

## บทที่ 2

### แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

โครงการเรื่อง การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

#### 2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

#### 2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS และ Bootstrap

2.2.5 ทฤษฎีเกี่ยวกับการ visualization

2.2.6 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ

#### 2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

2.3.2 การประเมินผลโมเดล (Decision Tree)

2.3.3 การแสดงผลแบบ visualization

#### 2.4 วรรณกรรมที่เกี่ยวข้อง

#### 2.5 บทสรุป

## 2.1 แนวคิด

### 2.2.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 - 2558 และการแสดงผลข้อมูลบนเว็บไซต์ คณะผู้จัดทำได้ศึกษาหลัก การ และทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

ความหมายของการวิเคราะห์ข้อมูล

เอ็ดมุนด์ หลินเจริญ (2555) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเชิงคุณภาพ นับเป็นขั้นตอนที่สำคัญในกระบวนการวิจัย วิธีการหลักที่ใช้ในการ วิเคราะห์ข้อมูลเชิงคุณภาพ เป็นวิธีการสร้างข้อสรุปจากการศึกษาจากข้อมูลจำนวนหนึ่ง ซึ่งมักไม่ใช้สถิติในการวิเคราะห์หรือถ้าใช้สถิติก็ไม่ได้ถือว่าสถิติเป็นวิธีการวิเคราะห์หลัก แต่จะ ถือเป็นข้อมูลเสริมในการวิเคราะห์ ข้อมูลเชิงคุณภาพนั้น คณะผู้จัดทำนับเป็นผู้ที่มีบทบาท สำคัญยิ่งและควรมีความรอบรู้ในเรื่องแนวคิดทฤษฎีอย่างกว้างขวาง มีความเป็นสหวิทยาการ อยู่ในตัวเอง มีความสามารถทางภาษาสามารถเชื่อมโยงข้อความและสร้างข้อสรุปเป็นกรอบ แนวคิดและตีความหมายของข้อมูลได้หลาย ๆ แบบ ทั้งนี้การวิเคราะห์ข้อมูลเชิงคุณภาพนั้น อาจไม่จำเป็นต้องใช้เฉพาะการวิจัยเชิงคุณภาพเท่านั้น แต่อาจใช้กับการวิจัยเชิงปริมาณที่มีการ เก็บรวบรวมข้อมูลเชิงคุณภาพ เช่น แบบสอบถามปลายเปิด การสัมภาษณ์ การสังเกต จุด บันท์ก็มาทำการวิเคราะห์เพื่อให้ข้อมูลที่รวบรวมมา มีความหมายและตอบคำถามหรือ จุดมุ่งหมายของการวิจัยสำหรับสาระในการวิเคราะห์ข้อมูลเชิงคุณภาพ โดยเทคนิคการ วิเคราะห์ข้อมูลเชิงคุณภาพที่สำคัญ ๆ และนักวิจัย นิยมใช้ ได้แก่ การจำแนกหรือการจัดกลุ่ม ข้อมูล การเปรียบเทียบเหตุการณ์ การวิเคราะห์ส่วนประกอบ การวิเคราะห์แบบอุปนัย และ การวิเคราะห์ข้อมูลที่เป็นเอกสาร

หัสพร ทองแดง, พิศุทธิภา เมธิกุล, สมคิด หุมวงศ์ และแพรว สมบัติใหม่ (2559) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเป็นการจำแนกข้อมูล ให้เป็นระบบ เพื่อทำให้เกิดความเข้าใจในความหลากหลาย ความหมายและความสัมพันธ์ของ ข้อมูลในบริบทของสังคมและวัฒนธรรมหนึ่ง ๆ ในการวิเคราะห์ข้อมูลเชิงคุณภาพนี้หลังจากที่มี การแยกข้อมูลออกได้เป็นหมวดหมู่แล้ว นักวิจัยต้องใช้วิธีการนำเสนอโดยการพรรณนาให้เห็น สภาพเงื่อนไข กระบวนการ ขั้นตอนการสัมพันธ์ต่าง ๆ รวมทั้งพยายามหาความหมายทาง วัฒนธรรมของปรากฏการณ์ในทัศนะของบุคคลหรือกลุ่มตัวอย่างซึ่งเป็นผู้ให้ข้อมูล และอาจมี ความเข้าใจและตีความแตกต่างกันออกไปหรือแตกต่างกันไปจากของผู้วิจัยเอง การวิเคราะห์ ข้อมูลสำหรับการวิจัยเชิงคุณภาพ จึงเป็นเสมือนการทำความเข้าใจในแบบแผนและความหมาย

ของพฤติกรรมมากกว่าที่จะมุ่งหาระดับความมากน้อยหรือระดับความเข้มข้นของความสัมพันธ์ของปรากฏการณ์จึงเป็นการวิเคราะห์ข้อมูลที่อาศัยเครื่องมือทางความคิด (Conceptual tools) มากกว่าที่จะใช้เครื่องมือทางสถิติ และที่สำคัญต้องอาศัยความสามารถของนักวิจัยที่จะอธิบายให้เห็นความสัมพันธ์และความหมายของปรากฏการณ์บนพื้นฐานความเข้าใจในวัฒนธรรมและสังคมที่เป็นบริบทของปรากฏการณ์ที่ศึกษา

จุมพล หนิมพานิช และวรพลวัฒน์ โจรจนพล (2561) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเชิงคุณภาพโดยทั่วไปหมายถึง การสรุปเนื้อหาของสาระของข้อมูล และหาแบบแผนของความสัมพันธ์ระหว่างข้อมูลหรือตัวแปร หรือหมายถึงการสรุปพรรณนาและการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูล หรือหมายถึง การนำเอาข้อมูลที่ได้จากการค้นคว้าวิจัยมาจัดกระทำให้เป็นระบบและหาความหมาย แยกแยะ องค์ประกอบ รวมทั้งเชื่อมโยงและหาความสัมพันธ์ของข้อมูลเพื่อให้สามารถนำไปสู่ความเข้าใจต่อการดำรงอยู่และการเปลี่ยนแปลงของปรากฏการณ์ที่ศึกษา หรือหมายถึง การแยกแยะ การตีความหมาย การเปรียบเทียบ การหาความสัมพันธ์เกี่ยวกับปรากฏการณ์ การหาแบบแผนการอธิบายและการสรุปเกี่ยวกับกิจกรรมหรือพฤติกรรมในสภาพสังคมหรือปรากฏการณ์ทางสังคมที่ศึกษา

### 2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

#### ความหมายของการประยุกต์ใช้ข้อมูลทางสถิติ

จากเอกสารประกอบการสอนการวิจัยสำหรับครู ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า คำว่าสถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistics มีรากศัพท์มาจาก Stat หมายถึง ข้อมูลหรือสารสนเทศ ซึ่งจะอำนวยความสะดวกต่อการบริหารประเทศในด้านต่าง ๆ เช่น การทำสำมะโนครัวเพื่อจะทราบจำนวนพลเมืองในประเทศทั้งหมด ในสมัยต่อมา คำว่า สถิติ ได้หมายถึง ตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวม เช่น จำนวนผู้ประสบอุบัติเหตุบนท้องถนน อัตราการเกิดของเด็กทารก ปริมาณน้ำฝนในแต่ละปี เป็นต้น สถิติในความหมายที่กล่าวมานี้เรียกอีกอย่างหนึ่งว่า ข้อมูลทางสถิติ (Statistical data)

1) ค่ากึ่งกลาง (Median) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางวิธีหนึ่งที่ใช้การเรียงค่าการสะท้อนของจุดภาพจากค่าน้อยที่สุดไปหาค่ามากที่สุด โดยค่ากึ่งกลางเป็นค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งหมด ค่ากึ่งกลางจึงเป็นตัวแทนค่าการสะท้อนของจำนวนจุดภาพทั้งหมดในช่วงคลื่นหนึ่งๆ ที่แสดงให้เห็นว่ามีจำนวนจุดภาพที่มีค่าการสะท้อนมากกว่าและน้อยกว่าค่ากึ่งกลางอยู่ประมาณร้อยละ 50

2) ค่าฐานนิยม (Mode) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางอีกวิธีหนึ่ง โดยดูจากจำนวนความถี่ของค่าการสะท้อนซึ่งมีความถี่สูงที่สุด นิยมนำมาใช้กับข้อมูลที่เป็นนามบัญญัติ

เช่น ค่าของประเภทข้อมูลหลังจากการจำแนกประเภทแล้ว ถือเป็นค่าการสะท้อนที่แสดงการใช้ที่ดินประเภทต่าง ๆ ไม่ใช่ค่าการสะท้อนของวัตถุอีกต่อไป

3) ความเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D.) เป็นการวัดการกระจายที่นิยมใช้มากที่สุดการคำนวณใช้วิธียกกำลังสองของผลต่างระหว่างค่าการสะท้อนของทุกจุดภาพในแต่ละช่วงคลื่นกับค่าเฉลี่ยเลขคณิตของช่วงคลื่นนั้น

ภทรธิดา ผลงาม (2558) ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า สถิติ คือ ตัวเลขต่าง ๆ ที่ได้มีการรวบรวมขึ้นเพื่อบอกข้อเท็จจริงเกี่ยวกับคุณสมบัติหรือลักษณะบางสิ่งบางอย่างที่สามารถแสดงออกเป็นตัวเลขได้สถิติแบ่งเป็นประเภทใหญ่ ๆ ได้ 2 ประเภท ดังนี้ คือ

1) สถิติพรรณนา (Descriptive statistics) คือสถิติที่บรรยายถึงลักษณะของข้อมูลเฉพาะกลุ่มนั้น ๆ โดยไม่สรุปอ้างอิงไปยังประชากรกลุ่มอื่น ๆ สถิติประเภทนี้นิยมศึกษาในกลุ่มเล็กหรือกลุ่มใหญ่ก็ได้ สถิติประเภทนี้เป็นสถิติที่บรรยายลักษณะของข้อมูล เช่น ค่าเฉลี่ย ร้อยละ มัธยฐาน พิสัย ส่วนเบี่ยงเบนมาตรฐาน และการวัดความสัมพันธ์ต่าง ๆ เช่น สหสัมพันธ์

2) สถิติอ้างอิงหรือสถิติอนุมาน (Inferential or inductive statistics) คือ สถิติที่นำค่าสถิติพรรณนามาสรุปอ้างอิงไปยังประชากร หรือเป็นสถิติที่ศึกษาจากกลุ่มตัวอย่างแล้วนำผลอ้างอิงไปยังกลุ่มประชากร เช่น การศึกษาความคิดเห็นต่อการเลือกตั้งของประชาชนไม่จำเป็นต้องศึกษาจากประชาชนทุกคน แต่สามารถเลือกศึกษาจากประชาชนบางกลุ่มซึ่งจะเป็นตัวแทนของประชาชนทั้งหมด แล้วจึงสรุปว่าประชาชนมีความคิดเห็นอย่างไรต่อการเลือกตั้งได้ ดังนั้นในการใช้สถิติอ้างอิงนี้จึงจำเป็นต้องเลือกกลุ่มตัวอย่างที่เหมาะสม สถิติอ้างอิงนี้ก่อนนำไปอ้างอิงกลุ่มประชากรต้องมีการทดสอบทางสถิติก่อนทุกครั้งจึงสามารถอ้างอิงประชากรได้อย่างมีประสิทธิภาพ

สุทิน ชนะบุญ (2555) ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า สถิติประกอบด้วย 2 องค์ประกอบ ดังนี้ สถิติเชิงพรรณนา (Descriptive Statistics) เป็นวิธีการทางสถิติเพื่อใช้ในการพรรณนาหรือบรรยายลักษณะของสิ่งที่ศึกษา เพื่อให้เข้าใจถึงลักษณะของข้อมูลที่เก็บรวบรวมมาได้ จะพรรณนาภายในขอบเขตของข้อมูลที่เก็บรวบรวมมาเท่านั้น ไม่สามารถจะคาดคะเนลักษณะต่าง ๆ ออกไปนอกเหนือจากข้อมูลที่มีอยู่ได้ หรือไม่มีการอ้างอิงหรืออนุมานไปถึงกลุ่มอื่น หากผู้วิจัยสามารถศึกษาทุกหน่วยของประชากรได้ ก็จะใช้สรุปหรือบรรยายลักษณะของประชากรที่ศึกษา หากผู้วิจัยไม่สามารถศึกษาสมาชิกทุกหน่วยของประชากรได้ สุ่มสมาชิกเพียงบางส่วนมาศึกษา(ศึกษาจากกลุ่มตัวอย่าง) ก็จะใช้สรุปหรือบรรยายเฉพาะกลุ่มตัวอย่างที่สุ่มมาศึกษาเท่านั้น สถิติเชิงอ้างอิงหรือสถิติอนุมาน (Inferential Statistics) เป็นวิธีการทางสถิติที่ใช้ในการสรุปลักษณะของประชากร จากผลการศึกษาข้อมูลในกลุ่ม ตัวอย่าง โดยอาศัย

ทฤษฎีความน่าจะเป็นในการอนุมานลักษณะประชากร โดยศึกษาจากกลุ่มตัวอย่าง แต่ อ้างอิงหรืออนุมานไปถึงประชากร ประกอบด้วย 2 องค์ประกอบ คือ การประมาณค่าพารามิเตอร์ (Parametric Estimation) และการทดสอบสมมติฐาน (Testing Hypothesis)

### 2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

#### ความหมายการแสดงผลข้อมูล

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

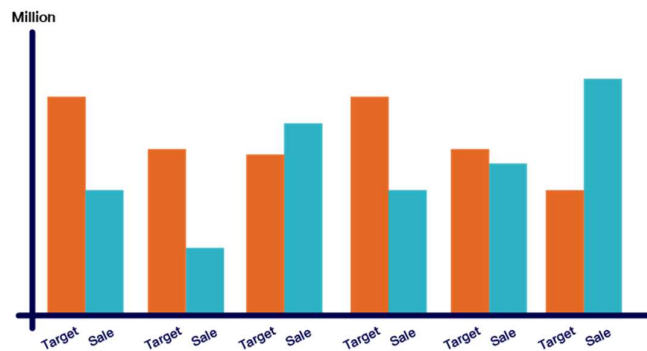
ชนาธิป ชื่นมนัส (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) คือวิธีการที่ใช้ในการนำข้อมูลที่เป็นนามธรรมมาแสดงให้เป็นภาพ เพื่อเพิ่มประสิทธิภาพในการติดต่อสื่อสาร หรือทำให้เข้าใจข้อมูลได้ง่ายขึ้น การใช้เทคนิค Visualization มีหลายวิธี ซึ่ง แต่ละวิธีเหมาะสมที่จะใช้ในการแสดงข้อมูลที่แตกต่างกัน การเลือกใช้เทคนิคใดนั้นขึ้นกับองค์ประกอบหลายอย่าง เช่น ชนิดของข้อมูลที่จะนำมาแสดง วิธีที่ต้องการแสดงผล เป็นต้น ในปัจจุบันมีงานวิจัยที่นำเอาเทคนิค Visualization มาใช้โดยมีจุดประสงค์เพื่อเพิ่มประสิทธิภาพในการแสดงผลการติดต่อสื่อสารภายในเครือข่ายคอมพิวเตอร์ และช่วยในการวิเคราะห์ความผิดปกติที่เกิดขึ้นในเครือข่าย

อาทิตย์ สิทธิบรรเจิด (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงผลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้งานสามารถเรียนรู้และจดจำข้อมูลผ่านการมองเห็นได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวได้ว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาก็คือได้กลายเป็นการนำภาพมาใช้ในการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล ในปัจจุบันเป็นยุคเทคโนโลยีเข้าถึงทุกคน ทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่าย และรวดเร็วมากขึ้น คนที่นำเสนอข้อมูลจึงต้องนำเสนอข้อมูลที่น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visualization ขึ้นมา Data Visualization เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ซึ่งอาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่น ๆ อีกมากมาย เพื่อให้เข้าใจได้

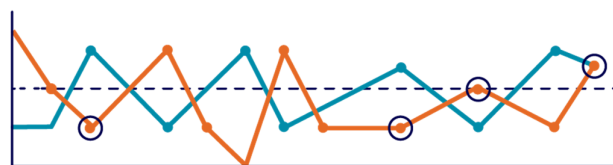
ง่าย และรวดเร็ว การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูลที่มีอยู่ เพื่อให้การนำเสนอข้อมูลน่าสนใจและมีประสิทธิภาพสูงสุด ซึ่งเลือกแผนภูมิชนิดนิยมที่ใช้กันอย่างแพร่หลาย ดังนี้

1) แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้งที่นิยมแสดงออกมาในรูปแท่งสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับการเปรียบเทียบจำนวนของข้อมูลในแต่ละชุดเช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริงเปรียบเทียบกับเป้าหมายยอดขายที่ตั้งไว้ เป็นต้นซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท • แผนภูมิแท่งแบบจัดกลุ่ม แผนภูมิแท่งแบบจัดกลุ่ม เหมาะกับการนำเสนอข้อมูลที่มีข้อมูลย่อย ๆ อยู่ภายใต้ข้อมูลใหญ่เป็นการเน้นให้เห็นข้อมูลย่อยนั้น ๆ แผนภูมิแท่งแบบวางซ้อนกัน แผนภูมิแท่งแบบวางซ้อนกัน เหมาะกับการนำเสนอข้อมูลที่ต้องการให้เห็นข้อมูลย่อยในแต่ละข้อมูลใหญ่และยังแสดงให้เห็นสัดส่วนของข้อมูลย่อยต่าง ๆ เหล่านั้นได้ด้วย ใช้แผนภูมินี้เมื่อมีชุดข้อมูล หลายชุดและต้องการเน้นผลรวมทั้งหมด



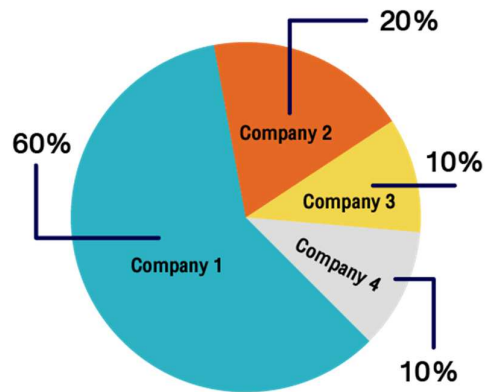
ภาพที่ 2.1 แผนภูมิแท่ง (Bar Charts)

2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั้นเองแผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดูแนวโน้ม เป็นต้น



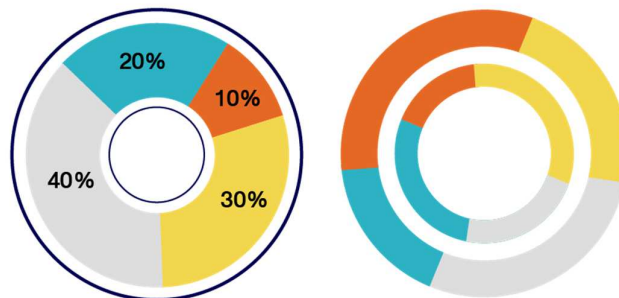
ภาพที่ 2.2 แผนภูมิเส้น (Line Charts)

3) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยาก เพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น



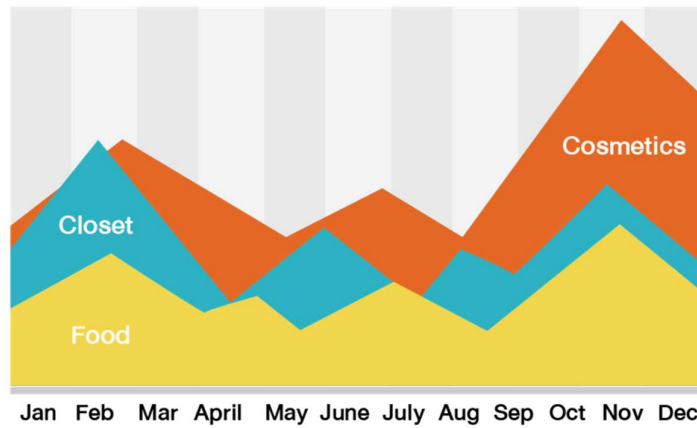
ภาพที่ 2.3 แผนภูมิวงกลม (Pie Charts)

4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลาย ๆ ชั้น นั่นเอง



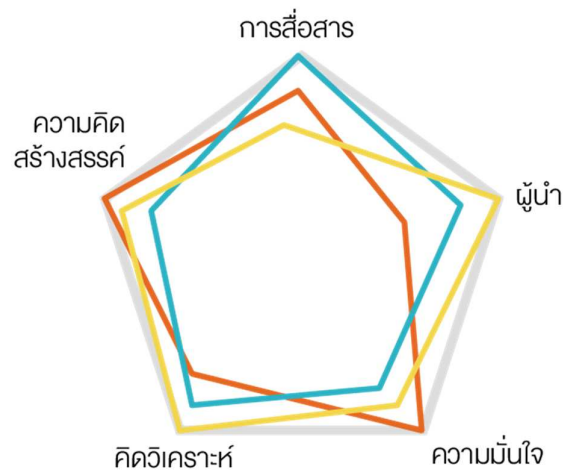
ภาพที่ 2.4 แผนภูมิโดนัท (Doughnut Charts)

5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่ละเดือน ตามหมวดหมู่ต่าง ๆ ไล่ไป เครื่องสำอางค์ เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

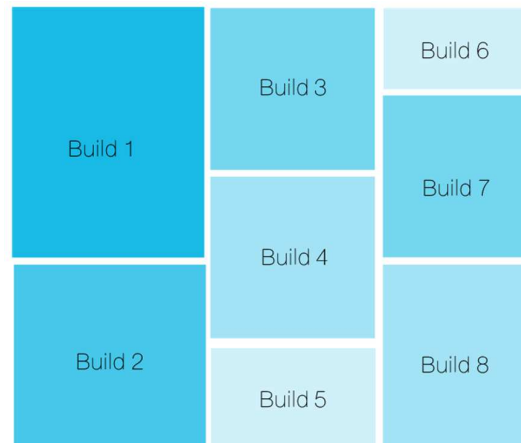
6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อนจุดแข็งของข้อมูลเช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อนจุดแข็งของแต่ละคน เป็นต้น



ภาพที่ 2.6 แผนภูมิเรดาร์ (Radar Charts)

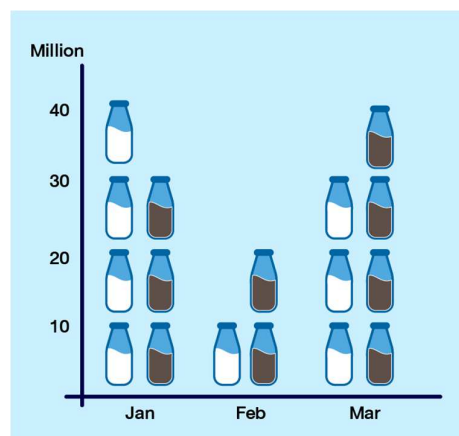
7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดงผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้





ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอน และ แกนตั้ง แต่เลือกใช้รูปภาพ หรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ เช่น การแสดงผลจำนวนของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละเดือนซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง แต่เปลี่ยนจากแท่งเป็นรูปภาพของนม 2 รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะนี้ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคน อาจจะแทนด้วยภาพไอคอนคนหรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงิน หรือ แบงค์แทน ก็จะทำให้คนดูเข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคยอยู่แล้วและยังดึงดูดความสนใจได้มากกว่าการใช้กราฟแท่งสีเหลี่ยมอีกด้วย



ภาพที่ 2.8 แผนภูมิรูปภาพ (Picture Graph)

## 2.2 ทฤษฎี

### 2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

#### 2.2.1.1 ข้อมูลขนาดใหญ่

ข้อมูลขนาดใหญ่ หมายถึงข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยากหรือเป็นไปได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมากเพื่อทำการวิเคราะห์มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วงต้นปีค.ศ. 2000 เมื่อค็อก ลานีย์ นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันในขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วย3Vs ดังนี้

Volume (ปริมาณ) : องค์กรต่าง ๆ รวบรวมข้อมูลจากหลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย และอื่น ๆ ในอดีต การจัดเก็บข้อมูลถือเป็นปัญหาใหญ่ – แต่เมื่อค่าใช้จ่ายในการจัดเก็บบนแพลตฟอร์มต่าง ๆ เช่น พื้นที่จัดเก็บข้อมูลส่วนกลาง (Data Lake) และ Hadoop ลดลง ภาระนี้จึงบรรเทาลง

Velocity (ความเร็ว) : ด้วยการเติบโตของ Internet of Things ข้อมูลจะถูกส่งไปยังธุรกิจต่าง ๆ ด้วยความเร็วที่ไม่เคยมีมาก่อนและต้องได้รับการจัดการในเวลาที่เหมาะสม แท็ก RFID, เซ็นเซอร์ และสมาร์ทมิเตอร์ช่วยผลักดันความต้องการในการจัดการกับกระแสข้อมูลเหล่านี้ในแบบเรียลไทม์

Variety (ความหลากหลาย) : ข้อมูลมีในทุกรูปแบบ นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ อีเมล วิดีโอ เสียง ข้อมูลหุ่น และธุรกรรมทางการเงิน

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

1) ลดต้นทุน : เทคโนโลยีข้อมูลขนาดใหญ่เช่น Hadoop และการวิเคราะห์บนคลาวด์นำมาซึ่งความได้เปรียบด้านต้นทุนอย่างมีนัยสำคัญเมื่อพูดถึงการจัดเก็บข้อมูลจำนวนมากรวมทั้งสามารถระบุวิธีการทำธุรกิจที่มีประสิทธิภาพมากขึ้น

2) เร็วกว่าและตัดสินใจดีกว่า : ด้วยความเร็วของ Hadoop และการวิเคราะห์ในหน่วยความจำรวมกับความสามารถในการวิเคราะห์แหล่งข้อมูลใหม่ๆ ของธุรกิจจะสามารถสร้างข้อมูลได้ทันทีและสามารถดำเนินการต่อได้ทันทีจากการวิเคราะห์นั้น ๆ

3) ผลิตภัณฑ์และบริการใหม่ : ด้วยความสามารถในการวัดความต้องการและความพึงพอใจของลูกค้าผ่านการวิเคราะห์นำมาซึ่งสิ่งที่ลูกค้าต้องการ ซึ่งยกตัวอย่างโดยดาเวนพอร์ท สามารถชี้ให้เห็นว่าด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ของบริษัทต่าง ๆ จะสามารถสร้างผลิตภัณฑ์ใหม่ ๆ เพื่อตอบสนองความต้องการของลูกค้าได้เป็นอย่างดี

### 2.2.1.2 การจัดการข้อมูลขนาดใหญ่

#### การรวบรวมข้อมูล

การรวบรวมข้อมูลของ Big Data เป็นการรวบรวมข้อมูลของจากหลากหลายทั้งที่มาและการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบดั้งเดิม ETL (extract, transform, and load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิค วิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาด เทราไบต์ และอาจจะเป็นระดับเพตาไบต์เลยทีเดียว ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล จัดรูปแบบ ให้เหมาะสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้น ๆ

#### การจัดการข้อมูล

ข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ on premises หรือแบบ cloud ขึ้นกับความต้องการหรือความสะดวกในการใช้ ซึ่งเราสามารถใช้และประเมินผลได้เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูล หรือข้อมูลบางอันต้องการความยืดหยุ่นสูงและไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยมกันเป็นอย่างมาก

#### การวิเคราะห์

การลงทุนสร้างข้อมูลขนาดใหญ่ หรือ Big data จะมีประโยชน์หรือคุ้มค่าก็ต่อเมื่อคุณใช้และวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่างและชัดเจนในชุดข้อมูลที่คุณมีอยู่ การสำรวจข้อมูลยังทำให้เราค้นพบสิ่งใหม่ แชร์สิ่งที่ค้นพบใหม่ๆต่อคนอื่น สร้างรูปแบบจำลองข้อมูล ด้วยการเรียนรู้ของเครื่องจักรและปัญญาประดิษฐ์ AI และนำข้อมูลเหล่านั้นไปใช้งาน

### 2.2.1.3 เทคโนโลยีการประมวลผลข้อมูล

การประมวลผลข้อมูลที่เป็น Big Data จะมีทั้งการวิเคราะห์ข้อมูลที่เป็น business intelligence (BI) เพื่อที่จะดึงข้อมูลมานำเสนอ หรือการทำ Predictive Analytics โดยใช้หลักการของ Data Science ความยากของการประมวลผลคือต้องการความเร็วในการประมวลผลข้อมูลที่นอกจากมีขนาดใหญ่แล้วบางครั้งยังเป็นข้อมูลที่ไม่มีโครงสร้าง ดังนั้นต้องจึงมีการนำเทคโนโลยีหรือภาษาต่าง ๆ มาเพื่อให้สามารถประมวลผลข้อมูลได้ ซึ่งในบางครั้งหน่วยงานอาจต้องพิจารณาต้องเลือกใช้

### 2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

2.2.2.1 การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

- 1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง
- 2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย
- 3) การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
- 4) การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

### 2.2.2.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้

- 1) Business Understanding เน้นไปที่การทำความเข้าใจในงาน ระบุโอกาส และหาปัญหาที่จะเกิดขึ้น กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์ ซึ่งต้องสามารถระบุผลลัพธ์ที่มีได้
- 2) Data Understanding ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์
- 3) Data Preparation ทำการแปลงข้อมูล (Raw Data) ให้กลายเป็นข้อมูลที่สามารถนำมาช่วยในการวิเคราะห์ต่อไปได้ ขั้นตอนนี้จะใช้เวลามากที่สุดในทุกขั้นตอน เพราะคุณภาพของงานที่ได้จะดีเพียงใดขึ้นอยู่กับคุณภาพข้อมูลที่จัดเตรียมในขั้นนี้ การเตรียมข้อมูลประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล
- 4) Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากขั้นตอนที่ 3 พร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้คำตอบที่ดีที่สุด บางครั้งอาจมีการย้อนกลับไปปรับการเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด
- 5) Evaluation การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับวัตถุประสงค์หรือเป้าหมายที่ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด หากไม่ได้ผลลัพธ์ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า
- 6) Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์จากทั้งหมด มาใช้ประโยชน์ตามวัตถุประสงค์ และทำการประเมินผลลัพธ์ที่ได้ว่ามีประสิทธิภาพตรงตามวัตถุประสงค์ที่ตั้งไว้หรือไม่

## 2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

### 2.2.3.1 หลักในการออกแบบเว็บไซต์

หน้าเว็บเป็นสิ่งที่ผู้ใช้จะได้เห็นขณะที่เปิดเข้าสู่เว็บไซต์ และยังเป็นสิ่งแรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วย หน้าเว็บจึงเป็นสิ่งสำคัญมาก เพราะเป็นสื่อกลางให้ผู้ชมสามารถใช้ประโยชน์จากข้อมูลของระบบงานของเว็บไซต์นั้นได้ โดยปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้น ระบบเนวิเกชั่น และองค์ประกอบอื่น ๆ ที่ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการใช้งานหลักสำคัญในการออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อความหมายเกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้ การออกแบบเว็บไซต์ ต้องคำนึงถึง

- 1) ความเรียบง่าย ได้แก่ มีรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานได้สะดวก ไม่มีกราฟิกหรือตัวอักษรที่เคลื่อนไหวอยู่ตลอดเวลา ชนิดและสีของตัวอักษรไม่มากเกินไปทำให้วุ่นวาย
- 2) ความสม่ำเสมอ ได้แก่ ใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ เช่น รูปแบบของหน้า สไตส์ของกราฟิก ระบบเนวิเกชันและโทนสี ควรมีความคล้ายคลึงกันตลอดทั้งเว็บไซต์
- 3) ความเป็นเอกลักษณ์ การออกแบบเว็บไซต์ควรคำนึงถึงลักษณะขององค์กร เพราะรูปแบบของเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กรนั้น ๆ เช่น ถ้าเป็นเว็บไซต์ของทาง ราชการ จะต้องดูน่าเชื่อถือไม่เหมือนสวนสนุก ฯลฯ
- 4) เนื้อหาที่มีประโยชน์ เนื้อหาเป็นสิ่งที่สำคัญที่สุดในเว็บไซต์ ดังนั้นควรจัดเตรียมเนื้อหาและข้อมูลที่ใช้ต้องการให้ถูกต้อง และสมบูรณ์ มีการปรับปรุงและเพิ่มเติมให้ทันเหตุการณ์อยู่เสมอ เนื้อหาไม่ควรซ้ำกับเว็บไซต์อื่น จึงจะดึงดูดความสนใจ
- 5) ระบบเนวิเกชันที่ใช้งานง่าย ต้องออกแบบให้ผู้ใช้เข้าใจง่ายและใช้งานสะดวก ใช้กราฟิกที่สื่อความหมายร่วมกับคำอธิบายที่ชัดเจน มีรูปแบบและลำดับของรายการที่สม่ำเสมอ เช่น วางไว้ ตำแหน่งเดียวกันของทุกหน้า
- 6) ลักษณะที่น่าสนใจ หน้าตาของเว็บไซต์จะต้องมีความสัมพันธ์กับคุณภาพขององค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่จะต้องสมบูรณ์ การใช้สี การใช้ตัวอักษรที่อ่านง่าย สบายตา การใช้โทนสีที่เข้ากันลักษณะหน้าตาที่น่าสนใจนั้นขึ้นอยู่กับความชอบของแต่ละบุคคล
- 7) การใช้งานอย่างไม่จำกัด ผู้ใช้ส่วนใหญ่สามารถเข้าถึงได้มากที่สุด เลือกใช้เบราว์เซอร์ชนิดใดก็ได้ในการเข้าถึงเนื้อหาสามารถแสดงผลได้ทุกระบบปฏิบัติการและความละเอียดหน้าจอต่าง ๆ กันอย่างไม่เป็นปัญหาเป็นลักษณะสำคัญสำหรับผู้ที่มีจำนวนมาก
- 8) คุณภาพในการออกแบบ การออกแบบและเรียบเรียงเนื้อหาอย่างรอบคอบ สร้างความรู้สึกว่าเว็บไซต์มี...คุณภาพ ถูกต้อง และเชื่อถือได้
- 9) ลิงค์ต่าง ๆ จะต้องเชื่อมโยงไปหน้าที่มีอยู่จริงและถูกต้อง ระบบการทำงานต่าง ๆ ในเว็บไซต์จะต้องมีความแน่นอนและทำหน้าที่ได้อย่างถูก

#### 2.2.3.2 องค์ประกอบของการออกแบบเว็บไซต์

- 1) โครงสร้างที่ชัดเจน ผู้ออกแบบเว็บไซต์ควรจัดโครงสร้างหรือจัดระเบียบของข้อมูลที่ชัดเจน แยกย่อยเนื้อหาออกเป็นส่วนต่าง ๆ ที่สัมพันธ์กันและให้อยู่ในมาตรฐานเดียวกัน จะช่วยให้การใช้งานและง่ายต่อการอ่านเนื้อหาของผู้ใช้

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

3) การเชื่อมโยงที่ดี ลักษณะไฮเปอร์เท็กซ์ที่ใช้ในการเชื่อมโยง ควรอยู่ในรูปแบบที่เป็นมาตรฐาน ทัวไปและต้องระวังเรื่องของตำแหน่งในการเชื่อมโยง การที่จำนวนการเชื่อมโยงมากและกระจัดกระจายอยู่ทัวไปในหน้าอาจก่อให้เกิดความล้นสน นอกจากนี้คำที่ใช้สำหรับการเชื่อมโยงจะต้องเข้าใจง่ายมีความชัดเจนและไม่ล้นจนเกินไป นอกจากนี้ในแต่ละเว็บเพจที่สร้างขึ้นมาควรมี จุดเชื่อมโยงกลับมายังหน้าแรกของเว็บไซต์ที่กำลังใช้งานอยู่ด้วย ทั้งนี้เพื่อว่าผู้ใช้เกิดหลงทาง และไม่ทราบว่าจะทำอย่างไรต่อไปจะได้มีหนทางกลับมาสู่จุดเริ่มต้นใหม่ ระวังอย่าให้มีหน้าที่ไม่มีการเชื่อมโยง (Orphan Page) เพราะจะทำให้ผู้ใช้ไม่รู้จะทำอย่างไรต่อไป

4) ความเหมาะสมในหน้าจอ เนื้อหาที่นำเสนอในแต่ละหน้าจควรล้นกระชับ และทันสมัย หลีกเลี้ยงการใช้หน้าจอที่มีลักษณะการเลื่อนขึ้นลง (Scrolling) แต่ถ้าจำเป็นต้องมี ควรจะให้ข้อมูลที่มี ความสำคัญอยู่บริเวณด้านบนสุดของหน้าจอ หลีกเลี้ยงการใช้กราฟิกด้านบนของหน้าจอ เพราะถึงแม้จะดูสวยงาม แต่จะทำให้ผู้ใช้เสียเวลาในการได้รับข้อมูลที่ต้องการ แต่หากต้องมีการใช้ภาพประกอบก็ควรใช้เฉพาะที่มีความสัมพันธ์กับเนื้อหาเท่านั้น นอกจากนี้การใช้รูปภาพเพื่อเป็นพื้นหลัง (Background) ไม่ควรเน้นสีล้นที่ฉูดฉาดมากนัก เพราะอาจจะไปลดความเด่นชัดของเนื้อหา ควรใช้ภาพที่มีสีอ่อน ๆ ไม่สว่างจนเกินไป รวมไปถึงการใช้เทคนิคต่าง ๆ เช่น ภาพเคลื่อนไหว หรือตัวอักษรวิ่ง (Marquees) ซึ่งอาจจะเกิดการรบกวนการอ่านได้ ควรใช้เฉพาะที่จำเป็นจริง ๆ เท่านั้นตัวอักษรที่นำมาแสดงบนจอภาพ ควรเลือกขนาดที่อ่านง่าย ไม่มีสีล้นและลวดลายมากเกินไป

5) ความรวดเร็ว ความรวดเร็วเป็นสิ่งสำคัญประการหนึ่งที่ส่งผลต่อการเรียนรู้ ผู้ใช้จะเกิดอาการเบื่อหน่ายและหมดความสนใจกับเว็บที่ใช้เวลาในการแสดงผลนาน สาเหตุสำคัญที่จะทำให้การแสดงผลนานคือการใช้ภาพกราฟิกหรือภาพเคลื่อนไหว ซึ่งแม้ว่าจะช่วยดึงดูดความสนใจได้ดี ฉะนั้นในการออกแบบจึงควรหลีกเลี้ยงการใช้ภาพขนาดใหญ่ หรือภาพเคลื่อนไหวที่ไม่จำเป็น และพยายามใช้กราฟิกแทนตัวอักษรธรรมดาให้น้อยที่สุด โดยไม่ควรใช้มากเกินไปกว่า 2 – 3 บรรทัดในแต่ละหน้าจอ

### 2.2.3.3 โครงสร้างเว็บไซต์

1) เว็บไซต์ที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure) เป็นโครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยมจัดด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การเรียงลำดับตามตัวอักษร ตระชนี สารานุกรม หรืออภิธานศัพท์ โครงสร้างแบบนี้ เหมาะกับเว็บไซต์ที่มีขนาดเล็ก เนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปทีละหน้า ทิศทางของการเข้าสู่เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรง โดยมี ปุ่มเดินทาง-ถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาของตนเองได้ ทำให้เสียเวลาเข้าสู่เนื้อ



ภาพที่ 2.9 เว็บไซต์ที่มีโครงสร้างแบบเรียงลำดับ

2) เว็บไซต์ที่มีโครงสร้างแบบลำดับชั้น (Hierarchical Structure) เป็นวิธีที่ดีที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วนต่างๆ และมีรายละเอียดย่อยๆ ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับ แผนภูมิองค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับโครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของ เว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดรวมจุดเดียว นั่นคือ โฮมเพจ (Homepage) และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง

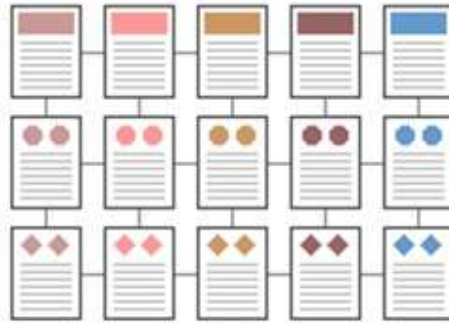


ภาพที่ 2.10 เว็บไซต์ที่มีโครงสร้างแบบลำดับชั้น

3) เว็บไซต์ที่มีโครงสร้างแบบตาราง (Grid Structure) โครงสร้างรูปแบบนี้มีความซับซ้อนมากกว่ารูปแบบที่ผ่านมา การออกแบบเพิ่มความยืดหยุ่น ให้แก่การเข้าสู่เนื้อหาของผู้ใช้ โดยเพิ่มการเชื่อมโยงซึ่งกันและกันระหว่างเนื้อหาแต่ละส่วน เหมาะแก่ การแสดงให้



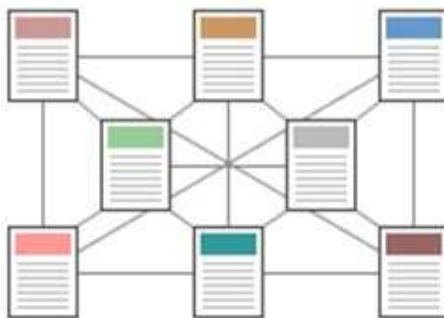
เห็นความสัมพันธ์กันของเนื้อหา การเข้าสู่เนื้อหาของผู้ใช้จะไม่เป็นลักษณะเชิงเส้นตรง เนื่องจากผู้ใช้สามารถเปลี่ยนทิศทางการเข้าสู่เนื้อหาของตนเองได้



ภาพที่ 2.11 เว็บที่มีโครงสร้างแบบตาราง

ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมีลักษณะที่เหมือนกัน และสามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมาบรรจุลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับโครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือก หัวข้อใด ก็จะไปสู่หน้าเนื้อหา (Topic Page) ที่แสดงรายละเอียดของหัวข้อนั้นๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้ารายละเอียดของหัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำ โครงสร้างแบบเรียงลำดับและแบบลำดับขั้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจจะสร้างความยุ่งยากในการเข้าใจได้ และอาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ใช้ได้เข้าใจถึงความสัมพันธ์ ระหว่างเนื้อหา ในส่วนของการออกแบบจำเป็นจะต้องมีการวางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้น ได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจเกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4) เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure) โครงสร้างประเภทนี้จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกัน ได้หมด เป็นการสร้างรูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วย ตนเอง การเชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกัน ของแต่ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็นรูปแบบที่ ไม่มีโครงสร้างที่แน่นอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัดเฉพาะเนื้อหาภายในเว็บนั้น ๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้



ภาพที่ 2.12 เว็บที่มีโครงสร้างแบบใยแมงมุม

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย กับข้อความที่มีโมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้จะปรากฏอยู่บริเวณใด บริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการเพื่อเลือกที่จะเข้าไปดูหน้าใด ๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการท่องเที่ยวนบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือ ถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอจะเป็นการยากในการ ปรับปรุง นอกจากนี้การเชื่อมโยงระหว่างข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสนและ เกิดปัญหาการคงค้างของหัวข้อ (Cognitive Overhead) ได้

### 2.2.3.3 การใช้สีในการออกแบบเว็บไซต์

การสร้างสีสันบนหน้าเว็บเป็นสิ่งที่สื่อความหมายของเว็บไซต์ได้อย่างชัดเจน การเลือกใช้สีให้เหมาะสม กลมกลืน ไม่เพียงแต่จะสร้างความพึงพอใจให้กับผู้ใช้ แต่ยังสามารถทำให้เห็นถึงความแตกต่างระหว่างเว็บไซต์ได้ สีเป็นองค์ประกอบหลักสำหรับการตกแต่งเว็บ จึงจำเป็นอย่างยิ่งที่จะต้องทำความเข้าใจเกี่ยวกับการใช้สีระบบสีที่แสดงบนจอคอมพิวเตอร์ มีระบบการแสดงผลผ่านหลอดลำแสงที่เรียกว่า CRT (Cathode ray tube) โดยมีลักษณะระบบสีแบบบวก อาศัยการผสมของของแสงสีแดง สีเขียว และสีน้ำเงิน หรือระบบสี RGB สามารถกำหนดค่าสีจาก 0 ถึง 255 ได้ จากการรวมสีของแม่สีหลักจะทำให้เกิดแสงสีขาว มีลักษณะเป็นจุดเล็ก ๆ บนหน้าจอไม่สามารถมองเห็นด้วยตาเปล่าได้ จะมองเห็นเป็นสีที่ถูกผสมเป็นเนื้อสีเดียวกันแล้ว จุดแต่ละจุดหรือพิกเซล (Pixel) เป็นส่วนประกอบของภาพบนหน้าจอคอมพิวเตอร์ โดยจำนวนบิตที่ใช้ในการกำหนดความสามารถของการแสดงสีต่าง ๆ เพื่อสร้างภาพบนจอขึ้นเรียกว่า บิตเด็ป (Bit-depth) ในภาษา HTML มีการกำหนดสีด้วยระบบเลขฐานสิบหก ซึ่งมีเครื่องหมาย (#) อยู่ด้านหน้าและตามด้วยเลขฐานสิบหกจำนวนอักษรอีก 6

หลัก โดยแต่ละไบต์ (byte) จะมีตัวอักษรสองตัว แบ่งออกเป็น 3 กลุ่ม เช่น #FF12AC การใช้ตัวอักษรแต่ละไบต์นี้เพื่อกำหนดระดับความเข้มของแม่สีแต่ละสีของชุดสี RGB โดย 2 หลักแรกแสดงถึงความเข้มของสีแดง 2 หลักต่อมา แสดงถึงความเข้มของสีเขียว 2 หลักสุดท้ายแสดงถึงความเข้มของสีน้ำเงิน

สีมีอิทธิพลในเรื่องของอารมณ์การสื่อความหมายที่เด่นชัด กระตุ้นการรับรู้ทางด้านจิตใจมนุษย์ สีแต่ละสีให้ความรู้สึก อารมณ์ที่ไม่เหมือนกัน สีบางสีให้ความรู้สึกสงบ บางสีให้ความรู้สึกตื่นเต้นรุนแรง สีจึงเป็นปัจจัยสำคัญอย่างยิ่งต่อการออกแบบเว็บไซต์ ดังนั้นการเลือกใช้โทนสีภายในเว็บไซต์เป็นการแสดงถึงความแตกต่างของสีที่แสดงออกทางอารมณ์ มีชีวิตชีวาหรือเศร้าโศก รูปแบบของสีที่สายตาของมนุษย์มองเห็น สามารถแบ่งออกเป็น 3 กลุ่ม คือ

1) สีโทนร้อน (Warm Colors) เป็นกลุ่มสีที่แสดงถึงความสุข ความปลอดภัย ความอบอุ่น และดึงดูดใจ สีกลุ่มนี้เป็นกลุ่มสีที่ช่วยให้หายจากความเฉื่อยชา มีชีวิตชีวามากยิ่งขึ้น

2) สีโทนเย็น (Cool Colors) แสดงถึงความที่ดูสุภาพ อ่อนโยน เรียบร้อย เป็นกลุ่มสีที่มีคนชอบมากที่สุด สามารถโน้มน้าวในระยะไกลได้

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

## 2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่งภาษาของการออกแบบเว็บไซต์

2.2.4.1 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

### ประโยชน์ของ CSS

- 1) CSS มีคุณสมบัติมากกว่า tag ของ html เช่น การกำหนดกรอบให้ข้อความ รวมทั้งสี รูปแบบของข้อความที่กล่าวมาแล้ว
- 2) CSS นั้นกำหนดที่ต้นของไฟล์ html หรือตำแหน่งอื่น ๆ ก็ได้ และสามารถมีผล กับเอกสารทั้งหมด หมายถึงกำหนด ครั้งเดียวจุดเดียวก็มีผลกับการแสดงผลทั้งหมด ทำให้เวลาแก้ไขหรือปรับปรุงทำได้สะดวก ไม่ต้องไล่ตามแก้ tag ต่างๆ ทั้งทั้งเอกสาร
- 3) CSS สามารถกำหนดแยกไว้ต่างหากจาก ไฟล์เอกสาร html และสามารถนำมาใช้ร่วม กับเอกสารหลายไฟล์ได้ การแก้ไขก็แก้เพียง จุดเดียวก็มีผลกับเอกสารทั้งหมด

CSS กับ HTML / XHTML นั้นทำหน้าที่คนละอย่างกัน โดย HTML / XHTML จะทำหน้าที่ในการวางโครงร่างเอกสารอย่างเป็นทางการ ถูกต้อง เข้าใจง่าย ไม่เกี่ยวข้องกับการแสดงผล ส่วน CSS จะทำหน้าที่ในการตกแต่งเอกสารให้สวยงาม เรียกได้ว่า HTML / XHTML คือส่วน coding ส่วน CSS คือส่วน design

2.2.4.2 Bootstrap คือชุดคำสั่งที่ประกอบด้วยภาษา CSS, HTML และ Javascript เป็นชุดคำสั่งที่ถูกพัฒนาขึ้นมาเพื่อกำหนดกรอบ หรือ รูปแบบการพัฒนาเว็บไซต์ในส่วนของการปฏิสัมพันธ์กับผู้ใช้งานเว็บไซต์ ( User Interface ) เราจึงสามารถเรียก Bootstrap ว่าเป็น Front-end framework คือใช้สำหรับ พัฒนาเว็บไซต์ส่วนการแสดงผล ซึ่งแตกต่างจากภาษาประเภท Server Side Script อย่าง PHP, Python หรือภาษาอื่น ๆ

#### จุดเด่นของ Bootstrap Framework

- 1) มี UI เริ่มต้นแบบที่สวยงามและใช้งานง่าย
- 2) มีการปรับปรุงและพัฒนาอย่างต่อเนื่อง ปัจจุบันเป็นเวอร์ชัน 3.3.0
- 3) เป็นที่นิยมของนักพัฒนาทั่วโลก ทำให้สามารถเรียนรู้ และแก้ปัญหาได้ง่าย
- 4) โค้ดหรือชุดคำสั่งต่าง ๆ ค่อนข้างสะอาดมีไฟล์เริ่มต้นแบบแค่ 3 ส่วน

คือ js, css, fonts

- 5) ประหยัดเวลาในการพัฒนาเว็บไซต์และนำไปพัฒนาต่อได้ง่าย
- 6) เป็น Responsive Framework พัฒนาเว็บไซต์ที่รองรับการแสดงผลได้

หลากหลาย Device

### 2.2.5 ทฤษฎีเกี่ยวกับการ visualization

Data Visualization หรือ Information Visualization คือ การนำข้อมูลในเชิงปริมาณ ทั้งที่จัดเก็บไว้ในรูปแบบของข้อมูลจำนวนน้อย และข้อมูลจำนวนมาก (Big Data) มาประมวลผล จากนั้นจึงนำมาแสดงผลในรูปแบบของกราฟ แผนภูมิอินฟอร์เมชันกราฟิก หรือแม้กระทั่งอินเทอร์แอคทีฟกราฟิก ที่ผู้บริโภครสามารถคลิกหรือมีปฏิสัมพันธ์กับกราฟิกนั้น ๆ ได้ ซึ่งปัจจุบันสื่อหลายสำนักในประเทศไทย เริ่มมีการนำเสนอข่าวโดยการนำข้อมูลแบบประยุกต์ ด้วยภาพ (Data Visualization) การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ ซึ่งคำว่า “ประสิทธิภาพ” ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้ว อาจทำให้เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์ เชื่อมโยงได้

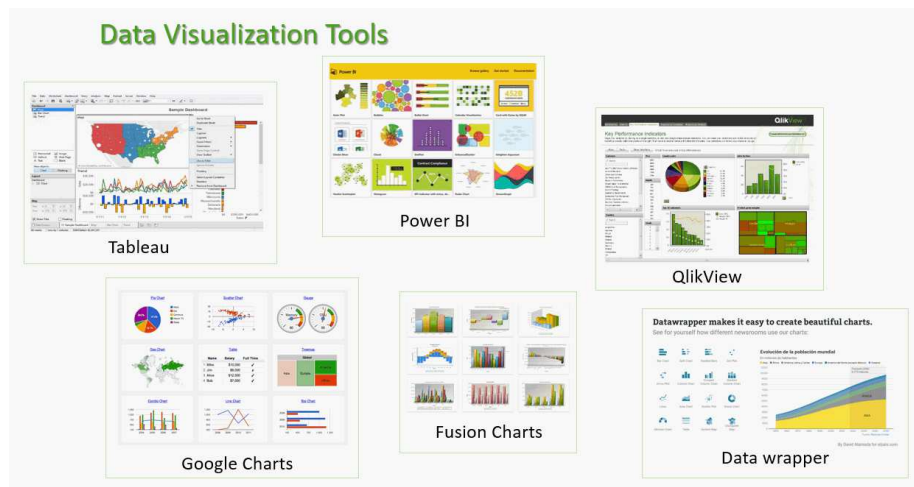
Visualization คือ การจินตนาการ หรือสร้างภาพขึ้นในความคิด ซึ่งเป็นกระบวนการ ทำให้เกิดความรู้ความเข้าใจสำหรับเรื่องที่จินตนาการยาก เข้าใจยาก วิธีการที่เป็นทางลัดก็คือ การสร้าง ภาพ ให้เป็น บันไดความคิด ไปสู่ การใช้ความคิดอีกระดับ บันไดนี้จะช่วยลดปริมาณ ข้อมูล ช่วยลดภาระการคำนวณหรือการนำไปผ่านหลากหลายกระบวนการความคิด เพื่อนำไปสู่คำตอบที่ต้องการได้โดยเร็วและถูกต้อง

Visualization System คือระบบ ซึ่งประกอบด้วยกระบวนการต่าง ๆ ที่ออกแบบมา เพื่อสร้าง รักษา นำไปใช้ และปรับปรุงทัศนสนเทศ เพื่อทำให้เกิดความรู้ความเข้าใจ อย่างถูกต้อง รวดเร็ว และได้ผลเป็นอย่างดี

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพในความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณ ให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

#### 2.2.5.1 โปรแกรมสำหรับการสร้างแดชบอร์ด (Data Visualization)

Tools ที่นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qlik View, Google Charts, Fusion Charts, Data wrapper และอื่น ๆ อีกมากมาย



ภาพที่ 2.13 Tools Data Visualization

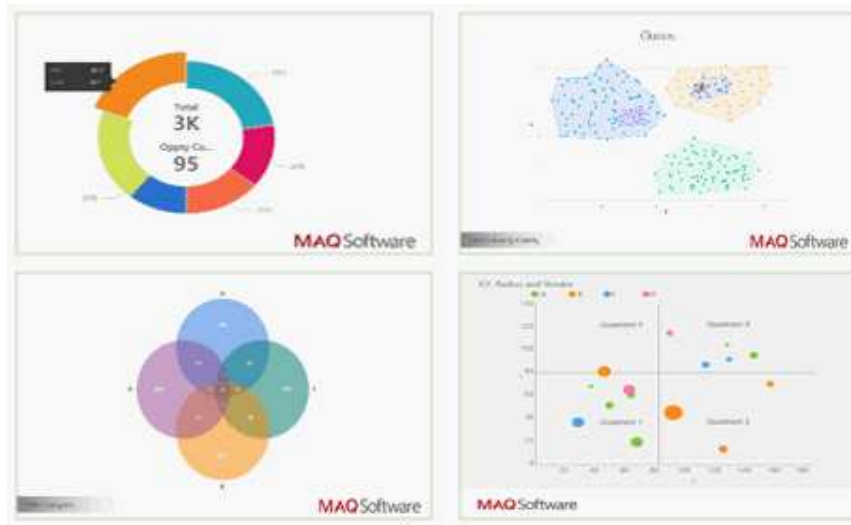
#### 2.2.5.2 รูปแบบในการใช้ Data Visualization

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



ภาพที่ 2.14 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

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



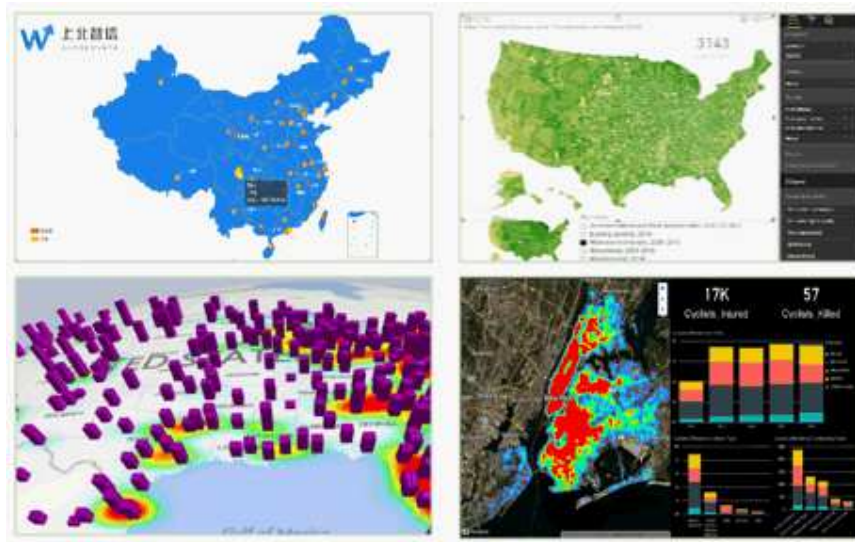
ภาพที่ 2.15 การนำเสนอแบบกลุ่มข้อมูล (Classification)

การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison) เหมาะสำหรับการนำเสนอที่ต้องการเปรียบเทียบผลการดำเนินงาน เช่น เทียบกับปีที่แล้ว (YoY) เทียบเทียบกับเป้าที่ตั้งไว้ (Target) ซึ่งกราฟที่เหมาะสมและมักนำมาใช้ เช่น KPI Indicator, Bullet Chart, Power BI Card with state เป็นต้น



ภาพที่ 2.16 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)

การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับการนำเสนอข้อมูลบนแผนที่ โดยสามารถที่จะนำยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อ Focus กลุ่มลูกค้าในแต่ละพื้นที่ที่เราสนใจ เช่น Globe Map, Google Map, Flow Map เป็นต้น



ภาพที่ 2.17 การนำเสนอรูปแบบแผนที่ (Geographical)

กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics) เราสามารถใช้ภาษา R หรือ Python ดึงข้อมูลในอดีตมาเพื่อวิเคราะห์เพื่อพยากรณ์อนาคต และนำเสนอข้อมูลที่ได้ในรูปแบบของกราฟ เช่น Association Rules, Clustering, Forecasting Time series, Calculation plot เป็นต้น



ภาพที่ 2.18 กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics)

ข้อดีของ Data Visualization

- ข้อมูลที่มีปริมาณมากมายมหาศาลยากต่อการเข้าใจ เราสามารถทำให้เข้าใจง่ายได้ด้วยรูปภาพ

- ช่วยจัดระเบียบความคิดวิเคราะห์ข้อมูลให้มีความน่าเชื่อถือ
- ประหยัดเวลาในการนำเสนอ



## 2.2.6 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ

### 2.2.6.1 ลักษณะต้นไม้การตัดสินใจ

ต้นไม้การตัดสินใจจะทำการจัดกลุ่ม (classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (node) ของต้นไม้การตัดสินใจคือตัวแปร (attribute) ต่าง ๆ ของชุดข้อมูล เช่นหากต้องการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณาคือทัศนียภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์จากต้นไม้คือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่า ซึ่งแต่ละตัวแปรนั้นก็จะมีค่าของตัวเอง (value) เกิดเป็นชุดของตัวแปร-ค่าของตัวแปร (attribute-value pair) เช่น ทัศนียภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่านั้นก็อาจมีค่าได้เป็นใช่ กับ ไม่ใช่ เป็นต้น การทำนายประเภทด้วยต้นไม้ตัดสินใจ จะเริ่มจากบัพราก โดยทดสอบค่าตัวแปรของบัพ แล้วจึงตามกิ่งของต้นไม้ที่กำหนดค่า เพื่อไปยังบัพลูกถัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอบัพใบซึ่งจะแสดงผลการทำนาย

#### 2.2.6.1 ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย

- 1) โหนด คือ คุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่จุดสูงสุดเรียกว่า โหนดราก (Root Node)
- 2) กิ่ง คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด
- 3) ใบ คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ

การเรียนรู้แบบต้นไม้ตัดสินใจ (อังกฤษ: decision tree learning) เป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูล

ในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (leaf node) จะแสดงประเภทของวัตถุ

ต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่าต้นไม้ตัดสินใจแบบจำแนก (classification trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (continuous values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (regression trees)

ต้นไม้การตัดสินใจในการบริหารธุรกิจ เป็นแผนผังต้นไม้ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และ ทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

## 2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

### 2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

#### 2.3.1.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

เป็นเทคนิคหนึ่งของ Classification ซึ่งเป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูล โดย Classification นั้นเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining)

การเรียนรู้แบบต้นไม้ตัดสินใจ (อังกฤษ: decision tree learning) เป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูล

ในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (leaf node) จะแสดงประเภทของวัตถุ

ต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่าต้นไม้ตัดสินใจแบบจำแนก (classification trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (continuous values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (regression trees)

ต้นไม้การตัดสินใจในการบริหารธุรกิจ เป็นแผนผังต้นไม้ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และ ทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

#### 2.3.1.2 ขั้นตอนในการสร้าง Decision Tree เพื่อใช้จำแนกข้อมูล มีดังนี้

- 1) เลือก Attribute ที่ทำหน้าที่เป็น Root Node

2) จาก Root Node สร้างเส้นเชื่อมโยงไปยังโหนดลูก จำนวนเส้นเชื่อมโยงจะเท่ากับจำนวนค่าที่เป็นไปได้ทั้งหมดของ Attribute ที่เป็น root node

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

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

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

เมื่อแทนค่า  $S$  คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

$p_s(j)$  คือ อัตราส่วนของกรณีใน  $S$  ที่ตัวแปรตามหรือผลลัพธ์มีค่า  $j$

โดยสำหรับต้นไม้การตัดสินใจที่มีผลลัพธ์เป็นแค่เพียงค่าตรรกะ (Boolean) ใช่กับไม่ใช่เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปี คือ

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

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

เกนความรู้ (Information Gain) ซึ่งจากการนิยามเอนโทรปีข้างต้น ทำให้เราสามารถนิยามลักษณะของตัวแปรต้นที่ดีได้ โดยตัวแปร  $A$  จะเป็นตัวแปรต้นที่ดีก็ต่อเมื่อหากว่าแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุดๆมีจำนวนชุดตามจำนวนค่าของ  $A$  ที่เป็นไปได้ เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า  $A$  เพียงค่าเดียวและค่าเฉลี่ยของเอนโทรปีของชุด

ข้อมูลที่ถูกแบ่งออก (partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปี หลังจากข้อมูลถูกแบ่งด้วย  $A$  ว่าเกนความรู้ของ  $A$  นิยามโดย

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

เมื่อแทนค่า  $S$  คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

$E$  คือเอนโทรปีของตัวอย่าง

$A$  คือตัวแปรต้นที่พิจารณา

$value(A)$  คือเซตของค่าของ  $A$  ที่เป็นไปได้

$S_v$  คือตัวอย่างที่  $A$  มีค่า  $v$  ทั้งหมด

จะเห็นว่าหากเกนความรู้ของ  $A$  ยิ่งมากแสดงว่าหลังจากแบ่งตัวอย่าง  $S$  ด้วย  $A$  แล้วในแต่ละชุดที่แบ่งได้จะมี Entropy เข้าใกล้ศูนย์มากยิ่งขึ้น ทำให้ใกล้ที่จะตัดสินใจได้มากขึ้น เกนความรู้จึงเป็นค่าที่ดีที่จะบอกความดีของตัวแปรต้นที่นำมาพิจารณา

### 2.3.2 การประเมินผลโมเดล (Decision Tree)

สถิติในการทดสอบหาความแม่นยำของอัลกอริทึม และทดสอบความถูกต้อง และความน่าเชื่อถือของโมเดลโดยใช้การประเมินค่าความแม่นยำ Confusion Matrix

Confusion Matrix คือ การประเมินผลลัพธ์การทำนาย (หรือผลลัพธ์จากโปรแกรม)

เปรียบเทียบกับผลลัพธ์จริงโดยที่

a, d คือ จำนวนข้อมูลที่มีในการทำนายถูก

b, c คือ จำนวนข้อมูลที่มีในการทำนายผิด

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a	b
	Class=No	c	d

ภาพที่ 2.19 Confusion Matrix)

แสดงเป็นสมการได้ ดังนี้

Sensitivity or Recall คือค่าที่บอกว่า โปรแกรมทำนายได้ว่าจริง เป็นอัตราส่วน เท่าไหร่ของจริงทั้งหมด คำนวณได้จากสมการ

$$Recall = \frac{a}{a+b}$$

Precision คือค่าที่บอกว่า โปรแกรมทำนายว่าจริงถูกต้องเท่าไร คำนวณได้จากสมการ

$$precision = \frac{a}{a+c}$$

Correctly Classified Instances คือค่าที่บอกว่ามีการทำนายข้อมูลถูกต้อง และมีค่าความแม่นยำเท่าไรในการทำนาย คำนวณได้จากสมการ

$$\text{Correctly Classified Instances} = \frac{(a+d)}{(a+b+c+d)}$$

### 2.3.3 การแสดงผลแบบ visualization

#### 2.3.3.1 การแสดงผลแบบ visualization ด้วย Tableau Public

ทาโบลิว (Tableau) Business Intelligence (BI) Software คือซอฟต์แวร์เพื่อการวิเคราะห์ข้อมูลอันหลากหลายเป็น Software ในระดับต้นๆของโลกที่ทำเรื่อง Data Visualization Tableau ทำหน้าที่แปลงข้อมูล (Data) ให้เป็นภาพ (Visualization) สามารถนำข้อมูลออกมาเป็น Chart แบบต่าง ๆ ได้อย่างง่ายดายรวมถึงการนำหลาย Chart มาสร้างเป็น Dashboard และ Story Teller ก็ได้เช่นกัน รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL, Oracle, MySQL เป็นต้น



ภาพที่ 2.20 การแสดงแดชบอร์ดของ Tableau Public

(ที่มา: <https://www.thestephaneandre.com/install-tableau-desktop-and-tableau-public/>)

Tableau นั้นถูกพัฒนาอย่างต่อเนื่องเพื่อให้แสดงผลได้บนทุก Platform อย่างง่ายดาย ไม่ว่าจะเป็น Desktop, Mobile หรือแม้แต่ Browser โดยองค์กรสามารถเลือกติดตั้งใช้งานภายในองค์กรเองก็ได้ หรือเช่าใช้ผ่านบริการ Tableau Cloud ก็ได้เช่นกัน ทำให้การทำ Business Intelligence และ Data Analytics ไม่ใช่เรื่องยากอีกต่อไป และง่ายขึ้นจนถึงขนาดผู้บริหารก็สามารถใช้งานด้วยตัวเองได้

### 2.3.3.2 ลักษณะเด่นของ Tableau Public

- รูปแบบการใช้งานที่ง่ายเพียงคลิกเมาส์ก็สามารถที่จะเปลี่ยนแปลงรายงานได้ตามความต้องการ
- รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot
- รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการนำเสนอต่อผู้บริหาร
- การเข้าถึงหรือการใช้งาน สามารถทำได้โดยง่าย ตั้งแต่ผู้บริหารระดับสูง ถึง พนักงานระดับปฏิบัติการ ทั้งนี้ทั้งนั้น สามารถเพิ่มทักษะและแนวคิดให้กับพนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติและการประยุกต์ใช้งานมากขึ้น
- สิ่งที่สำคัญ คือ Engine ของ Tableau Software นั้น พัฒนาจาก VizQL Technology เป็นเทคโนโลยีที่ช่วยให้เห็นภาพของข้อมูลขนาดใดก็ได้เพียงการลากและวาง (Drag and Drop) โดยใช้ข้อมูลจากฐานข้อมูล ตอบสนองในรูปแบบของ กราฟฟิก

## 2.4 วรรณกรรมที่เกี่ยวข้อง

อภิยศ เจริญวิวัฒน์ (2560) สำนักหอสมุด มหาวิทยาลัยเกษตรศาสตร์ มีแนวคิดในการประยุกต์ใช้ข้อมูลทางสถิติเพื่อเพิ่มประสิทธิภาพในการให้บริการแก่ผู้ใช้ โดยทำการศึกษาเกี่ยวกับตัวเลขของการให้บริการในด้านต่าง ๆ ที่กระจัดกระจายอยู่เป็นจำนวนมาก มีปริมาณข้อมูลสูง และยังไม่เคยผ่านการวิเคราะห์ทางด้านสถิติมาก่อน โดยการวิเคราะห์ข้อมูลนี้ จะวิเคราะห์ข้อมูลที่มีหลากหลาย ทั้งจากการเข้าใช้งานอาคาร ของสำนักหอสมุด และการใช้บริการพื้นฐานและบริการพิเศษต่าง ๆ โดยใช้วิธีวิเคราะห์ข้อมูล แสดงความสัมพันธ์ และนำมาแสดงผลในรูปแบบของ Data visualization โดยใช้ภาพ ตัวเลข แผนภูมิ กราฟ และอื่น ๆ ซึ่งช่วยเพิ่มความ

พึงพอใจแก่ผู้บริหารให้สามารถรับรู้ข้อมูลที่มากขึ้น และสามารถนำข้อมูลที่ได้ไปใช้ในการสนับสนุนการตัดสินใจ

รัตนา สุวรรณวิษนีย์และปราณี มณีรัตน์ (2560) ระบบบริหารจัดการงานวิจัยแห่งชาติ (National Research Management System : NRMS) ของสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) นำมาใช้งานในการบริหารจัดการข้อมูลงานวิจัยและงบประมาณวิจัยของหน่วยงานภาครัฐ และแหล่งทุนของหน่วยงานต่าง ๆ แต่ทาง วช. ยังขาดเครื่องมือในการวิเคราะห์ข้อมูลจำนวนมากในระบบ NRMS ให้ได้ทันเวลาและนำเสนอข้อมูลในหลากหลายมิติที่ตรงตามความต้องการใช้ประโยชน์ของผู้ใช้ เพื่อนำเสนอรัฐบาลหรือหน่วยงานนโยบายสำหรับใช้ในการตัดสินใจเชิงนโยบาย ทั้งนี้จึงนำเสนอการพัฒนาระบบรายงานรูปแบบหลายมิติเพื่อสนับสนุนการตัดสินใจเชิงนโยบายของสำนักงานคณะกรรมการวิจัยแห่งชาติ โดยใช้ระบบข่าวกรองทางธุรกิจ (Business Intelligence : BI) พัฒนารายงานรูปแบบหลายมิติตามฟิลต์ เช่น จำแนกตามปีแหล่งทุน กระทรวง หน่วยงาน นโยบายและยุทธศาสตร์การวิจัย และประเด็นการวิจัย เป็นต้น อีกทั้งการนำเสนอรายงานลักษณะแดชบอร์ด (Dashboard) สนับสนุนการตัดสินใจเชิงนโยบายของเจ้าหน้าที่และผู้บริหาร วช. เพื่อให้มองเห็นภาพรวมข้อมูลวิจัยของประเทศได้ การพัฒนาระบบรายงานงานวิจัยนี้ใช้ซอฟต์แวร์ทาบิล (Tableau) ในการพัฒนาระบบรายงาน ผลการประเมินผู้ใช้งานระบบที่พัฒนาขึ้น พบว่าเจ้าหน้าที่ วช.ที่มีความพึงพอใจต่อระบบอยู่ในระดับพอใจ ทำให้เจ้าหน้าที่ วช. สามารถนำเสนอรายงานต่อผู้บริหารได้รวดเร็วขึ้น และหลากหลายมิติมากขึ้น

วิไลลักษณ์ ตรีพีช (2562) การวิจัยนี้มีวัตถุประสงค์เพื่อการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล โดยใช้ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ ประจำปี 2551 - 2557 จำแนกตามโรงพยาบาลทั่วประเทศ จากเว็บไซต์ data.go.th ซึ่งผู้วิจัยได้ทำการเตรียมข้อมูลให้เหมาะสมสำหรับการจำแนกข้อมูลโดยตัดข้อมูลที่ไมเกี่ยวข้องออกไปและปรับให้จำนวนข้อมูลของคำตอบหรือผลลัพธ์มีสัดส่วนใกล้เคียงกัน จากนั้นทำการจำแนกข้อมูลโดยกำหนดให้เอทริบิวต์ผลการรักษาเป็น คำตอบที่ต้องการพยากรณ์สำหรับอัลกอริทึมที่ผู้วิจัยเลือกใช้ในการสร้างโมเดล ได้แก่ Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Decision Tree, Random Forest, และ Gradient Boosted Trees (XGBoost)

วิโรจน์ แสงมณ (2560) งานวิจัยนี้เป็นการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยมีวัตถุประสงค์เพื่อ 1. วิเคราะห์ปัจจัยที่ส่งผลต่อโอกาสการกลับมารักษาซ้ำของผู้ป่วยโรคเบาหวาน 2. สร้างโมเดลสำหรับ ทำนายโอกาสการกลับเข้ามารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำ

เหมือนข้อมูล 3. เปรียบเทียบประสิทธิภาพ ของโมเดลที่สร้างจากเทคนิค Decision tree ,Naive Bays และ K-Nearest Neighbors ผลการสร้างโมเดลในการทำนายโอกาสการกลับมา รักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล คือโมเดลที่สร้างโดยใช้เทคนิค Decision Tree มีประสิทธิภาพการทำนายสูงสุดคิดเป็นร้อยละ 85.5 ซึ่งสามารถ นำผลที่ได้ไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจ ในส่วนของแนวทางการรักษาของแพทย์ และผู้ป่วยได้ต่อไป

อุษา บิ๊กกินส์ (2557) จากการศึกษาเรื่อง ปัจจัยที่มีผลต่อพฤติกรรมเมาแล้วขับของผู้ขับขีรถจักรยานยนต์ในเขต กรุงเทพมหานคร ซึ่งมีวัตถุประสงค์เพื่อค้นหาปัจจัยที่มีผลต่อการขับขีรถจักรยานยนต์ในขณะมีเมามา และเพื่อศึกษาพฤติกรรมการเปิดรับสื่อของผู้ขับขีรถจักรยานยนต์ขณะมีเมามา พฤติกรรมการเปิดรับสื่อของผู้ที่เมาแล้วขับ วิทยุรุ่นเปิดรับโทรทัศน์ และอินเทอร์เน็ตมากที่สุด พวกเขารับรู้ ข้อความจากสติ๊กเกอร์ บิลบอร์ด โปสเตอร์ และโทรทัศน์ พวกเขาสนใจการ กระตุ้นความกลัวผ่านสื่อ ดังนั้น ควรมีวิธีการแก้ปัญหา 2 วิธีคือการขับเคื่อนนโยบาย และการ สื่อสารสาธารณะ ได้แก่ ประชาสัมพันธ์โครงการและกิจกรรมต่าง ๆ ทางเว็บไซต์ การนำเสนอข่าว การขับรถขณะมีเมามาและการใช้การรณรงค์สื่อสารการตลาด การรณรงค์ข้อมูลข่าวสาร และการรณรงค์ด้วยการประกาศโฆษณาในบริการสาธารณะ

ภูมิพัฒน์ ดวงกลาง และรัชนี เครือแก้ว (2562) แบบจำลองการทำนายแบบอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายอัตโนมัติ ผู้วิจัยได้สร้างแบบจำลองการทำนายแบบอากาศยานอัตโนมัติที่ไม่พึงพามนุษย์ โดยใช้ข้อมูลจากระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านของกรมข่าวทหารอากาศเป็นข้อมูลในการวิจัยเพื่อค้นหาแบบจำลองที่เหมาะสม ด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification) ได้แก่ Naive Bayes, Deep Learning, Decision Tree และ Random Forest ซึ่งผลการวิจัยพบว่า แบบจำลองที่เหมาะสมที่จะนำไปใช้งานกับกระบวนการวิเคราะห์และระบุเป้าหมายไม่ทราบฝ่ายมากที่สุด คือแบบจำลองที่ได้จากเทคนิค Random Forest เนื่องจากมีค่าความถูกต้องในการทำนายผลสูง ผู้วิจัยเชื่อมั่นว่าแบบจำลองดังกล่าวจะสามารถเพิ่มประสิทธิภาพให้กระบวนการตัดสินใจของผู้บังคับบัญชากระทำได้รวดเร็วและเหมาะสมยิ่งขึ้น อีกทั้งผลที่ได้จากงานวิจัยนี้จะสามารถแก้ปัญหาการขาดแคลนบุคลากรผู้เชี่ยวชาญ และลดการพึ่งพาความชำนาญของเจ้าหน้าที่เฉพาะทางลงได้ สอดคล้องกับแนวความคิดการใช้ระบบอาวุธที่ฉลาด (Smart Weapon Systems) ที่นำไปสู่การเลือกกลยุทธ์ที่ฉลาด (Smart Tactics) ของกองทัพอากาศอีกด้วย และทางวิจัยทำการสร้างแบบจำลองสำหรับวิเคราะห์และทำนายแบบอากาศยานจากข้อมูลของ



ระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านนั้น จะอ้างอิงกระบวนการมาตรฐาน CRISP-DM

หทัยพร หวังเซย และ ณัฐพร นันทจิระพงศ์ (2561) ระบบสารสนเทศเพื่อการบริหารจัดการรายงานการดำเนินงานของโรงแรมด้วย Power BI ถูกพัฒนาขึ้นโดยมีวัตถุประสงค์เพื่อเปลี่ยนรูปแบบรายงานการวิเคราะห์และสรุปผลการดำเนินงานของโรงแรมจากเอกสารเป็นสารสนเทศประกอบการตัดสินใจของผู้บริหารในรูปแบบของหน้ากระดานที่นำเสนอผลการวิเคราะห์ข้อมูลสำหรับผู้บริหารในมุมมองต่าง ๆ ประกอบด้วยแผนภูมิหลายชนิด หลายมิติ การทำงานแบ่งเป็น 4 ส่วน ได้แก่ 1) การลงชื่อเข้าใช้งานระบบ 2) การประมวลผลข้อมูล 3) การจัดเก็บข้อมูลบน cloud และ 4) การออกรายงาน ซึ่งแบ่งตามข้อกำหนดของฝ่ายบริหารและมิติทางธุรกิจได้เป็น 7 รูปแบบ โดยหลักการและเทคโนโลยีที่นำมาประยุกต์ใช้ในการพัฒนาระบบ ได้แก่ การวิเคราะห์ข้อมูลขนาดใหญ่ ธุรกิจอัจฉริยะ ระบบสารสนเทศเพื่อการบริหารคลังข้อมูล และกระบวนการสกัดข้อมูล ส่วนเครื่องมือที่ใช้ในการพัฒนาระบบ คือ Power BI desktop, Microsoft SQL server และ Microsoft Visual Studio ผลการพัฒนาระบบสามารถทำงานได้ตามวัตถุประสงค์และขอบเขตที่กำหนดไว้ ผลการสำรวจความพึงพอใจของผู้ใช้ที่มีต่อภาพรวมของระบบพบว่า อยู่ในระดับดี (ค่าเฉลี่ย เท่ากับ 4.5 และส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 0.71) จึงสรุปได้ว่า ระบบสามารถสรุปผลข้อมูลการดำเนินงานของโรงแรมและรายงานได้รวดเร็ว ทำให้ผู้บริหารสามารถวางแผน ตัดสินใจ แก้ไขปัญหา และกำหนดทิศทางการทำงานของโรงแรมได้อย่างมีประสิทธิภาพ

ปัทมา เทียงสมบุญ และ นิเวศ จิระวิชิตชัย (2561) การพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์และการตัดสินใจของผู้บริหาร กรณีศึกษากลุ่มโรงพยาบาล งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์และการตัดสินใจของผู้บริหาร กรณีศึกษากลุ่มโรงพยาบาล โดยใช้ระบบธุรกิจอัจฉริยะ (Business Intelligence) ผ่านโปรแกรมไมโครซอฟท์ เพาเวอร์ บีไอ (Microsoft Power BI) และการจัดการฐานข้อมูลด้วยออราเคิล ดาต้าเบส 11จี (Oracle Database 11g) เพื่อสนับสนุนการวิเคราะห์ข้อมูลการรักษาพยาบาลของผู้บริหาร โดยนำข้อมูลจากระบบสารสนเทศโรงพยาบาล (Hospital Information System) และข้อมูลโรคไม่ติดต่อเรื้อรัง (Non-Communicable Diseases : NCDs) มาสร้างรายงานอัจฉริยะ (Dashboard) โดยระบบสามารถพยากรณ์ข้อมูลและปรับเปลี่ยนมุมมองในการวิเคราะห์ช่วยในการตัดสินใจอย่างรวดเร็วและถูกต้อง จากการประเมินความพึงพอใจของระบบโดยผู้บริหารจำนวน 30 คน ได้ค่าเฉลี่ยเท่ากับ 4.15 และค่าส่วนเบี่ยงเบนมาตรฐาน

เท่ากับ 0.67 ซึ่งสรุปได้ว่าระบบธุรกิจอัจฉริยะที่พัฒนาขึ้นมาในงานวิจัยชิ้นนี้ใช้ในการสนับสนุนการตัดสินใจของผู้บริหารได้เป็นอย่างดี

อัจฉราภรณ์ จุฑาพาด (2559) งานวิจัยเรื่องการพัฒนาาระบบสารสนเทศเพื่อการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกต้นไม้ตัดสินใจ มีวัตถุประสงค์คือ 1) เพื่อพัฒนาาระบบสารสนเทศเพื่อใช้ในการพยากรณ์จำนวนนักศึกษาใหม่ ที่จะเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ (Decision Tree) และ 2) เพื่อวัดประสิทธิภาพและความถูกต้องของตัวต้นแบบที่พัฒนาขึ้นที่นำมาใช้ในการพยากรณ์จำนวนนักศึกษาใหม่ของคณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด ผลการวิจัยพบว่า 1) การพัฒนาาระบบสารสนเทศเพื่อใช้ในการพยากรณ์จำนวนนักศึกษาใหม่ที่จะเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ซึ่งจากการพัฒนาตัวต้นแบบในการพยากรณ์แบ่งออกเป็น 3 วิธีได้แก่ การตรวจสอบไขว้ การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ และการแบ่งชุดข้อมูลและการทดสอบออกจากกัน จำนวนหรือปริมาณของกฎขึ้นอยู่กับความน่าจะเป็นและการกำหนดค่าความเชื่อมั่นผู้วิจัยได้กำหนดไว้ 0.25 หรือ 25 เปอร์เซ็นต์ แสดงว่าวิธีการแบ่งชุดข้อมูลและการทดสอบออกจากกันสามารถนำไปใช้ในการพัฒนาตัวต้นแบบในการพยากรณ์นักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องแม่นยำสูง และเหมาะสมกว่าวิธีอื่น หากนำกฎการจำแนกจำนวน 315 กฎที่ได้มาจากตัวต้นแบบการพยากรณ์มาสร้างเพื่อการทำนาย โดยมีการนำชุดข้อมูลของผู้สมัครในปีถัดไปมาทดสอบกับตัวต้นแบบหรือกฎที่ได้จะช่วยให้สามารถทราบจำนวนนักศึกษาใหม่ในปีนั้นๆได้ และยังเป็นแนวทางกำหนดรูปแบบและวิธีการ การประชาสัมพันธ์ การแนะแนวหรือรับนักศึกษาใหม่ได้อย่างเหมาะสมกับกลุ่มเป้าหมายต่อไป และ 2) การวัดค่าประสิทธิภาพต่าง ๆ ที่วัดได้จะมีค่าใกล้เคียงกันหรือมีค่าเท่ากันในบางตัวต้นแบบ โดยตัวต้นแบบการพยากรณ์ที่พัฒนาด้วยวิธีการแบ่งชุดข้อมูลและการทดสอบออกจากกัน วัดค่าความถูกต้องได้เท่ากับร้อยละ 97.34 ค่าความแม่นยำเท่ากับร้อยละ 98.56 ค่าความระลึกรเท่ากับร้อยละ 97.00 และค่าความถ่วงดุลเท่ากับร้อยละ 97.13 ซึ่งมีค่าประสิทธิภาพสูงทุกค่า แสดงว่าตัวต้นแบบที่ใช้ในการพยากรณ์ที่ผู้วิจัยพัฒนาขึ้นมีความถูกต้องแม่นยำในการพยากรณ์ในการรับสมัครนักศึกษาใหม่ โดยที่ผู้วิจัยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจมากที่สุด

เบญจมาศ ปิยะ (2558) งานวิจัยเรื่องการศึกษาพฤติกรรมผู้ใช้งานตู้แช่แข็งพาณิชย์ด้วยเทคนิคต้นไม้ตัดสินใจ เนื่องจากปัญหาการกินพลังงานไฟฟ้าของตู้แช่ที่สูงทำให้ต้นทุนของร้านค้าสูงตามไปด้วยปัจจัยหลัก ๆ คือพฤติกรรมการใช้งานเปิดปิดตู้แช่ของลูกค้า ผู้วิจัยจึง

ทำการศึกษาพฤติกรรมการใช้งานตู้แช่ของผู้ใช้งานในร้านค้าโดยใช้อุปกรณ์ Arduino ในการบันทึกข้อมูลการใช้งานและข้อมูลการใช้พลังงานของตู้แช่ เพื่อนำข้อมูลที่ได้มาวิเคราะห์หาวิธีการที่จะลดการใช้พลังงานไฟฟ้าของตู้แช่ให้ได้อย่างเหมาะสมและนำเทคนิคการจำแนกข้อมูลต้นไม้การตัดสินใจมาเพื่อช่วยในการจำแนกข้อมูลแต่ละส่วน ผลที่ได้จากการดำเนินงานวิจัยนี้คือ จะได้ระบบแจ้งเตือนการเปิดประตูตู้แช่ด้วยเสียงเมื่อเปิดประตูตู้แช่เกินระยะเวลาที่กำหนดเอาไว้ จากการศึกษาพฤติกรรมการใช้งานเปิดปิดตู้แช่เพื่อนำมาปรับเปลี่ยนเสียงแจ้งเตือนผู้ใช้งานไม่ให้เปิดใช้งานเป็นเวลานาน และส่งผลกับการประหยัดพลังงานไฟฟ้าได้จริง โดยมีการเก็บข้อมูลการใช้งานจริงและนำเอาข้อมูลมาประมวลผลโดยโปรแกรม WEKA และได้ช่วงเวลาที่เหมาะสมกับการแจ้งเตือน เวลาที่ตั้งในการแจ้งเตือนคือไม่เกิน 2 นาทีผลที่ได้คือมีเวลาการเปิดใช้งานที่น้อยลงและทำให้ประหยัดพลังงานไฟฟ้า หลังจากได้ทำการตั้งเวลาแจ้งเตือนการเปิดปิดประตูตู้แช่ไปแล้วจากผลการเก็บข้อมูล พบว่ามีการเปิดปิดประตูตู้แช่ในระยะเวลาที่สั้นลงจากเดิมก่อนที่จะทำการติดตั้งสัญญาณเตือนการใช้งาน 18.75 % และมีผลทำให้การใช้พลังงานไฟฟ้าของตู้แช่ลดลงจากเดิม 4.4 %

## 2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น คณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม Weka 3.8.4 และโปรแกรม RapidMiner Studio 9.5.1 ในการสร้างโมเดล Decision Tree เพื่อทำการเปรียบเทียบ และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน web browser ที่เป็นที่นิยมในยุคอินเทอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา