บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

โครงงานเรื่อง การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลเพื่อใช้เป็นแนวทางการศึกษา ประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

2.1.1 การวิเคราะห์ข้อมูล (Data Analytic)

2.1.2 การประยุกต์ใช้ข้อมูลทางสถิติ

2.1.3 การแสดงผลข้อมูล (Data Visualization)

2.1.4 การเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

2.2 ทฤษฎี

2.2.1 การทำเหมืองข้อมูล (Data Mining)

2.2.2 หลักในการทำเหมืองข้อมูล

2.2.3 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

2.3.2 เทคนิคการถดถอยเชิงเส้น (Linear Regression)

2.3.3 เทคนิคเกาส์เซียน (Gussian Process)

2.3.4 เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression)

2.3.5 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks)

2.3.6 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP–DM)

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 การวิเคราะห์ข้อมูล (Data Analytic)

ในการดำเนินงานเรื่องการเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ทางผู้วิเคราะห์ข้อมูลได้ศึกษาหลักการและทฤษฎีต่างๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูลซึ่งมีรายละเอียด ดังนี้

Data Analytics คือการนำข้อมูลที่มีอยู่มาเข้าสู่กระบวนการวิเคราะห์ เพื่อนำข้อมูลไปใช้ประโยชน์ในด้านต่างๆที่ครอบคลุมในทุกอุตสาหกรรม ทั้งภาครัฐและภาคเอกชน ซึ่งจะต้องอาศัยเทคโนโลยีหรือซอฟต์แวร์เฉพาะด้านที่เข้ามาช่วยให้การประมวลผลและวิเคราะห์ข้อมูลเป็นไปอย่างแม่นยำและนำไปใช้งานได้อย่างมีประสิทธิภาพ โดยสามารถแบ่งประเภทของการวิเคราะห์ข้อมูล ได้ 4 รูปแบบ ตั้งแต่ Descriptive Analytics คือการวิเคราะห์ว่าเกิดอะไรขึ้น Diagnostic Analytics คือการวิเคราะห์ว่าสิ่งนั้น ๆ เกิดขึ้นเพราะอะไร, Predictive Analytics คือวิเคราะห์เพื่อคาดการณ์แนวโน้มของสิ่งที่จะเกิดต่อไป และ Prescriptive Analytics คือการวิเคราะห์สิ่งที่จะเกิดขึ้นจากการใช้ทางเลือกต่าง ๆ ในเชิงแนะนำ สำหรับรูปแบบของการวิเคราะห์ข้อมูล (Data Analytics) สามารถแบ่งได้ดังนี้

1. Descriptive Analytics คือการวิเคราะห์ข้อมูลแบบพื้นฐาน เพื่อแสดงผลที่เกิดขึ้น หรือกำลังจะเกิดขึ้น จากข้อมูลในอดีต ในลักษณะที่เข้าใจง่ายสามารถสร้างขึ้นได้ด้วยตนเอง เช่น รายงาน แผนภูมิ กราฟ ตาราง เป็นต้น ซึ่งจะช่วยให้เข้าใจการเปลี่ยนแปลงที่เกิดขึ้นกับองค์กรได้ดียิ่งขึ้น เช่น การเปลี่ยนแปลงของราคาสินค้ารายปี การเติบโตของยอดขายรายเดือน การเปรียบเทียบยอดขายในแต่ละสาขาหรือแต่ละช่องทาง การเปรียบเทียบจำนวนผู้ใช้งานเว็บไซต์ในแต่ละช่วงเวลา เป็นต้น ซึ่ง Descriptive Analytics คือการอธิบายสิ่งที่เกิดขึ้นตามวัตถุประสงค์และช่วงเวลาที่กำหนดนั่นเอง

2. Diagnostic Analytics คือการวิเคราะห์เชิงวินิจฉัย ซึ่งเป็นรูปแบบหนึ่งของการวิเคราะห์ขั้นสูงแบบเจาะลึก โดยการวิเคราะห์ข้อมูลที่ได้จากการทำ Descriptive Analytics เพื่อหาคำตอบว่าทำไมจึงเกิดสิ่งนั้น ๆ หรืออธิบายปัจจัยและตัวแปรที่เป็นสาเหตุของการเกิดสิ่งนั้น ๆ ขึ้น ซึ่งจะต้องอาศัยเทคนิคต่าง ๆ เข้ามาช่วย เช่น การทำ Data discovery หรือ Data mining เป็นต้น ตัวอย่างการวิเคราะห์ข้อมูลเชิงวินิจฉัย เช่น อธิบายสาเหตุที่ทำให้ยอดขายเพิ่มขึ้น จำนวน ผู้ใช้ที่เข้าชมเว็บไซต์เพิ่มขึ้น จำนวนลูกค้าที่เข้าใช้บริการที่หน้าร้านลดลง โปรโมชันที่ไม่ค่อยได้รับความนิยม เนื้อหาโฆษณาที่ได้ CTR% มากกว่าเนื้อหาอื่น ๆ หรือวิเคราะห์การทำงานของเครื่องคอมพิวเตอร์เพื่อตรวจจับความผิดปกติ เป็นต้น ข้อมูลเหล่านี้จะทำให้องค์กร

รู้ถึงความต้องการของตลาด เข้าใจพฤติกรรมของลูกค้า รู้สาเหตุของปัญหาด้านเทคโนโลยี รวมถึงสามารถปรับปรุงวัฒนธรรมองค์กร เพื่อการทำงานที่ดีขึ้น

3. Predictive Analytics คือการวิเคราะห์ข้อมูลทั้งข้อมูลในอดีตและปัจจุบัน ออกมาในเชิงคาดการณ์ ทำนาย หรือการพยากรณ์ เพื่อหาแนวโน้มที่จะเกิดสิ่งต่าง ๆ ขึ้นตามวัตถุประสงค์ที่กำหนด โดยการสร้างแบบจำลองทางสถิติ บวกกับการนำเทคโนโลยีปัญญาประดิษฐ์มาใช้ ซึ่งสามารถสร้างประโยชน์ได้มากมายในหลายแง่มุม เช่น การคาดการณ์ความเสี่ยงและโอกาส ยอดขาย
ภัยไซเบอร์ สภาพอากาศ การลงทุน หุ้น หรือผลการเลือกตั้ง เป็นต้น อย่างไรก็ตามการทำ Predictive Analytics ที่ถูกต้องและแม่นยำนั้น ขึ้นอยู่กับคุณภาพของข้อมูล ซึ่งเป็นสิ่งที่องค์กรควรให้ความสำคัญเป็นอันดับแรก โดยการเตรียมข้อมูลให้มีคุณภาพที่ดีและเหมาะสม ก่อนนำไปใช้วิเคราะห์เพื่อให้เกิดผลลัพธ์ที่มีประสิทธิภาพ ลดข้อผิดพลาด และเกิดประโยชน์อย่างแท้จริง

4. Prescriptive Analytics คือการวิเคราะห์แบบให้คำแนะนำ ซึ่งเป็นการวิเคราะห์ที่มีความซับซ้อนมากที่สุด ต่อเนื่องจากการทำ Predictive Analytics กล่าวคือ เมื่อได้ข้อมูลแนวโน้มที่จะเกิดบางสิ่งขึ้นแล้ว การทำ Prescriptive Analytics จะช่วยแนะนำแนวทางการดำเนินการในขั้นตอนต่อไปที่เหมาะสมที่สุด และวิเคราะห์ไปถึงผลที่จะเกิดขึ้นถ้าหากเลือกปฏิบัติตามแนวทางนั้น ๆ หรือแม้แต่แนะนำแนวทางในการรับมือและแก้ไขปัญหา การวิเคราะห์แบบให้คำแนะนำจึงถือเป็นเครื่องมือที่สำคัญอย่างมากสำหรับการตัดสินใจที่ขับเคลื่อนด้วยข้อมูล การวิเคราะห์แบบให้คำแนะนำคือทำงานร่วมกันระหว่าง Big data อัลกอริธึมของ Machine learning และเทคโนโลยี AI เพื่อช่วยในการวิเคราะห์ข้อมูลจำนวนมหาศาลที่มีความซับซ้อนเกินกว่าที่มนุษย์จะทำได้ ซึ่งการทำ Prescriptive Analytics ยังช่วยเพิ่มประสิทธิภาพในการตัดสินใจด้านต่างๆ ไม่ว่าจะเป็นด้านการลงทุน ด้านการตลาด ด้านการตรวจจับการฉ้อโกง (Fraud Detection) ในอุตสาหกรรมธนาคาร ด้านการพัฒนาและปรับปรุงผลิตภัณฑ์ เป็นต้น ซึ่งนับว่าการวิเคราะห์ที่ทรงพลังเป็นอย่างมาก และเป็นแรงขับเคลื่อนสำคัญที่จะช่วยให้องค์กรสามารถบรรลุเป้าหมายทางธุรกิจได้อย่างรวดเร็วและมีประสิทธิภาพ

2.1.2 การประยุกต์ใช้ข้อมูลทางสถิติ

สถิติ เป็นกระบวนการที่เกิดขึ้นเพื่อแก้ปัญหาเกี่ยวกับปรากฏการณ์ธรรมชาติทางวิทยาศาสตร์ ทางสังคม การบริหารธุรกิจ และการปกครอง จัดเป็นเทคนิคที่ใช้แก้ปัญหาโดยการวินิจฉัยข้อมูล ช่วยให้เกิดการตัดสินใจจากพื้นฐานความไม่แน่นอน เป็นวิธีทางวิทยาศาสตร์ที่ช่วยในการตัดสินใจในรูปของตัวเลข การรวบรวมข้อความจริงมาศึกษา จึงกล่าวได้ว่าสถิติเป็น

เทคนิคเกี่ยวกับตัวอย่าง ปัจจุบันวิทยาศาสตร์ทุกสาขา คือสถิตินั่นคือต้องมีการทดลองทางวิทยาศาสตร์ทุกแขนง ผลที่ได้จัดเป็นเพียงสถิติที่ใช้อ้างอิง ปัจจุบันกระบวนการทางวิทยาศาสตร์ใช้วิธีการสรุปผลงานที่ได้จากการทดลองในรูปของตัวอย่างและนำมากล่าว โดยอ้างหลักของความน่าจะเป็นหรือโอกาส (probability) โดยใช้หลักการตัดสินใจทางสถิติ (statistical decision) ตัวอย่างการใช้ข้อมูลสถิติสำหรับการพัฒนาในด้านต่างๆ

1. ด้านการศึกษา ในการกำหนดนโยบายและการวางแผนพัฒนาการศึกษาและการกระจายโอกาสทางการศึกษาของประชาชนในระดับการศึกษาต่างๆ ข้อมูลสำคัญที่ต้องการใช้ ได้แก่ ประชากรก่อนวัยเรียนและวัยเรียน บุคลากรทางการศึกษา ปริมาณการผลิตและพัฒนาครูในแต่ละสาขา จำนวนสถานศึกษา ค่าใช้จ่ายในแต่ละระดับการศึกษา เป็นต้น

2. ด้านการเกษตร ในการกำหนดนโยบายและวางแผนพัฒนาทางการเกษตรของประเทศ ข้อมูลที่ต้องการใช้ ได้แก่ ครัวเรือนที่ทำการเกษตร เนื้อที่การเพาะปลูก ผลิตผลทางการเกษตร จำนวนปศุสัตว์ ราคาสินค้าเกษตรกรรม เครื่องมือเครื่องใช้ทางการเกษตร ภาวะเศรษฐกิจและสังคมของครัวเรือนเกษตร การประมง การป่าไม้ ข้อมูลเกี่ยวกับแหล่งน้ำและการชลประทาน เป็นต้น

3. ด้านอุตสาหกรรม ใช้จัดทำแผนงานหรือกำหนดนโยบายและส่งเสริมอุตสาหกรรม ส่งเสริมการลงทุนและพัฒนาเทคโนโลยีทางด้านอุตสาหกรรม ได้แก่ ข้อมูลเกี่ยวกับปริมาณการผลิตทางอุตสาหกรรม ต้นทุนการผลิต จำนวนแรงงาน ค่าใช้จ่ายของสถานประกอบการ มูลค่าเพิ่ม ฯลฯ

4. ด้านรายรับ – รายจ่ายของครัวเรือน เป็นข้อมูลที่มีความสำคัญที่ใช้วัดความเจริญเติบโตทางเศรษฐกิจ การครองชีพและการกระจายรายได้ของประชากร ซึ่งเป็นเครื่องชี้วัดที่สำคัญของผลการพัฒนาประเทศ ข้อมูลสถิติที่สำคัญ ได้แก่ รายได้รายจ่ายของครัวเรือน ภาวะหนี้สิน สภาพความเป็นอยู่ ที่อยู่อาศัยของครัวเรือน เป็นต้น

5. ด้านสาธารณสุข การจัดทำแผนพัฒนาด้านสาธารณสุข การพัฒนางานวิชาการทางการแพทย์/สาธารณสุข เพื่อให้ประชาชนมีสุขภาพอนามัยที่ดี จำเป็นต้องใช้สถิติเกี่ยวกับการเกิด การตาย การเจ็บป่วยของประชาชน การรักษาพยาบาล ความเป็นอยู่และสภาพทางสังคมของประชากร การอนามัยและสุขาภิบาลพฤติกรรมด้านการบริโภค การสูบบุหรี่และดื่มสุรา เป็นต้น

6. ด้านคมนาคมและขนส่ง การปรับปรุงบริการและพัฒนาทางการคมนาคมขนส่ง และการสื่อสารของประเทศ เพื่อสนับสนุนการพัฒนาด้านต่างๆ และกระจายความเจริญไปสู่ภูมิภาค ข้อมูลที่ใช้ได้แก่ รายรับ – รายจ่ายของการประกอบการขนส่ง ปริมาณผู้ใช้บริการ

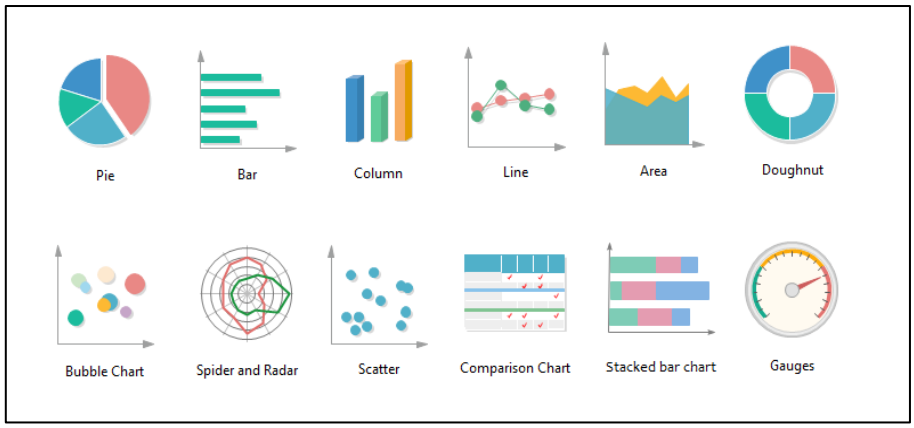
ในแต่ละเส้นทาง ปริมาณการขนส่งทางถนน ทางน้ำและทางอากาศ รายละเอียดเส้นทางคมนาคม ข้อมูลเกี่ยวกับการจัดสรรความถี่วิทยุ จำนวนครัวเรือนที่มีเครื่องรับวิทยุ โทรทัศน์ เป็นต้น

2.1.3 การแสดงผลข้อมูล (Data Visualization)

เฟิร์สคราฟท์ ดิจิทัล โซลูชั่น, 2563 ได้ให้ความหมายของ Data Visualization คือ การนำข้อมูลหรือ Data ที่ได้มาจากแหล่งต่างๆ มาวิเคราะห์ประมวลผลแล้วนำเสนอออกมาในรูปแบบที่มองเห็นและทำความเข้าใจได้ด้วยตา เช่น แผนภูมิ รูปภาพ แผนที่ กราฟแสดงเทรนด์ ตาราง วิดีโอ อินโฟกราฟิก (Infographic) แดชบอร์ด (DashBoard) จุดประสงค์สำคัญของการทำ Data Visualization คือ การนำเสนอข้อมูลให้ เข้าใจง่าย ผู้อ่านข้อมูลสามารถเข้าใจได้ทันทีว่าตัวชิ้นงาน (Media) ต้องการสื่อสารอะไร จุดสำคัญของเนื้อหา และชี้ข้อเปรียบเทียบให้เห็นอย่างชัดเจน ช่วยให้สังเกตเห็นจุดที่น่าสนใจของข้อมูลได้ง่ายขึ้น

Data Visualization มีหลากหลายรูปแบบและไม่จำกัดว่าต้องใช้รูปแบบต่อไปนี้ ในการนำเสนอข้อมูลเท่านั้น เพราะแต่ละรูปแบบก็มีฟังก์ชันเฉพาะของการนำเสนอข้อมูลบางรูปแบบใช้เปรียบเทียบข้อมูลแต่ละชุดได้ดี บางรูปแบบช่วยให้มองเห็นเทรนด์ได้ง่าย ช่วยเล่าข้อมูลที่ไกลตัวให้เข้าใจได้ง่ายโดยการเปรียบเทียบให้สอดคล้องกับสิ่งที่คุ้นเคยในชีวิตประจำวัน

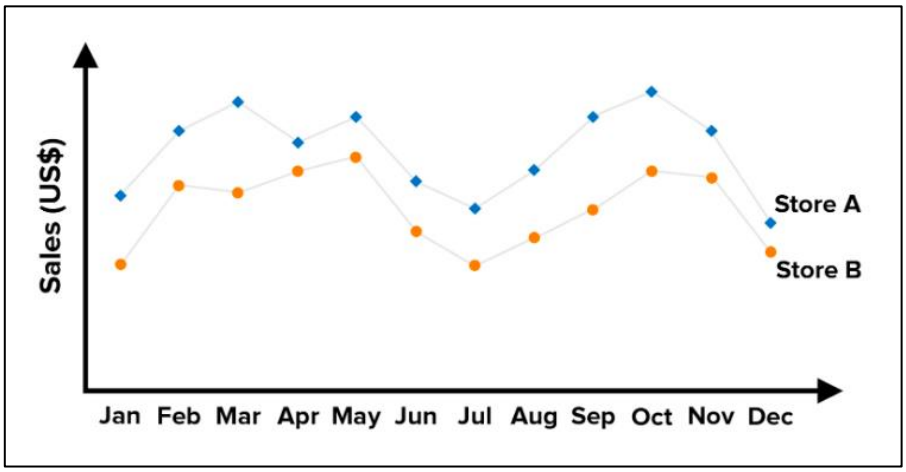
1) แผนภูมิ (Charts) ซึ่งเป็นรูปแบบที่น่าจะคุ้นเคยกันมากที่สุด และเป็นรูปแบบที่มีหลากหลายชนิดที่เหมาะกับการนำเสนอข้อมูลที่แตกต่างกันไปตามวัตถุประสงค์ เช่น Pie chart จะช่วยให้เราเห็นปริมาณความแตกต่างได้ชัดเจน, Comparison chart เหมาะสำหรับการเปรียบเทียบคุณสมบัติหลายๆ ข้อ มาตรวัด (Gauges) จะช่วยให้เห็นความเข้มข้น ความรุนแรง หรือน้ำหนัก



ภาพที่ 2.1 แสดงตัวอย่างแผนภูมิรูปแบบต่าง ๆ

(ที่มา: medium.com/@Lynia_Li/)

2) กราฟ (Graphs) คือ Subset หรือประเภทหนึ่งของแผนภูมิ โดยกราฟจะทำหน้าที่แสดงความสัมพันธ์ระหว่างข้อมูล 2 ตัวแปร ผ่านแกนแนวนอน (แกน X) และแกนแนวตั้ง (แกน Y) ช่วยให้เห็นเทรนด์สถานการณ์ประกอบกับบริบทได้เป็นอย่างดี



ภาพที่ 2.2 แสดงกราฟตัวอย่างยอดขายแต่ละเดือน

(ที่มา: mindtools.com)

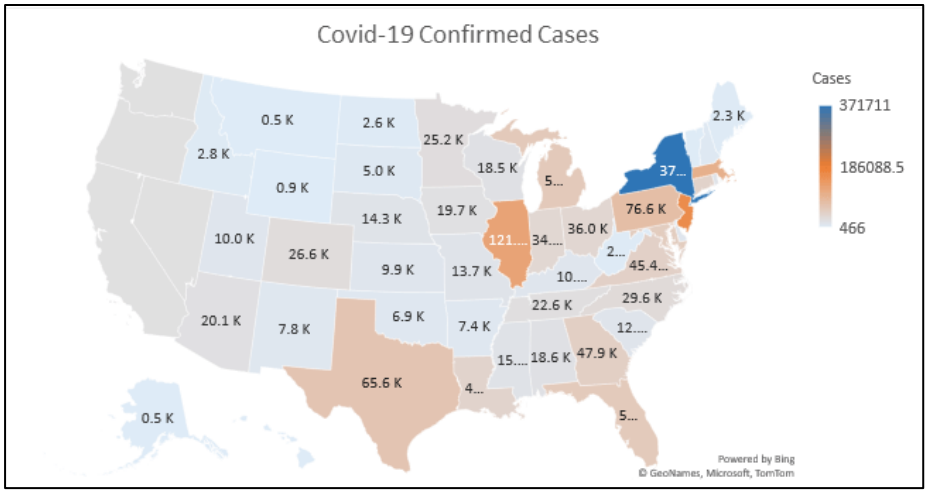
3) ตาราง (Tables) ก็เป็นอีกรูปแบบที่ใช้กันมากเพื่อนำเสนอข้อมูลให้ออกมา ดูง่าย ตารางประกอบไปด้วย 2 ส่วน ได้แก่ คอลัมน์และแถว ซึ่งช่วยจัดการข้อมูลให้เรียบร้อย ช่วยให้มองเห็นบริบทและความสัมพันธ์ของข้อมูลหลายๆ ชุดได้อย่างง่ายดาย

Marks	Number of Students		Total
	Males	Females	
30 – 40	8	6	14
40 – 50	16	10	26
50 – 60	14	16	30
60 – 70	12	8	20
70 – 80	6	4	10
Total	56	44	100

ภาพที่ 2.3 แสดงตัวอย่างตารางช่วงอายุ

(ที่มา: embibe.com)

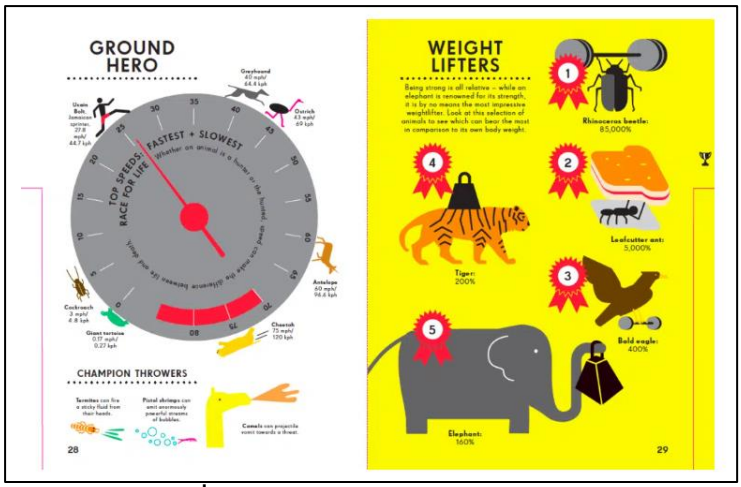
4) แผนที่ (Maps) เป็นการนำเสนอข้อมูลบนแผนที่เพื่อแสดงข้อมูลเกี่ยวกับ พื้นที่ต่างๆ ยกตัวอย่างเช่น การนำเสนอข้อมูลยอดผู้ติดเชื้อ Covid-19 ในแต่ละรัฐของประเทศสหรัฐอเมริกา ซึ่งนอกจากการใส่ข้อมูลลงไปยังพื้นที่ต่างๆ แล้ว ยังสามารถใช้สีสันเพื่อบอกช่วงปริมาณหรือความหนาแน่นของผู้ติดเชื้ออีกด้วย



ภาพที่ 2.4 แสดงแผนที่จำนวนผู้ป่วย Covid-19

(ที่มา: spreadsheetweb.com)

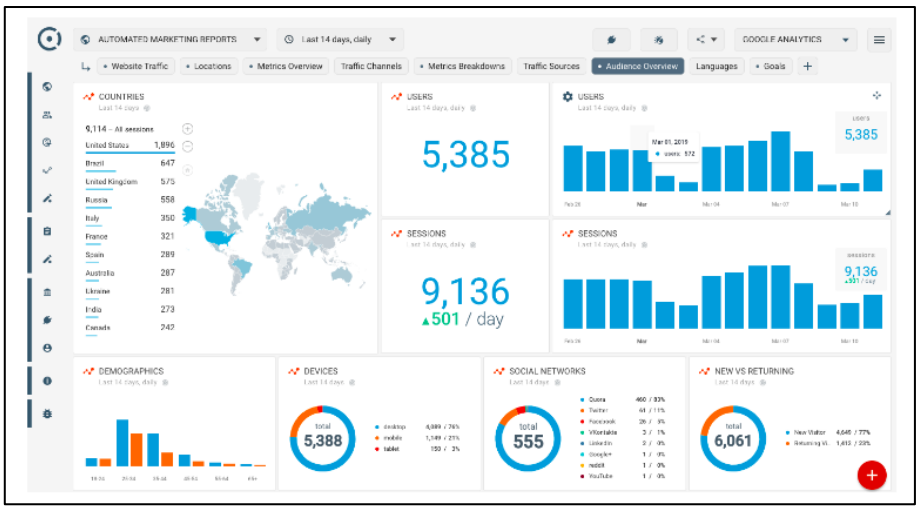
5) อินโฟกราฟิก (Infographic) คือการนำเสนอสารสนเทศ (Info: information) ด้วยภาพกราฟิก (Graphic) เป็นรูปแบบการนำเสนอข้อมูลที่ใช้ภาพสื่อแทน ทำให้ผู้อ่านข้อมูล เข้าใจข้อมูลได้ง่ายหรือสามารถทำความเข้าใจผ่านภาพแทนที่คุ้นเคย นอกจากนี้อินโฟกราฟิกยังเป็นรูปแบบการนำเสนอข้อมูลที่น่าสนใจ มีการนำเทคนิคการเล่าเรื่อง (Storytelling) มาใช้ทำให้ข้อมูลน่าสนใจน่าดึงดูดจึงมักจะใช้เพื่อนำเสนอเนื้อหา ความรู้ หรือเป็นสื่อการเรียนการสอน



ภาพที่ 2.5 แสดงภาพอินโฟกราฟิก

(ที่มา: simonrogers.net)

6) แดชบอร์ด (Dashboards) คือ การนำข้อมูลต่างๆ มาเรียบเรียงและสรุปเป็นภาพ โดยใช้แผนภูมิและกราฟต่างๆ มาใช้นำเสนอ ปัจจุบันแดชบอร์ดเป็น Data Visualization ที่นิยมใช้กับการนำเสนอข้อมูลแบบ Real-time ผ่านซอฟต์แวร์หรือเครื่องมือจัดการและวิเคราะห์ข้อมูลต่างๆ เช่น เครื่องมือการตลาด เครื่องมือบริหารจัดการข้อมูล เครื่องมือติดตามและดูแลเว็บไซต์



ภาพที่ 2.6 แสดงตัวอย่างหน้าแดชบอร์ด

(ที่มา: octoboard.com)

2.1.4 การเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

การเกิดอุบัติเหตุบนท้องถนนเป็นหนึ่งในปัญหาที่สำคัญของสังคมไทย โดยเฉพาะในช่วงเทศกาลสำคัญ เช่น เทศกาลปีใหม่ ที่ประชาชนมักเดินทางกลับภูมิลำเนาหรือออกท่องเที่ยว ทำให้จำนวนยานพาหนะบนท้องถนนเพิ่มขึ้นอย่างมาก ส่งผลให้ความเสี่ยงต่อการเกิดอุบัติเหตุเพิ่มสูงขึ้นตามไปด้วย นอกจากนี้สถิติการเกิดอุบัติเหตุในช่วงเทศกาลนี้ยังมีแนวโน้มเพิ่มขึ้นเรื่อยๆ โดยสาเหตุหลักมักจะมาจากการดื่มสุราแล้วขับขี่ การขับรถเร็วเกินกว่าที่กฎหมายกำหนด และความเหนื่อยล้าจากการเดินทางไกล ซึ่งนำไปสู่การสูญเสียทั้งชีวิตและทรัพย์สินจากสถิติในช่วงเทศกาลปีใหม่ พบว่ามีจำนวนอุบัติเหตุที่เกิดขึ้นและจำนวนผู้เสียชีวิตหรือบาดเจ็บเพิ่มขึ้นอย่างมีนัยสำคัญเมื่อเทียบกับช่วงเวลาปกติ ปัญหาการเกิดอุบัติเหตุบนท้องถนนในช่วงเทศกาลปีใหม่ยังเป็นประเด็นที่รัฐบาลและหน่วยงานที่เกี่ยวข้องให้ความสำคัญเป็นพิเศษ โดยมีการจัดตั้งศูนย์อำนวยการความปลอดภัยทางถนนเพื่อรณรงค์และลดอุบัติเหตุในช่วงดังกล่าว การเก็บสถิติจากกลุ่มป้องกันการบาดเจ็บจากการจราจรในประเทศไทยพบว่าในปี 2557 เพศชายมีการตายจากอุบัติเหตุทางถนนสูงกว่าเพศหญิงประมาณ 3 เท่า กลุ่มอายุ

ที่เสียชีวิตจากอุบัติเหตุทางถนนมากที่สุด คือ กลุ่มอายุ 15–19 ปี รถจักรยานยนต์เป็นพาหนะที่มีจำนวนการเสียชีวิตมากที่สุด ช่วงเวลาที่มีจำนวน ผู้เสียชีวิตมากที่สุดคือเดือนมกราคม ซึ่งบุคคลากรในประเทศชาติถือว่าเป็นส่วนสำคัญในการพัฒนาประเทศ นอกจากนี้รัฐบาลยังต้องสูญเสียงบประมาณ ในการบริหารประเทศเพื่อมารณรงค์ ป้องกัน แก้ไขปัญหา และรักษาผู้บาดเจ็บจากอุบัติเหตุที่นับวันจะยิ่งจะมีมากขึ้นเรื่อยๆ

เทศกาลไทย ไม่ว่าจะเป็นเทศกาลตามประเพณีไทยหรือสากล อย่างเช่น วันขึ้นปีใหม่ วันสงกรานต์ วันตรุษจีน หรือวันวาเลนไทน์ เป็นต้น ซึ่งเทศกาลเหล่านี้เป็นเทศกาลแห่งความสุข ความรื่นเริงและเป็นช่วงเวลาที่ได้เดินทางไปพบปะสังสรรค์กันระหว่างคนในครอบครัว/ญาติ หรือเพื่อน โดยเฉพาะในเทศกาลที่มีวันหยุดติดต่อกันหลายวัน อย่างไรก็ตามในการเฉลิมฉลองของเทศกาลแห่งความสุขเหล่านี้ยังคงมีความเชื่อที่ผิดๆ ที่ต้องสังสรรค์กันด้วยสุราหรือเครื่องดื่มที่มีแอลกอฮอล์ ซึ่งเป็นสาเหตุหลักของการเกิดอุบัติเหตุในระหว่างการเดินทางที่ส่งผลให้สูญเสียทั้งชีวิตและทรัพย์สิน โดยเฉพาะอย่างยิ่งอุบัติเหตุจากการจราจรทางบก

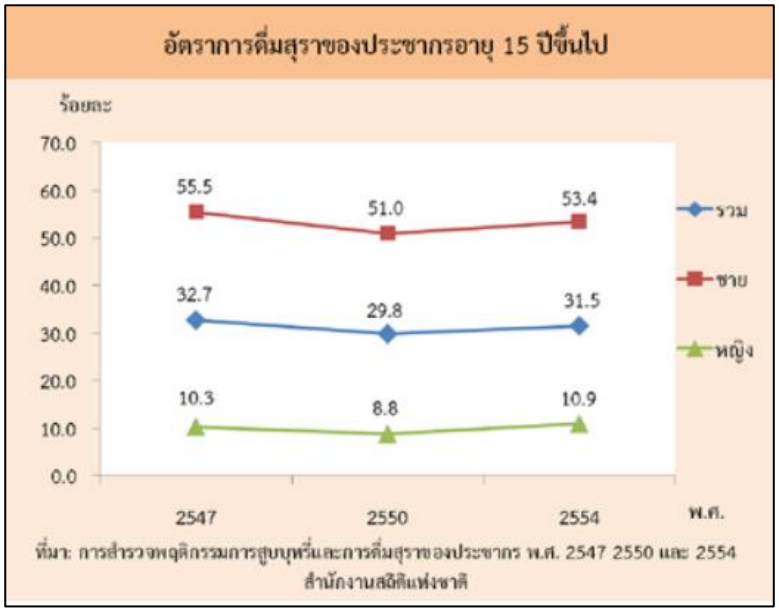
ภาพที่ 2.7 สถิติอุบัติเหตุจราจรทางบกในเทศกาลปีใหม่ (7 วันอันตราย)

(ที่มา: https://www.thaihealth.or.th/?p=231763)

สาเหตุของการเกิดอุบัติเหตุนั้น ส่วนใหญ่เกิดจากสุราหรือเครื่องดื่มที่มีแอลกอฮอล์มากที่สุด แม้หลายฝ่ายทั้งภาครัฐและเอกชนจะร่วมมือร่วมใจกันรณรงค์ “เมาไม่ขับ” เพื่อป้องกันการเกิดอุบัติเหตุก็ตามแต่ก็ยังคงเป็นสาเหตุอันดับต้นๆ ของการเกิดอุบัติเหตุ แม้จะมีแนวโน้มลดลงจากร้อยละ 25.6 ในปี 2552 เหลือร้อยละ 19.8 ในปี 2553 ก็ตาม เป็นที่น่าสังเกตว่า การขับรถเร็ว/ขับรถตัดหน้า/ขับรถตามกระชั้นชิด มีแนวโน้มเพิ่มขึ้นอย่างชัดเจน ซึ่งพฤติกรรมการขับรถแบบนี้มีผลต่อการเกิดอุบัติเหตุได้ทั้งสิ้น

ภาพที่ 2.8 สาเหตุการเกิดอุบัติเหตุการจราจรทางบก (5 อันดับแรก)

(ที่มา: https://www.thaihealth.or.th/?p=231763)



ภาพที่ 2.9 อัตราการดื่มสุราของประชากรอายุ 15 ปีขึ้นไป

(ที่มา: https://www.thaihealth.or.th/?p=231763)

2.2 ทฤษฎี

2.2.1 การทำเหมืองข้อมูล (Data Mining)

จรัสสรี รุ่งรัตนาอุบล, สำนักพิมพ์มหาวิทยาลัยนเรศวร, 2566 ได้ให้ความหมายของการทำเหมืองข้อมูลคือ เทคนิคที่ใช้คอมพิวเตอร์ช่วยในการวิเคราะห์เพื่อประมวลผลและสำรวจชุดข้อมูลขนาดใหญ่ เมื่อใช้เครื่องมือและวิธีการทำเหมืองข้อมูล องค์กรสามารถค้นพบรูปแบบ

และความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลของตน การทำเหมืองข้อมูลแปลงข้อมูลดิบเป็นความรู้เชิงปฏิบัติ บริษัทใช้ความรู้นี้ในการแก้ไขปัญหา วิเคราะห์ผลกระทบในอนาคตของการตัดสินใจทางธุรกิจ และเพิ่มขอบเขตกำไรของบริษัท

ขั้นตอนการทำเหมืองข้อมูล ที่นิยมใช้ในปัจจุบันมีหลายแบบแต่ที่จะนำมากล่าวคือวิธี Cross-Industry Standard Process for Data Mining (CRISP-DM) ที่มีการพัฒนาเป็น Workflow มาตรฐานสำหรับการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอน คือ

1. ความเข้าใจทางธุรกิจ

นักวิทยาศาสตร์ข้อมูลหรือผู้ทำเหมืองข้อมูลเริ่มต้นด้วยการระบุวัตถุประสงค์และขอบเขตของโครงการ พวกเขาร่วมมือกับผู้มีส่วนได้ส่วนเสียทางธุรกิจเพื่อระบุข้อมูลบางอย่าง จากนั้นพวกเขาใช้ข้อมูลนี้เพื่อกำหนดเป้าหมายการทำเหมืองข้อมูลและระบุทรัพยากรที่จำเป็นสำหรับการค้นพบความรู้

2. ความเข้าใจข้อมูล

เมื่อพวกเขาเข้าใจปัญหาทางธุรกิจแล้ว นักวิทยาศาสตร์ด้านข้อมูลจะเริ่มวิเคราะห์ข้อมูลเบื้องต้น พวกเขารวบรวมชุดข้อมูลจากแหล่งต่างๆ รับสิทธิ์การเข้าถึง และเตรียมรายงานคำอธิบายข้อมูล รายงานประกอบด้วยประเภทข้อมูล ปริมาณ และข้อกำหนดด้านฮาร์ดแวร์และซอฟต์แวร์สำหรับการประมวลผลข้อมูล เมื่อธุรกิจได้รับการอนุมัติแผนแล้ว พวกเขาจะเริ่มสำรวจและตรวจสอบข้อมูล พวกเขาจัดการข้อมูลโดยใช้เทคนิคทางสถิติพื้นฐาน ประเมินคุณภาพข้อมูล และเลือกชุดข้อมูลสุดท้ายสำหรับขั้นตอนต่อไป

3. การเตรียมข้อมูล

ผู้ทำเหมืองข้อมูลใช้เวลาส่วนใหญ่ในขั้นตอนนี้ เนื่องจากซอฟต์แวร์การทำเหมืองข้อมูลต้องการข้อมูลคุณภาพสูง กระบวนการทางธุรกิจรวบรวมและเก็บข้อมูลด้วยเหตุผลอื่นนอกเหนือจากการทำเหมือง และผู้ทำเหมืองข้อมูลต้องปรับแต่งก่อนที่จะใช้สร้างแบบจำลอง

4. การจำลองข้อมูล

ทำเหมืองข้อมูลป้อนข้อมูลที่เตรียมไว้ลงในซอฟต์แวร์การทำเหมืองข้อมูลและศึกษาผลลัพธ์ ในการทำเช่นนี้ พวกเขาสามารถเลือกเทคนิคและเครื่องมือการทำเหมืองข้อมูลได้หลากหลาย พวกเขายังต้องเขียนการทดสอบเพื่อประเมินคุณภาพของผลการทำเหมืองข้อมูลเพื่อจะจำลองข้อมูล

5. การประเมิน

หลังจากสร้างโมเดลแล้ว ผู้ทำเหมืองข้อมูลจะเริ่มวัดผลเทียบกับเป้าหมายธุรกิจเดิม พวกเขาแบ่งปันผลลัพธ์กับนักวิเคราะห์ธุรกิจและรวบรวมคำติชม โมเดลอาจตอบคำถามเดิมได้

ดีหรือแสดงรูปแบบใหม่ที่ไม่รู้จักมาก่อน ผู้ทำเหมืองข้อมูลสามารถเปลี่ยนรูปแบบ ปรับเป้าหมายธุรกิจ หรือ Revisit ข้อมูลได้ ขึ้นอยู่กับผลตอบรับของธุรกิจ การประเมิน คำติชม และการปรับเปลี่ยนอย่างต่อเนื่องเป็นส่วนหนึ่งของกระบวนการค้นพบความรู้

6.การติดตั้งใช้งาน

ในระหว่างการติดตั้งใช้งาน ผู้มีส่วนได้ส่วนเสียอื่นๆ ใช้โมเดลการทำงานเพื่อสร้างข่าวกรองธุรกิจ นักวิทยาศาสตร์ข้อมูลวางแผนกระบวนการติดตั้งใช้งาน ซึ่งรวมถึงการสอนผู้อื่นเกี่ยวกับฟังก์ชันของโมเดล การตรวจสอบอย่างต่อเนื่อง และการบำรุงรักษาแอปพลิเคชันการทำเหมืองข้อมูล นักวิเคราะห์ธุรกิจใช้แอปพลิเคชันนี้เพื่อสร้างรายงานสำหรับการจัดการแบ่งปันผลลัพธ์กับลูกค้า และปรับปรุงกระบวนการทางธุรกิจ

2.2.2 หลักในการทำเหมืองข้อมูล

จรัสสรี รุ่งรัตนาอุบล, สำนักพิมพ์มหาวิทยาลัยนเรศวร, 2566 ได้ให้ความหมายของการทำเหมืองข้อมูลคือ การจัดทำเหมืองข้อมูลจัดเป็นกระบวนการอัตโนมัติ เพื่อค้นพบข้อสนเทศหรือองค์ความรู้ รูปแบบ หรือแม้แต่ตัวแบบเพื่อการพยากรณ์จากฐานข้อมูลขนาดใหญ่ การทำเหมืองข้อมูลมีหลักการทำเหมืองข้อมูล 3 หลักการทำงานหลัก ๆ ดังนี้

1) การแบ่งกลุ่มข้อมูล (Clustering) เป็นการรวมข้อมูลที่มีลักษณะรูปแบบ และแนวโน้มที่เหมือนกันไว้ในกลุ่มเดียวกันโดยเริ่มต้นจากการาตัวแทนของกลุ่ม และทำการเปรียบเทียบข้อมูลตัวที่เหลือกับตัวแทนของแต่ละกลุ่ม ถ้าข้อมูลของตัวที่นำมาเปรียบเทียบ เหมือนคล้ายคลึง หรือสอดคล้องกับตัวแทนของกลุ่มไหนก็จะถูกจัดประเภทให้อยู่ในกลุ่มนั้น

2) การจำแนกประเภท (Classification) เป็นการจำแนกประเภทของข้อมูลโดยการเรียนรู้ข้อมูลเพื่อสร้างเป็นฐานความรู้ซึ่งแต่ละกลุ่มจะมีคุณสมบัติเฉพาะของตัวเองที่เรียกว่าตัวจำแนกข้อมูล (Classifier) จะมีหลักในการพัฒนาและอัลกอริทึมที่เกี่ยวข้องหลายตัวที่นิยมใช้ในปัจจุบัน โดยในที่นี้จะกล่าวถึงขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด k ตัว (Knearest Neighbor Algorithm) วิธีต้นไม้ตัดสินใจ (Decision Tree) การสร้างกฎ (Rule–based Classifier) วิธีเบย์อย่างง่าย (Naïve Bayes Classifier) และโครงข่ายประสาทเทียม (Artificial Neural Network) ซึ่งสามารถใช้ในการวิเคราะห์ และแยกแยะข้อมูลที่เข้ามาใหม่ได้

3) การพยากรณ์ (Prediction) เป็นการนำข้อมูลมาทำนายคำตอบแต่ค่าของการพยากรณ์หรือการทำนายจะเป็นค่าแบบต่อเนื่อง (Continuous Value) ซึ่งแตกต่างจากเทคนิคการจำแนกที่คำตอบของการทำนายจะเป็นค่าเต็มหน่วย (Discrete Value) หรือที่เรียกว่าคลาส (Class) ที่เป็นการสื่อถึงค่าคำตอบแบบเต็มหน่วย ขั้นตอนการพัฒนาตัวพยากรณ์จะมีความคล้ายคลึงกับการพัฒนาตัวจำแนก โดยจะมีการแบ่งข้อมูลเป็นข้อมูลฝึกสอนและข้อมูลทดสอบ

เหมือนกัน แต่สิ่งที่แตกต่างกันคือการวัดประสิทธิภาพของการพยากรณ์หรือความแม่นยำในการพยากรณ์ (Predicted Accuracy) ซึ่งจะใช้เกณฑ์การวัดค่าความแม่นยำอีกลักษณะหนึ่งที่ไม่ใช่การวัดร้อยละการจำแนกที่ถูกต้องและเมทริกซ์สับสนเหมือนเทคนิคการจำแนก โดยเกณฑ์การวัดประสิทธิภาพที่นิยมใช้กันเช่น รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) เป็นต้น

2.2.3 ทฤษฎีเกี่ยวข้องกับการสร้างเว็บไซต์

2.2.3.1 การสร้างเว็บไซต์สิ่งสำคัญอยู่ที่การออกแบบเว็บเพราะเว็บไซต์ที่มีรูปแบบสวยงามจะสามารถดึงดูดความสนใจจากผู้คนได้ดีกว่าทำให้ผู้คนเกิดความรู้สึกประทับใจอยากกลับมาใช้งานเว็บไซต์อีกครั้งในอนาคต ดังนั้นเริ่มแรกก่อนทำเว็บไซต์จึงจำเป็นต้องทำความเข้าใจกับหลักการออกแบบและรูปแบบโครงสร้างของเว็บก่อน การออกแบบเว็บไซต์เพื่อให้มีประสิทธิภาพและสามารถดึงดูดความสนใจของผู้คนได้ดีจะต้องมีองค์ประกอบของเว็บไซต์อย่างครบถ้วน ซึ่งได้แก่

1. ความเรียบง่าย เข้าใจง่าย การออกแบบเว็บไซต์ที่ดีจะต้องเน้นที่ความเรียบง่ายเป็นหลัก โดยเลือกนำเสนอเฉพาะสิ่งที่ต้องการนำเสนอจริงๆในรูปแบบที่หลากหลายโดยอาจจะเป็นสีสัน กราฟิก ภาพเคลื่อนไหวหรือตัวอักษร ที่สำคัญจะต้องมีการนำเสนอที่ไม่ดูรกหน้าเว็บจนเกินไปเพื่อไม่ให้เกิดความรู้สึกรกสายตาหรือสร้างความเบื่อหน่าย น่ารำคาญให้กับผู้ที่เข้าชมเว็บไซต์ มีตัวอย่างเว็บไซต์ที่มีการออกแบบโดยเน้นความเรียบง่ายได้ดี คือ Apple, Nokia และ Microsoft เป็นต้น

2. ความสม่ำเสมอ ไม่สับสน ควรออกแบบเว็บไซต์ด้วยความสม่ำเสมอคือจะต้องมีรูปแบบ กราฟิก โทนสีและการตกแต่งต่างๆให้แต่ละหน้าบนเว็บไซต์มีความคล้ายคลึงกันและเป็นแนวเดียวกันไปตลอดทั้งเว็บไซต์ ดังตัวอย่างเว็บไซต์ทั่วๆไปที่จะสังเกตเห็นได้ว่าทุกหน้าของเว็บไซต์นั้นจะเน้นการตกแต่งในรูปแบบเดียวกันทั้งหมดต่างก็แค่การนำเสนอของแต่ละหน้าเท่านั้น

3. สร้างความโดดเด่น เป็นเอกลักษณ์ การออกแบบเว็บไซต์เพื่อให้สามารถสื่อถึงจุดประสงค์ในการนำเสนอเว็บได้ดีจะต้องมีการสร้างความเป็นเอกลักษณ์และจุดเด่นให้กับเว็บไซต์ เพื่อให้สามารถสะท้อนถึงลักษณะขององค์กรได้มากที่สุด โดยการสร้างเอกลักษณ์ดังกล่าวนั้น อาจใช้ชุดสี รูปภาพ ตัวอักษรหรือกราฟิก นอกจากนี้ก็ต้องขึ้นอยู่กับว่าเป็นเว็บไซต์แบบบทางการหรือไม่เพื่อจะได้ออกแบบได้อย่างเหมาะสมที่สุด

4. เนื้อหาต้องดี ครบถ้วน เนื้อหาเป็นสิ่งที่สำคัญที่สุดของการสร้างเว็บไซต์ เพราะสิ่งที่ทำให้ผู้คนเกิดความสนใจและหมั่นติดตามเว็บไซต์เหล่านั้นอยู่เสมอ ก็คือเนื้อหาที่มีความสมบูรณ์และน่าสนใจ นอกจากนี้จะต้องมีการปรับปรุงพัฒนาเนื้อหาบนเว็บให้มีความทันสมัยอยู่เสมอรวมถึงข้อมูลต้องมีความถูกต้องที่สุด

5. ระบบเนวิเกชั่น ใช้ง่าย ระบบเนวิเกชั่นเป็นเสมือนป้ายบอกทางเพื่อให้ผู้ใช้งานไม่เกิดความสับสนในขณะใช้งานเว็บไซต์ ซึ่งการออกแบบเนวิเกชั่นก็จะต้องเน้นที่ความเรียบง่ายใช้งานสะดวก และมีความเข้าใจได้ง่าย ที่สำคัญจะต้องมีตำแหน่งการวางที่สม่ำเสมอเพื่อให้ดูเป็นแนวทางเดียวกันทำให้ผู้ใช้งานหรือผู้ชมรู้สึกประทับใจและจดจำเว็บไซต์ได้ง่ายขึ้น ส่วนใครที่มีการนำกราฟิกมาใช้ในระบบเนวิเกชั่นก็จะต้องเลือกกราฟิกที่สามารถสื่อความหมายได้ดีเช่นกัน

6. คุณภาพของเว็บไซต์ เว็บไซต์ที่ดีจะต้องมีคุณภาพ ทั้งสิ่งที่ปรากฏให้เห็นบนเว็บไซต์ ไม่ว่าจะเป็นกราฟิก ชนิดตัวอักษร รูปภาพหรือสีสันที่ใช้เนื้อหาที่นำมาแสดงผล ซึ่งหากเว็บไซต์มีคุณภาพก็จะสร้างความน่าเชื่อถือและเป็นจุดเด่นที่ทำให้ผู้คนส่วนใหญ่เกิดความสนใจได้ดี เพราะฉะนั้นห้ามละเลยในส่วนของคุณภาพเด็ดขาด

7. ความสะดวกในการเข้าใช้งาน เว็บไซต์ควรให้ความสะดวกสบายแก่ผู้ใช้งานได้ดี คือจะต้องมีการแสดงผลได้ในทุกระบบปฏิบัติการไม่ว่าจะเป็นเว็บเบราว์เซอร์ คอมพิวเตอร์โน๊ตบุ๊คหรือบนโทรศัพท์มือถือ ที่สำคัญจะต้องมีความละเอียดของการแสดผลและสามารถใช้งานได้โดยไม่มีปัญหาด้วย

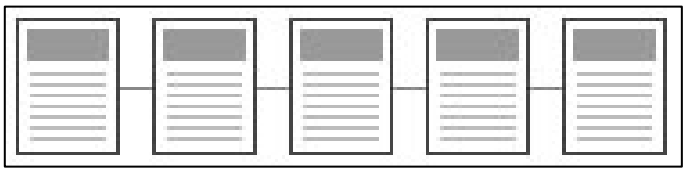
8. ความคงที่ของการออกแบบ การออกแบบเว็บไซต์ควรจะมีความคงที่ในการออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกันและมีการเรียบเรียงเนื้อหาอย่างรอบคอบ ทำให้เว็บมีความน่าเชื่อถือและดูมีคุณภาพช่วยสร้างความประทับใจให้กับผู้ใช้งานได้เป็นอย่างดี

9. ความคงที่ของการทำงาน ระบบการทำงานบนเว็บไซต์จะต้องมีความคงที่และสามารถใช้งานได้ดี ซึ่งนอกจากการออกแบบระบบการทำงานให้มีความทันสมัยและสร้างสรรค์แล้วก็จะต้องหมั่นตรวจสอบอยู่เสมอเพราะหากระบบการใช้งานมีความผิดปกติก็จะได้แก้ปัญหาได้ทัน นอกจากนี้อาจมีการอัพเดตดีไซน์ให้ทันสมัยขึ้นบ่อยๆ เพื่อให้ผู้ใช้งานรู้สึกสนุกไปกับการใช้งานเว็บไซต์

2.2.3.2 รูปแบบโครงสร้างของเว็บไซต์ การออกแบบโครงสร้างของเว็บไซต์สามารถทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอเพราะจะต้องออกแบบให้เหมาะกับการ

ใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1. โครงสร้างแบบเรียงลำดับ

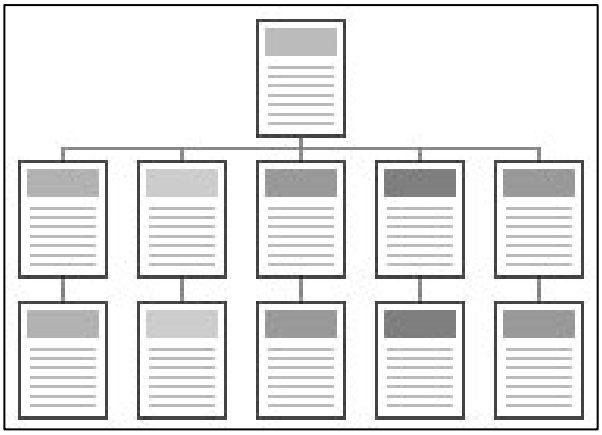


ภาพที่ 2.10 แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ

(ที่มา: https://www.1belief.com/article/website-design/)

โครงสร้างเว็บไซต์แบบเรียงลำดับจะเป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูลและสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็กมีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิ้งค์เนื้อหา ก็จะลิ้งค์ไปทีละหน้า มีทิศทางการเข้าสู่เนื้อหาต่างๆในแบบเส้นตรงใช้ปุ่มเดินหน้า-ถอยหลังในการกำหนดทิศทางจึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับก็มีข้อเสียคือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหาเพราะไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาด้วยตัวเองได้

2. โครงสร้างแบบลำดับขั้น



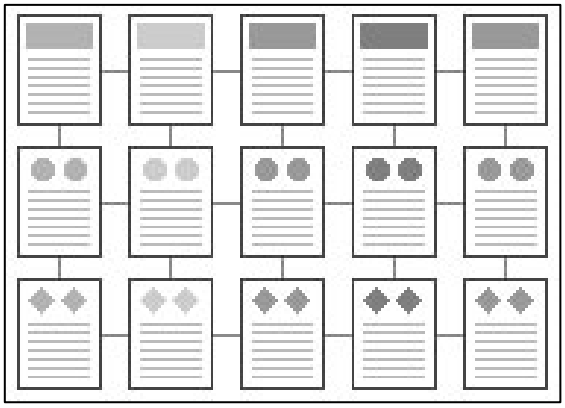
ภาพที่ 2.11 แสดงโครงสร้างแบบลำดับขั้น

(ที่มา: https://www.1belief.com/article/website-design/)

โครงสร้างแบบลำดับขั้นนิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่างๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆและมีการนำเสนอ

รายละเอียดย่อยๆที่ลดหลั่นกันมา ทำให้สามารถทำความข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีโฮมเพจเป็นจุดเริ่มต้นและจุดร่วมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง

3. โครงสร้างแบบตาราง

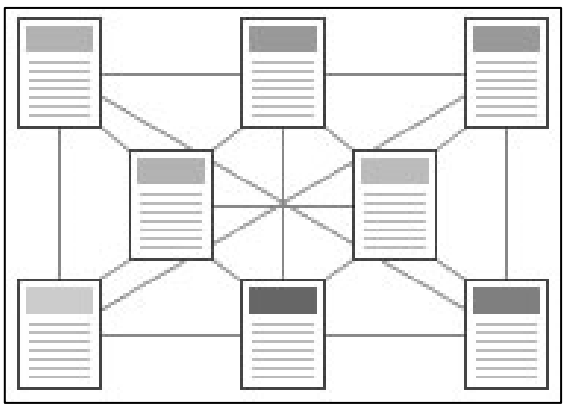


ภาพที่ 2.12 แสดงโครงสร้างแบบตาราง

(ที่มา: https://www.1belief.com/article/website-design/)

โครงสร้างแบบตาราง เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อนแต่ก็มีความยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่างๆได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งานสามารถเปลี่ยนทิศทางหรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลาแถมยังทำให้เว็บไซต์มีความทันสมัยขึ้น

4. โครงสร้างแบบใยแมงมุม



ภาพที่ 2.13 แสดงโครงสร้างแบบใยแมงมุม

(ที่มา: https://www.1belief.com/article/website-design/)

โครงสร้างแบบใยแมงมุม เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมากเพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถเข้าถึงหน้าเว็บเพจต่างๆ ที่ต้องการได้อย่างง่ายและมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี

2.2.3.3 ส่วนประกอบสำคัญบนหน้าเว็บเพจ บนหน้าเว็บเพจ จะมีส่วนประกอบสำคัญ ที่จำเป็นต้องมีอยู่ 3 ส่วน

1) ส่วนหัวของหน้า (Header) อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซต์ต่อไป ซึ่งส่วนใหญ่ก็มักจะมีการใส่ภาพกราฟิกให้ดูสวยงาม สิ่งสำคัญหลักๆเลยก็คือ โลโก้ ชื่อเว็บไซต์และเมนู หลักที่สามารถลิงค์ไปยังเนื้อหาในหน้าเว็บเพจต่างๆได้

2) ส่วนของเนื้อหา (Body) อยู่บริเวณตอนกลางของหน้าเว็บ โดยจะแสดงข้อมูลเกี่ยวกับเนื้อหาบนเว็บแบบคร่าวๆ ซึ่งก็จะมีข้อความกราฟิก ตารางข้อมูลหรือวิดีโอ ประกอบอยู่และหากมีเมนูแบบเฉพาะกลุ่มก็จะถูกจัดไว้ในหน้านี้เช่นกันและที่สำคัญเนื้อหาในส่วนนี้ควรจะมีความกระชับ เข้าใจง่าย มีการใช้รูปแบบตัวอักษรแบบเรียบง่ายและเป็นระเบียบ

3) ส่วนท้ายของหน้า (Footer) อยู่ล่างสุดของหน้าเว็บซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่างๆเพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูลเจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่างๆเกี่ยวกับการใช้งานเว็บไซต์อย่างถูกต้อง

2.2.3.4 วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์การเลือกใช้สีในการออกแบบ เว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้น การรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไรต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่ม

1) สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอบโยนและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2) สีโทนเย็น (Cool Colors) เป็นสีแห่งความสุภาพและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้นและยังสามารถใช้โน้มน้าวจากในระยะไกลได้อีกด้วย

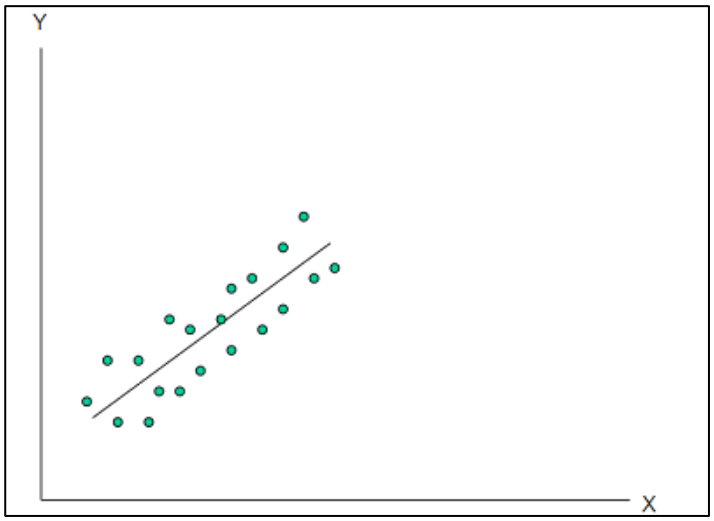
3) สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.2 เทคนิคการถดถอยเชิงเส้น (linear regression)

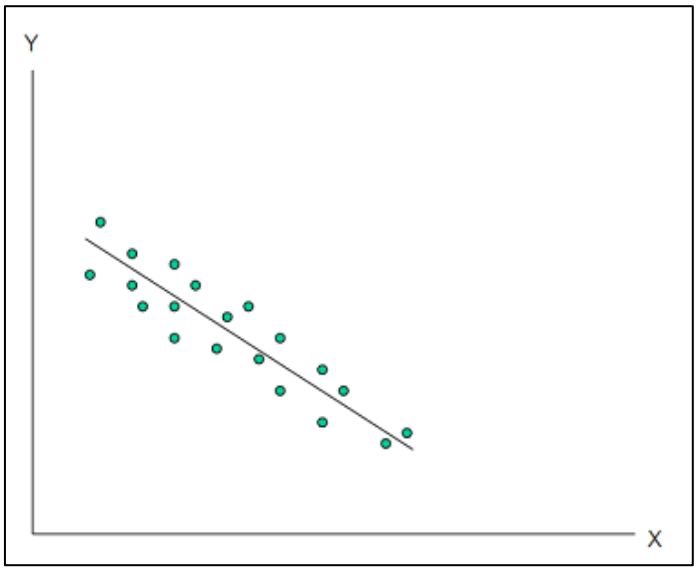
การวิเคราะห์การถดถอยเชิงเส้น เป็นการคำนวณหาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร คือ ตัวแปรที่เราทราบค่า (Predictor : x) และตัวแปรที่เราไม่ทราบค่า (Response : y) ซึ่งเป็นความสัมพันธ์แบบเชิงเส้น (Linear) โดยการคำนวณจากค่า x และ y ที่มีความสัมพันธ์กันจะใช้สมการของ Linear Regression คือ y=ax+b บางท่านอาจจะเรียกตัวแปร x , y ว่า x= ตัวแปรอิสระ ,y= ตัวแปรตามรูปแบบความชันและระยะตัดแกน x=ตัวแปรที่ทราบค่า I ตัวประมาณการ(Predictor) y=ตัวแปรที่เราไม่ทราบค่า I ตัวตอบสนอง (Response)

การวิเคราะห์การถดถอยเชิงเส้นต้องมีการเก็บจำนวน Sample space จำนวนมากพอหรือเขียนโปรแกรมกำหนดค่าให้ได้กลุ่มตัวอย่างของข้อมูลให้ได้เยอะที่สุดนั้นคือ มี x และ y ที่มีความสัมพันธ์กันหลายๆ ครั้งเพื่อนำมาหาสมการความสัมพันธ์ (สามารถ LAB Python Linear Regression ประกอบความเข้าใจได้) ประเภทของความสัมพันธ์แบ่งได้ดังนี้



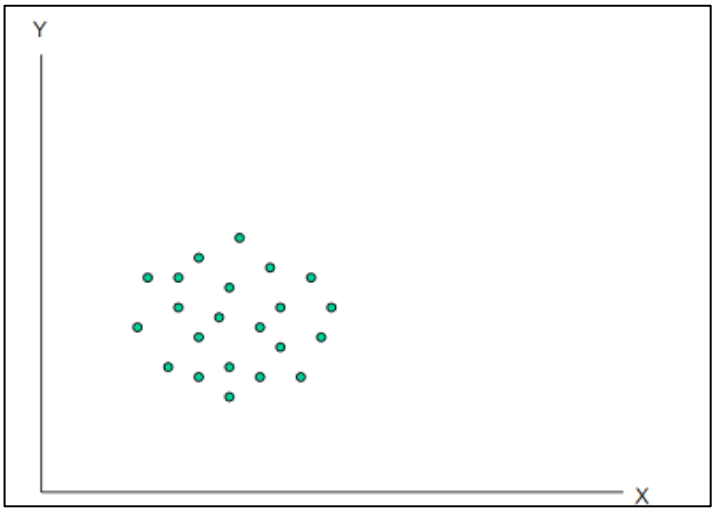
ภาพที่ 2.14 ตัวอย่างแผนภาพการกระจาย (Scatter Plot) ความสัมพันธ์เส้นตรงทางบวก (ที่มา: https://kongruksiam.medium.com/สรุป-machine-learning-ep-3-การวิเคราะห์การถดถอยเชิงเส้น-linear-regression-891260e4a957)

หมายถึงค่าของตัวแปร x , y มีทิศทางเดียวกันส่งผลให้เมื่อค่า x เพิ่มขึ้น ค่า y ก็เพิ่มตามไปด้วยหรือเรียกว่า การแปรผันตรง



ภาพที่ 2.15 ตัวอย่างแผนภาพการกระจาย (Scatter Plot) ความสัมพันธ์เส้นตรงทางลบ (ที่มา: https://kongruksiam.medium.com/สรุป-machine-learning-ep-3-การวิเคราะห์การถดถอยเชิงเส้น-linear-regression-891260e4a957)

หมายถึงค่าของตัวแปร x , y มีทิศทางตรงข้ามกันส่งผลให้เมื่อค่า x ลดลง ค่า y เพิ่มขึ้นหรือค่า x เพิ่มขึ้นแต่ค่า y ลดลงเรียกว่า การแปรผกผัน



ภาพที่ 2.16 ตัวอย่างแผนภาพการกระจาย (Scatter Plot) ไม่มีความสัมพันธ์ของข้อมูล (ที่มา: https://kongruksiam.medium.com/สรุป-machine-learning-ep-3-การวิเคราะห์การถดถอยเชิงเส้น-linear-regression-891260e4a957)

Coefficient ค่าสัมประสิทธิ์แสดงการตัดสินใจ คือ ตัวเลขที่บอกความสัมพันธ์ของสองตัวแปร หรือ ค่าที่แสดงว่าตัวแปร x มีอิทธิพลต่อตัวแปร y มากน้อยเพียงใดโดยมี ขอบเขตในช่วง –1 ถึง 1 Intercept คือ ค่าที่บ่งบอกจุดตัดแกน R-Square คือ ค่าความผันแปรของตัวแปร y มีค่าอยู่ระหว่าง 0% — 100%

– 0% หมายถึง ผลลัพธ์ที่ได้มานั้นไม่สามารถอธิบายความผันแปรของค่าตัวแปร y ต่างที่กระจายรอบค่าเฉลี่ยได้เลย

– 100% แสดงให้เห็นว่าผลลัพธ์ที่ได้มานั้นสามารถอธิบายความผันแปรของค่าตัวแปร y ต่างที่กระจายรอบค่าเฉลี่ยได้เป็นอย่างดี

การวัดประสิทธิภาพ

y_pred คือผลการทำนายผลลัพธ์จากโมเดล

y_test คือ ข้อมูลจริงที่ใช้ทดสอบการทำนายผลโมเดล

Loss Function คือ การคำนวน Error ว่า y_pred ที่โมเดลทำนายออกมา ต่างจาก y_test อยู่เท่าไร แล้วหาค่าเฉลี่ย เพื่อที่จะนำมาหา Gradient ของ Loss แล้วใช้อัลกอริทึม Gradient Descent เพื่อให้ Loss น้อยลงในการเทรนรอบถัดไป (Loss ยิ่งค่าน้อยยิ่งดี)

Loss Function ที่นิยมใช้ในงาน Regression ในปัจจุบัน มีหลายตัว เช่น Mean Absolute Error (MAE) หรือ L1 Loss , Mean Squared Error (MSE) หรือ L2 Loss , Root Mean Squared Error (RMSE)

1. Mean Absolute Error (MAE) การคำนวน Error ว่า y_pred ต่างจาก y_test อยู่เท่าไร ด้วยการนำมาลบกันตรง ๆ แล้วหาค่าเฉลี่ย โดยไม่สนใจเครื่องหมาย (Absolute) เพื่อหาขนาดของ Error โดยไม่สนใจทิศทาง

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - y_j|$$

2. Mean Squared Error (MSE) การคำนวน Error ว่า y_pred ต่างจาก y_test อยู่เท่าไร ด้วยการนำมาลบกัน แล้วยกกำลังสอง (Squared) เพื่อไม่ต้องสนใจค่าติดลบ (ถ้ามี) แล้วหาค่าเฉลี่ย

$$MSE = \frac{1}{N}\sum_{i}^{n}(Y_i - y_i)^2$$

3. Root Mean Squared Error (RMSE) คือ นำ MSE มาหา Squared Root

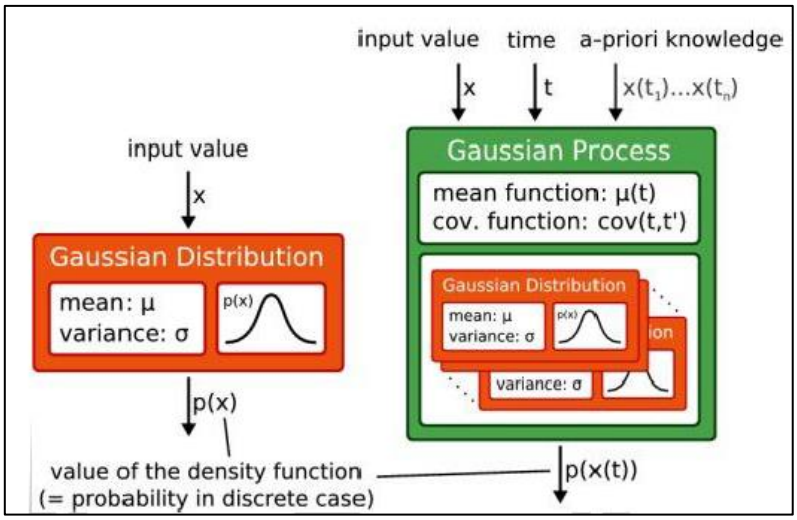
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

ค่า MSE, MAE อยู่ในช่วง 0 — Infinity เหมือนกัน ยิ่งน้อยคือยิ่งดีถ้าเป็น 0 คือ ไม่มีError เลย ดังนั้นถ้าค่าเท่ากับ 0 แปลว่าโมเดลทำนายค่า y_test ได้ถูกต้อง100% แต่ในทางปฏิบัติโอกาสที่จะเทรนโมเดลได้ loss = 0 เป็นไปได้ยากมาก

อาศัยส่วนของ scikit-learn เข้ามาทำงานร่วมด้วยเพื่อวัดผลจากค่าที่คาดการหรือพยากรณ์ได้จากการ predict ตัวโมเดล (y_pred) และค่าจริงจาก y_test

2.3.3 เทคนิคเกาส์เซียน (Gussian Process)

เป็นเทคนิคที่นำเอาระบบ Stochastic Process ซึ่ง ปัจจุบัน ถูกยอมรับแล้วว่าเป็นเครื่องมือที่มีประสิทธิภาพในการแก้ปัญหาประเภท Regression, Classification, และ Decision ใน Machine learning สามารถทำงานได้ดีถึงแม้ว่ามี Training Data น้อยและมีประสิทธิภาพและ Convergence rate ดีกว่า ARMA, NN และ SVR ตามลำดับ ซึ่งมีงานวิจัยของทัศนัย พลอยสุวรรณ และคณะ นำเสนอการพยากรณ์หาค่าความต้องการไฟฟ้า สูงสุด (Peak load) ระยะยาวระหว่างปี 2011-2012 โดยใช้ GaussianProcess ซึ่งผลการพยากรณ์ที่ดีกว่าเมื่อเปรียบเทียบกับกับการใช้โครงข่ายประสาทเทียม (Neural Networks)



ภาพที่ 2.17 หลักการทำงานของ Guassian Process

ในทฤษฎีความน่าจะเป็นและสถิติกระบวนการเกาส์เซียนเป็นกระบวนการสุ่ม (กลุ่มตัวแปรสุ่มที่จัดทำดัชนีตามเวลาหรือพื้นที่) โดยที่กลุ่มตัวแปรสุ่มจำกัดทุกกลุ่มมีการแจกแจงแบบปกติหลายตัวแปรการแจกแจงของกระบวนการเกาส์เซียนคือการแจกแจงร่วมของตัวแปรสุ่มทั้งหมด (จำนวนอนันต์) และด้วยเหตุนี้ จึงเป็นการแจกแจงของฟังก์ชันที่มีโดเมนต่อเนื่อง เช่น เวลาหรือพื้นที่

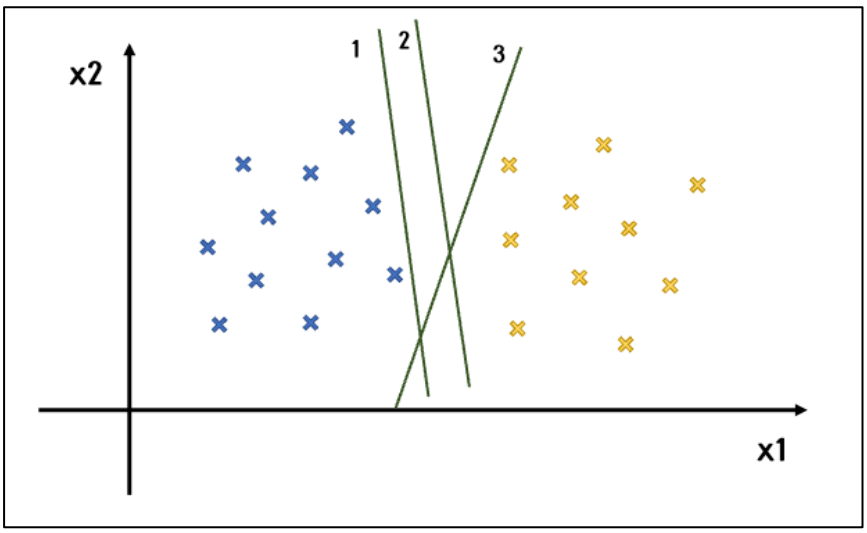
แนวคิดของกระบวนการแบบเกาส์เซียนได้รับการตั้งชื่อตามคาร์ล ฟรีดริช เกาส์ เนื่องจากแนวคิดนี้มีพื้นฐานมาจากแนวคิดของการแจกแจงแบบเกาส์เซียน (การแจกแจงแบบปกติ) กระบวนการแบบเกาส์เซียนสามารถมองได้ว่าเป็นการสรุปแบบมิติอนันต์ของการแจกแจงแบบปกติหลายตัวแปร

กระบวนการแบบเกาส์เซียนมีประโยชน์ในการสร้างแบบจำลองทางสถิติโดยได้รับประโยชน์จากคุณสมบัติที่สืบทอดมาจากการแจกแจงแบบปกติ ตัวอย่างเช่น หากจำลอง กระบวนการสุ่มเป็นกระบวนการแบบเกาส์เซียน การแจกแจงของปริมาณที่ได้มาต่างๆ ก็สามารถหาได้อย่างชัดเจน ปริมาณดังกล่าวรวมถึงค่าเฉลี่ยของกระบวนการในช่วงเวลาหนึ่งและข้อผิดพลาดในการประมาณค่าเฉลี่ยโดยใช้ค่าตัวอย่างในช่วงเวลาสั้นๆ แม้ว่าแบบจำลองที่แน่นอนมักจะปรับขนาดได้ไม่ดีเมื่อปริมาณข้อมูลเพิ่มขึ้น แต่ ก็มีการพัฒนา วิธีการประมาณ หลายวิธี ซึ่งมักจะรักษาความแม่นยำได้ดีในขณะที่ลดเวลาในการคำนวณลงอย่างมาก

2.3.4 เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression)

SVM หรือ Support Vector Machine ถือเป็นวิธีคลาสสิกที่น่าเรียนรู้มากทีเดียว เพราะไอเดียจากโมเดลนี้ก็เป็นหนึ่งในรากฐานสำคัญที่ทำให้เข้าใจโมเดลใหญ่ๆในปัจจุบันมากขึ้น อีกทั้งหนึ่งในผู้ที่คิดค้นวิธี Vladimir Naumovich Vapnik ซึ่งเป็นชาวรัสเซียและยังเป็นนักวิทยาศาสตร์ที่ยิ่งใหญ่ในวงการ Machine Learning อีกด้วย

Support Vector Machine ใช้มากในการแยกประเภท หรือการทำ classification ใช้การคำนวณไม่มากเมื่อเทียบกับ deep learning และยังเป็นวิธีที่ถือว่า robust ต่อข้อมูลที่มี noise พอสมควร



ภาพที่ 2.18 ตัวอย่างหาฟังก์ชั่นเส้นแบ่งข้อมูลออกจากกัน

(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765)

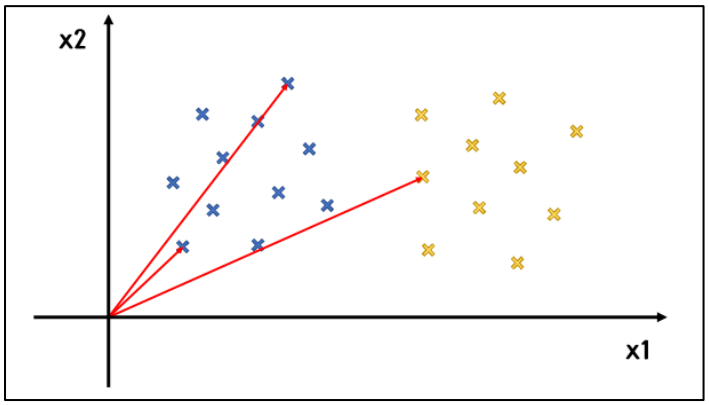
จากรูป สมมติว่าเรามีข้อมูล input สองกลุ่ม สีเหลืองและสีฟ้า มี 2 features ซึ่งก็คือ x1 และ x2 และเราต้องการจะหาฟังก์ชั่นเส้นแบ่งข้อมูลออกจากกัน เส้นไหนจะเป็นเส้นที่ดีที่สุด? ระหว่าง 1, 2 และ 3

เส้นที่ 1 อยู่ชิดกับข้อมูลกลุ่มสีฟ้ามากเกินไป

เส้นที่ 2 อยู่ตรงกลางพอดี (เส้นนี้ดีที่สุด)

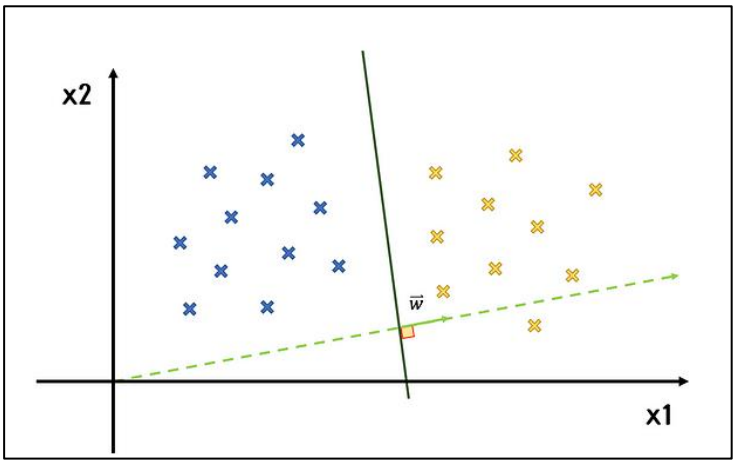
เส้นที่ 3 อยู่ติดกับทั้งข้อ

จะเห็นว่า ทั้งสามเส้นสามารถแยกข้อมูลได้อย่างถูกต้องทั้งหมด แต่เส้นที่ 2 จะดีที่สุด เพราะตำแหน่งและทิศทางของเส้นอยู่แนวตรงกลางระหว่างข้อมูลสองกลุ่มพอดี ฉะนั้น ไอเดียของ SVM จึงเป็นการมองข้อมูลในรูปแบบ Geometry หรือแบบเรขาคณิต ซึ่งมองว่าข้อมูลของเรามีทั้งขนาดและทิศทางดังภาพ เพื่อที่จะได้คำนวณเส้นแบ่งที่อยู่แนวกลางได้



ภาพที่ 2.19 แสดงตำแหน่งและทิศทางของเส้น

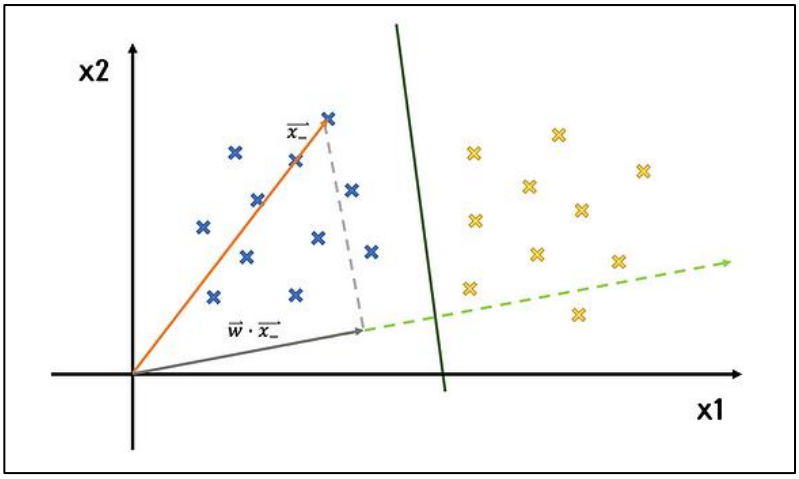
(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765



ภาพที่ 2.20 แสดงการกำหนดเส้นแบ่ง

(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765)

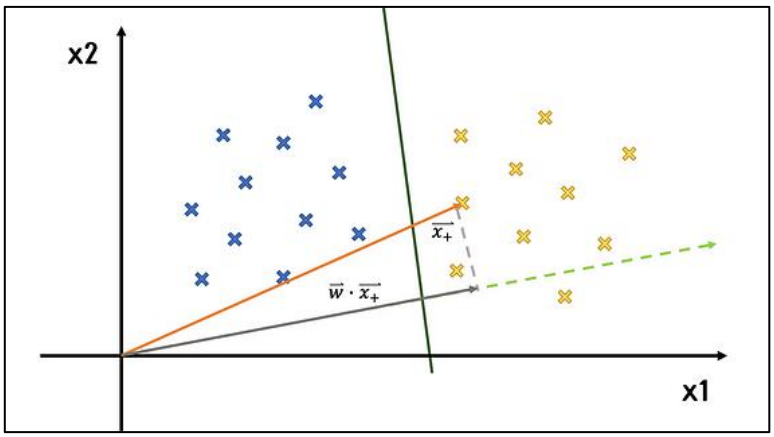
กำหนดว่าเส้นแบ่งของเราสามารถเขียนแทนได้ด้วยเวกเตอร์ขนาดหนึ่งหน่วย และตั้งฉากกับเส้นแบ่ง มีชื่อว่าเวกเตอร์ w ดังภาพ



ภาพที่ 2.21 ตัวอย่างการนำเอาข้อมูลมา dot กับเวกเตอร์ w

(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765)

การนำเอาข้อมูลมา dot กับเวกเตอร์ w จะเปรียบเสมือนว่าเรากำลังหาขนาดที่เป็นทิศทางเดียวกับเวกเตอร์ w ที่อยู่ในเวกเตอร์ของข้อมูล ซึ่งก็คือเวกเตอร์สีเทาในภา



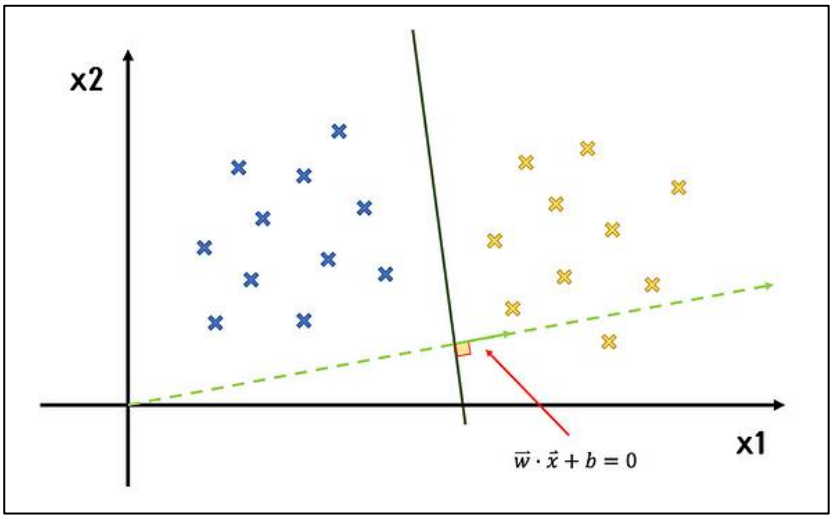
ภาพที่ 2.22 ตัวอย่างการนำเอาข้อมูลของฝั่งสีฟ้า (x-) มา dot กับเวกเตอร์ w

(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765)

จากภาพเป็นตัวอย่างการนำเอาข้อมูลของฝั่งสีฟ้า (x-) มา dot กับเวกเตอร์ w และหากนำเอาข้อมูลของฝั่งสีเหลือง (x+) มา dot กับเวกเตอร์ w ก็จะได้ผลลัพธ์เวกเตอร์สีเทาดังภาพข้างล่าง

จากทั้งสองกรณีเราจะสามารถสังเกตได้ว่า ผลการคำนวณ dot product ของข้อมูลทั้งสองกลุ่มจะไม่เท่ากัน ข้อมูลที่อยู่ฝั่งซ้ายจะได้ค่า dot น้อยกว่าข้อมูลที่อยู่ฝั่งขวาของเส้นแบ่ง

เสมอ เราจึงนำเอากฎเกณฑ์นี้มาสร้างสมการของโมเดล เพื่อทำนายว่าเป็นประเภทสีเหลืองได้ดังนี้



ภาพที่ 2.23 ตัวอย่างการนำเอาข้อมูลอยู่บนแนวเส้นแบ่งพอดี

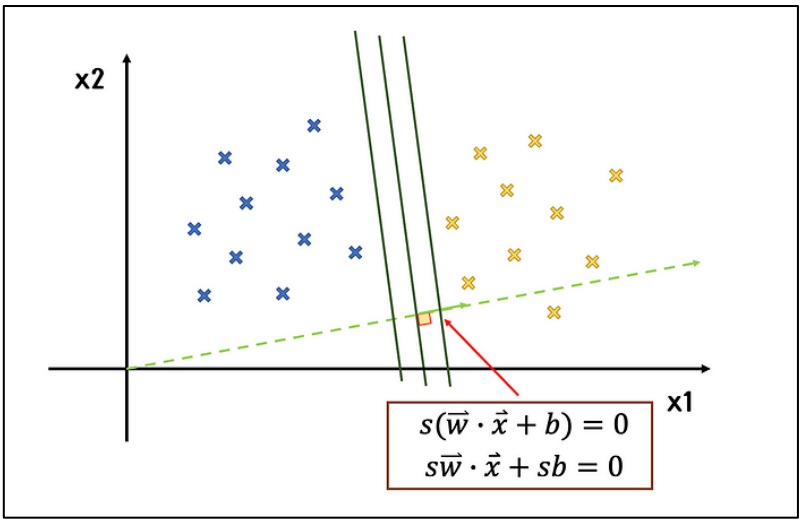
(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765)

หากข้อมูลอยู่บนแนวเส้นแบ่งพอดิบพอดี ก็จะได้สมการของเวกเตอร์ w dot กับข้อมูล x บวกด้วยค่าคงที่ b (ค่าคงที่ตัวนี้จะทำหน้าที่ย้ายเวกเตอร์ w จากจุด origin ไปยังตำแหน่งอื่นในทิศแนวเดียวกับ w ที่ชี้ออกจากจุด origin ค่ะ) จะเท่ากับ 0 พอดี ฉะนั้นหากเป็นฝั่งสีเหลือง (x+) ก็จะได้ค่ามากกว่า 0 และหากข้อมูลอยู่ฝั่งสีฟ้า(x−) ก็จะได้ค่าน้อยกว่า 0 ดังนี้

$$\vec{w} \cdot \overrightarrow{x_+} + b \geq 0$$

$$\vec{w} \cdot \overrightarrow{x_-} + b \leq 0$$

แต่ค่า w ที่เป็นคำตอบนั้น มีได้หลายค่า เพราะสามารถ scale ค่า w ด้วยค่าคงที่ s ค่าใดก็ได้ ก็ยังจะทำให้สมการเป็นจริง ดังรูป



ภาพที่ 2.24 ตัวอย่างสมการที่ได้

(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765)

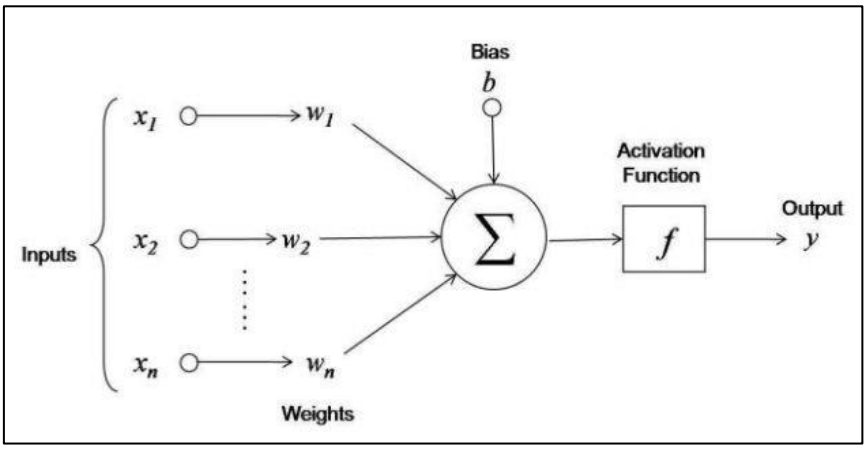
และ

$$y(\vec{w} \cdot \vec{x} + b) \geq 0$$

$$ys(\vec{w} \cdot \vec{x} + b) \geq 0$$

2.3.5 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks)

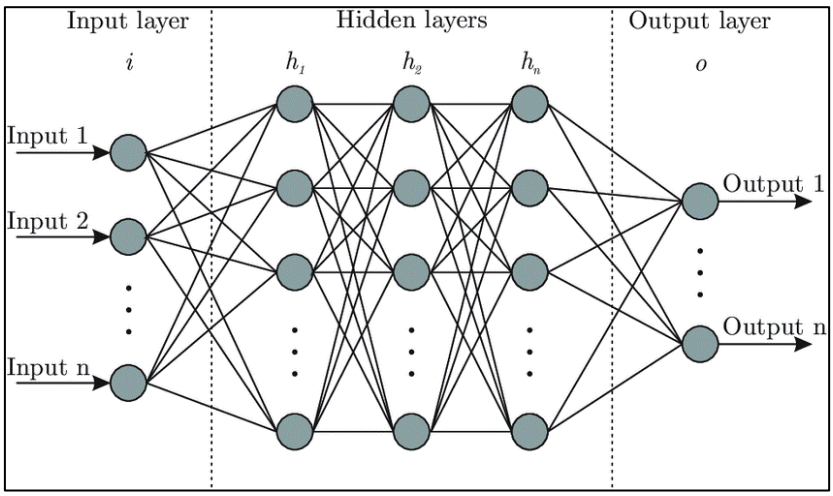
โครงข่ายประสาทเทียมถูกสร้างขึ้นเพื่อการจำลองลักษณะการประมวลผลของระบบประสาทมนุษย์ด้วยแบบจำลองเชิงคณิตศาสตร์และสถิติ (Mathematical and Statistical Model) ซึ่งประกอบด้วยส่วนของการประมวลผลที่เรียกว่า นิวรอน (Neuron) ทุกๆ นิวรอนสามารถมีข้อมูลป้อนเข้า (Input) ได้หลายค่า แต่ข้อมูลส่งออก (Output) มีได้เพียงค่าเดียว และทุกๆข้อมูลส่งออก (Output) จะเชื่อมโยงไปยังข้อมูลป้อนเข้าของนิวรอน (Input) อื่นๆภายในโครงข่าย สำหรับการเชื่อมโยงกันภายในระหว่างนิวรอนทุกๆข้อมูลป้อนเข้า (Input) จะมีค่าน้ำหนัก (Bias) เป็นตัวกำหนดกำลังของการเชื่อมโยงภายในนิวรอนจะมีฟังก์ชันกำหนดสัญญาณส่งออกที่เรียกว่า ฟังก์ชันถ่ายโอน (Transfer Function)



ภาพที่ 2.25 Artificial Neural Networks

(ที่มา: https://weather4thai.kmitl.ac.th/ANN.php)

โครงข่ายประสาทเทียมประกอบด้วยนิวรอนจำนวนมากเชื่อมต่อกัน ซึ่งการเชื่อมต่อแบ่งออกเป็นกลุ่มย่อย เรียกว่า ชั้น (Layer) ชั้นแรกเป็นชั้นข้อมูลป้อนเข้า (Input Layer) ชั้นสุดท้ายเป็นชั้นข้อมูลส่งออก (Output Layer) ส่วนชั้นที่อยู่ ระหว่างชั้นข้อมูลป้อนเข้าและชั้นข้อมูลส่งออก เรียกว่าชั้นซ่อน (Hidden Layer) ซึ่งโดยทั่วไปชั้นซ่อนอาจมีมากกว่า 1ชั้นก็ได้ ด้วยเหตุนี้จึงสามารถแบ่งประเภทตามโครงสร้างของโครงข่ายประสาทเทียมได้2แบบ คือ โครงข่ายประสาทเทียมแบบ ชั้นเดียว (Single Layer) และโครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer)



ภาพที่ 2.26 Multilayer

(ที่มา: https://weather4thai.kmitl.ac.th/ANN.php)

โดยทั่วไปการทำงานของโครงข่ายประสาทเทียมก็คือ การสอนให้โครงข่ายทำการคำนวนข้อมูลส่งออก (Output) พร้อมกับการปรับปรุงค่าน้ำหนัก (Bias) โดยอาศัยกระบวนการทำซ้ำ (Iterative) แบ่งออกเป็น 3 ประเภท คือ

1.การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือการสอนโครงข่ายโดยใช้ข้อมูลป้อนเข้า (Input)และข้อมูลส่งออก(Output)เป็นชุดฝึกสอนควบคู่(Training pair) โดยการสอนโครงข่ายนั้นจะใช้ชุดฝึดสอนหลายคู่จึงทำให้ข้อมูลส่งออกจริงกับข้อมูลส่งออก(Output)มีความคลาดเคลื่อนกัน โดยโครงข่ายจะต้องมีการปรับค่าน้ำหนัก(Bias)เพื่อลดค่าความแตกต่าง(Error)ระหว่างข้อมูลส่งออกจริงกับข้อมูลส่ง(Output)

2.การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) คือการสอนโครงข่ายโดยใช้ข้อมูลป้อนเข้า (Input)และหลักการทางสถิติหาค่าทางสถิติของชุดฝึกสอน ทำการจัดกลุ่มข้อมูลออกเป็นระดับต่างๆโดยโครงข่ายประสาทเทียมจะหาค่าข้อมูลออก(Output)

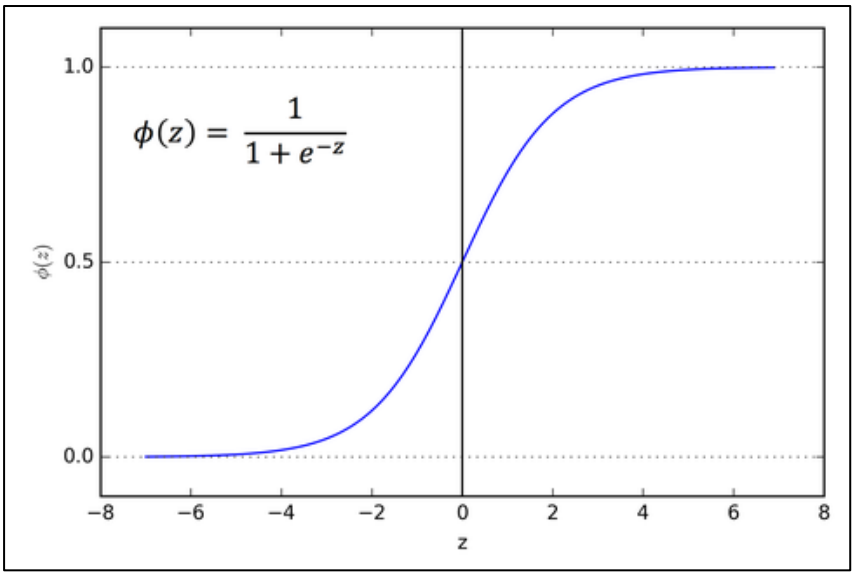
3.การเรียนรู้เชิงบังคับ (Reinforcement Learning)

การเรียนรู้เชิงบังคับ (Reinforcement Learning) คือการเรียนรู้แบบมีผู้สอนและไม่มีผู้สอนโดยจะใช้การเรียนรู้แบบไม่มีผู้สอนสำหรับข้อมูลป้อนเข้า (Input) และจะใช้การเรียนรู้แบบมีผู้สอนเมื่อได้ข้อมูลส่งออก(Output)แล้ว

สำหรับฟังก์ชันถ่ายโอน (Transfer Function) หรือฟังก์ชันการกระตุ้น (Activation Function)ถูกแบ่งเป็น 4 ประเภทได้แก่ (1) ฟังก์ชันเชิงเส้น (Linear function) (2) ฟังก์ชันไม่เชิงเส้น (Non-linear function) (3) ฟังก์ชันสมมาตร (Symmetrical function) และ (4) ฟังก์ชันไม่สมมาตร (Non-symmetrical function)

$$f(x) = \frac{1}{1+e^{-x}}$$

ฟังก์ชัน Sigmoid เป็นฟังก์ชันการกระตุ้น (Activation Function) แบบฟังก์ชันไม่เชิงเส้น (Non-linear function) ซึ่งจะมีฟังก์ชันอยู่ในรูปแบบ ข้อมูลส่งออกที่ได้จะเป็นเส้นโค้งรูปตัวเอสมีค่าอยู่ในช่วงระหว่าง 0 ถึง 1



ภาพที่ 2.27 ฟังก์ชันซิกมอยด์

(ที่มา: https://weather4thai.kmitl.ac.th/ANN.php)

โครงข่ายประสาทเทียม (ANN) ถูกแบ่งตามการเชื่อมต่อของปมประสาทและฟังก์ชันกระตุ้นแบ่งออกเป็น 5 ประเภท

1. ประเภทการคาดเดา (Prediction) คือการใช้ข้อมูลนำเข้า (Input) เพื่อเดาข้อมูลส่งออก(Output) เช่น Back-propagation, Delta Bar Delta, Extended Delta Bar Delta, Directed Random Search, Higher Order Neural Networks และ Self-organizing map into Back-propagation

2. ประเภทการจัดหมวดหมู่ (Classification) คือการใช้ข้อมูลนำเข้าเพื่อกำหนดการจัดหมวดหมู่ เช่น Learning Vector Quantization, Counter-propagation และ Probabilistic Neural Networks

3. ประเภทการเชื่อมโยงข้อมูล (Data Association) คือการใช้ข้อมูลนำเข้าเพื่อกำหนดการจัดหมวดหมู่แต่จะจดจำข้อมูลที่มีค่า Error เช่น Hopfield, Boltzamann Machine, Hamming Network และ Bidirectional associative Memory

4. ประเภทกระบวนการสร้างความคิด (Data Conceptualization) คือการวิเคราะห์ข้อมูลนำเข้า(Input)เพื่อจัดกลุ่ม เช่น Adaptive Resonance Network และ Self Organization Map

5. ประเภทการกลั่นกรองข้อมูล (Data Filtering) คือการทำให้ข้อมูลนำเข้ามีความสม่ำเสมอเช่น Recirculation

2.3.6 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM) หรือ Cross Industry Standard Process for Data Mining หมายถึง กระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูลเพื่อการวิเคราะห์และนำไปใช้ประโยชน์ในทางธุรกิจ พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ DaimlerChrysler, SPSS และ NCR ในกระบวนการ CRISP-DM ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่

1) การทำความเข้าใจในธุรกิจ (Business understanding) เป็นการกำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์เพื่อหาความได้เปรียบทางการตลาด เพื่อนำมาแก้ไขปัญหาองค์กรและต้องสามารถระบุผลลัพธ์ที่มีได้

2) การทำความเข้าใจในข้อมูล (Data Understander) โดยการรวบรวมข้อมูลที่เกี่ยวข้องให้อยู่รวมกลุ่มกัน คัดเลือกให้เหลือเพียงข้อมูลที่มีความสำคัญต่อการทำงาน

3) การเตรียมข้อมูล (Data Preparation) การจัดเตรียมข้อมูลประกอบด้วยการคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4) การสร้างแบบจำลอง (Modeling) นำผลที่ได้มาทดลองทำแบบจำลอง เพื่อใช้ถ่ายทอดข้อมูลหรือทำการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปสารสนเทศเป็นข้อมูลที่นำไปใช้ประโยชน์ได้จริงในทางธุรกิจ

5) การประเมิลผล (Evaluation) การวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ตั้งไว้ในตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใดเพื่อเปลี่ยนแปลงแก้ไขให้ได้ตามผลลัพธ์ที่ต้องการ

6) เผยแพร่ผลวิเคราะห์ (Deployment) การนำเอาข้อมูลที่เป็นผลลัพธ์มาลองปฏิบัติจริงกับธุรกิจ โดยแปลงแนวคิดที่มีให้เกิดเป็นการนำไปใช้ประโยชน์ และติดตามรวบรวมผลที่ได้เพื่อการปรับปรุง

2.4 วรรณกรรมที่เกี่ยวข้อง

ปทิตญา บุญรักษา (2561) ได้ศึกษาการพยากรณ์การเกิดอุบัติเหตุบนท้องถนนโดยใช้เทคนิคอนุกรมเวลา งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อสร้างแบบจำลองที่มีประสิทธิภาพในการพยากรณ์โดยใช้ เทคนิคอนุกรมเวลา คือ Linear regression (LR), Artificial Neural Networks: ANN (ANN), Sequential Minimal Optimization for Regression (SMOreg), Support Vector Machine for Regression (SVMR) และ Gussian Process (GP) และสามารถแสดงผลการพยากรณ์ผ่านทาง เว็บไซต์โดยใช้ข้อมูลจากงานอุบัติเหตุจราจรรายเดือนของ สถานีตำรวจอำเภอเมืองขอนแก่น ตั้งแต่ เดือนมกราคม 2554 ถึงเดือนมิถุนายน 2560 จำนวน 78 เดือน

การวัดประสิทธิภาพแบบจำลอง ได้ ทดลองแบ่งชุดข้อมูลออกเป็นชุดข้อมูลเรียนรู้และ ชุดข้อมูลทดสอบ ทั้งหมดจำนวน 7 รอบ โดยการ ใช้หลักการ Sliding window และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่า mean absolute error (MAE) และ root mean square error (RMSE) เพื่อวัดประสิทธิภาพของแบบจำลองนั้น จากการศึกษาพบว่า เทคนิค SVMR สามารถพยากรณ์การเกิดอุบัติเหตุบนท้องถนนด้วย เทคนิคอนุกรมเวลา ได้ค่าความผิดพลาดต่ำที่สุด จากยานพาหนะที่ 1.42 จากเพศที่ 8.99 และจาก การสูญเสีย 3.19 เมื่อเปรียบเทียบกับแบบจำลองที่สร้างด้วยเทคนิค LR,ANN, SMOreg และ Gussian Process เทคนิค SVMR มีความเหมาะสมมากที่สุดในการพยากรณ์การเกิดอุบัติเหตุบน ท้องถนน

ภูมินทร์ สุขโข (2566) ได้ศึกษาการวิเคราะห์ปัจจัยที่มีผลต่อความรุนแรงของอุบัติเหตุรถจักรยานยนต์บนทางหลวงในประเทศไทย วัตถุประสงค์ของการศึกษานี้คือการสำรวจปัจจัยที่มีอิทธิพลต่อความรุนแรงของอุบัติเหตุรถจักรยานยนต์บนทางหลวงในประเทศไทย การวิเคราะห์ถดถอยแบบลอจิสติกส์พหุนาม (MLR) ถูกนำมาใช้ในการวิเคราะห์บันทึกอุบัติเหตุรถจักรยานยนต์ตั้งแต่ปี 2018 ถึง 2020 เพื่อพิจารณาปัจจัยที่ส่งผลต่อความรุนแรงของอุบัติเหตุรถจักรยานยนต์ ปัจจัยที่วิเคราะห์ได้แบ่งออกเป็น 5 ประเภท ได้แก่ วันและเวลาที่เกิดอุบัติเหตุ สภาพถนน สภาพแวดล้อม ลักษณะเฉพาะของผู้ขับขี่รถจักรยานยนต์ และประเภทของอุบัติเหตุ ความรุนแรงของอุบัติเหตุรถจักรยานยนต์แบ่งออกเป็น 3 ประเภท ได้แก่ บาดเจ็บเล็กน้อย บาดเจ็บสาหัส และเสียชีวิต ผลการศึกษาพบว่า ปัจจัยที่เพิ่มโอกาสเสียชีวิต ได้แก่ พื้นคอนกรีต ถนนลาดชัน ผิวถนนเปียก เวลากลางคืน ผู้ขับขี่ชาย ผู้ขับขี่รถจักรยานยนต์อายุเท่ากับหรือมากกว่า 60 ปี อุบัติเหตุชนกับยานพาหนะขนาดใหญ่ รถจักรยานยนต์ชนกัน และการชนกันในทิศทางตรงกันข้าม ในทางกลับกัน ปัจจัยที่ลดโอกาสเสียชีวิต ได้แก่ ถนนหน้าบ้านและการชนกันจากการชนรถจักรยานยนต์คันอื่น นอกจากนี้ การศึกษานี้ยังได้ให้คำแนะนำในการลดความรุนแรงของอุบัติเหตุรถจักรยานยนต์บนทางหลวงในประเทศไทย

ศิรวิทย์ ชาวระ, ธเนศ เสถียรนาม, วิชุดา เสถียรนาม และชัยวุฒิ กาญจนะสันติสุข (2566) ได้ศึกษาปัจจัยที่มีอิทธิพลต่อความรุนแรงในการเกิดอุบัติเหตุย้อนศรบนทางหลวงของประเทศไทย งานวิจัยนี้วิเคราะห์อุบัติเหตุย้อนศรบนโครงข่ายถนนทางหลวงของประเทศไทยย้อนหลัง 10 ปี ระหว่างปี 2556–2565 ของกรมทางหลวง เพื่อค้นหาปัจจัยที่มีอิทธิพลต่อความรุนแรงของอุบัติเหตุย้อนศร ด้วยการวิเคราะห์หลักสถิติเชิงพรรณนา (Descriptive statistics) และ ใช้แบบจำลองสมการถดถอย โลจิสติคประเภทไบนารี่ (Binary logistic regression) ปัจจัยที่ศึกษาแบ่งได้เป็น 4 ลักษณะ ได้แก่ วันและช่วงเวลาในการเกิดอุบัติเหตุ, ลักษณะกายภาพของถนน, ลักษณะสิ่งแวดล้อม และลักษณะอุบัติเหตุ โดยกำหนดความรุนแรง

ของอุบัติเหตุของรถย้อนศรออกเป็น 2 ประเภท คือ การเสียชีวิต และการบาดเจ็บ ผลการวิเคราะห์พบว่า ปัจจัยที่มีนัยสำคัญต่อการเพิ่มโอกาสของการเสียชีวิตสูงของอุบัติเหตุย้อนศร ได้แก่ ช่วงเวลากลางคืน, ทางหลวงที่ไม่มีทางคู่ขนาน, การชนกันระหว่างรถจักรยานยนต์กับรถบรรทุกขนาดใหญ่ และการชนกันระหว่างรถจักรยานยนต์กับรถยนต์นั่งส่วนบุคคล สุดท้ายงานวิจัยนี้ยังได้เสนอแนวทางการลดความรุนแรงและการเสียชีวิตของอุบัติเหตุย้อนศรบนทางหลวงของประเทศไทย

เพ็ญพักตร์ พรชะตา (2565) ได้ศึกษาระดับการมีส่วนร่วมของภาคีเครือข่ายในการป้องกันอุบัติเหตุชุมชนในเขตอำเภอวังสามหมอ จังหวัดอุดรธานี เพื่อศึกษาวิเคราะห์ปัจจัยที่มีอิทธิพลต่อการมีส่วนร่วมของภาคีเครือข่ายในการป้องกันอุบัติเหตุชุมชนในเขตอำเภอวังสามหมอ จังหวัดอุดรธานี และ เพื่อศึกษาแนวทางพัฒนาการมีส่วนร่วมของภาคีเครือข่ายในการป้องกันอุบัติเหตุชุมชนในเขตอำเภอวังสามหมอ จังหวัดอุดรธานี กลุ่มตัวอย่างคือประชาชนผู้มีสิทธิเลือกตั้งที่อยู่ในเขตอำเภอวังสามหมอ จังหวัดอุดรธานี จำนวน 397 คน เครื่องมือที่ใช้ในการวิจัยคือแบบสอบถามและแบบสัมภาษณ์ สถิติที่ใช้ในการวิจัยเชิงปริมาณ ได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการวิเคราะห์ถดถอยพหุคูณโดยใช้วิธีการคัดเลือกตัวแปรแบบเป็นลำดับประกอบการอภิปรายผล และการวิเคราะห์ข้อมูลเชิงคุณภาพโดยใช้การพรรณนาวิเคราะห์ ผลการวิจัยพบว่า 1. ระดับการมีส่วนร่วมของภาคีเครือข่ายในการป้องกันอุบัติเหตุชุมชนในเขตอำเภอวังสามหมอ จังหวัดอุดรธานี โดยรวมทุกด้านระดับการมีส่วนร่วมของภาคีเครือข่ายในการป้องกันอุบัติเหตุชุมชนอยู่ในระดับปานกลาง 2. ค่าน้ำหนักความสำคัญของตัวแปรปัจจัยที่มีอิทธิพลต่อการมีส่วนร่วมของภาคีเครือข่ายในการป้องกันอุบัติเหตุชุมชนในเขตอำเภอวังสามหมอ จังหวัดอุดรธานี โดยรวมทุกด้าน ด้านบทบาทของภาคีเครือข่าย ด้านบทบาทของท้องถิ่น มีค่าสัมประสิทธิ์ของตัวพยากรณ์ในคะแนนดิบเท่ากับ 596, 358 ตามลำดับ

พิชญากร เลค (2565) ได้ศึกษาการเปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องจักรเพื่อทำนายการเกิดอุบัติเหตุ โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น แบบจำลองโครงข่าย ประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลในจังหวัดที่มีจำนวนการเกิดอุบัติเหตุมากที่สุด ได้แก่ จังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมาและจังหวัดสมุทรปราการ โดยรวบรวมข้อมูลตั้งแต่ปีพ.ศ. 2558–2562 จำนวน 60 เดือน จากการทดลองพบว่าแบบจำลองที่มีความเหมาะสมในการพยากรณ์การเกิดอุบัติเหตุในประเทศไทย คือแบบจำลองการถดถอยเชิงเส้น มีความเหมาะสมมากที่สุดกับชุด

ข้อมูลจังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมาและจังหวัดสมุทรปราการ มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับร้อยละ 16.29 19.29 และร้อยละ 19.64

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือ และวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นที่เกี่ยวข้องกับ การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการการวิเคราะห์ข้อมูลด้วย CRISP–DM ด้วยการวิเคราะห์อนุกรมเวลา (Time Series Analysis) ด้วยเทคนิคการถดถอยเชิงเส้น (linear regression) เทคนิคการหาค่าเฉลี่ยเคลื่อนที่ (moving average) เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression) และเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks) เพื่อต้องการทราบถึงข้อมูลเชิงลึกของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ในการวิเคราะห์ข้อมูลเพื่อให้หน่วยงานที่เกี่ยวข้องสามารถนำไปใช้ในทางปฏิบัติในการวางแผนและพัฒนากลยุทธ์เพื่อลดอุบัติเหตุในช่วงเทศกาลปีใหม่ และสร้างความปลอดภัยให้กับประชาชนบนท้องถนนในช่วงเวลาที่มีความเสี่ยงสูง