บทที่ 2

แนวคิด ทฤษฎีเอกสารที่เกี่ยวข้อง

โครงงานเรื่อง การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการเปรียบเทียบแบบจำลองที่เหมาะสมที่สุด เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ในบทนี้เป็นการนำเสนอ เกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการเปรียบเทียบแบบจำลองที่เหมาะสมที่สุด เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้ เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับดังนี้

2.1 แนวคิด

2.1.1 การวิเคราะห์ข้อมูล (Data analytics)

2.1.2 การพยากรณ์ข้อมูล (Forecasting data)

2.1.3 การเปรียบเทียบโมเดล (Model Comparison Concepts)

2.1.4 การทำความสะอาดข้อมูล (Data Cleaning)

2.1.5 การแสดงผลข้อมูล (Data visualization)

2.2 ทฤษฎี

2.2.1 การทำเหมืองข้อมูล (Data Mining)

2.2.2 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

2.3.2 เทคนิคแรนดอมฟอเรสต์ (Random Forest)

2.3.3 เทคนิคต้นไม้เสริมกำลังแบบไล่ระดับ (Gradient Boosted Trees)

2.3.4 การทดสอบประสิทธิภาพของตัวแบบโดยใช้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Absolute Error: MAE)

2.3.5 ค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root M

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 การวิเคราะห์ข้อมูล (Data Analytics)

ในการดำเนินงานเรื่อง การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการเปรียบเทียบแบบจำลองที่เหมาะสม ทางผู้วิเคราะห์ ข้อมูลได้ศึกษาหลักการและทฤษฎีต่างๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูลซึ่งมี รายละเอียด ดังนี้

Data Analytics คือ การจัดระเบียบและวิเคราะห์ข้อมูลที่มีอยู่ตั้งแต่อดีตจนถึงปัจจุบัน โดยอาศัยข้อมูลจาก “Big Data” เพื่อนำมาสรุปผลของข้อมูล หาแนวโน้มความน่าจะเป็นในอนาคตเพื่อแก้ไขหรือวางแผนทางการตลาด และนำข้อมูลที่ได้ไปใช้ประโยชน์ในด้านต่าง ๆ ต่อไป

1.การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive analytics) คือ การวิเคราะห์ข้อมูลขั้นแรก เพื่อตอบคำถามว่าเหตุการณ์ต่าง ๆ ที่เกิดขึ้นในอดีตเป็นอย่างไรและเกิดอะไรขึ้นบ้าง เพื่อนำมาประมวลผลลัพธ์ของสิ่งที่กำลังเกิดขึ้น ตัวอย่างเช่น รายงานการขาย รายงานผลดำเนินการ รายได้สุทธิในเดือนนี้ สินค้าตัวที่ทำยอดขายได้มากที่สุด

2.การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics) คือ การวิเคราะห์ข้อมูลเพื่อเจาะลึกว่าสิ่งที่เกิดขึ้นมีสาเหตุมาจากอะไร มีปัจจัยใดบางที่ส่งผลต่อความสัมพันธ์ของข้อมูลที่เกิดขึ้น โดยจะมีความลึกกว่าการวิเคราะห์ข้อมูลแบบพื้นฐาน ตัวอย่างเช่น ทำไมดอกไม้ถึงขายดีในช่วงเดือนกุมภาพันธ์ เทศกาลวาเลนไทน์มีผลหรือไม่

3.การวิเคราะห์แบบพยากรณ์ (Predictive analytics) คือ การวิเคราะห์ข้อมูลเชิงทำนาย โดยนำข้อมูลที่มีในอดีตมาวิเคราะห์และทำนายสิ่งจะเกิดขึ้นในอนาคต หรือมีความเป็นไปได้ว่าจะเกิดขึ้น จะให้ความสำคัญกับซอฟต์แวร์ที่ช่วยในการวิเคราะห์ข้อมูลเชิงทำนาย (Predictive Analytics Platform) เป็นอย่างมาก ยกตัวอย่างเช่น ยอดขายในอีก 5 ปีข้างหน้าจะเป็นอย่างไร กลุ่มเป้าหมายใดบ้างที่จะมาเป็นลูกค้า สินค้าหรือบริการแบบไหนที่จะทำให้แบรนด์ขายดีติดตลาด

4.การวิเคราะห์แบบให้คำแนะนำ (Prescriptive analytics) คือ การวิเคราะห์ข้อมูลแบบที่มีความซับซ้อนมากที่สุด ซึ่งเป็นการพยากรณ์สิ่งที่กำลังจะเกิดขึ้น เพื่อหาสาเหตุ ข้อดี ข้อเสีย รวมไปถึงระยะเวลาของสิ่งที่จะเกิดขึ้น เพื่อหาว่าควรปรับปรุง พัฒนา หรือแก้ไขปัญหาอะไรบ้าง และให้คำแนะนำของทางเลือกแต่ละแนวทางว่าจะมีผลลัพธ์แบบใด (datawow, 2567)

2.1.2 การพยากรณ์ข้อมูล (Forecasting data)

การพยากรณ์เป็นวิธีการทำนายเหตุการณ์หรือเงื่อนไขในอนาคตโดยการวิเคราะห์รูปแบบและเปิดเผยแนวโน้มในข้อมูลก่อนหน้าและปัจจุบัน ใช้วิธีการทางคณิตศาสตร์และใช้แบบจำลองทางสถิติเพื่อสร้างการคาดการณ์ การคาดการณ์คือการคาดการณ์ ซึ่งหมายความว่ามักจะไม่แม่นยำ 100% และกรอบเวลาสำหรับการพยากรณ์มีความสำคัญ—การคาดการณ์ระยะใกล้อาจแม่นยำกว่าเมื่อเทียบกับการคาดการณ์ระยะยาว นอกจากนี้ยังอาจช่วยในการรวบรวมข้อมูลหรือรวมเทคนิคต่างๆ เพื่อความแม่นยำยิ่งขึ้น และคิดว่าการคาดการณ์เป็นแนวทาง ไม่ใช่ปัจจัยกำหนดขั้นสุดท้ายในการตัดสินใจ กระบวนการพยากรณ์อาจดูแตกต่างกันไปในแต่ละองค์กร แต่โดยทั่วไป จะเกี่ยวข้องกับขั้นตอนเหล่านี้:

1) กำหนดสิ่งที่ต้องทำนาย บริษัทต่างๆ ระบุกรณีธุรกิจหรือตัวชี้วัดเฉพาะที่ต้องการคาดการณ์และคำนึงถึงสมมติฐานที่เกี่ยวข้องและตัวแปรที่เกี่ยวข้อง
2) รวบรวมข้อมูล ขั้นตอนนี้รวมถึงการรวบรวมข้อมูลที่จำเป็น หากมีข้อมูลในอดีตอยู่แล้ว ก็เป็นเรื่องของการกำหนดชุดข้อมูลที่เหมาะสมที่สุด
3) เลือกวิธีการพยากรณ์ เลือกเทคนิคการพยากรณ์ที่เหมาะสมที่สุดไม่เพียงแต่กรณีทางธุรกิจหรือตัวชี้วัดเท่านั้น แต่ยังรวมถึงตัวแปร สมมติฐาน และชุดข้อมูลที่เกี่ยวข้องด้วย
4) สร้างการคาดการณ์ ข้อมูลคือ วิเคราะห์แล้ว โดยใช้วิธีการที่เลือกและการคาดการณ์จะถูกสร้างขึ้นจากการวิเคราะห์นี้
5) ตรวจสอบการคาดการณ์ ตรวจสอบการคาดการณ์และดูว่าสามารถเพิ่มประสิทธิภาพเพื่อสร้างการคาดการณ์ที่แม่นยำยิ่งขึ้นได้หรือไม่

6) นำเสนอการคาดการณ์ การแสดงข้อมูลเป็นภาพ สามารถใช้เพื่อแสดงการคาดการณ์ในรูปแบบภาพมากขึ้นซึ่งผู้มีส่วนได้ส่วนเสียสามารถเข้าใจและนำไปใช้ในกระบวนการตัดสินใจได้ดีขึ้น

2.1.3 การเปรียบเทียบโมเดล (Model Comparison Concepts)

การเปรียบเทียบประสิทธิภาพระหว่างโมเดลวิเคราะห์ข้อมูลคือกระบวนการที่ใช้เพื่อ ทดสอบและวัดผลประสิทธิภาพของโมเดลเพื่อหาว่าแบบจำลองใดมีประสิทธิภาพที่ดีกว่าใน การแก้ไขปัญหาหรือทำนายผลข้อมูลที่ให้มา การเปรียบเทียบประสิทธิภาพของโมเดลอาจมี หลายแง่มุมต่าง ๆ และต้องพิจารณาในรูปแบบต่าง ๆ ตามลักษณะของงานและข้อมูลที่มีอยู่ ด้วยกัน ดังนี้

MAE วิธีการคำนวณ MAE นั้นตรงไปตรงมา แต่การทำความเข้าใจนั้นยากกว่า Mean Absolute Percentage Error เนื่องจากจำเป็นต้องพิจารณาควบคู่กับขนาด (scale) ของสิ่งที่ทำนายด้วย สมมติว่าค่า MAE ของโมเดลหนึ่งเท่ากับ 20 เซนติเมตร ผู้ตีความจำเป็นต้องพิจารณาว่าโมเดลนี้กำลังทำนายขนาดของอะไร ถ้าโมเดลนี้ใช้ทำนาย
ความสูงของตึกที่มีความสูงเฉลี่ย 10 เมตร ในกรณีนี้ถือว่าดี แต่ถ้าใช้ทำนายความสูงของมนุษย์ที่มีความสูงเฉลี่ย 1.5 เมตร ในกรณีนี้อาจจะถือว่าไม่ดีพอ ในกรณีนี้จำเป็นต้องหาค่า MAE ด้วยค่าเฉลี่ยของค่าจริงหรือที่เราจะเรียกว่า %MAE ของ MAE ดังสมการ

RMSE ย่อมาจาก Root Mean Square Error ตัวชี้วัด Root Mean Square Error **(RMSE)** นั้นมีความซับซ้อนมากกว่าการคำนวณ MAE (Mean Square Error) เล็กน้อย และจะให้น้ำหนักความเสียหาย (Penalize) กับกรณีที่ค่าความคลาดเคลื่อนมีความสุดโต่ง โดยตัวค่า RMSE นั้นสามารถหาเป็น % ได้เช่นเดียวกับ MAE ด้วยสูตรค่า RMSE นั้นจะได้รับความนิยมสูงในการใช้เพื่อฝึกฝน และเลือกแบบจำลอง
เนื่องจากตัวสูตรนั้นค่อนข้างง่ายต่อการหาค่าอนุพันธ์ (Derivative) ทำให้การปรับแบบจำลองนั้น สามารถทำได้อย่างเป็นธรรมชาติมากกว่า และเป็นค่าชี้วัดที่ใช้ในการแข่งขันสำหรับโจทย์ประเภท Regression จำนวนมากบน Platform Kaggle (Weerapat, Navavit ,2566)

2.1.4 การทำความสะอาดข้อมูล (Data Cleaning)

Data Cleaning การทำความสะอาดข้อมูล หรือ Data Cleansing คือ กระบวนการตรวจจับข้อมูล แก้ไข ลบ แทนที่ และจัดรูปแบบของข้อมูลที่ไม่สมบูรณ์ ซ้ำซ้อน ให้มีความถูกต้องและเป็นระเบียบ ตลอดจนนำไปใช้งานต่อได้อย่างมีประสิทธิภาพ จากการรวบรวมข้อมูลขนาดใหญ่ หรือ Big Data จึงต้องมีการล้างข้อมูลเพื่อคัดกรองให้เหลือแค่ข้อมูลที่นำไปใช้ได้จริง ซึ่งลักษณะของข้อมูลที่ต้องผ่านการ Data Cleansing ก่อนนำไปใช้ประโยชน์ มีดังนี้

1. ข้อมูลที่ไม่ได้อยู่ในรูปแบบเดียวกัน ในกรณีนี้อาจเกิดจากการที่มีข้อมูลจากหลาย Database ทำให้ข้อมูลที่รวบรวมมามีไฟล์คนละนามสกุลกัน เช่น .pdf, .doc, .xls หรือ .pptx เป็นต้น ทำให้ไม่สามารถใช้ในการประมวลผลด้วยกัน จึงต้องมีการแปลงไฟล์ให้อยู่ในนามสกุลเดียวกันเพื่อสามารถประมวลผลได้ และลดพื้นที่ในการจัดเก็บชุดข้อมูล

2. ข้อมูลที่ไม่ได้จัดเก็บในรูปแบบที่ต้องการเป็นข้อมูลที่ต้องทำให้อยู่ในรูปแบบที่สามารถนำไปใช้ในการวิเคราะห์ได้ บางข้อมูลที่ถูกรวบรวมมาอาจอยู่ในรูปแบบของรูปภาพ เช่น .jpg, .png, .tiff หรือ .bmp จึงต้องมีการแปลงไฟล์รูปภาพให้อยู่ในรูปแบบของไฟล์ข้อความหรือสคริปต์ เช่น .csv, .tsv, .json, และ .xml เป็นต้น

3. ข้อมูลที่ไม่ถูกต้อง โดยส่วนใหญ่แล้วการรวบรวมข้อมูลโดยคนอาจเกิดการผิดพลาด (Human Errors) เป็นเรื่องปกติ เช่น กรอกข้อมูลเกินความจริง กรอกข้อมูลในช่องที่ผิด หรือสะกดชื่อข้อมูลไม่ถูกต้อง ทำให้วิเคราะห์ข้อมูลเชิงลึก หรือ Insight ออกมาผิดพลาดไม่แม่นยำ จึงต้องทำ Data Cleansing เพื่อแก้ไขข้อมูลให้ถูกต้อง

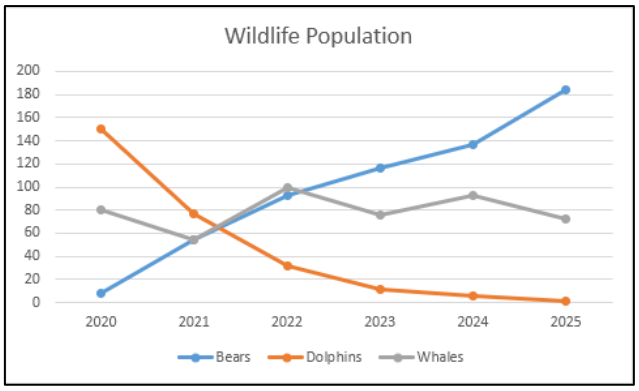
2.1.5 การแสดงผลข้อมูล (Data visualization)

Data Visualization คือ การแสดงข้อมูลที่ผ่านการวิเคราะห์และประมวลผลในรูปแบบที่เข้าใจง่ายและสื่อสารได้ชัดเจน เช่น แผนภูมิ แผนที่ อินโฟกราฟิก หรือรูปภาพ การนำเสนอเหล่านี้ช่วยให้ผู้ใช้งานสามารถมองเห็นแนวโน้ม รูปแบบ และข้อมูลเชิงลึกได้อย่างรวดเร็ว องค์กรจึงสามารถนำข้อมูลไปใช้ในการตัดสินใจที่แม่นยำและมีประสิทธิภาพมากขึ้น

เนื่องจากทุกองค์กรมีข้อมูลสำคัญและมีความซับซ้อน ซึ่งข้อมูลปริมาณมากอาจทำให้การวิเคราะห์เกิดความผิดพลาดได้ ด้วยเหตุผลนี้ทำให้องค์กรต้องมีการทำ Data Visualization

เพื่อช่วยให้ผู้ใช้งานมองเห็นภาพรวมได้ชัดเจน มีส่วนช่วยในการตัดสินใจทั้งในด้านการบริหารและการวางแผนธุรกิจ โดยการทำ Data Visualization นั้น มีรูปแบบการแสดงผลหลายประเภทด้วยกัน ซึ่งมี 5 รูปแบบที่เป็นที่นิยมนำมาใช้ในการทำงาน คือ (cheewin.j , 2568)

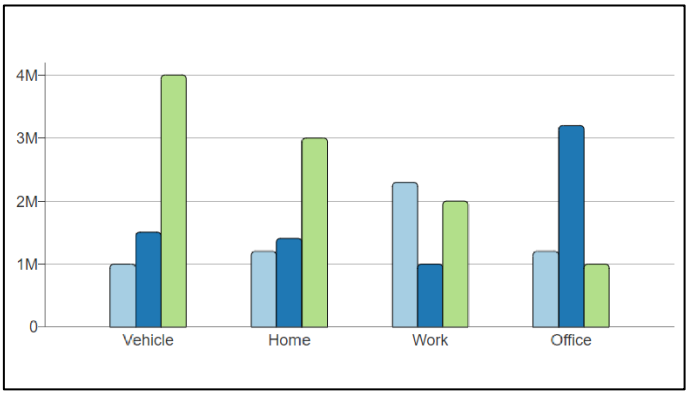
1) Line Chart คือแผนภูมิที่จะใช้เพื่อแสดงความสัมพันธ์ของข้อมูลที่เปลี่ยนแปลงตามช่วงเวลา เรียงตามแกน X และแกน Y โดยใช้จุดแทนข้อมูล และลากเส้นเชื่อมต่อจุดแต่ละจุดเพื่อให้เห็นการเปลี่ยนแปลง



ภาพที่ 2.1 แผนภูมิแบบเส้น

(ที่มา: https://www.excel-easy.com/examples/line-chart.html)

2) Bar Chart คือแผนภูมิที่จะใช้เพื่อแสดงเหมาะสำหรับการเปรียบเทียบข้อมูลระหว่างหมวดหมู่ต่างๆ โดยใช้แท่งสี่เหลี่ยมในการแสดงข้อมูลแทนจุด และใช้ความสูงต่ำเพื่อแสดงให้เห็นความแตกต่าง

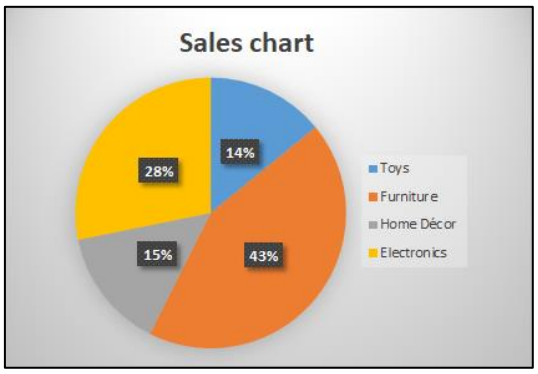


ภาพที่ 2.2 แผนภูมิแบบแท่ง

(ที่มา: https://www.cdc.gov/cove/data-visualization-types/bar-chart.html)

3) Pie Chart คือแผนภูมิที่อยู่ในรูปแบบวงกลม โดยจะมีการแบ่งวงกลม

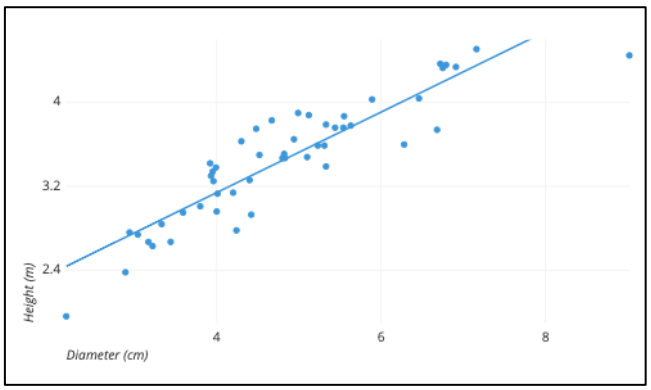
ออกเป็นส่วน ๆ ตามจำนวนข้อมูลที่ต้องการเปรียบเทียบ และใช้ตัวเลขมาแสดงปริมาณข้อมูลแต่ละส่วนเมื่อเทียบกับปริมาณข้อมูลทั้งหมด



ภาพที่ 2.3 แผนภูมิแบบพาย

(ที่มา: https://www.spotfire.com/glossary/what-is-a-pie-chart)

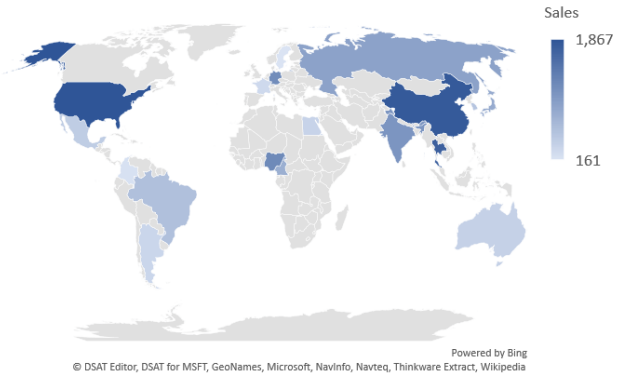
4) Scatter Plot คือกราฟที่แสดงความสัมพันธ์ระหว่างชุดข้อมูลสองชุด โดยใช้จุดแทนค่าข้อมูลบนแกน X และแกน Y จุดที่กระจายตัวบนกราฟช่วยให้เห็นรูปแบบหรือแนวโน้มความสัมพันธ์ระหว่างตัวแปร



ภาพที่ 2.4 กราฟกระจาย

(ที่มา: https://www.atlassian.com/data/charts/what-is-a-scatter-plot)

5) Map คือกราฟแสดงข้อมูลที่เกี่ยวข้องกับตำแหน่งบนพื้นที่ โดยใช้แผนที่เป็นเครื่องในการแสดงผล และปรับแต่งพื้นที่เพื่อแสดงค่าหรือความแตกต่าง



ภาพที่ 2.5 แสดงแผนที่

(ที่มา: https://reportingengineer.com/2018/04/06/excel-map-chart-world-thailand/)

2.2 ทฤษฎี

2.2.1 การทำเหมืองข้อมูล (Data Mining)

Data Mining หรือการทำเหมืองข้อมูล คือการเก็บรวบรวมข้อมูลมหาศาลเพื่อวิเคราะห์ออกมาเป็นชุดข้อมูลที่เอามาใช้ประโยชน์กับองค์กรได้ โดยผ่านกระบวนการคัดแยกแบ่งกลุ่ม และกระบวนการอื่น ๆ อีกมากมายเพื่อจะได้ชุดข้อมูลที่เหมาะกับการนำไปใช้งาน ผลลัพธ์ของการทำ Data Mining อาจจะออกมาในรูปแบบของความสัมพันธ์ของข้อมูลและแนวโน้ม

ในภาพรวม Data Mining ช่วยให้องค์กรค้นพบข้อมูลเชิงลึกซึ่งนำไปสู่การตัดสินใจได้อย่างแม่นยำ ส่งผลให้การทำงานขององค์กรบรรลุเป้าหมายที่วางไว้และช่วยให้องค์กรใช้ทรัพยากรได้อย่างมีประสิทธิภาพมากขึ้น ส่วนประโยชน์ของการใช้ Data Mining ในแต่ละประเภทธุรกิจนั้นก็แตกต่างกันออกไป กระบวนการทำ Data Mining มี 6 ข้อดังนี้

1. วิเคราะห์วัตถุประสงค์ ก่อนที่จะเริ่มทำ Data Mining นักวิเคราะห์ต้องรู้ก่อนว่าจะนำผลลัพธ์ของข้อมูลนี้ไปใช้เพื่ออะไร ถ้านักวิเคราะห์ไม่เข้าใจปัญหาและเป้าหมายของการทำงาน ก็อาจทำให้ดำเนินการผิดวิธีและเสียเวลาไปโดยไม่ตอบโจทย์

2.เก็บข้อมูล เมื่อทราบวัตถุประสงค์แล้วว่าองค์กรมีเป้าหมายอะไรในการทำ Data Mining นักวิเคราะห์ข้อมูลต้องรู้ว่าข้อมูลที่ต้องการคืออะไรบ้าง หาได้จากที่ไหน จะเก็บและรักษาข้อมูลอย่างไร ข้อมูลที่มีคุณภาพนั้นต้องเป็นข้อมูลที่แม่นยำและน่าเชื่อถือ โดยที่ไม่ขัดต่อหลักความปลอดภัย ความเป็นส่วนตัว และข้อจำกัดอื่น ๆ บางครั้ง นักวิเคราะห์ต้องเก็บ

ข้อมูลเป็นล้าน ๆ ชุดเพื่อที่จะได้เห็นภาพใหญ่ของสิ่งที่ต้องการ ซึ่งกลุ่มก้อนข้อมูลเหล่านี้จะถูกเรียกว่า *Big Data* เป็นเหมือนเหมืองขนาดใหญ่ที่มีแร่หลายชนิด

3. เตรียมข้อมูล กระบวนการนี้คือการนำข้อมูลที่มีอยู่มาคัดเฉพาะส่วนที่จำเป็นต้องใช้ในประเด็นย่อยต่าง ๆ เมื่อสกัดข้อมูลมาได้แล้วก็ต้องนำมาล้างทำความสะอาดให้พร้อมใช้มากขึ้น จากนั้นก็นำมาแบ่งหมวดหมู่และจัดเรียงภายใต้มาตรฐานเดียวกัน ตรวจสอบข้อผิดพลาด จนถึงประเมินขนาดของข้อมูลก่อนนำไปสู่ขั้นตอนถัดไป

4. เจียระไนข้อมูล หลังจากเตรียมข้อมูลเสร็จ นักวิเคราะห์จะนำข้อมูลเหล่านั้นมาทำให้พร้อมนำไปใช้งานได้มากขึ้นด้วยการเข้าโปรแกรมประมวลผลผ่าน Algorithm ที่ซับซ้อนเพื่อทำให้เห็นว่าข้อมูลเหล่านี้มีความสัมพันธ์หรือรูปแบบอย่างไงบ้าง ผลลัพธ์ที่ได้ออกมาเรียกว่า *Insight* ซึ่งอาจบอกได้ตั้งแต่พฤติกรรมผู้บริโภค เทรนด์ที่กำลังมา ไปจนถึงความเสี่ยงต่อธุรกิจ

5. นำเสนอข้อมูล ผลลัพธ์ที่ได้จะถูกนำมาสร้างเป็นโมเดลที่ทำให้คนทั่วไปเห็นภาพและเข้าใจข้อมูลมากขึ้น เปรียบเสมือนการนำอัญมณีที่สวยงามมาร้อยเรียงให้มีคุณค่ามากขึ้นทักษะที่สำคัญต่อการทำ Data Mining นั้นส่วนใหญ่จะเกี่ยวข้องกับระบบคอมพิวเตอร์และการคำนวณ

ทักษะทางสถิติ (Statistics) ในการทำ Data Mining นั้นต้องอาศัยความสามารถในการจัดการกับตัวเลขหลายล้านชุดเพื่อประมวลผลความสัมพันธ์ของข้อมูล และต้องสามารถใช้โปรแกรมทางสถิติในการสร้างและคำนวณผลลัพธ์ด้วย นั่นยังไม่รวมถึงวิธีการคิดวิเคราะห์ที่เป็นระบบ ความสามารถในการเข้าใจตรรกะและความเชื่อมโยงที่ถูกต้อง การตั้งคำถามและเก็บรายละเอียดได้เป็นอย่างดี

ปัญญาประดิษฐ์ (Artificial Intelligence หรือ AI) นักวิเคราะห์ต้องใช้ความสามารถของ AI ในการจัดการกับข้อมูลมหาศาลได้อย่างรวดเร็วมาลดเวลาในการประมวลผล และเพิ่มประสิทธิภาพในการทำงานเช่นกัน

ทักษะด้านภาษาโปรแกรม (Programming Languages) หรือการเขียนโค้ด นักวิเคราะห์ต้องใช้ภาษาโปรแกรมสื่อสารกับคอมพิวเตอร์เพื่อสร้างคำสั่งจำเพาะบางอย่าง เช่น ภาษา Python และภาษา R การมีทักษะนี้จะทำให้นักวิเคราะห์ สามารถสร้างคำสั่งแบบ

เฉพาะเจาะจงขึ้นมาเองได้ ทำให้การทำงานมีความสร้างสรรค์และให้ผลลัพธ์ที่ตอบโจทย์มากขึ้น

Machine Learning ในการประมวลผลข้อมูลอาจต้องใช้ Algorithm ที่สร้างมาเพื่อใช้ในการเจียระไนข้อมูลชุดนั้น ๆ โดยเฉพาะ หากนักวิเคราะห์สามารถทำความเข้าใจตรรกะของ Machine Learning ได้ก็จะช่วยอำนวยความสะดวกในการจัดการข้อมูลมากขึ้น

ทักษะด้านฐานข้อมูล (Database) เป็นทักษะที่จำเป็นเพราะนักวิเคราะห์ต้องดึงข้อมูลจากฐานข้อมูลหลายแหล่งมาวิเคราะห์และจัดการ เช่น Structured Query Language (SQL)

ทักษะด้าน Data Visualization ช่วยทำให้การนำเสนอข้อมูลเชิงลึกที่ค้นพบนั้นเข้าใจง่ายขึ้น เป็นประโยชน์ทั้งต่อนักวิเคราะห์ที่ต้องใช้อธิบายผลลัพธ์และต่อองค์กรที่ต้องการนำผลลัพธ์ไปใช้ประโยชน์ รูปแบบของ Data Visualization นั้นมีหลากหลาย เช่น การใช้กราฟชาร์ท และแผนที่แบบต่าง ๆ (TNIC , 2567)

2.2.2 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.2.2.1 การออกแบบเว็บไซต์ หรือ Web Design คือ กระบวนการในการสร้างหน้าตา และองค์ประกอบภายในเว็บไซต์ ตั้งแต่การจัดวางเนื้อหา การเลือกใช้สี ฟอนต์ การเลือกใช้ภาพ หรือกราฟิก และการออกแบบระบบนำทางผู้ใช้บนหน้าเว็บ ให้สวยงามสอดคล้องและสามารถใช้ งานได้ราบรื่น โดยใช้หลักการของ Responsive Web Design เข้ามาช่วยในการออกแบบ เพื่อให้สามารถแสดงผลได้ทุก ๆ อุปกรณ์ ซึ่งจะใช้ URL ร่วมกัน แต่การแสดงผลในแต่ละอุปกรณ์ จะแตกต่างกันไป การตอบสนองเป็นการปฏิบัติของการสร้างเว็บไซต์ที่เหมาะสมเพื่อทำงานบน ทุกอุปกรณ์ และทุกขนาดหน้าจอทั้งขนาดใหญ่ หรือขนาดเล็ก เช่น มือถือ หรือเดสก์ท็อป ฯลฯ การตอบสนองจะเน้นการใช้งานง่าย จึงต้องมีการออกแบบอย่างเป็นระบบ เว็บดีไซน์ที่ประสบความสำเร็จ จะต้องสอดคล้องกับแบรนด์ และเป้าหมายทางธุรกิจ สามารถสื่อสารคุณค่าของสินค้ากับบริการได้อย่างโดดเด่น ขณะเดียวกันก็ต้องเข้าใจความต้องการของกลุ่มเป้าหมาย เพื่อมอบประสบการณ์ใช้งานที่เหมาะสมกับผู้ใช้จริง

2.2.2.2 องค์ประกอบในการออกแบบเว็บไซต์ การออกแบบเว็บไซต์ที่มีประสิทธิภาพ นั้นต้องคำนึงถึงองค์ประกอบสำคัญ ดังต่อไปนี้

1) การออกแบบที่เรียบง่าย (Simplicity) ในยุคที่ผู้คนมีความสนใจในช่วงเวลา สั้นลง (Short attention span) และต้องการข้อมูลที่กระชับ และตรงประเด็น การออกแบบเว็บไซต์ แบบมินิมอล (Minimalist) ซึ่งเน้นความเรียบง่ายจึงเป็นแนวคิดที่ควรนำมาใช้เพื่อให้ผู้ใช้งาน สามารถมุ่งเน้นประเด็นสำคัญได้อย่างชัดเจน

ทั้งนี้ความเรียบง่ายไม่ควรถึงขั้นที่ทำให้ขาดจุดเด่นที่น่าสนใจ เพราะ การออกแบบหน้าเว็บไซต์ที่ซับซ้อนเกินไปอาจทำให้ผู้ใช้งานเกิดความสับสน และไม่สามารถค้นหาข้อมูลที่ต้องการได้อย่างสะดวก ดังนั้น ควรใช้กราฟิกและข้อความในปริมาณที่พอเหมาะ เน้นที่ความสะอาดตา และการใช้งานที่ราบรื่น โดยยังสามารถสื่อสารข้อมูลสำคัญของหน้าเว็บไซต์ ได้อย่างมีประสิทธิภาพ

2) การเลือกใช้สีที่เข้ากัน (Color Scheme) การเลือกใช้สีเป็นองค์ประกอบสำคัญในการออกแบบเว็บไซต์ (Web Design) เนื่องจากสีสามารถสื่อถึงอารมณ์ และความรู้สึก รวมทั้งส่งผลต่อการรับรู้ของผู้ใช้งาน หากเลือกใช้สีที่เหมาะสม และสอดคล้องกับลักษณะของเว็บไซต์ ย่อมสามารถสร้างความโดดเด่น และเอกลักษณ์ได้ดี การใช้สีไม่ควรเกิน 1 – 2 สีหลัก และสีรอง 2 – 3 สี เพื่อให้ดูเป็นระเบียบน่าสนใจ

นอกจากนี้ การเลือกสีควรคำนึงถึงความคมชัดในการแสดงผล ไม่ควรใช้ สีที่ฉูดฉาดเกินไปเพราะอาจทำให้การอ่านเนื้อหาลำบาก ควรเลือกใช้คู่สีที่เป็น Complementary colors หรือใช้ Color palette ที่มีโทนสีใกล้เคียงกัน เพื่อให้เว็บไซต์ดูสวยงามและน่าสบายตา

3) การเลือกฟอนต์ที่อ่านง่าย (Readable Fonts) ฟอนต์ และการจัดวางข้อความเป็นองค์ประกอบสำคัญในการออกแบบหน้าเว็บไซต์ (Web Page Design) เนื่องจากมีผล ต่อความสะดวกในการอ่าน ควรเลือกใช้ฟอนต์ที่อ่านง่าย และไม่ซับซ้อน เช่น ฟอนต์ Sans-serif อย่าง Arial Verdana Helvetica สำหรับเนื้อหาทั่วไป หรือฟอนต์ Serif เช่น Times New Roman Georgia ที่เหมาะกับเนื้อหาบทความยาว

ขนาดของฟอนต์ (Font size) และการเว้นระยะห่างระหว่างบรรทัด (Line-Height) ที่เหมาะสมจะช่วยเพิ่มความสะดวกในการอ่าน และนำเสนอข้อมูลได้อย่างมี ประสิทธิภาพ

4) การใช้พื้นที่ว่างอย่างเหมาะสม (White Space) หรือพื้นที่ว่าง เป็นส่วนที่ปราศจากข้อความหรือกราฟิก ซึ่งมักใช้เป็นพื้นหลังสีขาว หรือสีอ่อน เพื่อแบ่งแยกเนื้อหาให้เป็นสัดส่วน สร้างความเป็นระเบียบเรียบร้อย และช่วยให้ผู้ใช้เข้าใจเนื้อหาได้ง่ายขึ้น สีพื้น หลังไม่จำเป็นต้องเป็นสีขาว อาจใช้สีอ่อนหรือเส้นที่ไม่ทำให้ดูลายตา สิ่งสำคัญ คือ ต้องใช้ White Space อย่างพอเหมาะ ไม่มาก หรือน้อยเกินไป

5) การออกแบบเว็บไซต์ให้รองรับการแสดงผลบนมือถือ ด้วยการเพิ่มขึ้นของ ผู้ใช้งานอินเทอร์เน็ตผ่านสมาร์ทโฟน และแท็บเล็ต การออกแบบเว็บไซต์ที่รองรับการแสดงผล บนอุปกรณ์พกพา หรือ Responsive Design จึงกลายเป็นปัจจัยสำคัญ Responsive Design คือการออกแบบเว็บไซต์ให้ปรับขนาด และการจัดเรียงเนื้อหาได้โดยอัตโนมัติ ไม่ว่าจะเป็นคอมพิวเตอร์ สมาร์ทโฟน หรือแท็บเล็ต ทำให้ผู้ใช้สามารถใช้งานได้อย่างราบรื่น

6) การใช้ Visual Content ในการตกแต่งเว็บไซต์ Visual Content เช่น รูปภาพ ไอคอน อินโฟกราฟิก และวิดีโอ เป็นองค์ประกอบที่ช่วยเสริมสร้างความน่าสนใจให้กับเว็บไซต์ นอกจากการตกแต่งให้สวยงาม ยังช่วยดึงดูดความสนใจ และอธิบายเนื้อหาที่ยากให้เข้าใจง่ายขึ้น การใช้ Visual Content ควรอยู่ในปริมาณที่เหมาะสม และต้องคำนึงถึงขนาดไฟล์เพื่อไม่ให้กระทบ ต่อความเร็วในการโหลดเว็บไซต์

7) การวางตำแหน่งเนื้อหาหลักให้เด่นชัด (Highlight Key Content) การวางตำแหน่งเนื้อหาที่ต้องการเน้นให้เห็นได้ชัดเจนเป็นสิ่งสำคัญ เพราะผู้ใช้มักให้ความสนใจกับสิ่งที่เห็นชัดเจนที่สุดก่อน เช่น ข้อความขนาดใหญ่ หรือรูปภาพเด่นชัด ควรใช้หลักการจัดลำดับความสำคัญขององค์ประกอบ (Visual Hierarchy) เพื่อทำให้เนื้อหาสำคัญโดดเด่น และสะดุดตามากขึ้น

8) การออกแบบระบบนำทางเว็บไซต์ (Navigation) การออกแบบระบบนำทาง หรือเมนูที่ใช้งานง่ายเป็นหัวใจสำคัญ เพราะช่วยให้ผู้ใช้งานสามารถค้นหา และเข้าถึงข้อมูลได้อย่าง รวดเร็ว ควรวางโครงสร้างเมนูให้ชัดเจน และเป็นหมวดหมู่ การใช้ข้อความในเมนูควรกระชับ และเข้าใจง่าย

9) การใส่ Call-to-Action (CTA) เป็นองค์ประกอบที่กระตุ้นให้ผู้ใช้ดำเนิน การตามที่เว็บไซต์ต้องการ เช่น การกดสั่งซื้อ หรือการสมัครสมาชิก ปุ่ม CTA ควรมีขนาดใหญ่พอ ที่จะสังเกตเห็นได้ชัดเจน และข้อความควรกระชับ ชัดเจน เช่น "สั่งซื้อทันที" หรือ "ดาวน์โหลดฟรี"

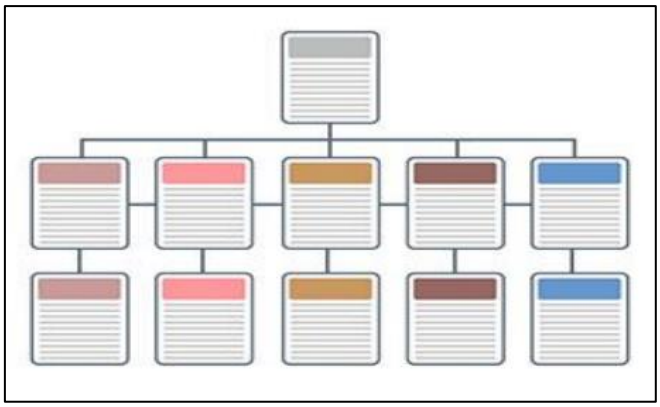
10) การทดสอบ และปรับปรุงการออกแบบเว็บไซต์อยู่เสมอ การออกแบบ เว็บไซต์เป็นกระบวนการที่ต้องมีการทดสอบ และปรับปรุงอย่างต่อเนื่อง เพื่อให้สอดคล้อง กับเทคโนโลยีที่เปลี่ยนแปลง และพฤติกรรมของผู้ใช้ การใช้เครื่องมืออย่าง Google Analytics ช่วยให้เห็นพฤติกรรมการใช้งาน และสามารถปรับปรุงเว็บไซต์ให้ตอบโจทย์ความต้องการ ได้อย่างมีประสิทธิภาพ

2.2.2.3 รูปแบบโครงสร้างของเว็บไซต์ การออกแบบโครงสร้างของเว็บไซต์สามารถ ทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบ และความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งาน ของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบ ไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างเว็บไซต์แบบเรียงลำดับ จะเป็นโครงสร้างแบบธรรมดาที่นิยม ใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราว ตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็น พวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิ้งค์เนื้อหา ก็จะลิ้งค์ไป ทีละหน้า มีทิศทางการเข้าสู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ปุ่มเดินหน้า – ถอยหลังในการกำหนด ทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับมีข้อเสีย คือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหาเพราะไม่สามารถกำหนดทิศทางการเข้า สู่เนื้อหาด้วยตัวเองได้

ภาพที่ 2.6 แสดงโครงสร้างแบบเรียงลำดับ
(ที่มา: https://asria.org/website-structure/)

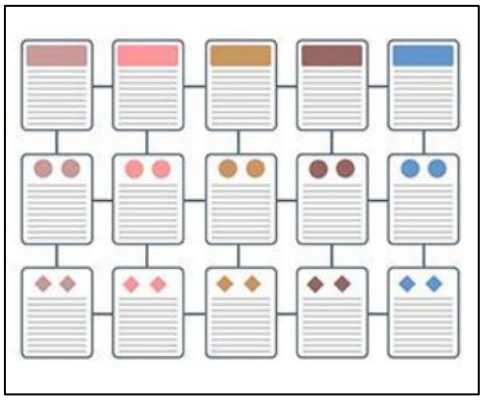
2) โครงสร้างแบบลำดับขั้น นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้ สามารถเข้าถึงข้อมูลต่าง ๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วน ๆ และมีการนำเสนอ รายละเอียดย่อย ๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีโฮมเพจเป็นจุดเริ่มต้น และจุดร่วมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับ จากบนลงล่าง



ภาพที่ 2.7 แสดงโครงสร้างแบบลำดับขั้น

(ที่มา: https://asria.org/website-structure/)

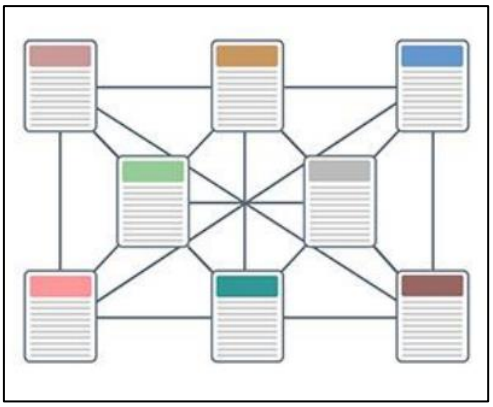
3) โครงสร้างแบบตาราง เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีความยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่าง ๆ ได้ง่ายขึ้น การออกแบบ ในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกัน และกัน ทำให้ผู้ใช้งานสามารถเปลี่ยน ทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลา และทำให้เว็บไซต์ มีความทันสมัยขึ้น



ภาพที่ 2.8 แสดงโครงสร้างแบบตาราง

(ที่มา: https://asria.org/website-structure/)

4) โครงสร้างแบบใยแมงมุม เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถเข้า ถึงหน้าเว็บเพจต่าง ๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถ เชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2.9 แสดงโครงสร้างแบบใยแมงมุม
(ที่มา: https://asria.org/website-structure/)

2.2.2.4 ส่วนประกอบหลักในหน้าเว็บไซต์ มีองค์ประกอบหลัก ๆ เช่น ชื่อเว็บไซต์ แถบเมนูนำทาง แถบเมนูย่อย ข้อความหัวเรื่อง ภาพประกอบตกแต่ง หรือภาพเนื้อหา ป้ายโฆษณา และอื่น ๆ ตามที่ต้องการนำเสนอ บางเว็บไซต์อาจจะแสดงภาพกราฟิกของสินค้า หรือบริการ แทนการแสดงข้อความ ทั้งนี้ก็ขึ้นอยู่กับการออกแบบหน้าเว็บไซต์ให้เหมาะสมกับเนื้อหาที่ต้องการ นำเสนอส่วนประกอบบนหน้าเว็บเพจ มีทั้งหมด 6 องค์ประกอบ ดังนี้

1) ส่วนหัว (Header) สิ่งสำคัญที่ช่วยสร้างการจดจำให้กับธุรกิจ คือ โลโก้ (Logo) ซึ่งทำหน้าที่เป็นสัญลักษณ์ที่สะท้อนถึงตัวตน และลักษณะเฉพาะของธุรกิจอย่างชัดเจน การออกแบบโลโก้ที่เหมาะสมจะช่วยให้ผู้เยี่ยมชมเว็บไซต์สามารถเข้าใจได้ทันทีว่าเว็บไซต์นั้นมีเนื้อหาหรือบริการเกี่ยวข้องกับอะไร โลโก้จึงถือเป็นองค์ประกอบที่สำคัญในการสื่อสารภาพลักษณ์ ของธุรกิจ ควรจัดวางโลโก้ให้อยู่บริเวณด้านบนสุดของหน้าแรกของเว็บไซต์ เพื่อให้ผู้ เข้าชม สามารถมองเห็นได้ทันทีที่เข้าสู่หน้าเว็บไซต์ และสร้างการจดจำธุรกิจของคุณอย่างมีประสิทธิภาพ

2) เมนู (Navigator) แถบเมนูเป็นองค์ประกอบสำคัญที่ช่วยอำนวยความสะดวก ในการใช้งานเว็บไซต์ให้กับผู้เยี่ยมชม ควรออกแบบแถบเมนูให้สอดคล้องกับโครงสร้างและการใช้ งานของเว็บไซต์ โดยตำแหน่งของเมนูสามารถอยู่ในตำแหน่งเดียวกับโลโก้ด้านบนของหน้าเว็บไซต์ หรือวางไว้ที่ด้านซ้ายของหน้าจอก็ได้ ทั้งนี้ เมนูควรถูกจัดเรียงตามลำดับความสำคัญ เพื่อให้ผู้เยี่ยม ชมสามารถเข้าถึงข้อมูลและบริการต่าง ๆ ของเว็บไซต์ได้อย่างรวดเร็ว และสะดวก อย่างไรก็ตาม ควรหลีกเลี่ยงการใส่เมนูมากเกินความจำเป็น เพราะอาจก่อให้เกิดความสับสนในการใช้งาน และส่งผลให้การนำทางภายในเว็บไซต์ดูซับซ้อน และไม่เป็นระเบียบ

3) เนื้อหา (Contents) ในหน้าแรกของเว็บไซต์ธุรกิจ ควรนำเสนอเนื้อหาที่ชัดเจน และครอบคลุมรายละเอียดเกี่ยวกับธุรกิจ หรือบริการของคุณ เพื่อให้ผู้เยี่ยมชมได้รู้จักธุรกิจของ คุณมากยิ่งขึ้น ข้อมูลที่ควรนำเสนอ ได้แก่ การแนะนำว่าสินค้าหรือบริการของคุณคืออะไร คุณสมบัติเด่นของสินค้า หรือบริการ ประโยชน์ที่ผู้ใช้จะได้รับ ข่าวสารหรือกิจกรรมสำคัญ รวมถึง โปรโมชันที่เกี่ยวข้อง ช่องทางการติดต่อที่สะดวกและเข้าถึงง่าย คำรับรองจากลูกค้าที่เคยใช้บริการ รวมถึงรางวัล หรือการรับรองคุณภาพบริการที่ธุรกิจของคุณได้รับ นอกจากนี้ หากต้องการเพิ่มโอกาสให้ธุรกิจของคุณถูกค้นพบได้ง่ายขึ้น ควรใส่ คำสำคัญ (Keywords) ที่เกี่ยวข้องกับสินค้า หรือบริการในเนื้อหาของเว็บไซต์ ซึ่งจะช่วยให้ กลุ่มเป้าหมายสามารถเข้าถึงเว็บไซต์ของคุณได้สะดวก และตรงกับความต้องการมากยิ่งขึ้น

4) ฟังก์ชันการค้นหา (Search Function) การจัดทำฟังก์ชันการค้นหาในเว็บไซต์ เป็นสิ่งสำคัญที่ช่วยให้ผู้เยี่ยมชมสามารถค้นหาสินค้าหรือบริการได้อย่างรวดเร็วและสะดวกยิ่งขึ้น โดยเฉพาะในกรณีที่เว็บไซต์ของคุณมีสินค้าหรือบริการจำนวนมาก ฟังก์ชันนี้จะช่วยให้ผู้ใช้สามารถ ค้นหาสิ่งที่ต้องการได้อย่างง่ายดาย การออกแบบ และพัฒนาฟังก์ชันการค้นหาที่มีประสิทธิภาพ จึงเป็นสิ่งที่ควรให้ความสำคัญ เพื่อให้ผู้เข้าชมสามารถเข้าถึงข้อมูลที่ต้องการได้โดยไม่ยุ่งยาก และเพิ่มประสบการณ์การใช้งานที่ดีขึ้นบนเว็บไซต์

5) ส่วนท้าย (Footer) ส่วนท้ายของเว็บไซต์ควรถูกออกแบบให้ช่วยอำนวยความ สะดวกแก่ผู้เข้าชมในการนำทางไปยังหน้าต่าง ๆ ภายในเว็บไซต์ พร้อมทั้งเชื่อมโยงไปยัง แพลตฟอร์ม โซเชียลมีเดีย (Social Media) ข้อมูลที่ควรแสดงในส่วนนี้ประกอบด้วย ชื่อของธุรกิจ ช่องทางการติดต่อ เช่น หมายเลขโทรศัพท์และอีเมล ที่ตั้งของบริษัท รวมถึงลิงก์ไปยังโซเชียลมีเดีย ต่าง ๆ และลิงก์ที่เชื่อมไปยังหน้าหลักอื่น ๆ ของเว็บไซต์ ทั้งนี้ การจัดวางข้อมูลในส่วนท้ายของเว็บไซต์ควรคำนึงถึงการใช้งานที่ง่ายและการเข้าถึงข้อมูลได้อย่างรวดเร็ว

2.2.2.5 ทฤษฎีสีในการออกแบบเว็บไซต์

การสร้างสีสันบนหน้าเว็บเป็นสิ่งที่สื่อความหมายของเว็บไซต์ได้อย่างชัดเจนการเลือกใช้สีให้เหมาะสม กลมกลืน ไม่เพียงแต่จะสร้างความพึงพอใจให้กับผู้ใช้ แต่ยังสามารถทำให้เห็นถึงความแตกต่างระหว่างเว็บไซต์ได้ สีเป็นองค์ประกอบหลักสำหรับการตกแต่งเว็บ จึงจำเป็นอย่างยิ่งที่จะต้องทำความเข้าใจเกี่ยวกับการใช้สีระบบสีที่แสดงบนจอคอมพิวเตอร์ มีระบบการแสดงผลผ่านหลอดลำแสงที่เรียกว่า CRT (Cathode ray tube) โดยมีลักษณะระบบสีแบบบวก อาศัยการผสมของของแสงสีแดง สีเขียว และสีน้ำเงิน หรือระบบสี RGB สามารถกำหนดค่าสีจาก 0 ถึง 255 ได้ จากการรวมสีของแม่สีหลักจะทำให้เกิดแสงสีขาว

มีลักษณะเป็นจุดเล็ก ๆ บนหน้าจอไม่สามารถมองเห็นด้วยตาเปล่าได้ จะมองเห็นเป็นสีที่ถูกผสมเป็นเนื้อสีเดียวกันแล้ว จุดแต่ละจุด หรือ พิกเซล (Pixel) เป็นส่วนประกอบของภาพบนหน้าจอคอมพิวเตอร์ โดยจำนวนบิตที่ใช้ในการกำหนดความสามารถของการแสดงสีต่าง ๆ เพื่อสร้างภาพบนจอนั้นเรียกว่า บิตเด็ป (Bit depth) ในภาษา HTML มีการกำหนดสีด้วยระบบเลขฐานสิบหก ซึ่งมีเครื่องหมาย (#) อยู่ด้านหน้าและตามด้วยเลขฐานสิบหกจำนวนอักษรอีก 6 หลัก โดยแต่ละไบต์ (byte) จะมีตัวอักษรสองตัว แบ่งออกเป็น 3 กลุ่ม เช่น #FF12AC การใช้ตัวอักษรแต่ละไบต์นี้เพื่อกำหนดระดับความเข้มของแม่สีแต่ละสีของชุดสี RGB โดย 2 หลักแรกแสดงถึงความเข้มของสีแดง 2 หลักต่อมา แสดงถึงความเข้มของสีเขียว 2 หลักสุดท้ายแสดงถึงความเข้มของสีน้ำเงิน

สีมีอิทธิพลในเรื่องของอารมณ์การสื่อความหมายที่เด่นชัด กระตุ้นการรับรู้ทางด้านจิตใจมนุษย์ สีแต่ละสีให้ความรู้สึกถึงอารมณ์ที่ไม่เหมือนกัน สีบางสีให้ความรู้สึกสงบ บางสีให้ความรู้สึกตื่นเต้นรุนแรง สีจึงเป็นปัจจัยส าคัญอย่างยิ่งต่อการออกแบบเว็บไซต์ ดังนั้นการเลือกใช้โทนสีภายในเว็บไซต์เป็นการแสดงถึงความแตกต่างของสีที่แสดงออกทางอารมณ์ มีชีวิตชีวาหรือเศร้าโศก รูปแบบของสีที่สายตาของมนุษย์มองเห็น สามารถแบ่งออกเป็น 3 กลุ่ม คือ

1) สีโทนร้อน (Warm Colors) เป็นกลุ่มสีที่แสดงถึงความสุข ความปลอบโยน ความอบอุ่น และดึงดูดใจ สีกลุ่มนี้เป็นกลุ่มสีที่ช่วยให้หายจากความเฉื่อยชา มีชีวิตชีวามากยิ่งขึ้น

2) สีโทนเย็น (Cool Colors) แสดงถึงความที่ดูสุภาพ อ่อนโยน เรียบร้อย เป็นกลุ่มสีที่มีคนชอบมากที่สุด สามารถโน้มนาวในระยะไกลได้

3) สีโทนกลาง (Neutral Colors) สีที่เป็นกลาง ประกอบด้วย สีดำ สีขาว สีเทา และสีน้ำตาล กลุ่มสีเหล่านี้คือ สีกลางที่สามารถนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีกลางขึ้นมา

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

โครงข่ายประสาทเทียม (Artificial Neural Networks) หรือที่มักจะเรียกสั้น ๆ ว่า โครงข่ายประสาท (Neural Networks หรือ Neural Net) เป็นหนึ่งในเทคนิคของการทำเหมืองข้อมูล (Data Mining) คือโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอน

เนคชันนิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจดจำรูปแบบ(Pattern Recognition) และการสร้างความรู้ใหม่ (Knowledge Extraction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท หรือ "นิวรอน" (Neurons) และ "จุดประสานประสาท" (Synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า "เดนไดรท์" (Dendrite) ซึ่งเป็น input และปลายในการส่งกระแสประสาท เรียกว่า "แอคซอน" (Axon) ซึ่งเป็นเหมือน output ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทางแอคซอนของมัน (Wittaya ,2551)

2.3.2 เทคนิคแรนดอมฟอเรสต์ (Random Forest)

Random Forest คือ Algorithm การเรียนรู้ของเครื่อง (Machine Learning) ที่เกิดจากการรวม Decision Tree หลาย ๆ ต้นเข้าด้วยกัน โดยแต่ละต้นจะถูกสร้างขึ้นจากคุณลักษณะของข้อมูล (Feature) ที่สุ่มมาเพียงบางส่วนRandom Forest เป็นเทคนิคการรวม Model หลาย ๆ Model เพื่อสร้าง Model ที่มีประสิทธิภาพสูงขึ้น (Ensemble Learning) ซึ่งใช้หลักการของ Bagging (Bootstrap Aggregating) เพื่อเพิ่มความแม่นยำและลดปัญหา Overfitting โดยการสุ่มตัวอย่าง (Random Sampling) แบบใส่คืนเพื่อสร้างชุดข้อมูลย่อยหลายชุดจากชุดข้อมูลต้นฉบับ ทำให้แต่ละชุดข้อมูลย่อยอาจมีตัวอย่างซ้ำกันได้ สำหรับนำมาสร้าง Model หลาย ๆ Model โดยแต่ละ Model จะใช้ชุดข้อมูลย่อยที่สุ่มมาในการสร้าง (Nuttachot, 2567)

2.3.3 เทคนิคต้นไม้เสริมกำลังแบบไล่ระดับ (Gradient Boosted Trees)

Gradient boosting เป็นอัลกอริทึมการเรียนรู้แบบ *ensemble* ที่สร้างการพยากรณ์ที่แม่นยำขึ้นโดยการรวมต้นไม้ตัดสินใจ (decision trees) หลายต้นเข้าด้วยกันเป็นโมเดลเดียว วิธีการสร้างโมเดลเชิงพยากรณ์นี้ ซึ่งถูกพัฒนาโดย Jerome Friedman ใช้โมเดลฐาน (base

models) เพื่อเสริมจุดแข็งของกันและกัน โดยแก้ไขข้อผิดพลาดและปรับปรุงความสามารถในการพยากรณ์ ด้วยการจับรูปแบบที่ซับซ้อนในข้อมูล gradient boosting จึงโดดเด่นในการทำงานด้านการพยากรณ์ที่หลากหลาย (Bryan, Fangfang , 2568)

2.3.4 การทดสอบประสิทธิภาพของตัวแบบโดยใช้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Absolute Error: MAE)

คือ การคำนวณหาค่าเฉลี่ยที่เป็น Absolute ของความต่างระหว่างค่า Predict กับค่า Actual วิธีนี้ใช้ประเมินประสิทธิภาพ Model ดีกว่าวิธีอื่นหาก Data ที่นำมาคำนวณมีค่าตัวเลขที่ต่างจากค่าส่วนใหญ่มากๆ (Outlier) เช่นค่าส่วนใหญ่เป็นหลักสิบ แต่มีค่าหลักพันผสมอยู่จำนวนหนึ่ง (Boonruethairat, J., 2567)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

ภาพที่ 2.10 ภาพของ Mean Absolute Error

(ที่มา: https://jysim.xyz/21)

2.3.5 ค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root Mean Square Error : RMSE)

คือ การคำนวณหาค่า Square Root ของ Metric MSE ข้างต้น แต่มีข้อดีกว่าตรงที่ตีความง่าย ใช้หน่วยประเมินประสิทธิภาพ Model แบบเดียวกับ Data เนื่องจากมี Square Root จึงไม่มีอิทธิพลของเลขยกกำลังเหมือนกับ MSE (Boonruethairat, J. , 2567)

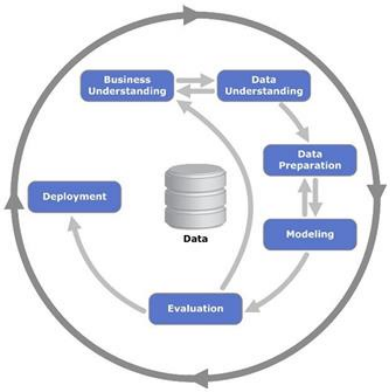
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2}$$

ภาพที่ 2.11 ภาพของ Root Mean Square Error

(ที่มา: https://bdi.or.th/big-data-101/mape-evaluation/)

2.3.6 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

CRISP-DM ย่อมาจาก Cross-industry standard process for data mining หรือกระบวนการมาตรฐานข้ามอุตสาหกรรมสำหรับการทำเหมืองข้อมูลเป็นรูปแบบกระบวนการมาตรฐานแบบเปิดที่ อธิบายแนวทางทั่วไปที่ ผู้เชี่ยวชาญด้านการทำเหมืองข้อมูลใช้ เป็นรูปแบบ การวิเคราะห์ที่ใช้กันอย่างแพร่หลายที่สุด



ภาพที่ 2.3.3 ขั้นตอนในกระบวนการ CRISP-DM

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

1. การทำความเข้าใจธุรกิจ (Business Understanding) ขั้นตอนแรกมุ่งไปที่การทำความเข้าใจธุรกิจ ปัญหาและวัตถุประสงค์ของโครงการจากมุมมองทางธุรกิจ จากนั้นแปลงปัญหาให้อยู่ในรูปของโจทย์สำหรับการวิเคราะห์ข้อมูล และวางแผนการดำเนินงาน เบื้องต้น

2. การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนนี้เริ่มต้นด้วยการรวบรวมข้อมูล จากนั้นทำความเข้าใจ ตรวจสอบคุณภาพ และเลือกข้อมูลที่เก็บรวบรวมมาว่าจะใช้ข้อมูลใดบ้างในการวิเคราะห์ ขั้นตอนที่ 1 และ 2 สามารถทำกลับไปมาได้ เนื่องจากการทำความเข้าใจธุรกิจทำให้เราเข้าใจข้อมูลมากขึ้น และการเข้าใจข้อมูลก็ทำให้เราเข้าใจธุรกิจมากขึ้นเช่นกัน

3. การเตรียมข้อมูล (Data Preparation) ขั้นตอนการเตรียมข้อมูล หมายถึง ขั้นตอนทั้งหมดที่จะทำเพื่อให้ข้อมูลดิบที่เรารวบรวมมา กลายเป็นข้อมูลสมบูรณ์ที่พร้อมจะเข้า สู่โมเดลในขั้นตอนที่ 4 เช่น การสร้างตาราง การลบข้อมูลที่ไม่ต้องการออก การแปลงข้อมูลให้อยู่ในรูปแบบที่ต้องการ

4. การสร้างโมเดล (Modeling) ในขั้นตอนนี้ เราจะเลือกและทดสองสร้างโมเดลหลายๆแบบที่น่าจะสามารถแก้ไขปัญหาที่ต้องการได้ จากนั้นค่อยๆปรับค่าพารามิเตอร์ในแต่ละโมเดล เพื่อให้ได้โมเดลที่เหมาะสมที่สุดมาใช้ในการแก้ไขปัญหา

5. การวัดประสิทธิภาพของโมเดล (Evaluation) เราจะทำการวัดประสิทธิภาพ ของโมเดลที่ได้จากขั้นตอนที่ 4 เพื่อวัดว่าโมเดลมีประสิทธิภาพเพียงพอต่อการนำไปใช้งานแล้วหรือไม่ ซึ่งโมเดลแต่ละประเภทก็จะมีตัววัดประสิทธิภาพที่แตกต่างกันออกไป

6. การนำโมเดลไปใช้งานจริง (Deployment) เป็นการนำโมเดลที่เหมาะสมที่สุด ไปใช้งานจริง เพื่อวิเคราะห์และแก้ปัญหาที่ต้องการ

2.4 วรรณกรรมที่เกี่ยวข้อง

Gammall และ Lai (2568) มุ่งเน้นการสร้างแบบจำลองพยากรณ์การอยู่รอดของผู้ป่วยมะเร็งครอบคลุมทั้งหมด 10 ชนิด โดยใช้ข้อมูลเวชระเบียนอิเล็กทรอนิกส์ (EHR) ร่วมกับข้อมูลพันธุกรรมเชิงลึก (genomic alterations และ mutational burden) รวมถึงตัวแปรทางคลินิกที่ต้องใช้เป็นมาตรฐาน ได้แก่ Cancer_Type, Cancer_Stage, Age, Gender, และตัวบ่งชี้ด้านพันธุกรรม เช่น TP53, TMB, รวมทั้ง polygenic risk signals อื่น ๆ งานวิจัยทำการเปรียบเทียบโมเดลสำหรับ Survival Analysis เช่น Elastic Net Cox, Random Survival Forest (RSF), Gradient Boosting Survival Model และ DeepSurv (Neural Network) โดยประเมินด้วย Harrell's C-index ซึ่งรายงานค่าเฉลี่ยประมาณ 0.72 แสดงให้เห็นว่าโมเดลเชิงลึกและ tree-based มีประสิทธิภาพสูง อย่างไรก็ตาม การเพิ่มข้อมูลพันธุกรรมพบว่าให้ประโยชน์เฉพาะในบางชนิดมะเร็งเท่านั้น โดยเฉพาะมะเร็งที่มี mutation burden สูงหรือมี driver mutation ชัดเจน

จากการศึกษาวรรณกรรมดังกล่าว ผู้ศึกษาเห็นว่าการใช้โมเดลกลุ่ม RSF/GBM/NN นั้นมีความเหมาะสมกับลักษณะข้อมูลในโครงงานที่ประกอบด้วยตัวแปรพื้นฐานคล้ายคลึงกัน ได้แก่ Cancer_Type, Cancer_Stage, Genetic_Risk, Age, Gender นอกจากนี้บทความยังสนับสนุนแนวคิดการผสานข้อมูลพันธุกรรมกับข้อมูลคลินิก ซึ่งช่วยเพิ่มความแม่นยำของการพยากรณ์ในหลายกรณี โดยเฉพาะเมื่อ predicting Survival_Years ที่ต้องอาศัยสัญญาณชีวภาพ เช่น

mutation load หรือ genetic risk score เพื่อให้โมเดลมีความแยกแยะสูงขึ้นในผู้ป่วยที่มี stage ใกล้กันแต่ความเสี่ยงแตกต่างกันอย่างแท้จริง

Wiegrebe และคณะ (2567) บทความนี้เป็น systematic review ครอบคลุมความก้าวหน้าของ Deep Learning ในงาน Survival Analysis โดยเน้นโมเดล Neural Networks เช่น DeepSurv, DeepHit, Dynamic–DeepHazard และ attention–based survival models พร้อมทั้งเปรียบเทียบกับโมเดลคลาสสิกอย่าง Random Survival Forest และ Gradient Boosting ในงานที่มีข้อมูลแบบ right–censored ผู้วิจัยสรุปว่าโมเดลเชิงลึกสามารถเรียนรู้ความสัมพันธ์เชิงซับซ้อนในข้อมูลคลินิกได้ดี โดยเฉพาะเมื่อมีตัวแปรหลายด้าน เช่น clinicopathologic, behavioral, imaging และ omics

นอกจากนี้ review ยังชี้ว่าตัวแปรมาตรฐานทางคลินิก เช่น Cancer_Stage, Age, Gender สามารถถูกขยายให้ครอบคลุมปัจจัยไลฟ์สไตล์ เช่น Smoking, Alcohol_Use, Obesity_Level เพื่อช่วยเสริมความแม่นยำของ survival models โดยเฉพาะอย่างยิ่งในงาน multi–cancer ที่ความเสี่ยงมีความแปรผันสูง

จากการศึกษาวรรณกรรมดังกล่าว ผู้ศึกษาได้แนวทางว่าโมเดล NN สามารถทำงานร่วมกับ RF/GBM ได้เป็นอย่างดีในฐานะ baseline และสามารถรองรับข้อมูลของโครงงานที่มีทั้งตัวแปรคลินิกและตัวแปรพฤติกรรม (Smoking, Alcohol_Use, Obesity_Level) บทความชี้ว่าสถาปัตยกรรม neural networks ที่เหมาะสมจะช่วยให้โมเดลรับมือกับ censoring และ non–linear interactions ที่พบในข้อมูลสุขภาพได้ดีกว่าวิธีดั้งเดิม

Dutta และคณะ (2568) ใช้ dataset ผู้ป่วยมะเร็งปอดจาก Kaggle ที่ประกอบด้วยตัวแปรไลฟ์สไตล์ เช่น Smoking, Alcohol_Use, Obesity_Level รวมทั้งปัจจัยประชากร (Age, Gender) และตัวแปรรองอื่น ๆ ทดลองโมเดล Decision Trees, KNN, Random Forest, Naïve Bayes, AdaBoost, Logistic Regression, SVM และ Neural Networks (1–3 ชั้น) โดยประเมินด้วย train/test split และ K–fold cross–validationผลลัพธ์พบว่า Neural Network ชั้นเดียว ให้

accuracy สูงสุด (~92.86%) และเน้นว่าการทำ feature selection และ normalization มีผลต่อความเสถียรของโมเดลอย่างมีนัยสำคัญ

จากการศึกษาวรรณกรรมดังกล่าว ผู้ศึกษาเห็นว่าในโครงงานของตน ซึ่งมีตัวแปร Smoking, Alcohol_Use, Obesity_Level, Age, Gender ควรใช้ NN และ RF/GBM เป็นโมเดลหลัก และจำเป็นต้องมีขั้นตอน preprocessing ที่รัดกุม เช่น standardization ของตัวแปรไลฟ์สไตล์ เพื่อให้โมเดลทำนายได้แม่นขึ้นและลด noise จากข้อมูลต้นทาง

Hosseini และคณะ (2561) ได้ศึกษาการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งปอดโดยใช้เทคนิคการเรียนรู้ของเครื่องแบบผสมผสาน ระหว่างการจำแนกประเภทและการพยากรณ์เชิงถดถอย (Regression) งานวิจัยนี้ใช้ชุดข้อมูลจากฐานข้อมูล SEER เพื่อพัฒนาตัวแบบในการทำนายระยะเวลาการอยู่รอดเป็นจำนวนเดือน โดยมีการเปรียบเทียบประสิทธิภาพของตัวแบบ Random Forest (RF), Gradient Boosted Machines (GBM) และ Linear Regression

จากการศึกษาวรรณกรรมดังกล่าว พบว่าตัวแบบประเภท Ensemble Learning โดยเฉพาะ Random Forest และ GBM ให้ประสิทธิภาพดีที่สุดในการพยากรณ์ โดยวัดประสิทธิภาพด้วยค่า Root Mean Square Error (RMSE) ซึ่งงานวิจัยระบุว่าตัวแบบกลุ่มนี้สามารถจัดการกับข้อมูลที่มีความซับซ้อนสูงได้ดี และให้ผลการพยากรณ์ที่แม่นยำกว่าการใช้สถิติเชิงเส้นแบบดั้งเดิม โดยเฉพาะในกลุ่มผู้ป่วยที่มีระยะเวลาการอยู่รอดสั้น

Mohamed และคณะ (2568) งานนี้ใช้ CatBoost (GBM) และ Random Forest เพื่อวิเคราะห์ข้อมูล lifestyle + genetics โดยตัวแปรสำคัญ ได้แก่ Genetic_Risk, Smoking, Alcohol_Use, Physical_Activity (เชื่อมโยงกับ Obesity_Level), Age, Gender ผลลัพธ์สูงมาก (Accuracy 98.75%, F1 = 0.982) สะท้อนว่าข้อมูลแบบตารางที่มีทั้ง genetic และ behavioral features เหมาะกับโมเดล tree-based และ boosting

จากการศึกษาวรรณกรรมดังกล่าว ผู้ศึกษาเห็นว่าโมเดล GBM (โดยเฉพาะ CatBoost) และ RF เป็นทางเลือกที่แข็งแรงมากสำหรับข้อมูลของโครงงาน ซึ่งประกอบด้วยตัวแปรลักษณะเดียวกัน เช่น Genetic_Risk + Lifestyle และสามารถรองรับความไม่เชิงเส้นของข้อมูลได้ดี

Yu และคณะ (2568) งานนี้นำเสนอวิธี Neural Network ที่ใช้การ re-weight ข้อมูลเพื่อจัดการ noise และความไม่สมดุลของข้อมูล survival รวมถึงการปรับ tuning ผ่าน grid optimization ผลลัพธ์เพิ่มขึ้นประมาณ 5% เทียบกับ baseline และใช้ข้อมูลคลินิก (Age, Gender, Cancer_Stage) ร่วมกับข้อมูลพันธุกรรมที่มีสัญญาณอ่อน

จากการศึกษาวรรณกรรมดังกล่าว ผู้ศึกษามองเห็นแนวทางปรับโครงสร้าง NN ในโครงงานของตนให้ทนทานต่อข้อมูลจริงที่อาจมี noise เช่น ค่า Genetic_Risk ที่ไม่สม่ำเสมอ หรือค่า Stage ที่มีความคลาดเคลื่อน โดยเฉพาะเมื่อทำการพยากรณ์ Survival_Years ที่มีความไวต่อตัวแปรผิดพลาดสูง

Nguyen (2563) งานนี้ใช้ component-wise gradient boosting สำหรับ survival prediction ในมะเร็งต่อมน้ำเหลือง พร้อมเสนอให้ใช้ Brier Score เป็น metric หลัก เนื่องจากไวต่อทั้ง calibration และ discrimination ของโมเดล โดยใช้ตัวแปรคลินิกสำคัญ ได้แก่ Cancer_Stage, Age, Gender ผลลัพธ์แสดงว่า GBM ให้โมเดลที่ตีความได้ดีและสามารถทำ feature selection ไปในตัว

จากการศึกษาวรรณกรรมดังกล่าว ผู้ศึกษาเห็นว่า GBM เป็นแกนกลางที่เหมาะกับโครงงาน และสามารถขยายตัวแปรจาก Cancer_Stage, Age, Gender ไปสู่ Smoking, Alcohol_Use, Obesity_Level, Genetic_Risk ได้โดยตรง ทั้งยังสามารถใช้ Brier Score เป็นตัววัดความแม่นยำเพิ่มเติมในชุดข้อมูลของโครงงาน

Ganie และคณะ (พ.ศ. 2568) ได้นำเสนอแนวทางการสร้างแบบจำลองเชิงพยากรณ์ขั้นสูงโดยใช้เทคนิค Stacking Ensemble ซึ่งเป็นการผสานพลังของอัลกอริทึมที่มีคุณลักษณะแตกต่างกัน ได้แก่ Random Forest (RF), Gradient Boosting (GBM) และ Neural Network (NN)

เพื่อเพิ่มขีดความสามารถในการพยากรณ์ระยะเวลาการรอดชีวิตของผู้ป่วยโรคมะเร็งหลายชนิด ผลการทดสอบประสิทธิภาพพบว่าแบบจำลองแบบรวมกลุ่มนี้สามารถลดความคลาดเคลื่อนในการพยากรณ์ได้อย่างมีนัยสำคัญ โดยให้ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) อยู่ที่ 0.82 และค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) อยู่ที่ 0.65 ซึ่งเป็นระดับความแม่นยำที่สูงกว่าการใช้โมเดลเดี่ยว (Single Model) ทั่วไปอย่างชัดเจน นอกจากนี้ยังมีการนำเทคนิค SHAP (Shapley Additive Explanations) มาใช้เพื่ออธิบายน้ำหนักความสำคัญของตัวแปร โดยพบว่าปัจจัยไลฟ์สไตล์ เช่น การดื่มแอลกอฮอล์ (Alcohol_Use) และปัจจัยทางประชากร เช่น อายุ (Age) มีผลต่อการพยากรณ์ในระดับที่สูงมาก

จากการศึกษาวรรณกรรมดังกล่าว ผู้วิจัยเห็นว่าการประยุกต์ใช้เทคนิค Ensemble (RF + GBM + NN) ที่พิสูจน์แล้วว่าให้ค่าความคลาดเคลื่อนต่ำ (RMSE 0.82 / MAE 0.65) มีความเหมาะสมอย่างยิ่งที่จะนำมาเป็นแกนหลักในโครงงานนี้ เนื่องจากสามารถยกระดับความเชื่อมั่นในการพยากรณ์ Survival_Years ให้สูงขึ้น โดยผู้วิจัยมีแผนที่จะนำผลลัพธ์ไปแสดงผลผ่านระบบ Dashboard หรือเว็บไซต์ และใช้เทคนิค SHAP เข้ามาช่วยในการแจกแจงอิทธิพลของตัวแปรสำคัญที่ปรากฏในชุดข้อมูล ได้แก่ ความเสี่ยงทางพันธุกรรม (Genetic_Risk), พฤติกรรมการสูบบุหรี่ (Smoking), ระดับความอ้วน (Obesity_Level) และชนิดของมะเร็ง (Cancer_Type) เพื่อให้ผู้ใช้งานสามารถทำความเข้าใจปัจจัยเสี่ยงส่วนบุคคลได้อย่างเป็นรูปธรรมและแม่นยำตามมาตรฐานงานวิจัยระดับสากล

Ganggayah และคณะ (2562) ได้ทำการศึกษาวิจัยเรื่องการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งปอด (Prediction of survival time of lung cancer patients) โดยใช้ตัวแบบการเรียนรู้ของเครื่อง (Machine Learning) งานวิจัยนี้มุ่งเน้นการเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการพยากรณ์ค่าต่อเนื่อง (Regression) เพื่อประมาณการระยะเวลาการรอดชีวิตที่แท้จริง ตัวแบบที่นำมาทดสอบประกอบด้วย Gradient Boosting Machines (GBM), Random Forest (RF) และ Support Vector Machines (SVM) โดยใช้ตัววัดประสิทธิภาพที่เป็น

เกณฑ์มาตรฐานทางสถิติ ได้แก่ Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) และค่าสัมประสิทธิ์การตัดสินใจ (R^2)

ผลการศึกษาพบว่า ตัวแบบในกลุ่ม Ensemble Learning ได้แก่ GBM และ RF มีประสิทธิภาพในการพยากรณ์ที่โดดเด่นกว่าตัวแบบอื่น โดยสามารถให้ค่าความคลาดเคลื่อน MAE และ RMSE ที่ต่ำกว่าอย่างมีนัยสำคัญ ซึ่งชี้ให้เห็นว่าการใช้ตัวแบบที่มีโครงสร้างแบบต้นไม้ตัดสินใจหลายต้นร่วมกัน (Tree-based Ensemble) มีความเหมาะสมอย่างยิ่งในการจัดการกับข้อมูลทางการแพทย์ที่มีความซับซ้อนและมีความสัมพันธ์ของตัวแปรที่ไม่เป็นเส้นตรง เพื่อให้ได้ผลพยากรณ์ระยะเวลาการอยู่รอดที่แม่นยำและนำไปใช้ประโยชน์ในเชิงคลินิกได้จริง

Yufan และคณะ (พ.ศ. 2568) ได้ทำการศึกษาการใช้ข้อมูลทางคลินิกและข้อมูลที่เกี่ยวข้องกับการรักษาในการสร้างแบบจำลอง Machine Learning เพื่อพยากรณ์การอยู่รอดของผู้ป่วยมะเร็งปอด โดยการศึกษานี้ใช้แบบจำลองหลายรูปแบบ เช่น Random Forest, Support Vector Machine และ Extreme Gradient Boosting (XGBoost) เพื่อเปรียบเทียบประสิทธิภาพของการพยากรณ์ผลลัพธ์ของผู้ป่วย ผลการศึกษาพบว่าการบูรณาการข้อมูลทางคลินิกหลายตัวแปรช่วยให้แบบจำลองสามารถทำนายผลลัพธ์ได้อย่างมีประสิทธิภาพ โดยมีการประเมินผลผ่านตัวชี้วัดด้านประสิทธิภาพของโมเดล เช่น ค่า AUC (Area Under the Curve) และการวิเคราะห์ ROC curve ซึ่งแสดงให้เห็นถึงความสามารถของโมเดลในการจำแนกและพยากรณ์การอยู่รอดของผู้ป่วยได้อย่างเหมาะสม

จากการวิเคราะห์ตัวแปรที่ใช้ในแบบจำลอง พบว่าตัวแปรพื้นฐานด้านประชากร เช่น อายุ (Age) และเพศ (Gender) รวมถึงข้อมูลทางคลินิกที่เกี่ยวข้องกับโรคมะเร็ง เป็นปัจจัยสำคัญที่มีผลต่อการพยากรณ์การอยู่รอดของผู้ป่วยมะเร็งปอด ซึ่งสะท้อนให้เห็นว่าการนำข้อมูลหลายมิติ เช่น ข้อมูลทางคลินิกและข้อมูลการรักษา มาประยุกต์ใช้ร่วมกับเทคนิค Machine Learning สามารถช่วยเพิ่มประสิทธิภาพของแบบจำลองในการพยากรณ์ผลลัพธ์ของผู้ป่วยได้อย่างมีนัยสำคัญ

เมื่อเชื่อมโยงกับชุดข้อมูลของผู้วิจัย พบว่ามีตัวแปรสำคัญหลายรายการที่สอดคล้องกับวรรณกรรมของ Yufan และคณะ ได้แก่ Gender, Age, Cancer_Type (Lung) และตัวแปร

ผลลัพธ์ Survival_Years ซึ่งเป็นตัวแปรสำคัญในการพยากรณ์ แม้ชุดข้อมูลของผู้วิจัยอาจไม่ได้รวมข้อมูลทางการแพทย์หรือข้อมูลการรักษาทั้งหมดที่ปรากฏในงานวิจัยดังกล่าว แต่ตัวแปรพื้นฐานด้านประชากรศาสตร์และลักษณะของโรคยังคงเป็นตัวพยากรณ์ที่มีความสำคัญในแบบจำลอง Machine Learning หลายประเภท

จากการศึกษาวรรณกรรมดังกล่าว ผู้วิจัยจึงเห็นว่าการนำแบบจำลอง Machine Learning เช่น Random Forest และ Gradient Boosting มาประยุกต์ใช้ในการพยากรณ์การอยู่รอดของผู้ป่วยมะเร็ง เป็นแนวทางที่เหมาะสม เนื่องจากแบบจำลองดังกล่าวสามารถจัดการกับข้อมูลที่มีความซับซ้อนและมีตัวแปรจำนวนมากได้ดี อีกทั้งยังสามารถนำผลการวิเคราะห์มาแสดงในรูปแบบ Dashboard เพื่อช่วยให้ผู้ใช้งานสามารถเข้าใจข้อมูลและแนวโน้มของผลการพยากรณ์ได้อย่างชัดเจนมากยิ่งขึ้น

Lynch และคณะ (2561) ได้ทำการศึกษาวิจัยเชิงเปรียบเทียบระหว่างอัลกอริทึมการเรียนรู้ของเครื่องที่หลากหลายเพื่อพยากรณ์การอยู่รอดของผู้ป่วยมะเร็งเต้านม งานวิจัยนี้ได้นำโมเดลยอดนิยมมาทดสอบประสิทธิภาพ ประกอบด้วย Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Neural Network (NN) และ Logistic Regression (LR) โดยใช้เกณฑ์การประเมินประสิทธิภาพในด้านความแม่นยำ (Accuracy) ความไว (Sensitivity) และความจำเพาะ (Specificity)

จากการศึกษาวรรณกรรมดังกล่าว ผู้วิจัยพบว่า ตัวแบบ Support Vector Machine (SVM) และ Random Forest (RF) มีประสิทธิภาพโดดเด่นที่สุดในการพยากรณ์การอยู่รอด โดยให้ค่าความแม่นยำสูงสุดเมื่อเทียบกับโมเดลอื่น งานวิจัยระบุว่าเทคนิคการเรียนรู้ของเครื่องสามารถวิเคราะห์ปัจจัยเสี่ยงที่มีความซับซ้อนและช่วยให้บุคลากรทางการแพทย์สามารถพยากรณ์ผลลัพธ์การรักษาได้อย่างแม่นยำยิ่งขึ้น ซึ่งส่งผลดีต่อการวางแผนการดูแลผู้ป่วยแบบเฉพาะเจาะจง

จากการศึกษาวรรณกรรมทั้งหมด พบว่ามีการใช้เทคนิค Random Forest , Gradient Boosted Tree และ Neural Networks อย่างแพร่หลายในการพยากรณ์ระยะเวลาการอยู่รอด ซึ่งสามารถนำมาพยากรณ์ได้โดยตรง โดยจะใช้ attribute ใน dataset เช่น Survival_Years , Age ,

Gender, Cancer Type, Cancer Stage ซึ่งเป็น attribute ส่วนใหญ่ที่เกี่ยวข้องกับวรรณกรรมแต่จะมี Age และ Cancer Type ที่มีในทุกวรรณกรรมที่ยกมา สามารถนำมาปรับใช้และนำไปคาดการณ์ระยะเวลาการอยู่รอดของผู้ป่วยได้

2.5 บทสรุป

จากการศึกษาแนวคิด ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง สามารถสรุปได้ว่า การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกจำเป็นต้องอาศัยกระบวนการวิเคราะห์ข้อมูลที่ถูกต้องและมีระบบ ตั้งแต่การทำความสะอาดข้อมูล (Data Cleaning) การพยากรณ์ข้อมูล (Forecasting) ไปจนถึงการเปรียบเทียบโมเดล (Model Comparison) เพื่อเลือกแบบจำลองที่เหมาะสมที่สุด ทั้งนี้ การเปรียบเทียบโมเดลถือเป็นหัวใจสำคัญ เนื่องจากช่วยให้ผู้วิจัยสามารถประเมินประสิทธิภาพของแบบจำลองได้รอบด้าน ทั้งด้านความแม่นยำ ความคงเส้นคงวา ความเร็วในการทำงาน และความสามารถในการจัดการกับข้อมูลที่ซับซ้อน