

## บทที่ 2

### แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

การจัดทำโครงการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลความคงอยู่ของลูกค้ากรณีศึกษาบริษัท Telco ผู้ศึกษาได้รวบรวมเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางในการศึกษา ซึ่งข้อมูลเหล่านี้เป็นสารสนเทศที่มีความสำคัญ เป็นอย่างยิ่งที่จะทำให้การพัฒนาโครงการประสบความสำเร็จลุล่วงไปได้ ตรงตามวัตถุประสงค์ที่กำหนดไว้ โดยมีรายละเอียด ดังต่อไปนี้

#### 2.1 แนวคิดที่เกี่ยวข้อง

##### 2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล

2.1.1.1 การวิเคราะห์ข้อมูล (Data Analytics) คือ การวิเคราะห์ข้อมูลที่มีอยู่ในปัจจุบัน อดีต และทำนายอนาคต เพื่อพัฒนาการตลาดให้ตรงกับความต้องการของลูกค้ามากยิ่งขึ้น Data Analytics เป็น เครื่องมือสำหรับธุรกิจ (Business Intelligence) Data Analytics นี้ไม่จำเป็นต้องเป็นธุรกิจขนาดใหญ่ ธุรกิจขนาดเล็กก็สามารถทำได้เหมือนกัน สามารถแบ่งเป็น 4 แบบคือ

1) Descriptive Analytics (การวิเคราะห์เชิงพรรณนา) เป็นรูปแบบการใช้ข้อมูลแบบพื้นฐานที่สุด โดยเน้นการอธิบายว่าเดิ่เกิดขึ้น หรืออาจจะเกิดอะไรขึ้น สามารถอธิบาย ถึงสาเหตุการเกิดต่าง ๆ ซึ่ง Descriptive Analytics ตัวอย่างคือรายงานธุรกิจ รายงานด้านการทำ Campaign หรือโฆษณา หรือรายงานผลดำเนินงานที่ผ่านมา เป็นข้อมูลพื้นฐานที่แสดงผลในการทำกิจกรรมต่าง ๆ

2) Diagnostic Analytics (การวิเคราะห์เชิงวินิจฉัย) เป็นการอธิบายถึงสาเหตุ ของสิ่งที่เกิดขึ้น ปัจจัยต่างๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่างๆ ที่มีความสัมพันธ์ต่อกันของ

สิ่งที่เกิดขึ้น ตัวอย่างเช่นความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่ช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

3) Predictive Analytics (การวิเคราะห์เชิงคาดการณ์) เป็นรูปแบบการใช้ข้อมูล ที่มีความซับซ้อนขึ้นมา โดยจะเป็นการ “พยากรณ์” หรือ “ทำนาย” สิ่งที่จะเกิดขึ้น โดยใช้ ข้อมูลในอดีต ร่วมกับโมเดลทางคณิตศาสตร์ต่าง ๆ หรือรวมกับการทำ Data Mining นอกจากนี้ Predictive Analytics ยังทำให้เราสามารถวิเคราะห์หาโอกาสและความเสี่ยงต่าง ๆ ที่ จะเกิดขึ้นในอนาคตได้ด้วย

4) Prescriptive Analytics (การวิเคราะห์เชิงข้อแนะนำ) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด เป็นทั้งการพยากรณ์สิ่งต่างๆ ที่จะเกิดขึ้น ข้อดี ข้อเสีย สาเหตุ และ ระยะเวลาของสิ่งที่จะเกิดขึ้น และการให้คำ แนะนำ ทางเลือกต่างๆที่มีอยู่ และผลของแต่ละ

### 2.1.2 แนวคิดเกี่ยวกับการถดถอย

การถดถอย (Regression) หมายถึง การศึกษาเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งที่เรียกว่าตัวแปรถูกอธิบาย (Explained Variable) หรือตัวแปรตาม (Dependent Variable) และอีกตัวแปรหนึ่งหรือมากกว่าหนึ่งตัวแปรที่เรียกว่าตัวแปรอิสระ (Independent Variable) หรือตัวแปรอธิบาย (Explanatory Variable) เช่นนักเศรษฐศาสตร์สนใจ ที่จะศึกษาความสัมพันธ์ระหว่างปริมาณความต้องการเสนอซื้อสินค้ากับราคาของสินค้านั้นและระดับรายได้ของผู้บริโภค หรือสนใจศึกษาว่าปริมาณการขายสินค้าจะมีความสัมพันธ์กับ รายจ่ายในการโฆษณาสินค้าอย่างไร เป็นต้น

เป้าหมายของการถดถอย คือ

- 1) เพื่อประมาณค่าเฉลี่ยของตัวแปรตามเมื่อเดหนดค่าของตัวแปรอิสระมาให้
- 2) เพื่อทดสอบสมมติฐานเกี่ยวกับธรรมชาติของตัวแปร โดยมีสมมติฐานภายใต้ทฤษฎีทางเศรษฐศาสตร์
- 3) เพื่อพยากรณ์หรือทำนายค่าเฉลี่ยของตัวแปรตาม เมื่อเดหนดค่าของตัวแปรอิสระมาให้

2.1.2.1 การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis) เป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ ระหว่างตัวแปรอิสระ (Independent Variable) กับตัวแปร

ตาม (Dependent Variable) จะเป็นการศึกษาความสัมพันธ์เชิงเส้นตรง (Linearity) ถ้าศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหนึ่ง ตัวกับตัวแปรตามหนึ่งตัว เรียกว่าการวิเคราะห์ถดถอยเชิงเส้นเชิงเดี่ยวหรือการวิเคราะห์ ถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis) ถ้าตัวแปรอิสระมีมากกว่าหนึ่ง ตัวกับตัวแปรตามหนึ่งตัว เรียกว่า การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Regression Analysis)

#### 2.1.2.2 วัตถุประสงค์ของการวิเคราะห์การถดถอย

- 1) เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม เช่น ศึกษาความสัมพันธ์ของอายุกับระดับคอเลสเตอรอล
- 2) เพื่อศึกษาปัจจัย (ตัวแปรอิสระ) ที่ร่วมกันทำนายหรือพยากรณ์ตัวแปรตาม เช่น การศึกษาปัจจัยที่มีผลต่อระดับน้ำตาลในเลือดของผู้ป่วยเบาหวาน

การวิเคราะห์การถดถอยเพื่อหาความสัมพันธ์หรือสร้างสมการทำนายหรือพยากรณ์ตัวแปรตาม (Y) หนึ่งตัว จากกลุ่มตัวแปรอิสระ (X) หลายตัวนั้น ตัวแปร อิสระที่นำมาวิเคราะห์จะต้องมีหลักฐานตามทฤษฎีหรือรายงานการวิจัยที่เกี่ยวข้องว่าเป็นตัวแปรต้นเหตุที่ส่งผลต่อตัวแปรตาม

2.1.2.3 การวิเคราะห์การถดถอยเชิงเส้นเชิงเดี่ยว (Simple Linear Regression) การวิเคราะห์การถดถอยเชิงเส้นเชิงเดี่ยวเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ 1 ตัว กับ ตัวแปรตาม 1 ตัว คล้ายกับการวิเคราะห์สหสัมพันธ์ แตกต่างกันที่การวิเคราะห์สหสัมพันธ์ไม่ได้ระบุว่าตัวแปรใดเป็นตัวแปรต้น ตัวแปรใดเป็นตัวแปรตามส่วนการวิเคราะห์การถดถอยเชิงเส้นเชิงเดี่ยวเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรที่เป็น ต้นเหตุ (ตัวแปรอิสระ) กับตัวแปรที่เป็นผล (ตัวแปรตาม) ซึ่งนอกจากจะทราบความสัมพันธ์ ระหว่างตัวแปรทั้งสองแล้ว ยังสามารถนำค่าของตัวแปรต้นเหตุไปทำนายหรือพยากรณ์ตัวแปร ที่เป็นผลได้ พร้อมบอกขนาดของความสามารถในการทำนายหรือความสามารถในการอธิบาย ตัวแปรผล ว่ามีมากน้อยเพียงใด

#### 2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล

การแสดงผลข้อมูล (Data Visualization) เป็นการนำ Data เชิงลึกที่ได้มาจากช่องทางต่าง ๆ มา วิเคราะห์และแสดงผลในรูปแบบของแผนภูมิ กราฟรูปแบบที่หลากหลาย วิธีโอ

ที่แสดงให้เห็นถึง การเปลี่ยนแปลงของข้อมูลเชิงปริมาณ ซึ่งจะทำให้ผู้อ่านเข้าใจได้ง่ายขึ้น ดู น่าสนใจมากกว่า การอ่านข้อมูลแบบตารางทั่วไป

1) แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้งที่นิยม แสดงออกมาในรูปแบบสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับการเปรียบเทียบ จำนวนของข้อมูลในแต่ละชุดเช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริงเปรียบเทียบกับ เป้าหมายยอดขายที่ตั้งไว้ เป็นต้นซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท แผนภูมิแท่ง แบบจัดกลุ่ม แผนภูมิแท่งแบบจัดกลุ่ม

2) แผนภูมिवงกลม (Pie Charts) แผนภูมिवงกลมเหมาะกับการนำเสนอข้อมูล ที่มี ส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทาง กลับกัน อาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่ง แยกยาก เพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูล แสดงส่วนผสมต่าง ๆ เป็นต้น

3) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่ง ประกอบด้วยแกนนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั้นเองแผนภูมิ ประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดง ให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์ แนวโน้ม ในอนาคตได้เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาสและนำมาวิเคราะห์เพื่อดูแนวโน้ม เป็นต้น

4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบ เช่นเดียวกับ แผนภูมिवงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกัน หลายๆ ชั้น

5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้ เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้นเหมาะสำหรับเน้น ความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่าง ระหว่างข้อมูล

เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่ละเดือนตามหมวดหมู่ต่าง ๆ ไล่ไป เครื่องสำอางค์ เสื้อผ้าแฟชั่น อาหาร ตามลำดับ

6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผล แบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึง ความต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หา จุดอ่อน จุดแข็งของข้อมูลเช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อ จุดอ่อน จุดแข็งของแต่ละคน เป็นต้น

7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดงผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอน และ แกนตั้ง แต่เลือกใช้รูปภาพ หรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ เช่น การแสดงผลจำนวน ของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละ เดือนซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง แต่เปลี่ยนจากแท่งเป็นรูปภาพิกของนม 2 รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะ นี้ต้องอาศัยความ คำนึงของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคน อาจจะใช้แทนด้วยภาพ ไอคอนคนหรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงิน หรือแบงค์แทน ก็จะทำให้คนดู เข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคยอยู่แล้วและยังดึงดูดความสนใจได้มากกว่าสนใจได้มากกว่า การใช้กราฟแท่งสีเหลี่ยมอีกด้วย

## 2.2 ทฤษฎีที่เกี่ยวข้อง

### 2.2.1 ทฤษฎีเกี่ยวกับฐานข้อมูล

โอบาส เอียมลิวรี่วศ์ (2559) ได้ให้ความหมายเกี่ยวกับฐานข้อมูล (Database) หมายถึง กลุ่มของข้อมูลที่มีความสัมพันธ์กันและถูกนำมารวมกันโดยมีโครงสร้างเดียวกัน ถูกควบคุมดูแล และจัดการโดยซอฟต์แวร์จัดการฐานข้อมูล (DBMS) เพื่อตอบสนองความต้องการสารสนเทศของ

องค์กรและเพื่อการใช้งานร่วมกันของผู้ใช้ เรียกองค์ประกอบทั้งหมดที่ ทำงานร่วมกันเหล่านี้ว่า “ระบบฐานข้อมูล (Database System)”

ในปัจจุบัน คำว่า “ฐานข้อมูล (Database)” ได้เข้ามามีบทบาทสำคัญต่อวงการ ทั่วไปเป็นอย่างมาก ดังจะเห็นได้ว่า แทบทุกที่ที่มีคอมพิวเตอร์ใช้งานเพื่อการประมวลผล ไม่ว่าจะเป็น การประยุกต์ใช้คอมพิวเตอร์สำหรับงานด้านธุรกิจ การศึกษา การแพทย์ วิทยาศาสตร์หรืองานด้านวิศวกรรม ก็ล้วนแต่เกี่ยวข้องกับฐานข้อมูลแทบทั้งสิ้น ดังนั้นอาจกล่าวได้ว่าอัตรา การเติบโตของการใช้คอมพิวเตอร์ ได้ส่งผลกระทบโดยตรงต่อเทคโนโลยีฐานข้อมูล จึงส่งผลให้ องค์กรค่อนข้างจำเป็นต้องใช้คอมพิวเตอร์เพื่อการเข้าถึงฐานข้อมูลที่ตนต้องการ ซึ่งเชื่อว่า จะ เป็นเพียงฐานข้อมูลใน ระดับภายในองค์กรเท่านั้น แต่ยังสามารถพัฒนาเป็นฐานข้อมูลทั้งภายใน และภายนอกมาอยู่รวมกันเป็น ฐานข้อมูลขนาดใหญ่ที่เรียกว่าคลังข้อมูล (Data Warehouse) สำหรับข้อมูลในยุคปัจจุบันอาจจะ เป็นชนิดข้อมูลแบบมีโครงสร้าง (Structure Data Types) หรือไม่มีโครงสร้าง (Unstructured Data Types) ซึ่งอาจจะมีการนำมาใช้ในฐานข้อมูลเดียวกัน

#### 2.2.1.1 ระบบจัดการฐานข้อมูล (Database Management Systems)

คือซอฟต์แวร์ที่ใช้เป็นเครื่องมือของผู้ใช้เพื่อโต้ตอบกับฐานข้อมูล ซึ่ง จะประกอบไปด้วยฟังก์ชันหน้าที่ต่าง ๆ ในการจัดการกับข้อมูล รวมทั้งภาษาที่ใช้ทำงานกับ ข้อมูล ซึ่งโดยมักใช้ ภาษาเอสคิวแอล (SQL) ในการโต้ตอบระหว่างกันกับผู้ใช้ด้วยการสร้าง การเรียกดู และการบำรุงรักษาฐานข้อมูล นอกจากนี้ ยังมีหน้าที่ในการรักษาความมั่นคง และ ความปลอดภัยของ ข้อมูล ด้วยการป้องกันมิให้ผู้ไม่มีสิทธิ์การใช้งานเข้ามาละเมิดข้อมูล ในฐานข้อมูลที่เป็นศูนย์กลาง ได้ รวมถึงการสำรองข้อมูล และการกู้คืนข้อมูล ในกรณีข้อมูล เกิดความเสียหาย เป็นต้น จึงกล่าว โดยสรุปว่า ระบบจัดการฐานข้อมูล เป็นซอฟต์แวร์ หรือ โปรแกรมที่ใช้สำหรับโต้ตอบกับผู้ใช้งาน โดยระบบจัดการฐานข้อมูล จะเป็นตัวกลางในการ โต้ตอบระหว่างผู้ใช้กับฐานข้อมูล ซึ่งผู้ใช้ สามารถโต้ตอบกับฐานข้อมูลผ่านระบบจัดการ ฐานข้อมูลโดยตรง หรือผ่านโปรแกรมประยุกต์ก็ได้ และเนื่องจากระบบฐานข้อมูลนั้น คือซอฟต์แวร์ที่อนุญาตให้ผู้ใช้ทำการสร้าง เรียกดู และ บำรุงรักษาฐานข้อมูล รวมถึงการ จัดการควบคุมการเข้าถึงข้อมูล โดยความหมายของ คำว่า ฐานข้อมูลที่หลายคนส่วนใหญ่เข้าใจ คือ เป็นการรวมกลุ่มของข้อมูลที่มีความสัมพันธ์กัน และโดย

ปกติฐานข้อมูลก็มักจะใช้เพื่อ อ้างอิงถึงข้อมูลในตัวเอง แต่ความเป็นจริงแล้วยังมีส่วนประกอบด้านสภาพแวดล้อมอื่น ๆ เพิ่มเติมอีกหลายส่วนด้วยกัน ครั้นเมื่อนำมาประกอบรวมกัน ก็จะทำให้ระบบจัดการฐานข้อมูล มีความสมบูรณ์ยิ่งขึ้น ส่วนประกอบด้านสภาพแวดล้อมของระบบจัดการฐานข้อมูล ประกอบด้วย 5 ส่วนด้วยกัน คือ

1) ฮาร์ดแวร์ (Hardware) หมายถึง ตัวเครื่องคอมพิวเตอร์และ อุปกรณ์รอบข้าง (Peripherals) ตัวเครื่องคอมพิวเตอร์ สิ่งที่จะนำมาพิจารณาก็คือหน่วย ประมวลผลกลาง (CPU) และหน่วยความจำหลัก หน่วยประมวลผลกลางจะเป็นตัวบ่งชี้ถึง ความเร็วในการประมวลผล ในขณะที่ขนาดของหน่วยความจำหลักในเครื่องคอมพิวเตอร์ที่ใช้ เป็นพื้นที่สำหรับจัดเก็บข้อมูลและโปรแกรมต่าง ๆ ที่นำมาประมวลผลร่วมกับฐานข้อมูล

2) ซอฟต์แวร์ (Software) จะประกอบไปด้วยโปรแกรมต่าง ๆ เช่น ระบบปฏิบัติการ โปรแกรมจัดการฐานข้อมูล และโปรแกรมประยุกต์กับโปรแกรมยูทิลิตี้ต่าง ๆ เป็นต้น

3) ข้อมูล (Data) ที่เป็นส่วนสำคัญของฐานข้อมูล โดย เปรียบเสมือนกับสะพานที่เชื่อมโยงระหว่างส่วนประกอบของเครื่องจักร และมนุษย์เข้าด้วยกัน สำหรับข้อมูลที่บันทึกอยู่ในฐานข้อมูลนั้น จะได้รับการออกแบบเพื่อการจัดเก็บจากนักออกแบบ ฐานข้อมูลอย่างมีระเบียบ

4) โพรซีเจอร์ (Procedure) ที่เกี่ยวข้องกับชุดคำสั่งและกฎระเบียบ เพื่อใช้สำหรับการออกแบบและใช้งานฐานข้อมูล โดยสามารถจัดทำขึ้นเพื่อเป็นเอกสารหรือ คู่มือการใช้งานว่าจะปฏิบัติตนอย่างไรเพื่อให้สามารถใช้งานหรือให้ระบบทำงานได้ ซึ่งอาจ ประกอบไปด้วยชุดคำสั่งต่าง ๆ ว่าจะต้องทำอย่างไร

5) ผู้ใช้งาน (Users) จะประกอบไปด้วยกลุ่มบุคคลต่าง ๆ ซึ่ง สามารถจำแนกออกเป็นตำแหน่ง แต่ละตำแหน่งจะมีภาระหน้าที่รับผิดชอบที่แตกต่างกัน

### 2.2.1.2 แบบจำลองข้อมูล

เป็นแหล่งรวมของแนวคิดที่น่าเสนอความเป็นจริงของวัตถุ ข้อมูล และเหตุการณ์ รวมถึงความสัมพันธ์ระหว่างข้อมูลที่มีความสอดคล้องตรงกัน โดยจุดประสงค์ ของแบบจำลองข้อมูล คือการนำแนวความคิดต่าง ๆ มานำเสนอให้เกิดเป็นรูปแบบจำลอง ขึ้นมา เพื่อใช้ในการ

สื่อสารระหว่างผู้ออกแบบฐานข้อมูลกับผู้ใช้ให้เกิดความเข้าใจตรงกัน สำหรับพื้นฐานการสร้างแบบจำลองข้อมูลโดยทั่วไปแล้วจะประกอบด้วยเอนทิตี แอตทริบิวต์ และความสัมพันธ์

1) เอนทิตี (Entities) คือ บุคคล สถานที่ สิ่งของ หรือเหตุการณ์ ที่เกี่ยวข้องกับข้อมูลที่เรารวบรวมไว้เพื่อการจัดเก็บ เอนทิตีถือเป็นตัวแทนของวัตถุในโลกแห่งความเป็นจริง หรืออาจเป็นนามธรรม (Abstractions) ก็ได้

2) แอตทริบิวต์ (Attribute) คือ คุณลักษณะของเอนทิตี

3) ความสัมพันธ์ (Relationship) คือ ความสัมพันธ์ระหว่างเอนทิตี โดยในแบบจำลองจะมีความสัมพันธ์อยู่ 3 ชนิด คือ ความสัมพันธ์แบบหนึ่งต่อหนึ่ง (one-to-one) ความสัมพันธ์แบบหนึ่งต่อกลุ่ม (one-to-many) และความสัมพันธ์แบบกลุ่มต่อกลุ่ม (many-to-many)

4) ข้อบังคับ (Constraints) คือ กฎเกณฑ์เพื่อการบรรจุข้อมูล ซึ่งเป็นสิ่งที่สำคัญ เพราะจะช่วยให้เกิดความมั่นใจเป็นอันหนึ่งอันเดียว เกิดความสอดคล้องของข้อมูล

การออกแบบฐานข้อมูล (Designing Databases) มีความสำคัญต่อการจัดการระบบ ฐานข้อมูล (DBMS) ทั้งนี้เนื่องจากข้อมูลที่อยู่ภายในฐานข้อมูลจะต้องศึกษาถึงความสัมพันธ์ของข้อมูล โครงสร้างของข้อมูล การเข้าถึงข้อมูล และกระบวนการที่โปรแกรมประยุกต์จะเรียกใช้ฐานข้อมูล ดังนั้น สามารถแบ่งวิธีการสร้างฐานข้อมูลได้ 3 ประเภท ดังนี้

1) รูปแบบข้อมูลแบบลำดับชั้น หรือโครงสร้างแบบลำดับชั้น (Hierarchical data model) วิธีการ จำกัดในปี 1980 ได้รับความนิยมมากในการพัฒนาฐานข้อมูลบนเครื่องคอมพิวเตอร์ขนาดใหญ่ และขนาดกลาง โดยที่โครงสร้างข้อมูลจะสร้างรูปแบบเหมือนต้นไม้ โดยความสัมพันธ์เป็นแบบหนึ่งต่อหลาย (One- to Many)

2) รูปแบบข้อมูลแบบเครือข่าย (Network data Model) ฐานข้อมูล แบบเครือข่ายมีความคล้ายคลึงกับฐาน ข้อมูลแบบลำดับชั้น ต่างกันที่โครงสร้างแบบเครือข่าย อาจจะมีการติดต่อหลายต่อหนึ่ง (Many-to-one) หรือ หลายต่อหลาย (Many-to-many) กล่าวคือลูก (Child) อาจมีพ่อแม่ (Parent) มากกว่าหนึ่ง สำหรับตัวอย่างฐานข้อมูลแบบ เครือข่ายให้ลองพิจารณาการจัดการข้อมูลของห้องสมุด ซึ่งรายการจะประกอบด้วย ชื่อเรื่อง ผู้แต่ง สำนักพิมพ์ ที่อยู่ ประเภท



3) รูปแบบความสัมพันธ์ข้อมูล (Relation data model) เป็นลักษณะ การออกแบบฐานข้อมูลโดยจัดข้อมูลให้อยู่ในรูปของตารางที่มีระบบคล้ายแฟ้ม โดยที่ข้อมูลแต่ละแถว (Row) ของตารางจะแทนเรคคอร์ด (Record) ส่วนข้อมูลแนวตั้งจะแทนคอลัมน์ (Column) ซึ่งเป็นขอบเขตของข้อมูล (Field) โดยที่ตารางแต่ละตารางที่สร้างขึ้นจะเป็นอิสระ ดังนั้น ผู้ออกแบบฐานข้อมูลจะต้องมีการวางแผนถึงตารางข้อมูลที่เป็นต้นตอใช้ เช่น ระบบฐานข้อมูลบริษัทแห่งหนึ่งประกอบด้วย ตารางประวัติพนักงาน ตารางแผนก และตารางข้อมูลโครงการ แสดงประวัติพนักงาน ตารางแผนก และตารางข้อมูลโครงการ

### 2.2.1 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

2.2.2.1 การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถ ค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยก ๆ เป็นข้อ ๆ ได้ดังนี้

- 1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและ เก็บ
- 2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัยฐานข้อมูล
- 3) การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือ
- 4) การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

## 2.2.2 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.2.2.1 หลักในการออกแบบเว็บไซต์หน้าเว็บเป็นสิ่งแรกๆที่ผู้ใช้จะมองเห็นขณะที่เปิดเข้าสู่เว็บไซต์ และยังเป็นสิ่งแรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วยหน้าเว็บ จึงเป็นสิ่งสำคัญมาก เพราะเป็นสื่อกลางให้ผู้ชมสามารถใช้ประโยชน์จากข้อมูลของระบบงาน ของเว็บไซต์นั้นได้ โดยปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้นระบบเนวิเกชัน และองค์ประกอบอื่น ๆ ที่ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการใช้งานหลักสำคัญในการออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อความหมาย เกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้งาน การออกแบบเว็บไซต์ ต้องคำนึงถึง

1) ความเรียบง่าย การออกแบบเว็บไซต์ที่ดี จะต้องเน้นที่ความเรียบง่ายเป็นหลัก โดยเลือกนำเสนอเฉพาะสิ่งที่ต้องการนำเสนอจริงๆ ในรูปแบบที่หลากหลาย โดยอาจจะ เป็นสีสังกะสี ภาพเคลื่อนไหวหรือตัวอักษร ที่สำคัญจะต้องมีการนำเสนอที่ไม่ดูรกหน้าเว็บจนเกินไป

2) ความสม่ำเสมอ คือจะต้องมีรูปแบบ กราฟิก โทนมสีและการตกแต่งต่างๆ ให้แต่ละหน้าบนเว็บไซต์มีความคล้ายคลึงกัน และเป็นแนวเดียวกันไปตลอดทั้งเว็บไซต์ เน้นการ ตกแต่งในรูปแบบเดียวกันทั้งหมด ต่างแค่การนำเสนอของแต่ละหน้าเท่านั้น

3) สร้างความโดดเด่น ต้องมีการสร้างความเป็นเอกลักษณ์และจุดเด่นให้กับ เว็บไซต์ เพื่อให้สามารถสะท้อนถึงลักษณะขององค์กรได้มากที่สุด โดยการสร้างเอกลักษณ์ ดังกล่าวนั้น อาจใช้ชุดสี รูปภาพ ตัวอักษรหรือกราฟิก นอกจากนี้ก็ต้องขึ้นอยู่กับว่า เป็นเว็บไซต์ แบบทางการหรือไม่ เพื่อจะได้ออกแบบได้อย่างเหมาะสมที่สุด

4) เนื้อหาครบถ้วน เนื้อหาเป็นสิ่งสำคัญที่สุดของการสร้างเว็บไซต์ เพราะ สิ่งที่ทำให้ผู้คนเกิดความสนใจ และหมั่นติดตามเว็บไซต์เหล่านั้นอยู่เสมอ ก็คือเนื้อหาที่มีความ สมบูรณ์และน่าสนใจ นอกจากนี้จะต้องมีการปรับปรุง พัฒนาเนื้อหาบนเว็บให้มีความทันสมัย อยู่เสมอ

5) ระบบเนวิเกชันที่ใช้งานง่าย ระบบเนวิเกชัน เป็นเสมือนป้ายบอกทางเพื่อให้ ผู้ใช้งาน ไม่เกิดความสับสนในขณะที่ใช้งานเว็บไซต์ ซึ่งการออกแบบเนวิเกชันก็จะต้องเน้นที่ความ เรียบง่าย ใช้งาน

สะดวก และมีความเข้าใจได้ง่าย ที่สำคัญจะต้องมีตำแหน่งการวางที่สม่ำเสมอ เพื่อให้ดูเป็นแนวทางเดียวกัน

6) คุณภาพของเว็บไซต์ เว็บไซต์ที่ดีจะต้องมีคุณภาพ ทั้งสิ่งที่ปรากฏให้เห็น บนเว็บไซต์ ไม่ว่าจะเป็นกราฟิก ชนิดตัวอักษร รูปภาพหรือสีเส้นที่ใช้ เนื้อหาที่นำมาแสดงผล ซึ่ง หากเว็บไซต์มีคุณภาพก็จะสร้างความน่าเชื่อถือ และเป็นจุดเด่นที่ทำให้ผู้คนส่วนใหญ่เกิดความสนใจได้ดี

7) ความสะดวกในงานเข้าใช้งาน เว็บไซต์ควรให้ความสะดวกสบายแก่ ผู้ใช้งานได้ดี คือ จะต้องมีการแสดงผลได้ในทุกระบบปฏิบัติการ ไม่ว่าจะเป็นเว็บเบราว์เซอร์ คอมพิวเตอร์ โน้ตบุ๊ก หรือบนโทรศัพท์มือถือ ที่สำคัญจะต้องมีความละเอียดของการแสดงผล และสามารถใช้งานได้โดยไม่มีปัญหาด้วย

8) ความคงที่ของการออกแบบ การออกแบบเว็บไซต์ควรมีความคงที่ใน การออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกัน และมีการเรียบเรียงเนื้อหาอย่าง รอบคอบ ทำให้เว็บมีความน่าเชื่อถือ และดูมีคุณภาพ ช่วยสร้างความประทับใจให้กับผู้ใช้งาน 9) ความคงที่ของการทำงาน ระบบการทำงานบนเว็บไซต์จะต้องมีความคงที่ และสามารถใช้งานได้ดี ซึ่งนอกจากการออกแบบระบบการทำงานให้มีความทันสมัยและ สร้างสรรค์แล้ว ก็จะต้องหมั่นตรวจสอบอยู่เสมอ เพราะหากระบบการใช้งานมีความผิดปกติ ก็จะได้แก้ปัญหาได้ทัน นอกจากนี้ถ้าจะมีการอัปเดต ดีไซน์ให้ทันสมัยขึ้นบ่อยๆ

2.2.2.2 รูปแบบโครงสร้างของเว็บไซต์ การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบ ซึ่งก็ ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างแบบเรียงลำดับ เป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมาก ที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้ เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิงค์เนื้อหา ก็จะลิงค์ไปที่หน้า มี ทิศทางการเข้า

สู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ง่ายดูหน้า-ถอยหลังในการdeหนดทิศทาง จึงทำให้การใช้งาน เป็นไปอย่างง่าย

2) โครงสร้างแบบลำดับขั้น นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้ สามารถ เข้าถึงข้อมูลต่างๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆ และมีการนำเสนอ รายละเอียด ย่อย ๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความเข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีไฮมเพจ เป็นจุดเริ่มต้น และจุดรวมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง

3) โครงสร้างแบบตาราง เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีความ ยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่างๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้ จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งาน สามารถเปลี่ยนทิศทาง หรือde หนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลาและยังทำให้เว็บไซต์มีความ ทันสมัย

4) โครงสร้างแบบในแมงมุม เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความ ยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถ เข้าถึงหน้าเว็บเพจ ต่างๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถ เชื่อมโยงไปสู่เว็บไซต์ ภายนอกได้ดี

## 2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

### 2.3.1 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

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

1) การทำความเข้าใจในธุรกิจ (Business understanding) เป็นการdeหนดขอบเขต ของ ข้อมูลที่จะนำวิเคราะห์เพื่อหาความได้เปรียบทางการตลาด เพื่อนำมาแก้ไขปัญหาคอร์

และต้องสามารถระบุผลลัพธ์ที่มีได้

2) การทำความเข้าใจในข้อมูล (Data Understanding) โดยการรวบรวมข้อมูลที่ เกี่ยวข้องให้ อยู่รวมกลุ่มกัน คัดเลือกให้เหลือเพียงข้อมูลที่มีความสำคัญต่อการทำงาน

3) การเตรียมข้อมูล (Data Preparation) การจัดเตรียมข้อมูลประกอบด้วย การ คัดเลือก ข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4) การสร้างแบบจำลอง (Modeling) นำผลที่ได้มาทดลองทำแบบจำลอง เพื่อใช้ ถ่ายทอด ข้อมูลหรือทำการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปสารสนเทศเป็นข้อมูลที่น่าไปใช้

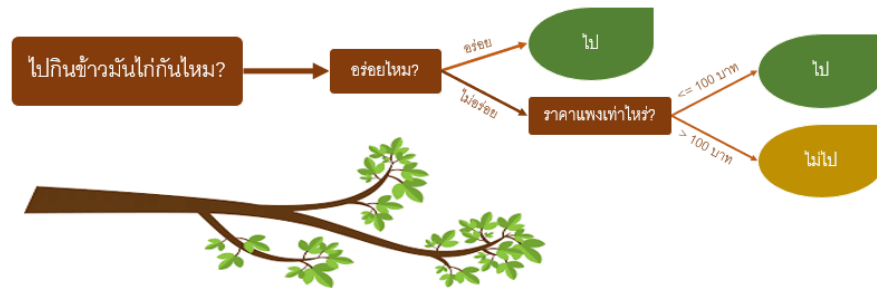
ประโยชน์ได้จริงในทางธุรกิจ

5) การประเมินผล (Evaluation) การวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับ วัตถุประสงค์ที่ตั้งไว้ในตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใดเพื่อเปลี่ยนแปลงแก้ไข ให้ ได้ตามผลลัพธ์ที่ต้องการ

6) เผยแพร่ผลวิเคราะห์ (Deployment) การนำเอาข้อมูลที่เป็นผลลัพธ์มาลอง ปฏิบัติจริง กับธุรกิจ โดยแปลงแนวคิดที่มีให้เกิดเป็นการนำไปใช้ประโยชน์ และติดตามรวบรวม ผลที่ได้เพื่อ การปรับปรุง

### 2.3.2 Decision Tree

Decision Tree หรือ ต้นไม้ตัดสินใจ คือ การจำลองวิธีการตัดสินใจของมนุษย์ ซึ่งการ ตัดสินใจแต่ละครั้งของมนุษย์ มนุษย์จะแตกโจทย์หลักออกเป็นโจทย์ย่อยหลาย ๆ โจทย์ก่อน เพื่อที่จะได้ง่ายต่อการตัดสินใจ หรือจะนำเอาปัจจัยต่าง ๆ ที่เกี่ยวข้องกับการตัดสินใจหรือ เกี่ยวข้องกับโจทย์หลักมาตั้งเป็นคำถามใหม่หรือแตกเป็นโจทย์ย่อย และถามตัวเองใหม่อีกครั้ง เช่น วันนี้ มีเพื่อนมาถามเราว่าไปกินข้าวมันไก่ด้วยกันไหม มนุษย์อย่างเราจะคิดก่อนว่าอร่อยหรือเปล่า ถ้าไม่อร่อยแล้วราคาแพงไหม? แล้วค่อยตัดสินใจว่าจะไปหรือไม่ไป



Model Decision Tree เป็น Rule-Based Model ที่จะสร้างเงื่อนไข If-else ขึ้นมาจากข้อมูลในตัวแปร เพื่อที่จะแบ่งข้อมูลออกเป็นกลุ่มใหม่ที่สามารถอธิบาย Target ได้ดีที่สุด โดยการสร้างเงื่อนไข If-else ในแต่ละตัวแปร จะถูกกำหนดด้วย Objective Function ซึ่ง Model Decision Tree มี Objective Function อยู่หลายตัว ตามประเภทของ Decision Tree นั้น ๆ

Decision Tree จะแบ่งออกเป็น 2 ประเภท คือ

Regression Tree คือ Decision Tree ที่ใช้สำหรับการทำโจทย์ Regression โดยมีค่า Residual sum of squares (RSS) เป็น Objective Function ในการหาจุดที่ดีที่สุดในการแบ่งข้อมูล (Split point) จากการ Minimize ให้ RSS มีค่าน้อยที่สุด

– Residual ( $e_i$ ) คือ ค่าความคลาดเคลื่อน หรือค่า Error ระหว่าง  $y$  ทุก ๆ จุดในข้อมูล กับ  $\hat{y}$  ที่ได้มาจากการประมาณค่าขึ้นจากการคำนวณ Residual ของข้อมูลตัวที่  $i$

$$e_i = y_i - \hat{y}_i$$

– Residual sum of squares (RSS) คือ การวัดค่า Residual หรือ ค่า Error ของทุก ๆ จุดในชุดข้อมูล และนำมายกกำลังสอง เพื่อให้ค่า Residual เป็นบวก และเป็นการทำ Normalize ด้วย เนื่องจากถ้า  $\hat{y}$  มีค่ามากกว่า  $y_i$  จะทำให้ค่า Residual ติดลบ

การคำนวณ Residual sum of squares

$$RSS = e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2$$

Classification Tree คือ Decision Tree ที่ใช้สำหรับการทำ Classification โดยจะใช้ Gini Impurity หรือ Entropy เป็น Objective Function ในการหาจุดที่ดีที่สุดในการแบ่งข้อมูล (Split point)

– Gini Impurity คือ การวัดค่า Impurity หรือ ค่าความไม่บริสุทธิ์ในการอธิบาย Target ของกลุ่มที่ถูกแบ่งออกมาจากตัวแปร นั้นหมายความว่าถ้าค่า Impurity ยิ่งต่ำก็ยิ่งแบ่งข้อมูลออกมาได้ดีนั่นเอง

– การคำนวณ Gini Impurity คือ การนำเอาผลรวมของค่าความน่าจะเป็นของเหตุการณ์ที่เราสนใจมาคูณด้วย (1 ลบ ค่าความน่าจะเป็นของเหตุการณ์ที่เราสนใจ)

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk})$$

หลังจากได้ค่า Gini Impurity ของแต่ละกลุ่มในทุก ๆ ตัวแปรแล้ว จะทำการหาค่า Weighted Gini Impurity เพื่อเลือกตัวแปรที่มีค่า Weighted Gini Impurity ต่ำที่สุดมาใช้ในการตัดสินใจก่อน เพราะสามารถ Split ข้อมูลได้ดีที่สุด

การคำนวณ Weighted Gini Impurity คือ การนำเอาผลรวมของค่า Gini ในเหตุการณ์ที่เราสนใจ คูณกับจำนวนข้อมูลของ Class ที่ i ในตัวแปรที่เราสนใจ และหารด้วยจำนวนข้อมูลทั้งหมดในตัวแปรที่เราสนใจ

$$GINI_{split} = \sum_{i=1}^k \frac{n_i}{n} GINI(i)$$

### 2.3.4 แผนภาพกระแสข้อมูล (Data Flow Diagram: DFD)

เป็นเครื่องมือที่ใช้กันอย่างแพร่หลายในการเขียนแบบระบบใหม่ในการเขียนแผนภาพจำลองการทำงานของกระบวนการ (Process) ต่าง ๆ ในระบบ โดยเฉพาะกับ ระบบที่ "หน้าที" ของระบบมีความสำคัญ และมีความสลับซับซ้อนมากกว่าข้อมูลที่ไหลเข้า เป็นเครื่องมือเชิงโครงสร้างที่ใช้บรรยายภาพรวมของระบบโดยแสดงขั้นตอนการทำงานของระบบ หรือโพรเซส (process) ระบุแหล่งกำเนิดของข้อมูล การไหลของข้อมูล ปลายทาง ข้อมูล การเก็บข้อมูลและการประมวลผลข้อมูล กล่าวง่าย ๆ คือดีเอฟดีจะช่วยแสดงแผนภาพ ว่าข้อมูลมาจากไหน จะไปไหน เก็บข้อมูลไว้ที่ไหน มีอะไรเกิดขึ้นกับข้อมูลระหว่างทาง เรียกว่าแผนภาพกระแสข้อมูลหรือ แผนภาพแสดงความเคลื่อนไหวของข้อมูลโดยดีเอฟดี

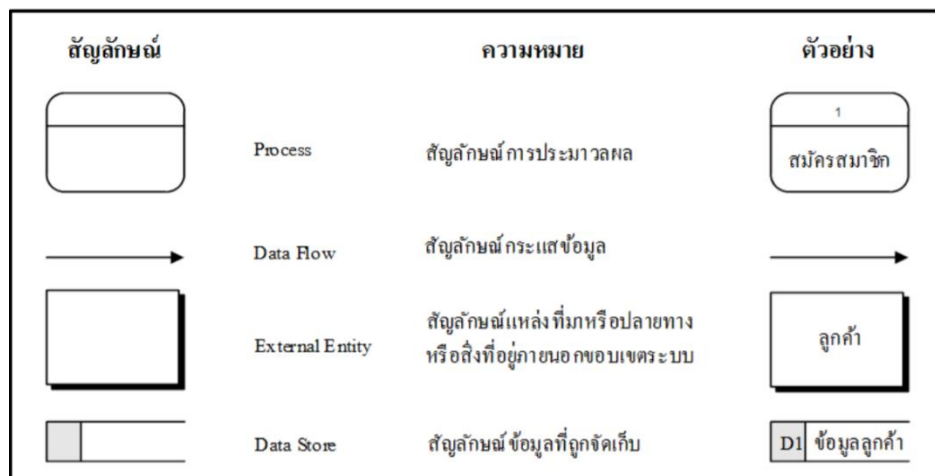
แผนภาพกระแสข้อมูล เป็นแบบจำลองที่ถูกนำมาใช้กับวิธีการ พัฒนาระบบแบบดั้งเดิมตามแนวทางการวิเคราะห์และออกแบบระบบเชิงโครงสร้าง และถือ เป็นหนึ่งในเทคนิคที่ถูกนำมาใช้งานบ่อยที่สุดในเรื่องของการสร้างแบบจำลองกระบวนการ และ หากเปรียบเทียบกับการพัฒนาซอฟต์แวร์กับการสร้างบ้านแล้ว จะพบว่าการสร้างบ้านขึ้นมา สักหลังหนึ่ง เจ้าของบ้านกับสถาปนิกจะต้องมีการตกลงถึงความเข้าใจระหว่างกันในเรื่องต้นว่า จะสร้างบ้านบนพื้นที่นี้ในรูปแบบอย่างไร แบบบ้านทรงไหน มีกี่ชั้น กี่ห้อง มีการแบ่งสัดส่วนห้อง ต่าง ๆ ไว้อย่างไร ซึ่งสิ่งเหล่านี้ก็คือความต้องการหรือ Requirement นั้นเอง จากนั้นสถาปนิกก็ จะรวบรวมความต้องการของลูกค้าในเรื่องต้น เพื่อนำไปเขียนเป็นแบบแปลน หรือ แบบพิมพ์เขียว (Blueprint) ขึ้นมา ซึ่งแบบพิมพ์เขียวนี้เองที่จะนำมาใช้เป็นข้อตกลงระหว่างกัน ว่าบ้านที่สถาปนิกออกแบบนั้นเป็นไปตามแบบที่ลูกค้าต้องการใช้หรือไม่ ทั้งนี้สถาปนิกอาจ จำเป็นต้องปรับปรุงแบบแปลนบางส่วนตามที่ลูกค้าต้องการ จนกระทั่งได้แบบแปลนที่สมบูรณ์ จากนั้นก็จะนำแบบแปลนดังกล่าวไปให้วิศวกรเพื่อควบคุมงานก่อสร้างต่อไป ดังนั้น แผนภาพกระแสข้อมูล จึงเปรียบเสมือนกับแบบพิมพ์เขียว โดยที่สถาปนิก เปรียบเสมือนกับ นักวิเคราะห์ระบบ และ วิศวกร เปรียบเสมือนกับนักวิศวกรรมซอฟต์แวร์หรือผู้จัดการโครงการ ที่ทำหน้าที่ควบคุมโครงการสร้างบ้านให้เป็นไปตามแผนและงบประมาณที่กำหนด ส่วนคนงาน ก่อสร้าง ก็เปรียบเสมือนกับโปรแกรมเมอร์ ที่จะต้องสร้างบ้านตามที่สถาปนิกออกแบบไว้



แผนภาพกระแสข้อมูล ก็เป็นแบบจำลองกระบวนการที่นำมาใช้กับ การวิเคราะห์และ ออกแบบระบบเชิงโครงสร้าง ที่มีการนำมาใช้ตั้งแต่ยุคที่มีการเริ่มใช้ ภาษาระดับสูง โดยแผนภาพ ดังกล่าวจะแสดงความสัมพันธ์ระหว่างกระบวนการทำงานกับ ข้อมูล (Process and Data) เพื่อให้ ทราบว่าข้อมูลมาจากไหน ข้อมูลไปที่ใด ข้อมูลเก็บไว้ที่ไหนและมีกระบวนการอะไรบ้างที่เกิดขึ้นใน ระบบ โดยที่แผนภาพกระแสข้อมูลจะแสดงภาพรวมของ ระบบ และรายละเอียดที่เกี่ยวข้องระหว่าง กระบวนการกับข้อมูล แต่ในบางครั้งนักวิเคราะห์ ระบบก็ต้องการทราบรายละเอียดอื่น ๆ ที่ นอกเหนือไปจากนี้ ก็จะต้องใช้เครื่องมืออื่นเข้าช่วย เช่น คำอธิบายการประมวลผล ตารางการ ตัดสินใจ หรือแบบจำลองข้อมูล ทั้งนี้ ขึ้นอยู่กับ ความต้องการในรายละเอียดเฉพาะในเรื่องนั้น ๆ เป็นสำคัญ ซึ่งขั้นตอนการสร้าง แผนภาพกระแสข้อมูล สามารถอธิบายได้ดังต่อไปนี้

- 1) ขั้นตอนที่ 1 ศึกษารูปแบบการทำงานระบบเดิม
- 2) ขั้นตอนที่ 2 ศึกษาตรรกะการทำงานของระบบงาน
- 3) ขั้นตอนที่ 3 ความต้องการใหม่ ๆ มีอะไรบ้าง
- 4) ขั้นตอนที่ 4 จะสร้างระบบงานใหม่ให้ตรงกับความต้องการ

อย่างที่ได้อธิบายไปแล้ว แผนภาพกระแสข้อมูล เป็นวิธีการนำเสนอภาพรวมด้าน ความ ต้องการหลัก ๆ ของระบบ อันประกอบด้วยอินพุต เอาต์พุต โปรเซส และ ข้อมูล โดยทุกคนใน ทีมงานพัฒนาระบบ สามารถมองเห็นระบบได้จากแผนภาพนี้ เพื่อนำไปใช้ เป็นแนวทางในการ ออกแบบระบบ สำหรับสัญลักษณ์ที่ใช้ในแผนภาพกระแสข้อมูลจะมีอยู่ 5 สัญลักษณ์ด้วยกัน คือ โปรเซส (Process) ดาตาโฟลว์ (Data Flow) เอ็กซ์เทอร์นัลเอนทิตี (External Entities) ดาตาสโตร์ (Data Store) และเรียลไทม์ลิงก์ (Real-time Link) ดังรูปภาพต่อไปนี้



ภาพที่ 2.1 สัญลักษณ์ที่ใช้สำหรับการเขียนแผนภาพกระแสข้อมูล

1) โพรเซส (Process) เป็นสัญลักษณ์แทนกิจกรรมที่เกิดขึ้นในระบบ หรือกระบวนการที่ต้องทำในระบบ ตามปกติแล้ว แผนภาพกระแสข้อมูลจะต้องมีอย่างน้อยหนึ่งโพรเซสเสมอ โดยดาตาโฟลว์ที่เดินทางออกจากโพรเซสก็จะเรียกว่าเอาต์พุต ดังนั้นกาตา โฟลว์ที่เอาต์พุตออกมานั้น ข้อมูลย่อมได้รับการเปลี่ยนแปลง

สัญลักษณ์โพรเซสจะต้องมีหมายเลขกำกับเสมอ ซึ่งเรียกว่า “หมายเลขโพรเซส” โดยมักจะกำหนดเป็นหมายเลขตามลำดับ แต่การลำดับของหมายเลข โพรเซส ไม่ได้หมายความว่า จะต้องดำเนินกิจกรรมตามเลขที่กำกับไว้ในโพรเซสแต่อย่างใด และที่สำคัญโพรเซสจะซ้ำกันไม่ได้ ส่วนชื่อที่นำมาใช้กำกับโพรเซส จะใช้คำกริยาซึ่งหมายถึง การกระทำ ซึ่งในโพรเซสของแผนภาพกระแสข้อมูล จะไม่มีการแสดงรายละเอียดเกี่ยวกับ วิธีการทำงาน แต่นำเสนอเพียงแต่ว่าทำหน้าที่อะไร โดยมีดาตาโฟลว์อะไรบ้างอินพุตเข้ามา และมีดาตาโฟลว์อะไรบ้างเอาต์พุตออกไป ส่วนรายละเอียดเกี่ยวกับวิธีการทำงานของแต่ละ โพรเซสจะปรากฏอยู่ในแบบจำลองอีกชนิดหนึ่ง

2) ดาตาโฟลว์ (Data Flows) คือกระแสข้อมูลที่ใช้สัญลักษณ์แทน ด้วยเส้นลูกศรที่ไปพร้อมกับข้อมูล กล่าวคือ กระแสข้อมูลก็คือเส้นทางที่ข้อมูลเคลื่อนที่นั่นเอง ทำให้เราได้ทราบถึงข้อมูลต่าง ๆ ที่เคลื่อนไหวไปมาระหว่างโพรเซส ดาตาสโตร์และเอนทิตี โดยทุก ๆ โพรเซสในแผนภาพกระแสข้อมูล เมื่อมีดาตาโฟลว์อินพุตเข้าไป จะต้องมี ดาตาโฟลว์เอาต์พุตออกมาเสมอ

3) เอ็กเทอร์นัลเอนทิตี (External Entities) จะทำหน้าที่ส่งข้อมูล อินพุตเข้ามายังโปรเซส เพื่อแสดงถึงแหล่งที่มาของข้อมูล (Source) รวมถึงการรับเอาต์พุตจาก โปรเซส เพื่อแสดงถึง จุดสิ้นสุดของกระแสข้อมูล (Sink) โดยเอ็กเทอร์นัลเอนทิตีสามารถเป็นได้ ทั้งบุคคล หน่วยงาน หรือระบบงาน เอ็กเทอร์นัลเอนทิตี จึงมักถูกจัดวางตำแหน่งให้อยู่บริเวณ ด้านนอกหรือรอบ ๆ แผนภาพ ทั้งนี้ก็เพื่อให้แผนภาพแลดูสวยงาม และง่ายต่อการตรวจสอบ

4) ดาตาสโตร์ (Data Stores) เป็นแหล่งเก็บข้อมูล ซึ่งจะไม่สนใจว่า ระบบจะใช้สื่อจัดเก็บ ข้อมูลประเภทใดก็ตาม ทุก ๆ ดาตาสโตร์จะต้องมีชื่อข้อมูลที่จัดเก็บ และ มีการลำดับไว้ โดยดาตาสโตร์เหล่านี้จะถูกใช้งานโดยโปรเซส และสามารถทำซ้ำได้ ส่วนที่มา ของดาตาสโตร์นั้น จะได้มาจากการสร้างแบบจำลองข้อมูล (Data Model)

5) เรียลไทม์ลิงก์ (Real-Time Link) เป็นการเชื่อมโยงสื่อสาร ระยะไกล ที่มีการโต้ตอบกัน ไปมาระหว่างเอ็กเทอร์นัลเอนทิตีกับโปรเซส โดยจะเป็นการสื่อสาร โต้ตอบแบบทันทีทันใด

จะเห็นได้ว่าสัญลักษณ์ของแผนภาพกระแสข้อมูลที่นำมาใช้ในการ พัฒนาระบบใช้มาตรฐานของ Gane and Sarson แผนภาพกระแสข้อมูล เป็นแผนภาพที่แสดง ภาพรวมความต้องการหลัก ๆ ของระบบในรูปแบบของไดอะแกรม ซึ่งประกอบด้วย อินพุต เอาต์พุต กระบวนการ และข้อมูล แผนภาพกระแสข้อมูลที่นิยมใช้งานจนถึงปัจจุบัน และจัดเป็น แผนภาพที่ดูง่ายต่อการทำความเข้าใจ

### 2.3.5 แผนผังแสดงความสัมพันธ์ระหว่างเอนทิตี (Entity-Relationship Diagram)

เป็นแบบจำลองข้อมูลเชิงแนวคิดที่ใช้แสดงลักษณะโดยรวมของ ข้อมูลในระบบ โดยนำเสนอในรูปแบบแผนภาพหรือไดอะแกรมที่มักเรียกกันว่าแผนภาพอีอาร์ หรือ อีอาร์ไดอะแกรม (E-R Diagram) ซึ่งจุดประสงค์ของแบบจำลองเชิงแนวคิด ก็คือ ต้องการ นำเสนอให้เกิดความเข้าใจระหว่างผู้ออกแบบและผู้ใช้งาน โดยนักออกแบบฐานข้อมูลซึ่งเป็นผู้มี ความรู้เชิงเทคนิคจะเป็นผู้สร้างแบบจำลองเชิงแนวคิดนี้ขึ้นมาใช้สำหรับสื่อสารกับผู้ใช้เพื่อให้ เกิดความเข้าใจตรงกันในเรื่องของข้อมูลที่มีอยู่ในระบบ

เนื่องจากอีอาร์ไดอะแกรม (E-R Diagram) เป็นแบบจำลองที่ นำเสนอด้วยแผนภาพที่ช่วยสื่อสารให้เกิดความเข้าใจร่วมกันระหว่างผู้วิเคราะห์และผู้ใช้ได้เป็น อย่างดี อีอาร์ไดอะแกรมจึงจะ เสนอรายละเอียดหรือข้อมูลต่าง ๆ ของธุรกิจในภาพรวม ซึ่ง แสดงถึงโครงสร้างของฐานข้อมูลที่ เป็นอิสระจากซอฟต์แวร์ที่จะใช้ในการพัฒนาฐานข้อมูล รวมทั้งรายละเอียดและความสัมพันธ์ ระหว่างข้อมูลในระบบในลักษณะที่เป็นภาพรวม ทำให้ เป็นประโยชน์อย่างมากต่อการรวบรวม และวิเคราะห์รายละเอียด นอกจากนี้ผู้ใช้ส่วนใหญ่ ที่ไม่เชี่ยวชาญความรู้เชิงเทคนิคก็สามารถทำ ความเข้าใจผ่านแผนภาพได้ง่าย แต่อย่างไรก็ตาม ผู้ใช้ที่ไม่มีความรู้เชิงเทคนิคเหล่านี้เองที่กลับมา มีบทบาทสำคัญในเรื่องของการให้ข้อมูลต่อนัก ออกแบบฐานข้อมูลที่เชี่ยวชาญ ด้านเทคนิคเป็น อย่างดี เพื่อนำไปพัฒนา อีอาร์ไดอะแกรมที่สมบูรณ์ต่อไป อย่างไรก็ตามในการจะสร้างอีอาร์ ไดอะแกรมได้นั้น จะต้อง เรียบรู้ถึงสัญลักษณ์ที่ใช้แทนความหมายต่าง ๆ ซึ่งประกอบด้วย รายละเอียดที่สำคัญดังต่อไปนี้

1) เอนทิตี (Entity) คือ สิ่งของหรือวัตถุที่เราสนใจ ซึ่งอาจจับต้องได้ และเป็นได้ทั้งนามธรรม อาจจะเป็นบุคคล สถานที่ วัตถุ เหตุการณ์ หรือแนวคิดที่ก่อให้เกิด กลุ่มของข้อมูลที่ต้องการ และ โดยทั่วไปเอนทิตีจะมีลักษณะที่แยกออกจากกันไป เช่น เอนทิตี พนักงาน จะแยกออกเป็นของ พนักงาน เอนทิตีเงินเดือนของพนักงานคนหนึ่งก็อาจเป็นเอนทิตี หนึ่งในระบบของโรงงาน และ ทั่วไปแล้วเอนทิตีจะมีกลุ่มที่บอกคุณสมบัติที่บอกลักษณะของ เอนทิตี เช่น พนักงานมีรหัสพนักงาน ชื่อ นามสกุล และแผนก เป็นต้น ซึ่งค่าของคุณสมบัติ บางกลุ่มที่ทำให้สามารถแยกเอนทิตีออกจาก เอนทิตีอื่นได้





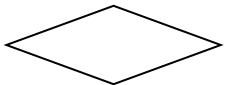
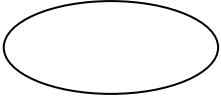
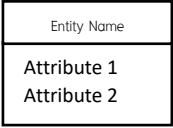
2) แอตทริบิวต์ (Attribute) คือ คุณสมบัติของเอนทิตี วัตถุหรือ สิ่งของที่เราสนใจ สัญลักษณ์แอตทริบิวต์จะมีลักษณะเป็นรูปวงรี โดยคุณสมบัตินี้มีอยู่ในทุก เอนทิตี เช่น ชื่อ นามสกุล ที่อยู่ แผนก เป็น Attribute ของเอนทิตีพนักงาน โดยทั่วไปแล้วโมเดลข้อมูล เรามักจะ พบว่า Attribute มีลักษณะข้อมูลพื้นฐานอยู่โดยที่ไม่ต้องมีคำอธิบายมากมาย และแอตทริบิวต์ก็ไม่ สามารถอยู่แบบโดด ๆ ได้โดยที่ไม่มีเอนทิตีหรือความสัมพันธ์

3) ความสัมพันธ์ (Relationship) คือ ความสัมพันธ์ระหว่างเอนทิตี โดยแต่ละความสัมพันธ์ควรมีชื่อระบุไว้เพื่อให้อธิบาย ซึ่งปกติจะใช้สัญลักษณ์สี่เหลี่ยมขนมเปียกปูน ที่ภายในระยะคำกริยาไว้เพื่ออธิบายความสัมพันธ์

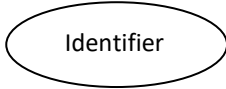
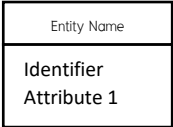
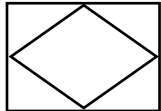
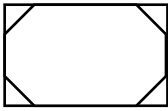
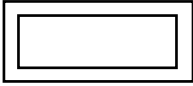
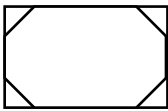
อย่างไรก็ตาม โดยธรรมชาติแล้วแบบจำลองเชิงสัมพันธ์จะไม่สนับสนุนความสัมพันธ์แบบกลุ่มต่อกลุ่ม แต่ในขั้นตอนการออกแบบเชิงแนวคิดนั้นสามารถมีขึ้นได้ แต่เมื่อมาถึงกระบวนการออกแบบเชิงตรรกะแล้ว ควรมีการปรับปรุงใหม่ เนื่องจาก ความสัมพันธ์แบบกลุ่มต่อกลุ่มนั้น มีการทับซ้อนกันของข้อมูลอยู่ ดังนั้นจึงจำเป็นต้องสร้างเอนทิตีขึ้นมาใหม่ เพื่อแปลงความสัมพันธ์จากกลุ่มต่อกลุ่มไปเป็นความสัมพันธ์ แบบหนึ่งต่อกลุ่ม

นอกจากนี้ในส่วนของการออกแบบความสัมพันธ์ของข้อมูล จะต้อง มีการใช้เครื่องมือสำหรับแสดงความสัมพันธ์ของข้อมูล ซึ่งใช้สัญลักษณ์ดังกล่าว สามารถแบ่ง ออกได้เป็นสัญลักษณ์ดังต่อไปนี้


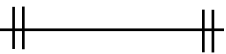

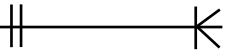
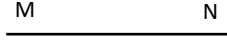
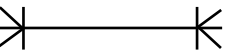
ตารางที่ 2.1 แสดงข้อมูลสัญลักษณ์ในการออกแบบความสัมพันธ์ของข้อมูล

Chen Model	Crow's Foot Model	ความหมาย
		ใช้แสดง Entity
		Relationship Line เส้นเชื่อม ความสัมพันธ์ระหว่าง Entity
	-	Relationship ใช้แสดงความสัมพันธ์ ระหว่าง Entity สำหรับ Crow's Foot Model ใช้ตัวอักษรเขียนแสดง
		Attribute ใช้แสดง Attribute ของ Entity

ตารางที่ 2.1 (ต่อ) แสดงข้อมูลสัญลักษณ์ในการออกแบบความสัมพันธ์ของข้อมูล

Chen Model	Crow's Foot Model	ความหมาย
		ใช้แสดงคีย์หลัก Identifier
		Associative Entity
		Weak Entity

ตารางที่ 2.2 แสดงข้อมูลสัญลักษณ์ในการออกแบบความสัมพันธ์ของข้อมูล

Chen Model	Crow's Foot Model	ความหมาย
		หนึ่ง-ต่อ-หนึ่ง (one-to-one)
		หนึ่ง-ต่อ-กลุ่ม (one-to-many)
		กลุ่ม-ต่อ-กลุ่ม (many-to-many)

### 2.3.6 พจนานุกรมข้อมูล (Data Dictionary)

พจนานุกรมข้อมูล (Data Dictionary) คือ พจนานุกรมข้อมูลที่แสดง รายละเอียด ตารางข้อมูลต่าง ๆ ในฐานข้อมูล (Database) ทำให้สามารถค้นหารายละเอียดที่ต้องการได้สะดวกมากยิ่งขึ้น พจนานุกรมข้อมูลเป็นการผสมผสานระหว่างรูปแบบของ พจนานุกรมโดยทั่วไปและรูปแบบของข้อมูลในระบบงานคอมพิวเตอร์ เพื่ออธิบายชนิดของ ข้อมูลแต่ละตัวว่าเป็น ตัวเลข อักขระ ข้อความ หรือวันที่ เป็นต้น เพื่อช่วยในการอธิบายรายละเอียดต่างๆ ในการอ้างอิงหรือ ค้นหาที่เกี่ยวกับข้อมูล หรือจะเรียกง่ายๆ ว่าเป็นเอกสาร ที่ใช้อธิบายฐานข้อมูลหรือการจัดเก็บ ฐานข้อมูล

ซึ่งพจนานุกรมข้อมูลเป็นเครื่องมือที่ช่วยในการจัดเก็บรายละเอียด ต่าง ๆ เกี่ยวกับข้อมูล ให้เป็นหมวดหมู่ ทำให้สามารถค้นหารายละเอียดที่ต้องการได้โดยสะดวก ตัวอย่างเช่น ผู้ใช้อาจเก็บ ข้อมูลเกี่ยวกับรายงานต่าง ๆ ไว้ภายในหมวดรายการชื่อ “Report” เป็นต้น ทั้งนี้ วัตถุประสงค์ของการจัดเก็บรายละเอียดเกี่ยวกับข้อมูลให้เป็นหมวดหมู่ ในพจนานุกรมข้อมูล คือ เพื่อให้สามารถ อธิบายความหมายของข้อมูลต่าง ๆ แก่ผู้ใช้งานได้ อย่างถูกต้อง และเป็นมาตรฐานเดียวกันทั้ง หน่วยงาน

ในการกำหนดโครงสร้างของฐานข้อมูล ระบบจัดการฐานข้อมูล (Data Base Management System : DBMS) ซึ่งเป็นซอฟต์แวร์ที่ทำหน้าที่สื่อกลางประสานงาน ระหว่างผู้ใช้กับฐานข้อมูล โดย ทำการควบคุม ดูแล และจัดการเรื่องต่าง ๆ ที่เกี่ยวข้องกับข้อมูล ภายในฐานข้อมูล ตัวอย่างเช่น การจัดเก็บและดูแลรักษาข้อมูล การปรับปรุงข้อมูล และการ เรียกใช้ข้อมูล เป็นต้น จะทำการเก็บ รวบรวม รายละเอียดและคำอธิบายต่าง ๆ ที่เกี่ยวข้องกับ ข้อมูล ตัวอย่างเช่น ชื่อตาราง (Table) ชื่อ ขอบเขตข้อมูล (Field) และคีย์ต่าง ๆ เป็นต้น ไว้ใน พจนานุกรมข้อมูลที่มีการสร้างขึ้นมาเป็นส่วน หนึ่งของฐานข้อมูล ซึ่ง Data Dictionary มี ประโยชน์ ดังนี้

1) ช่วยจัดเก็บรายละเอียดข้อมูล 2) ช่วยแสดงความหมายที่เกี่ยวกับระบบ 3) ช่วยทำ เอกสารที่บอกคุณลักษณะของระบบ 4) ช่วยหาข้อบกพร่องและสิ่งที่หายไปจากระบบ

นอกจากนี้ พจนานุกรมข้อมูล (Data Dictionary) ยังมีหน้าที่สำคัญ ในเรื่องอื่น ๆ อีกมากมาย ดังต่อไปนี้

1) การควบคุมการใช้ฐานข้อมูลพร้อมกันจากผู้ใช้หลายคน เนื่องจากในระบบฐานข้อมูล อาจมีผู้ใช้งานฐานข้อมูลพร้อมกันหลายคนในเวลาเดียวกัน ซึ่งผู้ใช้ แต่ละคนสามารถค้นหาและเรียกใช้ข้อมูลได้พร้อมกัน หากทว่าในการปรับปรุง การเพิ่ม หรือ การลบข้อมูลใด ๆ จะมีผู้ใช้เพียงคนเดียวเท่านั้นที่สามารถดำเนินการได้ เนื่องจากการใช้ หลักการล็อกข้อมูลเพื่อป้องกันการแก้ไขข้อมูล เดียวกันของผู้ใช้หลายคนในขณะเดียวกัน ซึ่ง รายละเอียดดังกล่าวจะมีการเก็บอยู่ในพจนานุกรมข้อมูล

2) การรักษาความปลอดภัยของข้อมูล เนื่องจากในระบบ ฐานข้อมูลจะมีผู้ใช้งานฐานข้อมูลพร้อมกันหลายคน ข้อมูลที่สำคัญบางอย่างจึงควรได้รับการป้องกันเพื่อไม่ให้ผู้ที่ไม่เกี่ยวข้องเข้ามาใช้งานหรือทำการแก้ไขเปลี่ยนแปลงได้ ซึ่งระบบจัดการ ฐานข้อมูลจะจัดการเก็บรวบรวมรายละเอียดเกี่ยวกับสิทธิของผู้ใช้งานแต่ละคนไว้ในฐานข้อมูล ว่า ใครบ้างที่มีสิทธิในการค้นหาข้อมูลได้เพียงอย่างเดียว ใครที่มีสิทธิในการปรับปรุง เพิ่มเติม หรือลบข้อมูลการควบคุมความบูรณภาพของข้อมูล เนื่องจากรายละเอียดต่าง ๆ เกี่ยวกับ ข้อมูลในฐานข้อมูลจะถูกเก็บรวบรวมไว้ในพจนานุกรมข้อมูล ดังนั้น ในการแก้ไขเปลี่ยนแปลง ข้อมูลใด ๆ พจนานุกรมข้อมูลจะทำการตรวจสอบว่า ข้อมูลที่ทำการแก้ไขเปลี่ยนแปลงนั้น แตกต่างไปจากขอบเขตที่ได้มีการกำหนดไว้แล้วหรือไม่ โดยจะยอมรับการแก้ไขเปลี่ยนแปลง เฉพาะข้อมูลที่อยู่ภายในขอบเขตที่กำหนดไว้เท่านั้น โดยโครงสร้าง ฐานข้อมูลของระบบ สารสนเทศภายใต้โปรแกรมฐานข้อมูลมายเอสคิวแอล (MySQL) โดยใช้ภาษาเอสคิวแอล (SQL) ในการจัดการฐานข้อมูล

## 2.4 เครื่องมือที่ใช้ในการพัฒนาระบบ

### 2.4.1 ซอฟต์แวร์ (Software)

ซอฟต์แวร์ (Software) หมายถึงชุดคำสั่งหรือโปรแกรมที่ใช้สั่งงานให้ คอมพิวเตอร์ทำงาน ซอฟต์แวร์จึงหมายถึงลำดับขั้นตอนการทำงานที่เขียนขึ้นด้วยคำสั่งของ คอมพิวเตอร์ คำสั่งเหล่านี้เรียงกันเป็นโปรแกรมคอมพิวเตอร์ จากที่ทราบมาแล้วว่า คอมพิวเตอร์ทำงานตามคำสั่ง การทำงานพื้นฐานเป็นเพียงการกระทำกับข้อมูลที่เป็นตัว เลขฐานสอง ซึ่งใช้แทนข้อมูลที่เป็นตัวเลข



ตัวอักษร รูปภาพ หรือแม้แต่เป็นเสียงพูด โปรแกรมคอมพิวเตอร์ที่ใช้สั่งงานคอมพิวเตอร์จึงเป็นซอฟต์แวร์ เพราะเป็นลำดับขั้นตอนการทำงานของ คอมพิวเตอร์ คอมพิวเตอร์เครื่องหนึ่งทำงานแตกต่างกันได้มากมายด้วยซอฟต์แวร์ที่แตกต่าง กัน ซอฟต์แวร์จึงหมายถึงรวมถึงโปรแกรมคอมพิวเตอร์ทุกประเภทที่ทำให้คอมพิวเตอร์ทำงานได้ โดยการพัฒนาโครงการได้ใช้ซอฟต์แวร์ดังนี้

#### 2.4.1.1 Adobe Photoshop CC 2015

สุนิสา คงทอง (2559) ได้กล่าวถึงโปรแกรม Adobe Photoshop CC 2015 (อะโดบี โฟโตชอป) เป็น โปรแกรมแต่งรูป ใช้สำหรับตัดต่อภาพ แต่งรูปภาพ ตัดต่อภาพถ่ายจากกล้อง dslr ตกแต่งภาพจากสมาร์ตโฟน แก้ไขรายละเอียดของภาพ ออกแบบโลโก้ ออกแบบป้ายโฆษณาต่างๆ ออกแบบภาพกราฟฟิกระดับมืออาชีพ แก้ไขภาพถ่ายที่มีความละเอียดสูงโดยไม่ทำลายไฟล์ภาพต้นฉบับ

Photoshop เป็นซอฟต์แวร์แก้ไขภาพที่มีประสิทธิภาพที่สุดในโลก ที่มีเครื่องมือในการแก้ไขรูปภาพมากมายที่คุณต้องการ เพื่อสร้าง และแก้ไขทุกรายละเอียดของภาพโดยยังคงรักษาคุณภาพของไฟล์ต้นฉบับเอาไว้ เป็นโปรแกรมมาตรฐานในการแก้ไขกราฟิกแบบแรสเตอร์ ช่วยให้คุณสามารถสร้างแก้ไขและเขียนภาพแรสเตอร์ได้หลากหลายรูปแบบ อีกทั้งยังรองรับรูปแบบไฟล์ภาพทุกชนิดและรูปแบบไฟล์วิดีโอที่หลากหลาย

#### 2.4.1.2 พีเอชพีมายแอดมิน (phpMyAdmin)

โอภาส เอี่ยมสิริวงศ์ (2559) ได้กล่าวถึงโปรแกรม พีเอชพีมาย แอดมิน (phpMyAdmin) คือ โปรแกรมที่ถูกพัฒนาโดยใช้ภาษา PHP เพื่อใช้ในการบริหาร จัดการฐานข้อมูล MySQL แทนการคีย์คำสั่ง เนื่องจากถ้าเราจะใช้ฐานข้อมูลที่เป็น MySQL บางครั้งจะมีความลำบากและยุ่งยากในการใช้งาน ดังนั้นจึงมีเครื่องมือในการจัดการฐานข้อมูล MySQL ขึ้นมาเพื่อให้สามารถจัดการ ตัวDBMS ที่เป็น MySQL ได้ง่าย และสะดวกยิ่งขึ้น โดย phpMyAdmin ก็ถือเป็นเครื่องมือชนิดหนึ่งในการจัดการนั่นเอง นอกจากนี้ phpMyAdmin เป็นส่วนต่อประสานที่สร้างโดยภาษาพีเอชพี (PHP) ซึ่งใช้จัดการฐานข้อมูล MySQL ผ่าน เว็บเบราว์เซอร์ โดยสามารถที่จะทำการสร้างฐานข้อมูลใหม่ หรือทำการสร้างตารางใหม่ ๆ และยังมีการทำงานที่ใช้สำหรับการทดสอบการ

query ข้อมูลด้วยภาษา SQL พร้อมกันนั้น ยังสามารถทำการ insert delete update หรือแม้กระทั่งใช้ คำสั่งต่าง ๆ เหมือนกับกับการใช้ ภาษา SQL ในการสร้างตารางข้อมูล

การแสดงผลของภาษาพีเอชพี (PHP) จะปรากฏในลักษณะของ ภาษาเอชทีเอ็มแอล (HTML) ซึ่งจะไม่แสดงคำสั่งที่ผู้ใช้เขียน ซึ่งเป็นลักษณะเด่นที่พีเอชพี แตกต่างจากภาษาในลักษณะโคลเอนต์ ไซด สคริปต์ เช่น ภาษาจาวาสคริปต์ ที่ผู้ชมเว็บไซต์ สามารถอ่าน ดูและคัดลอกคำสั่งไปใช้เองได้ นอกจากนี้พีเอชพี (PHP) ยังเป็นภาษาที่เรียนรู้ และเริ่มต้นได้ไม่ยาก ความสามารถการประมวลผลหลักของพีเอชพี (PHP) ได้แก่ การสร้างเนื้อหาอัตโนมัติจัดการคำสั่ง การอ่านข้อมูลจากผู้ใช้และประมวลผล การอ่านข้อมูลจากดาต้า เบส (Database) ความสามารถจัดการกับคุกกี้ ซึ่งทำงานเช่นเดียวกับโปรแกรมในลักษณะ CGI คุณสมบัติอื่น ๆ

คำสั่งของภาษาพีเอชพี สามารถสร้างผ่านทางโปรแกรมแก้ไข ข้อความทั่วไป เช่น โน้ตแพด (Notepad) ซึ่งทำให้การทำงานพีเอชพี (PHP) สามารถทำงานได้ใน ระบบปฏิบัติการหลักเกือบทั้งหมด สำหรับส่วนหลักของภาษาพีเอชพี (PHP) ยังมีโมดูลในการ รองรับ CGI มาตรฐาน ซึ่ง ภาษาพีเอชพี (PHP) สามารถทำงานเป็นตัวประมวลผล CGI ด้วย และ จากการที่ภาษาพีเอชพี (PHP) คุณมีอิสรภาพในการเลือก ระบบปฏิบัติการ และเว็บเซิร์ฟเวอร์ นอกจากนี้คุณยังสามารถใช้สร้างโปรแกรมโครงสร้าง สร้างโปรแกรมเชิงวัตถุ (OOP) หรือ สร้างโปรแกรมที่รวมทั้งสองอย่างเข้าด้วยกัน แม้ว่าความสามารถของคำสั่งโปรแกรมเชิงวัตถุ มาตรฐานในรุ่นนี้ยังไม่สมบูรณ์ แต่ตัวไลบรารีทั้งหลายของโปรแกรม และตัวโปรแกรมประยุกต์ รวมไปถึง PEAR library ได้ถูกเขียนขึ้นโดยใช้รูปแบบการเขียนแบบโปรแกรมเชิงวัตถุเท่านั้น นอกจากนี้ภาษาพีเอชพียังสามารถรองรับการสื่อสารกับการบริการในโพรโทคอลต่าง ๆ เช่น LDAP IMAP SNMP NNTP POP3 HTTP COM

#### 2.4.1.3 ชุดคำสั่งภาษาซีเอสเอส (CSS)

คณิน คล้ายทอง (2560) ได้กล่าวถึง ชุดคำสั่งภาษาซีเอสเอส เป็นภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสารภาษาเอชทีเอ็มแอล (HTML) โดย ที่ภาษาซีเอสเอส (CSS) กำหนดกฎเกณฑ์ในการระบุรูปแบบ หรือ “Style” ของเนื้อหาใน เอกสาร อันได้แก่ สีของข้อความ สีพื้นหลัง ประเภทตัวอักษร และการจัดวางข้อความ ซึ่งการ กำหนดรูปแบบ หรือ Style นี้ ใช้หลักการของการแยกเนื้อหาเอกสารภาษาเอชทีเอ็มแอล (HTML) ออกจากคำสั่งที่ใช้ในการ

จัดรูปแบบการแสดงผล deหนดให้รูปแบบของการแสดงผล เอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลล์์ของ เอกสารภาษาเอชทีเอ็มแอล (HTML) โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสาร บ่อยครั้ง หรือต้องการควบคุมให้รูปแบบการแสดงผลเอกสารภาษาเอชทีเอ็มแอล (HTML) มี ลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสาร ภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการ กำหนดรูปแบบ (Style) เอกสารภาษาเอชทีเอ็มแอล (HTML) ถูกเพิ่มเข้ามาครั้งแรกในภาษา เอชทีเอ็มแอล 4 (HTML4) เมื่อปีพุทธศักราช 2539 ในรูปแบบของภาษาซีเอสเอส (CSS) level 1 Recommendations ที่กำหนดโดย องค์การ World Wide Web Consortium หรือ W3C

#### 2.4.1.4 ชุดคำสั่งภาษาจาวาสคริปต์ (JAVA Script)

คณิน คล้ายทอง (2560) ได้กล่าวถึง ชุดคำสั่งภาษาจาวาสคริปต์ เป็น ภาษาคอมพิวเตอร์สำหรับการเขียนโปรแกรมบนระบบอินเทอร์เน็ต ที่deลังได้รับความนิยม อย่างสูง ภาษาจาวาสคริปต์ (JavaScript) เป็นภาษาสคริปต์เชิงวัตถุ ที่เรียกกันว่า “สคริปต์” (Script) ซึ่งในการสร้าง และพัฒนาเว็บไซต์ ใช้ร่วมกับภาษาเอชทีเอ็มแอล (HTML) เพื่อให้ เว็บไซต์ของเราดูมีการเคลื่อนไหว สามารถตอบสนองของผู้ใช้งานได้มากขึ้น ซึ่งมีวิธีการทำงานในลักษณะ “แปลความ และดำเนินงานไปที่ละคำสั่ง” (interpret) หรือเรียกว่า ออบเจ็กโอเรียนเต็ด (Object Oriented Programming) ที่มีเป้าหมายในการ ออกแบบและพัฒนาโปรแกรมในระบบ อินเทอร์เน็ต สำหรับผู้เขียนด้วยภาษาเอชทีเอ็มแอล (HTML) สามารถทำงานข้ามแพลตฟอร์มได้ โดยทำงานร่วมกับ ภาษาเอชทีเอ็มแอล (HTML) และภาษาจาวา (Java) ได้ทั้งทางฝั่งไคลเอนต์ (Client) และ ทางฝั่ง เซิร์ฟเวอร์ (Server)

ภาษาจาวาสคริปต์ (JavaScript) สามารถทำให้การสร้างเว็บเพจ มีลูกเล่น ต่าง ๆ มากมาย และยังสามารถโต้ตอบกับผู้ใช้ได้อย่างทันที เช่น การใช้เมาส์คลิก หรือ การกรอกข้อความในฟอร์ม เป็นต้น และเนื่องจากภาษาจาวาสคริปต์ (JavaScript) ช่วยให้พัฒนา สามารถสร้างเว็บเพจได้ตรงกับความต้องการ และมีความน่าสนใจมากขึ้น ประกอบกับเป็นภาษาเปิด ที่ใครก็สามารถนำไปใช้ได้ ดังนั้นจึงได้รับความนิยมเป็นอย่างสูง มีการใช้งานอย่างกว้างขวาง รวมทั้งได้ถูกdeหนดให้เป็นมาตรฐานโดย ECMA การทำงาน ของภาษาจาวาสคริปต์ (JavaScript) จะต้องมีการแปลความ

คำสั่ง ซึ่งขั้นตอนนี้จะถูกจัดการ โดยบราวเซอร์ เรียกว่าเป็น client-side script ดังนั้นภาษาจาวาสคริปต์ (JavaScript) จึงสามารถทำงานได้ เฉพาะบนบราวเซอร์ที่สนับสนุน ซึ่งปัจจุบันบราวเซอร์เกือบทั้งหมด ก็สนับสนุนภาษาจาวาสคริปต์ (JavaScript) อย่างไรก็ตาม สิ่งที่ต้องระวังคือภาษาจาวาสคริปต์ (JavaScript) มีการพัฒนาเป็นเวอร์ชันใหม่ ๆ ออกมาด้วย ดังนั้น ถ้านำโค้ดของเวอร์ชันใหม่ไปรันบนบราวเซอร์รุ่นเก่าที่ยังไม่สนับสนุน ก็อาจจะทำให้เกิดข้อผิดพลาดในการแสดงผลได้

การทำงานของภาษาจาวาสคริปต์ (JavaScript) เกิดขึ้นบน บราวเซอร์ เรียกว่าเป็น client-side Script ดังนั้น ไม่ว่าจะใช้เซิร์ฟเวอร์อะไร หรือที่ไหน ก็ยังคง สามารถใช้ภาษาจาวาสคริปต์ (JavaScript) ในเว็บเพจได้ต่างกับภาษาสคริปต์อื่น เช่น PHP หรือ ASP ซึ่งต้องแปลความและทำงานที่ตัวเครื่องเซิร์ฟเวอร์ เรียกว่า server-side script ดังนั้น จึงต้องใช้บนเซิร์ฟเวอร์ ที่สนับสนุนภาษาเหล่านี้เท่านั้น อย่างไรก็ตาม จากลักษณะดังกล่าว ก็ทำให้ภาษาจาวาสคริปต์ (JavaScript) มีข้อจำกัด คือ ไม่สามารถรับและส่งข้อมูลต่าง ๆ กับเซิร์ฟเวอร์โดยตรง เช่น การอ่านไฟล์จากเซิร์ฟเวอร์ เพื่อนำมาแสดงบนเว็บเพจ หรือรับข้อมูลจากผู้ชม เพื่อนำไปเก็บบนเซิร์ฟเวอร์ เป็นต้น

#### 2.4.1.5 Visual Studio วิศวกรรมสตูดิโอ

Visual Studio (วิศวกรรมสตูดิโอ) คือ โปรแกรมตัวหนึ่งที่เป็นเครื่องมือที่ช่วยพัฒนาซอฟต์แวร์และระบบต่างๆ ซึ่งสามารถติดต่อสื่อสารพูดคุยกับคอมพิวเตอร์ได้ในระดับหนึ่งแล้ว แต่ยังไม่สามารถพัฒนาเป็นระบบเองได้ เหมาะสมสำหรับภาษา VB และ VB.NET เนื่องจากไมโครซอฟต์ได้พัฒนาโปรแกรมและภาษาขึ้นมาควบคู่กันเพื่อให้ใช้งานได้ซึ่งกันและกัน ซึ่งนักโปรแกรมเมอร์จะนำเครื่องมือมาใช้ในการพัฒนาต่อยอดให้เกิดเป็นระบบต่างๆ หรือเป็นเว็บไซต์ และแอปพลิเคชันต่างๆ

ดังนั้น วิศวกรรมสตูดิโอจึงเป็นโปรแกรมตัวหนึ่งที่เป็นเครื่องมือที่ช่วยพัฒนาซอฟต์แวร์และระบบต่างๆ ซึ่งโปรแกรมได้มีการติดต่อสื่อสารกับคอมพิวเตอร์ในระดับหนึ่งแล้ว แต่ไม่สามารถพัฒนาเป็นระบบได้ด้วยตนเอง นักพัฒนาจะนำเครื่องมือของโปรแกรมมาช่วยพัฒนาต่อให้เกิดเป็นซอฟต์แวร์หรือระบบต่างๆ เพื่อช่วยอำนวยความสะดวก และลดเวลาการทำงานและข้อผิดพลาดได้เป็นอย่างมาก

#### 2.4.1.6 โปรแกรมถ่ายโอนไฟล์ระหว่างเครื่อง FileZilla

โปรแกรมถ่ายโอนไฟล์ระหว่างเครื่อง (FileZilla) เป็นโปรแกรมที่ใช้ติดต่อกับ FTP Server เพื่อดาวน์โหลดหรืออัปโหลดไฟล์ โดยเฉพาะกับเว็บไซต์ ทำให้เหมาะสำหรับนักออกแบบเว็บ นอกจากนี้ FileZilla ยังรองรับการถ่ายไฟล์อย่างปลอดภัยผ่าน SSH (SFTP) อีกด้วย FileZilla รองรับ การกลับมาถ่ายไฟล์ต่อในกรณีที่อัปโหลดหรือดาวน์โหลดล้มเหลว และทำงานได้ดีผ่านไฟร์วอลล์ และพร็อกซี FileZilla มีวิธีใช้งานที่สะดวก แบ่งวินโดว์ออกเป็นสองส่วนคือแสดงฝั่ง local กับ remote จากนั้น สามารถ drag and drop ไฟล์ระหว่างสองฝั่งเพื่ออัปโหลดหรือดาวน์โหลดได้ทันที สามารถ login อย่าง รวดเร็วโดยป้อนข้อมูลของไซต์คือที่อยู่ ผู้ใช้ และรหัสผ่าน ได้บนทูลบาร์เลยทันที นอกจากนี้ FileZilla ยังมี host manager ที่เก็บรายละเอียดของ เซิร์ฟเวอร์ที่ใช้เป็นประจำ เพื่อให้เชื่อมต่อได้อย่างง่ายและ รวดเร็ว FileZilla ทำงานได้เฉพาะระบบ Windows เท่านั้น (“โปรแกรมFileZilla”, 2555: ออนไลน์)

#### 2.4.1.7 โปรแกรม RapidMiner 9.10

Rapidminer คือซอฟต์แวร์ Data Science ใช้สำหรับการเตรียมข้อมูล การเรียนรู้เครื่อง การเรียนรู้ลึก การทำเหมืองข้อมูล และการวิเคราะห์การทำนาย (Predictive analysis) เป็นซอฟต์แวร์ที่ช่วยในการจัดส่งข้อมูล และลดข้อผิดพลาดจนแทบจะไม่จำเป็นต้องเขียนโค้ดเพิ่ม ปัจจุบันได้ก้าวเข้าไปสู่ยุคที่เรียกว่า “Big Data” หรือ “ข้อมูลอภิมหาศาล” เนื่องจากในแต่ละวันมีข้อมูลเกิดขึ้นมากมาย เช่น ข้อมูลสมาชิกของ Facebook ข้อมูลการซื้อขายสินค้าจากในซูเปอร์มาร์เก็ตต่างๆ และเพื่อให้เกิดประโยชน์มากที่สุดเราจำเป็นต้องนำข้อมูลอภิมหาศาลเหล่านี้มาทำการวิเคราะห์ (analyze) ซึ่งเทคนิคหนึ่งที่ได้รับนิยามอย่างสูงในปัจจุบัน คือเทคนิค Data Mining ซึ่งเป็นเทคนิคที่ค้นหาความสัมพันธ์ในข้อมูล

โปรแกรม Rapidminer สามารถใช้ในการทำ Data mining (ขุดข้อมูล) และ Machine learning ซึ่งรวมถึงการโหลดและการแปลงข้อมูล (ETL) การประมวลผลล่วงหน้าและการวาดภาพจากข้อมูล การวิเคราะห์เชิงพยากรณ์และการสร้างแบบจำลองทางสถิติ การประเมินผลและการปรับใช้ต่าง ๆ

#### 2.4.1.8 โปรแกรม Weka 3.9.6

Weka ย่อมาจาก Waikato Environment for Knowledge Analysis : เริ่มพัฒนามาตั้งตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จ ภาพประกอบประเภทฟรีแวร์ ซึ่งโปรแกรม Weka ได้ถูกพัฒนามาจากภาษาจาวาทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง หรือ Machine Learning และ การทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยโมดูลย่อย ๆ สำหรับใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่สามารถใช้ GUI หรือ Graphic User Interface และ ใช้คำสั่งในการให้ซอฟต์แวร์ประมวลผล และสามารถรัน ได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทำงานในด้านการทำ Data Mining ที่รวบรวมแนวคิดอัลกอริทึมมากมาย ซึ่งอัลกอริทึมสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคือจากชุดเครื่องมือที่มีอัลกอริทึมมาให้ หรือเลือกใช้จากอัลกอริทึมที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล ได้แก่ Pre-Processing , Classification , Regression, Clustering , Association rules , Selection และ Visualization ส่วนในเรื่องของการ Import ข้อมูล โดยข้อมูลนั้นจะต้องอยู่ในรูปแบบ ASCII อาจเป็น arff , csv , c45

#### 2.4.1.8.1 ความสามารถของ Weka

สนับสนุนเกี่ยวกับการทำเหมืองข้อมูล มีอัลกอริทึมที่รู้จักกันดีของการทำดาต้าไมน์นิ่งให้เลือกใช้อย่างครบถ้วน และสามารถเขียนฟังก์ชันเพิ่มเข้าไปในโปรแกรมเองได้

#### 2.4.1.8.2 ข้อดีของ Weka

1. สามารถทำงานได้ทุกระบบปฏิบัติการ
2. เชื่อมต่อ SQL Database โดยใช้ Java Database Connectivity
3. เป็นโปรแกรมฟรี
4. มีการเตรียมข้อมูลและเทคนิคในการสร้างแบบจำลองที่ครอบคลุม

#### 2.4.1.8.3 ข้อเสียของ Weka

1. ในฟังก์ชันบางตัวการประมวลผลของข้อมูลไม่ค่อยมีความแม่นยำ
2. การ Import ไฟล์เข้าโปรแกรมมีความยากถ้าข้อมูลมีความผิดพลาด

โปรแกรม Weka เป็นโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล เทคนิคเหมืองข้อมูลคือกระบวนการวิเคราะห์ข้อมูลจากข้อมูลจำนวนมากเพื่อหาความสัมพันธ์ รูปแบบ และแยกประเภทของข้อมูล และ weka นั้นยังได้รวบรวมเทคนิคในการวิเคราะห์ข้อมูลหลาย ๆ เทคนิคเข้าไว้ด้วยกัน แถมยังสามารถวิเคราะห์ข้อมูลเหล่านี้ได้ง่าย ๆ ผ่านทางหน้าจอ GUI (Graphic User Interface) ของ Weka อีกด้วย ซึ่งนับว่าทำให้ผู้ใช้สามารถทำงานได้อย่างสะดวกสบายมากขึ้น

#### 2.4.1.9 โปรแกรม SPSS

SPSS ย่อมาจาก Statistical Package for the Social Science คือ โปรแกรมที่ช่วยวิเคราะห์ข้อมูลทางสถิติให้ออกมาเป็นรูปแบบต่างๆ เช่น ตาราง แผนภูมิ หรือกราฟต่างๆ สามารถบันทึกและสร้างไฟล์ข้อมูลตามความต้องการ ไม่ซับซ้อน เหมาะสำหรับผู้ที่ต้องการประมวลผลข้อมูลที่แม่นยำ เรียกว่า เป็นโปรแกรมที่ช่วยในการเก็บรวบรวมข้อมูลมาวิเคราะห์ให้สามารถอ่านและทำความเข้าใจหรือนำไปต่อยอดได้ง่ายมากขึ้น

##### 2.4.1.9.1 การวิเคราะห์ข้อมูลของโปรแกรม SPSS ชั้นพื้นฐาน

การใช้งานโปรแกรม SPSS สามารถใช้งานได้ตั้งแต่เรื่องที่เกี่ยวข้องกับการใช้ชีวิตประจำวัน อย่างเช่น การทำบัญชีค่าใช้จ่ายของครอบครัว การวิเคราะห์ความพึงพอใจต่อสิ่งต่างๆ และการวิเคราะห์ทัศนคติ เป็นต้น โดยการใช้โปรแกรม SPSS ควรมีความรู้เกี่ยวกับเรื่องพื้นฐานด้านสถิติ อย่างเช่น ค่าความถี่ (Frequency) ค่าร้อยละ (Percentage) ค่าเฉลี่ยเลขคณิต (Mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เป็นต้น

โดยการใช้งานโปรแกรม SPSS สามารถจัดการและนำเสนอข้อมูล เช่น การวิเคราะห์ข้อมูล การนำเสนอข้อมูลในรูปแบบต่างๆ หรือการจัดเก็บข้อมูลในลักษณะต่างๆ อย่างการสร้างแฟ้มข้อมูล การปรับปรุงแก้ไขข้อมูล ทำให้ผู้ใช้งานสะดวกมากขึ้นได้ด้วย

##### 2.4.1.9.2 การใช้ SPSS ในการวิเคราะห์ข้อมูลทางสถิติ

รูปแบบการใช้งานโปรแกรม SPSS ในการวิเคราะห์ข้อมูลทางสถิติจะดูซับซ้อน เพราะเป็นการนำเอาเทคนิคทางสถิติเข้ามารวมใช้ด้วย โดยมีทั้งการวิเคราะห์แบบเบื้องต้น เช่น ค่าเฉลี่ย ร้อยละ ค่าฐานนิยม ค่าแสดงตำแหน่งของข้อมูลอย่างควอไทล์ หรือการนำเสนอในรูปแบบกราฟ การ

แจกแจงความถี่ในรูปแบบของตาราง ฯลฯ ไปจนถึงการวิเคราะห์ข้อมูลด้วยเทคนิคทางสถิติ เช่น การใช้ Factor Analysis, Discriminant Analysis เป็นต้น

#### 2.4.1.9.3 การจัดการข้อมูล และผลลัพธ์

โปรแกรม SPSS ยังมีความสามารถในการจัดเก็บหรือดำเนินการกับข้อมูลในรูปแบบต่างๆ นอกจากการ copy paste หรือ delete แล้วยังสามารถเปลี่ยนรูปข้อมูล เลือกรูปข้อมูล และส่งผลลัพธ์ที่ได้รับการดำเนินการแล้วไปอยู่ในรูปแบบ Text, รูปแบบ Graphic หรือรูปแบบ HTML ได้ด้วย

#### 2.4.1.9.4 การเตรียมข้อมูลให้พร้อมทำการวิเคราะห์

การเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์ คือ การจัดเก็บผลลัพธ์ของแบบสอบถามที่ได้ โดยจะมีการกำหนดชื่อรหัสให้กับข้อคำถามที่มีการจัดเก็บข้อมูลและชื่อของตัวแปร รวมถึงทำการกำหนดมาตรวัดต่างๆ ของตัวแปร เพราะการใช้โปรแกรม SPSS จะต้องมีการใส่ข้อมูลที่ได้มาจากแบบสอบถามลงในตารางของโปรแกรมเพื่อให้มีการวิเคราะห์สรุปกันต่อไป

#### 2.4.1.9.5 การเลือกใช้สถิติที่เหมาะสม

โปรแกรม SPSS สามารถตอบโจทย์การวิจัยได้หลากหลาย รวมไปถึงเรื่องของสถิติด้วย โดยการเลือกใช้สถิติที่เหมาะสมนั้นจะขึ้นอยู่กับที่ตั้งสมมติฐานงานวิจัยว่าเหมาะสมกับสถิติแบบใด

#### 2.4.1.9.6 ประเภทของสถิติ นักคณิตศาสตร์ได้แบ่งออกเป็น 2 สาขาใหญ่ด้วยกัน ได้แก่

สถิติพรรณนา หรือ Descriptive Statistics เป็นการบรรยายลักษณะของข้อมูลที่ได้มีการเก็บรวบรวมมา ซึ่งอาจจะแสดงอยู่ในรูปของค่าเฉลี่ย มัธยฐาน ฐานนิยม ร้อยละหรือส่วนเบี่ยงเบนมาตรฐาน เป็นต้น มักมีการแสดงผลเป็นแบบตารางเพื่อข้อมูลที่แน่ชัด

สถิติเชิงอนุมานหรือ Inferential Statistics เป็นสถิติที่มาจากการวิเคราะห์ข้อมูลที่รวบรวมมาจากกลุ่มตัวอย่าง โดยมักจะมีการใช้ทฤษฎีความน่าจะเป็นเข้ามาร่วมด้วย เช่น การประมาณค่าทางสถิติ การทดสอบข้อมูลทางสถิติ เป็นต้น



#### 2.4.1.9.7 การเลือกตัวแปรที่ต้องการนำมาวิเคราะห์

ในการเก็บข้อมูลงานวิจัยจะต้องมีการใช้ตัวแปรมากมายเพื่อให้ข้อมูลที่ตอบรับกับวัตถุประสงค์ของงานวิจัยที่ตั้งไว้ โดยในการนำข้อมูลไปวิเคราะห์หรือสรุปผลต่อ จะต้องให้ความสำคัญกับการเลือกตัวแปรที่จะเป็นคำตอบ พร้อมกับตรวจเช็คการใช้มาตรวัดเพื่อให้ผลที่สามารถนำไปใช้ได้ถูกต้องด้วย

#### 2.4.1.10 Google Chart

Google Charts เครื่องมือที่ใช้ในการสร้างแผนภูมิรูปภาพ หรือที่เรียกง่าย ๆ ว่ากราฟ (Graphs) หรือชาร์ต (Charts) จุดประสงค์หลักๆมีไว้เพื่อนำเสนอข้อมูลต่าง ๆ ซึ่ง Google Charts นี้สามารถเรียกใช้ในรูปแบบของ Visualization API ได้อีกด้วย

##### 2.4.1.10.1 รูปแบบของ Google Charts

Google Charts เป็นบริการที่ถูกสร้างขึ้นจาก API โดยพัฒนามาจาก JavaScript ทำให้สะดวกในการนำไปใช้กับเอกสารประเภท HTML และง่ายในการนำ Script ไปใช้ในการแสดงผลที่ต้องการนำไปใช้ร่วมกับเว็บไซต์อีกด้วย

#### 2.4.1.11 Microsoft Word

โปรแกรม Microsoft Word เป็นโปรแกรมสำหรับสร้างเอกสารในแบบต่างๆ ตั้งแต่เอกสารแบบง่าย ๆ เช่น รายงาน , ประกาศ , จดหมาย หรือ บทความต่างๆ ไปจนถึงเอกสารที่ซับซ้อน เช่น ใบโฆษณา หรือ บทความประกอบภาพ เป็นต้น โดยผู้ใช้งานสามารถพิมพ์ข้อความ , จัดรูปแบบข้อความ , แทรกข้อความ WordArt , ใส่เลขหน้า , แทรกชาร์ต , แทรกรูปภาพ หรือตาราง แล้วสามารถสั่งพิมพ์ออกจากทางเครื่องพิมพ์เพื่อนำไปใช้งาน ซึ่งโปรแกรม Microsoft Word 2013 นั้นเพียบพร้อมไปด้วยคุณลักษณะพิเศษต่าง ๆ พร้อมด้วย ตัวช่วยมากมายที่ช่วยให้เราสามารถสร้างผลงานระดับคุณภาพได้อย่างง่าย ๆ

## 2.5 วรรณกรรมที่เกี่ยวข้อง

จิรกฤต บุญหมื่นไวย, เจษฎา ตัณฑนุช, เบญจวรรณ โรจนดิษฐ์ (2563) ได้นำเสนองานวิจัยที่ศึกษาความสัมพันธ์ของตัวแปรที่มีผลต่อการขอยกเลิกใช้บริการโดยใช้การถดถอยโลจิสติกส์

จากนั้นนำตัวแปรที่ได้ไปสร้างแบบจำลองด้วยวิธีการต้นไม้ ตัดสินใจการถดถอยโลจิสติกส์และตัว  
จำแนกประเภทแบบเบย์อย่างง่าย จากการศึกษาพบว่าปัจจัยที่มีผลต่อการคงสถานะของลูกค้า  
ได้แก่ คะแนนเครดิต ที่อยู่อาศัย เพศ อายุ ยอดเงินคงเหลือในบัญชี จำนวนผลิตภัณฑ์และการเป็น  
สมาชิกของธนาคาร

Tony Yiu (2560) ได้อธิบายหลักการของ Random Forest ไว้ดังนี้ Random Forest สร้าง  
แบบจำลองจาก Decision Tree หลาย ๆ แบบจำลองแต่ละแบบจำลองย่อยจะถูกสอนจากข้อมูลชุด  
เดียวกันแต่ใช้ข้อมูลย่อยที่แตกต่างกัน โดยข้อมูลย่อยที่ถูกใช้ในการสอนแบบจำลองได้มาจากการ  
สุ่ม ทั้งคำตอบที่ได้จากการทำนายของ Random Forest จะเกิดจากการเฉลี่ยหรือการโหวตเลือก  
ผลลัพธ์จาก Decision Tree แล้วแต่ว่าเป็นการทำนายความน่าจะเป็นหรือทำนายประเภทข้อมูล

อสมมา กุลวานิชไชยนันท์ (2561) ในปัจจุบันหลาย ๆ ธุรกิจ เช่น บริษัทให้บริการสัญญา  
โทรศัพท์ ธุรกิจบัตรเครดิต บริษัทประกันภัยและกลุ่มค้าปลีก เป็นต้น มีความต้องการจะรักษา  
ลูกค้ากลุ่มเดิมไว้ ไม่ให้ยกเลิกใช้บริการ ซึ่งการทำนายลักษณะลูกค้าที่ de ลังจะยกเลิกบริการ  
(Churn Prediction) สามารถช่วยให้ธุรกิจตรวจสอบได้ล่วงหน้าเมื่อลูกค้ามีพฤติกรรมที่เปลี่ยนแปลง  
ไปและมีแนวโน้มที่ de ลังจะยกเลิกใช้บริการ ซึ่งหากบริษัทหรือบุคคลที่เกี่ยวข้องสามารถเข้าใจ  
กลไก ของการทำนายประเภทนี้ และนำผลลัพธ์ที่ได้จากตัวแบบไปใช้งาน เช่น มีการส่งโปรโมชั่น  
พิเศษเพื่อจูงใจลูกค้า มีการติดตามความพึงพอใจของลูกค้าที่พบว่ามีแนวโน้มที่จะยกเลิกใช้บริการ  
เพื่อกระชับความสัมพันธ์ ก็อาจจะช่วยลดความเสี่ยงจากการยกเลิกใช้บริการของลูกค้าได้ ดังนั้น  
สิ่งที่สำคัญก็คือทำอย่างไรจึงจะสามารถนำข้อมูลของลูกค้าที่มีอยู่มาสร้างเป็นตัวแบบเพื่อใช้ในการ  
วิเคราะห์ได้อย่างแม่นยำมากที่สุด ซึ่งการวิเคราะห์หาความสัมพันธ์ของตัวแปรที่จะนำมาพิจารณา  
ในการสร้างตัวแบบถือว่าเป็นขั้นตอนที่สำคัญมากอีกขั้นตอนหนึ่ง เพราะถ้าวิเคราะห์ไม่รอบคอบ  
อาจจะทำให้ตัวแบบที่ได้ไม่สามารถเอาไปใช้งานได้จริง

วทัญญูตา นิลภาตระกูล และ ชุตติมา เปี้ยวไข่มุก (2562) จากการศึกษาเรื่อง การศึกษา  
ปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพ ผลการศึกษาพบว่า  
งานวิจัยนี้การศึกษาเกี่ยวกับเทคนิคการทำเหมืองข้อมูล เพื่อลดปัญหาอัตราการลาออกของ  
พนักงานและลดค่าใช้จ่ายที่สูงขึ้นจากการสรรหาพนักงานใหม่ในบริษัทประกันภัยแห่งหนึ่ง ชุด

ข้อมูลที่ใช้วิเคราะห์เป็นข้อมูลของพนักงานลาออกและพนักงานที่ยังทำงานอยู่ ตั้งแต่ปีพ.ศ. 2556-2560 จำนวน 1,000 รายการ 11 แอตทริบิวต์ เทคนิคกฎความสัมพันธ์ถูกนำมาใช้หารูปแบบของความสัมพันธ์การลาออกของพนักงาน และสร้างตัวแบบพยากรณ์การลาออกของพนักงาน โดยใช้เทคนิคการวิเคราะห์ 5 เทคนิค คือ ต้นไม้ตัดสินใจ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคโครงข่ายประสาทเทียม เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้สุดและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์เหล่านั้น การทดสอบประสิทธิภาพของตัวแบบใช้วิธีไขว้ข้อมูล

พฤตพิงศ์ เพ็งศิริ (2559) จากการวิจัยเรื่อง การพยากรณ์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ผลการศึกษาพบว่า การพยากรณ์โดยการประยุกต์ใช้กระบวนการตัดสินใจด้วยเทคนิคต้นไม้ตัดสินใจเป็นเทคนิคอย่างหนึ่งในการทำเหมืองข้อมูล ซึ่งอาศัยความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาเป็นการบ่งชี้ถึงระดับผลการเรียนของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ซึ่งผลลัพธ์ที่ได้นั้นจะส่งผลทำให้เอื้อประโยชน์กับแนวทางในการปรับปรุงบริหารจัดการและส่งเสริมการวางแผนของหลักสูตรการศึกษาที่เกี่ยวข้อง โดยนำปัจจัยข้อมูลของนักศึกษา ได้แก่ ข้อมูลทั่วไป ข้อมูลส่วนตัว ข้อมูลการศึกษา ข้อมูลระดับผลการเรียนของนักศึกษากลุ่มที่ศึกษา มาหาความสัมพันธ์กับผลการเรียนที่ได้ โดยวิเคราะห์จากระดับผลการเรียน (เกรดเฉลี่ย) ของการเรียนในชั้นปีที่ 1 ผลการทดลองพบว่าปัจจัยข้อมูลของนักศึกษาที่เหมาะสมในการเรียน มีทั้งหมด 7 ตัวแปร จากข้อมูลนำเข้าทั้งหมด 12 ตัวแปร ทั้งนี้ 7 ตัวแปรมาจากต้นไม้ตัดสินใจที่ได้มาสามารถสรุปเลือกเฉพาะกิ่งที่มีผลมากที่สุด โดยวัดค่าความแม่นยำ (Accuracy) ได้ค่าสูงถึง 84.78% ซึ่งถือว่าอยู่ในเกณฑ์สูง ดังนั้นตัวแปรที่น่าจะเป็นปัจจัยที่ส่งผลกับระดับผลการเรียนที่จบการศึกษาสูงสุดคือ ความสม่ำเสมอการเข้าเรียนในการเรียน

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรม SPSS สำหรับการวิเคราะห์การถดถอยพหุคูณ ใช้โปรแกรม RapidMiner Studio 9.5.1 และ โปรแกรม Weka 3.9.6 ในการสร้างโมเดล Decision Tree เพื่อทำ

การเปรียบเทียบและเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน Web Browser ที่เป็นที่ยอมรับในยุคอินเทอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา