บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

2.2 ทฤษฎี

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ และการแสดงผลข้อมูลบนเว็บไซต์ ผู้วิเคราะห์ข้อมูลได้ศึกษาหลัก การและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

2.1.1.1 Data analytics เป็นศาสตร์ของการวิเคราะห์ข้อมูลต่าง ๆ จาก Big data เพื่อช่วยในด้านธุรกิจหรือตามวัตถุประสงค์อื่น ๆ ที่ต้องการ ซึ่งเริ่มต้นจากความสามารถของเรา ในการนำข้อมูลหล่านั้น มาให้อยู่ในรูปแบบที่พร้อมจะประมวลผลได้โดยเทคโนโลยีหรือชุดคำสั่งและแบบจำลองที่สร้างขึ้นเพื่อนำ ข้อมูลที่ได้ผ่านการวิเคราะห์แล้วมาใช้ หรือ แปลความหมายโดยบุคคลที่ได้รับการฝึกอบรมการใช้เทคโนโลยีเหล่านั้น รูปแบบของการวิเคราะห์ข้อมูลสามารถแบ่งได้ดังนี้

1) การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive analytics) เป็นการวิเคราะห์เพื่อแสดงผลของรายการทางธุรกิจ เหตุการณ์ หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้นในลักษณะที่ง่ายต่อการเข้าใจ หรือต่อการตัดสินใจตัวอย่างเช่นรายงานการขายรายงานผลการดำเนินงาน

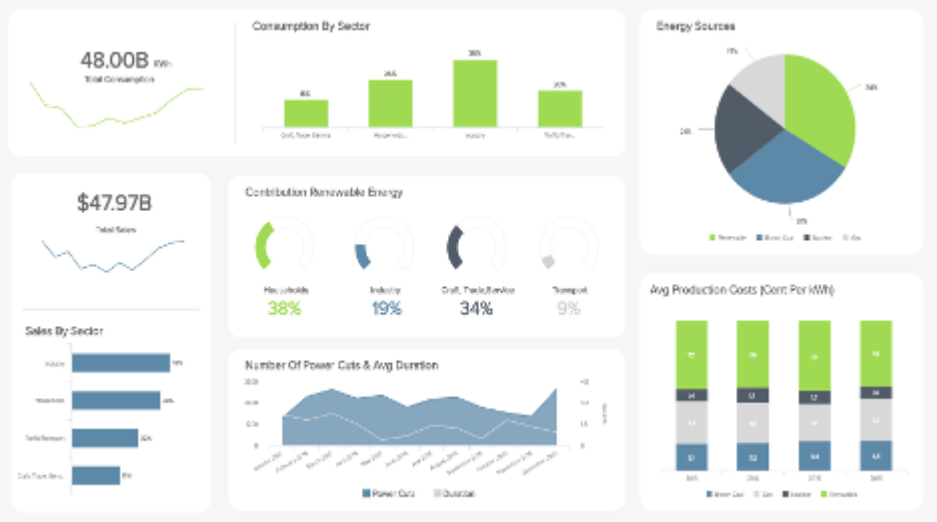
2) การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้นปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น ตัวอย่างเช่นความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่ช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

3) การวิเคราะห์แบบพยากรณ์ (Predictive analytics) เป็นการวิเคราะห์เพื่อพยากรณ์สิ่งที่ กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้น โดยใช้ข้อมูลที่ได้เกิดขึ้นแล้วกับแบบจำลองทางสถิติ หรือ เทคโนโลยีปัญญาประดิษฐ์ต่าง ๆ (Artificial intelligence) ตัวอย่าง เช่น การพยากรณ์ยอดขาย การพยากรณ์ผลประชามติ

4) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive analytics) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด เป็นทั้งการพยากรณ์สิ่งต่าง ๆ ที่จะเกิดขึ้นข้อดี ข้อเสีย สาเหตุ และระยะเวลาของสิ่งที่จะเกิดขึ้นและการให้คำแนะนำทางเลือกต่าง ๆ ที่มีอยู่และผลของแต่ละทางเลือก

2.1.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization เป็นเครื่องมือสำคัญที่ช่วยให้เห็นภาพของข้อมูลเชิงลึกที่ผ่านการเก็บข้อมูล การจัดการข้อมูล การวิเคราะห์ข้อมูล แล้วรวบรวมได้เป็นชุดข้อมูล หากต้องการประสบความสำเร็จในโลกธุรกิจที่มีการแข่งขันสูง ความสามารถในการตัดสินใจที่ถูกต้องในเวลาที่เหมาะสม โดยอาศัยข้อมูลเชิงลึกที่รัดกุมเป็นสิ่งสำคัญ ปัจจุบันองค์กรธุรกิจส่วนใหญ่ในประเทศไทย เริ่มใช้ประโยชน์จากการทำ Data Visualization เพราะว่าเครื่องมือนี้เป็นกุญแจสำคัญในการแสดงผลของข้อมูลให้เข้าใจง่าย ซึ่งจะช่วยให้ทุกแผนกในฝ่ายงานได้ใช้ประโยชน์ ในการหาแนวโน้มและความสัมพันธ์เชื่อมโยง อีกทั้งทำให้ทราบข้อมูลได้เร็วขึ้น ส่งผลให้การทำงานขององค์กรมีประสิทธิภาพเพิ่มขึ้น จนดำเนินการถึงการวางกลยุทธ์ภายในองค์กร และสามารถนำไปใช้กับการทำเว็บไซต์ได้ Data Visualization ช่วยให้การสื่อสารข้อมูล มีความรวดเร็วและเข้าใจง่ายมากขึ้น



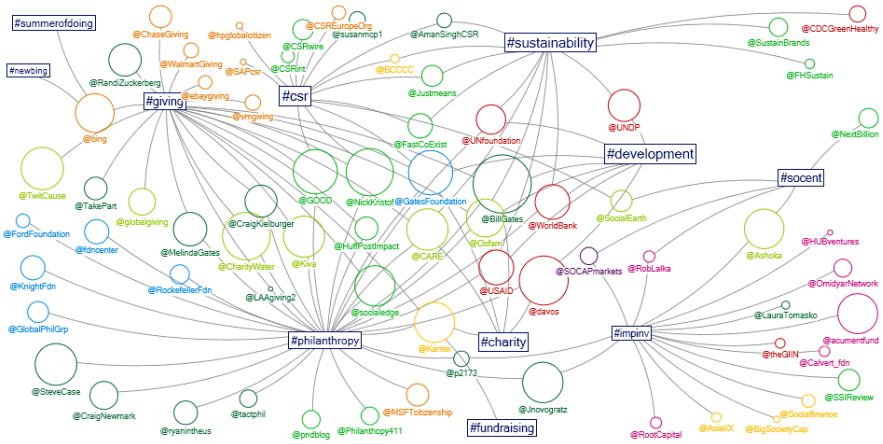
ภาพที่ 2. 1 แสดง Data Visualization

(ที่มา : https://bit.ly/31GJ4pE)

พื้นฐานของการสร้าง Data Visualization คือ การ Mapping ส่วนข้อมูลกับส่วนของ Graphic เข้าด้วยกัน ซึ่งตอนนี้มีโปรแกรมสำเร็จรูปในการสร้าง Data visualization หลากหลายโปรแกรม มีฟังก์ชันการใช้งานที่เข้าใจง่าย เช่น การสร้างฟิลเตอร์ การออกแบบเพื่อให้งานการวิเคราะห์ข้อมูลมีความยืดหยุ่น เป็นต้น ตัวอย่างรูปแบบ Data Visualization ที่แสดงผลให้อย่างสนใจ มีดังนี้

1. NETWORK/FLOW

Network Graph ใช้แสดงความเชื่อมโยงของเครือข่ายหรือความสัมพันธ์ในกลุ่ม

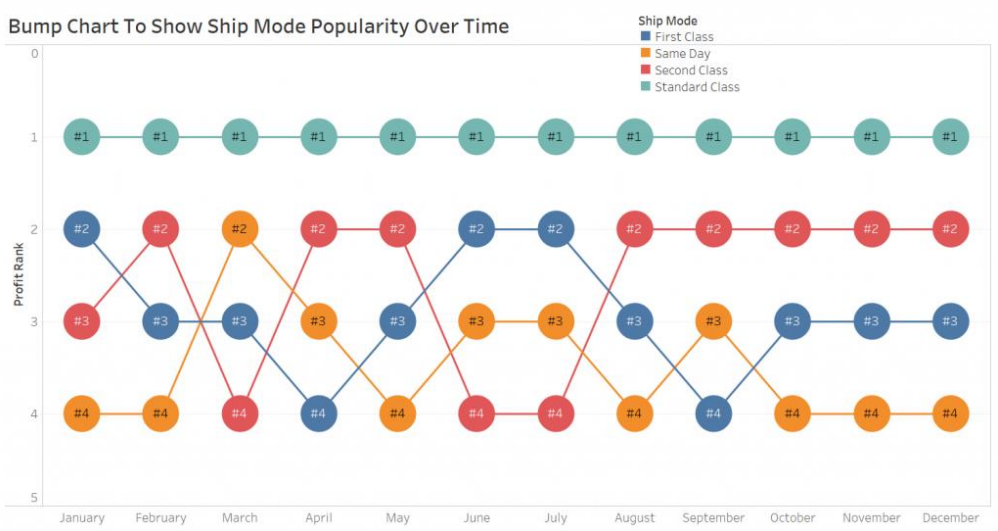


ภาพที่ 2. 2 แสดง Network Graph

(ที่มา : https://bit.ly/31GJ4pE)

2. RANKING

Bump Chart ใช้แสดงการเปลี่ยนแปลงลำดับของข้อมูลในหลายช่วงเวลา โดยใช้สีเพื่อช่วยให้อ่านข้อมูลได้ง่ายขึ้น

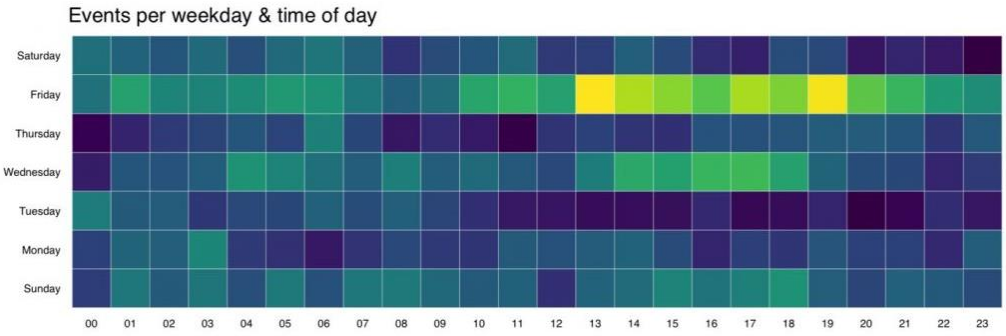


ภาพที่ 2. 3 แสดง Bump Chart

(ที่มา : https://bit.ly/31GJ4pE)

3. RELATIONSHIP

Heat map ใช้แสดงรูปแบบความสัมพันธ์ของข้อมูล โดยจะแสดงออกมาในรูปแบบของ “สี” ซึ่งแต่ละสีจะบ่งบอกถึงระดับความถี่ของพฤติกรรม แต่ใช้อ่านค่าความแตกต่างเล็กน้อยได้ยาก

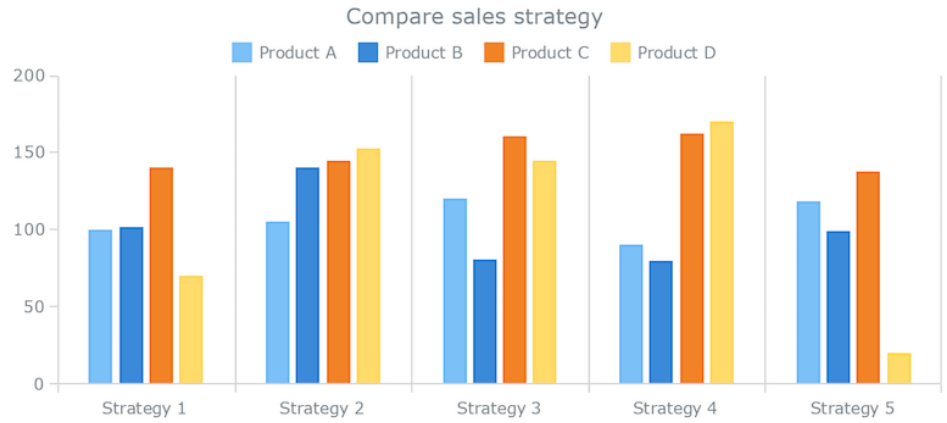


ภาพที่ 2. 4 แสดง Heat map

(ที่มา : https://bit.ly/31GJ4pE)

4. COMPARISON

Stacked Bar Chart ใช้เปรียบเทียบค่าผลรวมและสัดส่วนจากข้อมูลหลายกลุ่ม อาจอ่านสัดส่วนยากเมื่อมีข้อมูลหลายกลุ่มมากเรับประทานไป

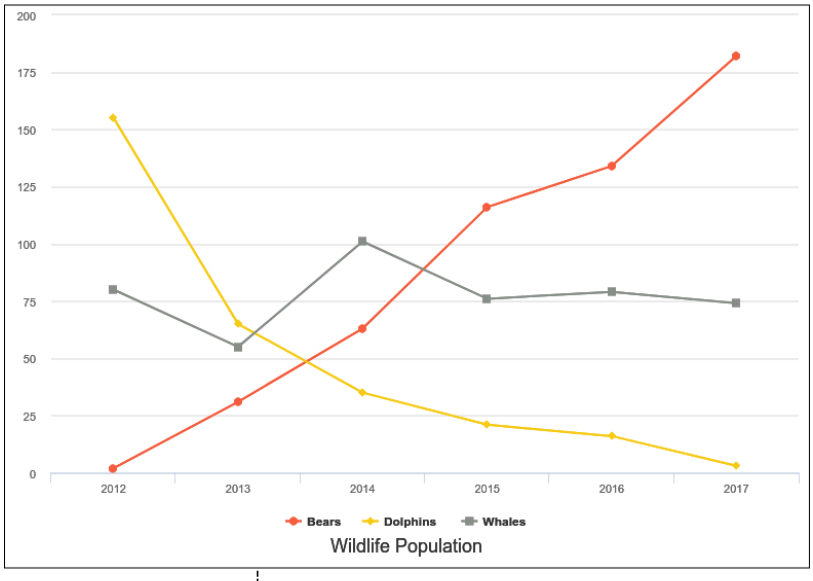


ภาพที่ 2. 5 แสดง Stacked Bar Chart

(ที่มา : https://bit.ly/31GJ4pE)

5. TIME–SERIES

Line Chart ใช้เปรียบเทียบข้อมูลเพื่อดูแนวโน้ม (Trends) โดยอาจจะเทียบกับมิติของเวลา time series

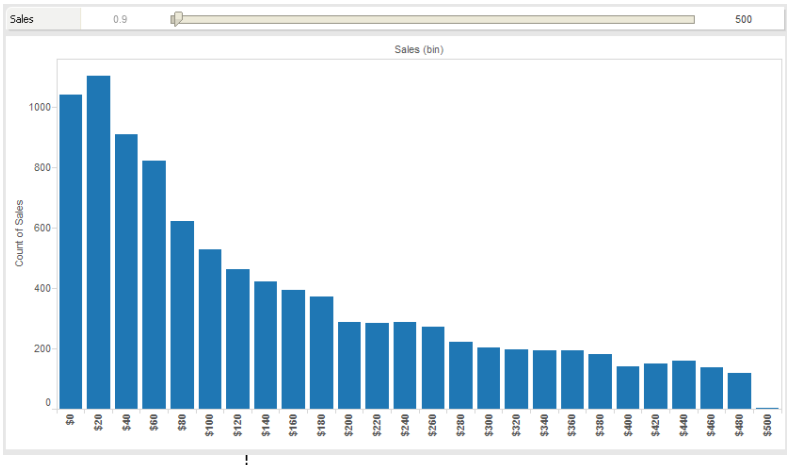


ภาพที่ 2. 6 แสดง Line Chart

(ที่มา : https://bit.ly/31GJ4pE)

6. DISTRIBUTION

Histogram กราฟแท่งแบบเฉพาะที่แสดงความสัมพันธ์ระหว่างข้อมูลเป็นหมวดหมู่ที่เรียกว่าชั้นข้อมูลกับความถี่ของข้อมูล เพื่อดูการกระจายของข้อมูล ลดระยะห่างระหว่างแท่งเพื่อให้เห็น “รูปทรง” ของข้อมูลชัดขึ้น

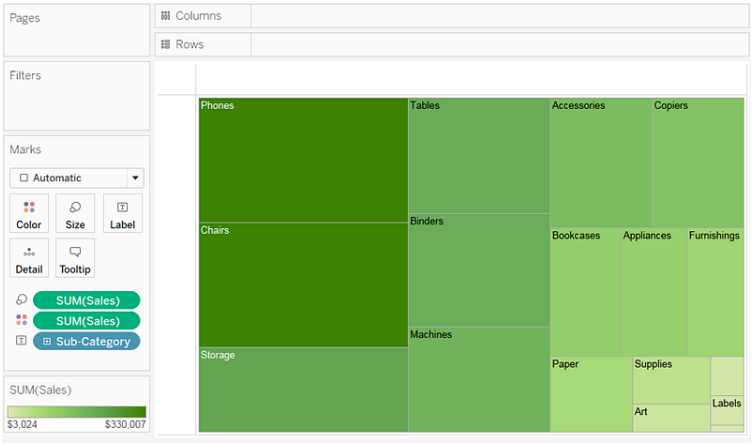


ภาพที่ 2. 7 แสดง Histogram

(ที่มา : https://bit.ly/31GJ4pE)

7. COMPOSITION

Tree map ใช้แสดงสัดส่วนของข้อมูลแบบมีลำดับชั้น (hierarchy) แสดงเป็นรูปสี่เหลี่ยม โดยจะแสดงทั้งในแง่ของสีเดียวกันจัดเป็นกลุ่มเดียวกัน อาจอ่านสัดส่วนยากเมื่อมีกลุ่มขนาดเล็กจำนวนมาก

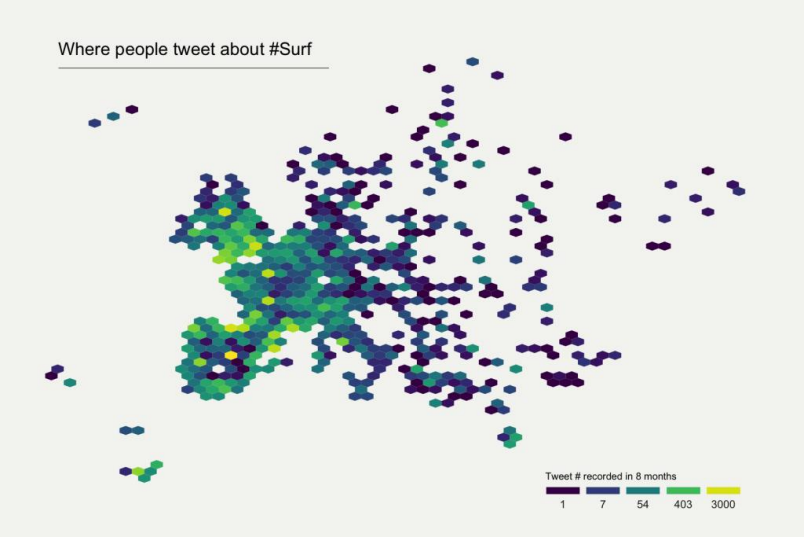


ภาพที่ 2. 8 แสดง Tree map

(ที่มา : https://bit.ly/31GJ4pE)

8. SPATIAL

Hexbin Map เป็นการแสดงผลข้อมูลด้วยรูปทรงหกเหลี่ยมที่มีขนาดเท่ากัน โดยกำหนดสีแทนระดับความถี่ของจำนวนข้อมูล ส่วนมากนำไปใช้ลงบนตำแหน่งทางภูมิศาสตร์ เพื่อให้เห็นภาพได้ชัดเจน



ภาพที่ 2. 9 แสดง Hexbin Map

(ที่มา : https://bit.ly/31GJ4pE)

จากที่ได้กล่าวไปจะเห็นได้ว่าข้อมูลแบบ Data Visualization มีพลังทางการตลาดอย่างมาก เพราะสามารถแสดงผลและเพิ่มความสนใจได้เป็นอย่างดี นอกจากนี้ยังทำให้ผู้ได้รับข้อมูลสามารถจินตนาการตามข้อมูลที่แสดงได้อีกด้วย หากองค์กรได้ลองนำการแสดงข้อมูลแบบ Data Visualization มาใช้ก็จะช่วยในการจัดการระดับสายงานแผนกอื่น ๆ ใน บริษัท ส่งผลให้การวิเคราะห์ข้อมูลและตัดสินใจในสายงานต่าง ๆ มีประสิทธิภาพ ซึ่งดีต่อธุรกิจและทีม หรือนำมาใช้ในหน้าแรกของเว็บไซต์ เพื่อทำการโปรโมทโปรโมชั่นหรือข้อเสนอทางการตลาดของธุรกิจ แล้วจะพบว่า Data Visualization ช่วยให้เกิดการตอบสนองกับข้อความมากขึ้น

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

การทำเหมืองข้อมูลหรือ Data Mining เทคโนโลยีเกี่ยวกับฐานข้อมูลได้เริ่มต้นขึ้นตั้งแต่ปี 1960 โดยพัฒนามาจาก file processing จากนั้นพัฒนามาในรูปแบบตารางและเริ่มมีการใช้ภาษาในการเรียกข้อมูล (Query Language) ข้อมูลที่เก็บเริ่มมีจำนวนมากขึ้นทำให้การจัดเก็บมีความซับซ้อนและมีประสิทธิภาพ การเพิ่มขึ้นของข้อมูลเป็นจำนวนมากและสะสมไว้ในฐานข้อมูลนั้นมีขนาดใหญ่เรับประทานกว่าที่กำลังคนจัดการได้ เป็นผลให้เกิดเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูลและหาความเป็นไปได้ของข้อมูลที่เป็นประโยชน์

วิวัฒนาการของการทำเหมืองข้อมูล

- ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่น่าเชื่อถือและป้องกันการสูญหายได้เป็นอย่างดี
- ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อประโยชน์ใน การนำไปวิเคราะห์ และการตัดสินใจอย่างมีคุณภาพ
- ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปในฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกด้านขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ
- ปี 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผลโดยการสร้างแบบจำลองและความสัมพันธ์ทางสถิติ

เนื่องด้วยปัจจุบันเป็นยุคที่ข้อมูลและสารสนเทศมีความสำคัญ การเผยแพร่และสื่อสารข้อมูล ข่าวสาร ที่ตรงกับความต้องการของผู้ใช้จึงเป็นสิ่งจำเป็น การประยุกต์เทคโนโลยีสารสนเทศเพื่อช่วยในการสื่อสารข้อมูลจำนวนมากให้แก่ผู้ ใช้ เช่น การให้บริการเว็บไซต์เพื่อเผยแพร่ข้อมูลข่าวสารและแลกเปลี่ยนความรู้ จึงเป็นเครื่องมือที่สำคัญในการสื่อสารข้อมูลถึงผู้ใช้จำนวนมาก ดังนั้นการศึกษาเกี่ยวกับพฤติกรรมของผู้ใช้บริการเว็บไซต์ จะช่วยให้องค์กรสามารถนำข้อมูลมาใช้ในการวางแผนพัฒนาเว็บไซต์ ให้ตรงกับความต้องการใช้งานหรือใช้ในการวางแผนกลยุทธ์ เพื่อสร้างความได้เปรียบทางการแข่งขัน

การทำเหมืองข้อมูล (Data Mining) คือ วิธีการค้นหาความรู้ความสัมพันธ์ และรูปแบบทั้งหมดที่ถูกซ่อนอยู่ในฐานข้อมูลที่มีปริมาณข้อมูลจำนวนมาก การทำเหมืองข้อมูลสามารถทำการสำรวจ และวิเคราะห์อย่างอัตโนมัติ จากข้อมูลที่มีอยู่ในรูปแบบความหมาย และรูปแบบของกฎ โดยที่ความสัมพันธ์แสดงให้เห็นถึงความรู้ต่าง ๆ ที่เป็นประโยชน์และน่าสนใจ ประโยชน์หลักของ

การทำเหมืองข้อมูลคือการสืบค้นความรู้ในฐานข้อมูลขนาดใหญ่เพื่อให้ได้ความรู้มาใช้ในการตัดสินใจ

การทำเหมืองข้อมูลหากผนวกเข้ากับระบบการจัดการความรู้ที่ได้จากฐานข้อมูลจะทำให้ได้ระบบการค้นหาข้อมูลและความรู้(Knowledge and Data Discovery Management System:KDDMS) เป็นการทำเหมืองข้อมูลที่ใช้เทคนิคการกลั่นกรอง วิเคราะห์ความรู้ที่ซ่อนอยู่ในฐานข้อมูลที่มีปริมาณมาก เพื่อได้ข้อมูลที่ต้องการและมีประโยชน์ สามารถนำข้อมูลที่ได้บริหารจัดการได้อย่างมีประสิทธิภาพ โดยผลลัพธ์มีอยู่หลายแบบ เช่น รูปแบบ (Pattern)ต้นแบบ (Model) กฎ (Rule) การทำนายหรือการคาดการณ์ล่วงหน้า (Prediction) ทั้งผลลัพธ์ที่ได้มาขึ้นอยู่กับการทำเหมืองข้อมูล

2.2.2 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

การออกแบบเว็บไซต์นั้นมีความสำคัญอย่างมากเนื่องจากจะทำให้ผู้ชมเกิดความประทับใจ ชื่นชอบ จนสามารถดึงดูดความสนใจได้จนกระทั่งมีความต้องการที่จะกลับเข้าชมเว็บนั้น ๆ อีกครั้งในอนาคต สำหรับการออกแบบเว็บไซต์นั้น หมายถึงการออกแบบพัฒนาส่วนหน้าแสดงของเว็บให้เกิดความประทับใจแก่ผู้ชม และนำเสนอความเป็นเอกลักษณ์ของหน่วยงานได้ด้วยโปรแกรมฟังก์ชันการใช้งานภายในเว็บไซต์ที่ทำให้ผู้ใช้อยากเข้ามามีปฏิสัมพันธ์ต่อเว็บไซต์นั้น โดยต้องให้ความสำคัญกับปัจจัยต่าง ๆ ดังต่อไปนี้

2.2.2.1 ความสำคัญของการออกแบบเว็บไซต์

เว็บไซต์เป็นสื่อที่อยู่ในการควบคุมของผู้ใช้โดยสมบูรณ์ ดังนั้นการออกแบบเว็บจึงมีความสำคัญที่จะดึงดูดให้ผู้ใช้เข้าชมและทำกิจกรรมร่วมในเว็บไซต์ของเราให้ได้มากที่สุด ซึ่งสามารถสรุปปัญหาต่าง ๆ เกี่ยวกับผู้ใช้ซึ่งเป็นองค์ประกอบหนึ่งในการกำหนดความสำเร็จหรือล้มเหลวของการสร้างเว็บไซต์ ดังต่อไปนี้

1. ผู้ใช้มักไม่อดทนกับอุปสรรคและปัญหาที่เกิดขึ้นจากการออกแบบการใช้งานผิดพลาด

2. ผู้ใช้มีทางเลือกมากขึ้นที่จะเข้าใช้บริการเว็บไซต์อื่น ถ้าหากการออกแบบเว็บไซต์นั้นนั้นเกิดความรู้สึกไม่ประทับใจ

3. ผู้ใช้หลายคนชื่นชอบเว็บไซต์ที่สวยงาม และมีความสะดวกในการใช้งานมากกว่าเว็บที่ดูสับสนวุ่นวาย มีข้อมูลจำนวนมากแต่ค้นหาข้อมูลที่ต้องการไม่พบ หรือใช้เวลามากในการแสดงหน้าเว็บ

4. หากผู้ใช้มีความประทับใจกับเว็บไซต์หนึ่ง ๆ แล้วจะทำให้อยากกลับมาเข้าใช้อีกครั้ง

2.2.2.2 องค์ประกอบของการออกแบบเว็บไซต์อย่างมีประสิทธิภาพ

การออกแบบเว็บไซต์ที่ดีนั้น จะต้องออกแบบให้เหมาะสมกับกลุ่มเป้าหมาย และ จะต้องคำนึงถึงความสะดวกของผู้ใช้งานเป็นหลัก โดยที่ผู้ออกแบบนั้นต้องให้ความสมดุลระหว่าง ความสวยงามและความสะดวก ซึ่งต้องนึกถึงองค์ประกอบดังต่อไปนี้

1) ความเรียบง่าย (Simplicity) คือ การสื่อเนื้อหาถึงผู้ใช้โดยจำกัดองค์ประกอบเสริมที่ เกี่ยวข้องกับการนำเสนอให้เหลือแต่สิ่งที่จำเป็นเท่านั้น ทั้งการแสดงออกในด้านเนื้อหา ตัวอักษร หรือแม้แต่สีที่ใช้ ทำให้ผู้ใช้รู้สึกใช้งานง่ายและสะดวกไม่ก่อให้เกิดความรู้สึกรำคาญใด ๆ

2) ความสม่ำเสมอ (Consistency) คือ การสื่อให้ผู้ใช้รู้ว่าเป็นเนื้อหาเดียวตลอดทั้งเว็บไซต์ โดยไม่ก็ให้เกิดความสับสนแก่ผู้ใช้ดังนั้นผู้ออกแบบเว็บไซต์ควรออกแบบระบบ Navigation และโทนสีให้มีความสม่ำเสมอคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ (Identity) คือ ต้องออกแบบเพื่อแสดงถึงความเป็นเอกลักษณ์ของ องค์กร ทั้งชุดสี ตัวอักษร รูปภาพกราฟิก เป็นสื่อถึงความเป็นเอกลักษณ์แทบทั้งสิ้น

4) เนื้อหาที่ประประโยชน์ (Useful content) เนื้อที่แสดงออกของเว็บไซต์นั้นเป็นสิ่งที่ สำคัญที่สุดซึ่งต้องแสดงออกถึงความถูกต้องและสมบูรณ์ โดยไม่ฟ้ากับเว็บไซต์อื่น และเป็นเนื้อหา ที่ทันต่อเหตุการณ์ เนื่องจากเป็นสิ่งที่ดึงดูดผู้เข้ามาใช้งานเว็บไซต์

5) ระบบ Navigation ที่ใช้งานง่าย (User–friendly navigation) เป็นส่วนที่สำคัญของ เว็บไซต์เนื่องจากจะทำให้ผู้ใช้ไม่เกิดความสับสนเวลาใช้งานเว็บไซต์ ซึ่งผู้ออกแบบจำเป็นต้องออกแบบให้ผู้ใช้ใช้งานได้ง่ายและสะดวก เช่นถ้าใช้ภาพกราฟิกที่สื่อความหมายควรมีคำอธิบายที่ชัดเจนร่วมด้วย แสดงไว้ส่วนบนหรือส่วนล่างของหน้าเว็บไซต์

6) มีลักษณะที่น่าสนใจ (Visual appeal) ควรสื่อให้เว็บไซต์มีคุณภาพสัมพันธ์กับ องค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่แสดงบนเว็บไซต์นั้นต้องสมบูรณ์ไม่มีร่องรอยของความเสียหายเป็นจุดด่าง หรือมีขอบขั้นบันได สำหรับตัวอักษรต้องเป็นตัวอักษรที่อ่านง่าย และ สบายตาร่วมกับโทนสีที่สวยงาม

7) การใช้งานได้อย่างไม่จำกัด (Compatible) การออกแบบเว็บไซต์นั้นต้องทำให้คนส่วน ใหญ่เข้าถึงให้ได้มากที่สุด โดยไม่มีการบังคับให้ติดตั้งโปรแกรมใด ๆ เพิ่มเติมสามารถแสดงผลได้ ทุกระบบปฏิบัติและที่ความละเอียดของหน้าจอต่าง ๆ กัน

8) คุณภาพการออกแบบ (Design Stability) ต้องออกแบบให้เว็บไซต์มีเนื้อหาที่น่าเชื่อถือ เพื่อดึงดูดให้มีผู้เข้าชมเว็บไซต์เกิดความประทับใจ และมีความเชื่อถือในตัวเว็บไซต์

9) ระบบการใช้งานที่ถูกต้อง (Functional stability) ต้องออกแบบให้เว็บไซต์นั้นมีความแน่นอน ทำหน้าที่ได้อย่างถูกต้อง และสามารถใช้งานได้จริง เช่น ถ้ามีแบบฟอร์มให้ผู้ใช้กรอก ข้อมูล ก็ต้องทดสอบก่อนว่าสามารถใช้งานได้จริง หรือลิงค์ต่าง ๆ ที่แสดงอยู่บนหน้าเว็บไซต์จะต้องเชื่อมโยงไปยังแหล่งข้อมูลนั้น ๆ ได้จริง

2.2.2.3 รูปแบบโครงสร้างของเว็บไซต์

การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างเว็บไซต์แบบเรียงลำดับ

จะเป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิงค์เนื้อหา ก็จะลิงค์ไปทีละหน้า มีทิศทางการเข้าสู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ปุ่มเดินหน้า–ถอยหลังในการกำหนดทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับก็มีข้อเสีย คือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหา เพราะไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาด้วยตัวเองได้

ภาพที่ 2. 10 แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ

(ที่มา : https://bit.ly/2VI4fDW)

2) โครงสร้างเว็บไซต์แบบลำดับขั้น

นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่างๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆ และมีการนำเสนอรายละเอียดย่อยๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีโฮมเพจเป็นจุดเริ่มต้นและจุดร่วมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง

ภาพที่ 2. 11 แสดงโครงสร้างเว็บไซต์แบบลำดับขั้น

(ที่มา : https://bit.ly/2VI4fDW)

3) โครงสร้างเว็บไซต์แบบตาราง

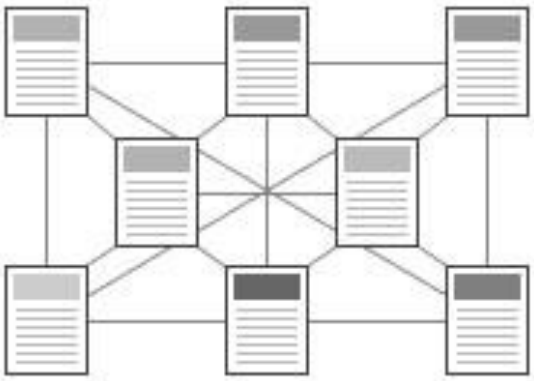
เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีความยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่าง ๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งานสามารถเปลี่ยนทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลา แถมยังทำให้เว็บไซต์มีความทันสมัยขึ้น

ภาพที่ 2. 12 แสดงโครงสร้างเว็บไซต์แบบตาราง

(ที่มา : https://bit.ly/2VI4fDW)

4) โครงสร้างแบบใยแมงมุม

เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถเข้าถึงหน้าเว็บเพจต่าง ๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2. 13 แสดงโครงสร้างแบบใยแมงมุม

(ที่มา : https://bit.ly/2VI4fDW)

ส่วนประกอบสำคัญบนหน้าเว็บเพจ

บนหน้าเว็บเพจ จะมีส่วนประกอบสำคัญที่จำเป็นต้องมีอยู่ 3 ส่วน ได้แก่

1. ส่วนหัวของหน้า (Header)

อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซต์ต่อไป ซึ่งส่วนใหญ่ก็มักจะมีการใส่ภาพกราฟิกให้ดูสวยงาม สิ่งสำคัญหลักๆ เลย ก็คือ โลโก้ ชื่อเว็บไซต์และเมนูหลักที่สามารถลิงค์ไปยังเนื้อหาในหน้าเว็บเพจต่าง ๆ ได้

2. ส่วนของเนื้อหา (Body)

อยู่บริเวณตอนกลางของหน้าเว็บ โดยจะแสดงข้อมูลเกี่ยวกับเนื้อหาบนเว็บแบบคร่าวๆ ซึ่งก็จะมีข้อความกราฟิก ตารางข้อมูลหรือวิดีโอประกอบอยู่ และหากมีเมนูแบบเฉพาะกลุ่มก็จะถูกจัดไว้ในหน้านี้เช่นกัน และที่สำคัญเนื้อหาในส่วนนี้ควรจะมีความกระชับ เข้าใจง่าย มีการใช้รูปแบบตัวอักษรแบบเรียบง่ายและเป็นระเบียบ

3. ส่วนท้ายของหน้า (Footer)

อยู่ล่างสุดของหน้าเว็บ ซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่าง ๆ เพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูลเจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่าง ๆ เกี่ยวกับการใช้งานเว็บไซต์อย่างถูกต้อง เป็นต้น

วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์

การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บ ว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไรต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่มดังต่อไปนี้

1. สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอบโยนและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2. สีโทนเย็น (Cool Colors) เป็นสีแห่งความสุภาพและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้โน้มน้าวจากในระยะไกลได้อีกด้วย

3. สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

สีสามารถสื่อถึงอารมณ์ความรู้สึกและสื่อความหมายของเว็บไซต์นั้น ๆ ได้อย่างชัดเจน ดังนั้นหากเลือกใช้สีไม่เหมาะสมกับเนื้อหาหรือจุดประสงค์ของเว็บ ก็จะทำให้เว็บดูไม่น่าสนใจ ขาดความน่าเชื่อถือและทำให้ผู้ที่เคยเข้ามาใช้บริการไม่คิดจะกลับมาใช้บริการอีก

2.2.3 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

CSS ย่อมาจาก Cascading Style Sheet มักเรียกโดยย่อว่า "สไตล์ชีต" คือภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบ (หรือ "Style") ของเนื้อหาในเอกสาร อันได้แก่ สีของข้อความ สีพื้นหลัง ประเภทตัวอักษรและการจัดวางข้อความ ซึ่งการกำหนดรูปแบบ หรือ Style นี้ใช้หลักการของการแยกเนื้อหาเอกสาร HTML ออกจากคำสั่งที่ใช้ในการจัดรูปแบบการแสดงผล กำหนดให้รูปแบบของการแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลลัพธ์ของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือต้องการ

ควบคุมให้รูปแบบการแสดงผลเอกสาร HTML มีลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสาร ภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการกำหนดรูปแบบ (Style) เอกสาร HTML ถูกเพิ่มเข้ามาครั้งแรกใน HTML 4.0 เมื่อปีพ.ศ. 2539 ในรูปแบบของ CSS level 1 Recommendations ที่กำหนดโดย องค์กร World Wide Web Consortium หรือ W3C

ไวยากรณ์ของ CSS

ไวยากรณ์ของ CSS นั้น มีโครงสร้างหลักของภาษาอยู่ 4 ส่วนด้วยกัน ซึ่งทั้งหมดมีความสำคัญต่อการกำหนดค่าและการใช้งาน CSS ซึ่งรูปแบบการกำหนดค่าจะประกอบไปด้วย property ตามด้วยเครื่องหมายโคลอน (:) เพื่อกำหนดค่าลงไปซึ่งอาจจะเป็นตัวเลข คีย์เวิร์ด หรือ URL และสุดท้ายของการกำหนดค่าคือ เซมิ-โคลอน (;) ที่ใช้ในการระบุว่าสิ้นสุดการกำหนดค่าให้กับproperty ดังตัวอย่างที่ 1

```
Selector {
Properties: Value;
}
```

ตัวอย่างที่ 1 ไวยากรณ์ของ CSS

ส่วนประกอบของโครงสร้างของ CSS มีดังนี้

1. Selector สามารถเป็น HTML Tag ต่าง ๆ ก็ได้เช่น <body>, <p> หรือเป็น Classname หรือ ID ที่ตั้งชื่อให้ก็ได้ซึ่งจะเรียนรู้ในหัวข้อถัดไป

2. เครื่องหมายปีกกาที่ครอบ Properties กับ Value ที่ต้องการกำหนดค่า

3. ชื่อของ Properties เป็นคุณสมบัติในการจัดรูปแบบการแสดงผล เช่น color สำหรับกำหนดสี, font-size สำหรับกำหนดขนาดตัวอักษร

4. ค่า Value ที่ต้องการจะกำหนดให้กับแต่ละ Properties เช่น color: black, fontsize:12px

ภาพที่ 2. 14 แสดงซอร์จโค้ดส่วนประกอบของโครงสร้างของ CSS

(ที่มา : https://bit.ly/2Z0dt0G)

จากซอร์จโค้ดข้างต้นนี้สามารถพิจารณาตามรูปแบบโครงสร้างของ CSS ได้ดังนี้คือ Selector ของ CSS คือ h1, Properties ของ CSS คือ color และ background, Value ของ CSS คือ red และ yellow

จะสังเกตเห็นได้ว่าจะมีเครื่องหมาย Colon (:) ที่เป็นตัวคั่นระหว่างชื่อ Properties กับ Value ที่เป็นการกำหนดค่าให้กับ Properties นั้น ๆ ซึ่งเมื่อสิ้นสุดแล้ว แต่ละ Properties จะจบด้วยเครื่องหมาย semi-colon (;) เสมอ ซึ่งเป็นข้อบังคับของการเขียน CSS

2.2.4 ทฤษฎีเกี่ยวกับการ Visualization

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

Visualization เป็นมากกว่าวิธีการทาง Computer Visualization เป็นการนำเสนอข้อมูลแบบหนึ่งที่ทำการแสดงข้อมูลให้อยู่ในรูปแบบของ Visual Form ซึ่งอาจจะเป็นการแสดงในรูปแบบของรูปภาพ ,กราฟ หรือ แผนภาพ ซึ่ง ผลลัพธ์ของการทำ Visualization คือ การนำข้อมูลต่าง ๆ ที่ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้สามารถสัมผัสได้ ซึ่งกระบวนการต่าง ๆ ในการแสดงหรือนำข้อมูลต่าง ๆ ออกมา อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือ ใช้ในการวิเคราะห์ข้อมูลโดยจุดมุ่งหมายของ Visualization

ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ เพื่อช่วยในการลดช่องว่างระหว่างผู้ใช้และตัวข้อมูล และยังช่วยผู้ใช้สามารถเรียนรู้ ได้อย่างมีประสิทธิภาพ

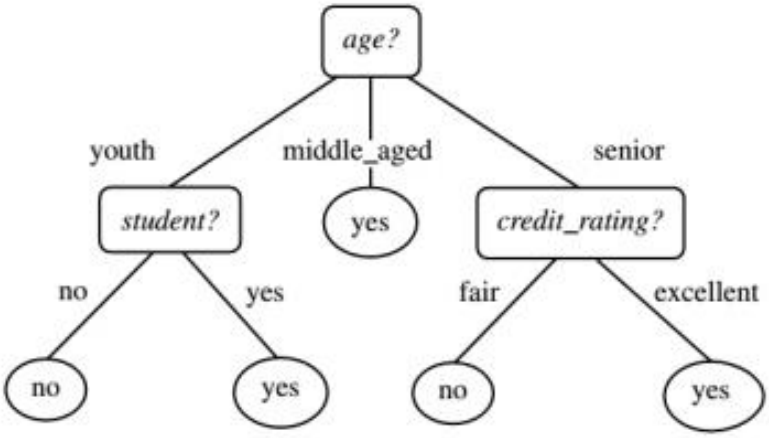
1) วิธีการของ Visualization

การใช้ Visualization เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้จะทำขั้นตอนหลัก ๆ อยู่ 3 ขั้นตอน คือOverview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้ต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้ก็จะทำการตัดสินใจเลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้ก็จะทำการเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็จะอ้างอิงหรือพัฒนาจากขั้นตอนเหล่านี้ซึ่งVisualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ใช้ต้องการโดยอาจจะใช้หลายๆวิธีการรวมกันเพื่อให้ได้ข้อมูลที่ผู้ใช้ต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามารถนำมาผ่านกระบวนการของVisualization มีลักษณะต่าง ๆ มากมายดังนี้ ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง,ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi Dimensional Data ได้แก่Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อข่าวต่าง ๆ และ Web Document,Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document ,Algorithms และ Softwareได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลาย

2.2.5 ทฤษฎีเกี่ยวกับการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ

การจำแนกข้อมูลด้วยต้นไม้ตัดสินใจจะเป็นกระบวนการสร้างต้นไม้ขึ้นเพื่อใช้ในการตัดสินใจจากข้อมูลที่มีหมวดหมู่ข้อมูลแนบอยู่ด้วย ต้นไม้ตัดสินใจจะประกอบไปด้วยโหนดต่าง ๆ (ที่ไม่ใช่โหนดใบ—nonleaf node) ที่ซึ่งถูกใช้ในการแสดงถึงเงื่อนไขหรือแอทริบิวหนึ่ง ๆ ของข้อมูล โดยที่แต่ละกิ่งก้านของโหนดหนึ่ง ๆ จะหมายถึงค่าที่เป็นไปได้จากการทดสอบกับแอทริบิวนั้น ๆ และจะประกอบไปด้วยโหนดใบ (leaf node) ที่ซึ่งจะมีหมวดหมู่ข้อมูลจัดเก็บอยู่ โดยตัวอย่างต้นไม้ตัดสินใจจะถูกแสดงในรูปที่ 2.14 ที่จะแสดงการทำนายคุณลักษณะของลูกค้าที่จะทำการซื้อคอมพิวเตอร์จากร้านขายอุปกรณ์ไฟฟ้า โดยโหนดต่าง ๆ ที่ไม่ใช่โหนดใบจะถูกแทนด้วยสี่เหลี่ยมและโหนดใบจะถูกแทนด้วยวงรีตามลำดับ จากรูปเราจะเห็นว่าโหนดใบจะเป็นโหนดที่ บ่งบอกถึงข้อมูลหมวดหมู่ของคำตอบที่เราต้องการ อาทิ เช่น “yes” หมายถึง ลูกค้าจะซื้อคอมพิวเตอร์ และ“no” หมายถึงลูกค้าจะไม่ซื้อคอมพิวเตอร์โดยต้นไม้ที่ถูกสร้างขึ้นอาจเป็นต้นไม้ที่

มีลักษณะเป็นไบนารี่หรืออาจจะไม่เป็นไบนารี่ก็ได้ หลังจากทำการสร้างต้นไม้ตัดสินใจแล้ว เราจะสามารถใช้ต้นไม้ตัดสินใจในการจำแนกข้อมูลได้โดยจะทำการจำแนกหมวดหมู่ของข้อมูลเรคคอร์ดหนึ่ง ๆ (ที่ประกอบไปด้วยแอทริบิวต่าง ๆ แต่เราจะไม่ทราบหมวดหมู่ข้อมูลในเรคคอร์ดนั้น ๆ) ด้วยการเปรียบเทียบแอทริบิวที่อยู่ในโหนดรากกับค่าของแอทริบิวในเรคคอร์ดที่พิจารณา โดยจะทำการเปรียบเทียบจากโหนดรากไปจนถึงโหนดใบ เมื่อเราทราบถึงโหนดใบจะทำให้เราทราบถึงหมวดหมู่ข้อมูลของเรคคอร์ดที่ทำการพิจารณา



ภาพที่ 2. 15 แสดงตัวอย่างต้นไม้ตัดสินใจสำหรับการจำแนกคุณลักษณะของลูกค้าที่ทำการซื้อคอมพิวเตอร์

(ที่มา : https://bit.ly/38r2EaK)

การจำแนกหมวดหมู่ข้อมูลด้วยต้นไม้ตัดสินใจได้รับความนิยมเป็นอย่างมาก และถูกประยุกต์ใช้ในหลาย ๆ งาน อาทิ เช่น การผลิตและการใช้ยา (medicine) การผลิตสินค้าต่าง ๆ (manufacturing and production) การวิเคราะห์ทางการเงิน (Financial analysis) ดาราศาสตร์ (astronomy) อณูชีววิทยา (molecular biology) เป็นต้น สาเหตุที่ต้นไม้ตัดสินใจได้รับความนิยมอันเนื่องมากจากเหตุผลหลายประการ ด้วยกัน เช่น 1) ไม่ต้องการองค์ความรู้ใด ๆหรือการกำหนดค่าพารามิเตอร์ใด ๆ เพื่อที่จะทำการสร้างต้นไม้ตัดสินใจ 2) สามารถจัดการกับข้อมูลที่มีหลายๆมิติหรือข้อมูลที่มีหลายๆแอทริบิวได้ 3) ผลลัพธ์ที่ได้จะเป็นต้นไม้ตัดสินใจที่อยู่ในรูปแบบที่เข้าใจง่าย 4) ขั้นตอนการสร้างต้นไม้ตัดสินใจค่อนข้างง่ายและสามารถทำงานได้อย่างรวดเร็ว และ 5) มักจะให้ผลการจำแนกข้อมูลที่มีความถูกต้องค่อนข้างสูง ที่ซึ่งอาจขึ้นอยู่กับคุณลักษณะ ของข้อมูลที่เราทำการพิจารณาด้วยเช่นกัน

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

ในโครงงานนี้ใช้เครื่องมือสำหรับทำกระบวนการเหมืองข้อมูลที่มีชื่อว่า Weka โดย Weka เป็นโปรแกรมที่มีเทคนิคต่าง ๆ มากมายพัฒนาอยู่ภายในเช่น อัลกอริทึมการจัดหมวดหมู่:K-NN , LR , C4.5 , Support Vector Machine , การคัดเลือกคุณสมบัติ: Information Gain ,Gain Ratio เป็นต้นซึ่ง Weka เป็นที่นิยมกันอย่างแพร่หลายในกลุ่มผู้จัดทำวิจัยที่วิจัยเกี่ยวกับเหมืองข้อมูล Weka ไม่ได้มีเพียงตัวโปรแกรมที่สามารถใช้งานได้ด้วยตัวเอง (Standalone) เพียงอย่างเดียวแต่ยังมี Library สำหรับใช้คู่กับการเขียน โปรแกรมในภาษาจาวา (Java Programming) อีกด้วยซึ่ง สะดวกมากสำหรับ ผู้วิจัยที่มีความจำเป็นจะต้องเขียนโปรแกรมเกี่ยวกับการทำเหมืองข้อมูล เช่น การคัดเลือกคุณสมบัติ, การหากฎความสัมพันธ์ และผู้วิเคราะห์ยังได้ใช้การVisualization ผ่านโปรแกรมTableau Public ในการนำเสนอข้อมูลผ่านเว็บไซต์อีกด้วย ทาโบล์ว (Tableau) Business Intelligence (BI) Software คือซอฟต์แวร์เพื่อการวิเคราะห์ข้อมูลอันหลากหลายเป็น Software ในระดับต้น ๆ ของโลกที่ทำเรื่อง Data Visualization Tableau ทำหน้าที่แปลงข้อมูล (Data) ให้เป็นภาพ (Visualization) สามารถนำข้อมูลออกมาเป็น Chart แบบต่าง ๆ ได้อย่างง่ายดายรวมถึงการนำหลาย Chart มาสร้างเป็น Dashboard และ Story Teller ก็ได้เช่นกัน รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และ สามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL, Oracle, MySQL เป็นต้น

2.4 วรรณกรรมที่เกี่ยวข้อง

วีณาวดี ม่วงอ้น และคณะ (2558) จากการวิจัยเรื่องระบบผู้เชี่ยวชำญสำหรับระบุพันธุ์เห็ดในประเทศไทย ผลการศึกษาพบว่า งานวิจัยนี้ได้นำเสนอระบบผู้เชี่ยวชาญ ที่มีความสามารถในการวิเคราะห์จำแนกพันธุ์เห็ด ในประเทศไทย ได้เหมาะสมตรงกับปัญหาของผู้ใช้ ระบบได้ทำการแทนค่าความรู้ในรูปแบบของกฎ ที่ถูกสร้างจากต้นไม้ตัดสินใจอัลกอริทึม ID3 และเก็บไว้ในฐานความรู้ ซึ่งเชื่อมโยงสัมพันธ์กับกฎต่าง ๆ ที่เกี่ยวข้องกับรูปร่างลักษณะพันธุ์เห็ด โดยระบบใช้วิธีการกลไกการอนุมานแบบเดินหน้าในการหาเหตุผลเพื่อค้นหาคำตอบเป้าหมายของผู้ใช้ และใช้ค่าความเชื่อมั่นในการคำนวณค่าความแตกต่างของความเชื่อมั่นในแต่ละคำตอบของผู้ใช้แล้วนำค่าความเชื่อมั่นนั้นแทนค่าออกมาเป็นร้อยละของความเป็นไปได้ที่จะเป็นสายพันธุ์เห็ดนั้นเพื่อใช้สำหรับเรียงลำดับผลลัพธ์ให้กับผู้ใช้ ระบบผู้เชี่ยวชาญระบุพันธุ์เห็ดนี้พัฒนาให้สามารถทำงานได้บนเว็บบราวเซอร์ โดยพัฒนาหน้าเว็บส่วนติดต่อผู้ใช้ด้วย ภาษา PHP HTML และ JAVASCRIPT ใช้

MySQL เป็นเครื่องมือช่วยในการจัดการฐานข้อมูลและใช้ภาษา Prolog ในการจัดเก็บฐานความรู้ของสายพันธุ์เห็ด ผลการทดสอบแสดงว่าระบบนี้สามารถให้คำตอบที่ถูกต้องอยู่ในเกณฑ์ดี

วทัญญุตา นีลาภาตระกูล และ ชุติมา เบี้ยวไข่มุข(2562) จากการวิจัยเรื่อง การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพ ผลการศึกษาพบว่า งานวิจัยนี้การศึกษาเกี่ยวกับเทคนิคการทำ เหมืองข้อมูล เพื่อลดปัญหาอัตราการลาออกของพนักงานและลดค่าใช้จ่ายที่สูงขึ้นจากการสรรหาพนักงานใหม่ในบริษัทประกันภัยแห่งหนึ่ง ชุดข้อมูลที่ใช้วิเคราะห์เป็นข้อมูลของพนักงานลาออกและพนักงานที่ยังทำงานอยู่ ตั้งแต่ปีพ.ศ. 2556-2560 จำนวน 1,000 รายการ11 แอตทริบิวต์ เทคนิคกฎความสัมพันธ์ถูกนำ มาใช้หารูปแบบของความสัมพันธ์การลาออกของพนักงาน และสร้างตัวแบบพยากรณ์การลาออกของพนักงาน โดยใช้เทคนิคการวิเคราะห์5เทคนิคคือต้นไม้ตัดสินใจเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคโครงข่ายประสาทเทียม เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้สุดและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์เหล่านั้น การทดสอบประสิทธิภาพของตัวแบบใช้วิธีไขว้ข้อมูลโดยแบ่งข้อมูลออกเป็น 5 ชุดเท่ากัน ผลของกฎความสัมพันธ์พบปัจจัยที่มีความสัมพันธ์กับการตัดสินใจลาออก คือ เงินเดือน อายุงาน ความก้าวหน้าในอาชีพ ผลการประเมินการปฏิบัติงาน และความสัมพันธ์กับหัวหน้างาน ผลการทดสอบตัวแบบพยากรณ์ที่เป็นต้นไม้ตัดสินใจ ให้ค่าความถูกต้องในการพยากรณ์สูงที่สุดที่ร้อยละ91.03ในขณะที่ตัวแบบพยากรณ์ที่สร้างจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องเท่ากับร้อยละ 90.93 เทคนิคโครงข่ายประสาทเทียม ให้ค่าความถูกต้องเท่ากับ ร้อยละ 90.75 เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้สุด ให้ค่าความถูกต้องเท่ากับ ร้อยละ 89.60 และ 82.10 ตามลำดับ ผลที่ได้จากการวิจัยสามารถนำผลมาประยุกต์ใช้ในการสร้างแผนกลยุทธ์สำ หรับออกแบบสวัสดิการที่เหมาะสมเพื่อรักษาพนักงานให้คงอยู่กับบริษัทและเป็นแนวทางในการสร้างระบบสารสนเทศในการสนับสนุนการตัดสินใจสำหรับงานด้านบริหารทรัพยากรบุคคลต่อไป

พีรพล คำพันธ์และคณะ(2562) จากการวิจัยเรื่องการประยุกต์ใช้เบย์เซียนและการประมวลผลภาพสำหรับการคัดแยกมะม่วง ผลการศึกษาพบว่า งานวิจัยนี้นำเสนอการประยุกต์เบย์เซียนและการประมวลผลภาพดิจิตอลเพื่อแยกประเภทมะม่วงตามลักษณะความสุก/ดิบ ด้วยวิธีการคัดแยกเชิงคุณลักษณะ (Feature-based) โดยที่คุณลักษณะที่ใช้สำหรับการแยกประเภทได้มาจากคุณลักษณะของภาพสี RGB และความเข้มข้นของค่าสีในเพลนสีแดงร่วมกับการวิเคราะห์ค่าทางสถิติ ในขั้นตอนการแยกประเภทผลมะม่วงได้ทำการทดลองเปรียบเทียบประสิทธิภาพ

ระหว่างการแยกประเภทด้วยเบย์เซียนและการแยกประเภทด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (SVM) จากผลการทดลองพบว่าการแยกประเภทด้วยเบย์เซียนให้ผลลัพธ์ความถูกต้องแม่นยำร้อยละ 88 มีความถูกต้องแม่นยำมากกว่าการแยกประเภทด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนซึ่งให้ผลลัพธ์ความถูกต้องแม่นยำร้อยละ 83

พิชัย ระเวงวังและพุธษดี ศิริแสงตระกูล(2560) จากการวิจัยเรื่องโมเดลเพื่อการพยากรณ์สถานภาพทางการศึกษาของนักศึกษา ผลการศึกพบว่า งานวิจัยนี้มุ่งเน้นศึกษาปัจจัยที่มีผลต่อสถานภาพการศึกษาของนักศึกษา เพื่อทำการสร้างโมเดลเพื่อพยากรณ์สถานภาพทางการศึกษา ของนักศึกษาในสถาบันอุดมศึกษาเอกชนที่ตั้งในภาคตะวันออกเฉียงเหนือ โดยการประยุกต์ใช้วิธีการเหมืองข้อมูล ในการศึกษาผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลปัจจัยที่คาดวาจะมีผลต่อสถานภาพทางการศึกษาจากงานวิจัยที่เกี่ยวข้องและได้รวมรวมข้อมูลของนักศึกษาระดับปริญญาตรีที่เข้าศึกษาระหวางปี การศึกษา 2551–2553 จำนวนทั้งสิ้น 2,272 คน และวิเคราะห์ความสัมพันธ์ของปัจจัยที่มีผลต่อสถานภาพของนักศึกษาด้วยวิธีการวิเคราะห์สหสัมพันธ์ ผลจากศึกษาพบวาปัจจัยที่มีผลต่อสถานภาพของนักศึกษามี9 ปัจจัย ได้แก่การกู้ยืมในระหว่างเรียน รายได้ ผู้ปกครอง สถานภาพสมรสของผู้ปกครอง เพศ ประเภทโรงเรียนเดิมที่จบมา วุฒิการศึกษาเดิม กลุ่มโรงเรียนเดิมที่จบ อาชีพของมารดา และที่อยู่นักศึกษาระหว่างเรียน ปัจจัยดังกล่าวจะถูกนำมาใช้ในการสร้างโมเดลเพื่อทำนายสถานภาพของนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ เบย์เซี่ยนบีลิฟเน็ตเวิร์ค และโลจิสติกรีเกรสชัน ผลการศึกษาพบว่า ค่าความถูกต้องมีค่า 82.85%, 78.98% และ 78.50% ด้วยเทคนิคต้นไม้ตัดสินใจ จิสติกรีเกรสชัน และ การเรียนรู้แบบเบย์เซี่ยน เน็ตเวิร์ค ตามลำดับ ซึ่งพบว่าโมเดลที่สร้างด้วยวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องดีที่สุด

พัฒนพงษ์ ดลรัตน์ และ จารี ทองคำ(2560) จากการวิจัยเรื่องการเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ ผลการศึกษาปัจจุบันประเทศไทยนั้นอยู่สภาวะขาดแคลนแรงงานฝีมือ กระทรวงศึกษาธิการจึงได้มีนโยบายในปีพุทธศักราช 2554 การเปลี่ยนแปลงสัดส่วนของผู้เรียนอาชีวศึกษาต่อผู้เรียนสายสามัญจากเดิม 40:60 เป็น 60:40 ภายในปีพุทธศักราช 2561 ซึ่งเป็นปีสิ้นสุดของการปฏิรูปการศึกษาทศวรรษที่ 2 ซึ่งสถานศึกษาอาชีวศึกษาเอกชนก็เป็นหน่วยงานหนึ่งที่ดำเนินการสนองนโยบายของรัฐบาลเพื่อแก้ไขปัญหาสภาวะขาดแคลนแรงงานฝีมือ เพิ่มจำนวนผู้สำเร็จการศึกษา พัฒนาการศึกษาให้เกิดประสิทธิภาพสูงสุด ดังนั้นงานวิจัยนี้จึงมีจุดประสงค์เพื่อการเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียน

ระดับประกาศนียบัตรวิชาชีพ งานวิจัยนี้ได้ใช้ 6 เทคนิคที่มีประสิทธิภาพในการสร้างแบบจำลอง คือ C4.5, Random Forest, Random Tree, Reduced Error Pruning (REP Tree), k–Nearest Neighbors (k–NN) และSupport Vector Machine(SVM) และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) จากการศึกษาพบว่า แบบจำลอง C4.5 มีประสิทธิภาพในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพมากที่สุดถึง 95.36%

พฤฒิพงศ์ เพ็งศิริ (2559)จากการวิจัยเรื่อง การพยากรณ์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ผลการศึกษาพบว่า การพยากรณ์โดยการประยุกต์ใช้กระบวนการตัดสินใจด้วยเทคนิคต้นไม้ตัดสินใจเป็นเทคนิคอย่างหนึ่งในการทำเหมืองข้อมูล ซึ่งอาศัยความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาเป็นการบ่งชี้ถึงระดับผลการเรียนของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ซึ่งผลลัพธ์ที่ได้นั้น จะส่งผลทำให้เอือ้ ประโยชน์กับแนวทางในการปรับปรุงบริหารจัดการและส่งเสริมการวางแผนของหลักสูตรการศึกษาที่เกี่ยวข้อง โดยนำปัจจัยข้อมูลของนักศึกษา ได้แก่ข้อมูลทั่วไป ข้อมูลส่วนตัว ข้อมูลการศึกษา ข้อมูลระดับผลการศึกษา ของนักศึกษากลุ่มที่ศึกษา มาหาความสัมพันธ์กับผลการศึกษาที่ได้ โดยวิเคราะห์จากระดับผลการเรียน (เกรดเฉลี่ย) ของการเรียนในชั้นปีที่ 1 ผลการทดลองพบว่าปัจจัยข้อมูลของนักศึกษาที่เหมาะสมในการเรียน มีทั้งหมด 7 ตัวแปร จากข้อมูลนำเข้าทั้งหมด 12 ตัวแปร ทั้งนี้ 7 ตัวแปรมาจากต้นไม้ตัดสินใจที่ได้มาสามารถสรุปเลือกเฉพาะกิ่งที่มีผลมากที่สุด โดยวัดค่าความแม่นยำ (Accuracy) ได้ค่าสูงถึง 84.78% ซึ่งถือว่าอยู่ในเกณฑ์สูง ดังนั้นตัวแปรที่น่าจะเป็นปัจจัยที่ส่งผลกับ ระดับผลการเรียนที่จบการศึกษาสูงสุดคือ ความสม่ำเสมอการเข้าเรียนในการเรียน

หนูเดือน เมืองแสนและคณะ(2556) จากการวิจัยเรื่องความหลากหลายของไลเคน เห็ดและราขนาดใหญ่ในพื้นที่ปกปักพันธุกรรมพืช อพ.สธ. เขื่อนน้ำพุง การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย ผลการศึกษาพบว่า เห็ดกินได้ (edible mushroom) เป็นเห็ดที่รับประทานได้ ลักษณะ ส่วนใหญ่เจริญในทุ่งหญ้า ก้านสั้น อ้วนปูอมและไม่โป่งพองออก ผิวเรียบไม่ขรุขระ ไม่มีสะเก็ด สีผิวของหมวกส่วนใหญ่เป็นสีขาวถึงสี น้ำตาล ผิวของหมวกเห็ดเรียบจนถึงเป็นเส้นใยและ เหมือนถูกกดจนเป็นแผ่นบาง ๆ ดึงออกยาก ครีบแยกออก จากกัน ในระยะแรกเป็นสีชมพู แล้วเปลี่ยนเป็นสีน้ำตาล สปอร์สีน้ำตาลอมม่วงแก่รูปกระสวยกว้าง เช่น เห็ด ฟาง เห็ดหูหนู เห็ดหอม เห็ดเผาะ เห็ดขมิ้น ฯลฯ เห็ดพิษ (toadstool หรือ poisonous mushroom) เห็ดที่นำมาบริโภคแล้วเป็นอันตรายต่อ

ผู้บริโภคอาจร้ายแรงถึงตายได้ ลักษณะส่วนใหญ่เจริญงอกงามในป่า ก้านสูง ลำต้นโป่งพองออก โดยเฉพาะที่ ฐาน กับที่วงแหวนเห็นชัดเจน สีผิวของหมวกมีได้หลายสี เช่น สีมะนาว ถึงสีส้ม สีขาว ถึงสีเหลือง ผิวของหมวก เห็ดส่วนมากมีเยื่อหุ้มดอกเห็ดเหลืออยู่ในลักษณะที่ดึงออกได้ หรือเป็นสะเก็ดติดอยู่ ครีบแยกออกจากกัน ชัดเจน มักมีสีขาว บางชนิดสีแดงหรือสีเขียวอมเหลือง สปอร์ใหญ่มีสีขาวหรือสีอ่อน มีลักษณะใส ๆ รูปไข่กว้าง ซึ่งโดยทั่วไปมักมีลักษณะแปลก ๆ เช่น เห็ดหัวกรวด เห็ดขี้ควาย ฯลฯ

พรรณพรและคณะ (2554) จากการวิจัยเรื่อง การสำรวจความหลากหลายชนิดและคุณลักษณะพิเศษเฉพาะตัวของเห็ดป่ารับประทานได้เพื่อการจำแนกแยกเห็ดพิษ และภูมิปัญญาการใช้ประโยชน์ในจังหวัดน่าน ผลการศึกษาพบว่า เมื่อเปรียบเทียบลักษณะสัณฐานของเห็ดระโงกหิน (เห็ดพิษ) กับเห็ดไข่ขาว (เห็ดรับประทานได้) พบว่าในระยะอ่อน เห็ดระโงกหินจะมีลักษณะเนื้อแน่น เหนียวคล้ายยางพารา ก้านและโคนตัน เมื่อลอกเปลือกหุ้มออกจะเห็ดหมวกสีขาวสว่าง ขอบหมวกเรียบ ในระยะโตเต็มวัย(ดอกบาน) เห็ดระโงกหินจะมีลักษณะ เนื้อแข็ง หมวกสีขาวสว่าง ขอบหมวกมีริ้วคล้ายชายกระโปรง ก้าน และโคนตัน เปลือกหุ้มโคนแนบชิดติดก้าน

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย 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 ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา