

## บทที่ 2

### แนวคิด ทฤษฎี และวรรณกรรมที่เกี่ยวข้อง

โครงการเรื่อง ปัจจัยที่มีอิทธิพลต่อความสำเร็จของนักศึกษาโดยใช้วิธีการเหมืองข้อมูล ในบทนี้เป็น การนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี และวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ปัจจัยที่มีอิทธิพลต่อความสำเร็จของนักศึกษาโดยใช้วิธีการเหมืองข้อมูลและการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง กับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

#### 2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

#### 2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดกลุ่มแบบ Clustering

2.2.2 ทฤษฎีเกี่ยวกับระบบสนับสนุนการตัดสินใจ Decision tree

2.2.3 ทฤษฎีเกี่ยวกับการ visualization

2.2.4 ทฤษฎีเกี่ยวกับกับชุดคำสั่งชุด HTML

2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่งชุด CSS

#### 2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม WEKA

2.3.2 โปรแกรม KNIME

2.3.3 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

#### 2.4 วรรณกรรมที่เกี่ยวข้อง

#### 2.5 บทสรุป

## 2.1 แนวคิด

### 2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเชิงคุณภาพนับเป็นขั้นตอนที่สำคัญในกระบวนการวิจัยวิธีการหลักที่ใช้ในการวิเคราะห์ข้อมูลเชิงคุณภาพ ประกอบด้วย การจำแนกและจัดระบบข้อมูลและการหาความสัมพันธ์ของข้อมูล มีจุดมุ่งหมายที่จะแยกแยะและอธิบายองค์ประกอบความหมาย และความสัมพันธ์ของปรากฏการณ์ ภายใต้เงื่อนไขและสภาพแวดล้อมทางสังคมและวัฒนธรรม เป็นวิธีการสร้างข้อสรุปจากการศึกษาจากข้อมูลจำนวนหนึ่ง ซึ่งมักไม่ใช่สถิติในการวิเคราะห์หรือถ้าใช้สถิติก็ไม่ได้ถือว่าสถิติเป็นวิธีการวิเคราะห์หลัก แต่จะถือเป็นข้อมูลเสริมในการวิเคราะห์ข้อมูลเชิงคุณภาพนั้น ผู้วิเคราะห์ข้อมูลนับเป็นผู้ที่มีบทบาทสำคัญยิ่ง และควรมีความรอบรู้ในเรื่องแนวคิด ทฤษฎีอย่างกว้างขวาง มีความเป็นสหวิทยาการอยู่ในตัวเอง มีความสามารถทางภาษาสามารถเชื่อมโยงข้อความ และสร้างข้อสรุปเป็นกรอบแนวคิดและตีความหมายของข้อมูลได้หลายแบบ วิธีการวิเคราะห์ข้อมูลในการวิจัยเชิงคุณภาพ อาจใช้การวิเคราะห์ข้อมูลแบบสร้างข้อสรุป หรือการวิเคราะห์เนื้อหา สามารถพิจารณาได้จากความเชื่อถือได้ การถ่ายโอนผลการวิจัย และการพึงพากับเกณฑ์อื่น และการยืนยันผลความหมายของคำศัพท์ที่สำคัญเกี่ยวกับการวิเคราะห์งานวิจัยเชิงคุณภาพ

### 2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

ความหมายของการประยุกต์ใช้ข้อมูลทางสถิติ จากเอกสารประกอบการสอนการวิจัย ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า คำว่าสถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistics มีรากศัพท์มาจาก Stat หมายถึง ข้อมูลหรือสารสนเทศ ซึ่งจะอำนวยความสะดวกต่อการบริหารประเทศในด้านต่าง ๆ เช่น การทำสำมะโนครัวเพื่อจะทราบจำนวนพลเมืองในประเทศทั้งหมดในสมัยต่อมา คำว่า สถิติ ได้หมายถึงตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวม เช่น จำนวนผู้ประสบอุบัติเหตุบนท้องถนน อัตราการเกิดของเด็กทารก ปริมาณน้ำฝนในแต่ละปี เป็นต้น สถิติในความหมายที่กล่าวมานี้ เรียกอีกอย่างหนึ่งว่า ข้อมูลทางสถิติ (Statistical data)

1.) ค่ากึ่งกลาง (Median) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางวิธีหนึ่งที่ใช้การเรียงค่า การสะท้อน

ของจุดภาพจากค่าน้อยที่สุดไปหาค่ามากที่สุด โดยค่ากึ่งกลางเป็นค่าที่อยู่ใน ตำแหน่งกึ่งกลางของข้อมูลทั้งหมด ค่ากึ่งกลางจึงเป็นตัวแทนค่าการสะท้อนของจำนวนจุดภาพ ทั้งหมดในช่วงคลื่นหนึ่งๆ

ที่แสดงให้ทราบว่า มีจำนวนจุดภาพที่มีค่าการสะท้อนมากกว่าและน้อยกว่าค่ากึ่งกลางอยู่ประมาณร้อยละ 50

2.) ค่าฐานนิยม (Mode) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางอีกวิธีหนึ่ง โดยดูจาก จำนวนความถี่ของค่าการสะท้อนซึ่งมีความถี่สูงที่สุด นิยมนำมาใช้กับข้อมูลที่เป็นนามบัญญัติเช่น ค่าของประเภทข้อมูลหลังจากการจำแนกประเภทแล้ว ถือเป็นค่าการสะท้อนที่แสดงการใช้ ที่ดินประเภทต่าง ๆ ไม่ใช่ค่าการสะท้อนของวัตถุอีกต่อไป

3.) สถิติพรรณนา (Descriptive statistics) คือสถิติที่บรรยายถึงลักษณะของข้อมูลเฉพาะกลุ่มนั้นๆ โดยไม่สรุปอ้างอิงไปยังประชากรกลุ่มอื่นๆ สถิติประเภทนี้นิยมศึกษาในกลุ่มเล็กหรือกลุ่มใหญ่ก็ได้

สถิติประเภทนี้เป็นสถิติที่บรรยายลักษณะของข้อมูล เช่น ค่าเฉลี่ย ร้อยละ มัธยฐาน พิสัย ส่วนเบี่ยงเบนมาตรฐาน และการวัดความสัมพันธ์ต่าง ๆ เช่น สหสัมพันธ์

4.) สถิติอ้างอิงหรือสถิติอนุมาน (Inferential or inductive statistics) คือ สถิติที่นำค่าสถิติพรรณนามาสรุปอ้างอิงไปยังประชากรหรือเป็นสถิติที่ศึกษาจากกลุ่มตัวอย่างแล้วนำผลอ้างอิงไปยังกลุ่มประชากร เช่น การศึกษาความคิดเห็นต่อการเลือกตั้งของประชาชนไม่จำเป็นต้องศึกษาจากประชาชนทุกคน แต่สามารถเลือกศึกษาจากประชาชนบางกลุ่มซึ่งจะเป็นตัวแทนของประชาชนทั้งหมด แล้วจึงสรุปว่าประชาชนมีความคิดเห็นอย่างไรต่อการเลือกตั้งได้ดังนั้นในการใช้สถิติอ้างอิงนี้จึงจำเป็นต้องเลือกกลุ่มตัวอย่างที่เหมาะสม สถิติอ้างอิงนี้ก่อนนำไปอ้างอิงกลุ่มประชากรต้องมีการทดสอบทางสถิติก่อนทุกครั้งจึงสามารถอ้างอิงประชากรได้อย่างมีประสิทธิภาพ

## 2.2 ทฤษฎี

### 2.2.1 ทฤษฎีเกี่ยวกับการจัดกลุ่มแบบ Clustering

#### 1. ความหมายของการวิเคราะห์กลุ่ม (Cluster Analysis)

การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคการแบ่งกลุ่มหน่วยข้อมูลหรือเป็นการแบ่งคน สัตว์ สิ่งของ องค์กร ฯลฯ ออกเป็นกลุ่มย่อยอย่างน้อย 2 กลุ่ม โดยมีหลักเกณฑ์ในการแบ่งดังนี้ “ให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่สนใจเหมือนกันหรือคล้ายกัน แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีลักษณะที่สนใจต่างกัน” คำว่าลักษณะที่สนใจอาจจะมีหลาย ๆ ตัว

แปร เช่น ถ้าสนใจความคิดเห็นทางการเมือง จะมีคำถามหลาย ๆ คำถามทางการเมือง และจะนำคำตอบเหล่านั้นมาแบ่งกลุ่ม (กัลยา วาณิชย์บัญชา, 2552 )

การจัด Case (หมายถึง คน สัตว์ สิ่งของ หรือ องค์กร ฯลฯ) หรือเป็นการจัดตัวแปรออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป Case ที่อยู่ในกลุ่มเดียวกันจะมีลักษณะที่เหมือนกันหรือคล้ายกัน ส่วน Case ที่อยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน ตัวแปรที่อยู่ในกลุ่มเดียวกัน จะมีความสัมพันธ์กันมากกว่าตัวแปรที่อยู่ต่างกลุ่มกัน ตัวแปรที่อยู่ต่างกลุ่มกัน จะมีความสัมพันธ์กันน้อยหรือไม่มีความสัมพันธ์กันเลย

2.วัตถุประสงค์ของการแบ่งกลุ่มข้อมูลด้วยเทคนิคการวิเคราะห์กลุ่ม (Cluster Analysis)

การวิเคราะห์กลุ่มเป็นเทคนิคที่ใช้ในการจัดกลุ่มโดยไม่ทราบมาก่อนว่าควรจะมีกี่กลุ่มแต่จะแบ่งตามค่าของตัวแปรที่นำมาใช้ในการแบ่ง โดยให้หน่วยที่อยู่ในกลุ่มเดียวกัน มีความคล้ายกันในตัวแปรที่ศึกษา แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีความต่างกัน

## 2.2.2 ทฤษฎีเกี่ยวกับระบบสนับสนุนการตัดสินใจ Decision tree

เทคนิคต้นไม้การตัดสินใจ คือ แบบจำลองทางคณิตศาสตร์ เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอนสามารถสร้างแบบจำลองการจัดการหมวดหมู่ได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ล่วงหน้า โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมา จัดหมวดหมู่ได้อีกด้วย (พยุณ,2548)

โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้ว ผลลัพธ์” เช่น

“If Income = High and Married = No THEN Risk = Poor”

“If Income = High and Married = Yes THEN Risk = Rich”

ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย(จิตฎุ,2550)

1) โหนด (Node) คือ คุณสมบัติต่างๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใดซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก(Root Node)

2) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจำนวนเท่ากับคุณสมบัติของโหนด

3) ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ

การสร้างต้นไม้ตัดสินใจ หลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจ เป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) คือเริ่มจากการสร้างรากของต้นไม้ก่อนแล้วจึงแตกกิ่ง ไปจนถึงใบ โดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจ ได้ดังนี้ (ฮานและแคมเบอร์, 2001)

- 1) ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก
- 2) ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น
- 3) ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่ จะต้องวัดค่าเกน (Gain) ของแต่ละแอททริบิวต์เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่างๆ ได้ดีที่สุด โดยแอททริบิวต์ที่มีค่าเกนมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือแอททริบิวต์ ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้
- 4) กิ่งของต้นไม้ ถูกสร้างขึ้นจากค่าต่างๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่างๆ ที่สร้างขึ้น
- 5) ทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกนมากที่สุดสำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์มาสร้างเป็นโหนดตัดสินใจต่อไปโดยแอททริบิวต์ถูกเลือกเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อไป
- 6) ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อยๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้ เป็นจริง

### 2.2.3 ทฤษฎีเกี่ยวกับเกี่ยวกับการ visualization

Power BI Desktop คือโปรแกรมที่ใช้ช่วยวิเคราะห์ สรุปผลข้อมูลจำนวนไม่จำกัด จากหลายๆ แหล่งข้อมูล ไม่ว่าจะเป็น Excel File, Microsoft Access Database, SQL Server, Oracle, Power Platform เป็นต้น ได้อย่างรวดเร็ว โดยผู้ใช้ไม่ต้องมีความรู้เชิงเทคนิคมากนัก พร้อมสามารถแสดงผลได้ทั้งรูปแบบ ตาราง สรุปผลด้วย Visualization กราฟในรูปแบบต่าง ๆ และยังติดตั้ง

เพิ่มเติมได้จาก Marketplace และไฮไลต์ก็คือ แสดงผลแบบเป็นแผนที่ได้อีกด้วย โดยสามารถแสดงผลผ่านเว็บไซต์ และอุปกรณ์ Mobile และ Tablet ได้อีกด้วยทำให้เรา ได้ข้อมูลที่ถูกต้องและรวดเร็ว พร้อมในการกำหนดกลยุทธ์ และตัดสินใจได้อย่างถูกต้องและแม่นยำในทางธุรกิจ

ข้อมูลที่เราสามารถนำเข้ามาใช้งานเพื่อการวิเคราะห์ข้อมูลมีหลายแหล่งดังนี้

1. นำเข้าจากไฟล์ข้อมูล ได้แก่ Excel, CSV, Text, XML, JSON เป็นต้น
2. นำเข้าจากฐานข้อมูล ได้แก่ SQL Server, Access Database, SQL Server Analysis Services Database, Oracle, MySQL, DB2, Sybase, PostgreSQL เป็นต้น
3. Power Platform ได้แก่ Power BI datasets, Power BI dataflows, Common Data Service (Legacy), Dataverse, Power Platform dataflows
4. Azure ได้แก่ Microsoft Azure SQL Database, Azure SQL Data Warehouse, Azure Marketplace, Azure Blob Storage, Azure Table Storage เป็นต้น
5. Online Services ได้แก่ SharePoint Online List, Dynamics 365 (online), Google Analytics, Google Sheets เป็นต้น
6. อื่น ๆ ได้แก่ Web, Active Directory, OLE DB เป็นต้น

Tableau Public เป็นซอฟต์แวร์ที่สามารถนำข้อมูลจำนวนมาก หรือมีความหลากหลายจากแหล่งข้อมูลในองค์กรมาทำการวิเคราะห์เพื่อเป็นข้อมูลเชิงธุรกิจช่วยให้ผู้บริหารมีข้อมูลเชิงลึกเพิ่มความสามารถในการตัดสินใจอย่างรวดเร็วและชาญฉลาด ในรูปแบบของ Data Visualization ซึ่งเป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ไม่ว่าจะเป็นตัวเลข แผนภูมิ กราฟ และอื่น ๆ อีกมากมาย คำว่า Data คือ ข้อมูล ส่วน Visualization คือการมองเห็นเมื่อนำมารวมกันแล้ว (Data Visualization) หมายถึง ข้อมูลที่มองเห็นได้ด้วยตา

โปรแกรม Tableau แบ่งเป็น 3 ประเภทหลักๆ

Tableau Public เป็นเวอร์ชันที่เปิดให้ใช้งานได้โดยไม่มีค่าใช้จ่าย ดังนั้นคุณสมบัติหลายอย่างจะถูกลดทอนลงไป เช่น ประเภทของแหล่งข้อมูล หรือ ข้อมูล ที่จะนำมาใช้ใน

โปรแกรมจะถูกจำกัดลง และจะมีข้อจำกัดในเรื่องบันทึกไฟล์ที่สร้างขึ้นจากโปรแกรมคือ ไม่สามารถที่จะบันทึกไฟล์ไว้ในเครื่องที่ใช้งานได้แต่สามารถที่จะบันทึกไฟล์งานไว้ที่ Tableau Public Server ได้

Tableau Desktop เป็นเวอร์ชันที่ต้องการชำระเงินในการใช้งาน แต่สามารถที่จะทดลองใช้งานได้ (Free Trial 14 days) ซึ่งคุณสมบัติต่าง ๆ ของโปรแกรมจะสามารถใช้งานได้เกือบครบทุกฟังก์ชัน (ในกรณีทดลองใช้งาน) และสามารถนำแหล่งข้อมูลมาใช้กับโปรแกรมได้อย่างหลากหลายและครอบคลุมในการใช้งาน สามารถที่จะบันทึกข้อมูลต่างๆ ไว้ภายในเครื่องที่ใช้งาน หรือสามารถ publish ไว้ใน Tableau Server, Tableau Public Server หรือ Tableau Online และสามารถให้บุคคลอื่นๆ สามารถเปิดดูได้ผ่าน Tableau Reader ได้

Tableau Public Server เป็นเวอร์ชันที่ทาง Tableau จัดเตรียมไว้สำหรับแสดงข้อมูลที่ได้จากการ publish จากโปรแกรม Tableau Public เพื่อให้ผู้ใช้งานคนอื่นสามารถเข้ามาชมข้อมูลใน Workbooks ที่อัปโหลดขึ้นมา โดยใช้งานผ่านทาง Web browser

## 2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่งชุด HTML

HTML ย่อมาจาก HyperText Markup Language เป็นภาษาคอมพิวเตอร์ที่ใช้สร้างหน้าเว็บ(WebPage) ในรูปแบบของไฟล์HTML (คือไฟล์ที่มีนามสกุลเป็น.htm หรือ.html) ซึ่งมีเว็บเบราว์เซอร์(WebBrowser) เป็นโปรแกรมที่ใช้แปลงไฟล์ HTML เพื่อแสดงผลในรูปแบบของหน้าเว็บ

ไฟล์ HTML เป็นไฟล์รหัสแอสกี (ASCII) ถูกบันทึกในรูปแบบของไฟล์เอกสาร(Text File) ที่สามารถถูกสร้างจากโปรแกรมสร้างไฟล์ข้อความ (Text Editor) เช่น Notepad หรือ Word Processingต่างๆ ไป ซึ่งลักษณะของไฟล์HTML ประกอบไปด้วยแท็ก (Tag) ต่างๆ ที่เป็นคำสั่งของ HTML ซึ่งแท็กจะอยู่ภายในเครื่องหมาย < และ >

แท็กใน HTML แบ่งเป็น 2 ประเภทคือคอนเทนเนอร์แท็ก (Container Tag) และแท็กเปล่า(Empty Tag) โดยที่คอนเทนเนอร์แท็ก ประกอบไปด้วยแท็กเปิด และแท็กปิด โดยที่แท็กปิดจะมีเครื่องหมาย / นำหน้าแท็ก เช่น<H1>. . ./H1> ส่วนแท็กเปล่าจะมีแท็กเปิดอย่างเดียว เช่น<HR> ซึ่งแท็กจะถูกเขียนด้วยตัวอักษรพิมพ์ใหญ่หรือพิมพ์เล็กก็ได้จะไม่มีผลต่อการแสดงผลของเว็บเบราว์เซอร์ เช่น <BR>, <br>, <Br> หรือ <br> เว็บเบราว์เซอร์จะแปลความหมายเหมือนกัน

โครงสร้างไฟล์ HTML แบ่งออกเป็น 2 ส่วน คือ ส่วนหัวเรื่อง (Head Section) และ ส่วนเนื้อหา (Body Section) โดยจะมีแท็ก <HTML> และ </HTML> เป็นตัวกำหนดขอบเขตไฟล์ซึ่ง ส่วนหัวเรื่อง มีไว้กำหนดข้อมูลเฉพาะของหน้าเว็บ เช่น ชื่อเรื่องของเว็บภายในแท็ก <HEAD> และ </HEAD> และสำหรับส่วนเนื้อหามีไว้กำหนดรายละเอียดต่างๆ ที่ต้องการแสดงบนหน้าเว็บ เช่น ข้อความ และรูปภาพภายในแท็ก <BODY> และ </BODY>

หน้าเว็บ คือ หน้าเอกสารที่ถูกแสดงโดย เว็บเบราว์เซอร์เพื่อแสดงข้อมูลต่างๆ ที่เป็นข้อความ รูปภาพ และสื่อผสมต่างๆซึ่งเนื้อหาของหน้าเว็บเป็นอย่างไร ขึ้นอยู่กับวัตถุประสงค์ของเจ้าของหน้าเว็บ ไม่ว่าจะเป็นเนื้อหาเกี่ยวกับการศึกษาธุรกิจ หรือความบันเทิง เป็นต้น

จากที่ได้กล่าวมาแล้วจะได้ว่า หน้าเว็บ คือ ไฟล์ HTML ซึ่งมีโครงสร้างไฟล์ตามภาพที่ 2 โดยหัวใจหลักของการเขียนไฟล์ HTML คือ แท็กคำสั่ง ที่ใช้ในการกำหนดรูปแบบต่างๆ ที่ต้องการแสดงในหน้าเว็บ

### บทสรุป

HTML เป็นภาษาคอมพิวเตอร์ที่มีบทบาทอย่างมากในปัจจุบัน เนื่องจากเป็นรูปแบบของภาษาสำหรับการสร้างหน้าเว็บที่มีลักษณะเป็นเอกสารแบบไฮเปอร์เท็กซ์ซึ่งมีคุณสมบัติที่สามารถเชื่อมโยงข้อมูลต่างๆ ไปยัง หน้าเว็บอื่นๆ ตามต้องการได้ทำให้การเชื่อมโยงข้อมูลในหน้าเว็บต่างๆ ในอินเทอร์เน็ตเป็นไปอย่างสะดวก และรวดเร็ว

ในปัจจุบัน แม้ว่าจะมีโปรแกรมประยุกต์สำหรับหน้าเว็บต่างๆ ออกมามากมาย เพื่อสร้างความสะดวกในการสร้างหน้าเว็บให้กับผู้ใช้ด้วยการแปลงสิ่งที่ผู้ใช้ออกแบบหน้าเว็บ ให้เป็นไฟล์ HTML โดยที่ผู้สร้างหน้าเว็บ ไม่จำเป็นต้องมีความรู้เกี่ยวกับ HTML แต่โปรแกรมประยุกต์ดังกล่าวมักจะแปลงสิ่งที่ออกแบบให้เป็นคำสั่ง HTML ที่มากเกินไปจนทำให้ไฟล์ HTML มีขนาดใหญ่และมีผลทำให้เวลาในการแสดงผลหน้าเว็บในเว็บเบราว์เซอร์นานขึ้น

ดังนั้น การเรียนรู้ HTML ยังเป็นสิ่งจำเป็น นอกจากจะใช้สร้างหน้าเว็บแล้ว ยังสามารถใช้ในการแก้ไขเนื้อหาหน้าเว็บที่มีอยู่แล้ว โดยไม่ต้องติดตั้งโปรแกรมประยุกต์ใดรวมทั้งสามารถเพิ่มสีสัน และความน่าสนใจให้กับหน้าเว็บ ด้วยการแทรกสคริปต์ต่างๆลงในไฟล์ HTML ได้อีกด้วย

### 2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่งชุด CSS



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

### CSS มีประโยชน์อย่างไร

ภาษา CSS (Cascading Style Sheets) มีประโยชน์หลายอย่างเลยทีเดียว ซึ่งทำให้การพัฒนาเว็บเพจด้วยภาษา HTML เป็นเรื่องที่ย่างมากขึ้น

1. ภาษา CSS จะช่วยในการจัดรูปแบบแสดงผลให้กับภาษา HTML ซึ่งจะช่วยลดการใช้ภาษา HTML ให้น้อยลง โดยเหลือเพียงแต่ส่วนที่เป็นเอกสารที่เป็นภาษา HTML เท่านั้นทำให้มีการแก้ไขและทำความเข้าใจได้ง่ายขึ้น
2. ทำให้ขนาดไฟล์ HTML น้อยลงเนื่องจาก ภาษา CSS จะช่วยลดการใช้ภาษา HTML ลงทำให้ขนาดไฟล์นั้นก็เล็กลงไปด้วยเช่นกัน
3. ภาษา CSS เป็นภาษา Style Sheets โดย Style Sheets ชุดเดียวสามารถใช้กำหนดรูปแบบการแสดงผลให้เอกสาร HTML ทั้งหมด หรือทุกหน้ามีผลเหมือนกันได้ จึงทำให้เวลาที่มีการแก้ไขก็จะแก้ไขได้ง่ายขึ้นเพียงแก้ไข Style Sheets ที่ใช้งานเพียงชุดเดียวเท่านั้น

4. ทำให้เว็บไซต์มีมาตรฐานเพราะการใช้งาน CSS นั้นจะทำให้การแสดงผลในสื่อต่าง ๆ ถูกปรับเปลี่ยนไปได้อย่างเหมาะสม เช่น การแสดงผลบนหน้าจอ และการแสดงผลในมือถือ
5. CSS สามารถที่จะใช้งานได้หลากหลาย เว็บเบราว์เซอร์ ทำให้การใช้งานนั้นสะดวกมากยิ่งขึ้น
6. CSS สามารถกำหนดแยกไว้ต่างหากจากไฟล์เอกสาร HTML และสามารถนำมาใช้รวม กับเอกสารหลายไฟล์ได้ การแก้ไขก็แก้เพียง จุดเดียวก็มีผลกับเอกสารทั้งหมด

CSS กับ HTML นั้นทำหน้าที่คนละอย่างกัน โดย HTML จะทำหน้าที่ในการวางโครงสร้างเอกสารอย่างเป็นรูปแบบ ถูกต้อง เข้าใจง่าย ไม่เกี่ยวข้องกับการแสดงผล ส่วน CSS จะทำหน้าที่ในการตกแต่งเอกสารให้สวยงาม เรียกได้ว่า HTML คือส่วน coding ส่วน CSS คือส่วน design

## 2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

### 2.3.1 โปรแกรม WEKA

โปรแกรม WEKA (Waikato Environment for Knowledge Analysis) เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จรูป อยู่ภายใต้การควบคุมของ GPL License โปรแกรม WEKA ได้ถูกพัฒนาจากภาษาจาวาทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และการทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยโมดูลย่อยๆ สำหรับใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่สามารถใช้ Graphic User Interface (GUI) และใช้คำสั่งในการให้ซอฟต์แวร์ประมวลผล และสามารถรัน run) ได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้เป็นเครื่องมือที่ใช้ทำงานในด้านการทำดาตาไมนนิ่งที่รวบรวมแนวคิดอัลกอริทึมมากมาย ซึ่งอัลกอริทึมสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคือจากชุดเครื่องมือที่มีอัลกอริทึมมาให้ หรือเลือกใช้จากอัลกอริทึมที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล ได้แก่ Pre-Processing, Classification, Regression, Clustering, Association rules, Selection และ Visualization

ข้อดีโปรแกรม WEKA

- เป็นซอฟต์แวร์ที่เป็นฟรีแวร์
- สามารถทำงานได้ทุก OS
- เชื่อมต่อ SQL Database โดยใช้ Java Database Connectivity
- มีลักษณะที่ง่ายต่อการใช้งาน
- สนับสนุนเกี่ยวกับการทำเหมืองข้อมูล (Data Mining)

#### ข้อเสียโปรแกรม WEKA

- หาโหลดใช้งานได้ยาก
- ฟังก์ชันอาจจะยังไม่ทันสมัย

การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบแนวทางและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ความรู้ที่ได้จากการทำเหมืองข้อมูลมีหลายรูปแบบ ได้แก่

#### กฎความสัมพันธ์ (Association rule)

แสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุ ที่เกิดขึ้นพร้อมกัน ตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยง เช่น การวิเคราะห์ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (POS) หรือร้านค้าออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าคนที่ซื้อเทปวิดีโอมักจะซื้อเทปการ์ดด้วย ร้านค้าก็อาจจะจัดร้านให้สินค้าสองอย่างอยู่ใกล้กัน เพื่อเพิ่มยอดขาย หรืออาจจะพบว่าหลังจากคนซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็สามารถนำความรู้นี้ไปแนะนำผู้ที่กำลังจะซื้อหนังสือ ก ได้

#### การจำแนกประเภทข้อมูล (Data classification)

หากกฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อคุณสมบัตินี้ของผู้ที่จะก่อกองหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้

### การแบ่งกลุ่มข้อมูล (Data clustering)

แบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกันจินตทัศน์(Visualization)

สร้างภาพคอมพิวเตอร์กราฟิกที่สามารถนำเสนอข้อมูลมากมายอย่างครบถ้วน แทนการใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ่อนเร้นเมื่อดูข้อมูลชุดนั้นด้วยจินตทัศน์

### ขั้นตอนการทำเหมืองข้อมูล

1. ทำความเข้าใจปัญหา
2. ทำความเข้าใจข้อมูล
3. เตรียมข้อมูล
4. สร้างแบบจำลอง
5. ประเมิน
6. นำไปใช้งาน

### ประโยชน์จากการทำเหมืองข้อมูล

การทำเหมืองข้อมูล จำเป็นต้องอาศัยบุคลากรจากหลายฝ่าย และต้องอาศัยความรู้จำนวนมาก ถึงจะได้รับประโยชน์อย่างแท้จริง เพราะสิ่งที่ได้จากขั้นตอนวิธีเป็นเพียงตัวเลขและข้อมูล ที่อาจจะนำไปใช้ประโยชน์ได้หรือใช้ประโยชน์อะไรไม่ได้เลยก็เป็นได้ ผู้ที่ศึกษาการทำเหมืองข้อมูลจึงควรมีความรู้รอบด้านและต้องติดต่อกับทุก ๆ ฝ่าย เพื่อให้เข้าใจถึงขอบเขตของปัญหาโดยแท้จริงก่อน เพื่อให้การทำเหมืองข้อมูลเกิดประโยชน์อย่างแท้จริง

### 2.3.2 โปรแกรม KNIME

KNIME (Konstanz Information Miner), เป็นแพลตฟอร์มการวิเคราะห์ การรายงาน และการรวมข้อมูลแบบโอเพนซอร์สฟรี KNIME ผสานรวมส่วนประกอบต่างๆ สำหรับการเรียนรู้ของเครื่องและการทำเหมืองข้อมูลผ่านแนวคิด "Lego of Analytics" ของการวางท่อข้อมูลแบบแยกส่วน ส่วนต่อประสานกราฟิกกับผู้ใช้และการใช้ JDBC ช่วยให้สามารถประกอบโหนดที่ผสมผสานแหล่งข้อมูลต่างๆ ได้ รวมถึงการประมวลผลล่วงหน้า ( ETL: Extraction, Transformation, Loading ) สำหรับการสร้างแบบจำลอง การวิเคราะห์ข้อมูล และการแสดงภาพโดยไม่ต้องใช้โปรแกรมหรือ

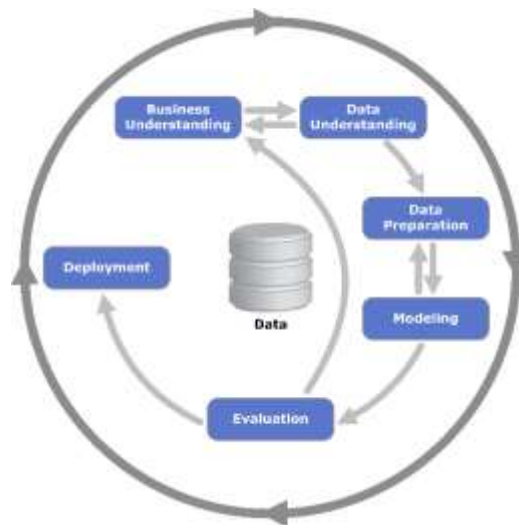
การเขียนโปรแกรมเพียงเล็กน้อย ตั้งแต่ปี 2006 KNIME ถูกนำมาใช้ในการวิจัยด้านเภสัชกรรม และ ยังใช้ในด้านอื่นๆ เช่นการวิเคราะห์ข้อมูลลูกค้า CRM ระบบธุรกิจอัจฉริยะ การทำเหมืองข้อความ และการวิเคราะห์ข้อมูลทางการเงิน เมื่อเร็ว ๆ นี้มีความพยายามในการใช้ KNIME เป็นเครื่องมืออัตโนมัติสำหรับกระบวนการอัตโนมัติ (RPA)

การพัฒนา KNIME เริ่มต้นในเดือนมกราคม 2547 โดยทีมวิศวกรซอฟต์แวร์ที่ มหาวิทยาลัยคอนสแตนซ์ในฐานะผลิตภัณฑ์ที่เป็นกรรมสิทธิ์ ทีมนักพัฒนาดังเดิมที่นำโดย Michael Berthold มาจากบริษัทใน Silicon Valley ที่ให้บริการซอฟต์แวร์สำหรับอุตสาหกรรมยา เป้าหมาย เริ่มต้นคือการสร้างแพลตฟอร์มการประมวลผลข้อมูลแบบแยกส่วน ปรับขนาดได้สูง และเปิดกว้าง ซึ่งอนุญาตให้รวมการไหลข้อมูล การประมวลผล การแปลง การวิเคราะห์ และโมดูลการสำรวจ ด้วยภาพต่างๆ ได้อย่างง่ายดาย โดยไม่ต้องเน้นไปที่การใช้งานเฉพาะใดๆ แพลตฟอร์มนี้มี จุดมุ่งหมายเพื่อเป็นแพลตฟอร์มการทำงานร่วมกันและการวิจัย และควรทำหน้าที่เป็นแพลตฟอร์ม การผสานรวมสำหรับโครงการวิเคราะห์ข้อมูลอื่นๆ

ในปี 2549 KNIME เวอร์ชันแรกเปิดตัวและบริษัทหลายแห่งเริ่มใช้ KNIME และผู้จำหน่ายซอฟต์แวร์ด้านวิทยาศาสตร์เพื่อชีวิตจำนวนหนึ่งเริ่มรวมเครื่องมือของตนเข้ากับ KNIME ใน ปีนั้น หลังจากบทความในนิตยสารเฮอร์มันน์ผู้ใช้จากพื้นที่อื่นๆ เข้าร่วมในเรือ ฅ ปี 2012 KNIME มี ผู้ใช้จริงมากกว่า 15,000 ราย (กล่าวคือไม่นับการดาวน์โหลด แต่ผู้ใช้เรียกข้อมูลอัปเดตเป็นประจำ เมื่อมีให้ใช้งาน) ไม่เพียงแต่ในวิทยาศาสตร์เพื่อชีวิตและในธนาคาร สำนักพิมพ์ ผู้ผลิตรถยนต์ telcos บริษัทที่ปรึกษา และ อุตสาหกรรมอื่น ๆ รวมถึงกลุ่มวิจัยจำนวนมากทั่วโลก การอัปเดตล่าสุด สำหรับเซิร์ฟเวอร์ KNIME และส่วนขยายข้อมูลขนาดใหญ่ของ KNIME ให้การสนับสนุนพื้นที่จัดเก็บ ข้อมูลประเภท Apache Spark 2.3, Parquet และ HDFS

### 2.3.3 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

CRISP-DM (Cross-Industry Standard Process for Data Mining) เป็นกระบวนการ หลักในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์และใช้ประโยชน์ในทางธุรกิจ โดยถือเป็น กระบวนการที่เป็นสากลในการวิเคราะห์ข้อมูลกระบวนการ CRISP-DM นี้จะประกอบด้วย 6 ขั้นตอน ดังรูปด้านล่าง



ภาพที่ 2.6 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

## 1. Business Understanding

เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทาง Data Mining พร้อมทั้งวางแผนในการดำเนินการคร่าวๆ เช่น ถ้าเรามีข้อมูล Monthly Report จากผู้แทนยาและข้อมูล Sale Order มา เริ่มแรกเราต้องมาทำความเข้าใจกับข้อมูลที่เราได้มาก่อน ซึ่งปัญหาจะอยู่ที่ Monthly Report จากทาง Sale มากกว่าที่หลายครั้งเนื้อหาที่ทางบริษัทให้กรอก นำไปใช้ประโยชน์ต่อไม่ได้ หรือ ข้อมูลที่กรอกมากก็ไม่ตรงกับความเป็นจริง

## 2. Data Understanding

ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์ ซึ่งปัญหาคือคนที่จะเป็นคนคัดเลือกข้อมูลต้องมีความเข้าใจทั้งเรื่องของที่มาของข้อมูล, ความเกี่ยวข้องของข้อมูล, ความรู้เชิงลึกในอุตสาหกรรมนั้นๆ มากกว่าแค่ นำข้อมูลที่หามา

### 3. Data Preparation

ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น

### 4. Modeling

ขั้นตอนนี้จะเป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทาง Data Mining ที่ได้แนะนำไปแล้ว เช่น การจำแนกประเภทข้อมูลหรือ การแบ่งกลุ่มข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปขั้นตอนที่ Data Preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย

ตัวอย่างเทคนิคในการวิเคราะห์ข้อมูลต่างๆ เช่น

- การแบ่งกลุ่มข้อมูล (Clustering)
- การหากฎความสัมพันธ์ (Association Rules)
- การจำแนกประเภทข้อมูล (Classification) ตัวอย่างเช่น
  - เทคนิค Decision tree
  - เทคนิค Naive Bayes
  - เทคนิค Neural Network
  - การวิเคราะห์การถดถอย Regression

## 5. Evaluation

ในขั้นตอนนี้เราจะได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคทาง Data Mining แล้วแต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปก็ต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้สำหรับการสร้างโมเดลด้วยเทคนิค Classification มีการทดสอบประสิทธิภาพของโมเดลอยู่ 3 แบบใหญ่ คือ

- Self-consistency test
- Split test
- Cross-validation test

## 6. Deployment

ในกระบวนการทำงานของ CRISP-DM นั้นไม่ได้หยุดเพียงแค่ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคทาง Data Mining เท่านั้นแม้ว่าผลลัพธ์ที่ได้จะแสดงถึงองค์ความรู้ที่มีประโยชน์แต่จะต้องนำองค์ความรู้ที่ได้เหล่านี้ไปใช้ได้จริงในองค์กรหรือบริษัท

โดยปกติเมื่อเราได้ทำตามทุกขั้นตอนแล้ว เรามักจะย้อนกลับไปทำซ้ำใหม่ เพื่อให้ Model ที่ได้ออกมา วิเคราะห์ข้อมูลได้ตรงและแม่นยำที่สุด โดยส่วนใหญ่ที่พบคือทำงานเสร็จครบทุกข้อแล้ว กลับพบว่าข้อมูลที่ได้มาไม่ตอบโจทย์ในสิ่งที่บริษัทต้องการ เพราะฉะนั้นก่อนที่จะทำ Data Mining ควรให้เวลากับข้อมูลที่เราให้มากๆ



## 2.4 วรรณกรรมที่เกี่ยวข้อง

ไกรสร เสงพะพรหม สุพจน์ เสงพะพรหม และสุวิมล มรรควิบูลย์ชัย (2559) การเลือกหลักสูตรเพื่อเข้าศึกษาต่อเป็นกระบวนการตัดสินใจที่ซับซ้อนซึ่งบุคคลที่จะตัดสินใจเลือกเข้าศึกษาต่อจำเป็นต้องรู้จักตัวเองในด้านความสนใจ ความสามารถ และความถนัดเป็นอย่างดี จึงจะทำให้ตัดสินใจได้อย่างถูกต้องเหมาะสม หากตัดสินใจเลือกเรียนในหลักสูตรที่ไม่เหมาะสมกับบุคลิกลักษณะอุปนิสัย ความรู้ ความสามารถ ความสนใจและความถนัดของตนเองและความต้องการของตลาดแรงงานอาจก่อให้เกิดผลเสียหายหลายประการ หากทำงานที่ไม่เหมาะสมกับบุคลิกลักษณะ ความสามารถ และความสนใจแล้วโอกาสที่จะรักงาน ตั้งใจทำงาน หรือรับผิดชอบในงานก็จะมึนอ้อย เกินความเบื่อหน่ายหรือทำงานอย่างไม่เต็มประสิทธิภาพ ทำให้มีโอกาสก้าวหน้าในการทำงานน้อย ทำให้ต้องเปลี่ยนงานบ่อยหรือต้องหางานใหม่ที่นาสนใจไปเรื่อยๆ จนกว่าจะถูกใจซึ่งเป็นการยากสำหรับนักเรียนที่เพิ่งสำเร็จการศึกษาระดับมัธยมศึกษาตอนปลายโดยทั่วไป เพราะนอกจากจะต้องรู้จักตัวเองแล้ว ผู้เลือกเรียนจะต้องรู้เนื้อหาของหลักสูตร คือ รู้เนื้อหาโดยละเอียดของหลักสูตรซึ่งอาจจะยากต่อการทำความเข้าใจสำหรับนักเรียนที่เพิ่งสำเร็จการศึกษาในระดับมัธยมศึกษาตอนปลาย

การทำเหมืองข้อมูลได้ถูกนำมาประยุกต์ใช้ในงานหลายประเภท เช่น ด้านธุรกิจ ด้านวิทยาศาสตร์และการแพทย์เพื่อวิเคราะห์สถิติการเกิดโรคและแนวทางในการรักษา รวมทั้งในด้านเศรษฐกิจและสังคม รวมถึงทางด้านการศึกษาดวย

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่า ทางผู้จัดทำจะนำแนวคิดและแนวทางมาใช้เป็นแนวทางในการวิเคราะห์ และสร้างแบบจำลองการวิเคราะห์ เพื่อเป็นส่วนช่วยในการสนับสนุนการตัดสินใจให้กับนักศึกษา

**ประกาศิต ช่างสุพรรณ (2553)** งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบสำหรับจำแนกระดับความมั่นใจของผู้เรียน โดยใช้เทคนิคต้นไม้ตัดสินใจอาศัยข้อมูลการทำแบบฝึกหัดวิชาฟิสิกส์ โดยรวมตัวแปรทั้งส่วนคะแนนการทำแบบฝึกหัดแต่ละคนและส่วนของวิเคราะห์แบบฝึกหัดแต่ละข้อ รวมทั้งสิ้น 11 ตัวแปร แยกชุดข้อมูลตามรายวิชา คือวิชาฟิสิกส์ 1 , วิชาฟิสิกส์ 3 และรวมกันทั้ง

ฟิลิกส์ 1 และฟิลิกส์ 3 จากนั้นแบ่งชุดข้อมูลตามระดับความมั่นใจของผู้เรียนเป็น 2 แบบ คือ แบบ 2 ระดับ (มั่นใจและไม่มั่นใจ) และ 3 ระดับ (มั่นใจสูง มั่นใจปานกลาง และมั่นใจต่ำ) รวมทั้งสิ้น 6 ชุดข้อมูล ทำการทดสอบตัวแบบที่ ได้ด้วยวิธี 10-fold Cross Validation ผลการทดลองพบว่า ชุดข้อมูลวิชาฟิลิกส์ 1 แบบ 2 ระดับ มีความถูกต้องสูงสุดเป็น 79.05 % และเมื่อใช้ข้อมูลคนละวิชา ที่มีระดับความมั่นใจเท่ากัน ตรวจสอบเพื่อหาค่าตัวแบบที่ดีที่สุด พบว่า ชุดข้อมูลที่รวมทั้งฟิลิกส์ 1 และฟิลิกส์ 3 แบบ 2 ระดับมีค่าเฉลี่ยความถูกต้องสูงสุด คือ 78.28 % ซึ่งอยู่ในระดับดี ปะสามารถที่จะใช้ตัวแบบดังกล่าวทั้งสองวิชาได้

ดังนั้นทางผู้จัดทำเห็นว่า การใช้เทคนิคต้นไม้การตัดสินใจของบทความดังกล่าวมีความคล้ายคลึงกับของผู้จัดทำที่จะนำเทคนิคต้นไม้ตัดสินใจมาทำการวิเคราะห์หาปัจจัยจากผลการเรียนแต่ละวิชาของนักศึกษา

**นายณฤทธิ์ จัตุรัสพันธ์ นายนิธิศ ขวรัตน์เจริญกิจ (2563)** วิเคราะห์ปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาของนักศึกษาภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยสยาม เนื่องจากมีจำนวนนักศึกษาที่เข้าศึกษาต่อในหลักสูตรวิทยาการคอมพิวเตอร์ มีอัตราตกออกและย้ายสาขาที่สูงเกินกว่าร้อยละ 50 จากจำนวนนักศึกษาที่เข้า ศึกษาต่อในปี การศึกษานั้นๆ และมีอัตราการสำเร็จการศึกษาตามเกณฑ์การศึกษา 4 ปี ที่ต่ำ ทางภาควิชาจึงต้องการทราบสาเหตุที่แท้จริง จึงได้มอบหมายให้คณะผู้จัดทำทำการวิเคราะห์หาสาเหตุดังกล่าว โดยได้กำหนดปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาในชั้นต้นไว้ 5 ปัจจัย ได้แก่ 1) วุฒิการศึกษาที่จบก่อนเข้าศึกษาต่อในหลักสูตรวิทยาการคอมพิวเตอร์ 2) เกรดเฉลี่ยสะสม (GPA) 3) รายวิชาบังคับ ที่ต้องเรียนก่อน (Pre-Requisite) 4) ผลการศึกษาตามรายวิชา และ 5) จำนวนหน่วยกิตสะสมที่ลงทะเบียนเรียนและผ่าน ในการวิเคราะห์ข้อมูลได้ใช้ข้อมูลการลงทะเบียนเรียนและผลการเรียนของนักศึกษาที่เข้าศึกษาต่อในหลักสูตรวิทยาการคอมพิวเตอร์ตั้งแต่ปี การศึกษา 2555 ถึง 2560 ขั้นตอนในการดำเนินการวิเคราะห์ประกอบด้วย 1) กำหนดวัตถุประสงค์ในการวิเคราะห์ข้อมูล 2) รวบรวมข้อมูล 3) เตรียมข้อมูล 4) จัดกลุ่มข้อมูล และ 5) นำเสนอแผนภาพข้อมูล เครื่องมือในการวิเคราะห์ข้อมูล ได้แก่ โปรแกรม Tableau ผลลัพธ์ที่ได้จากการวิเคราะห์ ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาภาควิชาวิทยาการคอมพิวเตอร์ ประกอบด้วย เกรดเฉลี่ยสะสม ผลการศึกษาตามรายวิชา

และจำนวนหน่วยกิตสะสมที่ลงทะเบียนเรียนและผ่าน โดยปัจจัยที่ไม่มีผลต่อการสำเร็จการศึกษาของนักศึกษาภาควิชาวิทยาการคอมพิวเตอร์ ประกอบด้วย วุฒิการศึกษาที่จบก่อนเข้าศึกษาต่อ และรายวิชาบังคับที่ต้องเรียนก่อน ซึ่งอาจารย์ที่ปรึกษาชั้นปี และอาจารย์ผู้สอนจะนำผลการวิเคราะห์ไปใช้เพื่อกำกับดูแลนักศึกษาและปรับปรุงวิธีการเรียนการสอนให้มีประสิทธิภาพยิ่งขึ้น

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่า จะนำกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM เพื่อเตรียมข้อมูลก่อนนำไปวิเคราะห์ได้อย่างเหมาะสม และการนำเทคนิคจัดกลุ่ม Clustering มาใช้ในการจัดกลุ่มรายวิชาเพื่อแบ่งรายวิชาที่ยาก ปลายทางง่าย

**สุมาลี ลิกเสน และมนตชัย เทียนทอง (2559)**งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ความถนัดของผู้เรียน ตามแนวทางพหุปัญญา ของ Howard Gardner ที่จำแนกความถนัดของผู้เรียนออกเป็น 3 กลุ่ม กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักศึกษามหาวิทยาลัยราชภัฏนครปฐม ในภาคเรียนที่ 2/2556 จำนวน 431คน ทำการวิเคราะห์ข้อมูลด้วยโปรแกรม WEKA โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ ภายใต้อัลกอริทึม j48 สร้างและทดสอบตัวแบบด้วยวิธี 10 fold Cross validation ผลการวิจัยพบว่า ตัวแปรต้นที่ส่งผลต่อความถนัดของผู้เรียน ตามแนวทางพหุปัญญาประกอบด้วย หลักสูตรเกรดเฉลี่ยวิชาที่ทา คะแนนได้สูงอยู่เสมอลักษณะการเรียนรู้ที่ชอบวิธีการเรียนที่มีความถนัดกิจกรรมที่ชอบและงานอดิเรกที่ชอบ ได้กฎการจำแนกความถนัดของผู้เรียนทั้งสิ้นจำนวน 54 กฎดังนี้ ความถนัดของผู้เรียนกลุ่มการวิเคราะห์ (Analytic) จำนวน 19 กฎ ความถนัดของผู้เรียนกลุ่มพินิจพิจารณา (Introspective) จำนวน 14 กฎ และความถนัดของผู้เรียนกลุ่มปฏิสัมพันธ์ (Interactive) จำนวน 21 กฎ ซึ่งมีค่าความถูกต้องคิดเป็นร้อยละ 85.90 ค่าความแม่นยำคิดเป็นร้อยละ 86.03 ค่าความระลึกคิดเป็นร้อยละ 85.56 และค่าถ่วงดุลคิดเป็นร้อยละ 85.89 โดยผลที่ได้จากงานวิจัยนี้ สามารถนำไปใช้ในการพัฒนาระบบ การวิเคราะห์ความถนัดของผู้เรียนตามแนวทางพหุปัญญาของ Howard Gardner โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องและแม่นยำในการทำนายความถนัดของผู้เรียนได้สำหรับแบ่งกลุ่มเพื่อจัดทำระบบโครงการออนไลน์ต่อไป

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่า ทางผู้จัดทำจะนำกฎการจำแนกด้วยเทคนิคต้นไม้ตัดสินใจมาใช้ในการสร้างแบบจำลองเพื่อเป็นมาตรการช่วยเหลือด้านการศึกษาสำหรับนักศึกษา โดยทำการวิเคราะห์ข้อมูลด้วยโปรแกรม WEKA

## 2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Clustering เพื่อจัดกลุ่มรายวิชา จากนั้นจะนำข้อมูลที่ผ่านมาหาปัจจัยที่มีผลต่อการสำเร็จของนักศึกษาโดยใช้เทคนิคต้นไม้ตัดสินใจ Decision tree โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือโปรแกรม WEKA 3.9.6 และโปรแกรม KNIME ในการสร้างโมเดล Clustering และ Decision tree เพื่อทำการเปรียบเทียบและเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Power Bi เผยแพร่บน Web Browser ที่เป็นที่ยอมรับในยุคอินเทอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา