บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลยอดขายสินค้าของร้าน พี.ที.เมดิคอลเซลล์แอนด์เซอร์วิส ในช่วง COVID-19 เพื่อพยากรณ์การซื้อของลูกค้า ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

- 2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล
- 2.1.2 แนวคิดการพยากรณ์
- 2.1.3 แนวคิดเกี่ยวกับอนุกรมเวลา
- 2.1.4 แนวคิดการวิเคราะห์อนุกรมเวลา

2.2 ทฤษฎี

- 2.2.1 ทฤษฎีการ visualization
- 2.2.2 ทฤษฎีเกี่ยวกับอนุกรมเวลา

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

- 2.3.1 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)
- 2.3.2 แบบจำลองอนุกรมเวลา (Time Series)
- 2.3.3 การแสดงผลแบบ Visualization
- 2.3.4 เทคนิคการพยากรณ์ (Forecasting Techniques)

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเชิงคุณภาพนับเป็นขั้นตอนที่สำคัญในกระบวนการวิจัย วิธีการหลักที่ใช้ในการวิเคราะห์ข้อมูลเชิงคุณภาพ ประกอบด้วย การจำแนกและจัดระบบข้อมูล และการหาความสัมพันธ์ของข้อมูล มีจุดมุ่งหมายที่จะแยกแยะและอธิบายองค์ประกอบ ความหมาย และความสัมพันธ์ของปรากฏการณ์ ภายใต้เงื่อนไขและสภาพแวดล้อมทางสังคม และวัฒนธรรม เป็นวิธีการสร้างข้อสรุปจากการศึกษาจากข้อมูลจำนวนหนึ่ง ซึ่งมักไม่ใช้สถิติในการวิเคราะห์หรือถ้าใช้สถิติก็ไม่ได้ถือว่าสถิติเป็นวิธีการวิเคราะห์หลัก แต่จะถือเป็นข้อมูลเสริม ในการวิเคราะห์ข้อมูลเชิงคุณภาพนั้น ผู้วิเคราะห์ข้อมูลนับเป็นผู้ที่มีบทบาทสำคัญยิ่ง และควรมีความรอบรู้ในเรื่องแนวคิดทฤษฎีอย่างกว้างขวาง มีความเป็นสหวิทยากรอยู่ในตัวเอง มีความสามารถทางภาษาสามารถเชื่อมโยงข้อความ และสร้างข้อสรุปเป็นกรอบแนวคิดและตีความหมายของข้อมูลได้หลายแบบ วิธีการวิเคราะห์ข้อมูลในการวิจัยเชิงคุณภาพ อาจใช้การวิเคราะห์ข้อมูลแบบสร้างข้อสรุป หรือการวิเคราะห์เนื้อหา สามารถพิจารณาได้จากความเชื่อถือได้ การถ่ายโอนผลการวิจัย และการพึ่งพากับเกณฑ์อื่น และการยืนยันผลความหมายของคำศัพท์ที่สำคัญเกี่ยวกับการวิเคราะห์งานวิจัยเชิงคุณภาพ

2.1.2 แนวคิดการพยากรณ์

การพยากรณ์ (Forecasting) หมายถึง การคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคต โดยอาศัยการวิเคราะห์ข้อมูลในอดีต ข้อมูลปัจจุบัน และจากประสบการณ์ สามารถนำไปใช้เพื่อให้ทราบถึงแนวโน้มการเปลี่ยนแปลงของสถานการณ์หรือสภาพแวดล้อมที่จะมีผลในอนาคตและทำให้สามารถที่จะวางแผนหรือกำหนดนโยบายเพื่อให้บรรลุวัตถุประสงค์ได้ ถ้าพิจารณาในมุมมองของนักธุรกิจ มักจะมีการพยากรณ์ยอดขายสินค้าของตนเอง พยากรณ์อัตราดอกเบี้ยเงินกู้พยากรณ์ราคาสินค้าของคู่แข่ง พยากรณ์ปริมาณการใช้วัตถุดิบ ถ้าเป็นนักการเงินจะต้องทำการพยากรณ์ราคาหุ้น พยากรณ์อัตราผลตอบแทนของหลักทรัพย์ หรือถ้าเป็นนักเศรษฐศาสตร์มักต้องพยากรณ์อัตราการเจริญเติบโตทางเศรษฐกิจ พยากรณ์อัตราเงินเฟ้อ พยากรณ์อัตราการว่างงาน

ในการพยากรณ์ตัวแปรใด ๆ ก็ตาม เราจะต้องใช้ข้อมูลของตัวแปรนั้น ในอดีตที่ผ่านมาเช่น หากนักธุรกิจต้องการพยากรณ์ยอดขายของบริษัทตนเองในเดือนหน้า ข้อมูลที่สำคัญที่สุดที่ต้องมีก็คือยอดขายของบริษัทที่ผ่านมาในอดีต จากนั้นผู้บริหารจะต้องทำการวิเคราะห์ข้อมูลยอดขายในอดีต แล้วจึงนำผลการวิเคราะห์ที่ได้ไปใช้พยากรณ์ข้อมูลนั้น

สาเหตุที่ต้องมีการรวบรวมข้อมูลยอดขายในอดีตเนื่องจากการวิเคราะห์ข้อมูลยอดขายในอดีตจะช่วยให้สามารถระบุถึงรูปแบบที่ค่าของตัวแปรยอดขายนั้นเป็นอยู่และการนำผลการวิเคราะห์(หรือรูปแบบที่ระบุได้) ไปใช้พยากรณ์ยอดขายของบริษัท ซึ่งจะต้องอยู่ภายใต้ข้อสมมุติว่า “รูปแบบที่ระบุได้จากข้อมูลยอดขายในอดีตนั้น ต้องเหมือนเดิมหรือไม่เปลี่ยนแปลงในอนาคต”

2.1.3 แนวคิดเกี่ยวกับอนุกรมเวลา

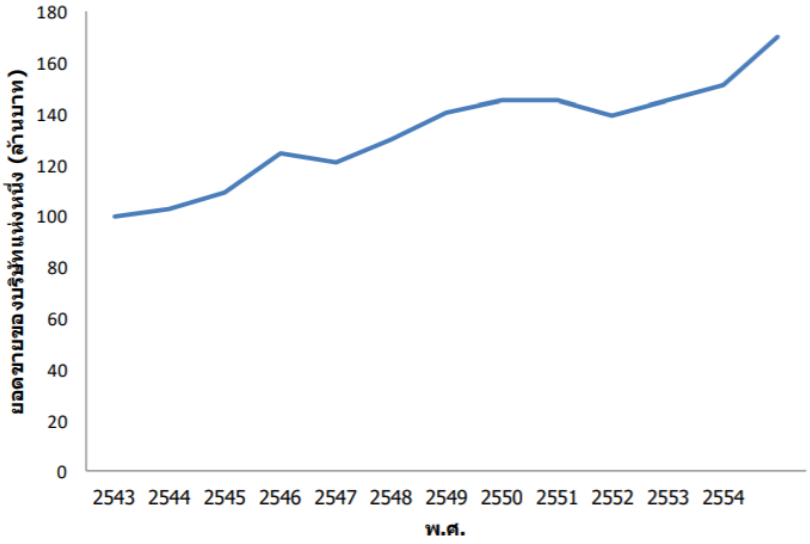
ข้อมูลอนุกรมเวลาเป็นค่าของข้อมูลที่ถูกเก็บควบคู่กับจุดเวลาที่เกิดขึ้น ดังนั้นทุกจุดข้อมูลในอนุกรมเวลาใด ๆ จะมีเวลากำกับเสมอ ซึ่งข้อมูลลักษณะนี้มักจะพบเห็นได้โดยทั่วไป ทั้งใกล้และไกลตัว เช่น ข้อมูลอนุกรมเวลาของตลาดหุ้น (stock market) นอกจากนี้ยังมีข้อมูลคลื่นไฟฟ้าสมอง (Electroencephalography หรือ EEG) ข้อมูลคลื่นไฟฟ้าหัวใจ (Electrocardiography หรือ ECG) ข้อมูลสภาพอากาศในแต่ละวัน หรือแม้แต่ข้อมูลรายได้ของแม่ค้าที่ได้รับในแต่ละชั่วโมงก็นับเป็นข้อมูลอนุกรมเวลาเช่นเดียวกัน จะเห็นว่าข้อมูลอนุกรมเวลานั้นเป็นข้อมูลที่พบได้โดยทั่วไปในชีวิตประจำวัน

ซึ่งนอกจากจะเป็นข้อมูลในลักษณะที่กล่าวไปข้างต้นแล้ว ข้อมูลอนุกรมเวลายังสามารถนำข้อมูลประเภทอื่นมาเปลี่ยนแปลงเป็นข้อมูลอนุกรมเวลาได้เช่นกัน ซึ่งประโยชน์ของการเปลี่ยนเป็นข้อมูลอนุกรมเวลานั้นคือการทำให้สามารถทำงานได้รวดเร็วยิ่งขึ้นในหลายชุดข้อมูล

2.1.4 แนวคิดการวิเคราะห์อนุกรมเวลา

การวิเคราะห์อนุกรมเวลาเป็นการวิเคราะห์ลักษณะหรือรูปแบบของอนุกรมเวลา โดยสังเกตจากการเปลี่ยนแปลงของข้อมูลในแต่ละช่วงเวลาว่ามีการเปลี่ยนแปลงไปในลักษณะใด มีการเคลื่อนไหวของข้อมูลอย่างไรเนื่องจากอนุกรมเวลาแต่ละประเภทมีลักษณะที่แตกต่างกัน ดังนั้นมีการกำหนดองค์ประกอบของอนุกรมเวลาเป็น 4 ส่วน ดังนี้

1) แนวโน้ม (Trend) คือ ส่วนที่ทำให้อนุกรมเวลามีค่าเพิ่มขึ้นเรื่อย ๆ หรือลดลงเรื่อย ๆ เมื่อเวลาผ่านไป เรามักใช้แนวโน้มในการบอกว่าอนุกรมเวลาที่เก็บข้อมูลมามีอัตราการเพิ่มขึ้นหรืออัตราการลดลงในระยะยาว เช่น ข้อมูลยอดขายรายเดือนของบริษัทแห่งหนึ่งแสดงได้ดังรูปต่อไปนี้



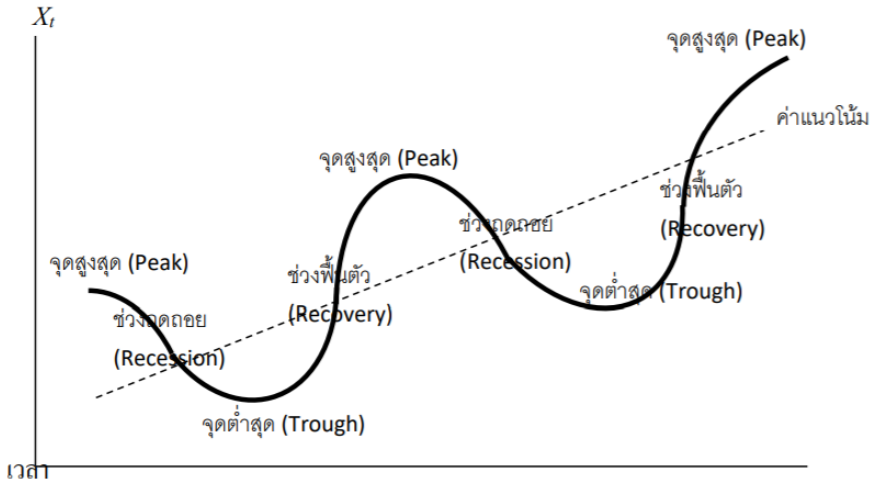
ภาพที่ 2.1 แสดงยอดขายรายเดือนของบริษัทแห่งหนึ่ง

จากรูปที่ 2.1 เรากล่าวได้ว่าแนวโน้มยอดขายสินค้าของบริษัทมีลักษณะเพิ่มขึ้นเรื่อย ๆ ซึ่งอาจมีสาเหตุมาจากประชากรในประเทศเพิ่มขึ้นเรื่อย ๆ รายได้ของคนในประเทศมากขึ้นเรื่อย ๆ หรือเทคโนโลยีการผลิตดีขึ้น บริษัทจึงสามารถขายสินค้าได้เพิ่มขึ้นเรื่อย ๆ เมื่อเวลาผ่านไป จากองค์ประกอบของงอนุกรมเวลาดังกล่าวทำ ให้สามารถสร้างแบบจำลองของข้อมูลอนุกรมเวลาได้ โดยมีแบบจำลองของอนุกรมเวลา ดังนี้

1) แบบจำลองผลบวก (Additive Model) เป็นแบบจำลองที่องค์ประกอบทั้ง 4 ส่วน ของอนุกรมเวลานั้น ๆ ไม่ขึ้นต่อกัน โดยเมื่อองค์ประกอบใด ๆ มีค่าเปลี่ยนแปลงไป จะไม่ส่งผลกระทบต่อการเปลี่ยนแปลงขององค์ประกอบอื่น ๆ 2) แบบจำลองผลคูณ (Multiplicative Model) เป็นแบบจำลองที่องค์ประกอบทั้ง 4 ส่วนของอนุกรมเวลานั้น ๆ สัมพันธ์กัน โดยเมื่อองค์ประกอบใด ๆ มีค่าเปลี่ยนแปลงไปจะส่งผลให้องค์ประกอบอื่น ๆ มีการเปลี่ยนแปลงตามไปด้วย

2) วัฏจักร (Cycle) คือ ส่วนที่ทำให้อนุกรมเวลาที่เก็บข้อมูลได้มีค่าเพิ่มขึ้น และลดลงสลับกันไปรอบ ๆ ค่าแนวโน้ม (ซึ่งแสดงด้วยเส้นประดังรูปที่ 2.2) การนับระยะเวลาของส่วนวัฏจักร จะนับจุดสูงสุดหนึ่ง (peak) ไปยังอีกจุดสูงสุดหนึ่ง หรือจากจุดต่ำสุดหนึ่ง (trough) ไปยังอีกจุดต่ำสุดหนึ่ง ซึ่งจะต้องกินเวลาตั้งแต่ 2 ปี ถึง 10 ปีขึ้นไป (หรือนานกว่านั้น) ส่วนของวัฏจักรจะเริ่ม ณ เวลาใดก็ได้ ตัวอย่างของวัฏจักรแสดงได้ในรูปที่ 2.2 เมื่อส่วนของวัฏจักรอยู่ในช่วงที่ทำให้อนุกรมเวลามีค่าลดลง เราจะเรียกว่าช่วงถดถอย (Recession) และหลังจาก

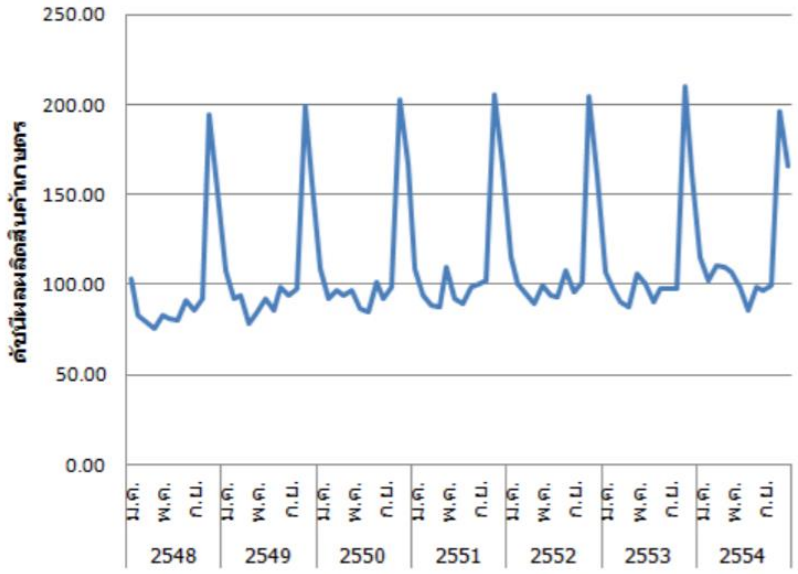
ผ่านจุดต่ำสุดไปแล้ว ส่วนของวัฏจักรที่ทำให้อนุกรมเวลามีค่าเพิ่มขึ้น เราจะเรียกว่า ช่วงฟื้นตัว (Recovery)



ภาพที่ 2.2 แสดงอนุกรมเวลาที่มีส่วนของวัฏจักร

3) ความผันแปรจากฤดูกาล (Seasonal Variations) คือ รูปแบบในช่วงเวลาหนึ่งของอนุกรมเวลาที่จะเป็นภายใน 1 ปีและจะเป็นแบบนี้ซ้ำกันทุกปี ตัวอย่างเช่น อุณหภูมิเฉลี่ยในเดือนเมษายน จะสูงกว่าอุณหภูมิเฉลี่ยในเดือนอื่น ๆ และจะเป็นเช่นนี้ซ้ำ ๆ กันทุกปี ค่าใช้ไฟฟ้าในเดือนพฤศจิกายน–ธันวาคม จะต่ำกว่าค่าใช้ไฟฟ้าในเดือนอื่น ๆ และเป็นเช่นนี้ทุกปี บริษัททัวร์จะมีรายรับในช่วงปิดเทอมสูงกว่าเดือนอื่น ๆ และเป็นเช่นนี้ทุกปี ยอดขายห้างสรรพสินค้าในเดือนธันวาคมจะสูงกว่ายอดขายเดือนอื่น ๆ และเป็นเช่นนี้ทุกปี

เมื่อพิจารณารูปที่ 2.3 ซึ่งแสดงข้อมูลดัชนีผลผลิตสินค้าเกษตรกรรมรายเดือนตั้งแต่ปี2548–2554 ของประเทศหนึ่ง เมื่อเราสังเกต ณ ปี 2548 จะพบว่าดัชนีผลผลิตสินค้าเกษตรกรรมในช่วงเดือนพฤศจิกายนและธันวาคม สูงกว่าเดือนอื่น ๆ และในปี อื่น ๆ ก็จะมีลักษณะเช่นนี้ดังนั้น เราก่าวได้ว่าดัชนีผลผลิตอุตสาหกรรมรายเดือนของประเทศนี้มีอิทธิพลของความผันแปรจากฤดูกาลเข้ามาเกี่ยวข้องในเดือนพฤศจิกายนและธันวาคมของทุกปีทั้งนี้อาจเป็นเพราะผลผลิตทางการเกษตรของประเทศนี้จะออกมาพร้อม ๆ กันในช่วงเวลาดังกล่าว



ภาพที่ 2.3 แสดงดัชนีผลผลิตสินค้าเกษตรของประเทศหนึ่งเป็นรายเดือนตั้งแต่ปี 2548–2554

4) ความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations) คือ ส่วนที่ทำให้อนุกรมเวลามีค่าที่ผิดปกติไปจากรูปแบบที่เคยเป็น มักเกิดจากเหตุการณ์ไม่คาดฝัน (Shock) เช่น แผ่นดินไหว สึนามิ ระเบิด การหยุดงานประท้วง ฯลฯ ส่วนความผันผวนจากเหตุการณ์ไม่ปกติคำนวณจากการนำค่าของส่วนแนวโน้ม ค่าของส่วนวัฏจักร และค่าของความผันแปรจากฤดูกาลไปหักล้างค่าอนุกรมเวลานั้น นั่นเอง

2.2 ทฤษฎี

2.2.1 ทฤษฎีการ Visualization

จินตทัศน์ (Visualization) หมายถึงการนำเสนอภาพแทนข้อมูลที่ต้องการจะขยายความก่อให้เกิดความรู้ความเข้าใจ ซึ่งหมายถึงภาพหรือกราฟิก การใช้สี รูปร่าง หรือรูปทรงสัญลักษณ์ เป็นการแสดงผลข้อมูลในระหว่างผู้ใช้งานและคอมพิวเตอร์ในรูปแบบของรูปภาพ โดยช่วยให้ผู้ใช้สามารถจดจำ และเรียนรู้ผ่านการมองเห็น (Ware,2004) ซึ่งได้มีการนำคอมพิวเตอร์เข้ามาช่วยในการนำเสนอและสามารถนำไปใช้สนับสนุนการตัดสินใจได้ทำให้ผู้ใช้งานเกิดการเรียนรู้และง่ายต่อการทำความเข้าใจ (Card et al., 1999)

การสร้างจินตทัศน์ หรือการแปลงข้อมูลดิบลงในรูปแบบของกราฟ หรือภาพทำให้ผู้ใช้สามารถเข้าใจข้อมูลที่แฝงอยู่ได้ง่ายขึ้น (Yeh, 2006) กระบวนการรับรู้รูปแบบนี้ง่ายต่อการทำความเข้าใจทำให้ผู้ใช้เข้าใจความหมายของข้อมูลมากขึ้น วิธีการสร้างจินตทัศน์เป็น

วิธีที่ช่วยให้ผู้ใช้มีความเข้าใจในข้อมูล ทำให้ผู้ใช้สามารถประเมินผล และวิเคราะห์ข้อมูลได้ง่ายขึ้น (Khan &Khan, 2011)

การสร้างจินตทัศน์มาใช้งานขึ้นอยู่กับชนิดข้อมูลและความต้องการแสดงผลข้อมูล โดยการนำเสนอข้อมูลในลักษณะนี้ส่งผลให้เกิดการรับรู้ด้วยภาพ ประกอบด้วยวัตถุ ฉาก หรือภาพคน และกระบวนการแสดงแนวความคิดกับวัตถุที่ได้ออกแบบสำหรับการสร้างคลาส (Class) ซึ่งเน้นไปที่การใช้งาน โดยการสร้างจินตทัศน์เป็นตัวแทนกราฟิกที่บ่งบอกถึงความคิดที่ซับซ้อนได้อย่างชัดเจนแม่นยำ และมีประสิทธิภาพ

จินตทัศน์ (Visualization) เป็นการนำเสนอข้อมูล ในรูปแบบของโครงสร้างภาพ (Visual form) ซึ่งเป็นการแสดงข้อมูลในรูปแบบของรูปภาพ กราฟ หรือแผนภาพ โดยผลลัพธ์ของจินตทัศน์เป็นการแสดงข้อมูลให้ผู้ใช้สามารถสัมผัสได้ซึ่งกระบวนการในการแสดงหรือการนำข้อมูลออกมาผู้ใช้อาจจะไม่ทันสังเกต แต่การสร้างจินตทัศน์มีจุดมุ่งหมายในการส่งข้อมูลไปยังระบบ ทำให้เกิดการรับรู้ โดยภาพของผู้ใช้งานระบบโดยจินตทัศน์เป็นสิ่งจำเป็นในการค้นหาข้อมูล และใช้ในการวิเคราะห์ข้อมูล เพื่อช่วยลดช่องว่างระหว่างผู้ใช้งาน และข้อมูล ซึ่งยังช่วยให้ผู้ใช้สามารถเรียนรู้ได้อย่างมีประสิทธิภาพ (Diehl, 2007)

2.2.2 ทฤษฎีเกี่ยวกับอนุกรมเวลา

อนุกรมเวลา (Time series) หมายถึง การเก็บรวบรวมข้อมูลของตัวแปรหนึ่งตามลำดับเวลา ตัวอย่างเช่น ข้อมูลราคาหุ้นรายวันตั้งแต่วัน ที่ 1 มกราคม 2554 – 30 มิถุนายน 2556 ข้อมูลอัตรา แลกเปลี่ยนรายสัปดาห์ตั้งแต่สัปดาห์ที่ 1 – สัปดาห์ที่ 52 ของปี 2555 ข้อมูลอัตราเงินเฟ้อรายเดือน ตั้งแต่เดือน มีนาคม 2540 — ตุลาคม 2554 ข้อมูล GDP รายไตรมาส ตั้งแต่ไตรมาสที่2ของปี 2525 — ไตรมาสที่ 4 ของปี 2554 ข้อมูลผลผลิตข้าวรายปีตั้งแต่ปี 2520 — 2554 และข้อมูลอัตราการว่างงาน รายปีตั้งแต่ปี 2530 — 2554 เป็นต้น ตัวแปรที่ยกตัวอย่างมาข้างตนนี้ล้วนเป็นตัวแปรสุ่ม (random or stochastic variables) ทั้งหมด ทั้งนี้เพราะในแต่ละช่วงเวลาข้อมูลดังกล่าวสามารถเพิ่มขึ้น หรือลดลง หรือเท่าเดิมก็ได้ซึ่งไม่อาจทราบล่วงหน้าได้ เมื่อกรณีนี้เกิดขึ้นเราจะเรียกว่าเป็น อนุกรมเวลาแบบสุ่ม (Stochastic Process หรือ Random Process) และหลังจากข้อมูลของตัวแปรที่สนใจถูกเก็บรวบรวมมาแล้ว ไม่ว่าจะเป็นรายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปี ค่าทางสถิติเบื้องต้น ได้แก่ ค่าเฉลี่ย (Mean) ความแปรปรวน (Variance) และความแปรปรวนร่วมระหว่างช่วงเวลา (Autocovariance) จะต้องสามารถคำนวณได้เสมอ อนุกรมเวลาแบบสุ่มของตัวแปรหนึ่งจะเขียนแทนด้วยสัญลักษณ์ X_t หรือ $\{X_t\}$ ก็ได้ โดยค่าทาง สถิติทั้ง 3 ค่าข้างต้น ได้แก่

- ค่าเฉลี่ยของ X_t เขียนแทนด้วยด้วย $E(X_t) = \mu_t$
- ค่าความแปรปรวนของ X_t เขียนแทนด้วย $var(X_t) = \sigma_t^2$
- ค่าความแปรปรวนร่วมระหว่างช่วงเวลา t_tและ t_2 ของ X_t เขียนแทนด้วย $\gamma_{t_1,t_2} = cov(X_{t_1}, X_{t_2}), t_1 \neq t_2$

โดย μ_t, σt^2 และ γ_{t_1,t_2} เรียกรวมกันว่าค่าพารามิเตอร์ (parameters) และเมื่อฟังก์ชันความน่าจะเป็นของ X_t เป็นแบบการแจกแจงแบบปกติ เราจะเรียก X_t ว่าเป็นอนุกรมเวลาแบบเกาส์เซียน (Gaussian process)

ในกรณี ที่ X_t ถูกเก็บรวมรวมต่อเนื่องกันมาเป็นจำนวน T ช่วงเวลา จะพบว่ามีค่าพารามิเตอร์ μ_t (t =1, 2, ..., T) จำนวน T ตัว ได้แก่ μ_1, μ_2, ..., μ_T และมีค่าพารามิเตอร์ σt^2 (t =1, 2, ..., T) จำนวน T ตัว ได้แก่ σ_1^2 , σ_2^2 , ... , σ_T^2 และมีค่าพารามิเตอร์ γ_{t_1,t_2} (t_1, t_2 =1, 2, ..., T และ $t_1 \neq t_2$) จำนวน $\frac{T(T+1)}{2}$ ตัว[2] รวมค่าพารามิเตอร์ทั้งหมดคือ $T + T + \frac{T(T+1)}{2}$ $\frac{T(T+3)}{2}$ ตัว นั่นคือหากเราเก็บรวบรวมข้อมูลอนุกรมเวลาแบบสุ่มจำนวน 120 เดือน จะพบว่ามี ค่าพารามิเตอร์ถึง 120 + 120 + 120(120–1) 2 = 120(120+3) 2 = 7,380 ตัว เพื่อที่จะลดจำนวน พารามิเตอร์ที่ไม่ทราบค่านี้เราสามารถทำได้โดยกำหนดข้อสมมุติคือให้ยอดขายรายเดือนดังกล่าวมีความนิ่ง (stationary[3])

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP–DM)

งานวิทยาการด้านข้อมูล (Data Science) ซึ่งกำลังมีบทบาทอย่างมากในโลกยุคปัจจุบัน และทวีความสำคัญยิ่งขึ้นในอนาคตก็มี CRISP–DM เป็นกระบวนการหลักในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์และใช้ประโยชน์ในทางธุรกิจ กระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่

1.รู้จักและเข้าใจในธุรกิจ (Business understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม หัวหน้าโครงการหรือที่ปรึกษาด้านการวางระบบวิเคราะห์ข้อมูล จะต้องทำการสัมภาษณ์ หรือรับฟังปัญหาความต้องการจากผู้บริหารองค์กร และหน่วยงานต่างๆ ที่จะนำผลการวิเคราะห์ข้อมูลไปใช้

ประโยชน์ โดยความต้องการทั้งหมดจะนำมาจัดลำดับความสำคัญ และกำหนดวัตถุประสงค์ที่จะนำไปสู่รูปแบบการวิเคราะห์ข้อมูลขององค์กร เช่น ผู้บริหารห้างสรรพสินค้า ต้องการรู้ว่าอะไรเป็นเหตุปัจจัยที่ทำให้ลูกค้าเป้าหมายตัดสินใจและเลือกที่จะเข้าห้าง ไม่ว่าจะเพื่อการจับจ่ายซื้อของ ใช้เป็นสถานที่นัดพบ/พักผ่อน หรือหาอาหารรับประทาน ร้านขายสินค้าออนไลน์อยากรู้ว่าผู้คนกำลังให้ความสนใจในสินค้าบริการประเภทใดอยู่ แหล่งข้อมูลออนไลน์ใดที่มีอิทธิพลต่อการตัดสินใจเลือกซื้อสินค้า เป็นต้น

2.สร้างฐานข้อมูลให้ครบ (Data understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมด หรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้ ในอดีตการศึกษาหาแนวโน้มความต้องการตลาด หรือพฤติกรรมผู้บริโภคในการตัดสินใจซื้อสินค้าเป็นเรื่องที่ยุ่งยากและต้องว่าจ้างบริษัทวิจัยสำรวจภาพรวม ควบคู่กับการพิจารณารายการสั่งซื้อสินค้าที่เก็บไว้ในฐานข้อมูลของบริษัท แต่ด้วยความก้าวหน้าทางด้านเทคโนโลยีในปัจจุบันและการทำธุรกรรมผ่านเครือข่ายอินเทอร์เน็ต ทำให้ข้อมูลมากมายมหาศาลวิ่งผ่านไปมาอยู่ในระบบเว็บไซต์ หรือแอพที่เป็นช่องทางในการทำธุรกรรมต่างๆ จึงเป็นแหล่งข้อมูลสำคัญ อีกทั้งยังได้ข้อมูลความสนใจของคนที่พร้อมยอมให้อย่างเต็มใจจากห้องแชทต่างๆ ที่มีการพูดคุยหารือกัน ปัจจุบันการแกะรอยหรือสะกดรอยตามคนได้ดีที่สุดเกิดขึ้นได้ง่ายมากจากออนไลน์ ไม่ว่าจะเป็นพิกัดตำแหน่งที่อยู่ของเราที่อนุญาตให้แอพต่างๆ เข้าถึง

3.เตรียมข้อมูลให้พร้อมใช้ (Data preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก ระบบการรับข้อมูลป้อนเข้าสู่ระบบที่ทันสมัยในปัจจุบันจะลดการคีย์ข้อมูลจากคนให้น้อยที่สุด แต่จะใช้วิธีการสแกน การติ๊กเลือก เพื่อลดความผิดพลาดให้น้อยที่สุด เพราะขั้นตอนใช้เวลามากกว่า 50% ของเวลารวมทั้งหมด การลดข้อผิดพลาดของข้อมูลได้มากเท่าใดก็จะยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น

4.จัดทำและเลือกโมเดลที่ใช่ (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์และสถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่างๆ ผสมผสานกัน อาทิ การจำแนก (Classification) การแบ่งกลุ่ม (Clustering) และการสร้างความสัมพันธ์ (Association rule) ในร้านสะดวกซื้อ จะนำข้อมูลการซื้อสินค้าของลูกค้าแต่ละรายมาหาความสัมพันธ์ เช่น คนที่ซื้อเครื่องดื่มแต่ละชนิดมักจะซื้อขนมหรือของกินอะไรร่วมอยู่ด้วย การ

ใช้จ่ายของแต่ละคนจะอยู่ที่ประมาณกี่บาท คนส่วนใหญ่ที่เข้ามาจะซื้อสินค้ากี่ชิ้นต่อคน และเพื่อให้ทราบข้อมูลของผู้ซื้อ ร้านค้ามักจะใช้การออกบัตรเติมเงินที่จูงใจให้ใช้จากส่วนลดหรือสะสมแต้ม ทำให้สามารถติดตามประวัติการใช้จ่ายได้ง่ายขึ้น ซึ่งปัจจุบันมีการนำกล้องจับภาพผู้ซื้อในการแยกแยะเพศ อายุ และไลฟ์สไตล์ของคน

5.ประเมินผลก่อนตัดสินใจ (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ทั้งนี้อาจต้องกลับไปทบทวนขั้นตอนที่ 2 – 4 ซ้ำอีกครั้ง ในกรณีที่ผลลัพธ์ไม่มีความน่าเชื่อถือเพียงพอ หรือไม่สอดรับกับวัตถุประสงค์

6.เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อมให้ฝ่ายต่างๆ นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่างๆ ในทางธุรกิจต่อไป

2.3.2 แบบจำลองอนุกรมเวลา (Time Series)

การพยากรณ์ ในการพยากรณ์อนุกรมเวลาแบบฉบับ ส่วนประกอบที่สำคัญที่มี อิทธิพลต่อการพยากรณ์ คือ ส่วนประกอบของการเปลี่ยนแปลงตามฤดูกาลแนวโน้ม และการเปลี่ยนแปลงตามวัฏจักร ส่วนการเปลี่ยนแปลงที่ผิดปกตินั้นเราไม่สามารถพยากรณ์ได้ เนื่องจากเป็นตัวแปรสุ่มมีรูปแบบที่ไม่แน่นอน ทำให้ไม่อาจคาดการณ์ได้ล่วงหน้าว่าจะเกิดอะไรขึ้น ณ เวลาใด และรุนแรงเพียงใด ดังนั้น ค่าพยากรณ์ Y ในหน่วยเวลา t จะหาค่าได้ดังนี้

$$Y = TxSxCxI$$

โดยที่ Y คือ ข้อมูลอนุกรมเวลา

T คือ คือค่าประมาณของส่วนประกอบแนวโน้มของหน่วยเวลา (Trend)

S คือค่าประมาณของส่วนประกอบการผันแปรตามฤดูกาล (Seasonal Variation)

C คือ การผันแปรตามวัฏจักร (Cyclical Variation)

I คือ การผันแปรผิดปกติ (Irregular Variation)

ในการพยากรณ์ค่าส่วนประกอบแนวโน้มและส่วนประกอบการเปลี่ยนแปลงตาม ฤดูกาลกระทำได้ไม่ยากนัก แต่การพยากรณ์ค่าของส่วนประกอบการเปลี่ยนแปลงตามวัฏจักร นั้นไม่ใช่สิ่งที่ง่ายนัก เพราะการเปลี่ยนแปลงตามวัฏจักรไม่มีรูปแบบ และความยาวของการเปลี่ยนแปลงตามวัฏจักรไม่แน่นอน ในการพยากรณ์การเปลี่ยนแปลงตามวัฏจักรในระยะสั้น หรืออาจใช้ความคิดเห็นของผู้เชี่ยวชาญในธุรกิจนั้น ๆ เช่น การประมาณอัตราการขยายตัวของธุรกิจโดยใช้ค่าเฉลี่ยของความสูง และความยาวของการเปลี่ยนแปลงตามวัฏจักรในอดีตเป็นต้น แต่ในทางปฏิบัติก็ไม่ใช่สิ่งที่จะทำได้โดยง่าย และการเปลี่ยนแปลงตามวัฏจักรในระยะสั้นก็มีความสำคัญต่อความแม่นยำของค่าพยากรณ์มากพอสมควร ดังนั้นการประมาณการเปลี่ยนแปลงตามวัฏจักรโดยวิธีการนำเสนอไปในหัวข้อ 4 เป็นแนวทางหนึ่งที่อาจใช้ประมาณค่าการเปลี่ยนแปลงตามวัฏจักรได้อย่างเป็นระบบ นอกเหนือไปจากวิธีหาค่าเฉลี่ยแบบเคลื่อนที่ ค่าประมาณของส่วนประกอบการเปลี่ยนแปลงตามวัฏจักรที่จะใช้พยากรณ์ในฤดูกาลหน้าจะเป็นค่าในฤดูกาลปัจจุบัน

หากข้อมูลอนุกรมเวลาที่ใช้ในการศึกษาครั้งนี้ ไม่พบการผันแปรตามวัฏจักร และการผันแปรผิดปกติของข้อมูลที่ชัดเจน ดังนั้นการพยากรณ์จึงมีเฉพาะองค์ประกอบที่เป็นค่าแนวโน้มและองค์ประกอบที่แสดงค่าการผันแปรตามฤดูกาล ดังนี้

$$Y = T \times S$$

การวัดความแม่นยำของตัวแบบที่นำมาใช้พยากรณ์ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Percentage Error : MAPE) สามารถคำนวณได้จากสูตร

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{e_t}{y_t} \right|}{n} \times 100$$

ภาพที่ 2.4 สูตรการหาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์

ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ยิ่งน้อย หมายถึง การพยากรณ์ยิ่งมีความแม่นยำ

2.3.3 การแสดงผลแบบ visualization

Tableau Public เป็นซอฟต์แวร์ที่สามารถนำข้อมูลจำนวนมาก หรือมีความหลากหลายจากแหล่งข้อมูลในองค์กรมาทำการวิเคราะห์เพื่อเป็นข้อมูลเชิงธุรกิจช่วยให้

ผู้บริหารมีข้อมูลเชิงลึกเพิ่มความสามารถในการตัดสินใจอย่างรวดเร็วและชาญฉลาด ในรูปแบบของ Data Visualization ซึ่งเป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ไม่ว่าจะเป็นตัวเลข แผนภูมิ กราฟ และอื่น ๆ อีกมากมาย คำว่า Data คือ ข้อมูล ส่วน Visualization คือ การมองเห็นเมื่อนำมารวมกันแล้ว (Data Visualization) หมายถึง ข้อมูลที่มองเห็นได้ด้วยตา

โปรแกรม Tableau แบ่งเป็น 3 ประเภทหลักๆ

Tableau Public เป็นเวอร์ชันที่เปิดให้ใช้งานได้โดยไม่มีค่าใช้จ่าย ดังนั้น คุณสมบัติหลายอย่างจะถูกลดทอนลงไป เช่น ประเภทของแหล่งข้อมูล หรือ ข้อมูล ที่จะนำมาใช้ในโปรแกรมจะถูกจำกัดลง และจะมีข้อจำกัดในเรื่องบันทึกไฟล์ที่สร้างขึ้นจากโปรแกรม คือ ไม่สามารถที่จะบันทึกไฟล์ไว้ในเครื่องที่ใช้งานได้แต่สามารถที่จะบันทึกไฟล์งานไว้ที่ Tableau Public Server ได้

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

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

2.4 วรรณกรรมที่เกี่ยวข้อง

บุษราภรณ์ โพธิ์ขวัญยืน รวิวงศ์ ศรีทองรุ่ง และ ภูกิจ ยลชญาวงศ์ (2017) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษา (1) พฤติกรรมการตัดสินใจซื้อเครื่องมือการแพทย์ของโรงพยาบาลเอกชนในเขตกรุงเทพมหานคร (2) ปัจจัยส่วนประสมทางการตลาดบริการของโรงพยาบาลเอกชนในเขตกรุงเทพมหานคร (3) ความสัมพันธ์ระหว่างส่วนประสมทางการตลาดบริการและพฤติกรรมการตัดสินใจซื้อเครื่องมือแพทย์ของโรงพยาบาลเอกชนในเขตกรุงเทพมหานคร กลุ่มตัวอย่างคือ ผู้มีอำนาจสั่งซื้อเครื่องมือแพทย์ได้แก่ นักเทคนิคการแพทย์ เจ้าพนักงานวิทยาศาสตร์การแพทย์ นักวิทยาศาสตร์ ผู้ใช้งานเครื่องของโรงพยาบาลเอกชนในเขตกรุงเทพมหานคร จำนวน 400 คน เครื่องมือที่ใช้ในการศึกษา คือ แบบสอบถามสถิติที่ใช้ในการ

วิเคราะห์ข้อมูล ไดแก่ สถิติร้อยละ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตราฐาน การทดสอบสมมติฐาน การเปรียบเทียบค่าเฉลี่ยด้วยสถิติ t-test , F-test และ LSD.

นลินี ชัยสุโรจน์ และ ผุสดี พลสารัมย์ (2562) การศึกษาค้นคว้าด้วยตนเอง เรื่อง การเพิ่มยอดขายการจำหน่ายเครื่องดื่ม ร้านชอบชา สาขาตลาดแจ่มแจ๋วมีวัตถุประสงค์ เพื่อวิเคราะห์สาเหตุของยอดขายเครื่องดื่มที่ลดลงและศึกษาแนวทางการเพิ่มยอดขาย ในการศึกษานี้ผู้ศึกษาได้ออกแบบการวิจัยเชิงสำรวจที่เป็นการศึกษาค้นคว้าด้วยตนเองโดยใช้เครื่องมือวิจัยทั้งเชิงปริมาณและเชิงคุณภาพ โดยดำเนินการตามกรอบของแนวคิดการแบ่งส่วนตลาด การกำหนดตลาดเป้าหมาย และการวางตำแหน่งผลิตภัณฑ์ (STP) การวิเคราะห์ปัจจัยแวดล้อมที่มีผลทางธุรกิจ ส่วนประสมทางการตลาด (Marketing Mix) การเทียบเคียงสมรรถนะกับคู่แข่ง การวิเคราะห์สถานการณ์(SWOT)

อนัญญา จำปาทอง (2559) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความสัมพันธ์ระหว่างปัจจัยส่วนบุคคล ปัจจัยภายในประเทศ และปัจจัยภายนอกประเทศที่มีผลต่อกระบวนการตัดสินใจนำเข้าเครื่องมือและอุปกรณ์ทางการแพทย์ของผู้ประกอบการในเขตกรุงเทพมหานครและปริมณฑล กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ ผู้ประกอบการนำเข้าเครื่องมือ และอุปกรณ์ทางการแพทย์ในเขตจังหวัดกรุงเทพมหานครและปริมณฑล โดยใช้แบบสอบถามในการเก็บข้อมูล จำนวน 350 ชุด สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ ค่าความถี่ ค่าร้อยละ ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตราฐานและทดสอบสมมติฐานด้วย One-way ANOVA, Simple Linear Regression, Multiple Linear Regression ที่ระดับนัยสำคัญทางสถิติ 0.05

ณัฐวดี หงษ์บุญมี และ ประภาสิริ ตรีพาณิชกุล (2019) งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลด้วยอัลกอริทึมเหมืองข้อมูลสามแบบคือ โครงข่ายประสาทเทียม การเรียนรู้แบบเบย์และต้นไม้ตัดสินใจ เพื่อให้ได้อัลกอริทึมที่มีประสิทธิภาพสูงสุดที่จะถูกนำมาวิเคราะห์หาปัจจัยที่ส่งผลต่อความเสี่ยงการเกิดโรคไฮเปอร์ไทรอยด์โดยการลดการนำเข้าทีละปัจจัย ซึ่งข้อมูลที่นำมาใช้ในการทดลองเป็นข้อมูลจากโรงพยาบาลในจังหวัดพิษณุโลกจำนวน 323 ชุดข้อมูล ข้อมูลสำหรับการวิเคราะห์มีจำนวน 12 ปัจจัย ผลการเปรียบเทียบพบว่าการจำแนกข้อมูลโดยใช้โครงข่ายประสามเทียมให้ค่า

ประสิทธิภาพความถูกต้อง 82.97% ซึ่งมากกว่าต้นไม้ตัดสินใจและการเรียนรู้แบบเบย์ที่มีค่าประสิทธิภาพความถูกต้อง 79.87% และ 68.11% ตามลำดับ

ศุภามณ จันทร์สกุล (2561) งานวิจัยนี้มีวัตถุประสงค์โดยการใช้เทคนิคเหมืองข้อมูลในการวิเคราะห์ข้อมูลทางการพยาบาล ซึ่งปัจจุบันมีข้อมูลต่างๆ และฐานข้อมูลที่เกี่ยวข้องในวิชาชีพการพยาบาลเป็นจำนวนมากแต่การนำข้อมูลต่างๆและฐานข้อมูลเหล่านั้นมาใช้ประโยชน์โดยค้นหารูปแบบความสัมพันธ์ของข้อมูลซึ่งเรียกว่า “การวิเคราะห์เหมืองข้อมูล” ยังมีอยู่น้อย วัตถุประสงค์ของบทความวิชาการเพื่อนำเสนอความรู้เบื้องต้นในการทำเหมืองข้อมูลสำหรับพยาบาลการวิเคราะห์เหมืองข้อมูลแบ่งออกเป็นสองกลุ่มใหญ่คือ (1) การวิเคราะห์เหมืองข้อมูลเพื่อการทำนายมีการเรียนรู้แบบมีการสอน และ (2) การวิเคราะห์เหมืองข้อมูลเพื่อการอธิบายมีการเรียนรู้แบบไม่มีการสอน เทคนิคเหมืองข้อมูล 3 เทคนิคที่ได้รับความนิยมได้แก่ การจำแนกประเภท (ต้นไม้ตัดสินใจ และเครือข่ายใยประสาท) การจัดกลุ่มข้อมูล และการค้นหากฎความสัมพันธ์ของข้อมูล ผลการวิเคราะห์ที่ได้รับจากเหมืองข้อมูลสร้างองค์ความรู้และเป็นข้อค้นพบที่เป็นประโยชน์ในการนำไปใช้เพื่อการตัดสินใจ

2.5 บทสรุป

จากการศึกษาแนวคิด ทฤษฎี และวรรณกรรมที่เกี่ยวข้องเป็นการกล่าวถึงการวิเคราะห์ข้อมูล รวมไปถึงโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูล วิธีการนำเสนอแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เพื่อแสดงผลลัพธ์จากการประมวล เพื่อนำไปแสดงผลบนเว็บไซต์