักษณะของความสัมพันธ์ตามจำนวนเอนทิตี้ที่สัมพันธ์กัน แบ่งออกได้เป็นดังนี้

1) ความสัมพันธ์ระหว่างสองเอนทิตี้ (Binary Relationship) ตัวอย่าง เช่น เอนทิตี้พนักงาน (Employee) มีความสัมพันธ์ กับเอนทิตี้แผนก (Department)



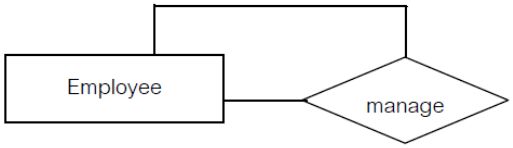
ภาพที่ 2.83 ความสัมพันธ์ระหว่างสองเอนทิตี้ (Binary Relationship)

2) ความสัมพันธ์ระหว่างเอนทิตี้มากกว่าสองเอนทิตี้ (Ternary Relationship)



ภาพที่ 2.84 ความสัมพันธ์ระหว่างเอนทิตี้มากกว่าสองเอนทิตี้ (Ternary Relationship)

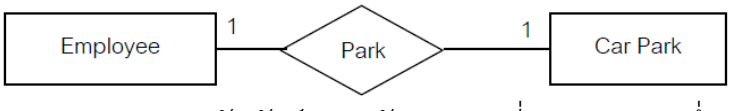
3) ความสัมพันธ์กับเอนทิตี้ตนเอง (Recursive Relationship หรือ Self Relationship หรือ Unary Relationship)



ภาพที่ 2.85 ความสัมพันธ์กับเอนทิตี้ตนเอง (Recursive Relationship หรือ Self Relationship หรือ Unary Relationship)

นอกจากนี้ การวิเคราะห์ความสัมพันธ์ระหว่างเอนทิตี้ จะต้องพิจารณาถึงจำนวนข้อมูลระหว่างเอนทิตี้ที่สัมพันธ์กัน (Cardinality Ratio) แบ่งออกได้ดังนี้ คือ

1) แบบหนึ่งต่อหนึ่ง (One-to-One Relationship) เป็นการแสดงความสัมพันธ์ของข้อมูลเอนทิตี้หนึ่งว่ามีความสัมพันธ์กับข้อมูลอย่างมากหนึ่งข้อมูลกับอีกเอนทิตี้หนึ่งในลักษณะที่เป็นหนึ่งต่อหนึ่ง ตัวอย่างเช่น พนักงานหนึ่งคนมีที่จอดรถเพียงหนึ่งที่และที่จอดหนึ่งที่สามารถรถจอดรถโดยพนักงานหนึ่งคน เท่านั้น ดังภาพที่ 2.86 หรือมีพนักงานเพียงหนึ่งคนที่จะเป็นผู้จัดการแผนก และแผนกหนึ่งผู้จัดการเพียงคนเดียว ความสัมพันธ์ลักษณะนี้เป็นความสัมพันธ์แบบหนึ่งต่อหนึ่ง ดังภาพที่ 2.87



ภาพที่ 2.86 แสดงความสัมพันธ์ของพนักงานและที่จอดรถแบบหนึ่งต่อหนึ่ง (One-to-One Relationship)

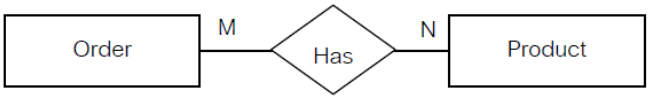
ภาพที่ 2.87

ภาพที่ 2.87 แสดงความสัมพันธ์ของพนักงานและผู้จัดการแผนกแบบหนึ่งต่อหนึ่ง (One-to-One Relationship)

2) แบบหนึ่งต่อกลุ่ม (One-to-Many Relationship) เป็นการแสดงความสัมพันธ์ระหว่างข้อมูลของเอนทิตี้หนึ่งว่ามี ความสัมพันธ์กับข้อมูลหลายข้อมูลของอีกเอนทิตี้หนึ่ง ตัวอย่างเช่น ลูกค้าหนึ่งคน มีคำสั่งซื้อหลายคำสั่งซื้อ ในขณะเดียวกัน คำสั่งซื้อที่ เกิดขึ้นในแต่ละครั้งจะเป็นของลูกค้าเพียงหนึ่งคนเท่านั้น

ภาพที่ 2.88 แสดงความสัมพันธ์แบบหนึ่งต่อกลุ่ม (One-to-Many Relationship)

3) ความสัมพันธ์แบบกลุ่มต่อกลุ่ม (Many-to-Many Relationship) เป็นการแสดงความสัมพันธ์ของข้อมูลระหว่างเอนทิตี้ในลักษณะกลุ่มต่อกลุ่ม ตัวอย่างเช่น การสั่งซื้อหนึ่งครั้งประกอบด้วยสินค้าอย่างน้อยหนึ่งชนิด และสินค้าอย่าง น้อยหนึ่งชนิดถูกสั่งซื้อในแต่ละครั้ง ดังรูป 6.6

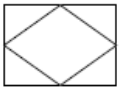
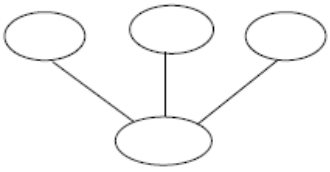
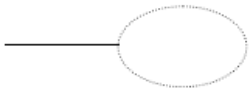


ภาพที่ 2.89 แสดงความสัมพันธ์แบบกลุ่มต่อกลุ่ม (Many-to-Many Relationship)

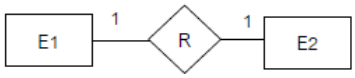
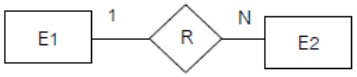
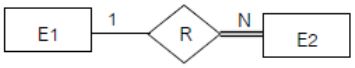
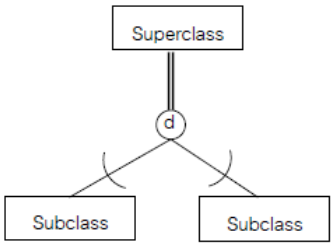
2.3.10.4 สัญลักษณ์ที่ใช้

E-R โมเดลจะมีการใช้สัญลักษณ์เพื่อแทนความหมายของเอนทิตี้ แอททริบิวต์ และความสัมพันธ์ของเอนทิตี้ในฐานข้อมูลที่ออกแบบ ดังนี้

1) สัญลักษณ์ในโมเดล E-R แบบ Chen

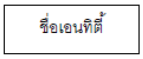
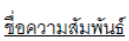
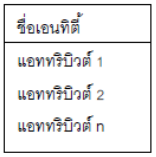
สัญลักษณ์	ความหมาย
	เอนทิตี้
	เอนทิตี้ชนิดอ่อนแอ (Weak Entity)
	ความสัมพันธ์
	ความสัมพันธ์แบบ Existence Dependency
	เป็น Composite Entity หรือ Gerund ที่จะแปลงความสัมพันธ์ของเอนทิตี้แบบ M:N ให้เป็น 1:N
	แอททริบิวต์
	แอททริบิวต์ที่มีหลายค่า (Multivalued Attribute)
	แอททริบิวต์ที่เป็นคีย์หลัก
	แอททริบิวต์ผสม (Composite Attribute)
	แอททริบิวต์ที่แปลค่ามา (Derived Attribute)

ภาพที่ 2.90 สัญลักษณ์ในโมเดล E-R แบบ Chen

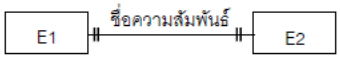
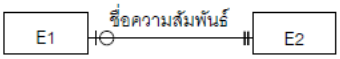
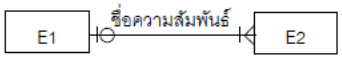
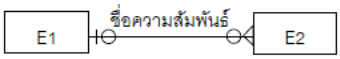
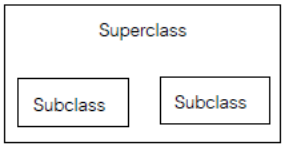
สัญลักษณ์	ความหมาย
	ความสัมพันธ์ของข้อมูลระหว่าง E1 กับ E2 แบบ 1:1 แบบ Partial Participation
	ความสัมพันธ์ของข้อมูลระหว่าง E1 กับ E2 แบบ 1:N แบบ Partial Participation
	ความสัมพันธ์ของเอนทิตี้ E1 และ E2 โดย E1 เป็น Partial Participation และ E2 เป็น Total Participation
	Generalization หรือ Specialization หาก สัญลักษณ์ในวงกลมเป็น d หมายถึง Disjoint (ตามรูปด้านซ้าย) แต่ถ้าสัญลักษณ์ ในวงกลมเป็น o หมายถึง Nondisjoint เส้นตรงคู่ที่ลากจาก Superclass มายัง วงกลมแสดงถึง Total Participation (Mandatory) ดังรูป แต่ถ้าเป็นเส้นเดียว หมายถึง Partial Participation

ภาพที่ 2.90 สัญลักษณ์ในโมเดล E-R แบบ Chen (ต่อ)

2) สัญลักษณ์ในโมเดล E-R แบบ Crow's Foot

สัญลักษณ์	ความหมาย
	เอนทิตี้ (Entity)
	ความสัมพันธ์ (Relationship)
	แอททริบิวต์จะแสดงข้างใต้ชื่อเอนทิตี้ โดยแอททริบิวต์ที่เป็นคีย์หลัก (Primary Key) จะขีดเส้นใต้

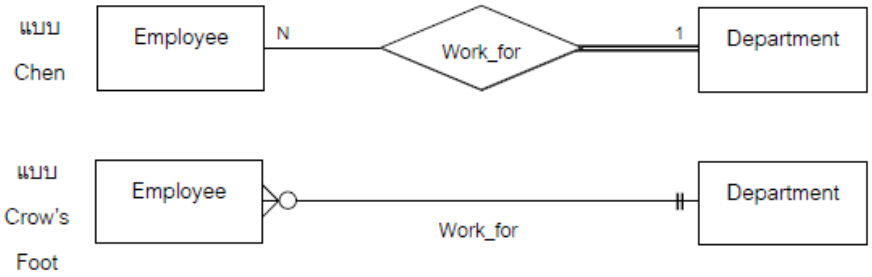
ภาพที่ 2.91 สัญลักษณ์ในโมเดล E-R แบบ Crow's Foot

สัญลักษณ์	ความหมาย
	ความสัมพันธ์แบบ 1:1 แบบ Total Participation (Mandatory) บางตำราจะใช้สัญลักษณ์ I แทนแบบ II
	ความสัมพันธ์แบบ 1:1 โดยที่ E1 เป็น Partial Participation (Optional) และ E2 เป็น Total Participation (Mandatory)
	ความสัมพันธ์แบบ 1:N โดย E1 เป็นแบบ Partial Participation (Optional) ส่วน E2 เป็น Total Participation (Mandatory)
	ความสัมพันธ์แบบ 1:N โดย E1 และ E2 เป็น Partial Participation (Optional)
	สี่เหลี่ยมใหญ่ใช้แสดง Superclass และสี่เหลี่ยมย่อยใช้แสดง Subclass (Generalization / Specialization)

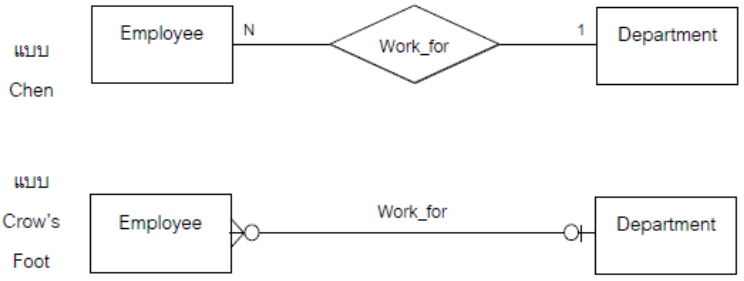
ภาพที่ 2.91 สัญลักษณ์ในโมเดล E-R แบบ Crow's Foot (ต่อ)

ตัวอย่างการเขียนความสัมพันธ์ระหว่างเอนทิตี้และข้อกำหนดที่เกี่ยวกับ คาร์ดินัลลิตี้ (Cardinality) แบบ Chen และแบบ Crow's Foot เป็นดังนี้

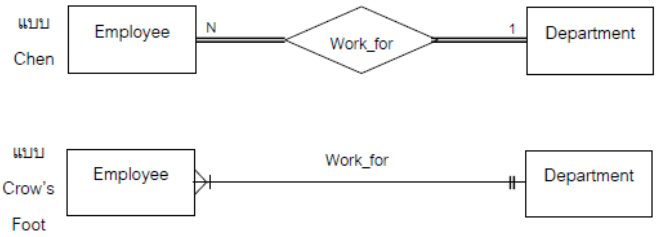
ตัวอย่างที่ 1 แผนกหนึ่งอาจจะมีพนักงานทำงานอยู่หรือไม่ก็ได้ แต่พนักงานทุกคนจะต้องมีแผนกสังกัดอยู่เพียงแผนกเดียวเท่านั้น การโมเดลจะเป็นดังนี้



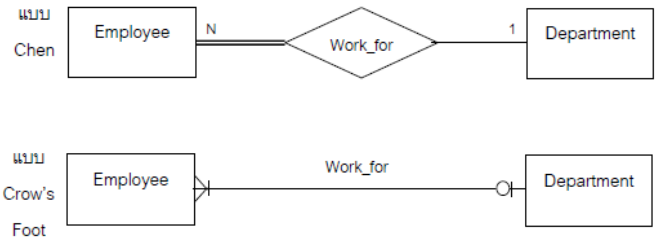
ตัวอย่างที่ 2 แผนกหนึ่งจะมีพนักงานสังกัดอยู่หรือไม่ก็ได้ พนักงานอาจจะยังไม่มีแผนก เนื่องจากอยู่ในช่วงทดลองงาน และหากคนใดมีแผนกสังกัดอยู่จะสังกัดเพียงหนึ่งแผนก



ตัวอย่างที่ 3 แผนกหนึ่งจะต้องมีพนักงานทำงานอยู่ และพนักงานจะต้องมีสังกัดเพียงแผนกเดี่ยว การโมเดลจะเป็นดังนี้



ตัวอย่างที่ 4 แผนกหนึ่งจะต้องมีพนักงานทำงานอยู่ พนักงานอาจจะยังไม่มีแผนกสังกัดอยู่ และถ้าหากพนักงานคนใดมีแผนกสังกัดอยู่ จะสังกัดเพียงแผนกเดี่ยวการโมเดลจะเป็นดังนี้



2.3.10.5 ขั้นตอนในการออกแบบฐานข้อมูลโดยใช้ E-R โมเดล

การออกแบบฐานข้อมูลโดยใช้ E-R โมเดล มีขั้นตอนที่เกี่ยวข้องหลายขั้นตอน และกระบวนการในการออกแบบ E-R โมเดลเป็นกระบวนการที่จะต้องพิจารณาซ้ำๆ หลาย

ครั้งเพื่อให้ได้ E-R โมเดลที่เหมาะสมและมีข้อมูลครบถ้วน ขั้นตอนในการออกแบบ ข้อมูลโดยใช้ E-R โมเดล ประกอบด้วยขั้นตอนดังนี้ คือ

1) ศึกษาถึงลักษณะหน้าที่งานของระบบ (Business Function) ว่า มีรายละเอียดของการทำงานและข้อมูลที่เกี่ยวข้องอะไรบ้าง มีข้อสมมุติฐาน (Business Rule) ของงานต่างๆ อะไรบ้าง

2) กำหนดเอนทิตี้ที่ควรจะมีอยู่ในฐานข้อมูล ฐานข้อมูลหนึ่ง ๆ ประกอบด้วยหลายเอนทิตี้ ในการกำหนดเอนทิตี้ที่ควรจะมีอยู่ในฐานข้อมูลหนึ่ง ๆ จะต้องคำนึงรวมไปถึงว่าเอนทิตี้นั้น ๆ เป็นเอนทิตี้ในประเภทใดบ้าง เช่น เอนทิตี้ประเภทอ่อนแอ (Weak entity)

3) กำหนดประเภทของความสัมพันธ์ระหว่างเอนทิตี้ว่ามีความสัมพันธ์อย่างไรบ้าง รวมถึงข้อกำหนดของความสัมพันธ์ โดยพิจารณาจากข้อสมมุติฐานของความสัมพันธ์ระหว่างเอนทิตี้ที่ได้ศึกษามาในข้อ 1 และข้อ 2

4) กำหนดคุณลักษณะของเอนทิตี้ว่าควรจะมีรายละเอียดอะไรบ้าง ซึ่งการกำหนดคุณลักษณะของเอนทิตี้ จะพิจารณาว่ารายละเอียดต่าง ๆ เป็นรายละเอียดที่มีคุณสมบัติเป็นคีย์ หรือเป็นรายละเอียดที่แปลค่ามา หรือเป็นรายละเอียดที่ประกอบด้วยรายละเอียดที่เป็นข้อมูลหลายอย่าง เช่น ที่อยู่ ประกอบด้วย บ้านเลขที่ ถนน เขต ตำบล จังหวัด รหัสไปรษณีย์ เป็นต้น

5) กำหนดคีย์ของแต่ละเอนทิตี้ว่า จะใช้รายละเอียดของข้อมูลใดเป็นคีย์ของเอนทิตี้นั้น ๆ ซึ่งจะต้องเป็นรายละเอียดของข้อมูลที่มีค่าเป็นเอกลักษณ์ หรือค่าเฉพาะไม่ซ้ำซ้อนในเอนทิตี้นั้น ๆ

6) นำรายละเอียดตั้งแต่ขั้นตอนที่ 2 ถึง 5 มาเขียน E-R โมเดล โดยใช้สัญลักษณ์ที่กล่าวมาข้างต้นเป็นการอธิบายความสัมพันธ์ระหว่างเอนทิตี้ หลังจากนั้นให้ทำการทบทวน E-R โมเดลว่าควรจะปรับเปลี่ยนเค้าร่างใหม่หรือไม่ (Refinement Primitive) โดยนำกล ยุทธ์ในการทบทวนโมเดล (Refinement Strategy) ที่ได้กล่าวในบทที่ 5 มาประยุกต์ใช้ เพื่อให้ได้ E-R โมเดลที่สมบูรณ์

2.3.11 ผังงาน Flowchart

Flow chart (โฟลวชาร์ต) หรือผังงาน คือ แผนภาพที่มีการใช้สัญลักษณ์รูปภาพและลูกศรที่แสดงให้เห็นถึงขั้นตอนการทำงานของระบบทีละขั้นตอนอย่างชัดเจน โดยแต่ละสัญลักษณ์ในแผนภาพ Flowchart นั้นจะหมายถึงการทำงานในหนึ่งขั้นตอน สัญลักษณ์ลูกศรจะแทนลำดับการทำงานในขั้นตอนต่าง ๆ และยังแสดงให้เห็นถึงทิศทางการไหลของข้อมูล ตั้งแต่เริ่มต้นการทำงานจนกระทั่งได้ผลลัพธ์ตามวัตถุประสงค์ที่ได้ตั้งไว้ โดยการที่จะเขียนผังงานได้นั้น น้อง ๆ จะต้องสามารถวิเคราะห์ระบบการทำงานอย่างเป็นลำดับขั้นตอนก่อน จากนั้นจึงนำขั้นตอนเหล่านั้นมาเขียนในรูปแบบของผังงานหรือ Flowchart

2.3.11.1 ผังงาน Flowchart

ผังงาน Flowchart เป็นผังงานที่แสดงให้เห็นถึงแนวคิดในการทำงานที่รอบคอบ มีการวางแผนไว้อย่างชัดเจนและเป็นลำดับขั้นตอน เพราะ Flow Chart เป็นเครื่องมือการจัดเรียงข้อมูลและเครื่องมือการสื่อสารที่ดี ที่นิยมใช้ในงานที่เกี่ยวข้องกับการพัฒนาระบบหรือการออกแบบวิธีการตัดสินใจต่าง ๆ กันอย่างแพร่หลาย แต่การเขียนผังงาน หรือ Flowchart ก็มีข้อจำกัด คือ ไม่เหมาะกับงานที่มีวิธีการซับซ้อน เช่น เป็นการทำงานที่มีเงื่อนไขในการทดสอบมากมาย ซึ่งหากเจอปัญหาในรูปแบบนี้มักจะใช้เครื่องมืออย่างตารางการตัดสินใจ (DECISION TABLE) เข้ามาช่วยมากกว่า

2.3.11.2 รูปแบบของผังงาน Flowchart

โครงสร้างของผังงาน หรือ Flowchart สามารถแบ่งได้เป็น 3 รูปแบบคือ การทำงานแบบตามลำดับ (Sequence Flowchart), การเลือกกระทำตามเงื่อนไข (Selection Flowchart), การทำซ้ำ (Iteration Flowchart) สามารถอธิบายโดยสังเขปได้ดังนี้

1) การทำงานแบบตามลำดับ (Sequence Flowchart) เป็นการเขียน Flowchart โดยจัดลำดับขั้นตอนการทำงานจากบนลงล่าง มีโครงสร้างในการเขียนคำสั่งเป็นบรรทัด และโดยทำงานตามคำสั่งทีละบรรทัดจากบรรทัดแรกที่เป็นการเริ่มต้นคำสั่งลงไปจนถึงบรรทัดล่างสุดเป็นการสิ้นสุดคำสั่ง

2) การทำงานแบบเลือกกระทำตามเงื่อนไข (Selection Flowchart) เป็นการตัดสินใจหรือการเลือกทำตามเงื่อนไข โดยการวิเคราะห์จากข้อมูลจะมีเหตุการณ์ให้ดำเนินการต่อไป 2 กระบวนการ คือ ถ้าหากเงื่อนไขเป็นจริงจะดำเนินการตามกระบวนการหนึ่ง และหากเงื่อนไขเป็นเท็จจะดำเนินการอีกกระบวนการหนึ่ง

3) การทำซ้ำ (Iteration Flowchart) เป็นการเขียน flowchart ให้กลับมาทำงานในขั้นตอนแบบเดิมซ้ำ ๆ ซึ่งจะเห็นว่า flowchart มีขั้นตอนการทำงานบางขั้นตอนได้รับการประมวลผลมากกว่า 1 ครั้ง ซึ่งเรียกว่า loop โครงสร้างผังงานแบบทำซ้ำนี้จะมีการใช้ “สัญลักษณ์การตัดสินใจ”เข้ามาเปรียบเทียบเงื่อนไขเพื่อให้มีการทำงานซ้ำ โดยแบ่งได้ 3 ลักษณะ ดังนี้

1) การทำซ้ำแบบลูป for (for Loop) เป็นการทำซ้ำที่มีการกำหนดจำนวนรอบที่แน่นนอน โดยเริ่มต้นการเปรียบเทียบกับเงื่อนไข

2) การทำซ้ำแบบลูป while (while Loop) เป็นคำสั่งการทำซ้ำที่จะเริ่มด้วยการตรวจสอบเงื่อนไข หากเงื่อนไขเป็น “จริง” จะมีการทำซ้ำต่อไป แต่ถ้าหากเงื่อนไขเป็น “เท็จ” ก็จะเลิกดำเนินการทำตามคำสั่ง

3) การทำซ้ำแบบลูป do...while (do...while Loop) เป็นคำสั่งการทำซ้ำที่จะเริ่มด้วยการทำงาน 1 รอบ แล้วจึงค่อยตรวจสอบเงื่อนไข ถ้าหากเงื่อนไขเป็น “จริง” ก็จะมีการทำซ้ำต่อ ถ้าเงื่อนไขเป็น “เท็จ” ก็จะเลิกดำเนินการทำคำสั่ง

2.3.11.3 หลักการเขียนผังงาน Flowchart

การเขียน Flowchart diagram มีโครงสร้างและวิธีการเขียนที่ค่อนข้างมาตรฐานในตัวอยู่แล้ว เนื่องจากมีการกำหนดสัญลักษณ์ต่าง ๆ ตามมาตรฐานสากลที่ใช้กันทั่วโลก ซึ่งการเขียน Flowchart ให้สามารถทำความเข้าใจได้ง่ายนั้น ไม่เพียงแค่ต้องใช้สัญลักษณ์ถูกต้องเพียงเท่านั้น แต่จะต้องมีหลักการต่าง ๆ ร่วมด้วย โดยหลักการหรือวิธีการเขียนผังงานที่ดี มีดังนี้

- การเขียน Flowchart ควรใช้สัญลักษณ์มาตรฐานตามหลักสากลที่กำหนดไว้โดย The American National Standard Institute, ANSI

- เขียน Flowchart โดยใช้ลูกศรแสดงทิศทางการไหลของข้อมูลตามลำดับจากบนลงล่างหรือจากซ้ายไปขวา

- การเขียน Flowchart จะต้องใช้คำอธิบายในแต่ละขั้นตอนให้มีความกระชับและเข้าใจง่ายเพื่อเขียนลงในสัญลักษณ์ต่าง ๆ

- การเขียนผังงานนั้น ในทุก ๆ สัญลักษณ์ Flowchart จะต้องมีลูกศรแสดงทิศทางเข้า–ออกของข้อมูลเพื่อความชัดเจน

- ในการเขียน Flowchart ไม่ควรโยงเส้นเชื่อมสัญลักษณ์ของผังงานที่อยู่ไกลกัน แต่ควรใช้สัญลักษณ์จุดเชื่อมต่อเพื่อสื่อสารแทน เนื่องจากการโยงเส้นเชื่อมกันของสัญลักษณ์ Flowchart ที่อยู่ไกลกันนั้นอาจทำให้สับสนในขั้นตอนการทำงานได้

- การเขียนผังงานหรือ Flowchart ควรมีการทดสอบความถูกต้องของอัลกอริทึมก่อนที่จะนำไปปฏิบัติงานหรือเขียนโปรแกรม

2.3.11.4 ลักษณะโครงสร้างของผังงาน Flowchart

การเขียนผังงานหรือ Flowchart จะใช้สัญลักษณ์สื่อสารความหมายให้เข้าใจตรงกันในระดับสากลซึ่งเป็นสัญลักษณ์ Flowchart ที่สถาบันมาตรฐานแห่งชาติอเมริกัน (The American National Standard Institute, ANSI) ได้กำหนดสัญลักษณ์เหล่านี้ไว้เป็นมาตรฐานในการเขียนผังงานดังนี้

1) สัญลักษณ์ Flowchart ที่แสดงขั้นตอนการทำงาน

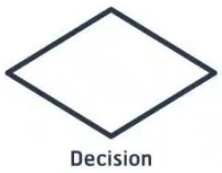
1.Start/Stop สัญลักษณ์กำหนดจุดเริ่มต้นของการทำงาน และแสดงจุดสิ้นสุดของการทำงาน

ภาพที่ 2.92 สัญลักษณ์ Start/Stop

2.Process สัญลักษณ์ Flowchart แสดงรายละเอียดของการทำงาน

ภาพที่ 2.93 สัญลักษณ์ Process

3.Decision สัญลักษณ์แสดงรายละเอียดการเปรียบเทียบในเงื่อนไขการทำงาน ขั้นตอนต่าง ๆ ใช้ใช้เมื่อจะต้องตัดสินใจว่า ใช่หรือไม่ใช จริงหรือเท็จ



ภาพที่ 2.94 สัญลักษณ์ Decision

4.Flowline สัญลักษณ์ที่มีลักษณะคล้ายลูกศรนี้แสดงทิศทางความสัมพันธ์ของการทำงานในระบบงานหรือลำดับงานในการเขียนผังงานโครงสร้าง (Structured Flowchart)

ภาพที่ 2.95 สัญลักษณ์ Flowline

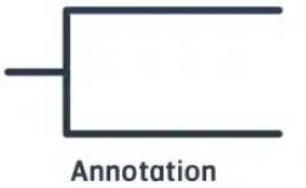
5.IN–Page Connector สัญลักษณ์ของผังงานที่แสดงถึงการกำหนดจุดอ้างอิงในการเชื่อมต่อ ในหน้ากระดาษ เดียวกันของการเขียน Flowchart

ภาพที่ 2.96 สัญลักษณ์ IN–Page Connector

6.Between–Page สัญลักษณ์โฟลวชาร์ตแสดงการกำหนดจุดอ้างอิงในการเชื่อมต่อ ระหว่าง หน้ากระดาษของการเขียนผังงาน

ภาพที่ 2.97 สัญลักษณ์ Between–Page

7.Annotation สัญลักษณ์ Flowchart ที่แสดงการระบุหมายเหตุเพื่อใช้อธิบายรายละเอียดเพิ่มเติมของการดำเนินงาน

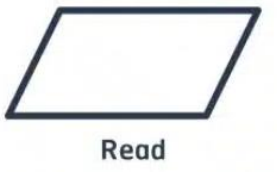


ภาพที่ 2.98 สัญลักษณ์ Annotation

8.SubProgram สัญลักษณ์แสดงคำสั่งการทำงานย่อยที่มีขอบเขตการดำเนินงานที่ชัดเจน โดยสามารถรับค่าข้อมูลที่ถูกส่งเข้าไปเพื่อเป็นข้อมูลประกอบการทำงาน และ/หรือสามารถส่งข้อมูลผลลัพธ์กลับมาเข้าสู่ระบบการดำเนินงานหลักเพื่อรับคำสั่งดำเนินการต่อไปได้

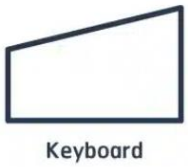
2) สัญลักษณ์ Flowchart ที่ใช้รับค่าข้อมูล

1.Read สัญลักษณ์ Flowchart แสดงถึงการรับค่าข้อมูลหรืออ่านข้อมูลเข้ามาโดยไม่ระบุอุปกรณ์ รับข้อมูล (Input Device)



ภาพที่ 2.99 สัญลักษณ์ Read

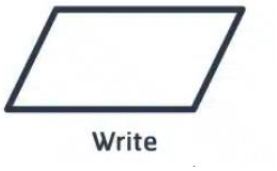
2.Keyboard สัญลักษณ์ Flowchart แสดงถึงการรับค่าข้อมูลหรืออ่านข้อมูลเข้ามาจากคีย์บอร์ด



ภาพที่ 2.100 สัญลักษณ์ Keyboard

3) สัญลักษณ์ Flowchart ที่ใช้แสดงผลของข้อมูล

1.Write สัญลักษณ์ Flowchart ที่ใช้แสดงรายละเอียดข้อมูลหรือผลลัพธ์ของการประมวลผล โดยไม่ระบุอุปกรณ์การแสดงผล (Output Device)

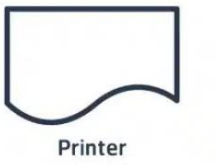


ภาพที่ 2.101 สัญลักษณ์ Write

2.Monitor แสดงรายละเอียดข้อมูลหรือผลลัพธ์ของการประมวลผลไปที่จอภาพ (Monitor)

ภาพที่ 2.102 สัญลักษณ์ Monitor

3.Printer แสดงรายละเอียดข้อมูลหรือผลลัพธ์ของการประมวลผลไปที่เครื่องพิมพ์(Printer)



ภาพที่ 2.103 สัญลักษณ์ Printer

2.3.11.5 ประโยชน์ของผังงาน Flowchart

การเขียน Flowchart เป็นการวางแผนการทำงานอย่างเป็นลำดับขั้นตอนชัดเจน ซึ่งส่งผลให้การทำงานนั้นมีประสิทธิภาพในหลาย ๆ ด้าน โดยประโยชน์ของการเขียน Flowchart มีดังนี้

1. ทำให้มองเห็นภาพรวมของระบบการดำเนินงานและโครงสร้างของคำสั่งโปรแกรมได้ทั้งหมดอย่างเป็นสัดส่วนขั้นตอน และช่วยให้ใช้เวลาในการเข้าใจการทำงานได้เร็วขึ้น

2. การเขียนผังงานมีการใช้สัญลักษณ์ Flowchart ตามมาตรฐานหลักสากล ดังนั้น การวางแผนงานในรูปแบบผังงานจึงสามารถนำไปเขียนอธิบายและสื่อความหมาย ความเข้าใจได้ในทุกภาษา

3. การวางแผนการดำเนินงานโดยเขียน Flowchart เป็นการเขียนขั้นตอนการดำเนินงานอย่างชัดเจน ทำให้ตรวจพบข้อผิดพลาดที่อาจก่อให้เกิดปัญหาได้ง่าย และสามารถแก้ไขจุดผิดพลาดได้อย่างรวดเร็ว

4. หากมีการขยายงานต่อจากผังงานเดิมก็สามารถเพิ่มเติมได้ โดยอาศัยได้ดูผังงานเดิมประกอบ จะทำให้เข้าใจได้รวดเร็วกว่าการดูรายละเอียดจากโปรแกรม หรือระบบงานเดิม

2.4 วรรณกรรมที่เกี่ยวข้อง

วรการ ใจดี และ นรินทร์ จิวิตัน (2566) ปัจจัยที่ส่งผลต่อการตัดสินใจเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและศิลปศาสตร์ โดยใช้เทคนิคการคัดเลือกคุณลักษณะและเทคนิคการทำเหมืองข้อมูล การวิจัยนี้ศึกษาปัจจัยที่มีผลต่อการตัดสินใจเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและศิลปศาสตร์ โดยวิเคราะห์ข้อมูลจากการรับสมัครนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาเชียงใหม่ ช่วงปีการศึกษา 2563–2565 ซึ่งมีทั้งหมด 2,509 รายการ โดยใช้เทคนิคการคัดเลือกคุณลักษณะ 3 เทคนิค และเทคนิคการทำเหมืองข้อมูล 6 เทคนิค ผลการศึกษาแสดงให้เห็นว่าปัจจัยสำคัญที่ส่งผลต่อการตัดสินใจ ได้แก่ สาขาวิชาหรือหลักสูตรที่เลือก สายการเรียนเดิม และวุฒิการศึกษาเดิม ส่วนเทคนิคการทำเหมืองข้อมูลที่มีความแม่นยำสูงสุดคือ เทคนิคการโหวต ด้วยความถูกต้อง 73.44% มากกว่าเทคนิคอื่น ๆ เช่น เทคนิคป่าสุ่มและเทคนิคการถดถอยโลจีสติก.

การใช้เทคนิคเหมืองข้อมูลเป็นการค้นหาองค์ความรู้ (Knowledge) จากข้อมูลที่มีอยู่มาทำการวิเคราะห์ปัจจัยที่ส่งผลต่อการตัดสินใจเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาเชียงใหม่ ซึ่งในการวิจัยครั้งต่อไป คณะผู้วิจัยมีความเห็นว่า 1. ควรเพิ่มปัจจัยอื่น ๆ ที่อาจจะมีผลต่อการเข้าศึกษาต่อในระดับปริญญาตรี เช่น ปัจจัยด้านผู้ปกครอง อาชีพ ผู้ปกครอง รายได้ผู้ปกครอง ปัจจัยด้านข้อมูลส่วนตัวนักศึกษา เช่น งานอดิเรก ความสามารถ อาชีพในอนาคต เป็นต้น 2. ควรมีการใช้เทคนิคการพยากรณ์อื่น ๆ เข้า

มาเปรียบเทียบ เช่น เทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble Model) ประเภทต่าง ๆ เพื่อเพิ่มประสิทธิภาพให้กับตัวแบบพยากรณ์ รวมถึงมีการปรับปรุงพารามิเตอร์ของแต่ละเทคนิคการพยากรณ์ให้ประสิทธิภาพมากขึ้น 3. ควรมีการใช้เทคนิคการปรับสมดุลข้อมูลที่หลากหลาย เพื่อเพิ่มประสิทธิภาพของตัวแบบพยากรณ์ เช่น วิธีการ สุ่มเพิ่ม (Oversampling) วิธีการผสมผสาน (Hybrid Method) เป็นต้น 4. ควรมีการทดสอบเทคนิคการคัดเลือกคุณลักษณะและเทคนิคการพยากรณ์หลายๆ ครั้ง เพื่อให้ได้ประสิทธิภาพมากขึ้นและเพื่อไม่ให้เกิดปัญหา overfiting

วรการ ใจดี และ นพณัฐ วรรณภีร์ (2563) การศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการคัดเลือกคุณลักษณะบนชุดข้อมูลที่ไม่สมดุล งานวิจัยนี้ได้จัดทำขึ้นโดยมีวัตถุประสงค์ 2 ประการคือ (1) เพื่อที่ศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษา ตามแผนของนักศึกษา และ (2) เพื่อศึกษาแนวทางและข้อปรับปรุงในการแก้ไขปัญหาการเรียนการสอนในแต่ละรายวิชาที่ ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษา โดยปัจจัยหรือข้อมูลหลักในการวิจัยนี้ ผู้วิจัยจะใช้ข้อมูลระดับคะแนน ในแต่ละวิชา ข้อมูลเกรดเฉลี่ย ข้อมูลเพศของนักศึกษาชั้นปี ที่ 1–2 โดยผู้วิจัยได้เก็บรวบรวมข้อมูลนักศึกษาจากสาขาวิชา ระบบสารสนเทศทางคอมพิวเตอร์ คณะบริหารธุรกิจและศิลป์ศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาเชียงใหม่ จำนวน 358 คน จากการรวบรวมข้อมูลเบื้องต้นพบว่า ชุดข้อมูลมีความไม่สมดุลกัน (Imbalanced Datasets) โดยมีจำนวนกลุ่มนักศึกษาที่สำเร็จการศึกษาตามแผนมากกว่ากลุ่มนักศึกษาที่ไม่สำเร็จการศึกษาตามแผน ดังนั้นผู้วิจัยจึงได้ทำการแก้ปัญหาโดยใช้วิธีการปรับสมดุลให้กับชุดข้อมูลโดยใช้เทคนิคการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (Synthetic Minority Over–sampling Technique : SMOTE) ก่อนที่จะนำเข้าสู่กระบวนการวิเคราะห์ปัจจัยโดยจะใช้ เทคนิคการคัดเลือกคุณลักษณะ (Feature Selection) 3 เทคนิค คือ (1) Chi–Square Feature Selection (2) Information Gain Feature Selection และ (3) Correlation Based Feature Selection จากการทดลองพบว่าปัจจัยที่มีความสำคัญมากที่สุดและเป็นอันดับหนึ่งในทุกเทคนิคการคัดเลือกคุณลักษณะคือ วิชาการเขียนโปรแกรมคอมพิวเตอร์ 1 ซึ่งวิชานี้เป็นวิชาพื้นฐานและเป็นวิชาบังคับของสาขาระบบสารสนเทศทางคอมพิวเตอร์และเป็นวิชาที่นักศึกษาส่วนใหญ่ได้ระดับคะแนนในวิชานี้ค่อนข้างน้อย ซึ่งอาจส่งผลให้รายวิชานี้ได้ลำดับความสำคัญมาเป็นอันดับหนึ่งในทุกเทคนิคการคัดเลือกคุณลักษณะและอาจเป็นปัจจัยหลักที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษาอย่างมีนัยสำคัญ.

เทคนิคการคัดเลือกคุณลักษณะทั้งสามเทคนิคที่ผู้วิจัยได้นำเสนอ นั้นเป็นเพียงบางเทคนิคของการคัดเลือกคุณลักษณะจากหลากหลายเทคนิคที่มีอยู่ในปัจจุบัน และยังมีข้อจำกัดในเรื่องของปัจจัยข้อมูลที่มีเพียงข้อมูลระดับคะแนนรายวิชา เกรดเฉลี่ยเพศของนักศึกษาเท่านั้น ซึ่งอาจจะมีปัจจัยในด้านอื่น ๆ ที่อาจ ส่งผลต่อการสำเร็จการศึกษาของนักศึกษาได้ เช่น ข้อมูลด้านสถานภาพทางครอบครัว สถานการณ์กู้ยืมเงิน ข้อมูลการศึกษาในอดีต เป็นต้น ในอนาคตหากมีการพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้น อาจจะต้องมีการเก็บรวมรวมข้อมูลเพื่อให้ได้ปัจจัยที่เพิ่มมากขึ้นและใช้เทคนิคในการคัดเลือกคุณลักษณะในหลายๆ เทคนิคมากขึ้นเพื่อสามารถนำผลของการทดลองไปใช้ได้อย่างมีประสิทธิภาพสูงสุด และในงานวิจัยนี้ผู้วิจัยยังไม่ได้มีกำดำเนินการในการนำคุณลักษณะที่คัดเลือกได้ไปหาค่าความแม่นยำในความถูกต้องของแต่ละเทคนิคการคัดเลือกคุณลักษณะ ที่ใช้ ซึ่งผู้วิจัยจะมีการดำเนินการในหัวข้อดังกล่าวในงานวิจัยถัดไป

จิรกิตติ์ จินะกัณฑ์ และ สุรางคณา หมื่นคำแปง (2564) การศึกษานี้วิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนระหว่างปี 2554–2563 โดยใช้กระบวนการ CRISP–DM และเทคนิคการทำเหมืองข้อมูลแบบ Classification กับโมเดล Decision Tree (J48) พบว่าความถูกต้องในการจำแนกประเภท "ดื่ม" และ "ไม่ดื่ม" คือ 67.37% พร้อมทั้งสร้างเว็บไซต์เผยแพร่ข้อมูลในรูปแบบแผนภาพผ่าน Google Data Studio โดยใช้ HTML, CSS, PHP และ JavaScript เว็บไซต์นี้มีประโยชน์ในการศึกษาและสามารถนำข้อมูลไปปรับปรุงเพื่อเพิ่มประสิทธิภาพความปลอดภัยในอนาคต.

การวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ที่จัดทำเสร็จสิ้นแล้ว แม้ว่าจะสามารถจัดทำและแสดงผลได้ตรงตามวัตถุประสงค์และขอบเขตที่กำหนดไว้ แต่ยังมีอีกหลายข้อจำกัดที่ยังสามารถพัฒนาต่อได้ ซึ่งหากจะมีผู้พัฒนาให้เว็บไซต์เผยแพร่ข้อมูลแสดงผลและทำงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้นผู้พัฒนาจะต้องปรับปรุงในส่วนต่าง ๆ 1. ควรมีการพัฒนาหน้าการแสดงผลต่าง ๆ ของเว็บไซต์ให้เหมาะสม สะดวก และเข้าใจง่ายมากยิ่งขึ้น เพื่อลดความผิดพลาดและเพิ่มประสิทธิภาพในการนำเสนอข้อมูล 90 2. ควรมีการพัฒนาในส่วนของเว็บไซต์ให้มีการรองรับบนโทรศัพท์ได้ในอนาคต 3. ควรมีการพัฒนาการปรับเปลี่ยนรูปแบบมุมมองการแสดงกราฟข้อมูลในรูปแบบต่าง ๆ ได้ 4. ควรทำให้ระบบสามารถแยกดูรายงานยอดสรุปของแต่ละจังหวัดได้

มนิศรา เทพฝั้น และศิริขวัญ ขันนาแล (2562) การจัดทำโครงงานนี้มีวัตถุประสงค์เพื่อวิเคราะห์ ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่ระหว่างปี 2551–2558 และเผยแพร่ข้อมูลผ่านเว็บไซต์ โดยใช้กระบวนการ CRISP–DM และเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยโมเดล Decision Tree (J48) ซึ่งมีความถูกต้อง 70.12% และสร้างกฎการจำแนกประเภทได้ 9 กฎ เว็บไซต์ที่สร้างขึ้นมีลักษณะการเผยแพร่ข้อมูลในรูปแบบแผนภาพผ่าน Tableau Public และใช้ HTML, CSS, PHP, และ JavaScript โดยประกอบด้วยข้อมูลเกี่ยวกับอุบัติเหตุ การวิเคราะห์ข้อมูล การสรุปผล รายปี และแนวทางป้องกันอุบัติเหตุ เว็บไซต์นี้สามารถนำไปปรับปรุงเพิ่มเติมเพื่อเพิ่มประสิทธิภาพในการส่งเสริมความปลอดภัยของบุคคลทั่วไปได้ในอนาคต.

การวิเคราะห์ข้อมูลผู้ประสบอุบัติเหตุในช่วงเทศกาลตั้งแต่ปี 2551 – 2558 เพื่อเผยแพร่บนเว็บไซต์ทีจัดทำเสร็จสิ้นแล้วนี้แม้จะสามารถทำงานและแสดงผลข้อมูลได้ถูกตองตรงตามวัตถุประสงค์และขอบเขตทีตั้งไว้ แต่ก็ยังมีข้อจำกัดหลายประการ ซึงหากจะพัฒนาใหเว็บไซต์เผยแพร่ขอมูลนี้แสดงผลขอมูลและทำงานได้อย่างมีประสิทธิภาพมากขึ้น ผู้พัฒนาควรจะตองปรับปรุงในส่วนต่าง ๆ ดังต่อไปนี้ 1. หากมีชุดข้อมูลเพิมเติมควรมีการปรับปรุงข้อมูลใหม่ ๆ อยู่เสมอ 2. ควรจัดเก็บข้อมูลด้วยตัวเองเพือให้ได้ข้อมูลทีตรงต่อความต้องการในการวิเคราะห์ข้อมูลในครั้งต่อไป 3. ควรมีการพัฒนาหน้าการแสดงผลต่าง ๆ ของเว็บไซต์ให้เหมาะสม สะดวก และ เข้าใจง่ายมากยิงขึ้น เพือลดความผิดพลาดและเพิมประสิทธิภาพในการนำเสนอข้อมูล

นรินทร์ จิวิตัน (2565) การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของข้อมูลการเสียชีวิตจากอุบัติเหตุบนโครงข่ายถนน โดยใช้เทคนิคการทำเหมืองข้อมูล งานวิจัยนี้มุ่งศึกษาและวิเคราะห์ปัจจัยที่ส่งผลต่อการเสี่ยงชีวิตจากอุบัติเหตุบนโครงข่ายถนน โดยใช้เทคนิคต่างๆ ในการจำแนกข้อมูลเพื่อทำนายความเสี่ยง มีการใช้เทคนิคหลายประการในการสร้างแบบจำลอง เช่น Gradient Boosting, Decision Trees, Naive Bayes, Random Forest, และ Neural Networks ซึ่งการศึกษาใช้กระบวนการ CRISP–DM และได้รวบรวมข้อมูลจากศูนย์เทคโนโลยีสารสนเทศและการสื่อสารสำนักงานปลัดกระทรวงคมนาคม ระหว่างวันที่ 1 มกราคม 2562 ถึง 30 มิถุนายน 2564 จำนวน 51,384 แถวการวิเคราะห์พบว่าเทคนิค Naive Bayes มีประสิทธิภาพต่ำที่สุดที่ 72.23% ขณะที่เทคนิคอื่นๆ เช่น Random Forest มีความแม่นยำรวม 73.35% และ Neural Networks มีค่าเฉลี่ยความแม่นยำ 72.10% โดยการใช้ K–Fold Cross Validation แบ่งข้อมูลออกเป็น 30 ส่วนเพื่อประเมินประสิทธิภาพของแต่ละเทคนิคผลการวิจัยสามารถใช้เป็นข้อมูล

พื้นฐานในการป้องกันอุบัติเหตุและเตรียมความพร้อมในการจัดการความเสี่ยงบนโครงข่ายถนนได้อย่างมีประสิทธิภาพ.

เพื่อให้ได้ข้อสรุปของผลการวิเคราะห์ข้อมูลที่ครอบคลุมมากขึ้นอาจเปรียบเทียบประสิทธิภาพโดยใช้อัลกอริทึมประเภทอื่นเพิ่มเติม เช่น วิธีเคเนียเรสเนเบอร์ วิธีโครงข่ายประสาทเทียม วิธีซัพพอร์ตเวกเตอร์แมชชีน เป็นต้น มีการเปรียบเทียบกับการใช้เครื่องมือสำเร็จรูปในการช่วยวิเคราะห์อื่น เช่น Microsoft Power BI, Tableau, Pentaho Big Data Integration

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น คณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification โดยคณะผู้จัดทำได้ใช้เทคนิคการทำเหมืองข้อมูลวิเคราะห์ปัจจัยการเกิดอุบัติเหตุ ได้แก่ 1) เทคนิค Chi-Square 2) เทคนิค Correlation Based Feature Selection (CFS) 3) เทคนิค Information Gain และเทคนิคการจำแนกข้อมูล ได้แก่ 1) เทคนิค Decision Tree 2) เทคนิค Random Forest 3) เทคนิค Voting 4) เทคนิค Naïve Bayes 5) เทคนิค Gradient Boosting และ 6) เทคนิค Deep Learning และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรมเผยแพร่บน web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใชภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา