บทที่ 2

แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องกับการสร้างโมเดลการพยากรณ์ ผลกระทบของปัจจัยทางสภาพภูมิอากาศ ต่อความเข้มข้นฝุ่นละอองขนาด 2.5 ไมครอน ในเขตอำเภอเมือง จังหวัดเชียงใหม่ โดยใช้การเรียนรู้ของเครื่อง ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง เพื่อใช้เป็นแนวทางการศึกษา ประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

2.1.2 แนวคิดเกี่ยบกับการพยากรณ์ข้อมูล (Forecasting data)

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

2.1.4 แนวคิดเกี่ยวกับการเปรียบเทียบโมเดล (Model Comparison Concepts)

2.1.5 แนวคิดเกี่ยวกับการการแบ่งข้อมูล (Data Splitting Concepts)

2.2 ทฤษฎี

2.2.1 ทฤษฎีการแสดงผล (visualization)

2.2.2 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

2.2.3 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง Java

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

2.3.2 โปรแกรม Weka

2.3.3 โปรแกรม Tableau

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

การวิเคราะห์ข้อมูล (Data Analytics) เป็นศาสตร์ที่มีความสำคัญอย่างยิ่งในศตวรรษที่ 21 เนื่องจากความสามารถในการนำข้อมูลมาวิเคราะห์เพื่อแก้ไขปัญหาหรือช่วยในการตัดสินใจทางธุรกิจ และวัตถุประสงค์อื่น ๆ ที่ต้องการ เริ่มต้นจากขั้นตอนพื้นฐานของการนำข้อมูลมาให้อยู่ในรูปแบบที่เหมาะสมสำหรับการประมวลผล ทั้งนี้ การวิเคราะห์ข้อมูลยังขึ้นอยู่กับเทคโนโลยีและชุดคำสั่งที่ใช้ รวมถึงแบบจำลองทางสถิติและปัญญาประดิษฐ์ที่ถูกสร้างขึ้นเพื่อใช้ในกระบวนการนี้ โดยผลลัพธ์จากการวิเคราะห์ข้อมูลที่ได้นั้นมีความสำคัญในการช่วยในการตัดสินใจและการใช้ประโยชน์จากข้อมูลเหล่านี้ นอกจากนี้ การวิเคราะห์ข้อมูลยังสามารถแบ่งออกเป็นหลายลักษณะตามวัตถุประสงค์และรูปแบบการวิเคราะห์ได้ดังนี้

1) การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive Analytics) การวิเคราะห์นี้เน้นการสร้างรายงานและกราฟเพื่อแสดงผลข้อมูลทางธุรกิจในลักษณะที่ง่ายต่อการเข้าใจ และช่วยในการตัดสินใจ โดยประยุกต์ใช้ข้อมูลปริมาณที่มีอยู่ เช่น รายงานการขายและผลการดำเนินงาน.
2) การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic Analytics) การวิเคราะห์นี้เน้นการค้นหาสาเหตุของเหตุการณ์และความสัมพันธ์ระหว่างปัจจัยหรือตัวแปรต่าง ๆ ที่มีผลต่อผลลัพธ์ โดยใช้ข้อมูลปริมาณที่มีอยู่ เช่น การวิเคราะห์ความสัมพันธ์ระหว่างยอดขายและกิจกรรมทางการตลาด.
3) การวิเคราะห์แบบพยากรณ์ (Predictive Analytics) การวิเคราะห์นี้เน้นการพยากรณ์สิ่งที่กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้น โดยใช้ข้อมูลปริมาณที่มีอยู่และแบบจำลองทางสถิติหรือปัญญาประดิษฐ์ เช่น การพยากรณ์ยอดขายหรือการพยากรณ์ผลประชามติ.
4) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive Analytics) การวิเคราะห์นี้เป็นการวิเคราะห์ข้อมูลที่ซับซ้อนที่สุด โดยไม่เพียงแค่พยากรณ์สิ่งที่จะเกิดขึ้น แต่ยังให้คำแนะนำทางเลือกต่าง ๆ และผลของแต่ละทางเลือก.

การนำแนวคิดและแบบจำลองต่าง ๆ ในการวิเคราะห์ข้อมูลทำให้เราสามารถเพิ่มความเข้าใจและการตัดสินใจที่ถูกต้องในด้านต่าง ๆ ของธุรกิจหรือหลายแวดวงตามวัตถุประสงค์ที่

ต้องการได้ ซึ่งการใช้ Data Analytics ที่ถูกต้องและมีประสิทธิภาพเป็นปัจจัยสำคัญในการสร้างความรู้และการแข่งขันในยุคปัจจุบันของธุรกิจและการเงิน.

2.1.2 แนวคิดเกี่ยบกับการพยากรณ์ข้อมูล (Forecasting data)

การพยากรณ์ (Forecasting) คือกระบวนการที่ใช้ข้อมูลปัจจุบันและอดีตเพื่อคาดคะเนหรือประมาณการเหตุการณ์หรือสิ่งหนึ่งที่จะเกิดขึ้นในอนาคต การพยากรณ์มีความสำคัญในการช่วยให้การตัดสินใจมีความถูกต้องและการวางแผนต่าง ๆ สำหรับอนาคตขององค์กรหรือการดำเนินงาน กระบวนการนี้อาศัยข้อมูลปัจจุบันและอดีตรวมทั้งการใช้วิจารณญาณ ความรู้และประสบการณ์ของบุคคลเพื่อกำหนดการประมาณการให้มีประโยชน์ในการตัดสินใจ.

การพยากรณ์ที่มีคุณภาพจะต้องมีลักษณะเด่นคือ:

1) ระบุวัตถุประสงค์ในการนำผลการพยากรณ์ไปใช้อย่างชัดเจนและกำหนดช่วงเวลาที่ต้องการในการพยากรณ์อย่างชัดเจน เพื่อเลือกใช้วิธีการที่เหมาะสมในการพยากรณ์.
2) การรวบรวมข้อมูลควรทำได้อย่างระบบและเรียบร้อย สามารถจำแนกชนิดและลักษณะของข้อมูลได้อย่างชัดเจนเพียงพอและตรงตามความเป็นจริงอย่างต่อเนื่อง เนื่องจากข้อมูลเป็นปัจจัยหลักในกระบวนการวิเคราะห์ตามวิธีที่เหมาะสมกับลักษณะข้อมูลนั้น.
3) การกำหนดนิยามเฉพาะของตัวแปรที่ใช้หรือเกี่ยวข้องในการพยากรณ์ รวมถึงการระบุข้อจำกัดและผลกระทบที่มีต่อการพยากรณ์.
4) การตรวจสอบความแม่นยำในการพยากรณ์อย่างต่อเนื่อง เนื่องจากสภาวการณ์ในช่วงเวลาต่าง ๆ อาจเปลี่ยนแปลง การตรวจสอบความแม่นยำในการพยากรณ์เป็นระยะจะเป็นตัวชี้วัดถึงวิธีการที่เหมาะสม.
5) การพยากรณ์ที่ดีจะต้องง่ายต่อการนำไปประยุกต์ใช้ สร้างความเข้าใจที่ตรงกันทั่วทั้งองค์กร การพยากรณ์ที่มีเป้าหมายที่แตกต่างกันตามลักษณะและการนำไปใช้

การพยากรณ์มีความสำคัญในการวางแผนธุรกิจ, การจัดการสต็อก, การวางแผนการผลิต, การบริหารคลังสินค้า, การวางแผนการทำงาน, และหลายกระบวนการอื่น ๆ ที่ต้องการการคาดคะเนหรือประมาณการที่แม่นยำในการตัดสินใจและการดำเนินงาน. การพิจารณาองค์ประกอบที่มีคุณภาพสูงในกระบวนการพยากรณ์เป็นสิ่งสำคัญที่จะช่วยให้ผลการพยากรณ์มีความเชื่อถือและสามารถนำไปใช้ในการวางแผนและการตัดสินใจได้อย่างมีประสิทธิภาพ.

2.1.3 แนวคิดเกี่ยวกับการเปรียบเทียบโมเดล (Model Comparison Concepts)

การเปรียบเทียบประสิทธิภาพระหว่างโมเดลวิเคราะห์ข้อมูลสองโมเดลคือกระบวนการที่ใช้เพื่อทดสอบและวัดผลประสิทธิภาพของโมเดลเพื่อหาว่าแบบจำลองใดมีประสิทธิภาพที่ดีกว่าในการแก้ไขปัญหาหรือทำนายผลข้อมูลที่ให้มา นี่คือขั้นตอนที่สำคัญในการวิเคราะห์ข้อมูลและการพัฒนาโมเดลสำหรับงานต่าง ๆ อย่างไรก็ตาม การเปรียบเทียบประสิทธิภาพของโมเดลอาจมีหลายแง่มุมต่าง ๆ และต้องพิจารณาในรูปแบบต่าง ๆ ตามลักษณะของงานและข้อมูลที่มีอยู่ด้วยกัน ดังนี้

1) ประสิทธิภาพในการทำนาย (Prediction Performance) : เมื่อเปรียบเทียบโมเดลสองระบบ โมเดลที่ให้ผลการทำนายที่แม่นยำและความคลาดเคลื่อนต่ำจะถือว่ามีประสิทธิภาพดีกว่า ค่าเชิงคณิตศาสตร์ที่ใช้ในการวัดอาจมีค่าเช่น Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2) หรืออื่น ๆ โดยทั่วไปค่าที่น้อยกว่าคือดีกว่า
2) ประสิทธิภาพในการจัดกลุ่ม (Classification Performance) : ในงานการจัดกลุ่ม (classification) การใช้ค่าคลาดเคลื่อนรวม (accuracy), precision, recall, F1-score, ROC-AUC, หรือค่าอื่น ๆ เป็นวิธีการวัดประสิทธิภาพที่พบบ่อย เพื่อประเมินความสามารถในการแยกแยะคลาสต่าง ๆ โมเดลที่ให้ผลดีในเรื่องนี้จะมีค่าใกล้เคียง 1 หรือใกล้เคียง 100% ในค่าที่เหมาะสม
3) ประสิทธิภาพในการใช้งาน (Operational Performance) : การประเมินประสิทธิภาพที่มีผลต่อการใช้งานจริงของโมเดล เช่น ความเร็วในการสร้างโมเดล (training time), ความเร็วในการทำนาย (inference time), และการใช้ทรัพยากรคอมพิวเตอร์ เป็นต้น โมเดลที่มีประสิทธิภาพดีแต่ใช้เวลานานหรือทรัพยากรมากอาจไม่เหมาะสมสำหรับงานบางประเภท
4) ประสิทธิภาพในการจัดการข้อมูล (Data Handling Performance) : การควบคุมความเสี่ยงและการจัดการข้อมูลที่หายไป (missing data) หรือข้อมูลที่ผิดปกติ (outliers) โดยการปรับปรุงและปรับแต่งข้อมูลอาจส่งผลต่อประสิทธิภาพของโมเดล
5) ความคงเส้นคงวง (Robustness) : โมเดลที่มีความสามารถในการทำงานแม้ในสภาวะที่ข้อมูลอาจมีความผิดปกติหรือเงื่อนไขที่ซับซ้อน หรือโมเดลที่ไม่ถูกกระทบโดยมากจาก

ความเปลี่ยนแปลงของข้อมูล (แยกจากการเทรนอาจเกิดการเรียนรู้เกินไปจากข้อมูลเฉพาะ)

6) การใช้งานจริง (Practical Usability) : การคำนึงถึงการทำงานในสภาพแวดล้อมที่จริง โดยพิจารณาปัญหาเชิงปฏิบัติที่อาจเกิดขึ้นในการใช้งานจริง เช่น ปัญหาความเร็วในการรับข้อมูลใหม่ หรือปัญหาความปลอดภัยและความเป็นส่วนตัว

ในการเปรียบเทียบประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลสองโมเดล ควรพิจารณาแง่มุมที่เหมาะสมกับงานและวัตถุประสงค์ของคุณ และใช้เครื่องมือทางสถิติและการทดสอบเพื่อวัดผลประสิทธิภาพที่ถูกต้องและทางวิเคราะห์ต่อไป

2.1.5 แนวคิดเกี่ยวกับการการแบ่งข้อมูล (Data Splitting Concepts)

การเลือกแบบการแบ่งข้อมูลที่เหมาะสม ขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของการวิเคราะห์ ยิ่งงานมีข้อมูลมากและซับซ้อนมาก เช่น งานที่เกี่ยวกับการเรียนรู้ของเครื่อง (Machine Learning) และ Deep Learning มักจะมีการใช้ Cross-Validation เพื่อประเมินโมเดลอย่างถูกต้อง ในทางอกจะมีการใช้ Holdout ในงานที่มีข้อมูลมากเพียงพอและมีเวลาจำกัด

การแบ่งข้อมูลสำหรับเทรนและทดสอบโมเดลการวิเคราะห์ข้อมูลมีหลายแบบ แต่ในงาน Machine Learning (ML) โดยทั่วไปมี 3 แบบหลักคือ Holdout, Cross-Validation, และ Time Series Split:

1. **การแบ่งข้อมูลแบบ Holdout (Train-Validation-Test Split):**
   1) **วิธีการทำ:** แบ่งข้อมูลเป็นชุด Train (สำหรับเทรนโมเดล), Validation (สำหรับปรับปรุงและวิเคราะห์โมเดล), และ Test (สำหรับทดสอบประสิทธิภาพ) โดยส่วนใหญ่ใช้สัดส่วน เช่น 70-15-15%.
   2) **ข้อดี:** ง่ายต่อการใช้งาน, เหมาะสำหรับการทดลองต่าง ๆ และงานที่มีข้อมูลมาก
   3) **ข้อเสีย:** อาจทำให้การประเมินโมเดลขึ้นอยู่กับการสุ่มแบ่งข้อมูล
2. **การแบ่งข้อมูลแบบ Cross-Validation (k-fold Cross-Validation):**
   1) **วิธีการทำ:** แบ่งข้อมูลเป็น k ชุดเทรนและทดสอบที่ไม่ซ้ำกัน แล้วประเมินโมเดล k ครั้ง โดยใช้ข้อมูลทดสอบแต่ละชุดเป็นครั้งกลุ่ม (fold) เพื่อประเมินค่าเฉลี่ย.

2) **ข้อดี:** ช่วยลดความเสี่ยงในการสูญเสียข้อมูล, ค่าประสิทธิภาพมีความเชื่อถือมากขึ้น
3) **ข้อเสีย:** มีความซับซ้อนและใช้เวลามากมาย

3. **การแบ่งข้อมูลแบบ Time Series Split:**
    1) **วิธีการทำ:** ใช้ในกรณีข้อมูลมีแนวเวลา โดยแบ่งข้อมูลเป็นชุด Train และชุด Test โดยให้ชุด Test อยู่หลังชุด Train.
    2) **ข้อดี:** เหมาะสำหรับข้อมูลแนวเวลา, ช่วยป้องกันการหลุดทิ้งข้อมูลที่สำคัญ
    3) **ข้อเสีย:** อาจไม่เหมาะสำหรับข้อมูลที่ไม่เกี่ยวข้องกับเวลา

การเลือกแบบการแบ่งข้อมูลที่เหมาะสำหรับงานวิจัย Machine Learning เกี่ยวกับปัจจัยสภาพอากาศจะขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของการวิเคราะห์

2.2 ทฤษฎี

2.2.1 ทฤษฎีการ visualization

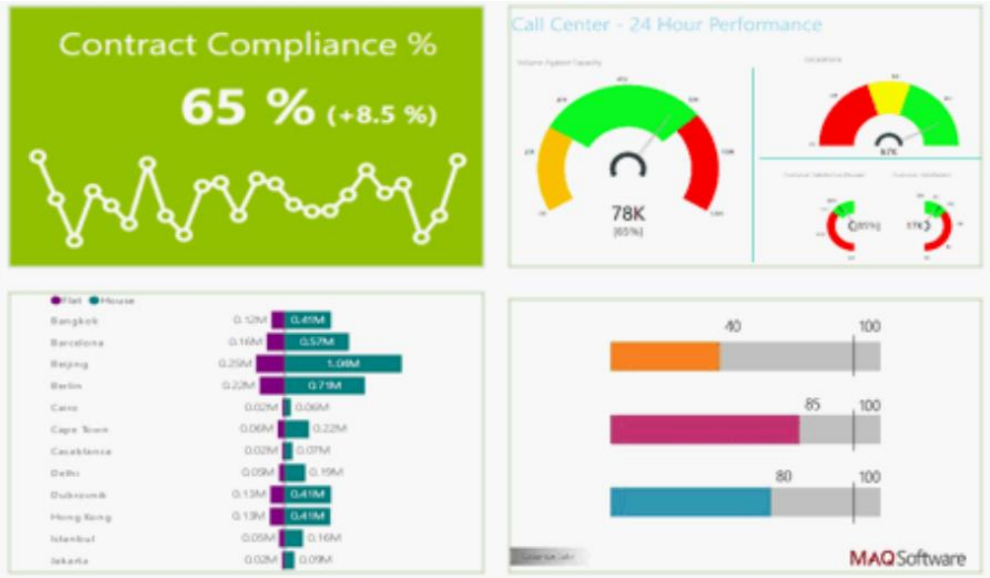
การวิเคราะห์ข้อมูล (Data) เพื่อนำมาใช้ในการตลาดยุคดิจิทัลถือว่าเป็นสิ่งที่สำคัญมากในการทำธุรกิจให้ประสบความสำเร็จในยุคปัจจุบัน ผู้ประกอบการที่ทำธุรกิจออนไลน์รู้ดีว่าการนำข้อมูลมาวิเคราะห์และใช้ให้เกิดประโยชน์ทำให้ธุรกิจได้เปรียบมากกว่าคู่แข่งในตลาด ข้อมูลที่นำมาประมวลผลสามารถแปลงเป็นรูปภาพ กราฟต่าง ๆ หรือแม้กระทั่งวิดีโอเพื่อนำเสนอผลงานในรูปแบบต่าง ๆ ทำให้เกิดการจดจำได้ง่ายกว่าการอ่านข้อมูลแบบตารางทั่วไป สิ่งนี้เรียกว่า Data Visualization คือการนำข้อมูลที่มีความซับซ้อนมาวิเคราะห์และแสดงผลในรูปแบบของแผนภูมิ กราฟและรูปแบบอื่น ๆ ที่หลากหลาย เพื่อให้ผู้ใช้งานเข้าใจข้อมูลได้ง่ายและน่าสนใจมากขึ้น ซึ่งเหมาะสำหรับการนำเสนอข้อมูลในธุรกิจและการตลาด

ภาพที่ 2.1 การนำเสนอข้อมูลแบบแนวโน้ม

การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending) ใช้กราฟที่แสดงผลแบบทิศทางหรือแนวโน้มเพื่อนำเสนอข้อมูลให้เห็นจำนวนข้อมูลที่เกิดขึ้นในแต่ละช่วงเวลา (period) โดยเน้นข้อมูลที่ต้องการนำเสนอ เช่น Line Chart, Bar Chart, Radar Chart, Area Chart เป็นต้น

ภาพที่ 2.2 การนำเสนอข้อมูลแบบกลุ่มข้อมูล

การนำเสนอแบบกลุ่มข้อมูล (Classification) การนำเสนอโดยการนำข้อมูลมาจัดเป็นกลุ่มๆ เช่น Donut Chart, Ring Chart, Pie Chart เพื่อให้ผู้ใช้งานเข้าใจข้อมูลในรูปแบบกลุ่ม



ภาพที่ 2.3 การนำเสนอข้อมูลเชิงเปรียบเทียบ

การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison) เหมาะสำหรับการนำเสนอที่ต้องการเปรียบเทียบผลการดำเนินงาน เช่นการเปรียบเทียบกับปีที่แล้ว (YoY) หรือการเปรียบเทียบกับเป้าที่ตั้งไว้ (Target) ซึ่งใช้กราฟเช่น KPI Indicator, Bullet Chart, Power BI Card with state

ภาพที่ 2.4 การนำเสนอข้อมูลแบบแผนที่

การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับการนำเสนอข้อมูลบนแผนที่ เช่นการแสดงยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อเน้นกลุ่มลูกค้าในแต่ละพื้นที่ที่สนใจ เช่น Globe Map, Google Map, Flow Map เป็นต้น

2.2.2 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

ภาษา HTML (Hyper Text Markup Language) เป็นภาษาที่ใช้ในการเขียนโปรแกรมและแสดงข้อมูลบนเว็บไซต์ ภาษา HTML ใช้ในการกำหนดรูปแบบและโครงสร้างของเนื้อหาที่แสดงบนเว็บ เช่น ข้อความ รูปภาพ เสียง และวิดีโอ ภาษา HTML เป็นภาษาที่ง่ายต่อการเรียนรู้และใช้งาน สามารถกำหนดรูปแบบและโครงสร้างได้อย่างง่าย และมีการพัฒนาต่อเนื่องเพื่อให้ใช้งานง่ายขึ้นและรองรับการแสดงผลในเว็บบราวเซอร์หลายรูปแบบ ภาษา HTML มักจะถูกบันทึกในรูปของไฟล์นามสกุล .html หรือ .htm

คำสั่ง (Tag) เป็นส่วนหนึ่งของ HTML ซึ่งใช้ในการกำหนดรูปแบบของข้อมูลที่แสดง โดยมีรูปแบบ <...> เพื่อแสดงว่าเป็นคำสั่ง HTML โดย HTML ประกอบด้วยสองส่วนหลัก:

1. ส่วนของคำสั่ง (Tag): เป็นส่วนที่กำหนดรูปแบบของข้อมูลที่แสดง ซึ่งเราเรียกว่า Tag และจะอยู่ในเครื่องหมาย <...>
2. ส่วนของบทความทั่ว ๆ ไป: เป็นส่วนของข้อมูลที่เราต้องการแสดงผล

ตัวอย่างการใช้งาน HTML คือการเริ่มต้นของเอกสาร HTML โดยใช้คำสั่ง <HTML> เพื่อเริ่มต้นการเขียนโปรแกรมและใช้คำสั่ง </HTML> เพื่อสิ้นสุดโปรแกรม HTML คำสั่งนี้จะไม่แสดงผลในโปรแกรมเว็บเบราเซอร์ แต่จำเป็นต้องเขียนเพื่อให้เกิดความเป็นระบบในงานและเพื่อระบุว่าเอกสารนี้เป็นเอกสารของภาษา HTML

ส่วนหัวเรื่องของเอกสาร HTML (Head Section) ใช้ในการระบุข้อมูลเกี่ยวกับหน้าเว็บ แบ่งเป็นส่วนของคำสั่ง <HEAD> ... </HEAD> และส่วนของบทความทั่ว ๆ ไป (Body Section) ที่ใช้ในการแสดงเนื้อหาหลักของหน้าเว็บ โดยแสดงผลข้อมูลที่ต้องการนำเสนอด้วย Tag ต่าง ๆ ตามลักษณะของข้อมูล การป้อนคำสั่งในส่วนนี้ไม่มีข้อจำกัดและสามารถป้อนติดกันหรือ 1 บรรทัดต่อ 1 คำสั่งก็ได้ แต่ส่วนใหญ่จะยึดรูปแบบที่ทำให้ง่ายต่อการอ่านโค้ด คือการทำย่อหน้าในชุดคำสั่งที่เกี่ยวข้องกัน ทั้งนี้ให้ป้อนคำสั่งทั้งหมดภายใต้ Tag ที่เหมาะสม

2.2.3 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS (Cascading Style Sheet) เป็นส่วนสำคัญที่ใช้ในการกำหนดรูปแบบและสไตล์ของเว็บเพจ โดย CSS ช่วยให้คุณสามารถควบคุมการแสดงผลของเอกสาร HTML ได้อย่างง่ายและมีประสิทธิภาพ นี่คือบางแนวคิดสำคัญในทฤษฎีเกี่ยวกับ CSS:

1) คำสั่ง CSS: CSS ใช้กลุ่มคำสั่งหรือ Selector เพื่อเลือกองค์ประกอบบนหน้าเว็บที่คุณต้องการจะกำหนดรูปแบบและสไตล์ เช่น คลาส (Class), ไอดี (ID), และแท็ก HTML โดย Selector ที่ใช้เรียกว่าเลือกเอกสารที่คุณต้องการแก้ไขสไตล์
2) คุณสมบัติ (Properties) และค่า (Values): ใน CSS, คุณสมบัติเป็นรายการของสไตล์ที่คุณต้องการให้เอกสาร HTML มี เช่น สีของตัวอักษร, สีพื้นหลัง, ขนาดตัวอักษร, และอื่น ๆ คุณสมบัติแต่ละอันจะมีค่าที่เกี่ยวข้องกับสไตล์นั้น ๆ
3) การกำหนดคุณสมบัติและค่า: คุณสมบัติและค่าใน CSS จะถูกกำหนดในรูปแบบ properties: value โดยคั่นด้วยเครื่องหมาย colon (:), และคุณสมบัติแต่ละคู่จะถูกคั่นด้วยเครื่องหมาย semi-colon (;) เช่น codecolor: blue; background-color: #FFFFFF;
4) กลไกการสรรหา: ในกรณีที่มี Selector หลายตัวที่กำหนดคุณสมบัติเดียวกันให้องค์ประกอบ HTML คุณสามารถใช้กลไกการสรรหาเพื่อกำหนดลำดับความสำคัญและลำดับการใช้งานของรูปแบบ CSS โดยใช้เกณฑ์พิเศษ

CSS เป็นเครื่องมือที่สำคัญสำหรับการออกแบบและสร้างเว็บไซต์ที่มีสไตล์และมีความสวยงาม การเรียนรู้ CSS จะช่วยให้คุณสามารถปรับแต่งเว็บไซต์ของคุณให้ดูดีและมีความสมดุลในการแสดงผลหน้าเว็บของคุณ.

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง Java

ภาษา Java เป็นภาษาโปรแกรมมิ่งที่มีความนิยมและถูกใช้งานอย่างแพร่หลายในโลกของพัฒนาซอฟต์แวร์ มันถูกสร้างขึ้นโดย Sun Microsystems (ที่ถูกซื้อโดย Oracle Corporation เมื่อภายหลัง) และถูกเปิดตัวครั้งแรกในปี 1995 ภาษา Java มีความพิเศษอย่างหนึ่งคือความแพล็ตฟอร์มอิสระ ซึ่งหมายความว่าโปรแกรมที่เขียนด้วย Java สามารถทำงานบนหลายระบบปฏิบัติการและอุปกรณ์ที่มี Java Virtual Machine (JVM) ที่ติดตั้งอยู่ โดยไม่จำเป็นต้องเขียนโค้ดใหม่สำหรับแต่ละระบบปฏิบัติการนั้นๆ

Java เป็นภาษาโปรแกรมมิ่งที่มีการจัดการหน่วยความจำและการจัดการข้อผิดพลาดที่เข้มงวด มีความปลอดภัยสูง และมีการจัดระเบียบโค้ดอย่างมีระเบียบ มันมีการสนับสนุนการเขียนโปรแกรมแบบออบเจ็กต์ (Object-Oriented Programming) และมีนวัตกรรมที่ช่วยให้สามารถพัฒนาแอปพลิเคชันขนาดใหญ่และซอฟต์แวร์ระบบได้อย่างมีประสิทธิภาพ นอกจากนี้ มีนวัตกรรม Java EE (Enterprise Edition) ที่ช่วยให้สามารถพัฒนาแอปพลิเคชันขนาดใหญ่และที่มีความซับซ้อนได้อีกด้วย

Java มีความสามารถในการใช้ในการพัฒนาแอปพลิเคชันเบาๆ และแอปพลิเคชันที่ต้องการประสิทธิภาพสูง เช่น แอปพลิเคชันบนเว็บ, แอปพลิเคชันมือถือ, ระบบเกม, ระบบควบคุมการเชื่อมต่อเครือข่าย และอื่น ๆ อีกมาก ซึ่งทำให้ Java เป็นภาษาโปรแกรมมิ่งที่มีความหลากหลายและสามารถใช้งานได้กับแอปพลิเคชันในหลายด้านของเทคโนโลยีและธุรกิจได้อย่างกว้างขวาง.

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

กระบวนการทำเหมืองข้อมูล (CRISP–DM) คือกระบวนการที่ถูกออกแบบมาเพื่อช่วยในการวิเคราะห์และการนำข้อมูลมาใช้ในการตัดสินใจ โดยมีขั้นตอนหลัก 6 ขั้นตอนดังนี้:

1. Business Understanding (เข้าใจธุรกิจ)
   1) ในขั้นตอนนี้คุณต้องเริ่มต้นด้วยการเข้าใจกับธุรกิจหรือประเภทของปัญหาที่คุณต้องการแก้ไขด้วยการทำเหมืองข้อมูล
   2) ระบุวัตถุประสงค์ของการทำเหมืองข้อมูลและคำถามที่คุณต้องการให้การทำเหมืองข้อมูลตอบ
2. Data Understanding (เข้าใจข้อมูล)
   1) ในขั้นตอนนี้คุณจะต้องรวบรวมข้อมูลที่เกี่ยวข้องกับปัญหาหรือวัตถุประสงค์ของคุณ
   2) ทำการสำรวจข้อมูล, ทราบลักษณะและโครงสร้างของข้อมูล รวมถึงการตรวจสอบคุณภาพข้อมูล
3. Data Preparation (เตรียมข้อมูล)
   1) ในขั้นตอนนี้คุณจะทำการคัดเลือกข้อมูลที่สำคัญและมีคุณภาพจากข้อมูลทั้งหมด
   2) การทำความสะอาดข้อมูล, การแปลงรูปแบบข้อมูล, และการสร้างชุดข้อมูลสำหรับการทดสอบและการฝึกอบรมโมเดล
4. Modeling (สร้างโมเดล)
   1) ในขั้นตอนนี้คุณจะสร้างแบบจำลอง (model) เพื่อทำการวิเคราะห์ข้อมูล
   2) การเลือกและการปรับแต่งอัลกอริทึมที่ใช้ในการสร้างโมเดล, การแบ่งข้อมูลเพื่อทดสอบและการฝึกอบรมโมเดล
5. Evaluation (ประเมิน)
   1) ในขั้นตอนนี้คุณจะประเมินความสามารถของโมเดลที่สร้างขึ้นโดยใช้ชุดข้อมูลทดสอบ (test data)
   2) การประเมินผลลัพธ์ของโมเดล, การวัดประสิทธิภาพ, และการทำการปรับปรุงหรือเลือกโมเดลที่ดีที่สุด

6. Deployment (นำไปใช้)

   1) ในขั้นตอนสุดท้ายคุณจะนำโมเดลหรือผลลัพธ์จากการทำเหมืองข้อมูลไปใช้ในการตัดสินใจหรือการวางแผนธุรกิจ
   2) การนำเอาการทำเหมืองข้อมูลมาใช้ในการเข้าใจลูกค้า, การปรับแผนการตลาด, หรือการปรับปรุงกระบวนการทางธุรกิจ

กระบวนการ CRISP–DM เป็นกระบวนการที่เป็นระเบียบและถูกใช้งานอย่างแพร่หลายในการวิเคราะห์ข้อมูลและการทำเหมืองข้อมูล เพื่อช่วยในการตัดสินใจและการวางแผนธุรกิจอย่างมีเสถียรภาพและมีประสิทธิภาพ.

2.4 วรรณกรรมที่เกี่ยวข้อง

กานต์กมล ทวีผล (2563) ได้ทำการศึกษาและวิเคราะห์เปรียบเทียบสมรรถนะระหว่างแบบจำลอง LSTM กับแบบจำลอง SARIMAX โดยผู้วิจัยได้เลือกศึกษาข้อมูลจากบริเวณพื้นที่สถานีตำรวจนครบาลโชคชัย และนำชุดข้อมูลดังกล่าวมาใช้ในการสร้างแบบจำลอง LSTM และ SARIMAX โดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อเป็นแนวทางในการพัฒนาการสร้างแบบจำลองในการทำนายปริมาณความหนาแน่นของฝุ่นละออง PM2.5 ในอีก 24 ชั่วโมงข้างหน้า โดยวัดประสิทธิภาพของแต่ละแบบจำลอง

จากผลการทดลองสรุปได้ว่าแบบจำลองที่มีประสิทธิภาพดีที่สุดคือแบบจำลองโครงข่ายประสาทเทียมแบบลึก LSTM ซึ่งได้ค่า RMSE = 6.04 และ MAE = 4.86

ขนิษฐา ชัยรัตนาวรรณ, ณัฐพศุตม์ ภัทธิราสินสิริ (2563) บทความวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาแหล่งกำเนิด ผลกระทบและแนวทางจัดการฝุ่นละออง PM2.5 บริเวณภาคเหนือของประเทศไทย ใช้วิธีการวิจัยเชิงเอกสาร โดยการสังเคราะห์ บทความวิจัยที่อยู่ในฐานข้อมูล TCI ISI และรายงานและเอกสารที่เผยแพร่โดยหน่วยงานภาครัฐ รวมทั้งหมด 15 เรื่องที่คณะผู้วิจัยได้คัดเลือกอย่างเฉพาะเจาะจง ใช้วิเคราะห์ข้อมูลในรูปแบบการวิเคราะห์เนื้อหา

ผลการสังเคราะห์พบว่า แหล่งกำเนิดของฝุ่นละออง PM2.5 อันดับแรกมาจากการเผาในที่โล่ง รองลงมาเกิดขึ้นจากไอเสียรถโดยเฉพาะดีเซล

ปรัชญา สิงหวรวงศ์ (2563)ในการวิจัยการพยากรณ์ความเข้มข้นฝุ่นละออง PM2.5 ใน 24 ชั่วโมงล่วงหน้า ซึ่งใช้ข้อมูล ความเข้มข้นฝุ่นละออง PM2.5 บริเวณพื้นที่สถานีศูนย์ราชการจังหวัดเชียงใหม่ และสถานีโรงเรียน ยุพราชวิทยาลัย โดยใช้เทคนิคการเรียนรู้เชิงลึก

แบบจำลองหน่วยความจำระยะสั้นแบบยาว (LSTM) ผู้วิจัยได้วัดความแม่นยำแบบจำลองของแต่ละสถานี และปรับปรุงประสิทธิภาพของ แบบจำลองในการพยากรณ์ให้แม่นยำมากขึ้นด้วยการเปรียบเทียบการเลือก Optimizer ทั้ง 3 แบบ ได้แก่ AdaGrad, RMSprop และ Adam เพื่อให้แบบจำลองมีประสิทธิภาพสูงสุดควรเลือกใช้ Optimizer ที่เหมาะสมกับแบบจำลอง และชุดของข้อมูล

จากการวิจัยพบว่า แบบจำลอง LSTM มีประสิทธิภาพในการพยากรณ์ค่าความเข้มข้น PM2.5 ได้อย่าง แม่นยำโดยแตกต่างกันที่การเลือกใช้ Optimizer ในชุดข้อมูลสถานีศูนย์ราชการจังหวัดเชียงใหม่ เป็นชุดข้อมูลที่มีค่าข้อมูลสูญหายไม่มาก ควรเลือกใช้ Optimizer เป็น Adam หรือ AdaGrad จะ ทำให้สามารถพยากรณ์ล่วงหน้า 24 ชั่วโมงได้ดีที่สุด ในส่วนชุดข้อมูลสถานีโรงเรียนยุพราชวิทยาลัยเป็นชุดข้อมูลที่มีค่าข้อมูลสูญหายมาก หรือร้อยละ 18.8 แบบจำลองควรเลือก Optimizer เป็น RMSprop จะทำให้สามารถพยากรณ์ล่วงหน้า 24 ชั่วโมงได้ดีที่สุด

อัญชลี พงศ์เกษตร และคณะ (2563) ปัจจัยทางสภาพภูมิอากาศที่มีความ สัมพันธ์กับปริมาณความเข้มข้นฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน: PM10 (Particulate Matter < 10 microns) และประเมินความเสี่ยงทางสุขภาพของประชาชนที่สัมผัส PM10 อำเภอเมือง จังหวัดยะลา โดยใช้ข้อมูลทุติยภูมิจากกรมควบคุมมลพิษ เพื่อหาปริมาณ PM10 และจากสถานีอุตุนิยมวิทยา จังหวัดยะลา เพื่อศึกษาสภาพภูมิอากาศ ศึกษาข้อมูลตั้งแต่ปี 2551 – 2560 หาความสัมพันธ์ด้วยสถิติสัมประสิทธ์สหสัมพันธ์ของเพียร์สัน วิเคราะห์การถดถอยพหุคูณแบบขั้นตอน โดยศึกษาตัวแปรทั้งหมด 5 ตัวแปร คือ 1. ปริมาณความเข้มข้น PM10 2. ปริมาณน้ำฝน 3. ความเร็วลม 4. อุณหภูมิ 5. ความชื้นสัมพัทธ์

จากการศึกษาพบว่า ความชื้นสัมพัทธ์ มีความสัมพันธ์ในทิศทางตรงกันข้ามกับปริมาณความเข้มข้น PM10 ($r = -0.214$, $P < 0.05$) และ สามารถทำนายปริมาณความเข้มข้น PM10 ได้ ร้อยละ 4.6 ($P < 0.05$) และเมื่อจำแนกตามฤดูกาล พบว่า ความชื้นสัมพัทธ์ในฤดูฝนสามารถทำนาย ปริมาณความเข้มข้น PM10 ในฤดูฝนได้ร้อยละ 11.4 ($P < 0.05$)

ชชนิดาภา วินาลัย และคณะ (2565) งานวิจัยนี้มีการเก็บรวบรวมข้อมูลตามพื้นที่ต่างๆ ซึ่งประกอบไปด้วยกรุงเทพมหานครนครปฐม นนทบุรี ปทุมธานี สมุทรปราการ และสมุทรสาคร มาใช้ในการศึกษา โดยใช้ข้อมูลค่าความ เข้มข้นของฝุ่นละออง PM2.5 รายเดือนตั้งแต่ปี พ.ศ. 2558 – 2563 จากกรมควบคุมมลพิษ ทั้งหมด 23 สถานี(Air Quality and Noise Management Bureau, 2020) และนำมาหาค่าเฉลี่ย รายเดือน และใช้ข้อมูลตัวแปรทางสภาพ

อากาศจาก Global Surface Summary of the Day (GSOD) ผู้วิจัยเลือกตัวแปรทางสภาพอากาศได้แก่ อุณหภูมิ ปริมาณน้ำฝน ความชื้นสัมพัทธ์ ความเร็วลม และความกดอากาศ โดยใช้แบบจำลองหน่วยความจำระยะสั้นแบบยาว(LSTM) เพื่อทำนายหาปริมาณความเข้มข้นฝุ่นละออง และศึกษาความสัมพันธ์กับตัวแปรทางสภาพอากาศ โดยนำข้อมูลจากกรมควบคุมมลพิษในพื้นที่กรุงเทพฯ และปริมณฑลมาใช้ในการศึกษา

ผลการทำนายของตัวแปรทางสภาพอากาศทั้งหมดพบว่า อุณหภูมิและ ปริมาณน้ำฝนมีความสัมพันธ์กันและสามารถให้ผลทำนายได้ดีที่สุด

ชาคริต โชติอมรศักดิ์ , ดวงนภา ลาภใหญ่ (2561) การวิเคราะห์ความสัมพันธ์ระหว่างปัจจัย ทางอุตุนิยมวิทยากับความเข้มข้น PM10 ที่สถานี ตรวจวัดศูนย์ราชการจังหวัดเชียงใหม่ด้วยวิธีการ วิเคราะห์ความสัมพันธ์จากเมตริกซ์ของกราฟ การกระจายในตัวแปรแต่ละคู่ และสัมประสิทธิ์ สหสัมพันธ์เพียร์สัน พบว่า ปัจจัยอุณหภูมิสูงสุด รายวันมีความสัมพันธ์เชิงบวกกับความเข้มข้น PM10 ส่วนปัจจัยอื่น ๆ ได้แก่ปริมาณน้ำฝน อุณหภูมิต่ำสุด ความเร็วลม ความชื้นสัมพัทธ์ และ ความกดอากาศ มีความสัมพันธ์เชิงลบกับความ เข้มข้น PM10 โดยที่ทุกปัจจัยยังมีค่าสัมประสิทธิ์ สัมพันธ์ไม่สูงมากนัก และข้อมูลมีลักษณะการ กระจัดกระจายแบบกลุ่ม

ผลการวิเคราะห์ เบื้องต้นได้ข้อสรุปว่าข้อมูลมีลักษณะการกระจัด กระจายแบบกลุ่ม วิธีการวิเคราะห์องค์ประกอบ หลัก (PCA) จึงนำมาใช้ศึกษาความสัมพันธ์ระหว่าง ปัจจัยทางอุตุนิยมวิทยากับความเข้มข้น PM10 ในลำดับต่อไป ซึ่งผลการวิเคราะห์องค์ประกอบ หลักพบว่า องค์ประกอบหลักที่ 1 มีปัจจัยอุณหภูมิ และความชื้นสัมพัทธ์ที่สามารถนำมาอธิบายสัมพันธ์ ระหว่างความเข้มข้น PM 10 ได้เป็นอย่างดี

กฤติกา ทิพย์คำมี และ คณะ (2566) เปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่อง สำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) โดยตัวแปรที่ใช้ในการสร้างตัวแบบฝุ่น PM2.5 จำนวน 16 ตัวแปร จากสถานีตรวจวัดคุณภาพอากาศและอุตุนิยมวิทยา ของกรมควบคุม มลพิษ จังหวัดอุดรธานี โดยนำมาวิเคราะห์ตามกระบวนการวิเคราะห์ข้อมูล เพื่อสร้างตัว แบบพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ด้วยเทคนิคการประมาณค่าข้อมูล (Regression Model) โดยในงานวิจัยในครั้งนี้ได้เลือกใช้เทคนิคการเรียนรู้ของเครื่องทั้งหมด 4 เทคนิค ได้แก่ เทคนิคการ ถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning)

จากผลการวิจัยพบว่าเทคนิคที่มีความเหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ ฝุ่นละอองขนาดเล็กในอากาศ คือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) โดยนี้สามารถอธิบายการเปลี่ยนแปลงของการเกิดฝุ่นละอองขนาดเล็กในอากาศ ได้ถึง 94.8%

2.5 สรุป

จากการศึกษาวรรณกรรมที่เกี่ยวข้อง พบว่าปัญหาของ PM 2.5 ส่วนใหญ่มาจากการเผาในที่โล่ง และ การเผาไหม้ของเครื่องยนต์ดีเซลตามลำดับ ปัจจัยทางสภาพอากาศที่มีผลต่อความเข้มข้นของ PM 2.5 คือปัจจัยด้าน ปริมาณน้ำฝน และ ความชื้นสัมพัทธ์ในอากาศ การทำนายค่าความเข้มข้นของฝุ่น PM 2.5 โดยใช้ปัจจัยทางสภาพอากาศ มีค่าความน่าเชื่อถือ และความแม่นยำ อยู่ในระดับที่สูง การใช้โมเดล LSTM ในการทำนายค่าความเข้มข้นของฝุ่น PM 2.5 สามารถททำนายได้อย่างมีประสิทธิภาพดีที่สุด เมื่อเทียบกับโมเดลในการทำนายอื่น ๆ