

## บทที่ 2

### แนวคิด ทฤษฎีเอกสารที่เกี่ยวข้อง

จากการศึกษาเอกสารงานการวิเคราะห์พฤติกรรมการณ์ซื้ออาหารญี่ปุ่น โดยใช้กฎความสัมพันธ์ด้วยอัลกอริทึมเอโพรออริกรณีศึกษาร้านอาหารญี่ปุ่น ผู้ศึกษาพบว่าเอกสารแนวความคิดและโครงการต่าง ๆ ที่เกี่ยวข้องกับการศึกษาสรุปได้ ดังนี้

#### 2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic)

2.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

#### 2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

2.2.6 ทฤษฎีเกี่ยวกับชุดคำสั่ง PHP

2.2.7 ทฤษฎีเกี่ยวกับการ Visualization

2.2.8 ทฤษฎีเกี่ยวกับกฎความสัมพันธ์ ด้วย Apriori Algorithm

2.2.9 ทฤษฎีเกี่ยวกับเทคนิคเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์ ด้วย

Pandas Library Python

#### 2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม Weka

2.3.2 โปรแกรม Sublime text 3

2.3.3 โปรแกรม Adobe XD

2.3.4 โปรแกรม Tableau

#### 2.4 วรรณกรรมที่เกี่ยวข้อง

#### 2.5 บทสรุป

## 2.1 แนวคิด

### 2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic)

Data analytics เป็นศาสตร์ของการวิเคราะห์ข้อมูลต่าง ๆ จาก Big data เพื่อช่วยในด้านธุรกิจ หรือตามวัตถุประสงค์อื่น ๆ ที่ต้องการซึ่งเริ่มต้นจากความสามารถของเราในการนำข้อมูลเหล่านั้นมาให้อยู่ในรูปแบบที่พร้อมจะประมวลผลได้ โดยเทคโนโลยีหรือชุดคำสั่งและแบบจำลองที่สร้างขึ้น เพื่อนำข้อมูลที่ได้ผ่านการวิเคราะห์แล้วมาใช้ หรือแปลความหมายโดยบุคคลที่ได้รับการฝึกอบรมการใช้เทคโนโลยีเหล่านั้น รูปแบบของการวิเคราะห์ข้อมูลสามารถแบ่งได้ดังนี้

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

2) การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้นปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น ตัวอย่างเช่น ความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่จะช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

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

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

### 2.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

การนำข้อมูลดิบมาเปลี่ยนเป็นกราฟ แผนภูมิ หรือแม้กระทั่งวิดีโอที่ช่วยอธิบายปริมาณ ตัวเลข และช่วยให้คุณได้ข้อมูลเชิงลึกจากข้อมูลเหล่านั้น การใช้ Data Visualization

เปลี่ยนแนวคิดของเราเกี่ยวกับการสร้างคุณค่าให้กับข้อมูล ทำให้เราค้นพบรูปแบบใหม่ ๆ และมองเห็นแนวโน้มของข้อมูลมากขึ้น

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



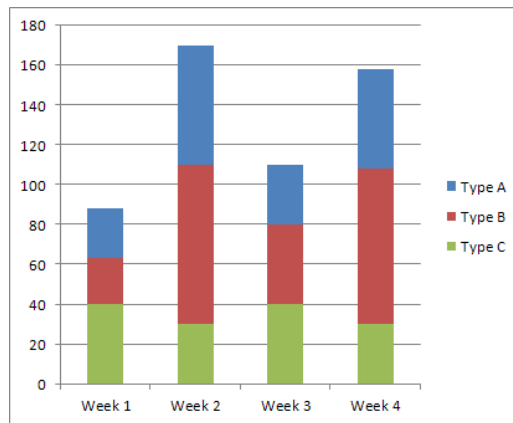
ภาพที่ 2.1 แสดง Data Visualization

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



## 3) Comparison

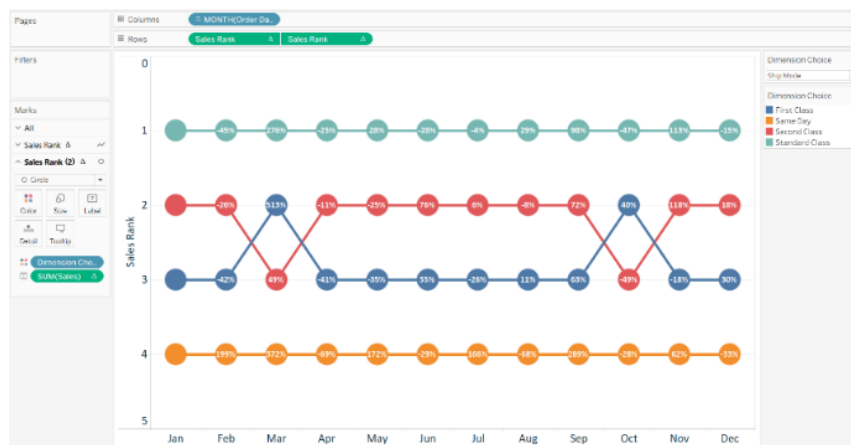
Stacked Bar Chart ใช้เปรียบเทียบค่าผลรวมและสัดส่วนจากข้อมูลหลายกลุ่ม อาจอ่านสัดส่วนยากเมื่อมีข้อมูลหลายกลุ่มมากเกินไป



ภาพที่ 2.4 แสดง Stacked Bar Chart

## 4) Ranking

Bump Chart ใช้แสดงการเปลี่ยนแปลงลำดับของข้อมูลในหลายช่วงเวลา โดยใช้สีเพื่อช่วยให้อ่านข้อมูลได้ง่ายขึ้น



ภาพที่ 2.5 แสดง Bump Chart

## 5) Time-Series

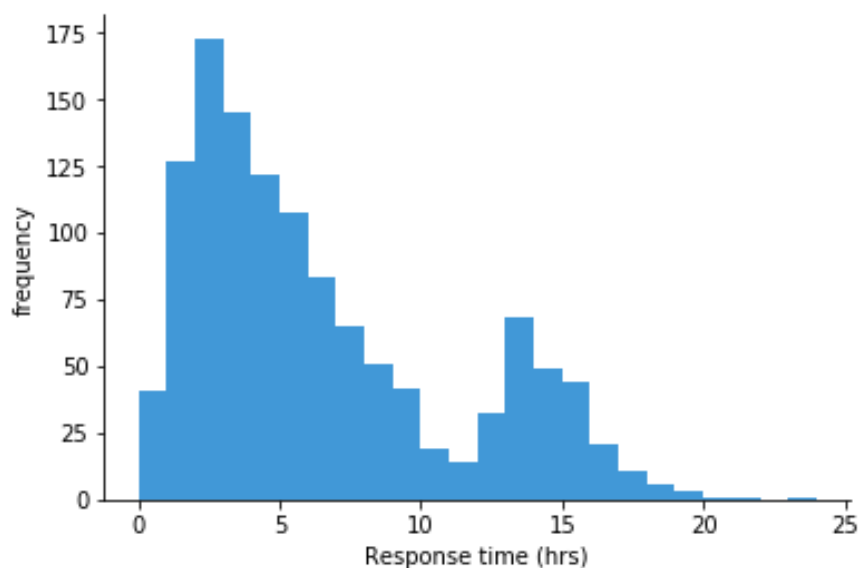
Line Chart ใช้เปรียบเทียบข้อมูลเพื่อดูแนวโน้ม (Trends) โดยอาจจะเทียบกับมิติของเวลา time



ภาพที่ 2.6 แสดง Line Chart

## 6) Distribution

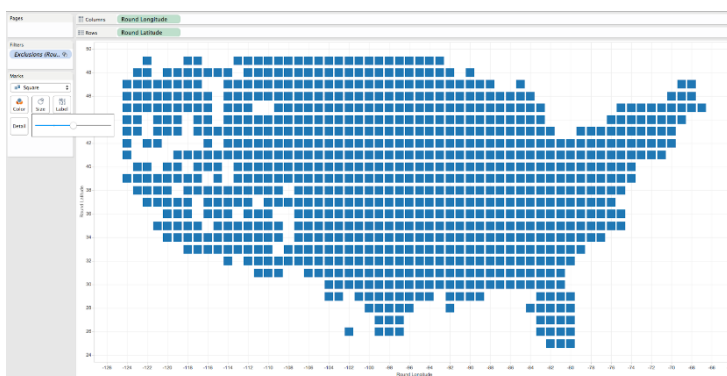
Histogram กราฟแท่งแบบเฉพาะที่แสดงความสัมพันธ์ระหว่างข้อมูลเป็นหมวดหมู่ที่เรียกว่าชั้นข้อมูลกับความถี่ของข้อมูล เพื่อดูการกระจายของข้อมูล ลดระยะห่างระหว่างแท่งเพื่อให้เห็น “รูปทรง” ของข้อมูลชัดเจน



ภาพที่ 2.7 แสดง Histogram

## 7) Spatial

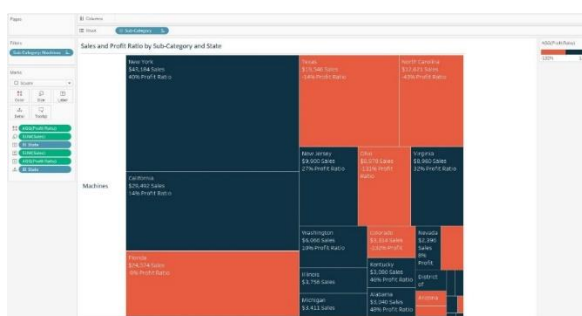
Hexbin Map เป็นการแสดงผลข้อมูลด้วยรูปทรงหกเหลี่ยมที่มีขนาดเท่ากัน โดยกำหนดสีแทนระดับความถี่ของจำนวนข้อมูล ส่วนมากนำไปใช้ลงบนตำแหน่งทางภูมิศาสตร์ เพื่อให้เห็นภาพได้ชัดเจน



ภาพที่ 2.8 แสดง Hexbin Map

## 8) Composition

Treemap ใช้แสดงสัดส่วนของข้อมูลแบบมีลำดับชั้น (hierarchy) แสดงเป็นรูปสี่เหลี่ยม โดยจะแสดงทั้งในแง่ของสีเดียวกันจัดเป็นกลุ่มเดียวกัน อาจอ่านสัดส่วนยากเมื่อมีกลุ่มขนาดเล็กจำนวนมาก



ภาพที่ 2.9 แสดง Treemap

จากที่ได้กล่าวไปจะเห็นได้ว่าข้อมูลแบบ Data Visualization มีพลังทางการสื่อสารอย่างมาก เพราะสามารถแสดงผลและเพิ่มความสนใจได้เป็นอย่างดี หากองค์กรได้ลองนำการแสดงผลข้อมูลแบบ Data Visualization มาใช้ก็จะช่วยในการจัดการระดับสายงานแผนกอื่น ๆ ในบริษัท จะส่งผลให้การวิเคราะห์ข้อมูลและตัดสินใจในสายงานต่าง ๆ ได้ง่าย สามารถช่วยให้มีประสิทธิภาพที่ดีต่อธุรกิจและทีม หรือนำมาใช้ในหน้าแรกของเว็บไซต์ เพื่อทำการโปรโมท

ไปโรมันหรือข้อเสนอทางการตลาดของธุรกิจ แล้วจะพบว่า Data Visualization ช่วยให้เกิดการตอบสนองกับข้อความมากขึ้น

## 2.2 ทฤษฎี

### 2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

#### 2.2.1.1 ข้อมูลขนาดใหญ่ (Big Data)

Big Data หรือข้อมูลขนาดใหญ่มหาดศาล ที่เกิดขึ้นแต่ละวัน เช่น ข้อมูลจากการทำธุรกิจ ข้อมูลจากการซื้อขายสินค้า ข้อมูลคนผ่านเข้าออกสถานีรถไฟ หรือ ข้อมูลคนที่ใช้บริการของ Facebook ข้อมูลที่มีปริมาณมหาศาลจนซอฟต์แวร์หรือฮาร์ดแวร์ธรรมดาไม่สามารถที่จะจัดการหรือวิเคราะห์ได้อย่างมีประสิทธิภาพ ดังนั้นต้องมีเทคโนโลยีและสถาปัตยกรรมไอทีรุ่นใหม่ ซึ่งอาจมาในรูปแบบซอฟต์แวร์ ที่สามารถรองรับการจัดการ การจัดการ กรองเลือกข้อมูล การวิเคราะห์ แสดงผล ซึ่งสามารถนำไปประยุกต์ใช้งานให้เป็นประโยชน์ต่อองค์กรต่าง ๆ สำหรับภาคเอกชนทั่วไปก็ต้องนำมาใช้ให้เกิดประโยชน์ต่อองค์กรมากที่สุด วัตถุประสงค์การใช้งานเพื่อองค์กรของตนเป็นหลัก แต่นอกจาก Big Data จะมีความสำคัญกับภาคเอกชนแล้วสำหรับภาครัฐนั้นก็มีความสำคัญอย่างยิ่งเช่นกัน เพราะรัฐจำเป็นต้องใช้ข้อมูลของ Big Data มากก่อให้เกิดประโยชน์ต่อประชาชนทุกคน เพื่อจะกระจายความช่วยเหลือไปถึงทุกที่ พร้อมทั้งสามารถนำข้อมูลที่มีการวิเคราะห์ไปใช้ประโยชน์ในการบริหาร วางแผนกำหนดทิศทางในด้านต่าง ๆ อย่างมีประสิทธิภาพที่สูงขึ้น คำว่า Big Data มักจะสื่อถึงชุดข้อมูลที่มีขนาดใหญ่ ซึ่งอาจจะแบ่งย่อยออกเป็นข้อมูล 3 รูปแบบคือ

1) ข้อมูลเชิงโครงสร้าง (structured data) คือตารางข้อมูลที่มีการจัดเรียงอย่างมีรูปแบบชัดเจนและเป็นระเบียบ สามารถนำมาใช้วิเคราะห์ได้เลย เช่น ข้อมูลที่เก็บไว้ในโปรแกรม spreadsheet อย่าง Microsoft Excel

2) ข้อมูลกึ่งโครงสร้าง (semi-structured data) คือข้อมูลที่ถูกจัดเก็บอย่างมีรูปแบบในระดับหนึ่งและข้อมูลที่สามารถค้นหา (search) หรือแท็ก (tag) ได้ เช่น เว็บเพจที่มีการระบุชื่อเพจ คำสำคัญในเพจ และวันที่อัปเดตข้อมูล

3) ข้อมูลที่ไม่มีโครงสร้างแน่นอน (Unstructured data) มักจะหมายถึงข้อความยาวๆ ข้อความเสียง รูปภาพ และวิดีโอ

คุณลักษณะของ Big Data ที่มีประสิทธิภาพ มี 6 ประการประกอบด้วย



1) ปริมาณมาก (Volume) ซึ่งได้จากการที่องค์กรต่าง ๆ รวบรวมข้อมูลจากหลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย ซึ่งสามารถเป็นได้ทั้งรูปแบบ Online และ Offline ซึ่งส่วนมากแล้วจะมีปริมาณมากกว่าหน่วย TB (Terabyte) ขึ้นไป

2) มีการเปลี่ยนแปลงอย่างรวดเร็ว (Velocity) ด้วยเทคโนโลยีที่ล้ำหน้าขึ้นทุกวัน ความเร็วในการได้รับข้อมูลที่เพิ่มขึ้นและต้องได้รับการจัดการในเวลาที่เหมาะสม ข้อมูลเหล่านี้ในแบบเรียลไทม์ ส่งผ่านข้อมูลกันอย่างต่อเนื่อง (Real-time) จนทำให้การวิเคราะห์ห่วยๆ เกินขีดจำกัด หรือไม่สามารถจับรูปแบบหรือทิศทางของข้อมูลได้

3) หลากหลายประเภทหรือแหล่งที่มา (Variety) หมายถึงรูปแบบของข้อมูลที่แตกต่างกันออกไป นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานะข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ ทั้งในรูปแบบ ตัวอักษร วิดีโอ รูปภาพ ไฟล์ต่าง ๆ และหลากหลายแหล่งที่มา เช่น Social Network ต่าง ๆ อีเมล วิดีโอ เสียง ข้อมูลหุ้น และธุรกรรมทางการเงิน

4) ยังไม่ผ่านการประมวลผล (Veracity) ข้อมูลที่มีระดับคุณภาพปะปนกันไป ยังไม่ผ่านการ Process ให้อยู่ในรูปแบบของข้อมูลที่สามารถใช้สร้างประโยชน์ต่อองค์กร ซึ่งเป็นข้อมูลที่ยังไม่สมบูรณ์จึงยังไม่สามารถนำไปประกอบการตัดสินใจได้

5) คุณค่า (Value) หมายถึง ข้อมูลมีประโยชน์และมีความสัมพันธ์ในเชิงธุรกิจ ซึ่งต้องเข้าใจก่อนว่าไม่ใช่ทุกข้อมูลจะมีประโยชน์ในการเก็บและวิเคราะห์ ข้อมูลที่มีประโยชน์จะต้องเกี่ยวข้องกับวัตถุประสงค์ทางธุรกิจ เช่นถ้าต้องการเพิ่มขีดความสามารถในการแข่งขันในตลาดของผลิตภัณฑ์ที่ขาย ข้อมูลที่มีประโยชน์ที่สุดน่าจะเป็นข้อมูลผลิตภัณฑ์ของลูกค้า

6) ความแปรผันได้ (Variability) หมายถึง ข้อมูลสามารถในการเปลี่ยนแปลงรูปแบบไปตามการใช้งาน หรือสามารถคิดวิเคราะห์ได้จากหลายแง่มุม และรูปแบบในการจัดเก็บข้อมูลก็อาจจะต่างกันไปในแต่ละแหล่งของข้อมูล

### 2.2.1.2 การจัดการข้อมูลขนาดใหญ่

การรวบรวมข้อมูลมาจัดเก็บ (Storage) เป็นการรวบรวมข้อมูลของจากหลากหลายทั้งที่มา และการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบดั้งเดิม ETL (Extract Transform and Load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิค วิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาด เทราไบต์ และเพตะ

ไบต์ ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล จัดรูปแบบ ให้เหมาะสมสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้น ๆ

1) การประมวลผล (Processing) ข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ On-premises หรือแบบ Cloud ขึ้นกับความต้องการ หรือความสะดวกในการใช้ ซึ่งเราสามารถใช่และประเมินผลได้เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูล หรือข้อมูลบางอันต้องการความยืดหยุ่นสูง และไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยมกันเป็นอย่างมาก

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

### 2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

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

กลุ่มอาศัยความคล้ายคลึงกันของข้อมูล Association rules อาศัยความสัมพันธ์ของข้อมูลที่เกิดขึ้นร่วมกัน สร้างเป็นกฎความสัมพันธ์ เช่น “ซื้อเบียร์แล้วจะ ซื้อผ้าอ้อมไปด้วย”

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

1) ความเข้าใจทางธุรกิจ (Business Understanding) เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทางดาต้าไมน์นิ่งพร้อมทั้งวางแผนในการดำเนินการคร่าว ๆ

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

3) การเตรียมข้อมูล (Data Preparation) ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) เช่น การแปลงข้อมูลให้อยู่ ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น โดยขั้นตอนนี้จะเป็นขั้นตอนที่ใช้ เวลามากที่สุดของกระบวนการ CRISP-DM

4) การพัฒนาแบบจำลอง (Modeling) ในขั้นตอนนี้จะเป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิ่ง ที่ได้แนะนำไปแล้ว เช่น การจำแนกประเภทข้อมูล หรือ การแบ่งกลุ่ม ข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปขั้นตอน Data Preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย เทคนิคในการวิเคราะห์ข้อมูลต่าง ๆ

5) การทดสอบแบบจำลอง (Evaluation) ในขั้นตอนนี้เราจะได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิ่งแล้วแต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปก็จะต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้

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

### 2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

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

เว็บไซต์ที่ได้รับการออกแบบอย่างสวยงาม มีการใช้งานที่สะดวก ย่อมได้รับความสนใจจากผู้ใช้งาน มากกว่าเว็บไซต์ที่ดูสับสนวุ่นวาย มีข้อมูลมากมายแต่หาอะไรไม่เจอ นอกจากนี้ยังใช้เวลาในการแสดงผลแต่ละหน้านานเกินไป ซึ่งปัญหาเหล่านี้ล้วนเป็นผลมาจากการออกแบบเว็บไซต์ไม่ดีทั้งสิ้น

ดังนั้น การออกแบบเว็บไซต์จึงเป็นกระบวนการสำคัญในการสร้างเว็บไซต์ ให้ประทับใจผู้ใช้ ทำให้เขาอยากกลับมาเข้ามาเว็บไซต์เดิมอีกในอนาคต ซึ่งนอกจากต้องพัฒนาเว็บไซต์ที่ดีมีประโยชน์แล้ว ยังต้องคำนึงถึงการแข่งขันกับเว็บไซต์อื่น ๆ อีกด้วย

#### 2.2.3.1 องค์ประกอบของการออกแบบเว็บไซต์

1) ความเรียบง่าย (Simplicity) หมายถึง การจำกัดองค์ประกอบเสริมให้เหลือเฉพาะองค์ประกอบหลัก กล่าวคือในการสื่อสารเนื้อหาให้กับผู้ใช้นั้น เราต้องเลือกเสนอสื่อที่เราต้องการนำเสนอจริง ๆ ออกมาในส่วนของกราฟิก สี สัน ตัวอักษรและภาพเคลื่อนไหว ต้องเลือกให้พอเหมาะ ถ้าหากมีมากเกินไปจะรบกวนสายตาและสร้างความรำคาญต่อผู้ใช้งาน ตัวอย่างเว็บไซต์ที่ได้รับการออกแบบที่ดี ได้แก่ เว็บไซต์ของบริษัทใหญ่ ๆ อย่างเช่น Apple Adobe Microsoft หรือ Kokia ที่มีการออกแบบเว็บไซต์ในรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานอย่างสะดวก

2) ความสม่ำเสมอ ( Consistency) หมายถึง การสร้างความสม่ำเสมอให้เกิดขึ้นตลอดทั้งเว็บไซต์ โดยอาจเลือกใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ก็ได้ เพราะถ้าหากว่าแต่ละหน้าในเว็บไซต์นั้นมีความแตกต่างกันมากจนเกินไป อาจทำให้ผู้ใช้เกิดความสับสนและไม่แน่ใจว่ากำลังอยู่ในเว็บไซต์เดิมหรือไม่ เพราะฉะนั้นการออกแบบเว็บไซต์ในแต่ละหน้าควรที่จะมีรูปแบบ สไตล์ของกราฟิก ระบบเนวิเกชัน (Navigation) และโทนสีที่มีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

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

4) เนื้อหา (Useful Content) ถือเป็นสิ่งสำคัญที่สุดในเว็บไซต์ เนื้อหาในเว็บไซต์ต้องสมบูรณ์และได้รับการปรับปรุงพัฒนาให้ทันสมัยอยู่เสมอ ผู้พัฒนาต้องเตรียมข้อมูลและเนื้อหาที่ผู้ใช้งานต้องการให้ถูกต้องและสมบูรณ์ เนื้อหาที่สำคัญที่สุดคือเนื้อหาที่ทีมผู้พัฒนาสร้างสรรค์ขึ้นมาเอง และไม่ไปซ้กับเว็บอื่น เพราะจะถือเป็นสิ่งที่ดึงดูดผู้ใช้ให้เข้ามาเว็บไซต์ได้เสมอ แต่ถ้าเป็นเว็บที่ลิงค์ข้อมูลจากเว็บอื่น ๆ มาเมื่อใดก็ตามที่ผู้ใช้ทราบว่ามีข้อมูลนั้นมาจากเว็บใด ผู้ใช้ก็ไม่จำเป็นต้องกลับมาใช้งานลิงค์เหล่านั้นอีก

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

6) คุณภาพของสิ่งที่ปรากฏให้เห็นในเว็บไซต์ (Visual Appeal) ลักษณะที่น่าสนใจของเว็บไซต์นั้น ขึ้นอยู่กับความชอบส่วนบุคคลเป็นสำคัญ แต่โดยรวมแล้วก็สามารถสรุปได้ว่าเว็บไซต์ที่น่าสนใจนั้นส่วนประกอบต่าง ๆ ควรมีคุณภาพ เช่น กราฟิกควรสมบูรณ์ไม่มี

รอยหรือขอบชั้นบันไดให้เห็น ชนิดตัวอักษรอ่านง่ายสบายตา มีการเลือกใช้โทนสีที่เข้ากันอย่างสวยงาม เป็นต้น

7) ความสะดวกของการใช้ในสภาพต่าง ๆ (Compatibility) การใช้งานของเว็บไซต์นั้นไม่ควรมีขอบจำกัด กล่าวคือ ต้องสามารถใช้งานได้ดีในสภาพแวดล้อมที่หลากหลาย ไม่มีการบังคับให้ผู้ใช้ต้องติดตั้งโปรแกรมอื่นใดเพิ่มเติม นอกเหนือจากเว็บเบราว์เซอร์ ควรเป็นเว็บที่แสดงผลได้ดีในทุกระบบปฏิบัติการ สามารถแสดงผลได้ในทุกความละเอียดหน้าจอ ซึ่งหากเป็นเว็บไซต์ที่มีผู้ใช้บริการมากและกลุ่มเป้าหมายหลากหลายควรให้ความสำคัญกับเรื่องนี้ให้มาก

8) ความคงที่ในการออกแบบ (Design Stability) ถ้าต้องการให้ผู้ใช้ใช้งานรู้สึก ว่าเว็บไซต์มีคุณภาพ ถูกต้อง และเชื่อถือได้ ควรให้ความสำคัญกับการออกแบบเว็บไซต์เป็นอย่างมาก ต้องออกแบบวางแผนและเรียบเรียงเนื้อหาอย่างรอบคอบ ถ้าเว็บที่จัดทำขึ้นอย่าง ลวก ๆ ไม่มีมาตรฐานการออกแบบและระบบการจัดการข้อมูล ถ้ามีปัญหามากขึ้นอาจส่งผลให้ เกิดปัญหาและทำให้ผู้ใช้หมดความเชื่อถือ

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

### 2.2.3.2 รูปแบบโครงสร้างของเว็บไซต์

การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างแบบเรียงลำดับ จะเป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิงค์เนื้อหา ก็จะเป็นลิงค์ไปที่หน้า มีทิศทางกรเข้าสู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ปุ่มเดินหน้า-ถอยหลังในการ

กำหนดทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับก็มีข้อเสีย คือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหาเพราะไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาด้วยตัวเองได้



ภาพที่ 2.10 แสดงโครงสร้างแบบเรียงลำดับ

2) โครงสร้างแบบลำดับชั้น นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่าง ๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆ และมีการนำเสนอรายละเอียดค่อยๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความเข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีไฮมเพจเป็นจุดเริ่มต้น และจุดรวมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง



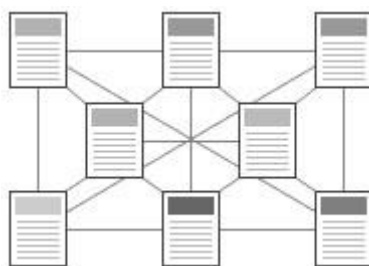
ภาพที่ 2.11 แสดงโครงสร้างแบบลำดับชั้น

3) โครงสร้างแบบตาราง เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีคามยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่าง ๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งานสามารถเปลี่ยนทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลาแถมยังทำให้เว็บไซต์มีความทันสมัยขึ้น



ภาพที่ 2.12 แสดงโครงสร้างแบบตาราง

4) โครงสร้างแบบใยแมงมุม เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถเข้าถึงหน้าเว็บเพจต่าง ๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2.13 แสดงโครงสร้างแบบใยแมงมุม

2.2.3.3 ส่วนประกอบสำคัญของหน้าเว็บเพจ บนหน้าเว็บเพจ จะมีส่วนประกอบสำคัญที่จำเป็นต้องมีอยู่ 3 ส่วน ได้แก่

1) ส่วนหัวของหน้า (Header) อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซต์ต่อไป ซึ่งส่วนใหญ่มักจะมีการใส่ภาพกราฟฟิกให้ดูสวยงาม สิ่งสำคัญหลักๆ เลย ก็คือ โลโก้ ชื่อเว็บไซต์และเมนูหลักที่สามารถลิงค์ไปยังเนื้อหาในหน้าเว็บเพจต่าง ๆ ได้

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

3) ส่วนท้ายของหน้า (Footer) อยู่ล่างสุดของหน้าเว็บ ซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่าง ๆ เพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูลเจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่าง ๆ เกี่ยวกับการใช้งานเว็บไซต์อย่างถูกต้อง เป็นต้น

#### 2.2.3.4 วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์

การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บ ว่าต้องการให้ผู้เข้าชมรู้สึก



อย่างไรต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่มดังต่อไปนี้

1) สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอดภัยและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2) สีโทนเย็น (Cool Colors) เป็นสีแห่งความสุภาพและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้โน้มน้าวจากในระยะไกลได้อีกด้วย

3) สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

สีสามารถสื่อถึงอารมณ์ความรู้สึกและสื่อความหมายของเว็บไซต์นั้น ๆ ได้อย่างชัดเจน ดังนั้นหากเลือกใช้สีไม่เหมาะสมกับเนื้อหาหรือจุดประสงค์ของเว็บ ก็จะทำให้เว็บดูไม่น่าสนใจ ขาดความน่าเชื่อถือและทำให้ผู้ที่เคยเข้ามาใช้บริการไม่คิดจะกลับมาใช้บริการอีก

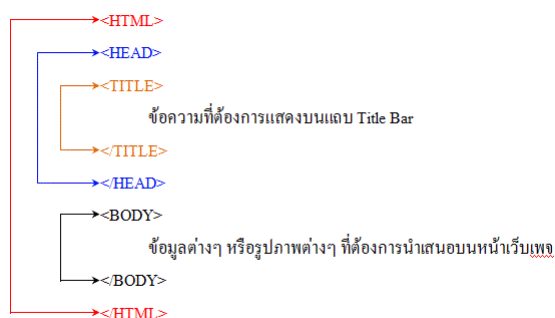
#### 2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

ภาษา HTML (Hyper Text Markup Language) เป็นภาษาที่ใช้ในการเขียนโปรแกรมข้อมูลที่แสดงผลบนเครือข่ายอินเทอร์เน็ตในลักษณะของข้อความ รูปภาพ เสียง และภาพเคลื่อนไหวต่าง ๆ ภาษา HTML เป็นภาษาที่ง่ายต่อการเรียนรู้ สามารถกำหนดรูปแบบและโครงสร้างได้ง่าย ทำให้ได้รับความนิยม และมีการพัฒนาอย่างต่อเนื่องเพื่อให้ใช้งานง่ายขึ้น และตอบสนองต่องานด้านกราฟิกมากยิ่งขึ้น และสนับสนุนการแสดงผลในเว็บเบราว์เซอร์มากมาย และบันทึกในรูปแบบของไฟล์นามสกุล html หรือ htm

(Tag) เป็น คำสั่งหลักของ HTML แทบจะพูดได้ว่า ทุกสิ่งทุกอย่างของ HTML จะขึ้นอยู่กับ tag ทั้งนี้ ไม่ว่าจะเป็น การเน้นข้อความ การแสดงภาพประกอบ หรือการสร้างจุดเชื่อมโยง (link) โปรแกรมเบราว์เซอร์จะตีพิมพ์เอกสารออกมาในรูปแบบใด ก็โดยคำสั่ง tag ทั้งสิ้น

### 2.2.4.1 ไวยากรณ์ของ HTML จะแบ่งออกเป็น 2 ส่วน คือ

- 1) ส่วนของคำสั่ง (Tag) เป็นส่วนที่กำหนดรูปแบบของข้อความที่แสดง ซึ่งเราเรียกว่า Tag โดยจะอยู่ในเครื่องหมาย < ... >
- 2) ส่วนของบทความทั่ว ๆ ไป เป็นส่วนของข้อความที่เราต้องการแสดงผล ตัวอย่างการใช้งานภาษา HTML



ภาพที่ 2.14 ไวยากรณ์ของ HTML

#### คำสั่งเริ่มต้นของเอกสาร HTML

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

<HEAD>.....</HEAD> Head Section เป็นส่วนที่ใช้อธิบายเกี่ยวกับข้อมูลเฉพาะของหน้านั้น ๆ เช่น ชื่อเรื่องของหน้าเว็บ (Title) ชื่อผู้จัดทำเว็บ(Author) คีย์เวิร์ดสำหรับการค้นหา (Keyword)

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

<BODY>.....</BODY> Body Section เป็นส่วนเนื้อหาหลักของหน้าเว็บ ซึ่งการแสดงผลจะต้องใช้ Tag จำนวนมาก ขึ้นอยู่กับลักษณะของข้อมูล เช่น ข้อความ, รูปภาพ, เสียง, วิดีโอ หรือไฟล์ต่าง ๆ ส่วนเนื้อหาเอกสารเว็บ เป็นส่วนการทำงานหลักของหน้าเว็บ ประกอบด้วย Tag มากมายตามลักษณะของข้อมูล ที่ต้องการนำเสนอ การป้อนคำสั่งในส่วนนี้ไม่มีข้อจำกัดสามารถป้อนติดกัน หรือ 1 บรรทัดต่อ 1 คำสั่งก็ได้ แต่ส่วนใหญ่จะยึดรูปแบบที่

อ่านง่าย คือ การทำย่อหน้าในชุดคำสั่งที่เกี่ยวข้องกัน ทั้งนี้ให้ป้อนคำสั่งทั้งหมดภายใต้ Tag <BODY> ... </BODY>

### 2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

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

#### 2.2.5.1 ไวยากรณ์ของ CSS

1) CSS จะประกอบไปด้วย ชื่อ Class หรือชื่อ ID หรือชื่อ Tag HTML ใน CSS จะเรียกว่า Selector ที่ใช้สำหรับการอ้างถึง

2) มีชื่อ Properties หรือชื่อคุณสมบัติ และตามด้วย Value เพื่อกำหนดค่าให้แก่คุณสมบัตินั้น และระหว่างชื่อ Properties กับ Value จะถูกคั่นด้วยเครื่องหมาย colon (properties: value)

3) คุณสมบัติภายในจะมีได้มากกว่า 1 คุณสมบัติตามที่เรากำหนด และ properties ในแต่ละ properties จะถูกคั่นด้วยเครื่องหมาย semi-colon (;)

4) ต้องเขียนคุณสมบัติของ CSS นั้นไว้ภายในเครื่องหมายปีกกา [...]

เท่านั้น

```
File Edit Format View Help
<html>
<head> ชื่อ Class (,Tag Selector, ID Selector)
<style type="text/css">
<!--
.font18red {
font-family: tahoma; font-weight: bold; font-size: 18px; color: #CC0000;
}
-->
</style> properties value
</head>
```

ภาพที่ 2.15 ไวยากรณ์ของ CSS

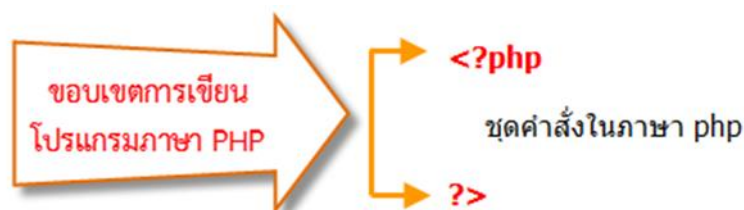
### 2.2.6 ทฤษฎีเกี่ยวกับชุดคำสั่ง PHP

ในช่วงแรกภาษาที่นิยมใช้งานบนระบบเครือข่าย คือ ภาษา HTML (Hypertext Markup Language) แต่ภาษา HTML มีลักษณะเป็น Static คือ ภาษาที่มีลักษณะของข้อมูลคงที่ ซึ่งไม่เพียงพอต่อความต้องการในปัจจุบันที่นิยมใช้ระบบเครือข่าย Internet เป็นศูนย์กลางในการติดต่อระหว่างกัน ทำให้ต้องการใช้เว็บไซต์ที่มีลักษณะเป็นแบบ Dynamic คือ เว็บไซต์ที่ข้อมูลสามารถเปลี่ยนแปลงได้โดยอัตโนมัติตามเงื่อนไขต่าง ๆ ที่ผู้เขียนเว็บไซต์เป็นผู้กำหนด และการควบคุมการทำงานเหล่านี้จะกระทำโดยโปรแกรมภาษาสคริปต์ เช่น ภาษา PHP ซึ่งเป็นภาษาหนึ่งที่ได้รับค่านิยมเป็นอย่างมากในปัจจุบัน

PHP ย่อมาจากคำว่า “Personal Home Page Tool” (ปัจจุบันได้เพิ่มเติมคำย่อใหม่ โดยรวมกับตัวย่อเป็น PHP : PHP Hypertext Preprocessor) ซึ่งเป็นภาษาประเภท Script Language ที่ทำงานแบบ Server Side Script กระบวนการทำงานจะทำงานแบบโปรแกรมแปลคำสั่ง interpreter คือแปลภาษาทุกครั้งที่มีคนเรียกสคริปต์ ข้อดี คือ ไม่ต้องนำไปประมวลผลใหม่ (Compiler) เมื่อนำโปรแกรมไปใช้งาน หรือจะอัปเดตเวอร์ชันของโปรแกรม สามารถอัปเดตขึ้นไปทับไฟล์เดิมแล้วใช้งานได้ทันที ข้อเสียที่ต่างกันอย่างชัดเจนก็คือ กรณี Syntax ผิดจะรู้ก็ต่อเมื่อมีผู้ใช้งานเจอกับัก

### 2.2.6.1 ไวยากรณ์ของ PHP

การเขียนโปรแกรมด้วยภาษา php จะต้องอยู่ภายในขอบเขตของการเขียนโปรแกรม php ภายใน แท็กคำสั่ง `<?php ... ?>` ดังภาพ



ภาพที่ 2.16 ไวยากรณ์ของ PHP

### 2.2.7 ทฤษฎีเกี่ยวกับการ Visualization

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองเห็นได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวได้ว่า Visualization ก็คือการสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการ นำภาพมาใช้ในการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

Visualization เป็นมากกว่าวิธีการทาง Computer Visualization เป็นการนำเสนอข้อมูลแบบหนึ่งที่ทำกรแสดงข้อมูลให้อยู่ในรูปแบบของ Visual Form ซึ่งอาจจะเป็นการแสดงในรูปแบบของรูปภาพ, กราฟ หรือแผนภาพ ซึ่งผลลัพธ์ของการทำ Visualization คือ การนำข้อมูลต่าง ๆ ที่ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้สามารถสัมผัสได้ซึ่งกระบวนการต่าง ๆ ในการแสดงหรือนำข้อมูลต่าง ๆ ออกมา อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือใช้ในการวิเคราะห์ข้อมูล โดยจุดมุ่งหมายของ Visualization ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ เพื่อช่วยในการลดช่องว่างระหว่างผู้ใช้และตัวข้อมูล และยังช่วยผู้ใช้สามารถเรียนรู้ ได้อย่างมีประสิทธิภาพ

### 2.2.7.1 วิธีการ Visualization

การใช้ Visual เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้จะทำขั้นตอนหลักๆอยู่ 3 ขั้นตอน คือ Overview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้ต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้ก็จะทำการตัดสินใจเลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้อาจจะเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็จะช่วยอำนวยความสะดวกจากขั้นตอนเหล่านี้ซึ่ง Visualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ใช้งานต้องการโดยอาจจะใช้หลายๆวิธีการรวมกันเพื่อให้ได้ข้อมูลที่ผู้ใช้งานต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามารถนำมาผ่านกระบวนการของ Visualization มีลักษณะต่าง ๆ มากมายดังนี้ ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง, ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi Dimensional Data ได้แก่ Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อข่าวต่าง ๆ และ Web Document, Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document ,Algorithms และ Software ได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลาย

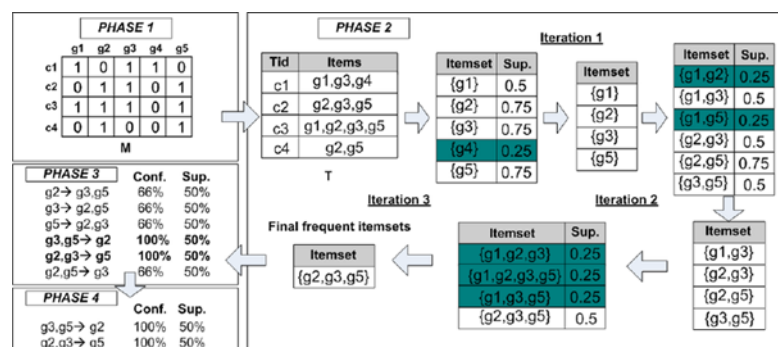
### 2.2.8 ทฤษฎีเกี่ยวกับกฎความสัมพันธ์ ด้วย Apriori Algorithm

ความสัมพันธ์ (Association) โดยหลักการทำงาน คือการหาความสัมพันธ์ของข้อมูลภายในกลุ่มข้อมูล เพื่อใช้ลักษณะของข้อมูลหนึ่งในการบอกถึงลักษณะที่จะเกิด ขึ้นกับอีกตัวหนึ่ง ซึ่งอาจจะเป็นการหาความสัมพันธ์ของข้อมูล ในกลุ่มเดียวกัน ตัวอย่างเช่น การระบุว่าในกลุ่มของนักเรียน ที่สมัครเรียนสาขาวิชาพยาบาลศาสตร์แล้วจะเป็นแพทย์หญิงนั้น มีโอกาสเกิดขึ้นร่วมกัน ผลลัพธ์ที่ได้จากการค้นหาความสัมพันธ์นั้นสามารถนำไปใช้ในการวิเคราะห์สนับสนุนหรือ พัฒนางานที่เกี่ยวข้องกับฐานข้อมูลนั้น ได้อย่างมีประสิทธิภาพมากยิ่งขึ้นซึ่งวิธีที่ได้รับความนิยมและเป็นที่ยอมรับ คือการค้นหาหาความสัมพันธ์ที่ปรากฏข้อมูลร่วมกันบ่อย

การค้นหาหาความสัมพันธ์ของข้อมูลในฐาน ข้อมูลได้พัฒนาขึ้นครั้งแรกโดยนักวิจัยจากศูนย์วิจัย IBM (International Business Machines Corporation) ประเทศสหรัฐอเมริกา มีจุดประสงค์เพื่อค้นหาความสัมพันธ์ที่น่าสนใจ ซึ่งซ่อนอยู่ในข้อมูลการซื้อขายสินค้าของลูกค้าว่าจะซื้อสินค้าใดบ้างร่วมกันในตะกร้ารถเข็นในห้างสรรพสินค้า (Market Basket Analysis) เพื่อให้ทราบถึงพฤติกรรมในการซื้อสินค้าของลูกค้า เช่นเมื่อลูกค้าซื้อนมแล้วจะซื้อขนมปังด้วย การ

ค้นหา กฎความสัมพันธ์มีขั้นตอนที่สำคัญที่สุดคือการค้นหากลุ่ม ข้อมูลที่ปรากฏร่วมกันบ่อยซึ่งจะมีขั้นตอนกระบวนการการทำงานนานที่สุด ดังนั้นในการเลือกขั้นตอนวิธีในการค้นหาความสัมพันธ์จึงเป็นสิ่งสำคัญเพื่อให้เหมาะสมกับลักษณะของ ข้อมูล เพื่อให้ลดระยะเวลา และเนื้อที่หน่วยความจำในการ ประมวลผลการทำงาน กระบวนการค้นหาความสัมพันธ์ มีขั้นตอนการค้นหาข้อมูลกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยทั้งหมด กลุ่มข้อมูลเหล่านั้นจะต้องมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่ผู้ใช้กำหนดจึงจะถือว่าเป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย และขั้นตอนการนำกลุ่มข้อมูลที่ปรากฏ บ่อยมาสร้างกฎความสัมพันธ์ซึ่งกฎความสัมพันธ์จะเป็นที่ ยอมรับได้หากตรวจสอบค่าความเชื่อมั่นของกฎนั้นมีค่า มากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่ผู้ใช้กำหนด

ขั้นตอนวิธี Apriori (Apriori Algorithm) เป็นขั้นตอนวิธีที่ได้รับการยอมรับและเป็นที่ยอมรับในการค้นหาความสัมพันธ์โดยเซตที่มีความถี่มากกว่าค่าสนับสนุนขั้นต่ำ เซตย่อยจะมีความถี่ที่มากกว่าและสามารถตัดเซตรายการ ที่มีความถี่ต่ำออก ขั้นตอนวิธีในการค้นหาเซตรายการที่ปรากฏร่วมกันบ่อยและนำ มาสร้างกฎความสัมพันธ์ข้อดีของขั้นตอนวิธี Apriori คือสามารถทำงานได้ดีหากกำหนดค่า สนับสนุนขั้นต่ำมีค่ามาก มีขนาดของฐานข้อมูลขนาดเล็ก และมีจำนวนของเซตรายการน้อย ส่วนข้อเสียคือต้องอ่านข้อมูลจากฐานข้อมูลหลายครั้ง



ภาพที่ 2.17 ขั้นตอนวิธี Apriori

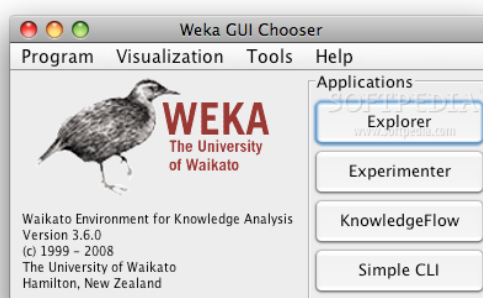
## 2.2.9 ทฤษฎีเกี่ยวกับเทคนิคเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์ ด้วย Pandas

Pandas เป็น Library ใน Python ที่ทำให้เราเล่นกับข้อมูลได้ง่ายขึ้น เหมาะมากสำหรับทำ Data Cleaning / Wrangling วิธีการใช้งาน Pandas คือ โหลดไฟล์ข้อมูล เช่น CSV เข้าไป แล้วเราจะได้ข้อมูลในรูปแบบตาราง (DataFrame) ที่แบ่งข้อมูลตามแถวและคอลัมน์ หรือเหมือน Excel ที่เราใช้กันนั่นเอง

## 2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

### 2.3.1 โปรแกรม Weka

โปรแกรม Weka (Waikato Environment for Knowledge Analysis) : เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จภาพประกอบประเภทฟรีแวร์ อยู่ภายใต้การควบคุมของ GPL License ซึ่งโปรแกรม Weka ได้ถูกพัฒนามาจากภาษาจาวาทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และ การทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยโมดูลย่อย ๆ สำหรับใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่สามารถใช้ Graphic User Interface (GUI) และ ใช้คำสั่งในการให้ซอฟต์แวร์ประมวลผล และสามารถรัน (run) ได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทำงานในด้านการทำคัต้าไมน์นิ่งที่รวบรวมแนวคิดอัลกอริทึมมากมาย ซึ่งอัลกอริทึมสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคือจากชุดเครื่องมือที่มีอัลกอริทึมมาให้ หรือเลือกใช้จากอัลกอริทึมที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล ได้แก่ Pre-Processing, Classification, Regression, Clustering, Association rules, Selection และ Visualization ส่วนของการ Import ข้อมูล โดยข้อมูลนั้นจะต้องอยู่ในรูปแบบ ASCII อาจเป็น arff , csv , c45



ภาพที่ 2.18 โปรแกรม Weka

โดยโปรแกรม Weka สามารถแยกประเภทการทำเหมืองข้อมูลได้ดังนี้ กฎความสัมพันธ์ (Association rule) แสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุ ที่เกิดขึ้นพร้อมกัน ตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยง เช่น การวิเคราะห์ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย(POS) หรือร้านค้าออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าคนที่ซื้อเทปวิดีโอมักจะซื้อเทปกาวยด้วย ร้านค้าก็อาจจะจัดร้านให้สินค้าสองอย่างอยู่ใกล้กัน



เพื่อเพิ่มยอดขาย หรืออาจจะพบว่าหลังจากคนซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็ สามารถนำความรู้นี้ไปแนะนำผู้ที่กำลังจะซื้อหนังสือ ก ได้

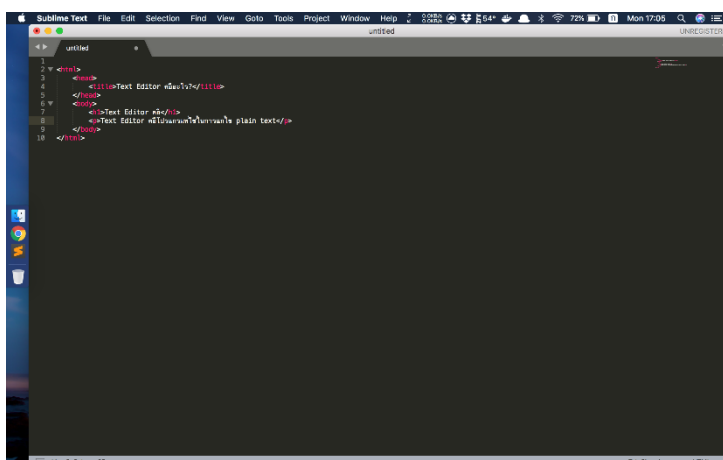
การจำแนกประเภทข้อมูล (Data classification) หากฎเพื่อระบุประเภทของวัตถุจาก คุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือ การวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อ ประกอบการพิจารณาการอนุมัติเงินกู้

การแบ่งกลุ่มข้อมูล (Data clustering) แบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หา สาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน

จินตทัศน์ (Visualization) สร้างภาพคอมพิวเตอร์กราฟิกที่สามารถนำเสนอข้อมูล มากมายอย่างครบถ้วนแทนการใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ้อน เร้นเมื่อดูข้อมูลชุดนั้นด้วยจินตทัศน์

### 2.3.2 โปรแกรม Sublime text 3

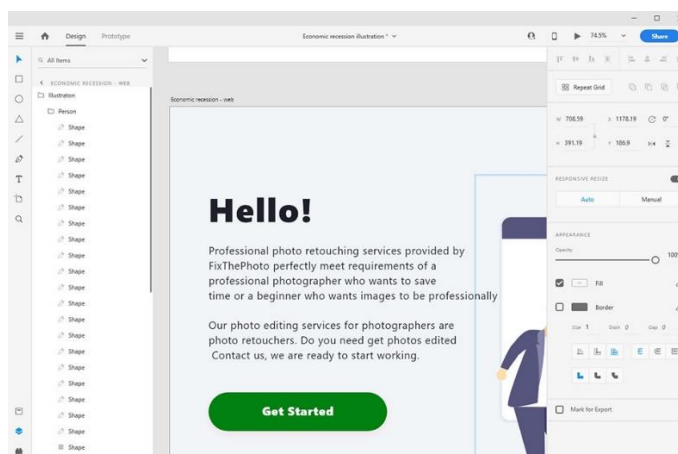
เป็นโปรแกรมแก้ไขข้อความและเขียนโค้ด สำหรับโปรแกรมเมอร์และนักพัฒนาที่มี ขนาดเล็ก เพียงแค่ 7MB เท่านั้น แต่ยังคงประสิทธิภาพและฟังก์ชันไว้มากมาย ใช้งานง่ายและ ทำงานได้อย่างรวดเร็ว ซึ่งสนับสนุนภาษาที่หลากหลาย C, C++, C#, CSS, D, Erlang, HTML, Groovy, Haskell, HTML, Java, JavaScript, LaTeX, Lisp, Lua, Markdown, Matlab, OCaml, Perl, PHP, Python, R, Ruby, SQL, TCL, Textile และ XML



ภาพที่ 2.19 โปรแกรม Sublime text 3

### 2.3.3 โปรแกรม Adobe XD

Adobe XD คือ แอปพลิเคชันบนคอมพิวเตอร์ ที่ช่วยให้นักออกแบบเริ่มทำงานตั้งแต่การสร้างไวร์เฟรม ไปจนถึงการสร้างตัวต้นแบบเพื่อทดลองใช้เพื่อเก็บข้อมูลกลับมาปรับปรุงการใช้งานให้ดีขึ้น เพื่อที่ผู้ใช้จะได้กลับมาใช้งานอีกมากขึ้น เวอร์ชันฟรีวิวแรกที่ทำให้ทดลองใช้ครั้งนี้ โหมดออกแบบ (Design) สามารถทำงานออกแบบได้อย่างง่ายดาย โหมดต้นแบบ (Prototype) ก็สามารถกำหนด Hotspots และ Transitions ได้ตามต้องการ โหมดพรีวิว (Desktop Preview) สำหรับทดลองใช้งานเพื่อดูความเปลี่ยนแปลงจริงได้ นอกจากนี้ยังสามารถแชร์พรีวิว ให้ผู้ตัดสินใจและคนอื่นได้ทดลองใช้งานได้ ในบราวเซอร์บนคอมพิวเตอร์หรือบนอุปกรณ์พกพาอีกด้วย นอกจากนี้ นักออกแบบยังสามารถนำงานที่สร้างขึ้นในแอปพลิเคชันสำหรับการออกแบบอย่าง Illustrator CC และ Photoshop CC มาใช้งานร่วมกันได้ และเมื่อการออกแบบและทดลองใช้มาถึงขั้นสุดท้ายแล้ว ทุกอย่างก็สามารถเอ็กซ์พอร์ตไปให้นักพัฒนาทำงานต่อไปได้อย่างง่ายดาย



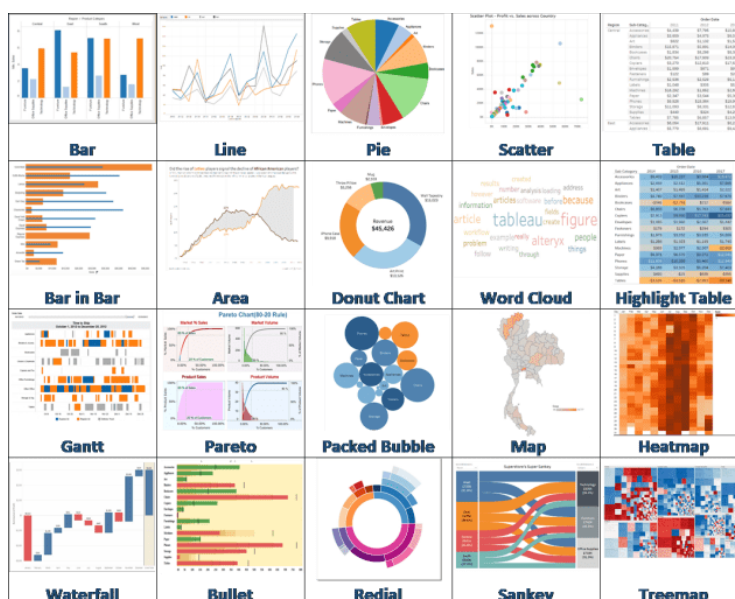
ภาพที่ 2.20 โปรแกรม Adobe XD

### 2.3.4 โปรแกรม Tableau

Tableau (อ่านว่า ทาโบลว หรือ แท็บโบลว) เป็นคำศัพท์ที่มีรากศัพท์มาจากภาษาฝรั่งเศส มีความหมายเกี่ยวกับ รูปภาพ ภาพวาด หรือการระบายสี เป็นต้น แต่ในบทความนี้จะพูดถึง Tableau ในมุมมองของซอฟต์แวร์ (Tableau Software) ซึ่งเป็นซอฟต์แวร์ที่ถูกกล่าวถึงเป็นอย่างมากในวงการ BI, Analytics และ Big Data ในบทความนี้จะขออธิบายว่าซอฟต์แวร์ Tableau คืออะไรในหลาย ๆ แง่มุม เพื่อให้ผู้อ่านเกิดความเข้าใจได้อย่างถูกต้องและครบถ้วน “Tableau ช่วยคนให้เห็นและเข้าใจข้อมูล” (Tableau helps people see and understand data) นี้

เป็นพันธกิจ (Mission) ที่มีมาตั้งแต่ต้น และไม่เคยเปลี่ยนแปลงของ Tableau บริษัทก่อตั้งในปี 2003 โดยผู้ก่อตั้ง 3 คนจากมหาวิทยาลัย Stanford คือ 1) Pat Hanrahan[1] ศาสตราจารย์ สาขา computer graphics มหาวิทยาลัย Stanford 2) Chris Stolte นักศึกษาซึ่งทำวิทยานิพนธ์ กับ Hanrahan เป็นผู้สร้าง Tableau และเป็น chief development officer และ 3) Christian Chabot นักศึกษา MBA ของ Stanford ผู้ที่เป็น CEO ในขณะก่อตั้งบริษัทตลาดของ Tableau คือทุกคนที่ทำงานกับข้อมูล ในตอนต้น Tableau ตั้งใจที่จะเสนอซอฟต์แวร์ที่ดีและมีประสิทธิภาพมากกว่า Microsoft Excel และใช้งานและเข้าถึงได้ง่ายกว่าซอฟต์แวร์ BI จากบริษัทยักษ์ใหญ่เช่น IBM, Oracle หรือ SAP ในช่วงแรกแล้วมีอุปสรรคมากมาย แต่เมื่อคนได้เห็นและทดลองใช้งาน รับรู้ถึงความง่ายและความสามารถทำให้เห็นและเข้าใจข้อมูลได้จริง ตลาดจึงยอมรับ Tableau จนกระทั่งตอนนี้ Tableau จึงประสบความสำเร็จอย่างสูง หลักฐานจาก Gartner[2] ที่จัดอันดับ Tableau ให้อยู่ใน Quadrant LEADERS เป็นปีที่ 6 ติดต่อกัน

Tableau ทำหน้าที่แปลงข้อมูล (Data) ให้เป็นภาพ (Visualization) ตัวอย่างของการแปลงข้อมูลเป็นภาพเป็นสิ่งที่เราค้นเคย และมีมาเนิ่นนานแล้วก็คือการสร้างกราฟนั่นเอง จะเห็นได้ว่าเรามีการสร้าง ตาราง กราฟแท่ง กราฟเส้น กราฟวงกลม (pie) มาก่อนแล้ว กราฟที่กล่าวมาถือเป็น Visualization อย่างหนึ่ง แต่ก็ยังมีกราฟรูปแบบอื่น ๆ อีกมากมายที่มีประสิทธิภาพและทรงพลัง ในการสื่อสารมากกว่า รูปด้านล่างคือตัวอย่าง Visualization ที่สร้างโดย Tableau



ภาพที่ 2.21 โปรแกรม Tableau

## 2.4 วรรณกรรมที่เกี่ยวข้อง

อังคณา พิจารโชติ (2552) ได้พัฒนาระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็นโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล เพื่อช่วยในการวิเคราะห์หาความสัมพันธ์ของปัจจัยเสี่ยงต่าง ๆ ที่ก่อให้เกิดโรคเบาหวาน เช่น พ่อแม่มีประวัติเป็นโรคเบาหวาน มีประวัติการปัสสาวะบ่อย เป็นต้น ข้อมูลความสัมพันธ์ของปัจจัยเสี่ยงที่ได้จากการวิเคราะห์จะเป็น สารสนเทศสำคัญที่เป็นตัวช่วยให้หน่วยงานด้านการแพทย์นำไปใช้สำหรับวางแผนป้องกันโรค โดยระบบที่พัฒนานี้จะประกอบด้วยสามส่วนคือ ส่วนของคลังข้อมูลใช้สำหรับเก็บข้อมูลปัจจัยเสี่ยงที่ได้จากการคัดกรองกลุ่มเสี่ยง ส่วนของเหมืองข้อมูลเป็นตัวค้นหาความสัมพันธ์ของปัจจัยเสี่ยงต่าง ๆ และส่วนการแสดงผลข้อมูลรายงานซึ่งจะจัดทำในรูปแบบของเว็บแอปพลิเคชันเพื่อให้ สะดวกต่อการเรียกใช้งาน ผลการค้นหาความสัมพันธ์ด้วยค่าสนับสนุนที่ 0.5 และค่าความเชื่อมั่นที่ 0.3 พบว่าประวัติครอบครัวเป็นปัจจัยเสี่ยงที่มีอัตราการเสี่ยงมากที่สุดซึ่งพบคู่กับค่าดัชนีมวลกาย มี ค่าความเชื่อมั่นที่ 0.94

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่าทางผู้จัดทำจะนำการสนับสนุนการตัดสินใจสำหรับวิเคราะห์ ปัจจัยเสี่ยงการเป็นโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล เพื่อช่วยในการวิเคราะห์หาความสัมพันธ์ของปัจจัยเสี่ยงต่าง ๆ เพื่อมาเป็นตัวช่วยในการสร้างการวิเคราะห์พฤติกรรมกรรมการซื้ออาหารญี่ปุ่นโดยใช้กฎความสัมพันธ์ด้วยอัลกอริทึมเอ็พริออริ

ศุภกรใจ วุฒิกิจโกศล (2551) ใช้เทคนิคการทำเหมืองข้อมูลในผู้ป่วยข้อไหล่ติด โรงพยาบาลพระนั่งเกล้าเพื่อหาปัจจัยที่มีผลต่อการรักษาและสร้างกฎความสัมพันธ์ที่น่าสนใจ เพื่อ เป็นแนวทางช่วยสนับสนุนการตัดสินใจเลือกเทคนิคการรักษาสำหรับนักกายภาพบำบัด โรงพยาบาลพระนั่งเกล้า โดยใช้ข้อมูลผู้ป่วยโรคข้อไหล่ติดที่ไม่มีโรคประจำตัวและไม่ได้เกิดจากอุบัติเหตุมารับการรักษาทางกายภาพบำบัดโรงพยาบาลพระนั่งเกล้าที่รักษาหายในช่วงปี พ.ศ. 2548- 2550 จำนวน 550 ราย มีการเตรียมข้อมูลเพื่อให้ได้ข้อมูลที่ถูกต้อง และหาปัจจัยที่มีผลต่อการรักษา จากนั้นวิเคราะห์โดยการทำเหมืองข้อมูลด้วยเทคนิค Clustering และ Association Rule เลือกกฎ ความสัมพันธ์ที่น่าสนใจด้วยค่าสนับสนุนของกฎไม่ต่ำกว่าร้อยละ 20 และค่าความเชื่อมั่นไม่ต่ำกว่า ร้อยละ 90 ผลการศึกษาพบว่า มีปัจจัย อายุ ระยะเวลาที่เป็น อองศาการยกแขน การไขว้หลัง ระดับ ความเจ็บปวด และเทคนิคการรักษาที่มีผลต่อการรักษา ตัวอย่างเช่น เมื่อใช้เทคนิค Association Rule หากกฎความสัมพันธ์ของศาการยกแขน 90-120 เลือกเทคนิคการรักษาด้วยแผ่นประคบความร้อน บริเวณคอ ความร้อนคลื่นเหนือเสียงบริเวณ

ไหล่ด้านหน้า ความร้อนคลื่นเหนือเสียงบริเวณ กล้ามเนื้อ Infraspinatus และกระตุ้นไฟฟ้าแบบ Surge บริเวณไหล่ด้านหน้า-หลังจะใช้จำนวนครั้ง ในการรักษาช่วง 19-24 ครั้งด้วยค่า สัมประสิทธิ์ของกฎ 43.48 เปอร์เซนต์และค่าความเชื่อมั่นที่ 100%

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่าทางผู้จัดทำจะนำการใช้เทคนิคการ ทำเหมืองข้อมูลในผู้ป่วยข้อไหล่ติด โรงพยาบาลพระนั่งเกล้าเพื่อหาปัจจัยที่มีผลต่อการรักษา และสร้างกฎความสัมพันธ์ที่น่าสนใจเพื่อ เป็นแนวทางช่วยสนับสนุนการตัดสินใจเลือกเทคนิค การรักษาสำหรับนักกายภาพบำบัด โรงพยาบาลพระนั่งเกล้ามาใช้ เพื่อมาเป็นตัวช่วย และเป็น แนวทางในการสร้างการวิเคราะห์พฤติกรรมการซื้ออาหารญี่ปุ่นโดยใช้กฎความสัมพันธ์ด้วย อัลกอริทึมเอโพรออริ

สมฤทัย คูหาอยู่สุขสม (2551) งานวิจัยชิ้นนี้นำเสนอการสร้างฐานข้อมูลโภชนาการที่ สำหรับใช้เป็นในชีวิตประจำวันบนเว็บเพื่อใช้ในการทำนายรายการอาหารที่ควรหาเพิ่มเติมซึ่งใช้ เทคนิคของคาน่าไมน์นิ่งแบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ซึ่งจะมี ประสิทธิภาพมากกว่าตัวจำแนกประเภทข้อมูลที่ใช้เทคนิคแบบอื่นสำหรับในกรณีที่ต้องการ จำแนกข้อมูลออกเป็นหลายประเภทจากผลการวิเคราะห์ระบบพบว่าระบบการคำนวณ สารอาหารมีความถูกต้องแม่นยำสามารถแสดงผลลัพธ์ออกมาเป็นตารางและกราฟได้อย่าง เข้าใจง่ายและสวยงามส่วนระบบวิเคราะห์สามารถให้คำตอบเป็นประเภทอาหารได้และยัง จำแนกเป็นรายการอาหารได้หลากหลายอีกด้วย

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่าทางผู้จัดทำจะนำการสร้าง ฐานข้อมูลโภชนาการที่สำหรับในชีวิตประจำวันบนเว็บเพื่อใช้ในการทำนายรายการอาหารที่ควร หาเพิ่มเติม เพื่อมาเป็นตัวช่วย และเป็นแนวทางในการสร้างการวิเคราะห์พฤติกรรมการซื้อ อาหารญี่ปุ่นโดยใช้กฎความสัมพันธ์ด้วยอัลกอริทึมเอโพรออริ

สุขสันต์ บัวฉิม (2550) งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์หาความสัมพันธ์ระหว่าง รายวิชาจากผลการเรียนของนักศึกษานั้นคือเมื่อนักศึกษามีผลการเรียนในรายวิชาหนึ่งแล้ว จะต้องมีการเรียนต่่าในรายวิชาอื่นด้วยซึ่งในงานวิจัยนี้ได้นำเทคนิคแอสโซซิเอชันรูลมมาใช้ใน การหากฎความสัมพันธ์โดยมีการเปลี่ยนลักษณะจากปกติคือเมื่อไอเท็ม A ดีแล้วไอเท็ม B จะดี ด้วยเป็น (ไอเท็ม A ไม่ดีแล้วไอเท็ม B จะดี) หรือ (ไอเท็ม A ดีแล้วไอเท็ม B จะไม่ดี) หรือ (ไอเท็ม A ไม่ดีแล้วไอเท็ม B จะไม่ดีด้วย) เป็นการเพิ่มลักษณะการหาความสัมพันธ์ของข้อมูลที่น่าสนใจ

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่าทางผู้จัดทำจะนำงานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์หาความสัมพันธ์ระหว่างรายวิชาจากผลการเรียนของนักศึกษา เพื่อมาเป็นตัวช่วย และเป็นแนวทางในการสร้างการวิเคราะห์พฤติกรรมการซื้ออาหารญี่ปุ่นโดยใช้กฎความสัมพันธ์ด้วยอัลกอริทึมเอโพรออริ

ดังนั้น วรรณกรรมข้างต้นนั้นจะเป็นการทราบถึงข้อมูลการวิเคราะห์หาความสัมพันธ์ของงานวิจัยของบุคคลอื่น ๆ ที่ได้หยิบยกเทคนิคการค้นหาความสัมพันธ์ เพื่อศึกษาและนำมาเป็นแบบอย่างในการวิเคราะห์ข้อมูลของคณะผู้จัดทำ และเพื่อให้เป็นแนวทางในการตัดสินใจได้ว่าควรที่จะทำการวิเคราะห์ปัญหาไปในทิศทางใดและควรวิเคราะห์ส่วนใดบ้าง เพื่อความถูกต้องของข้อมูลและวิเคราะห์หรือคาดคะเนข้อมูลในอนาคตต่อไปได้

## 2.5 บทสรุป

ข้อมูลได้เลือกใช้การทำเหมืองข้อมูลแบบ Association rule ด้วยเทคนิค Apriori เพื่อสร้างชุดข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม Weka 3.8.4 และชุดคำสั่ง Python ในการสร้าง Association rule และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผล แบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน Web Browser ที่เป็นที่นิยมในยุคอินเทอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา