บทที่ 2

แนวคิด ทฤษฎี เครื่องมือและวรรณกรรม

โครงการเรื่องการพัฒนาแบบจำลองการพยากรณ์แรงงานไทยในต่างประเทศด้วยเทคนิคเหมืองข้อมูล ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือ และวรรณกรรมที่เกี่ยวข้อง ของการพัฒนาแบบจำลองการพยากรณ์แรงงานไทยในต่างประเทศด้วยเทคนิคเหมืองข้อมูล ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง เพื่อใช้เป็นแนวทางการศึกษา ประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับแรงงานไทยในต่างประเทศ (Concept of Thai Labour Migration Abroad)

2.1.2 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analysis Concept)

2.1.3 แนวคิดเกี่ยวกับการพยากรณ์ (Forecasting Concept)

2.1.4 แนวคิดเกี่ยวกับการทำเหมืองข้อมูล (Data Mining Concept)

2.1.5 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization Concept)

2.2 ทฤษฎี

2.2.1 ทฤษฎีอนุกรมเวลา (Time Series Theory)

2.2.2 ทฤษฎีการทำเหมืองข้อมูล (Data Mining Theory)

2.2.3 ทฤษฎีการพยากรณ์ (Forecasting Theory)

2.2.4 ทฤษฎีเกี่ยวกับเทคนิค การถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression)

2.2.5 ทฤษฎีเกี่ยวกับเทคนิค Gradient Boosted Trees (GBT)

2.2.6 ทฤษฎีเกี่ยวกกับเทคนิค Random Forest Regression

2.2.7 ทฤษฎีเกี่ยวกับเทคนิค k–Nearest Neighbors (kNN)

2.2.8 ทฤษฎี CRISP–DM (Cross Industry Standard Process for Data Mining)

2.2.9 ทฤษฎีการประเมินผล (Evaluation)

2.2.10 ทฤษฎีการแสดงผลข้อมูล (Data Visualization Theory)

2.3 วรรณกรรมที่เกี่ยวข้อง

2.4 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับแรงงานไทยในต่างประเทศ (Concept of Thai Labour Migration Abroad) แรงงานถือเป็นทรัพยากรมนุษย์ที่สำคัญในการขับเคลื่อนระบบเศรษฐกิจและสังคม การเคลื่อนย้ายแรงงานจากประเทศหนึ่งไปยังอีกประเทศหนึ่ง หรือที่เรียกว่า “แรงงานข้ามชาติ (Migrant Workers)” เป็นปรากฏการณ์ที่เกิดขึ้นทั่วโลกมาอย่างต่อเนื่อง ปัจจัยสำคัญที่ทำให้เกิดการเคลื่อนย้ายแรงงานสามารถอธิบายได้ด้วย ทฤษฎีแรงงานข้ามชาติ เช่น ปัจจัยผลัก (Push Factors) ได้แก่ รายได้ต่ำ การขาดแคลนงานในประเทศต้นทาง ภาวะเศรษฐกิจตกต่ำ และปัจจัยดึง (Pull Factors) ได้แก่ โอกาสการทำงาน รายได้ที่สูงกว่า และสวัสดิการในประเทศปลายทาง (Todaro, 1976)

สำหรับประเทศไทย การส่งแรงงานไทยไปทำงานต่างประเทศมีความสำคัญต่อเศรษฐกิจภายในประเทศ ทั้งในด้านการลดแรงกดดันการว่างงาน การพัฒนาทักษะฝีมือแรงงาน และโดยเฉพาะอย่างยิ่งคือ เงินโอนกลับประเทศ (Remittances) ซึ่งเป็นแหล่งรายได้สำคัญที่ช่วยเพิ่มการหมุนเวียนทางเศรษฐกิจ (ธนาคารแห่งประเทศไทย, 2565)

แนวคิดทางเศรษฐศาสตร์แรงงาน เช่น ทฤษฎีทุนมนุษย์ (Human Capital Theory) ของ Becker (1964) อธิบายว่า แรงงานเปรียบเสมือนการลงทุนในทุนมนุษย์ ซึ่งเมื่อแรงงานย้ายถิ่นฐานไปทำงานในประเทศที่มีค่าแรงสูงกว่า ย่อมได้รับผลตอบแทนที่คุ้มค่ากว่า ขณะที่แนวคิด Dual Labour Market Theory ของ Piore (1979) ชี้ว่าตลาดแรงงานในประเทศพัฒนาแล้วมี “งานที่ไม่เป็นที่นิยม” ซึ่งต้องอาศัยแรงงานข้ามชาติเข้ามาเติมเต็ม

ในบริบทของแรงงานไทย การเคลื่อนย้ายแรงงานไปต่างประเทศมีลักษณะเป็น วงจรซ้ำ (Cyclical) โดยบางช่วงมีแรงงานออกไปมากกว่าปกติ เช่น ในช่วงที่เศรษฐกิจไทยชะลอตัว หรือในช่วงที่ประเทศปลายทางเปิดรับแรงงานจำนวนมาก ปรากฏการณ์นี้สะท้อนว่าข้อมูลแรงงานไทยในต่างประเทศมี แนวโน้ม (Trend) และ ฤดูกาล (Seasonality) ที่สามารถนำมาวิเคราะห์และพยากรณ์ได้

2.1.2 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analysis Concept) คือ กระบวนการจัดระเบียบข้อมูลให้อยู่ในรูปที่เข้าใจง่ายและเหมาะสมต่อการตีความ โดยอาจใช้วิธีการทางสถิติหรือการอธิบายเชิงตรรกะ เพื่อให้ได้ข้อสรุปที่เชื่อถือได้และสอดคล้องกับวัตถุประสงค์ของ

การวิจัย การวิเคราะห์ข้อมูลถือเป็นขั้นตอนสำคัญในงานวิจัยทุกประเภท เพราะช่วยให้ข้อมูลดิบที่ได้จากการเก็บรวบรวมมีความหมาย สามารถสรุป อธิบาย และนำไปใช้ประโยชน์ได้อย่างเหมาะสม ทั้งยังเป็นเครื่องมือในการตรวจสอบสมมติฐาน ตอบคำถามวิจัย และค้นหาความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลประเภทของการวิเคราะห์ข้อมูล บุญชม ศรีสะอาด (2553) ได้อธิบายว่าการวิเคราะห์ข้อมูลสามารถจำแนกได้หลายลักษณะ ขึ้นอยู่กับวัตถุประสงค์ของผู้วิจัยและลักษณะของข้อมูลที่นำมาวิเคราะห์ ซึ่งสามารถแบ่งออกได้ดังนี้

1) **การวิเคราะห์เชิงพรรณนา (Descriptive Analysis)** การวิเคราะห์เชิงพรรณนาเป็นการสรุปและอธิบายลักษณะทั่วไปของข้อมูลที่เก็บรวบรวมมา เพื่อให้เห็นภาพรวมของข้อมูลได้อย่างชัดเจน ตัวอย่างเช่น การหาค่ากลางของข้อมูล ได้แก่ ค่าเฉลี่ย (Mean) ค่ามัธยฐาน (Median) และฐานนิยม (Mode) รวมถึงการวัดการกระจายของข้อมูล เช่น ค่าพิสัย (Range) ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) และค่าความแปรปรวน (Variance) นอกจากนี้ยังรวมถึงการหาค่าร้อยละ ความถี่ และการสร้างตารางแจกแจงความถี่ ซึ่งช่วยให้ผู้วิจัยเข้าใจข้อมูลเบื้องต้นได้ง่ายขึ้น การวิเคราะห์ประเภทนี้มักใช้เป็นขั้นตอนแรกของการวิจัยเพื่อทำความเข้าใจข้อมูลก่อนเข้าสู่การวิเคราะห์เชิงลึก

2) **การวิเคราะห์เชิงอนุมาน (Inferential Analysis)** การวิเคราะห์เชิงอนุมานเป็นการใช้ข้อมูลจากกลุ่มตัวอย่างเพื่อสรุปไปยังประชากร โดยอาศัยทฤษฎีความน่าจะเป็นและสถิติอนุมาน ตัวอย่างของการวิเคราะห์ประเภทนี้ เช่น การทดสอบสมมติฐาน (Hypothesis Testing) การทดสอบค่าเฉลี่ยของประชากร (t-test) การวิเคราะห์ความแปรปรวน (ANOVA) การวิเคราะห์การถดถอย (Regression Analysis) และการหาสหสัมพันธ์ (Correlation) จุดประสงค์สำคัญของการวิเคราะห์เชิงอนุมาน คือการตรวจสอบว่าสมมติฐานที่ตั้งไว้สามารถยืนยันได้หรือไม่ รวมถึงการประมาณค่าของประชากรจากข้อมูลของกลุ่มตัวอย่าง ซึ่งช่วยเพิ่มความน่าเชื่อถือให้กับผลการวิจัย

3) **การวิเคราะห์เชิงพยากรณ์ (Predictive Analysis)** การวิเคราะห์เชิงพยากรณ์เป็นการนำข้อมูลในอดีตมาใช้สร้างแบบจำลองทางสถิติหรือคณิตศาสตร์ เพื่อคาดการณ์ค่าของข้อมูลที่จะเกิดขึ้นในอนาคต การวิเคราะห์ประเภทนี้นิยมใช้ในงานที่เกี่ยวข้องกับการวางแผนเชิงกลยุทธ์ เช่น การคาดการณ์แนวโน้มเศรษฐกิจ การคาดการณ์ยอดขาย หรือการคาดการณ์จำนวนแรงงานที่เดินทางไปทำงานต่างประเทศ วิธีการที่ใช้มักเป็นการวิเคราะห์อนุกรมเวลา (Time Series Analysis) การสร้างสมการถดถอยเชิงเส้น หรือแม้แต่การประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งช่วยเพิ่มความแม่นยำในการทำนาย

4) การวิเคราะห์เชิงอธิบายและสรุปผล (Interpretive Analysis) การวิเคราะห์ประเภทนี้มุ่งเน้นไปที่การตีความและเชื่อมโยงข้อมูลเข้ากับบริบทของการวิจัย เพื่อให้ได้ข้อสรุปที่มีความหมายและสอดคล้องกับวัตถุประสงค์ของงาน ตัวอย่างเช่น การอธิบายว่าทำไมจำนวนแรงงานที่เดินทางไปทำงานต่างประเทศจึงมีแนวโน้มเพิ่มขึ้นหรือลดลงในบางช่วงเวลา การวิเคราะห์เชิงอธิบายจึงไม่เพียงแต่ใช้ค่าทางสถิติมาแสดง แต่ยังรวมถึงการตีความเชิงเหตุผลเพื่อตอบคำถามการวิจัยด้วย

2.1.3 แนวคิดเกี่ยวกับการพยากรณ์ (Forecasting Concept) หมายถึง กระบวนการคาดคะเนหรือตีความแนวโน้มของข้อมูลที่จะเกิดขึ้นในอนาคต โดยอาศัยข้อมูลในอดีตและปัจจุบันเป็นพื้นฐาน การพยากรณ์จึงเป็นเครื่องมือสำคัญสำหรับการวางแผนและการตัดสินใจทั้งในระดับองค์กรและระดับประเทศ เนื่องจากสามารถช่วยให้หน่วยงานหรือผู้เกี่ยวข้องเตรียมความพร้อมต่อการเปลี่ยนแปลงที่อาจเกิดขึ้นได้อย่างเหมาะสม (เกรียงศักดิ์ เจริญวงศ์ศักดิ์, 2552) การพยากรณ์สามารถแบ่งออกได้หลายประเภท ได้แก่

1) การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting) ใช้การประเมินจากผู้เชี่ยวชาญหรือข้อมูลเชิงความคิดเห็น เหมาะกับกรณีที่ไม่มีข้อมูลเชิงปริมาณเพียงพอ ตัวอย่างเช่น การใช้เทคนิค Delphi เพื่อรวบรวมความเห็นของผู้เชี่ยวชาญหลายฝ่าย

2) การพยากรณ์เชิงปริมาณ (Quantitative Forecasting) ใช้ข้อมูลเชิงตัวเลขที่เก็บรวบรวมมาอย่างเป็นระบบเพื่อนำมาสร้างแบบจำลองทางสถิติ เช่น การใช้ค่าเฉลี่ยเคลื่อนที่ (Moving Average) วิธีการปรับเรียบแบบเอ็กซ์โพเนนเชียล (Exponential Smoothing) และแบบจำลองอนุกรมเวลา (Time Series Models) (สุชาติ พงษ์ประกอบ, 2556)

3) การพยากรณ์ตามช่วงเวลา เช่น ระยะสั้น ระยะกลาง และระยะยาว โดยขึ้นอยู่กับเป้าหมายและลักษณะการใช้งาน

ความสำคัญของการพยากรณ์การพยากรณ์มีความสำคัญในหลายด้าน เช่น **ด้านเศรษฐกิจ**การคาดการณ์อัตราการเติบโตทางเศรษฐกิจ การว่างงาน และอัตราเงินเฟ้อ **ด้านการบริหารจัดการ** การวางแผนการผลิตและการจัดสรรทรัพยากร **ด้านแรงงาน** การคาดการณ์จำนวนแรงงานไทยที่จะเดินทางไปทำงานต่างประเทศ เพื่อสนับสนุนการวางแผนนโยบายแรงงานและการคุ้มครองแรงงาน

2.1.4 แนวคิดเกี่ยวกับการทำเหมืองข้อมูล (Data Mining Concept) การทำเหมืองข้อมูล (Data Mining) หมายถึง กระบวนการสกัดหาความรู้ (Knowledge Discovery) จากข้อมูลจำนวนมาก ด้วยการใช้วิธีการทางสถิติ คณิตศาสตร์ และเทคนิคการเรียนรู้ของเครื่อง เพื่อค้นหารูปแบบ ความสัมพันธ์ และแนวโน้มที่ซ่อนอยู่ในข้อมูล ข้อมูลที่ผ่านการวิเคราะห์แล้วสามารถ

นำไปใช้ประกอบการตัดสินใจ การวางแผนเชิงกลยุทธ์ และการพยากรณ์อนาคตได้อย่างมีประสิทธิภาพ (เกษม จันทร์แก้ว, 2557) กระบวนการทำเหมืองข้อมูลตามแนวคิดของนักวิชาการไทย สามารถอธิบายได้เป็นขั้นตอนดังนี้

1) **การทำความเข้าใจปัญหา (Problem Understanding)** ระบุเป้าหมายของการวิเคราะห์ เช่น การพยากรณ์จำนวนแรงงานไทยที่ไปทำงานต่างประเทศ

2) **การทำความเข้าใจข้อมูล (Data Understanding)** ศึกษาลักษณะของข้อมูล เช่น ประเภท ขอบเขต ความครบถ้วน และคุณภาพ

3) **การเตรียมข้อมูล (Data Preparation)** จัดการข้อมูลให้อยู่ในรูปแบบที่เหมาะสม เช่น การลบค่าที่หายไป การจัดการค่าผิดปกติ และการสร้างตัวแปรใหม่

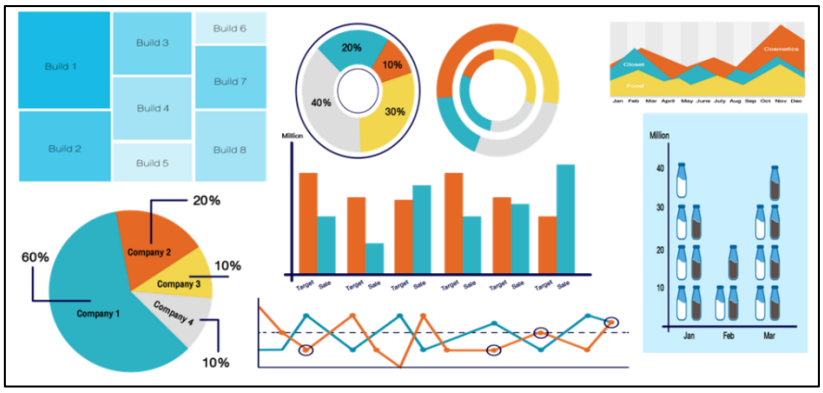
4) **การสร้างแบบจำลอง (Modeling)** ใช้เทคนิคการทำเหมืองข้อมูล เช่น การจำแนก (Classification) การถดถอย (Regression) หรือการจัดกลุ่ม (Clustering) เพื่อหาความรู้จากข้อมูล

5) **การประเมินผล (Evaluation)** ตรวจสอบประสิทธิภาพของแบบจำลองด้วยเกณฑ์มาตรฐาน เช่น Accuracy, RMSE หรือ MAPE

6) **การนำไปประยุกต์ใช้ (Deployment):** นำผลลัพธ์ที่ได้ไปใช้จริงในการตัดสินใจหรือนำเสนอในรูปแบบที่ผู้ใช้เข้าใจง่าย

2.1.5 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization Concept) การแสดงผลข้อมูล (Data Visualization) หมายถึง การนำข้อมูลที่มีอยู่ โดยเฉพาะข้อมูลเชิงปริมาณและข้อมูลที่มีความซับซ้อน มาถ่ายทอดในรูปแบบที่เข้าใจง่ายและสื่อความหมายได้ชัดเจน เช่น ตาราง แผนภูมิ กราฟ หรือแผงควบคุม (Dashboard) จุดมุ่งหมายสำคัญคือช่วยให้ผู้ใช้สามารถตีความข้อมูลได้รวดเร็ว เห็นแนวโน้ม ความสัมพันธ์ และความแตกต่างของข้อมูลในเชิงเปรียบเทียบได้อย่างชัดเจน (พรพิมล กองทิพย์, 2559)

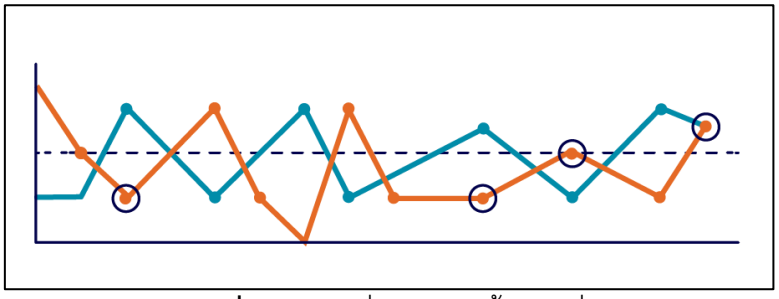
1) แผนภูมิ (Chart) แผนภูมิสามารถแสดงภาพข้อมูลประเภทใดก็ได้เพื่อการวิเคราะห์และตีความอย่างมีประสิทธิภาพ แผนภูมิมีหลากหลายชนิด เช่น แผนภูมิพื้นที่ แผนภูมิเส้น แผนภูมิแท่ง แผนภูมิซ้อน แผนภูมิวงกลม แผนภูมิกระจาย แผนภูมิผสม แผนภูมิกรวย แผนภูมิใยแมงมุม ฯลฯ แผนภูมิแต่ละแผนภูมิช่วยให้มีตัวเลือกการโต้ตอบกับผู้ใช้ที่หลากหลายสำหรับการวิเคราะห์เชิงลึกและเชิงบริบท



ภาพที่ 2.1 ตัวอย่างแผนภูมิรูปแบบต่าง ๆ

(ที่มา : https://www.9experttraining.com/articles/)

2) กราฟ (Graph) โดยทั่วไปหมายถึง แผนภาพที่แสดงข้อมูลเชิงปริมาณหรือความสัมพันธ์ระหว่างข้อมูล ในรูปแบบภาพ เช่น เส้น จุด หรือแท่ง เพื่อช่วยให้เข้าใจข้อมูลได้ง่ายขึ้น ในความหมายเฉพาะทางคณิตศาสตร์และวิทยาการคอมพิวเตอร์ กราฟ หมายถึงโครงสร้างข้อมูลที่ประกอบด้วย "จุดยอด" (vertex) และ "เส้นเชื่อม" (edge) ที่เชื่อมโยงจุดยอดเหล่านั้นเข้าด้วยกัน



ภาพที่ 2.2 ตัวอย่างกราฟเส้นตัวอย่าง

(ที่มา : https://www.9experttraining.com/articles/)

3) ตาราง (Table) เป็นรูปแบบการนำเสนอข้อมูลที่ประกอบด้วย Row หรือแถวแนวนอน และ Column หรือแถวแนวนอน เป็นการนำเสนอข้อมูลในการเปรียบเทียบตัวแปรหลากหลายตัว ใช้นำเสนอข้อมูลจำนวนมากแบบมีโครงสร้างได้

ภาค	ทั้งหมด			ถูกกฎหมาย			ผิดกฎหมาย		
	รวม	ชาย	หญิง	รวม	ชาย	หญิง	รวม	ชาย	หญิง
กรุงเทพฯ และปริมณฑล	344,334	211,331	133,003	286,587	180,800	105,787	57,747	30,531	27,216
ภาคกลาง	328,384	194,869	133,515	263,061	156,976	106,085	65,323	37,893	27,430
ภาคเหนือ	139,958	73,709	66,249	105,662	55,773	49,889	34,296	17,936	16,360
ภาคตะวันออกเฉียงเหนือ	30,502	15,865	14,637	23,328	12,861	10,467	7,174	3,004	4,170
ภาคใต้	290,673	171,621	119,052	261,893	153,801	108,092	28,780	17,820	10,960

ภาพที่ 2.3 ตัวอย่างตาราง

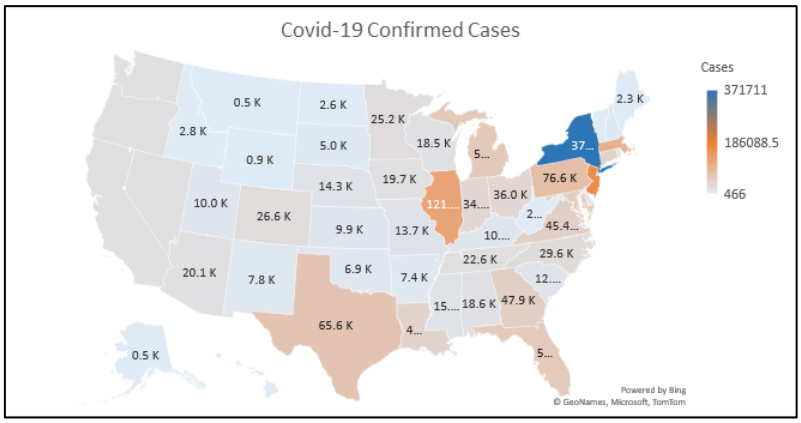
(ที่มา : https://www.m-society.go.th/ewtadmin/)

4) อินโฟกราฟิก การนำข้อมูลหรือความรู้มาสรุปเป็นสารสนเทศ ในลักษณะของกราฟิกที่ออกแบบ เป็นภาพนิ่งหรือภาพเคลื่อนไหว ดูแล้วเข้าใจง่ายในเวลารวดเร็วและชัดเจน สามารถสื่อให้ผู้ชมเข้าใจ ความหมายของข้อมูลทั้งหมดได้โดยไม่จำเป็นต้องมีผู้นำเสนอมาช่วยขยายความเข้าใจอีก

ภาพที่ 2.4 ตัวอย่างอินโทรกราฟิก

(ที่มา : https://www.kroochut.com/infographic/)

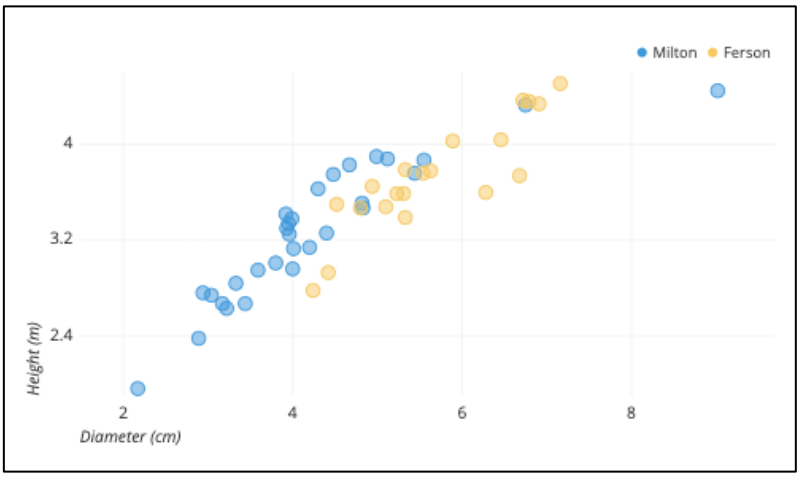
5) แผนที่ (Maps) เป็นการนำเสนอข้อมูลเชิงปริมาณและเชิงพื้นที่โดยใช้แผนที่ภูมิศาสตร์เป็นพื้นหลัง เพื่อแสดงแนวโน้ม, ความหนาแน่น, หรือการกระจายตัวของข้อมูลตามตำแหน่งทางภูมิศาสตร์ เช่น ประเทศ, จังหวัด, หรือรหัสไปรษณีย์. ข้อมูลจะถูกแสดงด้วยสี, สัญลักษณ์, หรือเฉดสีที่แตกต่างกัน เพื่อบ่งบอกถึงความแตกต่างของค่าหรือประเภทข้อมูลในแต่ละพื้นที่



ภาพที่ 2.5 ตัวอย่างแผนที่

(ที่มา https://spreadsheetweb.com/excel-map-chart/)

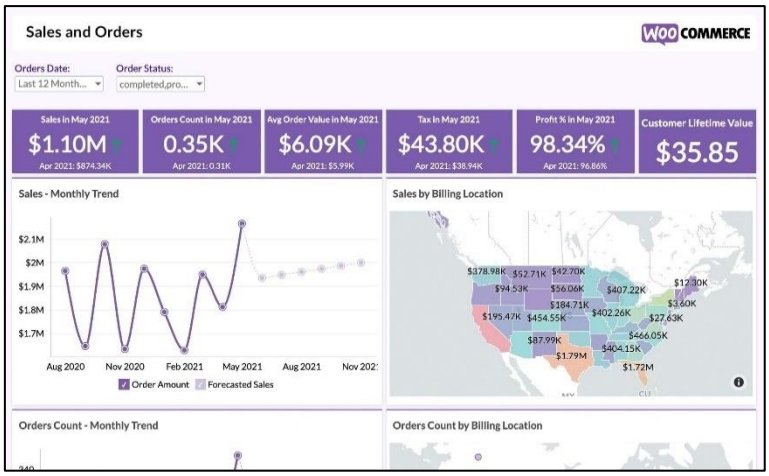
6) **Scatter Plot** เป็นการลงจุดลงบนกราฟ เอาค่าข้อมูลมา Plot หาความสัมพันธ์ บนสองตัวแปรข้อมูล จุดที่ Plot หรือลงไปในกราฟคือจุดตัดของค่าแนวแกน X และ Y จะช่วยให้เห็นความสัมพันธ์ นัยยะบางอย่างของสองตัวแปรนั้นสัมพันธ์กันยังไง



ภาพที่ 2.6 ตัวอย่าง Scatter Plot

(ที่มา : https://www.niwat.blog/graph)

7) แดชบอร์ด ช่วยให้คุณสามารถรวมรายงานหลายรายการและวิดเจ็ต KPI ในหน้าเดียวเพื่อดูภาพรวมอย่างรวดเร็ว สามารถจัดระเบียบรายงานและวิดเจ็ตได้อย่างสะดวกโดยใช้เค้าโครงที่ยืดหยุ่น นอกจากนี้ความสามารถในการกรองแบบไดนามิกยังช่วยให้สามารถดูข้อมูลที่ต้องการและข้อมูลเฉพาะได้



ภาพที่ 2.7 ตัวอย่างอินโทรกราฟิก

(ที่มา : https://www.zoho.com/th/analytics/)

2.2 ทฤษฎี

2.2.1 ทฤษฎีอนุกรมเวลา (Time Series Theory)

ทฤษฎีอนุกรมเวลาเป็นพื้นฐานสำคัญของการวิเคราะห์และการพยากรณ์ข้อมูลที่มีลำดับเชิงเวลา อนุกรมเวลาหมายถึงชุดข้อมูลที่ถูกรวบรวมและบันทึกตามช่วงเวลาต่อเนื่อง เช่น รายวัน รายเดือน หรือรายปี โดยมีองค์ประกอบหลัก ได้แก่ ระดับ (Level) แนวโน้ม (Trend) ฤดูกาล (Seasonality) และความไม่แน่นอนหรือความผันผวน (Noise) การทำความเข้าใจองค์ประกอบเหล่านี้ช่วยให้ผู้วิจัยสามารถเลือกใช้วิธีการพยากรณ์ที่เหมาะสมกับลักษณะของข้อมูลแต่ละชุดได้ (อารี เพ็ชรสม, 2554)

เนื่องจากข้อมูลทางธุรกิจมีการเปลี่ยนแปลงตลอดเวลา ผู้นำทางธุรกิจหรือองค์กรต้องหาวิธีพัฒนาต่าง ๆ ที่สามารถนำไปใช้ประกอบการตัดสินใจวางแผน เกี่ยวกับผลที่เกิดจากความเปลี่ยนแปลงในการดำเนินการอย่างใดอย่างหนึ่ง ดังนั้นการวิเคราะห์อนุกรมเวลาจึงเข้ามามีบทบาทช่วยในการตัดสินใจ เทคนิคอย่างหนึ่งที่ใช้ช่วยในการควบคุมการดำเนินการในปัจจุบันและในการวางแผนความต้องการในอนาคต คือ การพยากรณ์ **(forecasting)** ซึ่งการพยากรณ์นั้นทำได้หลายวิธี แต่ละวิธีต่างมีเป้าหมายเดียวกัน คือ ทำนายเหตุการณ์ในอนาคต

การวิเคราะห์อนุกรมเวลาผู้วิเคราะห์จะแยกองค์ประกอบต่างๆที่ประกอบกันขึ้นเป็นอนุกรมเวลา โดยจะมีการเปลี่ยนแปลงไปตามอิทธิพลต่าง ๆ เช่น การเปลี่ยนแปลงการผลิต เทคโนโลยี สภาวะอากาศ เป็นต้น ในการหาคุณลักษณะของอนุกรมเวลาเราสามารถใช้แบบจำลองได้หลายแบบ แบบจำลองที่ใช้โดยนักเศรษฐศาสตร์แบบหนึ่ง คือ แบบจำลองแบบ

คลาสสิก (classical model) เป็นการอธิบายถึงองค์ประกอบของการแปรผันของอนุกรมเวลา 4 ส่วน ดังนี้

1) ค่าแนวโน้ม (Secular trend) แทนด้วย T t

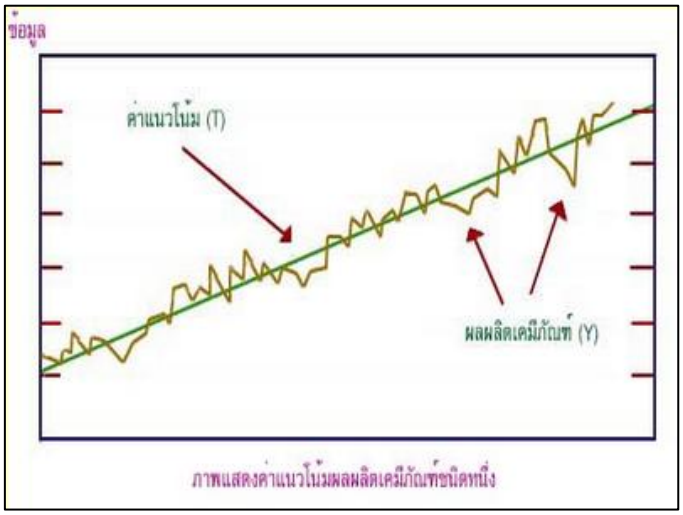
2) การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล (Seasonal Variation) แทนด้วย S t

3) การเปลี่ยนแปลงหรือความผันแปรตามวัฏจักร (Cyclical Variation) แทนด้วย C t

4) การเปลี่ยนแปลงหรือความผันแปรเนื่องจากเหตุการณ์ผิดปกติ (Irregularl Variation) แทนด้วย It

ค่าแนวโน้ม (Secular trend) แทนด้วย Tt

เป็นการเปลี่ยนแปลงข้อมูลมีลักษณะราบเรียบ แนวโน้ม อาจมีลักษณะเป็นเส้นตรงหรือเส้นโค้งในทางเพิ่มขึ้นหรือลดลง ค่าแนวโน้มของข้อมูลเป็นการเคลื่อนไหวในช่วงระยะเวลาที่ค่อนข้างนานพอสมควร ควรเป็นข้อมูลรายปี และควรมีข้อมูลอย่างน้อย 15 ปี ซึ่งจะแสดงทิศทางของอนุกรมเวลา



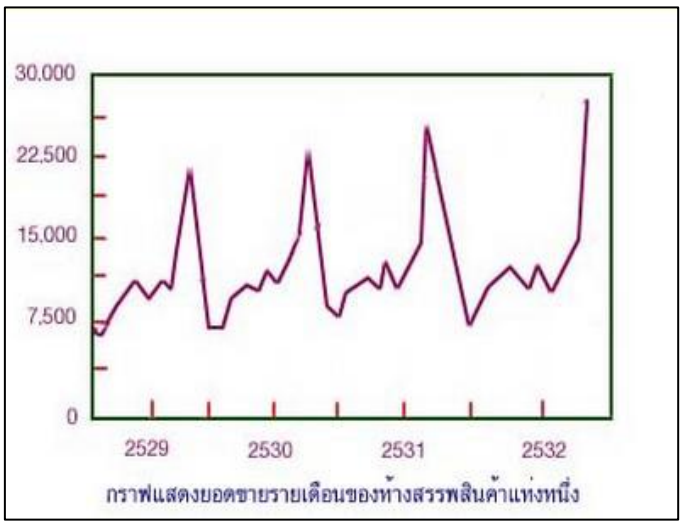
ภาพที่ 2.8 กราฟแสดงค่าแนวโน้มของผลผลิตเคมีภัณฑ์ชนิดหนึ่ง

(ที่มา : http://www.fpo.go.th/S-I/Source/ECO/ECO24.htm)

จากกราฟ Y แทนข้อมูลอนุกรมเวลาของผลผลิตเคมีภัณฑ์ชนิดหนึ่ง ในช่วงเวลา 15 ปี ค่าแนวโน้มแทนด้วยกราฟเส้นตรง ซึ่งซ้อนอยู่บนเส้นกราฟของ Y เส้นกราฟแสดงแนวโน้มนอกจากจะมีลักษณะเป็นเส้นตรงดังรูปแล้ว อาจมีลักษณะเป็นเส้นโค้ง เช่น เส้นโค้งเอ็กโพเนนเชียลหรือพาราโบลาก็ได้

การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล (Seasonal Variation) แทนด้วย St

เป็นการเปลี่ยนแปลงข้อมูลมีลักษณะการเพิ่มขึ้น หรือลดลงในลักษณะเดียวกันของรอบระยะเวลาหนึ่งที่แน่นอน เรียกว่า การเปลี่ยนแปลงตามฤดูกาล หน่วยของระยะเวลาสำหรับข้อมูลอาจเป็นรายชั่วโมง รายวัน รายสัปดาห์ รายเดือน รายไตรมาส สำหรับข้อมูลรายปีไม่มีการแปรผันตามฤดูกาล การเปลี่ยนแปลงตามฤดูกาลนั้นกำหนดระยะเวลาการเกิดซ้ำในรอบหนึ่ง ๆ ได้ค่อนข้างแน่นอน



ภาพที่ 2.9 ตัวอย่างยอดขายรายเดือนของห้างสรรพสินค้าแห่งหนึ่ง

(ที่มา : http://www.fpo.go.th/S-I/Source/ECO/ECO24.htm)

จากกราฟ จะเห็นว่ายอดขายของห้างสูงประมาณเดือนธันวาคมของทุกปี ซึ่งเป็นเทศกาลคริสต์มาส และปีใหม่ ประชาชนจึงมีการจับจ่ายใช่สอยมาก ส่วนในราวเดือนพฤษภาคมของทุกปี ยอดขายจะต่ำกว่าในเดือนอื่น ๆ ที่เป็นเช่นนี้เพราะเป็นช่วงเปิดภาคเรียน ประชาชนต้องเตรียมเงินไว้สำหรับค่าใช้จ่ายในการศึกษาของบุตรหลาน

2.2.2 ทฤษฎีการทำเหมืองข้อมูล (Data Mining Theory)

การทำเหมืองข้อมูล (Data Mining) หรือ การค้นหาองค์ความรู้ในฐานข้อมูล (Knowledge Discovery In Databases หรือ KDD) เป็นเทคนิคการคน้ หาองค์ความรู้ที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่เพื่อค้นหาแนวโน้ม รูปแบบรวม ความสัมพันธ์หรือความรู้ใหม่อื่น ๆ โดยอาศัยข้อมูลในอดีต ความรู้ที่ได้ทำให้เข้าใจ ลักษณะบางอย่างของข้อมูล ซึ่งจะช่วยให้สามารถ ทำนายแนวโน้มของข้อมูลใหม่ที่จะเกิดขึ้นในอนาคตได้ ผลที่ได้จะมีลักษณะของข้อมูลอยู่ 3 แบบคือ

ข้อมูลแบบที่ไม่ทราบมาก่อน (Unknown) ไม่มีความชัดเจนและไม่สามารถตั้งสมมติฐาน ก่อนได้ว่าควรเป็นแบบใด ข้อมูลแบบที่มีเหตุผล (Valid) และข้อมูลแบบที่สามารถนำไปใช้ได้ (Actionable)

กระบวนการทางานของการทำเหมืองข้อมูลประกอบด้วยขั้น ตอนต่าง ๆ ดังนี้

1) การทำความสะอาดข้อมูล (Data Cleaning) เป็นขั้น ตอนในการกำจัดข้อมูลที่ผิดปกติ หรือตัดรายละเอียดที่ไม่จำเป็นออกไป เพื่อแก้ไขข้อมูลที่บกพร่อง

2) การรวมข้อมูล (Data Integration) เป็นขั้นตอนการรวบรวมข้อมูลที่มาจากหลายแหล่งให้รวมกันเป็นชุดข้อมูลเดียว และมีการตรวจสอบรูปแบบของการจัดเก็บ ข้อมูลว่าเหมือนหรือต่างกันเพื่อลดความซ้ำซ้อน ของข้อมูลด้วย

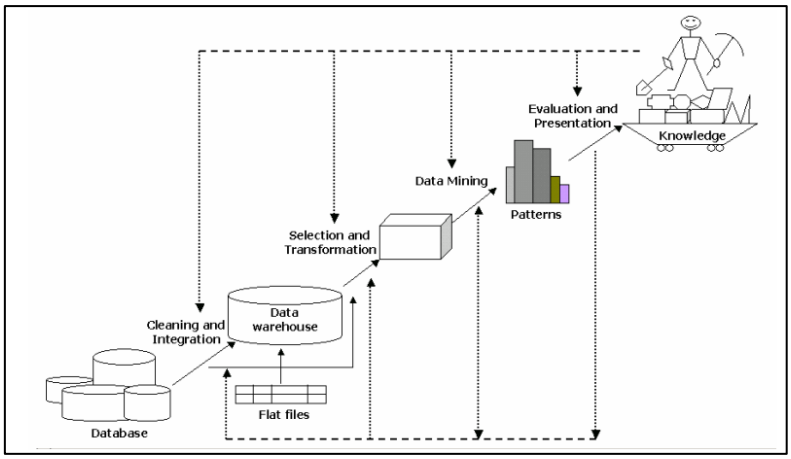
3) การคัดเลือกข้อมูล (Data Selection) เป็นขั้นตอนในการคัดเลือกข้อมูลที่เกี่ยวข้องกับ วัตถุประสงค์ที่ตั้งไว้เพียงบางส่วนจากแหล่งข้อมูลที่บันทึกไว้เพื่อให้ผลการวิเคราะห์ข้อมูลสามารถนำไปประกอบการตัดสินใจได้ตรงตามความต้องการ

4) การแปลงข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลเพื่อใช้กับ แบบจำลองต่าง ๆ ของกระบวนการทำเหมืองข้อมูล ซึ่งมีวิธีแปลงข้อมูลในรูปแบบ ต่าง ๆ เช่น การนอร์มัลไลซ์(Normalization) การดิสครีต (Discretization) การเข้ารหัส (1 Of N Coding) และการเจนเนอร์รัลไลซ์ (Generalization)

5) การทำเหมืองข้อมูล (Data mining)เป็นขั้นตอนในการค้นหารูปแบบที่เป็นประโยชน์ จากข้อมูลที่มีอยู่โดยมีเทคนิคหลายรูปแบบ เช่น การแบ่งกลุ่มขอ้ มูล (Clustering), การจำแนกประเภทข้อมูล (Classification) และ การค้นหากฎความสัมพันธ์ (Association)

6) การประเมินผล (Evaluation) เป็นขั้น ตอนสา หรับการวิเคราะห์ข้อมูลและประเมินผลที่ได้จากการทำเหมืองข้อมูล ซึ่งจะต้องเก็บผลลัพธ์ของกระบวนการทำเหมืองข้อมูลมาสรุปความหมายและตีความของผลลัพธ์ให้ออกมาเป็นข้อมูลความรู้ใหม่

7. อภิปรายผล (Knowledge Presentation) เป็นขั้น ตอนการนา เสนอความรู้ที่ได้



ภาพที่ 2.10 กระบวนการค้นหาความรู้

(ที่มา : https://archive.lib.cmu.ac.th/full/)

2.2.2.1 หลักในการทำเหมืองข้อมูลหลักในการทำเหมืองข้อมูลนั้นมี 3 หลักการทำงานหลัก ๆ คือ

1. การแบ่งกลุ่มข้อมูล(Clustering) เป็นการรวมข้อมูลที่มีลักษณะรูปแบบและ แนวโน้มที่เหมือนกนั ไว้ในกลุ่มเดียวกันโดยเริ่มต้น จากตัวแทนของกลุ่มและทำการเปรียบเทียบข้อมูลตัวที่เหลือกับตัวแทนของแต่ละกลุ่ม ถ้าข้อมูลของตัวที่นำมาเปรียบเทียบเหมือน คล้ายคลึง หรือสอดคล้องกับ ตัวแทนของกลุ่มไหนก็จะถูกจัดประเภทให้อยู่ในกลุ่มนั้น

2. การจำแนกประเภท (Classification) เป็นการจำแนกประเภทของข้อมูลโดย การเรียนรู้ข้อมูลเพื่อสร้างเป็นฐานความรู้ซึ่งแต่ละกลุ่มจะมีคุณสมบัติเฉพาะ ของตัวเองที่เรียกว่าตัวจำแนกข้อมูล (Classifier) ซึ่งสามารถใช้ในการวิเคราะห์และแยกแยะข้อมูลที่เข้ามาใหม่ได้

3. การพยากรณ์ (Prediction) เป็นการวิเคราะห์และพยากรณ์สิ่งที่สนใจหรือสิ่ง ที่ยังไม่รู้จากฐานความรู้เดิมที่ได้ ซึ่งในการศึกษานี้ได้เลือกวิธีการพัฒนาแบบจำลองการพยากรณ์แรงงานไทยในต่างประเทศ

2.2.3 ทฤษฎีการพยากรณ์ (Forecasting Theory)

2.2.3.1. การพยากรณ์ (Forecasting) หมายถึง การคาดการณ์ถึงสิ่งใดสิ่งหนึ่งที่จะเกิดขึ้นใน ช่วงเวลาในอนาคต และนำค่าพยากรณ์ที่ได้นั้นมาใช้ประโยชน์เพื่อการตัดสินใจใด ๆ การพยากรณ์นั้น มีบทบาทที่สำคัญกับทุกด้าน ทั้งหน่วยงานของรัฐบาลและเอกชน การพยากรณ์ทำหน้าที่เป็นเครื่องมือ ในการวางแผนเพื่อช่วยองค์กรหรือธุรกิจเตรียมความพร้อมในการรับมือกับความไม่แน่นอนที่อาจ เกิดขึ้นในอนาคต ช่วยในการตัดสินใจเกี่ยวกับการดำเนินงานในทุกธุรกิจและทุกสาขาอาชีพการ พยากรณ์ที่ทำกันโดยทั่วไปมีอยู่ 3 วิธี คือ

1) พยากรณ์โดยอาศัยประสบการณ์และความชำนาญ

2) พยากรณ์โดยอาศัยเหตุการณ์และหลักฐาน

3) การพยากรณ์ทางสถิติเป็นการพยากรณ์โดยใช้ข้อมูลสถิต

2.2.3.2. ช่วงเวลาที่ใช้พยากรณ์ หมายถึงช่วงเวลาในอนาคตที่ต้องการพยากรณ์ (Forecasting period) ช่วงเวลาที่กล่าวถึงนี้ อาจมีหน่วยเป็นวัน สัปดาห์ เดือน ไตรมาส หรือปี สำหรับเทคนิคการพยากรณ์ที่แตกต่างกันจะ เหมาะสมกับช่วงเวลาที่ต้องการพยากรณ์ต่างกัน ช่วงเวลาที่ใช้พยากรณ์แบ่งออกเป็น 4 ระยะ คือ (ศิริ ลักษณ์ สุวรรณวงศ์, 2535)

1) ช่วงเวลาที่สั้นมาก (Immediate – term Forecasting) เป็นการพยากรณ์ที่มีช่วงเวลา น้อยกว่าหรือเท่ากับ 1 ช่วงเวลาในอนาคตที่ต้องการพยากรณ์ เช่นถ้าข้อมูลเป็นรายไตรมาสการ พยากรณ์ในช่วงเวลาที่สั้นมากจะเป็นการพยากรณ์ข้อมูลในช่วงเวลาที่ไม่เกินไตรมาสถัดไป 20

2) ช่วงเวลาที่สั้น (Short – Term Forecasting) เป็นการพยากรณ์ในช่วงเวลา 1– 3 ช่วงเวลาในอนาคตที่ต้องการพยากรณ์

3) ช่วงเวลาระยะปานกลาง (Medium – Term Forecasting) เป็นการพยากรณ์ในช่วงเวลา ที่มากกว่า 3 เดือนจนถึง 2 ปีโดยข้อมูลเป็นรายเดือน ถ้าข้อมูลมีลักษณะเป็นอย่างอื่น เช่น ราย สัปดาห์ หรือรายไตรมาส ผู้พยากรณ์ต้องพิจารณาว่าช่วงเวลาเท่าใดเป็นระยะเวลากลาง การพยากรณ์ ช่วงเวลาระยะกลางที่พบบ่อย เช่น การวางแผนการผลิตของบริษัทแห่งหนึ่ง

4) ช่วงระยะยาว (Long – Term Forecasting) เป็นการพยากรณ์ ในช่วงเวลา 2 ปีขึ้นไป การพยากรณ์ระยะยาวใช้ในการวางแผนระยะยาว

2.2.3.3. รูปแบบของข้อมูล การเลือกวิธีการพยากรณ์จะต้องคำนึงถึงรูปแบบของข้อมูลในอดีต ซึ่งถ้าสังเกตข้อมูลอนุกรม เวลา แต่ละชุดจะมองเห็นถึงการเปลี่ยนแปลงขึ้นๆ ลงๆ ซึ่งสาเหตุของการเปลี่ยนแปลงที่เกิดขึ้นกับ ข้อมูลนั้น เนื่องจากอิทธิพลขององค์ประกอบต่าง ๆ 4 ประการ (พิภพ ลลิตาภรณ์, 2549) คือ

1) องค์ประกอบของแนวโน้ม (Trend) เป็นองค์ประกอบที่ แสดงถึงทิศทางของข้อมูล แต่ละ ชุด ตั้งแต่อดีตจนถึงระยะเวลาสุดท้ายของข้อมูลที่รวบรวมได้ซึ่งทิศทางของข้อมูลนั้นอาจจะพุ่งไปใน แนวที่สูงขึ้น หรือลดต่ำลง ข้อมูลอนุกรมเวลาที่มีองค์ประกอบของค่าแนวโน้มส่วนใหญ่จะเกี่ยวข้อง กับความเคลื่อนไหวของข้อมูลในระยะเวลา

ที่ค่อนข้างยาวนาน เช่น อุปสงค์สินค้า การใช้พลังงาน เป็น ต้น ลักษณะของแนวโน้มอาจจะเป็นเส้นตรงเส้นโค้งหรืออื่น ๆ ก็ได้

2) องค์ประกอบของการเปลี่ยนแปลงตามฤดูกาล (Seasonal) หมายถึงการที่ข้อมูล อนุกรม เวลามีรูปแบบการเคลื่อนไหวขึ้นหรือลงทำนองเดียวกันในช่วงเวลาเดียวกันของรอบเวลาหนึ่ง ซึ่งส่วน ใหญ่จะไม่เกิน 1 ปีโดยที่ หน่วยของระยะเวลาอาจจะเป็นราย 3 เดือน 5 เดือน รายเดือน รายสัปดาห์ รายวัน หรือแม้แต่รายชั่วโมงก็ได้ข้อมูลที่มักได้รับผลกระทบจากความเคลื่อนไหว หรือ เปลี่ยนแปลง ตามฤดูกาล ได้แก่การผลิต การขาย เป็นต้น สำหรับรูปแบบของดัชนีฤดูกาล โดยทั่วไปมี6 รูปแบบ ดังนี้

3) องค์ประกอบของการผันแปรตามวัฏจักร (Cyclical) เป็นลักษณะการเคลื่อนไหว ของ ข้อมูลที่ขึ้น ๆ ลง ๆ คล้ายกับลูกคลื่นที่มีผลกระทบกระเทือนต่อธุรกิจโดยทั่ว ๆ ไปรูปแบบของ การผัน แปรตามวัฏจักรนี้แตกต่างจากการผันแปรตามฤดูกาล คือเราจะไม่ทราบว่าช่วงของการเกิดวัฏจักร หนึ่ง ๆ นั้นว่าจะใช้ระยะเวลายาวนานเท่าใด เนื่องจากการเปลี่ยนแปลงตามวัฏจักรส่วนใหญ่ เป็น ปรากฏการณ์ที่เกิดขึ้นในด้านธุรกิจและเศรษฐศาสตร์ ดังนั้นการผันแปรตามวัฏจักรโดยทั่ว ๆ ไป จะ แสดงถึงภาวการณ์เกิดซ้ำกันของภาวะธุรกิจเฟื่องฟูถดถอย และตกต่ำภาวะต่าง ๆ เหล่านี้อาจจะสั้น หรือยาวก็ได้

4) องค์ประกอบความผันแปรเชิงสุ่ม ซึ่งเป็นผลอันเนื่องมาจากความผิดปกติ (Irregular) เป็นการเปลี่ยนแปลงของข้อมูลอนุกรมเวลาที่เกิดจากปัจจัยที่ไม่อาจคาดคะเนได้ล่วงหน้า เช่น การ เกิดภาวะผิดปกติทางดินฟ้าอากาศ การเกิดน้ำท่วม การนัดหมายหยุดงานของบุคลากร และ การเกิด สงคราม เป็นต้น ซึ่งเป็นปรากฏการณ์ที่เราไม่อาจทำนายได้ล่วงหน้า

2.2.3.4. เทคนิคการพยากรณ์ โดยทั่วไปแล้วสามารถแบ่งออกเป็น 2 ประเภท ได้แก่ การพยากรณ์เชิงคุณภาพ (Qualitative forecasting methods) และการพยากรณ์เชิงปริมาณ (Quantitative forecasting methods)

1) การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting Methods) เป็นเทคนิคการ พยากรณ์ที่ขึ้นอยู่กับประสบการณ์ การประเมินความคิดเห็น ของผู้เชี่ยวชาญ (Expert opinion) และ ใช้ดุลยพินิจของบุคคลเพื่อทำนายการเปลี่ยนแปลงในระยะยาว เป็นการพยากรณ์ที่ใช้วิจารณญาณ (Judgmental Forecasting) โดยจะถือเกณฑ์วิจารณญาณส่วนบุคคลหรือมีการตกลงกันของ คณะกรรมการเกี่ยวกับเหตุการณ์ หรือสถานการณ์ในอนาคต โดยทั่วไปแล้วเทคนิคนี้จะถูกนำมาใช้ สำหรับการพยากรณ์ระยะยาว (Long–range Projection) หรือเมื่อองค์กรมีข้อมูลอยู่จำกัด ไม่ สามารถหาได้ หรือข้อมูลที่มีอยู่ในปัจจุบันไม่เกี่ยวข้อง

หรือเมื่อข้อมูลที่มีอยู่ไม่สามารถนำมาใช้ได้ นอกจากนั้นเทคนิคนี้ยังเหมาะกับการใช้แนะนำผลิตภัณฑ์ใหม่เข้าสู่ตลาด หรือมีการเปลี่ยนแปลง เทคโนโลยีใหม่ เนื่องจากไม่มีข้อมูลอยู่

2) การพยากรณ์เชิงปริมาณ (Quantitative Forecasting Methods) จะใช้เทคนิคทาง คณิตศาสตร์บนพื้นฐานของข้อมูล ปริมาณความต้องการที่ เก็บรวบรวมไว้ในอดีต (Historical Data) รวมทั้งข้อมูลที่เกี่ยวข้องอื่น ๆ เพื่อใช้ในการพยากรณ์ โดยจะจำแนกวิธีการพยากรณ์ออกเป็น 2 วิธี ใหญ่ ๆ คือ 1) การพยากรณ์ด้วยเทคนิคอนุกรมเวลา (Time 16 Series Forecasting) ซึ่งจะมีข้อ สมมุติที่ว่า ค่าพยากรณ์ที่เกิดขึ้นจะขึ้นกับข้อมูลที่ผ่านมาในอดีต ดังนั้นวิธีการนี้จึงจะใช้เฉพาะข้อมูล เชิงปริมาณที่เก็บรวบรวมไว้ในอดีตมาพยากรณ์ และ 2) การพยากรณ์เชิงสาเหตุ (Causal or Associating Forecasting) จะสมมุติว่าปัจจัยอื่น ๆ ตั้งแต่ 1 ตัวแปรขึ้นไป (ตัวแปรอิสระ) มี ความสัมพันธ์กับปริมาณความต้องการ ซึ่งจะนำเข้ามาใช้ในตัวแทนที่จะพยากรณ์ความต้องการใน อนาคต เนื่องจากการพยากรณ์เชิงปริมาณนั้นขึ้นกับข้อมูลในอดีต ดังนั้นค่าการพยากรณ์จะมีความ เชื่อถือลดลงเมื่อระยะเวลาการพยากรณ์เพิ่มขึ้น ดังนั้นหากองค์กรใดต้องการที่จะพยากรณ์ในระยะ ยาว ควรจะนำเอาการพยากรณ์ทั้งเชิงปริมาณและคุณภาพเข้ามาร่วมวิเคราะห์ ด้วย (ทรงศิริแต้สมบัติ, 2539)

2.2.3.5. การเลือกเทคนิคในการพยากรณ์ วิธีการพยากรณ์มีผู้พัฒนาขึ้นหลายวิธี โดยแต่ละวิธีก็จะเหมาะสมกับข้อมูลที่มีลักษณะการ เคลื่อนไหวที่แตกต่างกันออกไป รวมทั้งขึ้นอยู่กับวัตถุประสงค์ของการนำเอาค่าพยากรณ์ที่ได้ไปใช้งาน และจำเป็นอย่างยิ่งที่จะต้องเลือกวิธีการพยากรณ์ให้เหมาะสมกับการนำไปใช้ ดังนั้น ก่อนที่จะ ดำเนินการพยากรณ์ ผู้พยากรณ์จะต้องทราบและตระหนักถึงรายละเอียดต่าง ๆ ที่เกี่ยวข้อง ดังต่อไปนี้

1) ระยะเวลาในการพยากรณ์ล่วงหน้า ผู้พยากรณ์มักจะพยากรณ์การเกิดขึ้นของเหตุการณ์ แตกต่างกัน ขึ้นอยู่กับวัตถุประสงค์ของการนำค่าพยากรณ์ไปใช้งาน ระยะเวลาในการพยากรณ์ ล่วงหน้าสามารถแบ่งได้เป็น ระยะสั้น ระยะกลาง และ ระยะยาว การพยากรณ์ระยะสั้นจะเป็น ช่วงเวลาที่จะพยากรณ์ล่วงหน้าไม่เกิน 3 เดือน ระยะกลาง เป็นช่วงเวลาตั้งแต่ 3 เดือนขึ้นไป จนถึง 2 ปีและระยะยาว เป็นช่วงเวลาที่ต้องการพยากรณ์ไปล่วงหน้าตั้งแต่ 2 ปีขึ้นไป ระยะเวลาในการ 22 พยากรณ์ล่วงหน้านี้จะส่งผลถึงเทคนิคที่จะเลือกใช้ในการพยากรณ์ โดยแต่ละช่วงเวลาก็จะเหมาะสม กับเทคนิคการพยากรณ์ที่แตกต่างกันออกไป เช่น หากผู้พยากรณ์ต้องการที่จะพยากรณ์ในระยะยาว แล้วการพยากรณ์เชิงคุณภาพจะมีความเหมาะสมมากกว่า ข้อมูลที่ใช้ในการพยากรณ์นั้น อาจจะมี หน่วยวัดเป็นรายชั่วโมง รายวัน รายสัปดาห์ รายเดือน หรือรายไตรมาส ก็ได้ขึ้นกับประเด็นของเรื่องที่ ศึกษา เช่น ห้างสรรพสินค้าแห่งหนึ่งพิจารณาจำนวนลูกค้าที่เข้ามารับบริการในแต่ละวันที่ห้างเปิด

บริการเพื่อจัดสรรพนักงานให้บริการได้สอดคล้องกับจำนวนลูกค้าในแต่ละช่วงเวลา หรือบริษัทแห่ง หนึ่งผลิตเสื้อผ้าสำเร็จรูปต้องการพยากรณ์ยอดขายสินค้าเสื้อผ้าสำเร็จรูปในแต่ละเดือน เป็นต้น อย่างไรก็ตามข้อมูลที่นำมาใช้ในการวิเคราะห์นั้นจะต้องมีขนาดใหญ่พอสมควร รวมทั้งมีความทันสมัย ด้วย

2) รูปแบบของข้อมูล ส่วนประกอบของข้อมูลจะเป็นตัวกำหนดเทคนิคที่ใช้ในการพยากรณ์ ดังนั้นก่อนที่ผู้พยากรณ์จะเลือกวิธีที่จะใช้เทคนิคในการพยากรณ์ จำเป็นอย่างยิ่งที่จะต้องนำข้อมูลที่มี อยู่มาพล็อตกราฟลงจุด เพื่อดูลักษณะการเคลื่อนไหวของข้อมูลเบื้องต้นก่อน

3) ค่าใช้จ่ายในการพยากรณ์ ค่าใช้จ่ายเป็นหลักเกณฑ์สำคัญที่ใช้ในการตัดสินเลือกเทคนิค การพยากรณ์ซึ่งประกอบด้วยค่าใช้จ่ายที่เกิดขึ้น เพื่อพัฒนาตัวแบบที่ใช้ในการพยากรณ์ ค่าใช้จ่ายที่ เกิดขึ้นจากการเก็บข้อมูลที่ให้ตัวแบบที่สร้างขึ้นมีความถูกต้องแม่นยำตลอดเวลา และค่าใช้จ่ายที่ เกิดขึ้นจากความผิดพลาดที่มาจากการพยากรณ์ ความสำคัญของค่าใช้จ่ายแต่ละประเภทจะขึ้นอยู่กับ วิธีการและสภาวการณ์ เช่น วิธีการพยากรณ์เชิงคุณภาพและเชิงเทคโนโลยีส่วนมากจะต้องประมาณ ค่าใช้จ่ายแยกเป็นส่วนๆ รวมทั้งจะต้องมีการประมาณปัจจัยนำเข้าทางทรัพยากรมนุษย์ การได้ข้อมูล จากภายนอกจะเป็นค่าใช้จ่ายส่วนใหญ่ที่เกิดขึ้น และ ค่าใช้จ่ายเหล่านี้จะเกิดขึ้นอีกทุกครั้งที่จะ พยากรณ์ใหม่ สำหรับการพยากรณ์เชิงปริมาณแล้ว ค่าใช้จ่ายส่วนใหญ่ไม่ขึ้นอยู่กับสภาวการณ์ ทางด้านการบริการจัดการ วิธีการพยากรณ์เชิงปริมาณที่ใช้กันในองค์กรส่วนใหญ่จะใช้คอมพิวเตอร์ ค่าใช้จ่ายในด้านการพัฒนาจึงเป็นเรื่องของการเขียนและดัดแปลงโปรแกรมที่จะใช้ในการพยากรณ์ ซึ่ง จะรวมถึงทรัพยากรมนุษย์ที่ต้องใช้เพื่อพัฒนาโปรแกรม และค่าใช้จ่ายของเวลาคอมพิวเตอร์เพื่อ จัดระบบการทำงานของวิธีการที่ใช้ เป็นต้น

4) ระดับความแม่นยำในการพยากรณ์ ระดับความแม่นยำของการพยากรณ์และความคลาด เคลื่อนที่เกิดขึ้นจะขึ้นกับสถานการณ์ที่ต้องการจะพยากรณ์ ในบางกรณี ความผิดพลาด 20% ก็ สามารถยอมรับได้ ในขณะที่บางกรณีความผิดพลาด 1% ก็ก่อให้เกิดความเสียหายมากมายต่อองค์กร และหากพิจารณาในแง่ของการวิเคราะห์การตัดสินใจแล้วก็จะพบว่า มีความแตกต่างระหว่างการ ตัดสินใจที่ดี (Good Decision) และผลลัพธ์ที่ดี (Good Outcome) ถ้าหากผู้พยากรณ์สามารถออก แบบจำลองสภาวการณ์ได้ท่ามกลางความไม่แน่นอน ก็น่าที่จะส่งเสริมสนับสนุนการพยากรณ์โดยไม่ ต้องคำนึงถึงระดับความแม่นยำมากนัก

5) ข้อมูลในอดีตที่เก็บรวบรวมไว้ ข้อมูลในอดีตที่มีอยู่จะเป็นตัวตัดสินใจ หลักในการที่จะ เลือกเทคนิคการพยากรณ์ นอกจากนั้นแล้วความถูกต้องของข้อมูลก็เป็นปัจจัยที่สำคัญอย่างยิ่งในการ พยากรณ์ ดังนั้นหากองค์กรใดยังไม่ได้จัดเก็บข้อมูลในอดีตจำเป็นอย่างยิ่งที่จะต้องรีบดำเนินการ

6) ความง่ายในการที่ผู้ปฏิบัติจะนำไปใช้ต่อความง่ายของเทคนิคการพยากรณ์ในการที่ผู้ ปฏิบัติจะนำไปใช้ต่อเป็นปัจจัยที่สำคัญอีกปัจจัยหนึ่ง หากผู้บริหารหรือผู้ที่เกี่ยวข้องขาดความรู้ความ 23 เข้าใจในเทคนิคการพยากรณ์ที่นำมาใช้การพยากรณ์ที่ทำขึ้นก็จะไม่มีความหมาย หรือถูกนำไป ประยุกต์ใช้อย่างไม่ถูกต้อง เช่น วิธีการของบ๊อก–เจนกินส์ (Box and Jenkins Method) ไม่เป็นที่ นิยมในหลายองค์กร เนื่องจากเทคนิคดังกล่าวยุ่งยากเกินไปสำหรับผู้ใช้ ที่จะเข้าใจแนวความคิด พื้นฐานของวิธีการในระดับที่จะมั่นใจได้ว่า วิธีการดังกล่าวสามารถนำไปประยุกต์ใช้ได้อย่างถูกต้อง

2.2.3.6. กระบวนการพยากรณ์

กระบวนการพยากรณ์ (Forecasting Process) หมายถึง ขั้นตอนการเลือกเทคนิคการ พยากรณ์ตั้งแต่หนึ่งวิธีขึ้นไป ที่สามารถประยุกต์ใช้กับข้อมูลที่จำเป็นต่อการพยากรณ์ ซึ่งประกอบด้วย 9 ขั้นตอน ต่อไปนี้ (กิตติพงศ์อินทร์ทอง, 2556)

1) การกำหนดวัตถุประสงค์ของการพยากรณ์ (Specific Objectives) เป็น การกำหนด วัตถุประสงค์ให้ชัดเจนว่าการพยากรณ์จะไปใช้ในการตัดสินใจอย่างไร เช่น ใช้เพื่อตัดสินใจ ลงทุน (การพยากรณ์ระยะยาว) หรือเพื่อวางแผนกลยุทธ์ (การพยากรณ์ระยะกลาง)

2) การกำหนดสิ่งที่จะพยากรณ์ให้ชัดเจน (Determine what to forecast) เช่น พยากรณ์ ยอดขายเป็นหน่วยสินค้า หรือเป็นตัวเงิน (บาทหรือดอลลาร์) การพยากรณ์เป็นยอดขายรวม ยอดขาย สายผลิตภัณฑ์ยอดขายของแต่ละภูมิภาค ยอดขายในประเทศ หรือยอดขายต่างประเทศเป็นต้น

3) การเก็บรวบรวมข้อมูล (Data Collection) เป็นการเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับการ พยากรณ์อย่างเหมาะสม และเป็นข้อมูลที่ถูกต้อง ขั้นตอนนี้เป็นขั้นตอนที่ยากและใช้เวลามากที่สุดใน การเก็บรวบรวมข้อมูลจะต้องพิจารณาถึงข้อกำหนดด้านเวลา (Identify time dimensions) โดย พิจารณา 2 ประการคือ ช่วงระยะเวลาการพยากรณ์ (Length and periodicity) เช่น ประจำปี ประจำไตรมาส ประจำเดือน ประจำวัน และความเร่งด่วนในการพยากรณ์ (Urgency) ถ้ามีความ จำเป็นเร่งด่วน วิธีที่ใช้ในการพยากรณ์จะมีความซับซ้อนน้อยกว่า และข้อกำหนดเกี่ยวกับข้อมูล (Data considerations) การพิจารณาจาก

ปริมาณและประเภทของข้อมูลที่มีเป็นข้อมูลภายในหรือภายนอก บริษัทเป็นข้อมูลรายปีรายเดือน เป็นข้อมูลที่เป็นตัวเงินหรือหน่วยสินค้า

4) การลดข้อมูล (Data Reduction) บางครั้งข้อมูลที่เก็บรวบรวมมีมากเกินไป และทำให้ การพยากรณ์มีความถูกต้องน้อยลง จึงจำเป็นต้องลดข้อมูลบางตัวที่อาจจะไม่เกี่ยวข้องกับการ พยากรณ์ลง

5) การเลือกแบบจำลองในการพยากรณ์ (Model Selection) การเลือกวิธีการพยากรณ์ ขึ้นอยู่กับรูปแบบของข้อมูล จำนวนข้อมูลที่มีและระยะเวลาการพยากรณ์การเลือกวิธีการพยากรณ์ที่ เหมาะสมกับข้อมูลจะช่วยลดความผิดพลาดในการพยากรณ์รูปแบบการพยากรณ์ ที่ยิ่งง่ายจะยิ่งดีต่อ การยอมรับของผู้ตัดสินใจ วิธีการพยากรณ์จะต้องมีความสมดุลระหว่างความถูกต้อง และเป็นวิธีที่ง่าย ต่อความเข้าใจ

6) การพยากรณ์ (Model Extrapolation) เป็นการพยากรณ์เหตุการณ์ที่ผ่านไป โดยใช้ ข้อมูลจริงที่มีอยู่ และประเมินว่าวิธีใดเหมาะสม (fit) กับข้อมูลในอดีตก่อน โดยการวัดค่าคลาดเคลื่อน ที่เกิดขึ้น

7) การเตรียมการพยากรณ์ (Forecast Preparation) เมื่อมีวิธีการพยากรณ์มากกว่า 2 วิธี ขึ้นไปที่ เหมาะสม การรวมค่าการพยากรณ์จากวิธีเหล่านั้นจะทำให้ค่าพยากรณ์ดีขึ้นกว่าการใช้วิธี เดียว 24

8) การนำเสนอผลการพยากรณ์ (Forecast Presentation) การนำเสนอค่าพยากรณ์ ให้กับ ผู้บริหารหรือผู้ใช้ด้วยการเขียนเป็นลายลักษณ์อักษรหรือนำเสนอด้วยวาจา (Written/oral) เป็นขั้น ตอนนี้มีความสำคัญมากเช่นกัน เพราะสามารถสร้างความเข้าใจให้กับผู้บริหารหรือผู้ใช้ได้

9) การตรวจสอบผลการพยากรณ์ (Tracking Results) การติดตามผลอย่างต่อเนื่อง ว่าผล การพยากรณ์เมื่อเปรียบเทียบกับค่าจริงแล้วมีความถูกต้องอย่างไร ซึ่งวิธีที่เคยพยากรณ์ได้ดีที่สุด อาจ มีความถูกต้องลดลง เนื่องจากสภาพแวดล้อมเปลี่ยนไปอาจจะต้องหาวิธีอื่นมาแทนการพยากรณ์ สามารถเรียนรู้ได้จากความผิดพลาด การทบทวนค่าคลาดเคลื่อนของการพยากรณ์อย่างรอบคอบจะ ช่วยให้สามารถเข้าใจถึงสาเหตุของความเบี่ยงเบนระหว่างค่าจริงกับค่าพยากรณ์ได้ดีขึ้น

2.2.3.7. การพิจารณาเลือกตัวแบบพยากรณ์ ในการเลือกตัวแบบหรือวิธีการพยากรณ์ลำดับแรกผู้พยากรณ์จำเป็นต้องศึกษารูปแบบ ของ ชุดข้อมูลอย่างละเอียดก่อน เช่น ทำการตรวจสอบว่าชุดข้อมูลมีรูปแบบอย่างไร รูปแบบองค์ประกอบ ของแนวโน้ม วัฏจักร ฤดูกาล หรือว่ามีเพียงตัวแปรสุ่มเพียงอย่างเดียว ซึ่งวิธีการที่จะทำให้ทราบถึง องค์ประกอบ

ของข้อมูลเหล่านี้สามารถทำได้จากการวาดกราฟและการวิเคราะห์ค่าสัมประสิทธิ์ สหสัมพันธ์ เมื่อทราบรูปแบบของชุดข้อมูลแล้ว จึงนำไปเลือกตัวแบบหรือวิธีการการพยากรณ์โดย เกณฑ์ในการเลือก วิธีการพยากรณ์ที่เหมาะสมมีดังนี้ (กิตติพงศ์อินทร์ทอง, 2556)

1) ชุดข้อมูลที่มีลักษณะคงที่ (Stationary Data) คือ อนุกรมที่มีค่าเฉลี่ยไม่เปลี่ยนแปลง เมื่อเวลาผ่านไป วิธีการพยากรณ์จะใช้ข้อมูลในอดีตเป็นค่าพยากรณ์ในอนาคต โดยที่เทคนิคการ พยากรณ์สำหรับข้อมูลที่มีลักษณะคงที่จะใช้เมื่อ

ก) ข้อมูลที่ไม่ค่อยเปลี่ยนแปลง เช่น จำนวนของเสียต่อสัปดาห์ ซึ่งมีอัตราเดียวกัน ทุกสัปดาห์

ข) ต้องการรูปแบบง่ายๆ เพราะขาดข้อมูล หรือเพื่อให้ง่ายต่อการอธิบาย หรือการ ปฏิบัติเช่น ธุรกิจหรือองค์กรใหม่และมีข้อมูลในอดีตเล็กน้อย

ค) ข้อมูลที่มีลักษณะเป็นแนวโน้ม อาจมีการเปลี่ยนรูปเป็นข้อมูลคงที่ เช่น การ เปลี่ยนรูปอนุกรมโดยวิธีถอดรากที่สองหรือการหาผลต่าง

ง) ข้อมูลที่เป็นกลุ่มของค่าคลาดเคลื่อนจากการพยากรณ์ เทคนิควิธีการพยากรณ์ที่ เหมาะสมกับชุดข้อมูลที่มีลักษณะคงที่ ได้แก่ วิธีนาอีฟ (Native Methods) วิธีค่าเฉลี่ยอย่างง่าย (Simple Average Methods) วิธีค่าเฉลี่ยเคลื่อนที่ (Moving Average) วิธีบ๊อกซ์และเจนกินส์ (BoxJenkins Methods)

2) ชุดข้อมูลที่มีลักษณะเป็นแนวโน้ม (Data with a Trend) คือ อนุกรมเวลาที่มีการเพิ่มขึ้น หรือลดลงเมื่อเวลาเพิ่มขึ้นในระยะยาว หรือกล่าวได้ว่าอนุกรมเวลาที่มีลักษณะเป็นแนวโน้ม ค่าเฉลี่ย จะมีการเปลี่ยนแปลงเมื่อเวลาเพิ่มขึ้น และสามารถคาดได้ว่าจะเพิ่มขึ้นหรือลดลงในช่วงเวลาที่ พยากรณ์เวลาใด เทคนิคการพยากรณ์สำหรับข้อมูลที่มีแนวโน้มจะใช้เมื่อประสบสถานการณ์ดังนี้

ก) มีการเพิ่มขึ้นของผลิตผลและเทคโนโลยีใหม่ที่ทำให้รูปแบบการดำรงชีวิต (Lifestyle) ของผู้บริโภคเปลี่ยนแปลงไป

ข) เมื่อมีการเพิ่มขึ้นหรือลดลงของจำนวนประชากร ทำให้ความต้องการสินค้า หรือ บริการเปลี่ยนแปลงไปในทิศทางเดียวกัน 25

ค) เมื่ออำนาจการซื้อได้รับผลกระทบจากตัวแปรทางเศรษฐกิจเนื่องจากเงินเฟ้อ

ง) เมื่อผู้บริโภครู้จักหรือยอมรับผลิตภัณฑ์มากขึ้น เทคนิควิธีการพยากรณ์ที่เหมาะสมกับชุดข้อมูลที่มีลักษณะแบบแนวโน้มคือ วิธี ค่าเฉลี่ยเคลื่อนที่ (Moving Average) วิธีปรับเรียบเอ็กซ์โปแนนเชียลของโฮล์ท (Holt's Exponential Smoothing

Method) วิธีการวิเคราะห์ความถดถอย (Simple Regression) วิธีปรับให้เรียบเอ็กซ์ โปแนนเชียลซ้ำสองครั้งหรือวิธีของบราวน์ (Double Exponential Smoothing) วิธีบ๊อกซ์และเจน กินส์ (Box–Jenkins Methods)

3) ชุดข้อมูลที่มีลักษณะฤดูกาล (Seasonal Data) คือ อนุกรมฤดูกาลเป็นอนุกรมเวลาที่มี รูปแบบที่เปลี่ยนแปลงซ้ำเดิมในช่วงเวลาเดียวกันทุกปีการพัฒนาเทคนิคการพยากรณ์สำหรับอนุกรม ฤดูกาล มักเป็นวิธีที่เกี่ยวข้องกับการแยกส่วนประกอบอนุกรมเวลา โดยมีการประมาณค่าดัชนีฤดูกาล จากอนุกรมในอดีต ค่าดัชนีเหล่านี้ใช้เพื่อเพิ่มหรือขจัดค่าฤดูกาลในการพยากรณ์ออกจากค่าสังเกต เทคนิคการพยากรณ์สำหรับข้อมูลที่มีลักษณะฤดูกาลจะใช้เมื่อประสบสถานการณ์ดังนี้

ก) สภาพของอากาศมีอิทธิพลต่อข้อมูลที่สนใจ เช่น ยอดขายเครื่องปรับอากาศในฤดูร้อน กิจกรรมในฤดูร้อนหรือฤดูหนาว (เช่น การว่ายน้ำ) เสื้อผ้า และผลิตผลเกษตรตามฤดูกาล

ข) เวลาตามปฏิทินมีผลต่อข้อมูลที่สนใจ เช่น ยอดขายร้านค้าปลีกในวันหยุด วันปีใหม่ เทคนิควิธีพยากรณ์ที่ใช้สำหรับชุดข้อมูลที่มีลักษณะแบบฤดูกาลได้แก่ วิธีแยกองค์ประกอบอนุกรม เวลา (Classical Decomposition) วิธี Census X–12 วิธีปรับเรียบเอ็กซ์โปแนนเชียลวินเตอร์ (Winter's Exponential Smoothing) วิธีการวิเคราะห์ถดถอยพหุคูณ (Multiple Regression) และ วิธีบ๊อกซ์และเจนกินส์ (Box–Jenkins Methods)

4) ชุดข้อมูลที่มีลักษณะที่เคลื่อนไหวตามวัฏจักร (Cyclical Series) คือ อนุกรมเวลาที่มีการ เคลื่อนไหวตามวัฏจักร มีลักษณะการเคลื่อนไหวขึ้นลงคล้ายรูปคลื่นรอบๆ เส้นแนวโน้ม โดยลักษณะ ของข้อมูลจะเกิดขึ้นซ้ำกันทุก 2–3 ปีหรือมากกว่านั้น การสร้างแบบจำลองของรูปแบบวัฏจักรทำได้ ยาก เพราะมีรูปแบบไม่แน่นอน และขนาดของการเคลื่อนไหวมักจะแตกต่างกัน โดยสามารถนำวิธี แยกส่วนประกอบอนุกรมเวลามาวิเคราะห์ข้อมูลที่มีลักษณะที่เคลื่อนไหวตามวัฏจักรได้เนื่องจากวัฏ จักรจะมีลักษณะที่ไม่ปกติการวิเคราะห์ส่วนประกอบของวัฏจักรจำเป็นต้องหาตัวชี้นำทางเศรษฐกิจ

ก) เทคนิคการพยากรณ์สำหรับข้อมูลที่มีลักษณะที่เคลื่อนไหวตามวัฏจักร จะใช้เมื่อ สถานการณ์ดังนี้ วงจรของธุรกิจมีอิทธิพลต่อข้อมูลที่สนใจ เช่น ปัจจัยทางเศรษฐกิจ การตลาดหรือการแข่งขัน เกิดการเปลี่ยนแปลงในรสนิยม เช่น แฟชั่น ดนตรี อาหาร เกิดการเปลี่ยนแปลงของประชากร เช่น เกิดสงคราม อดอยาก โรคระบาด และภัยธรรมชาติ เกิดการเปลี่ยนแปลงในวงจรชีวิตของผลิตภัณฑ์เช่น ขั้นแนะนำ ขั้นเจริญเติบโต ขั้นอิ่มตัว และขั้นถดถอย

ข) เทคนิควิธีการพยากรณ์ที่ใช้สำหรับชุดข้อมูลที่เคลื่อนไหวตามวัฏจักร ได้แก่ วิธีแยก องค์ประกอบอนุกรมเวลา (Classical Decomposition) วิธีการวิเคราะห์ถดถอยพหุคูณ (Multiple Regression) วิธีบ๊อกซ์และเจนกินส์ (Box–Jenkins Methods) ตัวชี้วัดภาวะทางเศรษฐกิจ (Economic Indicators) แบบจำลองทางเศรษฐมิติ (Econometric Models) 26

ค) เทคนิคหรือวิธีการพยากรณ์ที่ได้กล่าวมานี้ จะสัมพันธ์กับระยะเวลา สำหรับการพยากรณ์ ในระยะสั้นและระยะกลางจะสามารถใช้เทคนิคการพยากรณ์ได้หลากหลายวิธีแต่เมื่อช่วงระยะเวลา ในการพยากรณ์เพิ่มขึ้น จำนวนเทคนิคที่จะนำมาประยุกต์ใช้จะลดน้อยลง เช่น เทคนิคค่าเฉลี่ย เคลื่อนที่ และวิธีการปรับเรียบจะใช้คาดการณ์เกี่ยวกับเศรษฐกิจได้ไม่ดีนัก ในขณะที่แบบจำลองทาง เศรษฐมิติ (Econometric Models) จะใช้ได้ดีกว่า โดยสามารถจำแนกวิธีการพยากรณ์ให้มีความ เหมาะสมกับระยะเวลาได้ดังนี้ วิธีการวิเคราะห์ถดถอย เหมาะสำหรับการพยากรณ์ในระยะสั้น ระยะกลางและระยะยาว วิธีค่าเฉลี่ยเคลื่อนที่การแยกส่วนประกอบอนุกรมเวลา เหมาะสำหรับการพยากรณ์ระยะ สั้นและระยะกลาง แบบจำลองทางเศรษฐมิติเหมาะสำหรับการพยากรณ์ระยะสั้นและระยะกลาง ส่วนการพยากรณ์เชิงคุณภาพมักใช้ในการพยากรณ์ระยะยาว โดยผู้ที่ต้องการพยากรณ์ จะคาดการณ์โดยอาศัยประสบการณ์

2.2.3.8. การเปรียบเทียบความคลาดเคลื่อนของตัวแบบพยากรณ์ ในการพยากรณ์ใด ๆ ย่อมต้องการให้มีความคลาดเคลื่อนน้อยที่สุด ดังนั้น ในการเลือกใช้แบบจำลองการพยากรณ์จึงควรพิจารณาว่าการพยากรณ์ที่ได้นั้น มีความแม่นยำมากน้อยเพียงใดวิธีการประเมินความ แม่นยำของการพยากรณ์มีหลายวิธี เช่น Root Mean Square Error (RMSE) Mean AbsoluteError (MAE) Mean Absolute percent Error (MAPE) และ Mean Square Error (MSE) เป็นต้น แต่ในที่นี้จะกล่าวถึง เพียงวิธีเดียว คือ วิธีMean Absolute Percentage Error: MAPE คือค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ ความคลาดเคลื่อนของผลต่างระหว่างค่าที่เกิดขึ้นจริงกับค่าพยากรณ์ ซึ่งเป็นค่าใช้วัดความถูกต้องของการ พยากรณ์ที่เหมาะกับการเปรียบเทียบอนุกรมเวลาหลายชุด เมื่อใช้วิธีการพยากรณ์เดียวกัน หรือเปรียบเทียบ วิธีการพยากรณ์หลายวิธีเมื่อใช้อนุกรมเวลาชุดเดียวกัน ซึ่งตัวแบบที่เหมาะสมที่สุดจะให้ค่า MAPE ต่ำสุด สามารถคำนวณได้จากสมการ ดังนี้ (สุพรรณี อึ้งปัญสัต วงศ์, 2555)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100$$

เมื่อ Yt แทนอนุกรมเวลา ณ เวลา t

เมื่อ $\hat{Y}t$ แทนค่าพยากรณ์ของอนุกรมเวลา ณ เวลา t

เมื่อ n แทนจำนวนข้อมูลที่ศึกษา

และมีเกณฑ์ที่ใช้สำหรับการพยากรณ์ ดังนี้

ถ้าค่า MAPE น้อยกว่า 10%	จัดว่าการพยากรณ์ค่อนข้างแม่นยำ
ถ้าค่า MAPE อยู่ระหว่าง 10% – 20%	จัดว่าการพยากรณ์ใช้ได้ดี
ถ้าค่า MAPE อยู่ระหว่าง 20% – 50%	จัดว่าการพยากรณ์ค่อนข้างแม่นยำ
ถ้าค่า MAPE มากกว่า 50%	จัดว่าการพยากรณ์ไม่แม่นยำ

2.2.4 ทฤษฎีเกี่ยวกับเทคนิค การถดถอยเชิงเส้นแบบพหุ(Multiple Linear Regression) การวิเคราะห์ถดถอยพหุคูณ (Multiple regression Analysis) เป็นวิธีการวิเคราะห์ข้อมูล เพื่อหาความสัมพันธ์ระหว่างตัวแปรตาม (Y) หรือตัวแปรเกณฑ์ (Criterion Variable) จำนวน 1 ตัว กับตัวแปรอิสระ (X) หรือตัวแปรพยากรณ์ หรือตัวแปรทำนาย (Predictor Variable) ตั้งแต่ 2 ตัวขึ้น ไป เป็นเทคนิคทางสถิติที่อาศัยความสัมพันธ์เชิงเส้นตรงระหวางตัวแปรมาใช้ในการทำนาย โดย เมื่อทราบค่าตัวแปรหนึ่งก็สามารถทำนายอีกตัวแปรหนึ่งได้ ผลการวิเคราะห์ข้อมูลที่ได้อยู่ในรูป ของสมการทำนาย สิ่งสำคัญที่ต้องการหาในการวิเคราะห์การถดถอยพหุคูณ คือ สัมประสิทธิ์ สหสัมพันธ์พหุคูณ สมการพยากรณ์ในรูปคะแนนดิบ หรือในรูปคะแนนมาตรฐาน หรือทั้งคู่ และ ความคลาดเคลื่อนมาตรฐานในการพยากรณ์

หลักการและแนวคิดของการวิเคราะห์การถดถอยเชิงเส้น

การวิเคราะห์การถดถอยพหุคูณเชิงเส้น เป็นการศึกษาความสัมพันธ์ ระหว่างตัวแปรทำนายหลายตัว กับตัวแปรตาม 1 ตัว เพื่อศึกษาว่ามีตัวแปรทำนายตัวใดบ้างที่ร่วมกัน ทำนายหรือพยากรณ์หรืออธิบายการผันแปรของตัวแปรตามได้โดยเขียนความสัมพันธ์ในรูปแบบของสมการได้ ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

เมื่อ

Y_i คือ ค่าของตัวแปรตาม (Dependent Variable)

X_{ij} คือ ค่าตัวที่ i ของตัวแปรทำนายตัวที่ j (Independent Variable)

p คือ จำนวนตัวแปรทำนายในสมการถดถอย

β_0 คือ ระยะตัดแกน Y หรือค่าของ Y เมื่อ X มีค่าเป็นศูนย์

β_i คือ ค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) ของตัวแปรทำนาย Xi

ε_i คือ ความคลาดเคลื่อนเชิงสุ่ม (Random Error or Random Residual)

ความคลาดเคลื่อนมีข้อกำหนดว่าต้องเป็นตัวแปรสุ่มที่มีการแจกแจง ปรกติโดยมีค่าเฉลี่ยหรือ $E(\varepsilon)$ เท่ากับ 0 และความแปรปรวนหรือ เท่ากับ $\sigma^2(\varepsilon_i)$ เท่ากับ σ^2 จากข้อตกลงของความคลาดเคลื่อนดังกล่าวส่งผลให้ตัวแปรตาม Y แต่ละค่ามีความเป็นอิสระต่อกันและมี การแจกแจงปกติ

2.2.5 ทฤษฎีเกี่ยวกับเทคนิค Gradient Boosted Trees (GBT)

Gradient Boosted Trees (GBT) เป็นอัลกอริทึมการเรียนรู้ของเครื่องในกลุ่ม Ensemble Learning ที่ผสานแนวคิดของ Boosting เข้ากับ Decision Tree โดยมีวัตถุประสงค์เพื่อเพิ่มความแม่นยำของแบบจำลองผ่านการรวมแบบจำลองย่อยหลายตัวในลักษณะของแบบจำลองเชิงบวก (additive model) แนวคิดของ Boosting มีรากฐานมาจากงานของ Freund และ Schapire (1997) ซึ่งแสดงให้เห็นว่าแบบจำลองที่มีความสามารถจำกัด (weak learners) สามารถนำมารวมกันเพื่อสร้างแบบจำลองที่มีประสิทธิภาพสูงขึ้นได้

ต่อมา Friedman (2001) ได้นำเสนอแนวคิด Gradient Boosting โดยมองปัญหาการเรียนรู้ของเครื่องเป็นปัญหาการลดค่าฟังก์ชันความสูญเสีย (loss function) ด้วยวิธีการคล้าย gradient descent ในปริภูมิของฟังก์ชัน และเมื่อเลือกใช้ Decision Tree เป็นแบบจำลองฐาน จึงเกิดเป็นอัลกอริทึมที่เรียกว่า Gradient Boosted Trees

Gradient Boosted Trees สามารถอธิบายได้ในรูปของแบบจำลองเชิงบวกของสมการได้ ดังนี้

$$F(x) = \sum_{m=1}^{M} nh_m(x)$$

โดยที่ $F(x)$คือแบบจำลองพยากรณ์สุดท้าย, $h_m(x)$ คือ decision tree ลำดับที่ m, M คือจำนวนต้นไม้ทั้งหมด และ η คืออัตราการเรียนรู้ (learning rate) ซึ่งใช้ควบคุมอิทธิพลของต้นไม้แต่ละต้นต่อแบบจำลองรวม

2.2.6 ทฤษฎีเกี่ยวกับเทคนิค Random Forest Regression (สำหรับอนุกรมเวลา)

Random Forest เป็นเทคนิคการเรียนรู้ของเครื่องที่อยู่ในกลุ่ม Ensemble Learning โดยพัฒนาโดย Leo Breiman (2001) มีพื้นฐานจากแนวคิด Bagging (Bootstrap Aggregating) และ Decision Tree Theory จุดมุ่งหมายหลักคือการเพิ่มความแม่นยำและลดความแปรปรวน (variance) ของแบบจำลองการเรียนรู้ โดยอาศัยการสร้าง “ป่า” ของต้นไม้ตัดสินใจจำนวนมากแล้วรวมผลลัพธ์เข้าด้วยกัน (AnalytixLabs, 2023)

หลักการทำงานของ Random Forest

1) การสร้างต้นไม้ด้วย Bootstrap Sampling

1.1 สุ่มตัวอย่างข้อมูลจากชุดข้อมูลต้นฉบับแบบมีการคืนค่า (bootstrap)

1.2 ใช้แต่ละชุด bootstrap ในการสร้าง Decision Tree หนึ่งต้น

2) Feature Subset Selection (Random Subspace Method)

2.1 เมื่อแตก node ในแต่ละต้นไม้ จะไม่พิจารณาคุณลักษณะ (features) ทั้งหมด แต่จะสุ่มเลือกเฉพาะบางส่วน

2.2 ทำให้ต้นไม้ที่สร้างขึ้นมีความแตกต่างและลดการเกิด *correlation* ระหว่างต้นไม้

3) การรวมผลลัพธ์ (Ensemble Aggregation)

3.1 สำหรับการถดถอย (Regression) ผลลัพธ์สุดท้ายคำนวณจากค่าเฉลี่ยของผลลัพธ์จากทุกต้นไม้

$$\hat{f}(x) = \frac{1}{B}\sum_{b-1}^{B} T_b\,(x)$$

โดยที่ $\hat{f}(x)$ = ค่าทำนายสุดท้าย, $T_b(x)$ = ค่าทำนายจากต้นไม้ที่ b และ B = จำนวนต้นไม้ทั้งหมด

Random Forest สำหรับอนุกรมเวลา ไม่ได้ถูกออกแบบมาโดยตรงสำหรับอนุกรมเวลา (Time Series) แต่สามารถนำมาประยุกต์ใช้ได้ผ่านการสร้าง **คุณลักษณะเชิงเวลา (Temporal Features)** ได้แก่ :

Lag Features : ใช้ค่าที่ผ่านมา เช่น $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ เพื่อทำนายค่า y_t

Rolling Statistics : เช่น ค่าเฉลี่ยเคลื่อนที่ ความแปรปรวนเคลื่อนที่ เพื่อจับแนวโน้มระยะสั้นและความผันผวน

Seasonal Dummies : ตัวแปรแทนเดือน ไตรมาส หรือฤดูกาล เพื่อช่วยจับ pattern ที่ซ้ำกัน

เมื่อสร้าง features แบบนี้ Random Forest Regression จะสามารถพยากรณ์อนุกรมเวลาที่มีความซับซ้อนได้ดี โดยเฉพาะเมื่อความสัมพันธ์ไม่เป็นเชิงเส้น (*non–linear*)

2.2.7 ทฤษฎีเกี่ยวกับเทคนิค k–Nearest Neighbors (kNN)

k–Nearest Neighbors (kNN) เป็นอัลกอริทึมการเรียนรู้ของเครื่องในกลุ่ม **Instance–Based Learning** หรือ **Lazy Learning** ซึ่งแตกต่างจากอัลกอริทึมแบบสร้างแบบจำลองเชิงพารามิเตอร์ (parametric model) เนื่องจาก kNN ไม่มีขั้นตอนการฝึกแบบจำลองอย่างชัดเจน แต่จะเก็บข้อมูลฝึกทั้งหมดไว้ และทำการพยากรณ์เมื่อมีข้อมูลใหม่เข้ามาโดยอาศัยความคล้ายคลึงของข้อมูล

แนวคิดพื้นฐานของ kNN คือ การพิจารณาว่าข้อมูลที่มีลักษณะคล้ายกันมักจะมีผลลัพธ์หรือคลาสเดียวกัน หลักการนี้ถูกนำมาใช้ทั้งในงานจำแนกประเภท (classification) และงานถดถอย (regression) และถือเป็นหนึ่งในอัลกอริทึมพื้นฐานที่ถูกนำมาใช้เป็นจุดเริ่มต้นในการศึกษาการเรียนรู้ของเครื่อง (Cover & Hart, 1967)

หลักการทำงานของ k–Nearest Neighbors

เมื่อมีข้อมูลตัวอย่างใหม่ x เข้ามา ระบบจะคำนวณระยะห่างระหว่าง x กับข้อมูลทุกตัวอย่างในชุดข้อมูลฝึก จากนั้นจะเลือกข้อมูลจำนวน k ตัวอย่างที่อยู่ใกล้ที่สุด ซึ่งเรียกว่าเพื่อนบ้านใกล้เคียง (nearest neighbors)

ในกรณีของงานจำแนกประเภท ค่าผลลัพธ์จะถูกกำหนดจากคลาสที่ปรากฏมากที่สุดในกลุ่มเพื่อนบ้านทั้ง k ตัวอย่าง ขณะที่ในงานถดถอย ค่าผลลัพธ์จะคำนวณจากค่าเฉลี่ยของผลลัพธ์ของเพื่อนบ้านเหล่านั้น

การคำนวณระยะห่าง (Distance Metrics)

หัวใจสำคัญของเทคนิค kNN คือการวัดระยะห่างระหว่างข้อมูล โดยตัววัดที่นิยมใช้ ได้แก่

ระยะห่างแบบยูคลิด (Euclidean Distance)

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{p} (x_{il} - x_{jl})^2}$$

โดยที่

x_i, x_j คือเวกเตอร์ข้อมูล

p คือจำนวนคุณลักษณะ

นอกจากนี้ยังสามารถใช้ระยะห่างแบบ Manhattan, Minkowski หรือ Cosine distance ได้ ขึ้นอยู่กับลักษณะของข้อมูลและปัญหาที่ศึกษา (Hastie et al., 2009)

การเลือกค่าพารามิเตอร์ kkk

ค่าของ k มีผลโดยตรงต่อประสิทธิภาพของ kNN หากเลือกค่า k ที่มีค่าน้อยเกินไป แบบจำลองอาจมีความไวต่อ noise และเกิด overfitting ในทางตรงกันข้าม หากเลือกค่า k ที่มีค่ามากเกินไป แบบจำลองอาจไม่สามารถจับรายละเอียดของข้อมูลได้เพียงพอและเกิด underfitting ดังนั้นการเลือกค่า k ที่เหมาะสมจึงมักทำโดยใช้เทคนิค cross-validation

kNN สำหรับงานถดถอย ในกรณีของ kNN regression ค่าผลลัพธ์สามารถคำนวณได้จากค่าเฉลี่ยของผลลัพธ์ของเพื่อนบ้านทั้ง k ตัวอย่าง ดังสมการ

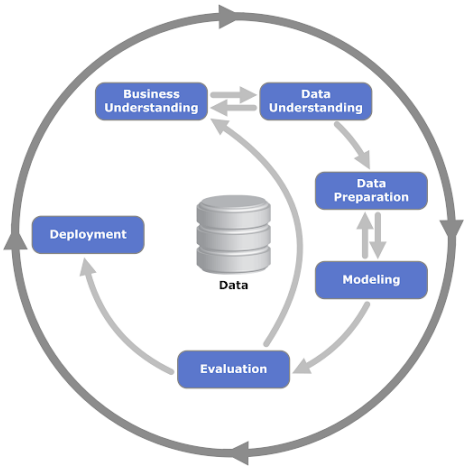
$$\hat{y}(x) = \frac{1}{k} \sum_{i \epsilon N_k(x)} y_i$$

โดยที่ $N_k(x)$ คือเซตของเพื่อนบ้านใกล้เคียง k ตัวอย่างของข้อมูล x

2.2.8 ทฤษฎี CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) เป็นกรอบการทำงานมาตรฐานสำหรับการทำเหมืองข้อมูลและโครงการวิเคราะห์ข้อมูล ซึ่งถูกพัฒนาตั้งแต่ช่วงปลายทศวรรษ 1990 เพื่อเป็นแนวทางให้สามารถดำเนินงานด้านข้อมูลได้อย่างเป็นระบบและสามารถปรับใช้ได้กับทุกอุตสาหกรรม (Coforge, 2025) ลักษณะเด่นของ CRISP-DM คือความยืดหยุ่นและการ

ทำงานแบบเชิงวนรอบ (*iterative process*) ที่อนุญาตให้ผู้ปฏิบัติย้อนกลับไปแก้ไขหรือปรับปรุงขั้นตอนก่อนหน้าได้หากจำเป็น (Boonchob, 2020)



ภาพที่ 2.11 The LSTM Cell

(ที่มา : https://kamboonchob.medium.com)

กระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูล เพื่อทำการวิเคราะห์และนำไปใช้ประโยชน์ทางธุรกิจ มีอยู่ 6 ขั้นตอน คือ

1) **การทำความเข้าใจธุรกิจ (Business Understanding)** ขั้นตอนแรกมุ่งไปที่การทำความเข้าใจธุรกิจ ปัญหาและวัตถุประสงค์ของโครงการจากมุมมองทางธุรกิจ จากนั้นแปลงปัญหาให้อยู่ในรูปของโจทย์สำหรับการวิเคราะห์ข้อมูล และวางแผนการดำเนินงานเบื้องต้น

2) **การทำความเข้าใจข้อมูล (Data Understanding)**ขั้นตอนนี้เริ่มต้นด้วยการรวบรวมข้อมูล จากนั้นทำความเข้าใจ ตรวจสอบคุณภาพ และเลือกข้อมูลที่เก็บรวบรวมมาว่าจะใช้ข้อมูลใดบ้างในการวิเคราะห์

3) **การเตรียมข้อมูล (Data Preparation)**ขั้นตอนการเตรียมข้อมูลหมายถึง ขั้นตอนทั้งหมดที่จะทำเพื่อให้ข้อมูลดิบที่เรารวบรวมมา กลายเป็นข้อมูลสมบูรณ์ที่พร้อมจะเข้าสู่โมเดลในขั้นตอนที่ 4 เช่น การสร้างตาราง การลบข้อมูลที่ไม่ต้องการออก การแปลงข้อมูลให้อยู่ในรูปแบบที่ต้องการ

4) **การสร้างโมเดล (Modeling)** ในขั้นตอนนี้ เลือกและทดสองสร้างโมเดลหลายๆแบบที่น่าจะสามารถแก้ไขปัญหาที่ต้องการได้ จากนั้นค่อย ๆ ปรับค่าพารามิเตอร์ในแต่ละโมเดล เพื่อให้ได้โมเดลที่เหมาะสมที่สุดมาใช้ในการแก้ไขปัญหา

5) การวัดประสิทธิภาพของโมเดล (Evaluation) ทำการวัดประสิทธิภาพของโมเดลที่ได้จากขั้นตอนที่ 4 เพื่อวัดว่าโมเดลมีประสิทธิภาพเพียงพอต่อการนำไปใช้งานแล้วหรือไม่ ซึ่งโมเดลแต่ละประเภทก็จะมีตัววัดประสิทธิภาพที่แตกต่างกันออกไป

6) การนำโมเดลไปใช้งานจริง (Deployment) เป็นการนำโมเดลที่เหมาะสมที่สุดไปใช้งานจริง เพื่อวิเคราะห์และแก้ปัญหาที่ต้องการ

2.2.9 ทฤษฎีการประเมินผล (Evaluation)

ตัวชี้วัดการประเมินผล (*Evaluation Metrics*) คือเครื่องมือทางสถิติที่ใช้วัดความแม่นยำและประสิทธิภาพของโมเดลการพยากรณ์หรือโมเดลการถดถอย การเลือกตัวชี้วัดที่เหมาะสมมีความสำคัญอย่างยิ่ง เพราะโมเดลเดียวกันอาจมีผลลัพธ์ที่แตกต่างกันขึ้นอยู่กับเกณฑ์ที่ใช้วัด โดยตัวชี้วัดสำคัญหลายตัวในการประเมินผลโมเดลประเมินแบบจำลองการถดถอย ได้แก่ ค่าค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) สำหรับแต่ละตัวชี้วัด (Wohlwend, 2023; Hyndman & Athanasopoulos, 2018)

4) RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t-1}^{n}(y_t - \hat{y}_t)^2}$$

รากที่สองของ MSE ทำให้ค่ากลับมาอยู่ในหน่วยเดียวกับข้อมูลจริง อ่านค่าและตีความได้ง่ายกว่า MSE แต่ยังคงให้โทษหนักกับ error ขนาดใหญ่

5) MAE (Mean Absolute Error)

$$MAE = \frac{1}{n}\sum_{t-1}^{n}|y_t - \hat{y}_t|$$

ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย ไม่ยกกำลังสอง ไม่เน้นลงโทษความผิดพลาดใหญ่เหมือน MSE/RMSE เหมาะเมื่อข้อมูลมี outliers ที่ไม่อยากให้มีผลมากเกินไป

2.2.10 ทฤษฎีการแสดงผลข้อมูล (Data Visualization Theory)

เป็นการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในรูปแบบของรูปภาพ ซึ่งผู้ใช้สามารถเรียนรู้และจดจำ ข้อมูลผ่านการมองเห็นได้เป็นอย่างดีโดยเป็นการสร้างมโนภาพของสิ่งต่าง ๆ ที่สนใจ

ขึ้นมาเพื่อช่วยให้ เข้าใจได้ดีขึ้น ต่อมาได้มีการนำภาพมาใช้กับการนำเสนอหรืออธิบายกรอบความคิด เพื่อสนับสนุนการตัดสินใจ โดยในบทความของ Ware (Ware, 2004) กล่าวถึง Visualization มีข้อดี คือ

1) ช่วยในการแสดงข้อมูลที่มีปริมาณสูง

2) ช่วยในการรับรู้หรือคาดคะเนสิ่งที่อาจจะเกิดขึ้นในอนาคต

3) Visualization ไม่เพียงแต่แสดงรายละเอียดข้อมูลในตัวเองแต่ยังสามารถแสดงปัญหาที่ เกิดขึ้นได้ด้วย

4) ช่วยให้ผู้ใช้สามารถเข้าใจหรือวิเคราะห์ข้อมูลได้สะดวกขึ้นไม่ว่า ขอบเขตข้อมูลจะมีขนาดใหญ่หรือเล็ก

5) ช่วยให้ผู้ใช้สามารถสันนิษฐานข้อมูลได้สะดวกขึ้น

Diehl (Stephan Diehl, 2007)ได้สรุปไว้ว่า Visualization เป็นมากกว่าวิธีการทาง Computer Visualization เป็นการนำเสนอข้อมูลแบบหนึ่งของการแสดงข้อมูลให้อยู่ในรูปแบบของ Visual Form ซึ่งอาจจะเป็นการแสดงในรู ปแบบของรูปภาพ กราฟ หรือแผนภาพ ซึ่งผลลัพธ์ของการทำ Visualization คือการนำข้อมูลต่าง ๆ ที่ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้สามารถสัมผัสได้ ซึ่งกระบวนการในการแสดงข้อมูลต่าง ๆ อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือใช้ในการวิเคราะห์ข้อมูลโดยจุดมุ่งหมาย ของ Visualization ก็คือการถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ (อาทิตย์ สิทธิ บรรเจิด, 2552)

การจำแนกจัดกลุ่มของ Visualization สามารถทำโดยยึดตามกลุ่มหรือยึดตามโครงสร้างก็ได้โดย เลือกกลุ่มที่มีความหมายเหมือนกันหรือใกล้เคียงกัน จัดไว้ในกลุ่มเดียวกัน โดย Lengler (Lengler & Eppler, 2007; รุ่งรัศมี สุวรรณวัฒนา, 2558)ได้การสรุปรูปแบบออกเป็น 6 กลุ่ม ได้แก่

1) Data Visualization คือการแสดงข้อมูลในเชิงปริมาณ เช่น กราฟวงกลม กราฟพื้นที่ หรือ กราฟเส้น เป็นต้น กราฟเหล่านี้มีวัตถุประสงค์สำหรับรู้และเข้าใจภาพรวมของข้อมูลเป็นหลัก ได้แก่ Table, Pie chart, Bar chart, Area Chart, Histogram

2) Information Visualizationคือการใช้ภาพเพื่อเพิ่มความรู้ความเข้าใจในข้อมูลโครงสร้าง การทำงานของข้อมูลมากขึ้น เช่น โครงสร้างเครือข่ายเน็ตเวิร์ค ที่มีการเปลี่ยนแปลงอย่างต่อเนื่องตามการทำงานของผู้ใช้งาน เป็นต้น ได้แก่ TimeLine, TreeMaps, Flow–chart, Clustering, Parallel Coordinates

3) Concept Visualizationคือแผนผังหรือแผนภูมิแนวคิด ส่วนใหญ่จะเป็นข้อมูลเชิงคุณภาพ เช่น การคิดวางแผน การวิเคราะห์แล้วนำ มาแสดงในรูปแบบของแผนภาพ โดยจะนำเสนอ แนวคิดออกมาในรูปแบบของกล่องรูปภาพ มีเส้นเชื่อมโยงความสัมพันธ์เข้าด้วยกัน ได้แก่ Mindmap, Gantt-chart, CPM-critical-pathmethod, Flight-plan, Concept-map

4) Metaphor Visualizationคือ การเปลี่ยนแปลงภาพจากภาพหนึ่งไปยังอีกภาพหนึ่ง ทำให้เกิดความรู้ความเข้าใจในแนวคิดของข้อมูลและลักษณะโครงสร้างเชิงลึกได้ยิ่งขึ้น ได้แก่ Metro-map, Temple Story-template Tree

5) Strategy Visualizationคือ การแสดงแผนภาพเพื่อใช้ประกอบแนวคิด วิเคราะห์ และ ปรับปรุง รวมถึงแผนกลยุทธ์ในองค์กร สามารถดำเนินการตามแผนที่วางไว้ ได้แก่ Supplydemand-chain,Organizationchart Failure-tree, Portfolio-diagram, Decisiondiscoverydiagram 13

6) Compound Visualizationคือ การนำภาพหรือข้อความต่าง ๆ ที่มีความหลากหลายมา ประกอบกัน เป็นให้เกิดเป็นภาพเดียวได้แก่ Graphic-facilitation, Cartoon Knowledge-map, Cognitive-mapping

2.3 วรรณกรรมที่เกี่ยวข้อง

Giang และคณะ (2565). ได้ทดลองใช้ Random Forest Regression (RF) ซึ่งเป็นเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) สำหรับการพยากรณ์จำนวนแรงงานเวียดนามที่ออกไปทำงานต่างประเทศ ข้อมูลที่ใช้เป็นอนุกรมเวลารายเดือนยาวนานหลายปี โดยผู้วิจัยได้สร้างตัวแปรใหม่ในรูปแบบ lag features เช่น จำนวนแรงงานย้อนหลัง 1 เดือน, 2 เดือน และ 3 เดือน เพื่อนำมาใช้เป็นตัวแปรอิสระให้โมเดลเรียนรู้ความสัมพันธ์กับค่าปัจจุบัน จากนั้นได้สร้างและทดสอบแบบจำลอง Random Forest ควบคู่ไปกับโมเดล Machine Learning อื่น ๆ ได้แก่ k-Nearest Neighbors (kNN) และ Backpropagation Neural Network (BPNN) พร้อมทั้งประเมินผลด้วยตัวชี้วัด RMSE, MAE และค่าสัมประสิทธิ์กำหนด (R^2) ผลการทดลองพบว่าโมเดล BPNN มีความแม่นยำสูงสุด โดยให้ค่า MAPE เพียง 4–5% ขณะที่ kNN มีค่า MAPE ประมาณ 6–7% ส่วน Random Forest Regression แม้จะให้ผลด้อยกว่าสองโมเดลแรก แต่ก็ยังให้ค่า MAPE ~7–8% และมีค่า R^2 ประมาณ 0.85 ซึ่งถือว่ามีความแม่นยำและเสถียรในระดับที่ดี เมื่อเทียบกับโมเดลเชิงสถิติแบบดั้งเดิม จุดแข็งของ Random Forest คือความสามารถในการจัดการกับความสัมพันธ์ที่ไม่เชิงเส้น (non-linear) และทนทานต่อค่าผิดปกติ (outliers) ได้ดีกว่า งานวิจัยจึงสรุปว่าแม้ Random Forest จะไม่ใช่โมเดลที่แม่นยำที่สุด แต่ก็เป็น baseline ที่สำคัญของการ

ใช้ Machine Learning ในการพยากรณ์แรงงาน และสามารถใช้เปรียบเทียบกับโมเดลเชิงลึกเพื่อยืนยันผลการวิเคราะห์ได้อย่างน่าเชื่อถือ

Kim (2025). ได้ศึกษาและประยุกต์ใช้ LSTM Neural Network เพื่อพยากรณ์จำนวนตำแหน่งงานว่างในตลาดแรงงานสหรัฐ โดยใช้ข้อมูลจาก Job Openings and Labor Turnover Survey (JOLTS) ซึ่งเผยแพร่รายเดือนและครอบคลุมช่วงเวลากว่า 20 ปี ข้อมูลดังกล่าวถูกจัดให้อยู่ในรูปแบบลำดับเวลา (sequence) โดยใช้ข้อมูลย้อนหลัง 12 เดือนเป็น input ให้โมเดล LSTM ทำการเรียนรู้ความสัมพันธ์ทั้งในระยะสั้นและระยะยาว จากนั้นได้เปรียบเทียบผลลัพธ์กับแบบจำลองเชิงสถิติ ได้แก่ ARIMA, SARIMA และ Holt–Winters รวมทั้ง Random Forest ที่เป็น baseline ฝั่ง Machine Learning การประเมินใช้ตัวชี้วัด RMSE, MAE และ MAPE ผลการวิจัยชี้ว่า LSTM ให้ค่า MAPE เพียง 4.2–4.5% และ RMSE อยู่ที่ประมาณ 200,000–220,000 ซึ่งต่ำกว่า SARIMA (MAPE ~9–10%, RMSE ~400,000), Holt–Winters (MAPE ~12–14%, RMSE >500,000) และ Random Forest (MAPE ~6–7%, RMSE ~300,000) อย่างชัดเจน การลดค่าความคลาดเคลื่อนลงกว่า 30% เมื่อเทียบกับ SARIMA และ Holt–Winters แสดงถึงศักยภาพของ LSTM ในการเรียนรู้ pattern ที่ซับซ้อนและความไม่เชิงเส้นของข้อมูลแรงงาน งานวิจัยนี้สรุปว่า LSTM เป็นโมเดลที่มีความแม่นยำสูงสุดในบรรดาเทคนิคทั้งหมด และมีศักยภาพสูงสุดในการนำไปใช้พยากรณ์ตลาดแรงงานเชิงคาดการณ์เพื่อสนับสนุนการวางแผนเชิงยุทธศาสตร์ในอนาคต

Dawson และคณะ (2020). ได้ทำการศึกษาการพยากรณ์ภาวะการขาดแคลนทักษะในตลาดแรงงาน โดยมีวัตถุประสงค์เพื่อพัฒนาแบบจำลองเชิงพยากรณ์ที่สามารถทำนายสถานะการขาดแคลนทักษะในระดับอาชีพได้อย่างแม่นยำ งานวิจัยใช้ข้อมูลขนาดใหญ่ด้านอุปสงค์แรงงานจากประกาศรับสมัครงานออนไลน์จำนวนกว่า 7.7 ล้านรายการ ร่วมกับข้อมูลอุปทานแรงงานและตัวชี้วัดตลาดแรงงานอย่างเป็นทางการของประเทศออสเตรเลียในช่วงปี ค.ศ. 2012–2018 ตัวแปรอิสระประกอบด้วยข้อมูลด้านค่าจ้าง ชั่วโมงการทำงาน ระดับการศึกษา ประสบการณ์ทำงาน อัตราการจ้างงาน และตัวแปรเชิงโครงสร้างอื่น ๆ ขณะที่ตัวแปรตามเป็นสถานะของอาชีพที่อยู่ในภาวะขาดแคลนทักษะหรือไม่ ซึ่งจัดอยู่ในรูปแบบการจำแนกประเภท (Binary Classification) แบบจำลองหลักที่ใช้คือ XGBoost ซึ่งเป็นเทคนิค Ensemble Learning แบบ Gradient Boosting ที่มีความสามารถในการจัดการความสัมพันธ์ที่ไม่เชิงเส้นและข้อมูลจำนวนมากได้อย่างมีประสิทธิภาพ ผลการทดลองพบว่าแบบจำลอง XGBoost ให้ค่า Macro–F1 score สูงสุดประมาณร้อยละ 83 ซึ่งถือว่าอยู่ในระดับสูงเมื่อเทียบกับงานวิจัยด้านตลาดแรงงานในระดับอาชีพ นอกจากนี้ การวิเคราะห์ความสำคัญของตัวแปร (Feature

Importance) ยังพบว่าชั่วโมงการทำงาน ค่าจ้างเฉลี่ย ระดับการศึกษา และประสบการณ์ทำงาน เป็นปัจจัยที่มีอิทธิพลสูงต่อการพยากรณ์ภาวะการขาดแคลนทักษะ งานวิจัยสรุปว่าเทคนิค Machine Learning โดยเฉพาะ XGBoost สามารถเพิ่มความแม่นยำในการพยากรณ์และช่วยอธิบายโครงสร้างเชิงลึกของตลาดแรงงานได้ดีกว่าวิธีเชิงสถิติแบบดั้งเดิม และมีศักยภาพในการนำไปใช้สนับสนุนการกำหนดนโยบายด้านแรงงานและการศึกษาในระยะยาว

Kim (2025). ได้ศึกษาการพยากรณ์อัตราการว่างงานของสหรัฐอเมริกาโดยมุ่งเน้นการเปรียบเทียบประสิทธิภาพของแบบจำลองเชิงสถิติ แบบจำลองการเรียนรู้ของเครื่อง และแบบจำลองเชิงลึก (Deep Learning) งานวิจัยใช้ข้อมูลอนุกรมเวลารายเดือนตั้งแต่เดือนมกราคม ค.ศ. 2020 ถึงเดือนธันวาคม ค.ศ. 2024 โดยกำหนดตัวแปรอิสระจำนวน 30 ตัวแปร ซึ่งครอบคลุมตัวชี้วัดเศรษฐกิจมหภาค ตัวชี้วัดตลาดแรงงาน และดัชนีความเชื่อมั่นของผู้บริโภค แบบจำลองที่นำมาเปรียบเทียบประกอบด้วย Linear Regression, SGDRegressor, Random Forest, XGBoost, CatBoost, Support Vector Regression (SVR) และ Long Short–Term Memory (LSTM) ผลการประเมินประสิทธิภาพด้วยตัวชี้วัด RMSE, MAE และ MAPE พบว่าแบบจำลอง LSTM ให้ผลการพยากรณ์ดีที่สุด โดยสามารถลดค่าความคลาดเคลื่อนในการพยากรณ์ลงได้อย่างชัดเจนเมื่อเทียบกับแบบจำลองอื่น ขณะที่แบบจำลองกลุ่ม Ensemble Learning เช่น CatBoost และ XGBoost ให้ค่า RMSE และ MAE ต่ำกว่า Linear Regression และ SGDRegressor อย่างมีนัยสำคัญ ส่วน Random Forest และ SVR ให้ผลการพยากรณ์อยู่ในระดับปานกลาง งานวิจัยชี้ให้เห็นว่า LSTM มีความสามารถโดดเด่นในการเรียนรู้รูปแบบเชิงเวลา ความสัมพันธ์ระยะยาว และการเปลี่ยนแปลงเชิงโครงสร้างของตลาดแรงงาน ซึ่งเป็นข้อจำกัดสำคัญของแบบจำลองเชิงสถิติแบบดั้งเดิม งานวิจัยสรุปว่า Deep Learning โดยเฉพาะ LSTM เป็นแบบจำลองที่มีศักยภาพสูงสุดในการพยากรณ์อัตราการว่างงานในสภาวะที่ข้อมูลมีความผันผวนและมีความซับซ้อนสูง

สุรีนาฎ มะโนลา และคณะ (2568). ได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์อัตราการว่างงานของประเทศไทยในช่วงปี พ.ศ. 2564–2567 โดยใช้ข้อมูลรายจังหวัดจำนวน 77 จังหวัดจากกองบริหารข้อมูลตลาดแรงงาน แบบจำลองที่นำมาใช้ประกอบด้วยแบบจำลองเชิงสถิติและเชิง Machine Learning ได้แก่ ARIMA, Holt–Winters, K–Nearest Neighbors (K–NN) และ Linear Regression ผลการประเมินด้วยตัวชี้วัด MAE, RMSE และ MAPE พบว่าแบบจำลอง ARIMA ให้ค่า MAE ต่ำที่สุดเท่ากับ 6,658.515 และค่า RMSE ต่ำที่สุดเท่ากับ 8,578.801 แสดงถึงความแม่นยำและเสถียรภาพของการพยากรณ์ในเชิงตัวเลข

ขณะที่ Holt–Winters ให้ค่า MAPE ต่ำที่สุดประมาณร้อยละ 4 ซึ่งสะท้อนถึงความแม่นยำเชิงสัมพัทธ์ที่ดี อย่างไรก็ตาม แบบจำลอง K–NN ให้ค่า MAPE อยู่ในระดับปานกลางประมาณร้อยละ 6 แต่มีค่า RMSE สูงมากถึง 62,036.073 ซึ่งสะท้อนถึงความไม่เสถียรของแบบจำลอง ส่วน Linear Regression แม้จะให้ค่า MAPE เท่ากับร้อยละ 0 แต่กลับมีค่า RMSE และ MAE สูงผิดปกติ แสดงถึงปัญหาการเรียนรู้เกินข้อมูล งานวิจัยจึงสรุปว่าแบบจำลอง ARIMA ยังคงเป็นทางเลือกที่เหมาะสมสำหรับการพยากรณ์อัตราการว่างงานของประเทศไทยในบริบทที่ข้อมูลมีขนาดจำกัดและมีลักษณะเป็นอนุกรมเวลาอย่างชัดเจน

สุนันชณา ภูแก้วงาม และคณะ (2566). ได้ศึกษาการสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติในประเทศไทยโดยใช้เทคนิคเหมืองข้อมูลและการวิเคราะห์ทางสถิติ ข้อมูลที่ใช้เป็นข้อมูลอนุกรมเวลารายเดือนตั้งแต่ปี พ.ศ. 2554–2564 รวมระยะเวลา 132 เดือน โดยกำหนดตัวแปรอิสระด้านเศรษฐกิจและตลาดแรงงานจำนวน 9 ตัวแปร และใช้กระบวนการ CRISP–DM ในการพัฒนาแบบจำลอง แบบจำลองที่นำมาเปรียบเทียบ ได้แก่ Multiple Linear Regression, Decision Tree และ Support Vector Regression (SVR) ผลการประเมินประสิทธิภาพพบว่าแบบจำลอง SVR ให้ค่า RMSE เท่ากับ 15.7566 ค่า MAE เท่ากับ 7.7770 และมีค่า R^2 เท่ากับ 0.8847 ซึ่งอยู่ในระดับสูงเมื่อเทียบกับแบบจำลองอื่น งานวิจัยสรุปว่า SVR มีความสามารถในการเรียนรู้ความสัมพันธ์ที่ไม่เชิงเส้นได้ดี และมีความเหมาะสมในการนำไปใช้พยากรณ์จำนวนแรงงานข้ามชาติในประเทศไทยเพื่อสนับสนุนการกำหนดนโยบายด้านแรงงานในอนาคต

2.4 บทสรุป

จากการทบทวนแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง พบว่าการพยากรณ์จำนวนแรงงานไทยในต่างประเทศด้วยเทคนิคเหมืองข้อมูลและการเรียนรู้ของเครื่อง จำเป็นต้องอาศัยองค์ประกอบหลายประการร่วมกัน ตั้งแต่การทำความเข้าใจลักษณะของข้อมูล การเตรียมและปรับปรุงคุณภาพข้อมูล การเลือกตัวแปรที่เหมาะสม ไปจนถึงการเลือกแบบจำลองพยากรณ์ที่สอดคล้องกับโครงสร้างของข้อมูลและวัตถุประสงค์ของการวิจัย แนวคิดและทฤษฎีด้านอนุกรมเวลา (Time Series) รวมถึงแบบจำลองเชิงสถิติ ช่วยอธิบายความสัมพันธ์เชิงเส้นและแนวโน้มของข้อมูลในอดีต ขณะที่เทคนิคการเรียนรู้ของเครื่อง ได้แก่ Multiple Linear Regression ,Gradient Boosted Trees (GBT), Random Forest และ k–Nearest Neighbors (kNN) มีความสามารถในการจัดการกับความสัมพันธ์ที่ไม่เชิงเส้น ความซับซ้อนของข้อมูล และความผันผวนที่เปลี่ยนแปลงตามเวลาได้ดียิ่งขึ้น

นอกจากนี้ กรอบแนวคิดกระบวนการทำเหมืองข้อมูลตามมาตรฐาน CRISP–DM ยังช่วยให้การดำเนินการวิจัยเป็นไปอย่างเป็นระบบ มีขั้นตอนที่ชัดเจน ตั้งแต่การทำความเข้าใจปัญหา การทำความเข้าใจและเตรียมข้อมูล การสร้างแบบจำลอง การประเมินผล และการนำแบบจำลองไปใช้งานจริง ควบคู่กับการใช้ตัวชี้วัดในการประเมินประสิทธิภาพของแบบจำลอง เช่น ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (MAE) ซึ่งช่วยให้สามารถเปรียบเทียบความแม่นยำและความเหมาะสมของแบบจำลองแต่ละประเภท