

บทที่ 2

แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

โครงการเรื่อง ระบบวิเคราะห์ข้อมูลและพยากรณ์ผลผลิตพืชผักอายุสั้นตระกูลสลัด เพื่อส่งเสริมเกษตรกรใน โครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่ โดยใช้แบบจำลอง การถดถอยด้วยเทคนิคแบบการจำลองการถดถอยพหุคูณ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

2.1.2 แนวคิดเกี่ยวกับแบบจำลองการถดถอย (Regression Model)

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

2.1.4 แนวคิดเกี่ยวกับการพยากรณ์ (Forecasting data)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวข้องกับระบบฐานข้อมูลขนาดใหญ่ (Big Data)

2.2.2 ทฤษฎีเกี่ยวข้องกับเทคนิควิธีเหมืองข้อมูล (Data mining)

2.2.3 ทฤษฎีเกี่ยวข้องกับภาษา SQL (Structured Query Language)

2.2.4 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

2.3.2 กระบวนการทำงานด้วย Streamlit

2.3.3 โปรแกรมจัดการข้อมูลด้วยโพสต์เกรสควิเอล (PostgreSQL)

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล

ในการดำเนินงานเรื่องระบบวิเคราะห์ข้อมูลและพยากรณ์ผลผลิตพืชผักอายุสั้น ครอบคลุมสัปดาห์เพื่อส่งเสริมเกษตรกรใน โครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่ โดยใช้แบบจำลอง การถดถอยด้วยเทคนิคแบบการจำลองการถดถอย พหุคูณ คณะผู้จัดทำได้ศึกษาหลักการและทฤษฎีต่าง ๆ องค์ประกอบที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

ความหมายของการวิเคราะห์ข้อมูล

Data Analytics คือ การวิเคราะห์ข้อมูลที่มีอยู่ในปัจจุบัน อดีต และทำนายอนาคต เพื่อพัฒนาการตลาด ให้ตรงกับความต้องการของลูกค้ามากยิ่งขึ้น Data Analytics เป็นเครื่องมือสำหรับธุรกิจ (Business Intelligence) Data Analytics นี้ไม่จำเป็นต้องเป็นธุรกิจขนาดใหญ่ ธุรกิจขนาดเล็กก็สามารถทำได้เหมือนกัน สามารถแบ่งเป็น 4 แบบคือ

1) Descriptive Analytics (การวิเคราะห์เชิงพรรณนา) เป็นรูปแบบการใช้ข้อมูลแบบพื้นฐานที่สุด โดยเน้นการอธิบายว่ากำลังเกิดขึ้น หรืออาจจะเกิดอะไรขึ้น สามารถอธิบายถึงสาเหตุการเกิดต่าง ๆ ซึ่ง Descriptive Analytics ตัวอย่างคือรายงานธุรกิจ รายงานด้านการทำ Campaign หรือโฆษณา หรือรายงานผลดำเนินงานที่ผ่านมา เป็นข้อมูลพื้นฐานที่แสดงผลในการทำกิจกรรมต่าง ๆ

2) Diagnostic Analytics (การวิเคราะห์เชิงวินิจฉัย) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้น ปัจจัยต่างๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่างๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น ตัวอย่างเช่นความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่ช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

3) Predictive Analytics (การวิเคราะห์เชิงคาดการณ์) เป็นรูปแบบการใช้ข้อมูลที่มีความซับซ้อนขึ้นมา โดยจะเป็นการ “พยากรณ์” หรือ “ทำนาย” สิ่งที่กำลังเกิดขึ้น โดยใช้ข้อมูลในอดีต ร่วมกับโมเดลทางคณิตศาสตร์ต่าง ๆ หรือร่วมกับการทำ Data Mining นอกจากนี้ Predictive Analytics ยังทำให้เราสามารถวิเคราะห์หาโอกาสและความเสี่ยงต่าง ๆ ที่ จะเกิดขึ้นในอนาคตได้ด้วย

4) Prescriptive Analytics (การวิเคราะห์เชิงข้อกำหนด) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด เป็นทั้งการพยากรณ์สิ่งต่างๆ ที่จะเกิดขึ้น ข้อดี ข้อเสีย สาเหตุ และระยะเวลาของสิ่งที่เกิดขึ้น และการให้คำแนะนำ ทางเลือกต่างๆที่มีอยู่ และผลของแต่ละทางเลือก

2.1.2 แนวคิดเกี่ยวกับการถดถอย

ความหมายของการถดถอย (The Meaning of Regression)

การถดถอย (Regression) หมายถึง การศึกษาเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งที่เรียกว่าตัวแปรถูกอธิบาย (Explained Variable) หรือตัวแปรตาม

(Dependent Variable) และอีกตัวแปรหนึ่งหรือมากกว่าหนึ่งตัวแปรที่เรียกว่าตัวแปรอิสระ (Independent Variable) หรือตัวแปรอธิบาย (Explanatory Variable) เช่น นักเศรษฐศาสตร์สนใจที่จะศึกษาความสัมพันธ์ระหว่างปริมาณความต้องการเสนอซื้อสินค้ากับราคาของสินค้านั้น และระดับรายได้ของผู้บริโภค หรือสนใจศึกษาว่าปริมาณการขายสินค้าจะมีความสัมพันธ์กับรายจ่ายในการโฆษณาสินค้าอย่างไร เป็นต้น

เป้าหมายของการถดถอย คือ

- 1) เพื่อประมาณค่าเฉลี่ยของตัวแปรตามเมื่อกำหนดค่าของตัวแปรอิสระมาให้
- 2) เพื่อทดสอบสมมติฐานเกี่ยวกับธรรมชาติของตัวแปร โดยมีสมมติฐานภายใต้ทฤษฎีทางเศรษฐศาสตร์
- 3) เพื่อพยากรณ์ หรือทำนายค่าเฉลี่ยของตัวแปรตาม เมื่อกำหนดค่าของตัวแปรอิสระมาให้

2.1.2.1 การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression)

การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) กับตัวแปรตาม (Dependent Variable) จะเป็นการศึกษาความสัมพันธ์เชิงเส้นตรง (Linearity) ถ้าศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหนึ่งตัวกับตัวแปรตามหนึ่งตัว เรียกว่าการวิเคราะห์ถดถอยเชิงเส้นเชิงเดี่ยวหรือการวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis) ถ้าตัวแปรอิสระมีมากกว่าหนึ่งตัวกับตัวแปรตามหนึ่งตัว เรียกว่า การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression)

2.1.2.2 วัตถุประสงค์ของการวิเคราะห์การถดถอย

- 1) เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม เช่น ศึกษาความสัมพันธ์ของอายุกับระดับคลอเลสเตอรอล

2) เพื่อศึกษาปัจจัย (ตัวแปรอิสระ) ที่ร่วมกันทำนายหรือพยากรณ์ตัวแปรตาม เช่น การศึกษาปัจจัยที่มีผลต่อระดับน้ำตาลในเลือดของผู้ป่วยเบาหวาน

การวิเคราะห์การถดถอยเพื่อหาความสัมพันธ์หรือสร้างสมการทำนายหรือพยากรณ์ตัวแปรตาม (Y) หนึ่งตัว จากกลุ่มตัวแปรอิสระ (X) หลายตัวนั้น ตัวแปรอิสระที่นำมาวิเคราะห์จะต้องมีหลักฐานตามทฤษฎีหรือรายงานการวิจัยที่เกี่ยวข้องว่าเป็นตัวแปรต้นเหตุที่ส่งผลต่อตัวแปรตาม

2.1.2.3 การวิเคราะห์การถดถอยเชิงเส้นเชิงเดียว (Simple Linear Regression Analysis)

การวิเคราะห์การถดถอยเชิงเส้นเชิงเดียวเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ 1 ตัว กับ ตัวแปรตาม 1 ตัว คล้ายกับการวิเคราะห์สหสัมพันธ์ แตกต่างกันที่การวิเคราะห์สหสัมพันธ์ไม่ได้ระบุว่าตัวแปรใดเป็นตัวแปรต้น ตัวแปรใดเป็นตัวแปรตาม ส่วนการวิเคราะห์การถดถอยเชิงเส้นเชิงเดียวเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรที่เป็นต้นเหตุ (ตัวแปรอิสระ) กับตัวแปรที่เป็นผล (ตัวแปรตาม) ซึ่งนอกจากจะทราบความสัมพันธ์ระหว่างตัวแปรทั้งสองแล้ว ยังสามารถนำค่าของตัวแปรต้นเหตุไปทำนายหรือพยากรณ์ตัวแปรที่เป็นผลได้ พร้อมบอกขนาดของความสามารถในการทำนายหรือความสามารถในการอธิบายตัวแปรผล ว่ามีมากน้อยเพียงใด

2.1.2.4 การวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression Analysis)

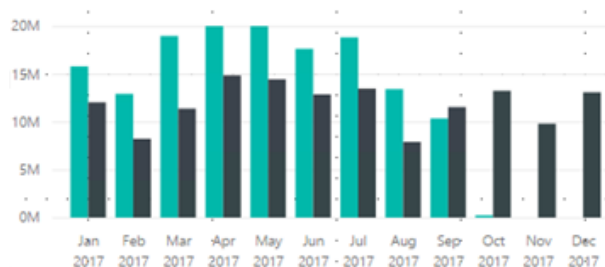
การวิเคราะห์การถดถอยเชิงพหุคูณเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตาม 1 ตัว เพื่อศึกษาว่ามีตัวแปรอิสระตัวใดบ้างที่ร่วมกันทำนายหรือพยากรณ์หรืออธิบายการผันแปรของตัวแปรตามได้

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

ความหมายของการแสดงผลข้อมูล

Data Visualization เป็นการนำ Data เชิงลึกที่ได้มาจากช่องทางต่าง ๆ มาวิเคราะห์และแสดงผลในรูปแบบของแผนภูมิ กราฟรูปแบบที่หลากหลาย วิดีโอที่แสดงให้เห็นถึงการเปลี่ยนแปลงของข้อมูลเชิงปริมาณ ซึ่งจะทำให้ผู้อ่านเข้าใจได้ง่ายขึ้น ดูน่าสนใจมากกว่าการอ่านข้อมูลแบบตารางทั่วไป

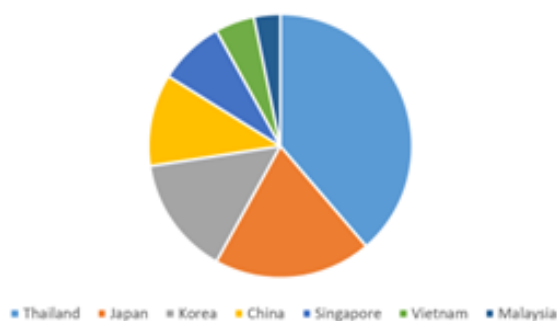
1) แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้งที่นิยมแสดงออกมาในรูปแบบสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับการเปรียบเทียบจำนวนของข้อมูลในแต่ละชุดเช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริงเปรียบเทียบกับเป้าหมายยอดขายที่ตั้งไว้ เป็นต้นซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท แผนภูมิแท่งแบบจัดกลุ่ม แผนภูมิแท่งแบบจัดกลุ่ม



ภาพที่ 2.1 แผนภูมิแท่ง (Bar Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

2) แผนภูมिवงกลม (Pie Charts) แผนภูมिवงกลมเหมาะกับการนำเสนอข้อมูลที่มี ส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกัน อาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยาก เพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูล แสดงส่วนผสมต่าง ๆ เป็นต้น



ภาพที่ 2.2 แผนภูมिवงกลม (Pie Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

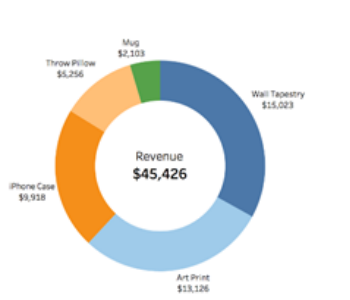
3) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั้นเองแผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้เช่น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



ภาพที่ 2.3 แผนภูมิเส้น (Line Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

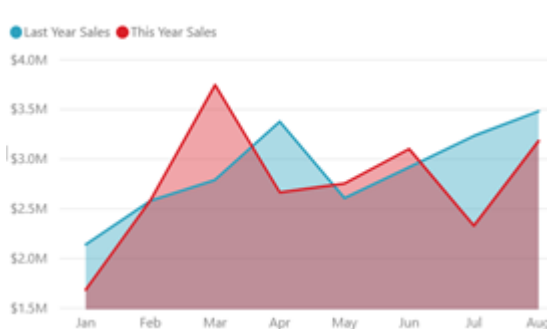
4) แผนภูมิโดนัท (Donut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลายๆ ชั้น



ภาพที่ 2.4 แผนภูมิโดนัท (Donut Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

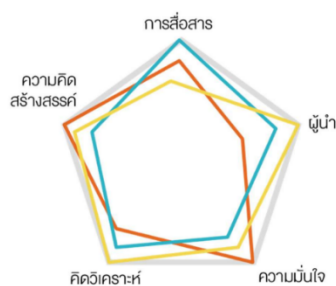
5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้ เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะ สำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่าง ระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่ละเดือน ตามหมวดหมู่ต่าง ๆ ไล่ไป เครื่องสำอางค์ เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

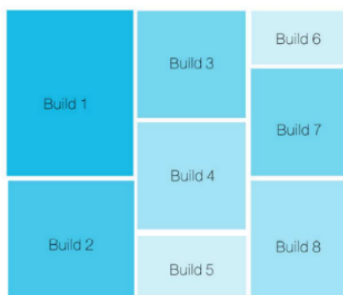
6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผล แบบ วงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึง ความ ต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หา จุดอ่อน จุดแข็งของข้อมูลเช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดู จุดอ่อน จุดแข็งของแต่ละคน เป็นต้น



ภาพที่ 2.6 แผนภูมิเรดาร์ (Radar Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

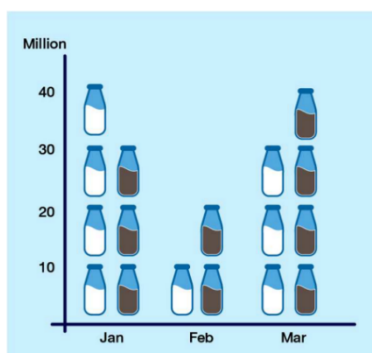
7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดงผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้



ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

ที่มา : blog.1moby.com(ม.ป.ป.)

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



ภาพที่ 2.8 แผนภูมิรูปภาพ (Picture Graph)

ที่มา : blog.1moby.com(ม.ป.ป.)

2.1.4 แนวคิดเกี่ยวกับการพยากรณ์

ความหมายของการพยากรณ์ (Forecasting data) การพยากรณ์ หมายถึงการ คาดการณ์ (Predict) เกี่ยวกับลักษณะหรือแนวโน้มของสิ่งที่จะเกิดขึ้นในอนาคต เพื่อใช้เป็นสารสนเทศ (Information) ประกอบการตัดสินใจซึ่งการพยากรณ์จะต้องดำเนินการเป็นส่วน แรกสุดก่อนการวางแผน หรือการเตรียมการที่จะเริ่มกระทำการใด ๆ เพื่อความถูกต้องและ แม่นยำในการตัดสินใจ ดังนั้นในการดำเนินธุรกิจภายใต้ความไม่แน่นอนและการเปลี่ยนแปลงที่ รวดเร็วเช่นปัจจุบันจึงจำเป็นที่จะต้องทราบถึงความเป็นไปในอนาคต โดยอาศัยเทคนิคการ พยากรณ์ต่าง ๆ เพื่อเพิ่มประสิทธิภาพในการดำเนินงานทางธุรกิจสามารถตอบสนองต่อความ ต้องการของลูกค้า เพิ่มผลตอบแทนทางธุรกิจ ลดต้นทุนและความสูญเสียต่าง ๆ ที่อาจเกิดขึ้น เป็นต้น

การพยากรณ์เป็นทั้งศาสตร์และศิลป์ในการทำนายเหตุการณ์ในอนาคต ซึ่งอาจ นำหลาย ๆ วิธีมาใช้แล้วแต่สถานการณ์ เช่น นำข้อมูลในอดีตมาพยากรณ์อนาคตโดยอาศัย หลักการทางคณิตศาสตร์เข้าช่วย หรือใช้ดุลยพินิจของผู้พยากรณ์เพียงอย่างเดียว หรืออาจใช้ หลาย ๆ วิธีร่วมกันเพื่อให้การพยากรณ์มีความแม่นยำมากขึ้น

2.1.4.1 ช่วงเวลาของการพยากรณ์

สามารถแบ่งตามระยะของการพยากรณ์ได้ 3 ประเภทคือ

1) การพยากรณ์ในระยะสั้น เป็นการพยากรณ์เหตุการณ์ที่ไม่เกิน 1 ปี โดยทั่วไปมักจะอยู่ในช่วงไม่เกิน 3 เดือน เช่น การพยากรณ์การวางแผนจัดซื้อ การจัดตาราง การทำงาน การมอบหมายงาน การพยากรณ์ยอดขาย และการพยากรณ์ระดับการผลิต

2) การพยากรณ์ระยะกลาง เป็นการพยากรณ์เหตุการณ์ที่อยู่ในช่วง 3 เดือน ถึง 3 ปี จะใช้มากในการพยากรณ์การวางแผนการขาย การวางแผนการผลิตการ วางแผนด้านงบประมาณเงินสด และการวิเคราะห์การวางแผนการดำเนินงานต่าง ๆ

3) การพยากรณ์ระยะยาว เป็นการพยากรณ์เหตุการณ์ที่มากกว่า 3 ปี ขึ้นไปมักใช้สำหรับการวางแผนผลิตภัณฑ์ใหม่ ค่าใช้จ่ายในการลงทุน การขยายทำเลที่ตั้ง และการวิจัยพัฒนา

2.1.4.2 วิธีการพยากรณ์

วิธีการพยากรณ์แบ่งออกเป็น 2 ประเภท คือ

1) การพยากรณ์เชิงคุณภาพ (Qualitative method) เป็นการพยากรณ์ที่ใช้ อารมณ์ ความรู้สึก ซึ่งอาจจะไม่มีการใช้ข้อมูลในอดีต เนื่องจากไม่ได้มีการเก็บรวบรวมข้อมูลในอดีตไว้หรือมีไม่พอเพียงพอต่อการนำมาสร้างรูปแบบในการพยากรณ์ ดังนั้น ความถูกต้องของการพยากรณ์เชิงคุณภาพจะขึ้นอยู่กับประสบการณ์และความสามารถของผู้พยากรณ์เป็นหลัก

2) การพยากรณ์เชิงปริมาณ (Quantitative methods) จะเป็นเทคนิคที่ต้องใช้ข้อมูลในอดีตมาสร้างรูปแบบการพยากรณ์ในรูปแบบของสมการคณิตศาสตร์ ดังนั้นความถูกต้องของการพยากรณ์นี้จะขึ้นอยู่กับความแม่นยำของข้อมูลที่มีอยู่

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการฐานข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

ข้อมูลขนาดใหญ่ หมายถึงข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยากหรือเป็นไปได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมากเพื่อทำการวิเคราะห์มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วงต้นปี ค.ศ. 2000 เมื่อดัก ลานีย์ นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันใน ขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วย 5Vs 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 และสมาร์ทมิเตอร์ช่วยผลักดันความต้องการในการจัดการกับกระแสข้อมูลเหล่านี้ในรูปแบบเรียลไทม์

Variety (ความหลากหลาย) ข้อมูลมีในทุกรูปแบบ นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ อีเมล วิดีโอ เสียง ข้อมูล หุ่น และธุรกรรมทางการเงิน

Veracity (ความถูกต้อง) คุณภาพหรือความเที่ยงตรงของข้อมูล ข้อมูลที่เข้าสู่สภาพแวดล้อม Big Data จะต้องได้รับการประเมินคุณภาพซึ่งอาจนำไปสู่กิจกรรมการประมวลผลข้อมูลเพื่อแก้ไขข้อมูลที่ไม่ถูกต้องและลบสัญญาณรบกวน

Value (มูลค่า) ประโยชน์ของข้อมูลสำหรับองค์กร ลักษณะมูลค่ามีความสัมพันธ์โดยนัยกับลักษณะความถูกต้องที่ว่า ยิ่งความเที่ยงตรงของข้อมูลสูงเท่าใดมูลค่าก็จะยิ่งมากขึ้นสำหรับธุรกิจ

2.2.1.2 การจัดการข้อมูลขนาดใหญ่

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

การจัดการข้อมูลข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ on premises หรือ แบบ cloud ขึ้นกับความต้องการหรือความสะดวกในการใช้ ซึ่งเราสามารถใส่และประเมินผลได้ เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูลหรือข้อมูลบางอัน ต้องการความยืดหยุ่นสูงและไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยม กันเป็นอย่างมาก

การวิเคราะห์การลงทุนสร้างข้อมูลขนาดใหญ่ หรือ Big data จะมีประโยชน์หรือคุ้มค่า ก็ต่อเมื่อคุณใช้และวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่างและชัดเจนในชุด ข้อมูลที่คุณมีอยู่ การสำรวจข้อมูลยังทำให้เราค้นพบสิ่งใหม่ แสร้งสิ่งที่คุณพบใหม่ ๆ ต่อคนอื่น สร้างรูปแบบจำลองข้อมูล ด้วยการเรียนรู้ของเครื่องจักรและปัญญาประดิษฐ์ AI และนำข้อมูล เหล่านั้นไปใช้งาน

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

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

- 1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและ เก็บข้อมูลที่เกี่ยวข้อง
- 2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย
- 3) การสกัดหรือแยกข้อมูลที่ เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
- 4) การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

2.2.3 ทฤษฎีเกี่ยวข้องกับภาษา SQL

SQL ย่อมาจาก structured query language คือภาษาที่ใช้ในการเขียนโปรแกรมเพื่อจัดการกับฐานข้อมูลโดยเฉพาะ เป็นภาษามาตรฐานบนระบบฐานข้อมูลเชิงสัมพันธ์และเป็นระบบเปิด (open system) หมายถึงสามารถใช้คำสั่ง sql กับฐานข้อมูลชนิดใดก็ได้ และคำสั่งงานเดียวกันเมื่อส่งงานผ่าน ระบบฐานข้อมูลที่แตกต่างกันจะได้ผลลัพธ์เหมือนกัน ทำให้เราสามารถเลือกใช้ฐานข้อมูล ชนิดใดก็ได้โดยไม่ติดขัดกับฐานข้อมูลใดฐานข้อมูลหนึ่ง นอกจากนี้แล้ว SQL ยังเป็นชื่อโปรแกรมฐานข้อมูล ซึ่งโปรแกรม SQL เป็นโปรแกรมฐานข้อมูลที่มีโครงสร้างของภาษาที่เข้าใจง่าย ไม่ซับซ้อน มีประสิทธิภาพการทำงานสูง สามารถทำงานที่ซับซ้อนได้โดยใช้คำสั่งเพียงไม่กี่คำสั่ง โปรแกรม SQL จึงเหมาะที่จะใช้กับระบบฐานข้อมูลเชิงสัมพันธ์ และเป็นภาษาหนึ่ง ซึ่งแบ่งการทำงานได้เป็น 4 ประเภท ดังนี้

- 1) Select query ใช้สำหรับดึงข้อมูลที่ต้องการ
- 2) Update query ใช้สำหรับแก้ไขข้อมูล
- 3) Insert query ใช้สำหรับการเพิ่มข้อมูล
- 4) Delete query ใช้สำหรับลบข้อมูล

2.2.3.1 ประเภทของคำสั่งภาษา SQL แบ่งออกเป็น 3 ประเภท ดังนี้

1) ภาษานิยามข้อมูล (Data Definition Language : DDL) เป็นคำสั่งที่ใช้ในการสร้างฐานข้อมูล กำหนดโครงสร้างข้อมูลว่ามี Attribute ใดชนิดของข้อมูล รวมทั้งการเปลี่ยนแปลงตาราง และการสร้างดัชนี คำสั่ง : CREATE, DROP, ALTER

2) ภาษาจัดการข้อมูล (Data Manipulation Language : DML) เป็นคำสั่งที่ใช้ในการเรียกใช้ เพิ่ม ลบ และแก้ไขเปลี่ยนแปลงข้อมูลในตาราง คำสั่ง : SELECT, INSERT, UPDATE, DELETE

3) ภาษาควบคุมข้อมูล (Data Control Language : DCL) เป็นคำสั่งที่ใช้ในการกำหนดสิทธิการอนุญาต หรือ ยกเลิก การเข้าถึงฐานข้อมูล เพื่อป้องกันความปลอดภัยของฐานข้อมูล คำสั่ง : GRANT, REVOKE

2.2.3.2 รูปแบบการใช้คำสั่ง SQL สามารถใช้ได้เป็น 2 รูปแบบ ดังนี้

1) คำสั่ง SQL ที่ใช้เรียกดูข้อมูลแบบโต้ตอบ (Interactive SQL) เป็นการ
ใช้คำสั่ง SQL สั่งงานบนจอภาพเพื่อเรียกดูข้อมูลจากฐานข้อมูลได้โดยตรงในขณะที่ทำงาน

2) คำสั่ง SQL ที่ใช้เขียนร่วมกับโปรแกรมอื่น ๆ (Embedded SQL) เป็น
การนำคำสั่ง SQL ไปใช้ร่วมกับชุดคำสั่งงานที่เขียนโดยภาษาต่าง ๆ เช่น COBOL PASCALPL/1
ฯลฯ

เกณฑ์การใช้งาน SQL ตามที่ได้กล่าวมาข้างต้นว่า SQL เป็นภาษาฐานข้อมูลที่สามารถ
ใช้ในเรือ่งของการนิยามข้อมูลการเรียกใช้ หรือการควบคุมคำสั่งเหล่านี้จะช่วย
ประหยัดเวลาในการพัฒนาระบบงาน หรือนำไปใช้ในส่วนของการสร้างฟอร์ม (FORM) การทำ
รายงาน (REPORT) ของระบบงานต่าง ๆ ได้รวดเร็วขึ้น

ข้อตกลงมาตรฐานการใช้คำสั่ง SQL รูปแบบของคำสั่ง SQL ที่ใช้ในตำราเล่มนี้ยึด
รูปแบบคำสั่งของ SQL ที่สามารถนำไปประยุกต์ใช้กับ ORACLE ซึ่งเมื่อเข้าใจถึงคำสั่งเหล่านี้
ผู้ใช้สามารถนำไปประยุกต์ใช้กับคำสั่ง SQL ที่ใช้ในระบบจัดการฐานข้อมูลต่าง ๆ หรือสามารถ
นำไปศึกษาเพิ่มในรายละเอียดด้วยตัวเองได้ โดยทั่วไป คำสั่ง SQL หนึ่งจะจบด้วยเครื่องหมาย ;
รูปแบบคำสั่งแต่ละคำสั่งใน SQL มีสัญลักษณ์ที่ใช้แทนความหมายตามตารางที่ 2.1 ดังต่อไปนี้

ตารางที่ 2.1 ตารางแสดงข้อมูลสัญลักษณ์ที่ใช้แทนความหมาย

| สัญลักษณ์ | ความหมาย |
|--------------|--|
| ตัวพิมพ์ใหญ่ | คำสั่ง |
| <> | ชื่อต่าง ๆ หรือนิพจน์ที่ผู้ใช้กำหนดค่าขึ้นมา |
| (.....) | สามารถระบุเพิ่มเติมอีกตามสิ่งที่ระบุมาแล้วข้างหน้า |
| [] | คำสั่งนั้น ๆ จะมีสิ่งที่ระบุไว้ในเครื่องหมายนี้หรือไม่ก็ได้ |
| ::= | ส่วนที่อยู่ด้านขวาของเครื่องหมายนี้เป็นคำอธิบาย หรือคำนิยามของสิ่งที่อยู่ด้านซ้ายของเครื่องหมายนี้ |
| | อาจเลือกใช้สิ่งใดสิ่งหนึ่งที่อยู่ด้านซ้าย หรือด้านขวาของเส้นนี้ได้ |

ตารางที่ 2.2 ตารางแสดงข้อมูลประเภทข้อมูลชนิดตัวอักษร

| ลำดับที่ | ชื่อประเภทข้อมูล | รายละเอียด | เนื้อที่เก็บข้อมูล |
|----------|------------------|--|----------------------|
| 1 | VARCHAR(M) | สำหรับเก็บข้อมูลประเภทตัวอักษร ชนิดของฟิลด์ประเภทนี้จะต้องมีข้อกำหนดความยาวของข้อมูลลงไปด้วย สามารถกำหนดค่าได้ตั้งแต่ 1-255 ฟิลด์ | ขนาดข้อมูลจริง+1byte |
| 2 | CHAR(M) | สำหรับเก็บข้อมูลประเภทตัวอักษรที่ถูกจำกัดความยาวเอาไว้ คือ 255 ตัวอักษร ไม่สามารถปรับเปลี่ยนได้เหมือนกับ VARCHAR หากทำการสืบค้นโดยเรียงตามลำดับก็จะเรียงข้อมูล | ตามจำนวนอักษรที่ระบุ |
| 3 | TINYTEXT | ในกรณีที่ต้องการความยาว ๆ หรือต้องการที่จะค้นหาข้อความ โดยอาศัยพีเจอร์ FULLTEXT SEARCH ของ MySQL อาจเลือกที่จะไม่เก็บข้อมูลลงในฟิลด์ประเภท VARCHAR ที่มีข้อจำกัด | ขนาดข้อมูลจริง+1byte |
| 4 | TEXT | สำหรับเก็บข้อมูลประเภทตัวอักษร เช่นเดียวกับ TINYTEXT แต่สามารถเก็บได้มากขึ้น โดยสูงสุดคือ 65,535 ตัวอักษรหรือ 64KB | ขนาดข้อมูลจริง+2byte |
| 5 | MEDIUMTEXT | เก็บข้อมูลประเภทตัวอักษรเช่นเดียวกับ TINYTEXT แต่เก็บข้อมูลได้ 16,777,215 ตัวอักษร | ขนาดข้อมูลจริง+3byte |
| 6 | LONGTEXT | เก็บข้อมูลประเภทตัวอักษรเช่นเดียวกับ TINYTEXT แต่เก็บข้อมูลได้ 4,294,967,295 ตัวอักษร | ขนาดข้อมูล |
| 7 | ENUM | เก็บข้อมูลประเภทระบุค่าที่ต้องการ ถ้าไม่มีจะให้ค่า Null สามารถกำหนดค่าได้ 65,535 ตัวอักษร | ขนาดข้อมูล |
| 8 | BYTE | สำหรับเก็บข้อมูลที่มีความยาวแบบยี่ดหุ่่น | ขนาดข้อมูลจริง+1byte |
| 9 | BOOLEAN | สำหรับเก็บข้อมูลจริง หรือ เท็จ | ขนาดข้อมูลจริง+2byte |

ที่มา : (อนรรฆนงค์ คุณมณี, 2555: หน้า 203)

ตารางที่ 2.3 ตารางแสดงข้อมูลประเภทข้อมูลจำนวนเต็ม

| ลำดับที่ | ชื่อประเภทข้อมูล | ค่าตัวเลขแบบมีเครื่องหมาย | ค่าตัวเลขแบบไม่มีเครื่องหมาย | เนื้อที่เก็บข้อมูล |
|----------|---------------------------|---|-----------------------------------|--------------------|
| 1 | TINYINT(M) | -128 ถึง 127 0 | 0 ถึง 255 | 1 byte |
| 2 | SAMLLINT(M) | -32768 ถึง 32767 | 0 ถึง 65535 | 2 byte |
| 3 | MEDIUMINT(M) | -8388608 ถึง 8388607 | 0 ถึง 1677215 | 3 byte |
| 4 | INT(M) หรือ INTEGER(M) | -214783648 ถึง 2147483647 | 0 ถึง 4294967295 | 4 byte |
| 5 | BIGINT(M) | 9223372036854775808 ถึง 9223372036854775807 | 0 ถึง 1844674407370 9551615 | 5 byte |

ที่มา : (อนรรฆวงศ์ คุณมณี, 2555: หน้า 203)

ตารางที่ 2.4 ตารางแสดงข้อมูลประเภทข้อมูลชนิดจำนวนทศนิยม

| ลำดับที่ | ชื่อประเภทข้อมูล | ค่าตัวเลขแบบมีเครื่องหมาย | ค่าตัวเลขแบบไม่มีเครื่องหมาย | เนื้อที่เก็บข้อมูล |
|----------|------------------|--|---|--|
| 1 | FLOAT(M, D) | -3.402823466E+38 ถึง -17.175494351E-38 | 0 และ 1.175494351E38 ถึง 3.402823466E+38 | 4 byte |
| 2 | DOUBLE(M, D) | -1.79763931348623157E+308 ถึง -2.2250738585072014E - 308 | 2.250738585072014E 308 ถึง 1.7976931348623157E +308 | 8 byte |
| 3 | DECIMAL(M, D) | เก็บค่าเลขทศนิยมแบบระบุจำนวนหลัก M ทุกหลักรวมจุดทศนิยมและ D หลักหลังทศนิยม เช่น 123.34 ให้กำหนดเป็น DECIMAL(3,2) | เก็บค่าเลขทศนิยมแบบระบุจำนวนหลัก M ทุกหลักรวมจุดทศนิยมและ D หลักหลังทศนิยม เช่น 123.34 ให้กำหนด เป็น DECIMAL(3,2) | ถ้า d = 0 ขนาดที่เก็บ คือ m+1byte ถ้า d > 0 ขนาดที่เก็บ คือ m+2byte |

ที่มา : (อนรรฆวงศ์ คุณมณี, 2555: หน้า 204)

ตารางที่ 2.5 ตารางแสดงข้อมูลประเภทข้อมูลชนิดวันและเวลา

| ลำดับที่ | ชื่อประเภทข้อมูล | รายละเอียด | เนื้อที่เก็บข้อมูล |
|----------|------------------|---|--------------------|
| 1 | DATE | สำหรับเก็บข้อมูลประเภทวันที่ โดยเก็บได้จาก 1 มกราคม ค.ศ. 1000 ถึง 31 ธันวาคม ค.ศ.9999 โดย จะแสดงผลในรูปแบบ YYYY-MMDD | 3 byte |
| 2 | DATETIME | สำหรับเก็บข้อมูลประเภทวันที่ และเวลา โดยจะเก็บได้ ตั้งแต่ 1 มกราคม ค.ศ. 1000 เวลา 00:00:00 ไป จนถึง 31 ธันวาคม ค.ศ. 9999 เวลา 23:59:59 โดย รูปแบบการแสดงผลจะเป็น YYYY-MM-DD HH:MM:SS | 8 byte |
| 3 | TIMESTAMP(M) | สำหรับเก็บข้อมูลประเภทวันที่ และเวลาเช่นกันแต่จะเก็บในรูปแบบของ YYYYMMDDHHMMSS หรือ YMMDDHHMMSS หรือ YYYYMMDD หรือ YMMDD แล้วแต่ว่าจะระบุค่า M เป็น 14, 12, 8 หรือ 6 ตามลำดับ สามารถเก็บได้ตั้งแต่วันที่ 1 มกราคม ค.ศ. 1000 ไปจนถึงปี ค.ศ. 2037 | 8 byte |
| 4 | TIME | สำหรับเก็บข้อมูลประเภทเวลามีค่าได้ตั้งแต่ - 838:59:59 ไปจนถึง 838:59:59 โดยแสดงผลออกมาในรูปแบบ HH:MM:SS | 8 byte |
| 5 | YEAR(2/4) | สำหรับเก็บข้อมูลประเภทปี ในรูปแบบ YYYY หรือ YY แล้วแต่ว่าจะเลือก 2 หรือ 4 (หากไม่ระบุจะถือว่า เป็น 4 หลัก) | 1 byte |

ที่มา : (อนรรฆวงศ์ คุณมณี, 2555: หน้า 204)

2.2.4 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.2.4.1 หลักในการออกแบบเว็บไซต์หน้าเว็บเป็นสิ่งแรกๆ ที่ผู้ใช้จะได้เห็นขณะที่เปิดเข้าสู่เว็บไซต์ และยังเป็นสิ่งแรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วยหน้าเว็บจึงเป็นสิ่งสำคัญมาก เพราะเป็นสื่อกลางให้ผู้ชมสามารถใช้ประโยชน์จากข้อมูลของระบบงานของเว็บไซต์นั้นได้ โดยปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้นระบบเนวิเกชัน และองค์ประกอบอื่น ๆ ที่ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการใช้งานหลักสำคัญในการออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อความหมาย เกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้งาน การออกแบบเว็บไซต์ ต้องคำนึงถึง

1) ความเรียบง่าย การออกแบบเว็บไซต์ที่ดี จะต้องเน้นที่ความเรียบง่ายเป็นหลัก โดยเลือกนำเสนอเฉพาะสิ่งที่ต้องการนำเสนอจริงๆ ในรูปแบบที่หลากหลาย โดยอาจจะเป็นสีสัน กราฟิก ภาพเคลื่อนไหวหรือตัวอักษร ที่สำคัญจะต้องมีการนำเสนอที่ไม่ดูรกหน้าเว็บจนเกินไป

2) ความสม่ำเสมอ คือจะต้องมีรูปแบบ กราฟิก โทนสีและการตกแต่งต่างๆ ให้แต่ละหน้าบนเว็บไซต์มีความคล้ายคลึงกัน และเป็นแนวเดียวกันไปตลอดทั้งเว็บไซต์ เน้นการตกแต่งในรูปแบบเดียวกันทั้งหมด ต่างแค่การนำเสนอของแต่ละหน้าเท่านั้น

3) สร้างความโดดเด่น ต้องมีการสร้างความเป็นเอกลักษณ์และจุดเด่นให้กับเว็บไซต์ เพื่อให้สามารถสะท้อนถึงลักษณะขององค์กรได้มากที่สุด โดยการสร้างเอกลักษณ์ดังกล่าวนี้ อาจใช้ชุดสี รูปภาพ ตัวอักษรหรือกราฟิก นอกจากนี้ก็ต้องขึ้นอยู่กับว่า เป็นเว็บไซต์แบบทางการหรือไม่ เพื่อจะได้ออกแบบได้อย่างเหมาะสมที่สุด

4) เนื้อหาครบถ้วน เนื้อหาเป็นสิ่งสำคัญที่สุดของการสร้างเว็บไซต์ เพราะสิ่งที่ทำให้ผู้คนเกิดความสนใจ และหมั่นติดตามเว็บไซต์เหล่านั้นอยู่เสมอ ก็คือเนื้อหาที่มีความสมบูรณ์และน่าสนใจ นอกจากนี้จะต้องมีการปรับปรุง พัฒนาเนื้อหาบนเว็บให้มีความทันสมัยอยู่เสมอ

5) ระบบเนวิเกชันที่ใช้ง่าย ระบบเนวิเกชัน เป็นเสมือนป้ายบอกทางเพื่อให้ผู้ใช้งาน ไม่เกิดความสับสนในขณะที่ใช้งานเว็บไซต์ ซึ่งการออกแบบเนวิเกชันก็จะต้องเน้นที่ความเรียบง่าย ใช้งานสะดวก และมีความเข้าใจได้ง่าย ที่สำคัญจะต้องมีตำแหน่งการวางที่สม่ำเสมอ เพื่อให้ดูเป็นแนวทางการเดียวกัน

6) คุณภาพของเว็บไซต์ เว็บไซต์ที่ดีจะต้องมีคุณภาพ ทั้งสิ่งที่ปรากฏให้เห็นบนเว็บไซต์ ไม่ว่าจะเป็นกราฟิก ชนิดตัวอักษร รูปภาพหรือสีสันทันทีใช้ เนื้อหาที่นำมาแสดงผล ซึ่งหากเว็บไซต์มีคุณภาพก็จะสร้างความน่าเชื่อถือ และเป็นจุดเด่นที่ทำให้ผู้คนส่วนใหญ่เกิดความสนใจได้ดี

7) ความสะดวกในงานเข้าใช้งาน เว็บไซต์ควรให้ความสะดวกสบายแก่ผู้ใช้งานได้ดี คือจะต้องมีการแสดงผลได้ในทุกระบบปฏิบัติการ ไม่ว่าจะเป็นเว็บเบราว์เซอร์ คอมพิวเตอร์ โน้ตบุ๊กหรือบนโทรศัพท์มือถือ ที่สำคัญจะต้องมีความละเอียดของการแสดงผล และสามารถใช้งานได้โดยไม่มีปัญหาด้วย

8) ความคงที่ของการออกแบบ การออกแบบเว็บไซต์ควรจะมี ความคงที่ในการออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกัน และมีการเรียบเรียงเนื้อหาอย่างรอบคอบ ทำให้เว็บมีความน่าเชื่อถือ และดูมีคุณภาพ ช่วยสร้างความประทับใจให้กับผู้ใช้งาน

9) ความคงที่ของการทำงาน ระบบการทำงานบนเว็บไซต์จะต้องมีความคงที่ และสามารถใช้งานได้ดี ซึ่งนอกจากการออกแบบระบบการทำงานให้มีความทันสมัยและสร้างสรรค์แล้ว ก็จะต้องหมั่นตรวจสอบอยู่เสมอ เพราะหากระบบการใช้งานมีความผิดปกติก็จะได้แก้ปัญหาได้ทัน นอกจากนี้อาจมีการอัปเดตดีไซน์ให้ทันสมัยขึ้นบ่อยๆ

2.2.4.2 รูปแบบโครงสร้างของเว็บไซต์

การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างแบบเรียงลำดับ เป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิ้งค์เนื้อหา ก็จะลิ้งค์ไปที่ละหน้า มีทิศทาง การเข้าสู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ปุ่มเดินหน้า-ถอยหลังในการกำหนดทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย



ภาพที่ 2.9 โครงสร้างแบบเรียงลำดับ

ที่มา : <https://www.1belief.com/article/website-design/>

2) โครงสร้างแบบลำดับชั้น นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่างๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆ และมีการนำเสนอรายละเอียดค่อยๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความเข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีโฮมเพจเป็นจุดเริ่มต้น และจุดรวมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง



ภาพที่ 2.10 โครงสร้างแบบลำดับชั้น

ที่มา : <https://www.1belief.com/article/website-design/>

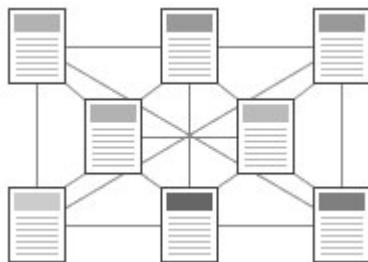
3) โครงสร้างแบบตาราง เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีความยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่างๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งานสามารถเปลี่ยนทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลา และยังทำให้เว็บไซต์มีความทันสมัย



ภาพที่ 2.11 โครงสร้างแบบตาราง

ที่มา : <https://www.1belief.com/article/website-design/>

4) โครงสร้างแบบในแมงมุม เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถเข้าถึงหน้าเว็บเพจต่างๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2.12 โครงสร้างแบบใยแมงมุม

ที่มา : <https://www.1belief.com/article/website-design/>

2.2.4.3 การเลือกใช้สีสำหรับการออกแบบเว็บไซต์

การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บ ว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไรต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่มดังต่อไปนี้

1) สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอดภัยและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2) สีโทนเย็น (Cool Colors) เป็นสีแห่งความสงบและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้โน้มน้าวจากในระยะไกลได้อีกด้วย

3) สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

สีสามารถสื่อถึงอารมณ์ความรู้สึกและสื่อความหมายของเว็บไซต์นั้นๆ ได้อย่างชัดเจน ดังนั้นหากเลือกใช้สีไม่เหมาะสมกับเนื้อหาหรือจุดประสงค์ของเว็บ ก็จะทำให้เว็บดูไม่น่าสนใจ ขาดความน่าเชื่อถือและทำให้ผู้ที่เคยเข้ามาใช้บริการไม่คิดจะกลับมาใช้บริการอีก

2.3 เครื่องมือในการออกแบบ และวิเคราะห์ข้อมูล

2.3.1 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM) หรือ Cross Industry Standard

Process for Data Mining หมายถึง กระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูลเพื่อการวิเคราะห์และนำไปใช้ประโยชน์ในทางธุรกิจ พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ DaimlerChrysler, SPSS และ NCR ในกระบวนการ CRISP-DM ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่

1) การทำความเข้าใจในธุรกิจ (Business understanding) เป็นการกำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์เพื่อหาความได้เปรียบทางการตลาด เพื่อนำมาแก้ไขปัญหาองค์กร และต้องสามารถระบุผลลัพธ์ที่มีได้

2) การทำความเข้าใจในข้อมูล (Data Understanding) โดยการรวบรวมข้อมูลที่เกี่ยวข้องให้อยู่รวมกลุ่มกัน คัดเลือกให้เหลือเพียงข้อมูลที่มีความสำคัญต่อการทำงาน

3) การเตรียมข้อมูล (Data Preparation) การจัดเตรียมข้อมูลประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4) การสร้างแบบจำลอง (Modeling) นำผลที่ได้มาทดลองทำแบบจำลอง เพื่อใช้ถ่ายทอดข้อมูลหรือทำการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปสารสนเทศเป็นข้อมูลที่น่าไปใช้ประโยชน์ได้จริงในทางธุรกิจ

5) การประเมินผล (Evaluation) การวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ตั้งไว้ในตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใดเพื่อเปลี่ยนแปลงแก้ไขให้ได้ตามผลลัพธ์ที่ต้องการ

6) เผยแพร่ผลวิเคราะห์ (Deployment) การนำเอาข้อมูลที่เป็นผลลัพธ์มาลงปฏิบัติจริงกับธุรกิจ โดยแปลงแนวคิดที่มีให้เกิดเป็นการนำไปใช้ประโยชน์ และติดตามรวบรวมผลที่ได้เพื่อการปรับปรุง

2.3.2 กระบวนการทำงานด้วย Streamlit

2.3.2.1 ความหมายของ Streamlit

Streamlit คือ Library ของ Python ที่เป็นแบบ open-source ง่ายต่อการสร้างและพัฒนาเว็บแอป สามารถเรียนรู้เครื่องมือได้ในเวลาอันสั้นและใช้งานได้อย่างมีประสิทธิภาพ Streamlit คือการทดลองทำสิ่งต่างๆด้วยตนเอง การทำงานของ Streamlit ทุกครั้งที่เราเพิ่มองค์ประกอบใหม่ ๆ ลงในสคริปต์จะบันทึก UI ของ Streamlit การทำงานของ Streamlit ช่วยในการโต้ตอบที่รวดเร็ว ดังนั้นเป้าหมายของการใช้งาน Streamlit คือการสร้างแอปแบบการโต้ตอบสำหรับข้อมูล รวมถึงการทำแบบจำลอง

2.3.3 โปรแกรมจัดการข้อมูลด้วยโพสต์เกรสคิวเอล (PostgreSQL)

“ข้อมูลสารสนเทศ” เป็นอีกหนึ่งปัจจัยที่ทำให้องค์กรก้าวหน้าไปได้อย่างรวดเร็ว การที่องค์กรนั้นมีข้อมูลมากกว่าคู่แข่งจะช่วยให้การพัฒนาหน้าองค์กรอื่นๆ ได้อย่างมีประสิทธิภาพ การจัดเก็บฐานข้อมูลจึงเป็นส่วนสำคัญ

PostgreSQL เป็นระบบการจัดการฐานข้อมูลเชิงวัตถุ-สัมพันธ์ (Object Relational Database Management หรือ ORDBMS) แบบ Open Source เริ่มต้นมาจากโครงการ University Ingres ตั้งแต่ปี ค.ศ. 1977 หลังจากนั้นได้พัฒนาเวอร์ชัน POSTGRES 4.2 โดยมหาวิทยาลัยแคลิฟอร์เนีย (Berkeley Computer Science department, University of California.)

โดยมีรูปแบบคำสั่งภาษา SQL ซึ่งเมื่อไม่นานมานี้ได้ออกเวอร์ชัน 13 มีการปรับปรุงประสิทธิภาพ ขนาดฐานข้อมูล การออกแบบดีไซน์ การควิรี่ระบบให้ดีขึ้น และยังรองรับฟังก์ชัน gen_random_uuid() โดยไม่ต้องเปิดใช้งานปลั๊กอินใด ๆ

คุณสมบัติของ PostgreSQL แบ่งออกได้ดังนี้

- 1) สามารถทำงานบนระบบปฏิบัติการได้ทั้ง Linux, UNIX (AIX, BSD, HP-UX, SGI Irix, Mac OS X, Solaris, Tru64) และ Windows
- 2) มีความรวดเร็วในการควิรี่ข้อมูล
- 3) สามารถทำงานแบบขนาดทำให้เรียก Autovacuum ได้
- 4) ลดการใช้ข้อมูลซ้ำใน B-tree index ทำให้ไม่สูญเสียทรัพยากรในการใช้งาน

2.4 วรรณกรรมที่เกี่ยวข้อง

ปรีชา รัตน์ง (2562) ได้กล่าวถึงอายุการเจริญเติบโตของพันธุ์ผักอายุสั้นแต่ละชนิดจะมีอายุในการเจริญพร้อมเก็บเกี่ยวแต่ละชนิดดังนี้ ผักบุ้ง 20-25 วัน, ผักกวางตุ้ง 25-30 วัน, คะน้า 30-35 วัน, ผักกาดขาวปลีพันธุ์เบา 25-30 วัน, ผักชี 40-50 วัน และอายุการเก็บเกี่ยวผักสลัดแต่ละชนิด คอสมอส 45-50 วัน, บัตเตอร์เฮด 45-50 วัน, กรีนไคค 45-50 วัน, เรดไคค 45-50 วัน, เรดโครอล 45-50 วัน, กรีนโครอล 45-50 วัน, ผักกาดหอม 45 วัน

อติศา ปานคง และ เพ็ญนภา ปาละปิ่น (2561) จากการวิจัยเกี่ยวกับปัจจัยที่เป็นอุปสรรคในการปลูกพืชหมุนเวียนในพื้นที่เกษตรกรรม อำเภอบ้านดุง จังหวัดอุดรธานี พบว่าปัจจัยที่เป็นอุปสรรคในการปลูกพืชหมุนเวียนในพื้นที่ได้แก่ สภาพปัญหาดินเค็ม น้ำเค็มในพื้นที่ส่งผลให้เกษตรกรไม่สามารถปลูกพืชหมุนเวียนได้หลากหลายชนิดพันธุ์ สภาพปัญหาน้ำแล้งฝนแล้งเป็นอุปสรรคทำให้ขาดแคลนน้ำที่จะนำมาใช้ในการปลูกพืชหมุนเวียนตลอดระยะเวลาของการเพาะปลูก การขาดแคลนทุนทรัพย์และขาดแคลนกำลังคน

ปิยะพร แซ่ลิ้ม (2562) จากการวิจัยการพยากรณ์ผลผลิตข้าวหอมมะลิ ในจังหวัดนครราชสีมาพบว่า ตัวแบบพยากรณ์ที่ได้จากเทคนิคการพยากรณ์เชิงสถิติทั้ง 2 วิธี พิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ พบว่า ตัวแบบการถดถอยพหุคูณ (Multiple Regression Analysis) โดยวิธีการถดถอยแบบลำดับขั้น (Stepwise Regressions) เป็นตัวแบบที่เหมาะสมกับการพยากรณ์ผลผลิตข้าวหอมมะลิจังหวัดนครราชสีมา และเมื่อนำตัวแบบมาพยากรณ์ผลผลิตข้าวหอมมะลิ จังหวัดนครราชสีมาล่วงหน้าอีก 1 คาบเวลา คือปี 2560 จากผลการพยากรณ์สรุปได้ว่า ปริมาณผลผลิตข้าวหอมมะลิจังหวัดนครราชสีมา จะมีแนวโน้มเพิ่มสูงขึ้นจากปี 2559 ประมาณ 7.429%

อสมมา กุลวานิชไชยนันท์ (2561, น.200-201) ได้กล่าวถึงการวิเคราะห์ถดถอยเชิงเส้นพหุคูณไว้ว่า เมื่อใดก็ตาม ที่ต้องการความสัมพันธ์ของปัจจัยหลายตัวที่มีผลต่อผลลัพธ์ หรือ Target จะเรียก Linear Regression Model นั้นว่า Multiple Linear Regression (MLR) แปลเป็นไทยว่า การวิเคราะห์ถดถอยเชิงเส้นพหุคูณ ข้อดีของการทำ MLR นอกจากเป็นการหาความสัมพันธ์ (หรือได้สมการออกมาให้เห็นอย่างชัดเจน) แล้ว ยัง สามารถวิเคราะห์ความสำคัญของปัจจัยแต่ละตัวที่มีผลต่อผลลัพธ์นั้นได้อีกด้วย โดยการเช็คค่า P-Value ค่าความน่าจะเป็นที่จะได้ผลลัพธ์เท่ากับหรือเกินกว่าที่สังเกตได้สมมุติฐานหลัก ซึ่งหาก ค่า P-value มากกว่า 0.05 จะบ่งบอกว่าตัวแปรนั้นอาจจะไม่มีความสำคัญต่อระบบสมการ) อีกทั้งยังกล่าวอีกด้วยว่า ข้อดีของการใช้ Machine Learning ก็คือใส่ข้อมูลอะไรก็ได้คำตอบ แต่ข้อเสียของ Machine Learning ก็คือใส่ข้อมูลอะไรก็ได้คำตอบ เพราะฉะนั้นความสำคัญของการสร้าง Machine Learning จึงคือต้องเลือกด้วยว่าจะใช้ข้อมูลอะไรใส่เข้าไป เพื่อให้ได้คำตอบอย่างไร และตรงนี้เองที่ทำให้ MLR มีข้อดีเหนือ Algorithm ประเภทอื่น ๆ เพราะ MLR มี Pvalue ซึ่งทำให้ทราบได้ว่า ปัจจัยแต่ละตัว มีความสำคัญกับ Model มากน้อยแค่ไหน สามารถตัดออกไปได้หรือไม่

จุมพล หนีมพานิช และวรวรลัญช์ โรจนพล (2561) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเชิงคุณภาพโดยทั่วไปหมายถึง การสรุปเนื้อหาของระขของข้อมูล และหาแบบแผนของความสัมพันธ์ระหว่างข้อมูลหรือตัวแปร หรือหมายถึงการสรุปพรรณนาและการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูล หรือหมายถึง การนำเอาข้อมูลที่ได้จากการค้นคว้าวิจัยมาจัดกระทำให้เป็นระบบและหาความหมาย แยกแยะ องค์ประกอบ รวมทั้งเชื่อมโยงและหาความสัมพันธ์ของข้อมูล

ดังนั้น วรรณกรรมที่เกี่ยวข้องทั้งหมดที่ได้กล่าวมาข้างต้นนั้น คณะผู้จัดทำโครงการได้เล็งเห็นความสำคัญของการทำเทคโนโลยี และการเลือกเครื่องมือมาประยุกต์ใช้ในการวิเคราะห์และพยากรณ์ผลผลิตพืชผักอายุสั้นตระกูลสลัดเพื่อส่งเสริมเกษตรกรใน โครงการพัฒนาพื้นที่สูงแบบโครงการหลวงปางเป๋ อำเภอมะแม่ง จังหวัดเชียงใหม่ โดยใช้แบบจำลองการถดถอยด้วยเทคนิคแบบจำลองการถดถอยพหุคูณ เนื่องจากเทคโนโลยีปัจจุบันได้เข้ามามีส่วนช่วยในการจัดการข้อมูล การนำเสนอ และการดำเนินงานต่าง ๆ ให้มีประสิทธิภาพและความถูกต้องมากยิ่งขึ้น

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้น ที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลและพยากรณ์ผลผลิตพืชผักอายุสั้นตระกูลสลัดเพื่อส่งเสริมเกษตรกรในโครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่ ทั้งหมดนั้นคณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย กระบวนการวิเคราะห์ CRISP-DM และทำการสร้างโมเดลแบบจำลองการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression Model) ทำการทดสอบโมเดล เพื่อหาปัจจัยที่มีผลต่อผลลัพธ์ จากนั้นนำข้อมูลมาทำการแสดงผลแบบ Visualization ในรูปแบบของแดชบอร์ด และสร้างส่วนติดต่อประสานระหว่างผู้ใช้กับระบบการทำเหมืองข้อมูล โดยใช้ชุดคำสั่งภาษา Python ในการเขียนเว็บไซต์