การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้ม และวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุ ในช่วงเทศกาลปีใหม่

นางสาวกฤตยา คงดี
นางสาวปรีชญา อินต๊ะชุ่ม

หลักสูตรระบบสารสนเทศทางธุรกิจ
สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปะศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาเชียงใหม่
ปีการศึกษา 2567

การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้ม และวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุ ในช่วงเทศกาลปีใหม่

นางสาวกฤตยา คงดี
นางสาวปรีชญา อินต๊ะชุ่ม

หลักสูตรระบบสารสนเทศทางธุรกิจ
สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปะศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาเชียงใหม่
ปีการศึกษา 2567

ชื่อโครงงาน การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

โดย นางสาวกฤตยา คงดี รหัส 66541207024-5

นางสาวปรีชญา อินต๊ะชุ่ม รหัส 66541207031-0

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์หฤทัย อาษากิจ

หลักสูตร ระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ

คณะบริหารธุรกิจและศิลปศาสตร์

ปีการศึกษา 2567

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ เพื่อพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ผู้วิเคราะห์ข้อมูลนำชุดข้อมูลผู้วิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่เป็นข้อมูลที่ได้จากได้รวบรวมข้อมูลจากเว็บไซต์ศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th/ 2567) ของสถาบันการแพทย์ฉุกเฉินแห่งชาติ และได้ทำการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining) โดยทำการสร้างโมเดล 4 แบบ คือ 1.เทคนิคการถดถอยเชิงเส้น (linear regression) 2.เทคนิคเกาส์เซียน (Gussian Process) 3.เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression) 4.เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks) ใช้ชุดข้อมูลช่วงเทศกาลปีใหม่ตั้งแต่ปี 2551 ถึง 2558 โดยแบ่งข้อมูลออกเป็นชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30% โดยใช้ค่า RMSE และ MAE เป็นตัวเปรียบเทียบประสิทธิภาพ ได้ผลสรุปว่า โมเดล Linear Regression มีคา RMSE น้อยที่สุด เท่ากับ 5.758 และ โมเดล SVMR มีคา MAE นอยที่สุด เท่ากับ 4.656 ผู้วิเคราะห์จึงใชโมเดล SVMR และ โมเดล Linear Regression ในการเปรียบเทียบโมเดลอีกครั้ง โดยใชขอมูลช่วงเทศกาลปีใหม่ปี 2558 ในการคำนวณคา RMSE และ MAE ที่นอยที่สุด ผลพบว่าโมเดล Linear Regression มีค่า MAE เทากับ 3.716 และคา RMSE เท่ากับ 5.142 โมเดล SVMR มีค่า MAE เท่ากับ 3.908 มีค่า RMSE เทากับ 5.343 ผู้วิเคราะห์จึงนำเอาโมเดล Linear Regression ที่มีประสิทธิภาพดีกว่าในการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่าง

ช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ และพัฒนาเว็บไซต์เพื่อเผยแพร่ข้อมูลการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ เพื่อให้หน่วยงานที่เกี่ยวข้องสามารถนำไปใช้ในทางปฏิบัติในการวางแผนและพัฒนากลยุทธ์เพื่อลดอุบัติเหตุในช่วงเทศกาลปีใหม่ และสร้างความปลอดภัยให้กับประชาชนบนท้องถนนในช่วงเวลาที่มีความเสี่ยงสูงนี้

กิตติกรรมประกาศ

โครงงานในครั้งนี้ สามารถสำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความกรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ หฤทัย อาษากิจ อาจารย์ที่ปรึกษาโครงงาน ผู้ช่วยศาสตราจารย์สัจจธรรม สุภาจันทร์ หัวหน้าหลักสูตรระบบสารสนเทศทางธุรกิจ และคณะอาจารย์ในหลักสูตรระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาทุกท่าน ที่ได้ให้คำปรึกษา และคำแนะนำในการจัดทำโครงงาน เพื่อบรรลุวัตถุประสงค์ของโครงงานให้สำเร็จสมบูรณ์ รวมถึงช่วยแก้ไขเอกสารโครงงาน ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบคุณเว็บไซต์ data.go.th ที่ได้รวบรวมชุดข้อมูลเกี่ยวกับข้อมูลเกี่ยวกับการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในประเทศไทย และนำมาเผยแพร่บนอินเทอร์เน็ต ทำให้คณะผู้จัดทำได้มาซึ่งข้อมูลที่ต้องการ เพื่อนำมาวิเคราะห์ตามกระบวนการและขั้นตอนต่าง ๆ และขอขอบคุณองค์กรณ์ สถาบันการแพทย์ฉุกเฉินแห่งชาติ ที่ได้เผยแพร่ข้อมูลเกี่ยวกับการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ทำให้คณะผู้จัดทำได้มา ซึ่งข้อมูลเป็นแหล่งเรียนรู้ในเรื่องข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ที่เกี่ยวข้องกับโครงการ

ท้ายสุดนี้ คณะผู้วิเคราะห์ข้อมูลโครงงานขอขอบคุณผู้มีส่วนเกี่ยวข้องทุกท่านที่กรุณาประสิทธิ์ประสาทวิชาความรู้และประสบการณ์อันมีค่ายิ่งแก่ศิษย์ และคอยชี้แนะแนวทางจัดทำโครงงานพร้อมทั้ง ช่วยส่งเสริมสนับสนุนให้คณะผู้จัดทำโครงงานเกิดแรงผลักดัน เป็นกำลังใจให้ความช่วยเหลือต่าง ๆ อีกทั้งยังมอบข้อคิดเห็นที่เป็นประโยชน์ในการดำเนินโครงงานนี้ อันส่งผลให้สำเร็จลุล่วงไปด้วยดี ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ปรีชญา อินต๊ะชุ่ม

กฤตยา คงดี

สารบัญ

	หน้า
บทคัดย่อ	ก
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญภาพ	ซ
บทที่ 1 บทนำ	
1.1 ความสำคัญและที่มาของปัญหา	1
1.2 วัตถุประสงค์ของโครงงาน	3
1.3 ประโยชน์ที่คาดว่าจะได้รับจากการทำโครงงาน	3
1.4 ขอบเขตการนำเสนอข้อมูล	3
1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม	4
1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล	5
1.7 ระยะเวลาในการดำเนินการ	6
1.8 นิยามศัพท์เฉพาะ	6
บทที่ 2 แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง	
2.1 แนวคิด	9
2.2 ทฤษฎี	17
2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล	25
2.4 วรรณกรรมที่เกี่ยวข้อง	38
2.5 บทสรุป	41
บทที่ 3 วิธีดำเนินงาน	
3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM	42
3.2 การวิเคราะห์และออกแบบระบบ	47
บทที่ 4 ผลการดำเนินงาน	
4.1 ผลการวิเคราะห์ข้อมูลด้วย CRISP-DM	50
4.2 ผลการวิเคราะห์และออกแบบระบบ	71

สารบัญ (ต่อ)

	หน้า
บทที่ 5 สรุปผลการดำเนินงาน และข้อเสนอแนะ	
5.1 บทสรุปผลโครงงาน	74
5.2 ข้อจำกัดเว็บไซต์	75
5.3 ปัญหาและอุปสรรคของโครงงาน	75
5.4 ข้อเสนอแนะ	75
บรรณานุกรม	77
ภาคผนวก ก คู่มือการใช้งานเว็บไซต์	80
ภาคผนวก ข ประวัติผู้จัดทำ	88

สารบัญตาราง

ตารางที่		**หน้า**
1.1	ตารางแสดงระยะเวลาดำเนินงาน	6
3.1	ตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดล	46
4.1	ตารางแปลงค่าข้อมูลเวลาเกิดเหตุ	55
4.2	ผลตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดล	67
4.3	ตารางแสดงคา MAE และ RMSE และเปรียบเทียบกันระหวาง SVMR และ Linear Regression เมื่อใชขอมูลช่วงเทศกาลปีใหม่ปี 2558 ล่าสุด ในการทดสอบ	67

สารบัญภาพ

ภาพที่		หน้า
2.1	แสดงตัวอย่างแผนภูมิรูปแบบต่าง ๆ	12
2.2	แสดงกราฟตัวอย่างยอดขายแต่ละเดือน	13
2.3	แสดงตัวอย่างตารางช่วงอายุ	13
2.4	แสดงแผนที่จำนวนผู้ป่วย Covid-19	14
2.5	แสดงภาพอินโฟกราฟิก	14
2.6	แสดงตัวอย่างหน้าแดชบอร์ด	15
2.7	สถิติอุบัติเหตุจราจรทางบกในเทศกาลปีใหม่ (7 วันอันตราย)	16
2.8	สาเหตุการเกิดอุบัติเหตุการจราจรทางบก (5 อันดับแรก)	17
2.9	อัตราการดื่มสุราของประชากรอายุ 15 ปีขึ้นไป	17
2.10	แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ	22
2.11	แสดงโครงสร้างแบบลำดับขั้น	22
2.12	แสดงโครงสร้างแบบตาราง	23
2.13	แสดงโครงสร้างแบบใยแมงมุม	23
2.14	ตัวอย่างแผนภาพการกระจาย (Scatter Plot) ความสัมพันธ์เส้นตรงทางบวก	25
2.15	ตัวอย่างแผนภาพการกระจาย (Scatter Plot) ความสัมพันธ์เส้นตรงทางลบ	26
2.16	ตัวอย่างแผนภาพการกระจาย (Scatter Plot) ไม่มีความสัมพันธ์ของข้อมูล	26
2.17	หลักการทำงานของ Guassian Process	29
2.18	ตัวอย่างหาฟังก์ชั่นเส้นแบ่งข้อมูลออกจากกัน	30
2.19	แสดงตำแหน่งและทิศทางของเส้น	31
2.20	แสดงการกำหนดเส้นแบ่ง	31
2.21	ตัวอย่างการนำเอาข้อมูลมา dot กับเวกเตอร์ w	32
2.22	ตัวอย่างการนำเอาข้อมูลของฝั่งสีฟ้า (x-) มา dot กับเวกเตอร์ w	32
2.23	ตัวอย่างการนำเอาข้อมูลอยู่บนแนวเส้นแบ่งพอดี	33
2.24	ตัวอย่างสมการที่ได้	33

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
2.25	Artificial Neural Networks	34
2.26	Multilayer	35
2.27	ฟังก์ชันซิกมอยด์	35
3.1	หน้าแสดงผลของเว็บไซต์ข้อมูล	37
3.2	แสดงข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ. 2558	43
3.3	แสดงโครงร่างการออกแบบหน้าแรกของเว็บไซต์	43
3.4	แสดงผลข้อมูลที่มาและความสำคัญ	47
3.5	แสดงผลสถานการณ์ผู้บาดเจ็บและเสียชีวิตของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558	47
3.6	แสดงผล Dashboard ของข้อมูลการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่	48
3.7	แสดงผลข้อมูลผู้จัดทำ	48
4.1	หน้าแสดงผลของเว็บไซต์ข้อมูล	49
4.2	การเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ที่ได้มาจากศูนย์กลางข้อมูลเปิดภาครัฐของประเทศไทย	51
4.3	ข้อมูลเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ. 2558	51
4.4	แสดงการคัดเลือกข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่	52
4.5	แสดงข้อมูลการแก้ไขด้วยการปรับรูปแบบวันที่	54
4.6	แสดงข้อมูลการแทนค่าด้วยจังหวัด 77 จังหวัดให้เป็นภาษาอังกฤษ	55
4.7	แสดงข้อมูลการแทนค่าข้อมูลเวลาเกิดเหตุให้ป็นตัวเลข	56
4.8	แสดงข้อมูลการแทนค่าข้อมูลของเพศให้เป็นภาษาอังกฤษ	56
4.9	แสดงข้อมูลการแทนค่าข้อมูลของถนนที่เกิดเหตุให้เป็นภาษาอังกฤษ	57
4.10	แสดงข้อมูลการแทนค่าข้อมูลของสถานะให้เป็นภาษาอังกฤษ	57

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
4.11	แสดงข้อมูลการแทนค่าข้อมูลของรถผู้บาดเจ็บให้เป็นภาษาอังกฤษ	57
4.12	แสดงข้อมูลการแทนค่าข้อมูลของรถคู่กรณีให้เป็นภาษาอังกฤษ	58
4.13	แสดงข้อมูลการแทนค่าข้อมูลของมาตรการให้เป็นภาษาอังกฤษ	58
4.14	แสดงข้อมูลการแทนค่าข้อมูลของการดื่มสุราให้เป็นภาษาอังกฤษ	59
4.15	แสดงข้อมูลการแทนค่าข้อมูลของการผลการรักษาให้เป็นภาษาอังกฤษ	59
4.16	แสดงกระบวนการประมวลผลและสร้างโมเดลเทคนิคการถดถอยเชิงเส้น (linear regression)	60
4.17	การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล linear regression	61
4.18	แสดงกระบวนการประมวลผลและสร้างโมเดลเทคนิคเกาส์เซียน (Gussian Process)	61
4.19	การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล Gussian Process	62
4.20	แสดงกระบวนการประมวลผลและสร้างโมเดล เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression)	62
4.21	การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล SVM	63
4.22	แสดงกระบวนการประมวลผลและสร้างโมเดลเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks)	63
4.23	การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล ANN	64
4.24	ผลการวัดประสิทธิภาพของโมเดล Linear Regression	65
4.25	ผลการวัดประสิทธิภาพของโมเดล Gussian Process	65
4.26	ผลการวัดประสิทธิภาพของโมเดล Support Vector Machine Regression	66
4.27	ผลการวัดประสิทธิภาพของโมเดล Artificial Neural Networks	66

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
4.28	กราฟผลการเปรียบเทียบระหว่างค่าจริงและค่าพยากรณ์ที่ได้ในโมเดลLinear Regression ของจำนวนผู้เสียชีวิตและผู้บาดเจ็บของช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ในปี พศ.2551-2558	68
4.29	แสดงกราฟผลการเปรียบเทียบระหว่างค่าจริงของปี 2007-2015 และค่าพยากรณ์	70
4.30	แสดงโครงร่างการออกแบบหน้าแรกของเว็บไซต์	71
4.31	แสดงหน้าข้อมูลที่มาและความสำคัญ	72
4.32	แสดงผลสถานการณ์ผู้บาดเจ็บและเสียชีวิตของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558	72
4.33	แสดงผล Dashboard ของข้อมูลการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่	73
4.34	แสดงผลข้อมูลผู้จัดทำ	73

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหา

ปัจจุบันอุบัติเหตุบนถนนกลายเป็นสาเหตุที่ทำให้เด็กและวัยรุ่นทั่วโลกเสียชีวิตมากที่สุด องค์การอนามัยโลก (World Health Onganization – WHO) ระบุว่าทวีปแอฟริกา มีอัตราการเสียชีวิตจากอุบัติเหตุบนถนนสูงที่สุดในโลกจากรายงานสถานการณ์ความปลอดภัยทางถนน ปี 2015 (Giobal status report) พบว่ามีผู้เสียชีวิตจากอุบัติเหตุทางถนน ทั่วโลกสูงถึง 1.25 ล้านคน บาดเจ็บจำนวน 20 – 50 ล้านคนในแต่ละปีในขณะที่การเสียชีวิตจากอุบัติเหตุทางถนนในภูมิภาคเอเชียใต้และตะวันออกมีอัตราตายอยู่ที่ 17 ต่อประชากรแสนคนในปี พ.ศ. 2557 ประเทศไทยเป็นประเทศที่มีอัตราการเสียชีวิตจากอุบัติเหตุบนถนนสูงสุดในเอเชียตะวันออกเฉียงใต้มีผู้เสียชีวิตจากอุบัติเหตุทางถนนจำนวน 24,237 คนเป็นอันดับ 2 ของโลก และเมื่อเปรียบเทียบกับประเทศอื่น ๆ ในภูมิภาคเอเชียตะวันออกเฉียงใต้โดยแบ่งตามภูมิศาสตร์ พบว่าประเทศไทยยังคงมีอัตราการเสียชีวิตบนท้องถนนสูงสุด โดยอยู่ที่ 32.7 คน ต่อประชากร 1 แสนคน ตามมาด้วยเวียดนาม 26.4 คน, มาเลเซีย 23.6 คน, เมียนมา 19.9 คน, กัมพูชา 17.8 คน, ลาว 16.6 คน, ปาปัวนิวกินี 14.2 คน, ติมอร์เลสเต 12.7 คน, ฟิลิปปินส์ 12.3 คน, อินโดนีเซีย 12.2 คน, สิงคโปร์ 2.8 คน ต่อประชากร 1 แสนคน ส่วนบรูไนไม่มีข้อมูล เมื่อพิจารณาอัตราการเสียชีวิตจากอุบัติเหตุบนท้องถนนแบ่งตามประเภทผู้ใช้ยานพาหนะ พบว่าไทยมีอัตราการเสียชีวิตของคนใช้รถยนต์, คนใช้จักรยานยนต์ และคนใช้จักรยาน อยู่ที่ 4 คน, 24.3 คน และ 1.1 คน ต่อประชากร 1 แสนคน ตามลำดับ ส่วนอัตราการเสียชีวิตของคนเดินเท้าในไทยอยู่ที่ 2.5 คน ต่อประชากร 1 แสนคน เป็นรอง ปาปัวนิวกินี, ศรีลังกา และเมียนมา ซึ่งอยู่ที่ 6.7 คน, 4.3 คน และ 2.8 คน ต่อประชากร 1 แสนคน ตามลำดับ สมัชชาสหประชาชาติจึงได้ประกาศเจตนารมณ์ในปฏิญญามอสโกให้ในปีพ.ศ. 2554–2563 เป็นทศวรรษแห่งความปลอดภัยทางถนน (Decade of Action for Read Safety) ประเทศไทยในฐานะประเทศสมาชิกสมัชชาสหประชาชาติได้รวมขับเคลื่อนวาระความปลอดภัยทางถนนของโลก เป็นทศวรรษแห่งความปลอดภัยทางถนนจำแนกเพื่อลดความลำเอียง (bias)

การเกิดอุบัติเหตุบนท้องถนนเป็นหนึ่งในปัญหาที่สำคัญของสังคมไทย โดยเฉพาะในช่วงเทศกาลสำคัญ เช่น เทศกาลปีใหม่ ที่ประชาชนมักเดินทางกลับภูมิลำเนาหรือออกท่องเที่ยว ทำให้จำนวนยานพาหนะบนท้องถนนเพิ่มขึ้นอย่างมาก ส่งผลให้ความเสี่ยงต่อการ

เกิดอุบัติเหตุเพิ่มสูงขึ้นตามไปด้วย นอกจากนี้ สถิติการเกิดอุบัติเหตุในช่วงเทศกาลนี้ยังมีแนวโน้มเพิ่มขึ้นเรื่อยๆ โดยสาเหตุหลักมักจะมาจากการดื่มสุราแล้วขับขี่ การขับรถเร็วเกินกว่าที่กฎหมายกำหนด และความเหนื่อยล้าจากการเดินทางไกล ซึ่งนำไปสู่การสูญเสียทั้งชีวิตและทรัพย์สินจากสถิติในช่วงเทศกาลปีใหม่ พบว่ามีจำนวนอุบัติเหตุที่เกิดขึ้นและจำนวนผู้เสียชีวิตหรือบาดเจ็บเพิ่มขึ้นอย่างมีนัยสำคัญเมื่อเทียบกับช่วงเวลาปกติ ปัญหาการเกิดอุบัติเหตุบนท้องถนนในช่วงเทศกาลปีใหม่ยังเป็นประเด็นที่รัฐบาลและหน่วยงานที่เกี่ยวข้องให้ความสำคัญเป็นพิเศษ โดยมีการจัดตั้งศูนย์อำนวยการความปลอดภัยทางถนนเพื่อรณรงค์และลดอุบัติเหตุในช่วงดังกล่าว การเก็บสถิติจากกลุ่มป้องกันการบาดเจ็บจากการจราจรในประเทศไทยพบว่าในปี 2557 เพศชายมีการตายจากอุบัติเหตุทางถนนสูงกว่าเพศหญิงประมาณ 3 เท่า กลุ่มอายุที่เสียชีวิตจากอุบัติเหตุทางถนนมากที่สุด คือ กลุ่มอายุ 15–19 ปี รถจักรยานยนต์เป็นพาหนะที่มีจำนวนการเสียชีวิตมากที่สุด ช่วงเวลาที่มีจำนวน ผู้เสียชีวิตมากที่สุดคือเดือนมกราคม ซึ่งบุคคลากรในประเทศชาติถือว่าเป็นส่วนสำคัญในการพัฒนาประเทศ นอกจากนี้รัฐบาลยังต้องสูญเสียงบประมาณ ในการบริหารประเทศเพื่อมารณรงค์ ป้องกัน แก้ไขปัญหา และรักษาผู้บาดเจ็บจากอุบัติเหตุที่นับวันจะยิ่งจะมีมากขึ้นเรื่อยๆ

การวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ คือนำข้อมูลที่ได้มารวบรวมไว้ และเผยแพร่ผ่านเว็บไซต์ ศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th/ 2567) ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่างๆ ภายในประเทศซึ่งเป็นประโยชน์ข้อมูลที่สามารถเปิดเผยได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาซึ่งการนำข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่มาทำการวิเคราะห์นั้น เนื่องจากต้องการทราบถึงข้อมูลเชิงลึกของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ซึ่งการวิเคราะห์ปัจจัยต่างๆ จากข้อมูลที่ได้มานั้นถือเป็นเรื่องสำคัญต่อการศึกษาหาผลสรุปแต่ด้วยจำนวนข้อมูลที่มีในปริมาณที่มากนั้น ทำให้ยากต่อการนำเสนอให้เข้าใจได้ง่าย ซึ่งถือเป็นข้อจำกัดหากต้องการนำข้อมูลชุดนี้ไปใช้เพื่อเผยแพร่ให้แก่บุคคลภายนอกหรือกลุ่ม ผู้ใช้งานที่ต้องการข้อมูลซึ่งข้อมูลดังกล่าวจะสามารถนำมาเป็นประโยชน์ต่อการป้องกันและลดความเสี่ยงในการเกิดอุบัติเหตุได้อีกด้วย

ในการศึกษานี้ผู้ศึกษาจึงมีความสนใจที่จะศึกษา การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ โดยการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining) ในการวิเคราะห์โดยนำชุดข้อมูลผู้บาดเจ็บและเสียชีวิต ในช่วงเทศกาล ปี 2551–2558 ที่ได้มาจากศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th/ 2567) เพื่อให้หน่วยงานที่เกี่ยวข้องสามารถนำไปใช้ในทางปฏิบัติใน

การวางแผนและพัฒนากลยุทธ์เพื่อลดอุบัติเหตุในช่วงเทศกาลปีใหม่ และสร้างความปลอดภัยให้กับประชาชนบนท้องถนนในช่วงเวลาที่มีความเสี่ยงสูงนี้

1.2 วัตถุประสงค์ของโครงงาน

1.2.1 เพื่อศึกษาโมเดลที่เหมาะสมสำหรับการพยากรณ์แนวโน้มการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

1.2.2 เพื่อวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

1.2.3 เพื่อเผยแพร่ข้อมูลผ่านเว็บไซต์

1.3 ประโยชน์ที่คาดว่าจะได้รับจากการทำโครงงาน

1.3.1 ได้โมเดลที่เหมาะสมสำหรับการพยากรณ์แนวโน้มการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

1.3.2 ได้ผลวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

1.3.3 ได้เผยแพร่ข้อมูลผ่านเว็บไซต์เพื่อวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

1.4 ขอบเขตการนำเสนอข้อมูล

1.4.1 ขอบเขตของวิเคราะห์ข้อมูล

1.4.1.1 ข้อมูลที่นำมาวิเคราะห์ เป็นข้อมูลที่ได้จากได้รวบรวมข้อมูลจากเว็บไซตศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th/ 2567) ของสถาบันการแพทย์ฉุกเฉินแห่งชาติ ข้อมูลประมาณ 214,951 แถว 19 คอลัมน์

1.4.1.2 ผู้วิเคราะห์ข้อมูลได้ทำการรวบรวมข้อมูลที่ได้จากการวิเคราะห์ข้อมูลในประเด็นการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

1.4.1.3 ผู้วิเคราะห์ใช้กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM (Cross Industry Standard Process for Data Mining)

1.4.1.4 สามารถพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ โดยการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining) ด้วยเทคนิคการถดถอยเชิงเส้น (linear regression) เทคนิคเกาส์เซียน (Gussian Process) เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression) และเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks) โดยใช้โปรแกรมการวิเคราะห์ข้อมูล เช่น โปรแกรม SPSS, โปรแกรม RapidMiner Studio, โปรแกรม Power BI, โปรแกรม Microsoft Excel หรือโปรแกรมอื่นๆ ที่เหมาะสมกับการวิเคราะห์ข้อมูล

1.4.1.5 ทดสอบประสิทธิภาพของโมเดลโดยใช้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Square Error : MSE) และ ค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root Mean Square Error : RMSE)

1.4.1.6 ผู้วิเคราะห์สามารถนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของแผนภูมิรูปภาพโดยใช้โปรแกรมเข้ามาช่วย เช่น โปรแกรม Power BI, โปรแกรม GoogleData Studio, โปรแกรม Endless loop, โปรแกรม Tableau, โปรแกรม Fine Report, หรือโปรแกรมอื่นๆ ที่เหมาะสมกับการนาเสนอข้อมูลในรูปแบบของแผนภูมิรูปภาพ

1.4.1.7 สามารถนำเสนอเผยแพร่ข้อมูลสารสนเทศนี้บน Web Browser โดยการใช้ภาษา HTML, PHP

1.4.2 ขอบเขตการนำเสนอข้อมูล

1.4.2.1 ขอบเขตผู้ใช้งานทั่วไปบนเว็บไซต์

1.4.2.1.1 สามารถดูแนวโน้มและการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่จากการพยากรณ์ในลักษณะแผนภูมิ

1.4.2.1.2 สามารถดูผลความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่จากการวิเคราะห์ข้อมูลในลักษณะแผนภูมิ

1.4.2.1.3 สามารถดูข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในแต่ละปี

1.4.2.1.4 สามารถดูข้อมูลประเภทยานพาหนะที่เกิดอุบัติเหตุในแต่ละพื้นที่

1.4.2.2 ขอบเขตของผู้ดูแลระบบ

1.4.2.2.1 สามารถเผยแพร่ข้อมูลสารสนเทศ

1.4.2.2.2 สามารถอัปโหลดข้อมูลเกี่ยวกับการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

1.4.2.2.3 ผู้ดูแลระบบสามารถเพิ่ม ลบ แก้ไข ข้อมูล

1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 Notebook Acar Laptop –KFVU008EAMD Ryzen 7

1.5.2 Software

1.5.1.2 โปรแกรม Microsoft Office 365 ใช้ในการจัดทำเอกสารต่าง ๆ

1.5.1.3 โปรแกรมจัดการฐานข้อมูล เช่น Microsoft SQL Server, โปรแกรม MySQLหรือโปรแกรมอื่น ๆ ที่เหมาะสมกับการจัดการฐานข้อมูล

1.5.1.4 เครื่องมือการวิเคราะห์ข้อมูล และเครื่องมือการจัดการฐานข้อมูลขนาดใหญ่เช่นโปรแกรม Microsoft Excel ,โปรแกรม Rapid Miner Studio, โปรแกรม R Studio, โปรแกรม Weka, โปรแกรม Tableau, โปรแกรม SAS, โปแกรม SPSS หรือซอฟต์แวร์อื่น ๆ ที่เหมาะสมกับการวิเคราะห์ข้อมูล และเหมาะสำหรับการจัดการฐานข้อมูลขนาดใหญ่

1.5.1.5 การนำเสนอข้อมูลภาพ เช่น โปรแกรม Microsoft Power BI, โปรแกรม Tableau, เช่นโปรแกรม Microsoft Excel ,โปรแกรม Rapid Miner Studio, โปรแกรม R Studio, โปรแกรม Weka, โปรแกรม Tableau, โปรแกรม SAS, โปแกรม SPSS หรือซอฟต์แวร์อื่น ๆ ที่เหมาะสมกับการวิเคราะห์ข้อมูล และเหมาะสำหรับการจัดการฐานข้อมูลขนาดใหญ่

1.5.1.6 การออกแบบเว็บไซต์ เช่น โปรแกรม Adobe XD, เว็บไซต์ Figma.com หรือซอฟต์แวร์อื่น ๆ ที่เหมาะสมกับการออกแบบเว็บไซต์

1.5.1.7 ภาษาที่ใช้ในการพัฒนาเว็บไซต์ เช่น ภาษา HTML, ภาษา Java Script, ภาษา PHP, ภาษา CSS และภาษาอื่น ๆ ที่เหมาะสมกับการพัฒนาเว็บไซต์

1.5.1.8 Microsoft PowerPoint ใช้ในการนำเสนอ

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา 128 ถนนห้วยแก้ว ตำบล ช้างเผือก อำเภอ เมือง จังหวัดเชียงใหม่ 50300

1.7 ระยะเวลาในการดำเนินการ

ตารางที่ 1.1 ตารางแสดงระยะเวลาดำเนินงาน

แผนการดำเนินการ	2567					2568				
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.
1. ศึกษาและกำหนดความต้องการ										
2.ทำความเข้าใจ										
3.การเตรียมข้อมูลและตรวจสอบความถูกต้อง										
4.การสร้างแบบจำลองข้อมูล										
5.การประเมินแบบจำลองข้อมูล										
6.การออกแบบระบบ										
7.การเขียนและทดสอบระบบ										
8.การติดตั้งระบบ										
9.จัดทำเอกสารประกอบโครงการ										

1.8 นิยามศัพท์เฉพาะ

1.8.1 ช่วงเทศกาลปีใหม่ หมายถึง ช่วงเวลาที่ครอบคลุมทั้งวันส่งท้ายปีเก่าและวันขึ้นปีใหม่ ซึ่งโดยทั่วไปจะเริ่มตั้งแต่วันที่ 27 ธันวาคมถึงวันที่ 5 มกราคม โดยในหลายประเทศรวมถึงประเทศไทย ช่วงเทศกาลปีใหม่มักจะมีการหยุดงานและการเฉลิมฉลองต่อเนื่องเป็นเวลา 2–3 วันหรือนานกว่านั้น ช่วงเทศกาลปีใหม่ยังรวมถึงวันหยุดยาวที่รัฐบาลกำหนดเพื่อให้ประชาชนมีโอกาสเดินทางกลับภูมิลำเนาหรือไปท่องเที่ยว ทำให้ช่วงนี้มีการเดินทางบนท้องถนนหนาแน่นและมีการจัดงานเฉลิมฉลองมากมาย ส่งผลให้ความเสี่ยงต่อการเกิดอุบัติเหตุเพิ่มขึ้นอย่างมาก

1.8.2 อุบัติเหตุบนท้องถนน หมายถึง อุบัติเหตุเกี่ยวกับยานยนต์ ซึ่งเป็นสาเหตุการตายและ บาดเจ็บ ที่ผู้ประสบเหตุบนท้องถนน ด้วยจักรยานยนต์ หรือยานพาหนะต่างๆ รวมไปถึงคนเดินเท้าที่ ถูกรถจักรยานยนต์หรือพาหนะอื่นๆ เฉี่ยวชน

1.8.3 วิเคราะห์ มาจากคำว่า พิเคราะห์ ซึ่งเป็นการเปลี่ยน พ เป็น ว ในภาษาไทยซึ่งแปลความหมายได้ว่า การพินิจพิเคราะห์ การพิจารณา การใคร่ครวญ การไต่สวนความหรือเรื่องราว “วิเคราะห์” ตรงกับภาษาอังกฤษว่า “Analysis” ซึ่งแปลว่า การแยกสิ่งที่ประกอบกัน

ออกเป็นส่วน ๆ เช่น การแยกระบบใหญ่ออกเป็นส่วนย่อย ๆ คือ เป็นการแยกปัญหาออกเป็นส่วน ๆ เพื่อสะดวกในการพิจารณาหรือตัดสินใจ จามความหมายของคำว่าวิเคราะห์ดังกล่าวนี้จะเห็นว่า การวิเคราะห์ระบบงานไม่ใช่เรื่องที่ยุ่งยากหรือเรื่องที่สลับซับซ้อนแต่ประการใด

1.8.4 เว็บไซต์ คือ สื่อนำเสนอข้อมูลบนเครื่องคอมพิวเตอร์ หรือคือการรวบรวม หน้าเว็บเพจหลายหน้า ซึ่งเชื่อมโยงกันผ่านทางไฮเปอร์ลิงก์ ซึ่งต้องเปิดด้วยโปรแกรมเฉพาะทางที่เรียกว่า Web Browser โดยถูกจัดเก็บไว้ในเวิลด์ไวด์เว็บ และเว็บไซต์นั้นถูกสร้างขึ้นด้วยภาษาทางคอมพิวเตอร์ที่เรียกว่า HTML (Hyper Text Markup Language) และได้มีการพัฒนาและนำภาษาอื่นๆเข้ามาร่วมด้วย เพื่อให้มีความสามารถมากขึ้น เช่น PHP , SQL , Java ฯลฯ

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

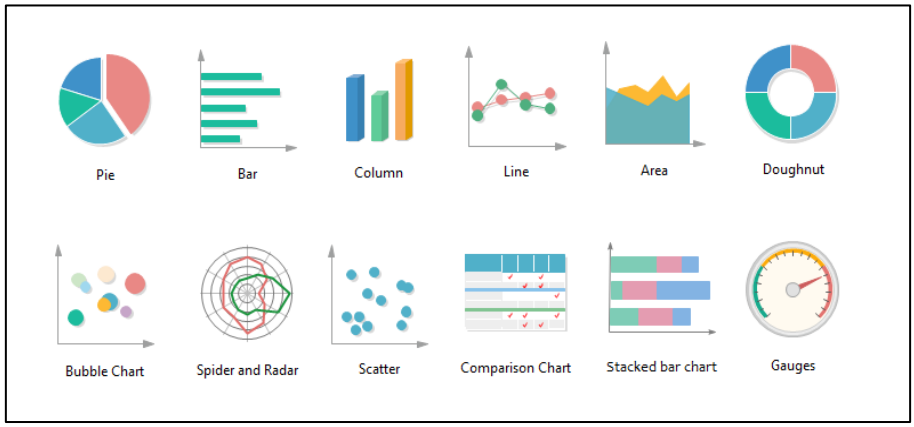
ในแต่ละเส้นทาง ปริมาณการขนส่งทางถนน ทางน้ำและทางอากาศ รายละเอียดเส้นทางคมนาคม ข้อมูลเกี่ยวกับการจัดสรรความถี่วิทยุ จำนวนครัวเรือนที่มีเครื่องรับวิทยุ โทรทัศน์ เป็นต้น

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

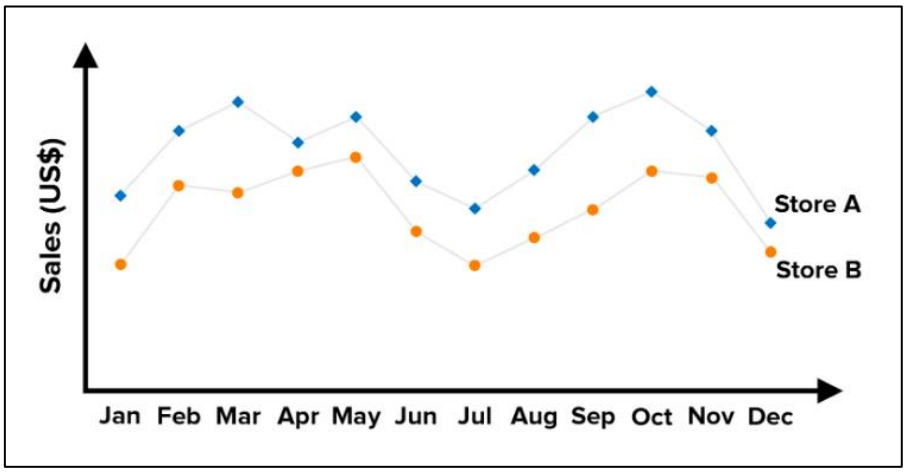
1) แผนภูมิ (Charts) ซึ่งเป็นรูปแบบที่น่าจะคุ้นเคยกันมากที่สุด และเป็นรูปแบบที่มีหลากหลายชนิดที่เหมาะกับการนำเสนอข้อมูลที่แตกต่างกันไปตามวัตถุประสงค์ เช่น Pie chart จะช่วยให้เราเห็นปริมาณความแตกต่างได้ชัดเจน, Comparison chart เหมาะสำหรับการเปรียบเทียบคุณสมบัติหลายๆ ข้อ มาตรวัด (Gauges) จะช่วยให้เห็นความเข้มข้น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

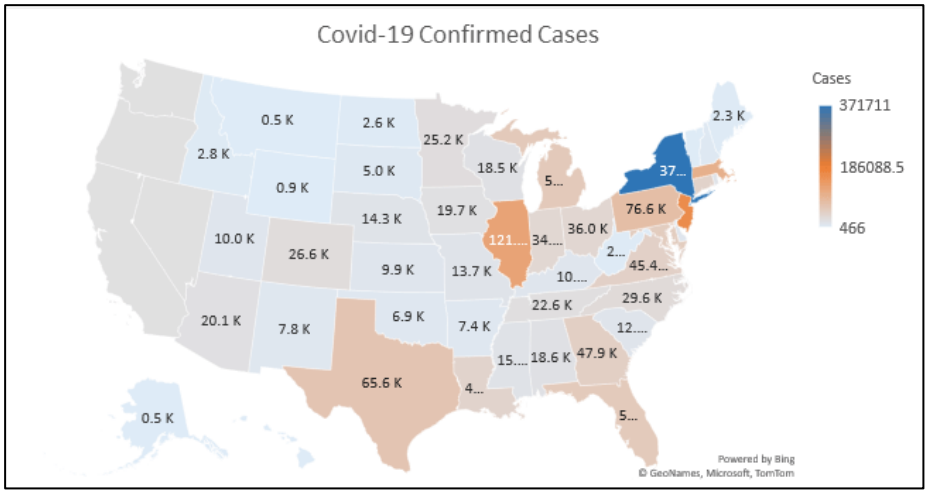
3) ตาราง (Tables) ก็เป็นอีกรูปแบบที่ใช้กันมากเพื่อนำเสนอข้อมูลให้ออกมา ดูง่าย ตารางประกอบไปด้วย 2 ส่วน ได้แก่ คอลัมน์และแถว ซึ่งช่วยจัดการข้อมูลให้เรียบร้อย ช่วยให้มองเห็นบริบทและความสัมพันธ์ของข้อมูลหลายๆ ชุดได้อย่างง่ายดาย

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)

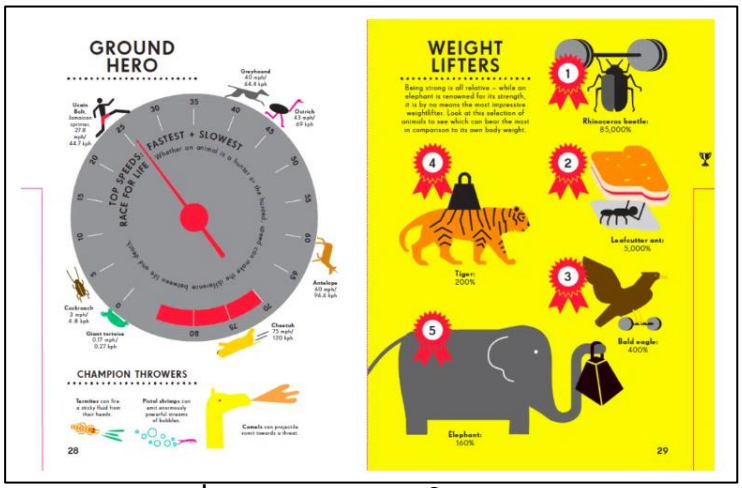
4) แผนที่ (Maps) เป็นการนำเสนอข้อมูลบนแผนที่เพื่อแสดงข้อมูลเกี่ยวกับ พื้นที่ต่างๆ ยกตัวอย่างเช่น การนำเสนอข้อมูลยอดผู้ติดเชื้อ Covid-19 ในแต่ละรัฐของประเทศสหรัฐอเมริกา ซึ่งนอกจากการใส่ข้อมูลลงไปยังพื้นที่ต่างๆ แล้ว ยังสามารถใช้สีสันเพื่อบอกช่วงปริมาณหรือความหนาแน่นของผู้ติดเชื้ออีกด้วย



ภาพที่ 2.4 แสดงแผนที่จำนวนผู้ป่วย Covid-19

(ที่มา: spreadsheetweb.com)

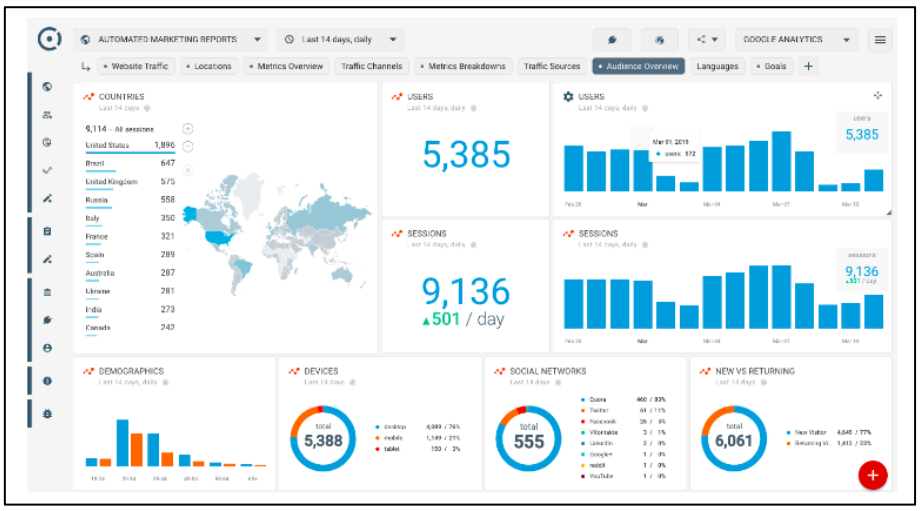
5) อินโฟกราฟิก (Infographic) คือการนำเสนอสารสนเทศ (Info: information) ด้วยภาพกราฟิก (Graphic) เป็นรูปแบบการนำเสนอข้อมูลที่ใช้ภาพสื่อแทน ทำให้ผู้อ่านข้อมูล เข้าใจข้อมูลได้ง่ายหรือสามารถทำความเข้าใจผ่านภาพแทนที่คุ้นเคย นอกจากนี้อินโฟกราฟิกยังเป็นรูปแบบการนำเสนอข้อมูลที่น่าสนใจ มีการนำเทคนิคการเล่าเรื่อง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

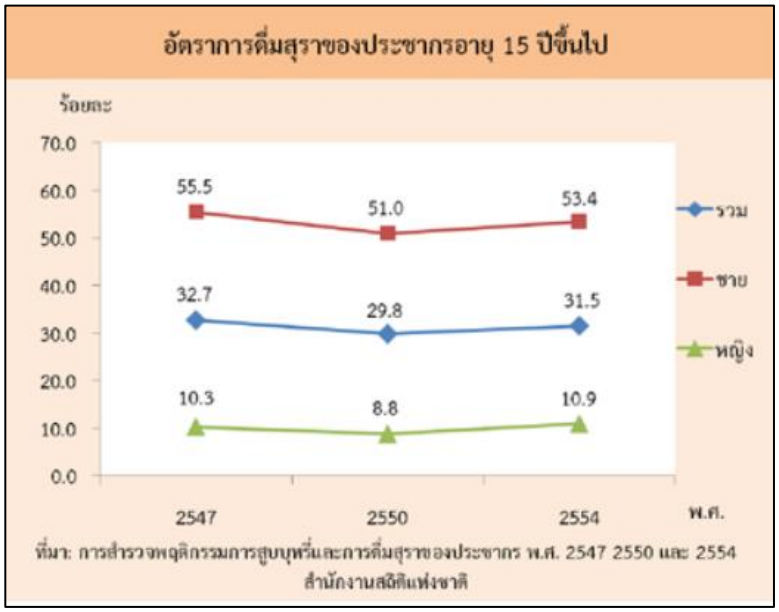
ภาพที่ 2.7 สถิติอุบัติเหตุจราจรทางบกในเทศกาลปีใหม่ (7 วันอันตราย)

(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

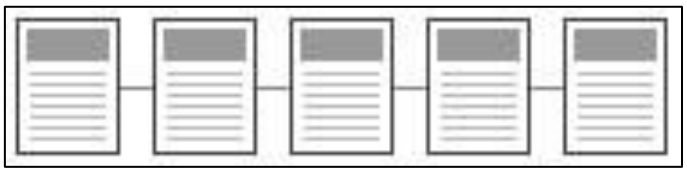
8. ความคงที่ของการออกแบบ การออกแบบเว็บไซต์ควรจะมีความคงที่ในการออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกันและมีการเรียบเรียงเนื้อหาอย่างรอบคอบ ทำให้เว็บมีความน่าเชื่อถือและดูมีคุณภาพช่วยสร้างความประทับใจให้กับผู้ใช้งานได้เป็นอย่างดี

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

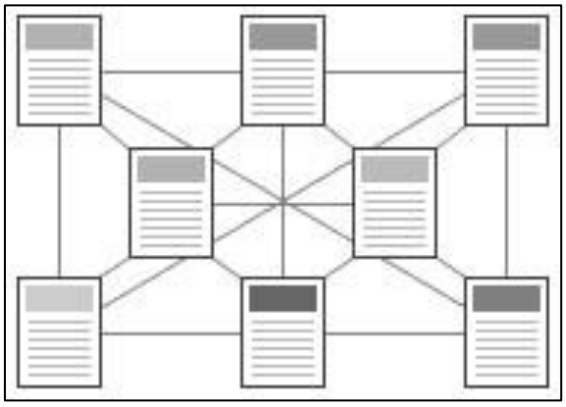
3. โครงสร้างแบบตาราง

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

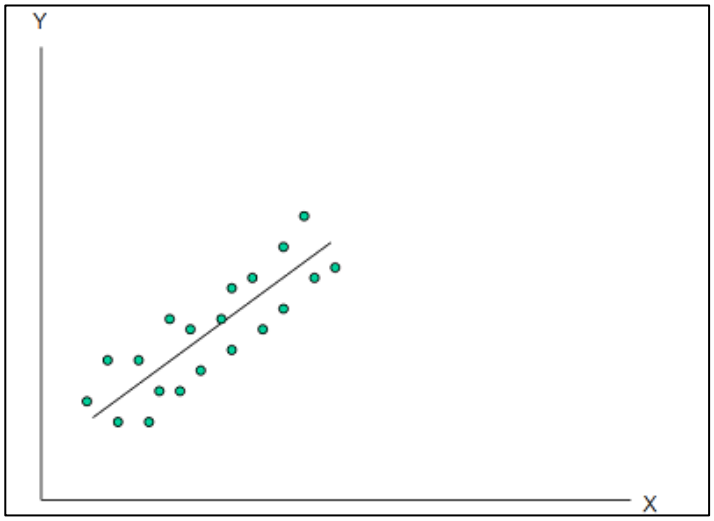
3) สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.2 เทคนิคการถดถอยเชิงเส้น (linear regression)

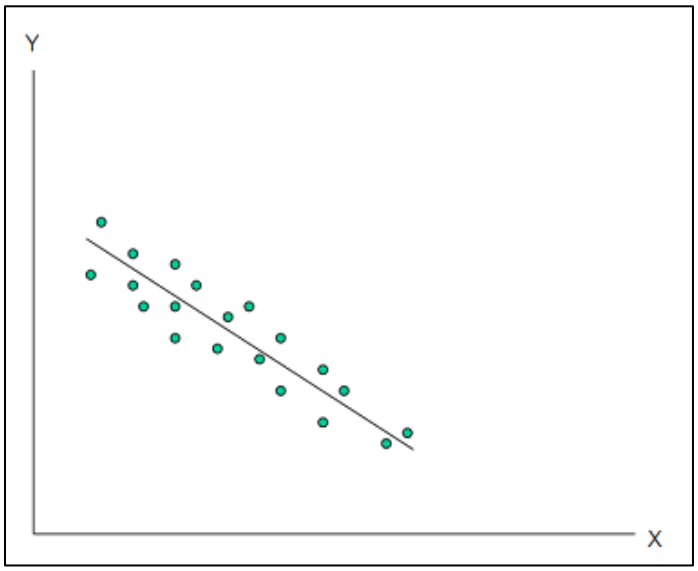
การวิเคราะห์การถดถอยเชิงเส้น เป็นการคำนวณหาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร คือ ตัวแปรที่เราทราบค่า (Predictor : x) และตัวแปรที่เราไม่ทราบค่า (Response : y) ซึ่งเป็นความสัมพันธ์แบบเชิงเส้น (Linear) โดยการคำนวณจากค่า x และ y ที่มีความสัมพันธ์กันจะใช้สมการของ Linear Regression คือ y=ax+b บางท่านอาจจะเรียกตัวแปร 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



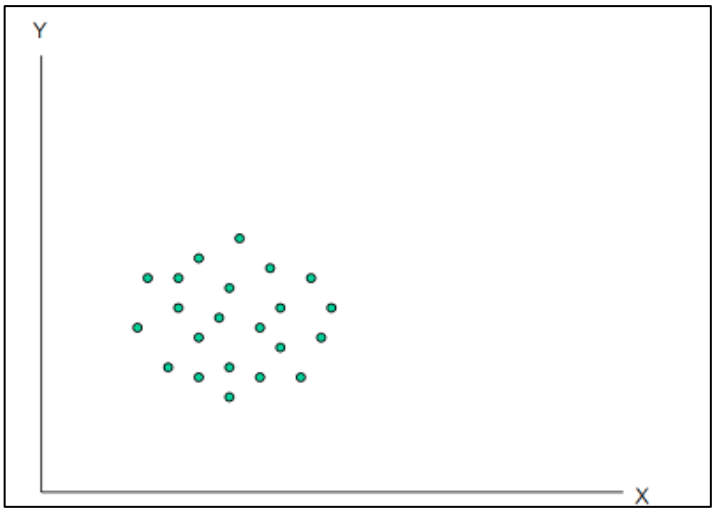
ภาพที่ 2.14 ตัวอย่างแผนภาพการกระจาย (Scatter Plot) ความสัมพันธ์เส้นตรงทางบวก (ที่มา: https://kongruksiam.medium.com/สรุป-machine-learning-ep-3-การวิเคราะห์การถดถอยเชิงเส้น-linear-regression-891260e4a957)

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

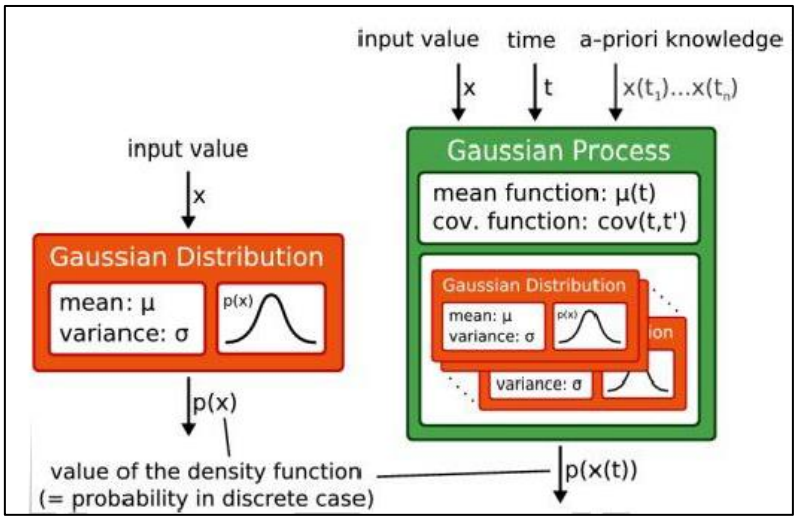
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

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

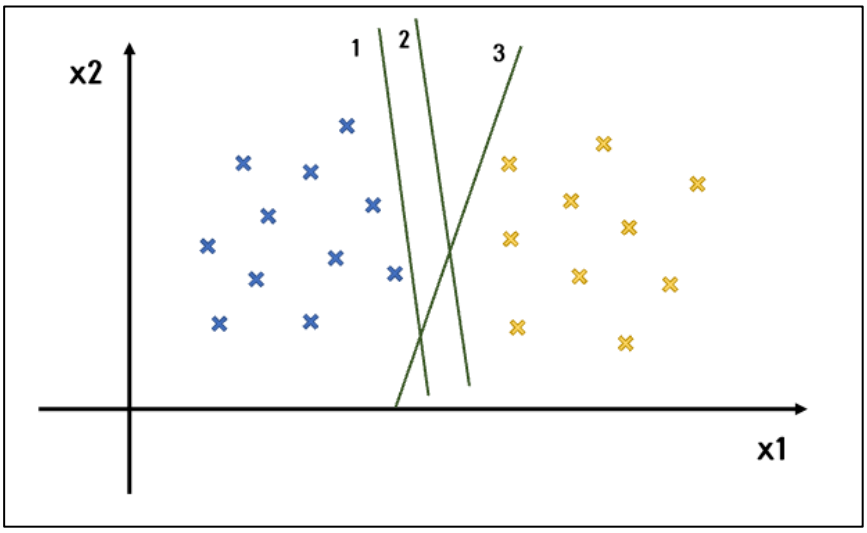
แนวคิดของกระบวนการแบบเกาส์เซียนได้รับการตั้งชื่อตามคาร์ล ฟรีดริช เกาส์ เนื่องจากแนวคิดนี้มีพื้นฐานมาจากแนวคิดของการแจกแจงแบบเกาส์เซียน (การแจกแจงแบบปกติ) กระบวนการแบบเกาส์เซียนสามารถมองได้ว่าเป็นการสรุปแบบมิติอนันต์ของการแจกแจงแบบปกติหลายตัวแปร

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)

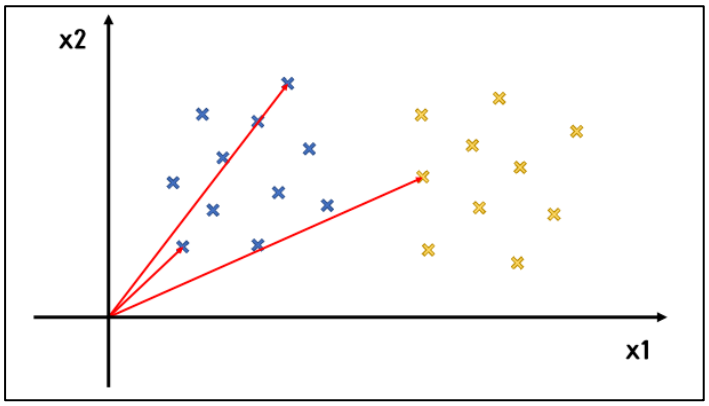
จากรูป สมมติว่าเรามีข้อมูล input สองกลุ่ม สีเหลืองและสีฟ้า มี 2 features ซึ่งก็คือ x1 และ x2 และเราต้องการจะหาฟังก์ชั่นเส้นแบ่งข้อมูลออกจากกัน เส้นไหนจะเป็นเส้นที่ดีที่สุด? ระหว่าง 1, 2 และ 3

เส้นที่ 1 อยู่ชิดกับข้อมูลกลุ่มสีฟ้ามากเกินไป

เส้นที่ 2 อยู่ตรงกลางพอดี (เส้นนี้ดีที่สุด)

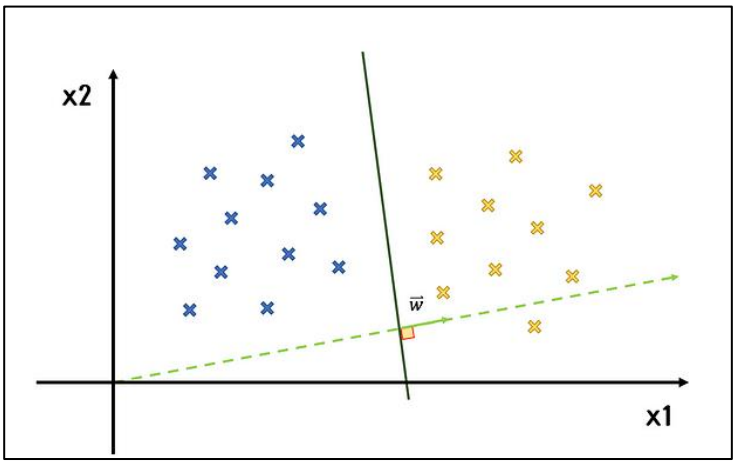
เส้นที่ 3 อยู่ติดกับทั้งข้อ

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



ภาพที่ 2.19 แสดงตำแหน่งและทิศทางของเส้น

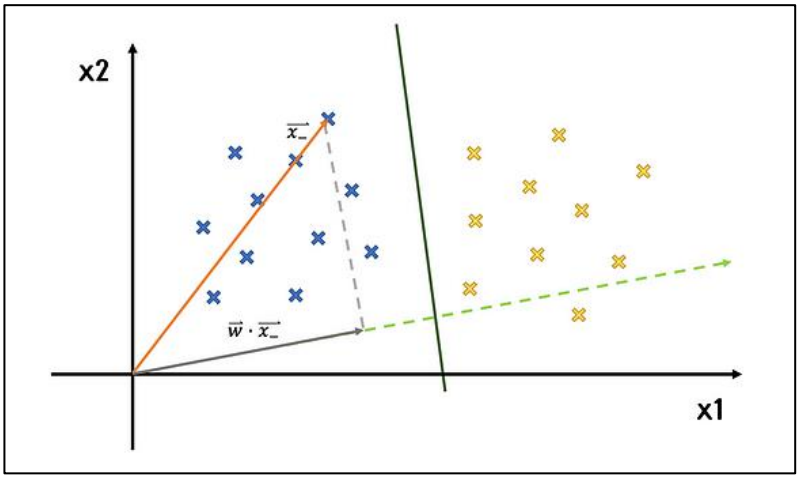
(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765



ภาพที่ 2.20 แสดงการกำหนดเส้นแบ่ง

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

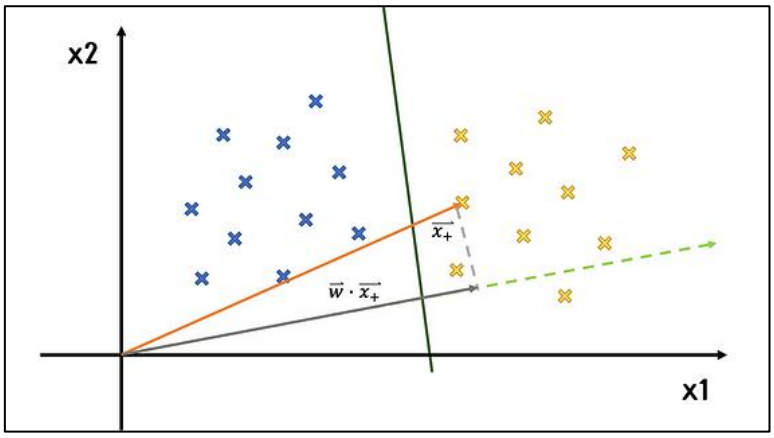
กำหนดว่าเส้นแบ่งของเราสามารถเขียนแทนได้ด้วยเวกเตอร์ขนาดหนึ่งหน่วย และตั้งฉากกับเส้นแบ่ง มีชื่อว่าเวกเตอร์ w ดังภาพ



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



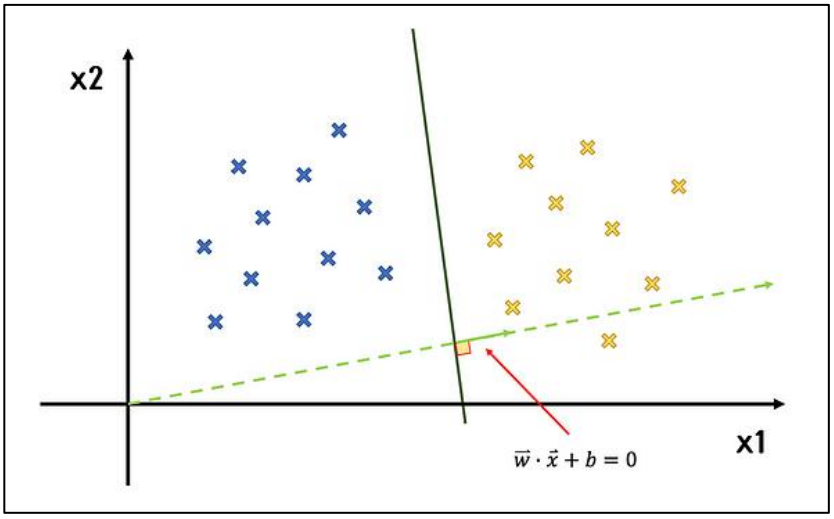
ภาพที่ 2.22 ตัวอย่างการนำเอาข้อมูลของฝั่งสีฟ้า (x-) มา dot กับเวกเตอร์ w

(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

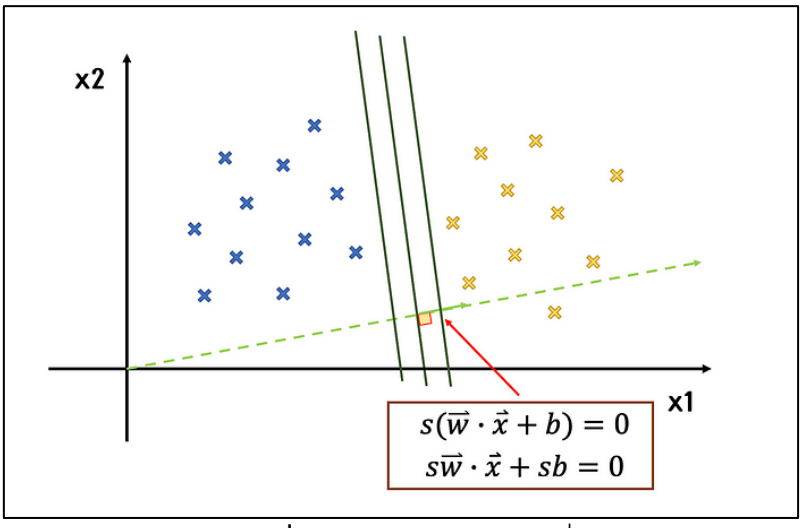
(ที่มา: https://beeying.medium.com/support-vector-machine-svm-78d1ce6da765)

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)

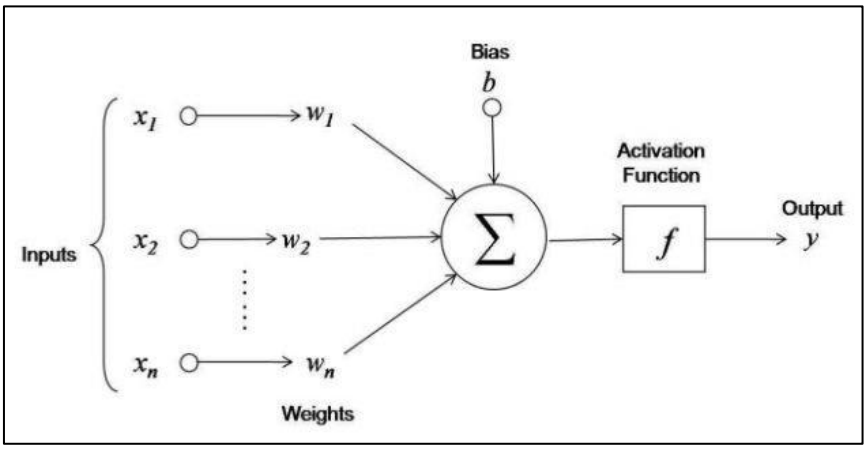
และ

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

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

2.3.5 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks)

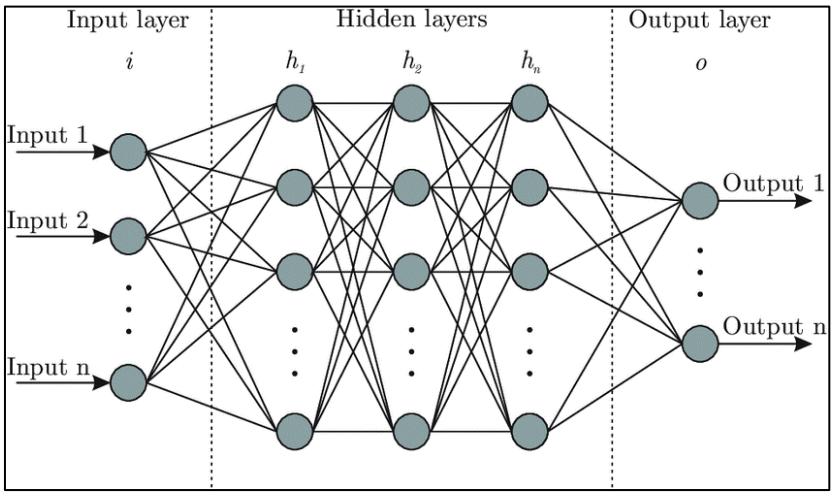
โครงข่ายประสาทเทียมถูกสร้างขึ้นเพื่อการจำลองลักษณะการประมวลผลของระบบประสาทมนุษย์ด้วยแบบจำลองเชิงคณิตศาสตร์และสถิติ (Mathematical and Statistical Model) ซึ่งประกอบด้วยส่วนของการประมวลผลที่เรียกว่า นิวรอน (Neuron) ทุกๆ นิวรอนสามารถมีข้อมูลป้อนเข้า (Input) ได้หลายค่า แต่ข้อมูลส่งออก (Output) มีได้เพียงค่าเดียว 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)

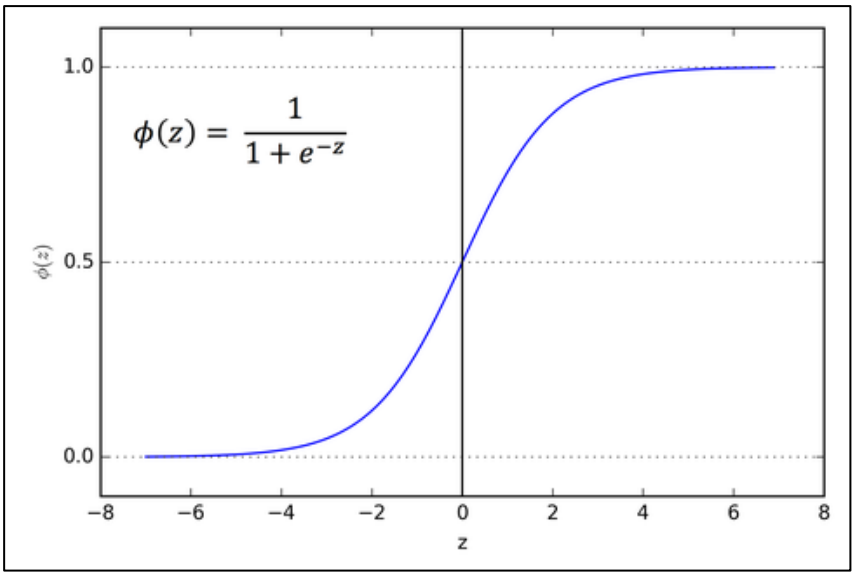
3.การเรียนรู้เชิงบังคับ (Reinforcement Learning)

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) การประเมินผล (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

บทที่ 3

วิธีดำเนินงาน

โครงงานเรื่อง การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ โดยการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining) โดยทำการสร้างโมเดล 4 แบบ คือ 1.เทคนิคการถดถอยเชิงเส้น (linear regression) 2.เทคนิคเกาส์เซียน (Gussian Process) 3.เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression) 4.เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks) และนำมาเปรียบเทียบประสิทธิภาพของโมเดล โดยใช้ค่า RMSE และ MAE เป็นตัวเปรียบเทียบ จากนั้นทำการนำโมเดลที่ได้ไปทำการพัฒนาเว็บไซต์ สำหรับทำการทำนาย และแสดงผล

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ที่มีการพัฒนาเป็น Work flow มาตรฐานสำหรับการทำเหมืองข้อมูลประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม หัวหน้าโครงการหรือที่ปรึกษาด้านการวางระบบวิเคราะห์ข้อมูล จะต้องทำการสัมภาษณ์หรือรับฟังปัญหา/ความต้องการจากผู้บริหารองค์กรและหน่วยงานต่างๆ ที่จะนำผลการวิเคราะห์ข้อมูลไปใช้ประโยชน์ โดยความต้องการทั้งหมดจะนำมาจัดลำดับความสำคัญ และกำหนดวัตถุประสงค์ที่จะนำไปสู่รูปแบบการวิเคราะห์ข้อมูลขององค์กร โดยการวิเคราะห์ข้อมูลในประเด็นนี้ ซึ่งมีข้อมูลมาก ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว

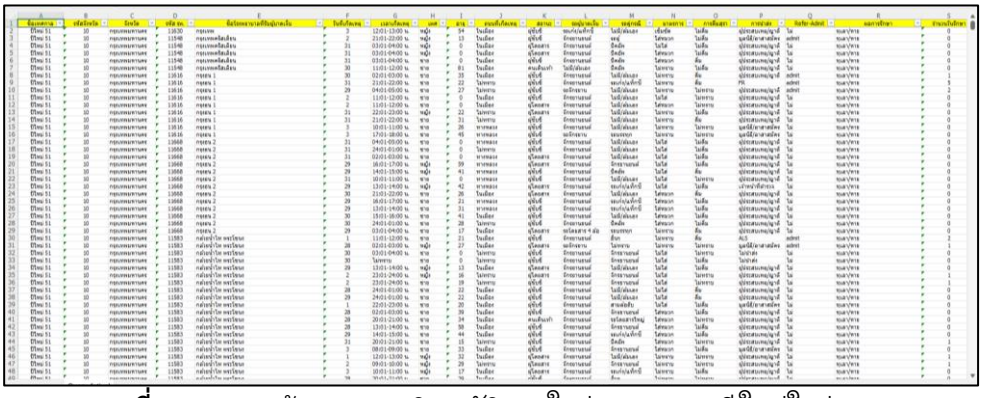
3.1.2 การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูล วันที่เกิดเหตุ เวลาเกิดเหตุ เพศ อายุ ถนนที่เกิดเหตุ สถานะ มาตรการ การดื่มสุรา รถผู้บาดเจ็บ รถคู่กรณี ผลการรักษา เพื่อพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ เป็นข้อมูลที่ได้จากได้รวบรวมข้อมูลจากเว็บไซต์ ศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th/ 2567) ของสถาบันการแพทย์ฉุกเฉินแห่งชาติ ซึ่งเป็นข้อมูลเปิดเผยต่อสาธารณะ

ภาพที่ 3.1 หน้าแสดงผลของเว็บไซต์ข้อมูล

(ที่มา: https://data.go.th/)

ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ โดยข้อมูลที่นำมาวิเคราะห์จะมีข้อมูลตั้งแต่ปีพ.ศ.2551 –2558 ของแต่ละจังหวัดในประเทศไทย จำนวน 77 จังหวัด มีข้อมูลประมาณ 214,951 แถว 19 คอลัมน์



ภาพที่ 3.2 แสดงข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ. 2558

3.1.3 การเตรียมข้อมูล (Data Preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมแล้วเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก ระบบการรับข้อมูลป้อนเข้าสู่ระบบที่ทันสมัยในปัจจุบันจะลดการคีย์ข้อมูลจากคนให้น้อยที่สุด การลดข้อผิดพลาดของข้อมูลได้มากเท่าใดก็จะยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น โดยมีขั้นตอนดังนี้

3.1.2.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลเหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูลผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูลและทำการ Data Cleaning ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ. 2551 – พ.ศ. 2558 ของประเทศไทย โดยตัดคอลัมน์ในส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม

3.1.2.2 การทำความสะอาดข้อมูล (Data Cleaning)

1) จัดการข้อมูลที่มีค่าไม่ทราบ

จัดการกับข้อมูลโดยการลบแถวที่มีค่าไม่ทราบ หรือเรียกว่าค่า Missing (Missing Values) ที่ไม่สามารถสามารถทดแทนได้และมีผลกระทบต่อการตีความผลลัพธ์จากข้อมูลได้ออกจากชุด

2) ข้อมูลที่เป็นไปไม่ได้หรือขัดแย้งกับความเป็นจริง

จัดการกับข้อมูลโดยการลบแถวที่เป็นไปไม่ได้หรือขัดแย้งกับความเป็นจริง ที่ไม่สมเหตุสมผลหรือไม่สอดคล้องกับลักษณะทางสถิติและบริบทของข้อมูลออกจากชุดข้อมูล

3.1.2.3 การแปลงข้อมูล (Data Transformation)

ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) ในข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558 จากการวิเคราะห์พบว่าที่ได้มาเป็นภาษาไทยซึ่งทำให้ไม่รองรับต่อการนำไปวิเคราะห์ ดังนั้นผู้วิเคราะห์ข้อมูลจึงได้ทำการแปลงค่าจากข้อมูลจากตัวอักษรภาษาไทยเป็นภาษาอังกฤษทั้งหมด

3.1.4 การสร้างโมเดล (Modeling)

3.1.4.1 การสร้างโมเดลการทำนายโดยใช้โมเดล Linear regression

โมเดล Linear regression จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

1) เลือกชุดข้อมูลที่นำเข้าสู่โปรแกรม Rapid Miner

2) ใช้ Parameter Nominal to Numerical ประเภท unique integers

3) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%

4) สร้างโมเดล Linear regression โดยใช้ Parameter Linear regression

3.1.4.2 การสร้างโมเดลการทำนายโดยใช้โมเดล Gussian Process

โมเดล Gussian Process จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

1) เลือกชุดข้อมูลที่นำเข้าสู่โปรแกรม Rapid Miner

2) ใช้ Parameter Nominal to Numerical ประเภท unique integers

3) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%

4) สร้างโมเดล Gussian Process โดยใช้ Parameter Gussian Process

3.1.4.3 การสร้างโมเดลการทำนายโดยใช้โมเดล Support Vector Machine Regression

โมเดล Support Vector Machine Regression จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

1) เลือกชุดข้อมูลที่นำเข้าสู่โปรแกรม Rapid Miner

2) ใช้ Parameter Nominal to Numerical ประเภท unique integers

3) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%

4) สร้างโมเดล Support Vector Machine Regression โดยใช้ Parameter Support Vector Machine Regression

3.1.4.2 การสร้างโมเดลการทำนายโดยใช้โมเดล Artificial Neural Networks

โมเดล Artificial Neural Networks จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

1) เลือกชุดข้อมูลที่นำเข้าสู่โปรแกรม Rapid Miner

2) ใช้ Parameter Nominal to Numerical ประเภท unique integers

3) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%

4) สร้างโมเดล Artificial Neural Networks โดยใช้ Parameter Neural Net

3.1.5 การวัดประสิทธิภาพของโมเดล (Evaluation)

3.5.1.1 การวัดประสิทธิภาพของโมเดล

การวัดประสิทธิภาพของโมเดล linear Regression โมเดล Gussian Process โมเดล Support Vector Machine Regression และโมเดล Artificial Neural Network

1) วัดผลประสิทธิภาพ โดยคำนวณค่า RMSE และ MAE

3.5.1.2 การเปรียบเทียบประสิทธิภาพของโมเดล

การเปรียบเทียบประสิทธิภาพระหว่าง โมเดล linear Regression โมเดล Gussian Process โมเดล Support Vector Machine Regression และโมเดล Artificial Neural Networks

1) เปรียบเทียบระหว่างทั้ง 4 โมเดล โดยทำการเลือกโมเดลที่มีค่าน่าพึงพอใจที่สุด

ตารางที่ 3.1 ตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดล

	MAE	RMSE
Linear Regression		
Gussian Process		
SVMR		
ANN		

3.1.6 การนำโมเดลไปใช้งานจริง (Deployment) ในการนำโมเดลที่ผ่านการพัฒนาและเปรียบเทียบมาใช้งานจริง มีการเลือกโมเดลที่เหมาะสมจากการวิเคราะห์โมเดลด้วยโปรแกรม RapidMiner ได้ทำการเปรียบเทียบค่าความคลาดเคลื่อน (Error) ของโมเดลทั้งหมดและเลือกโมเดลที่มีค่าความคลาดเคลื่อนต่ำที่สุด โดยใช้เกณฑ์ Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) โมเดลที่ได้รับการเลือกจะถูกนำมาใช้ในการพยากรณ์ข้อมูลในอนาคต โดยอาศัยข้อมูลจากชุดข้อมูลใหม่ เพื่อสร้างผลลัพธ์ที่มีความแม่นยำและสอดคล้องกับเป้าหมายของงานที่กำหนดไว้ และนำเสนอผลลัพธ์จากการพยากรณ์ด้วยโมเดลจะถูกนำมาแสดงในรูปแบบ กราฟเส้น (Line Chart) บนเว็บไซต์ โดยใช้ภาษา PHP และ HTML

3.2 การวิเคราะห์และออกแบบระบบ

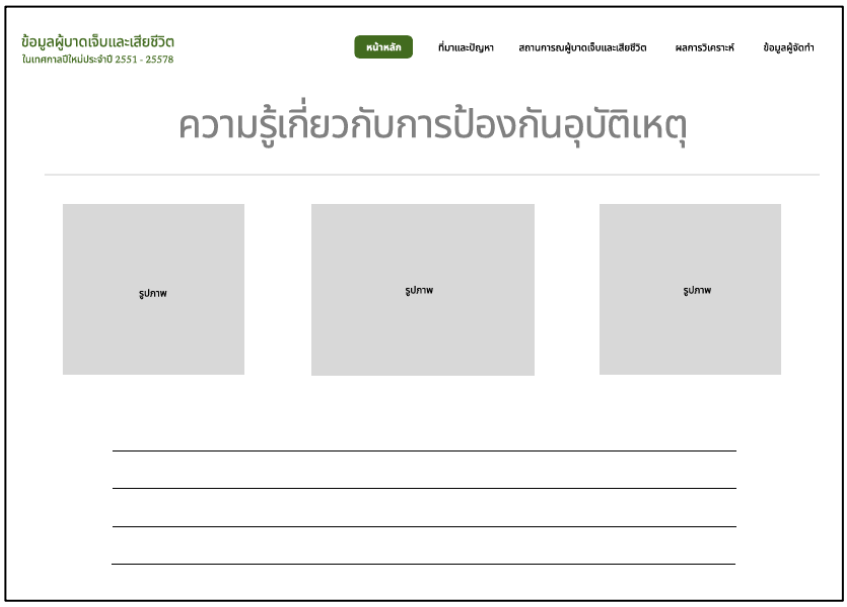
3.2.1 การวิเคราะห์และออกแบบฐานข้อมูล

ระบบจัดการข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ได้มีระบบมีฟังก์ชันหลักต่างๆ ได้แก่การแสดงจำนวนข้อมูลทั้งหมด การเพิ่มข้อมูล แก้ไข ลบ ค้นหา การกรองข้อมูลปีและจังหวัด นอกจากนี้ ยังมีระบบล็อกอินเพื่อกำหนดสิทธิ์การเข้าถึงของผู้ใช้งาน ซึ่งช่วยให้สามารถจัดการข้อมูลได้อย่างเป็นระบบและปลอดภัย

ข้อมูลพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ระบบต้องมีการจัดเก็บข้อมูลผู้ใช้งานที่สามารถล็อกอินเข้าสู่ระบบเพื่อกำหนดสิทธิ์ในการเข้าถึงและจัดการข้อมูล โดยสิทธิ์การใช้งานของระบบแบ่งออกเป็น 2 ระดับหลัก ได้แก่ ผู้ดูแลระบบ (Admin) และผู้ใช้งานทั่วไป (User)

3.2.2 การออกแบบเว็บไซต์

1) หน้าแรกของเว็บไซต์แสดงเมนูต่างๆของหน้าเว็บและเนื้อหาที่เกี่ยวกับข้อมูลความรู้เกี่ยวกับการป้องกันอุบัติเหตุ



ภาพที่ 3.3 แสดงโครงร่างการออกแบบหน้าแรกของเว็บไซต์

2) หน้าแสดงผลข้อมูลที่มาของการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558

ภาพที่ 3.4 แสดงผลข้อมูลที่มาและความสำคัญ

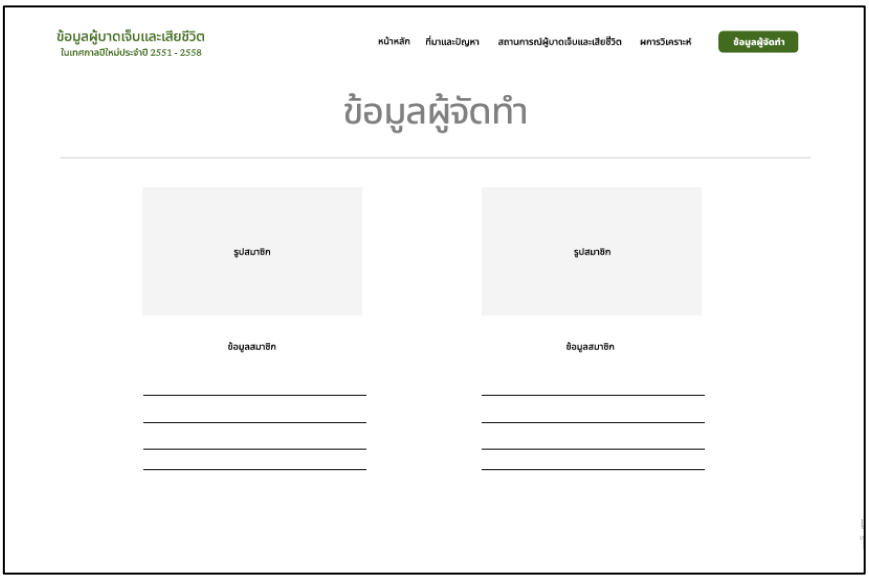
3) หน้าแสดงผลสถานการณ์ผู้บาดเจ็บและเสียชีวิตของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558

ภาพที่ 3.5 แสดงผลสถานการณ์ผู้บาดเจ็บและเสียชีวิตของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558

4) หน้าแสดงผล Dashboard ของข้อมูลการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ. 2558

ภาพที่ 3.6 แสดงผล Dashboard ของข้อมูลการวิเคราะห์ข้อมูลการเกิด อุบัติเหตุในช่วงเทศกาลปีใหม่

5) หน้าแสดงผลข้อมูลผู้จัดทำ



ภาพที่ 3.7 แสดงผลข้อมูลผู้จัดทำ

บทที่ 4
ผลการดำเนินงาน

ผลการดำเนินงาน ของโครงงานเรื่อง การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ โดยการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining)

4.1 ผลการวิเคราะห์ข้อมูลด้วย CRISP-DM

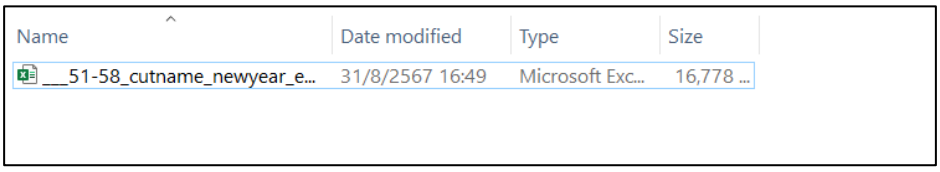
4.1.1 ผลกระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding) กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวมเป็น ขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ ข้อมูลจะถูกจัดเก็บโดยศูนย์กลางข้อมูลเปิดภาครัฐของประเทศไทย (Government Open Data Center – GDC) ดำเนินการโดยสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) หรือ สพร. หน่วยงานราชการ ซึ่งข้อมูลจะถูกจัดเก็บไว้จากทั้งประเทศโดยแยกออกเป็นจังหวัดต่าง ๆ

4.1.2 ผลการทำความเข้าใจข้อมูล (Data Understanding) ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูล วันที่เกิดเหตุ เวลาเกิดเหตุ เพศ อายุ ถนนที่เกิดเหตุ สถานะ มาตรการ การดื่มสุรา รถผู้บาดเจ็บ รถคู่กรณี ผลการรักษา เพื่อพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ เป็นข้อมูลที่ได้จากได้รวบรวมข้อมูลจากเว็บไซต์ ศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th/ 2567) ของสถาบันการแพทย์ฉุกเฉินแห่งชาติ ซึ่งเป็นข้อมูลเปิดเผยต่อสาธารณะ

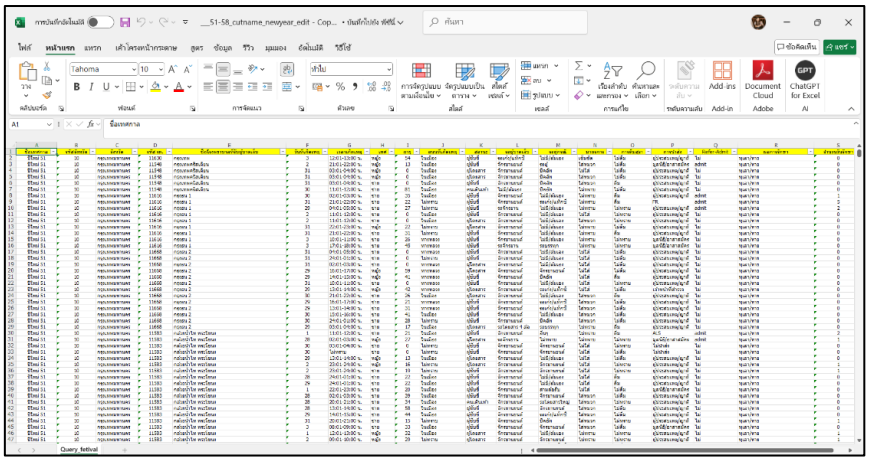
ภาพที่ 4.1 หน้าแสดงผลของเว็บไซต์ข้อมูล

(ที่มา: https://data.go.th/)

โดยเลือกว่า จะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้ ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ตั้งแต่ปี 2551 – 2558 โดยมีข้อมูลรวบรวมไว้ในไฟล์เดียว



ภาพที่ 4.2 การเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ที่ได้มาจากศูนย์กลางข้อมูลเปิดภาครัฐของประเทศไทย

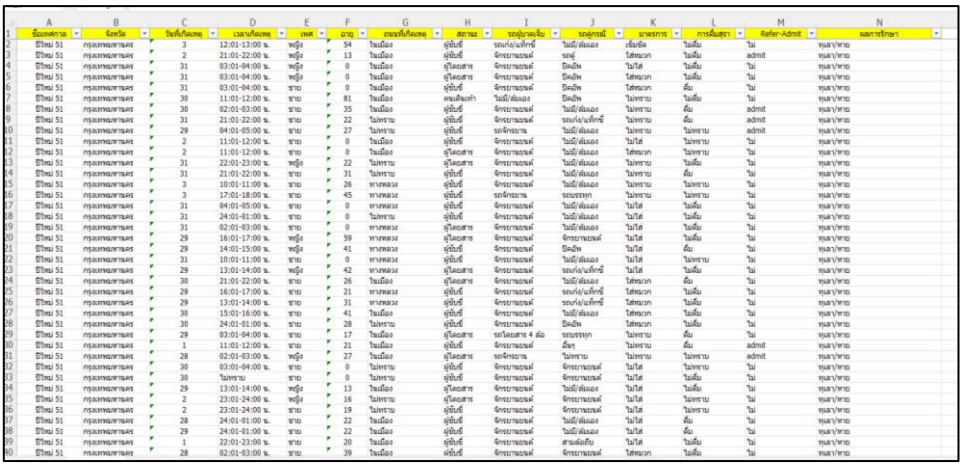


ภาพที่ 4.3 ข้อมูลเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ. 2558

ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ โดยข้อมูลที่นำมาวิเคราะห์จะมีข้อมูลตั้งแต่ปีพ.ศ.2551 –2558 ของแต่ละจังหวัดในประเทศไทย จำนวน 77 จังหวัด มีข้อมูลประมาณ 214,951 แถว 19 คอลัมน์

4.1.3 ผลการเตรียมข้อมูล (Data Preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมแล้วเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก ระบบการรับข้อมูลป้อนเข้าสู่ระบบที่ทันสมัยในปัจจุบันจะลดการคีย์ข้อมูลจากคนให้น้อยที่สุด การลดข้อผิดพลาดของข้อมูลได้มากเท่าใดก็จะยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น โดยมีขั้นตอนดังนี้

4.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลเหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูลผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ. 2558 ของประเทศไทย โดยตัดคอลัมน์ในส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 14 แอทริบิวท์ ได้แก่ year (ปี), Province (จังหวัด), Date (วันที่เกิดเหตุ), time of the incident (เวลาเกิดเหตุ), Gender (เพศ), Age (อายุ),road_location (ถนนที่เกิดเหตุ), Status (สถานะ), Injured car (รถผู้บาดเจ็บ), The car of the party (รถคู่กรณี), measures (มาตรการ), Drinking alcohol (การดื่มสุรา),Treatment effect (ผลการรักษา) และคัดแถวของข้อมูลที่มีข้อมูลไม่สมบูรณ์ออกจากชุดข้อมูล หลังจากที่มีการจัดการข้อมูลแล้วทำให้เหลือข้อมูลทั้งหมด 189,499 แถว



ภาพที่ 4.4 แสดงการคัดเลือกข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

4.1.3.2 การทำความสะอาดข้อมูล (Data Cleaning)

1) จัดการข้อมูลที่มีค่าไม่ทราบ

จัดการกับข้อมูลโดยการลบแถวที่มีค่าไม่ทราบ หรือเรียกว่าค่า Missing (Missing Values) ที่ไม่สามารถสามารถทดแทนได้และมีผลกระทบต่อการตีความผลลัพธ์จากข้อมูลได้ออกจากชุดข้อมูล จำนวน 7 แอทริบิวท์ ได้แก่ time of the incident (เวลาเกิดเหตุ) พบค่าไม่ทราบจำนวน 1,331 แถว,road_location (ถนนที่เกิดเหตุ) พบค่าไม่ทราบจำนวน 5,981 แถว, Status (สถานะ),พบค่าไม่ทราบจำนวน 1,471 แถว, Injured car (รถผู้บาดเจ็บ) พบค่าไม่ทราบจำนวน 713 แถว, The car of the party (รถคู่กรณี) พบค่าไม่ทราบจำนวน 5,204 แถว, measures (มาตรการ) พบค่าไม่ทราบจำนวน 10,040 แถว และ Drinking alcohol (การดื่มสุรา) พบค่าไม่ทราบจำนวน 7,128 แถว พบข้อมูลที่มีค่าไม่ทราบทั้งหมด 31,871 แถว

2) ข้อมูลที่เป็นไปไม่ได้หรือขัดแย้งกับความเป็นจริง

จัดการกับข้อมูลโดยการลบแถวที่เป็นไปไม่ได้หรือขัดแย้งกับความเป็นจริง ที่ไม่สมเหตุสมผลหรือไม่สอดคล้องกับลักษณะทางสถิติและบริบทของข้อมูลออกจากชุดข้อมูล ทางผู้วิเคราะห์ข้อมูลได้ดำเนินการดังนี้

– Age (อายุ) 0–14 ปี ตามกฎหมายของประเทศไทยที่อ้างอิงจากการขอรับใบขับขี่ กำหนดเกณฑ์อายุผู้ขอรับใบขับขี่รถจักรยานยนต์ส่วนบุคคลต้องมีอายุไม่ต่ำกว่า 15 ปีบริบูรณ์ พบทั้งหมด 11,325 แถว

– Age (อายุ) 0–17 ปี ตามกฎหมายของประเทศไทยที่อ้างอิงจากการขอรับใบขับขี่ กำหนดเกณฑ์อายุผู้ขอรับใบขับขี่รถยนต์ส่วนบุคคลต้องมีอายุไม่ต่ำกว่า 18 ปีบริบูรณ์ พบทั้งหมด 145 แถว

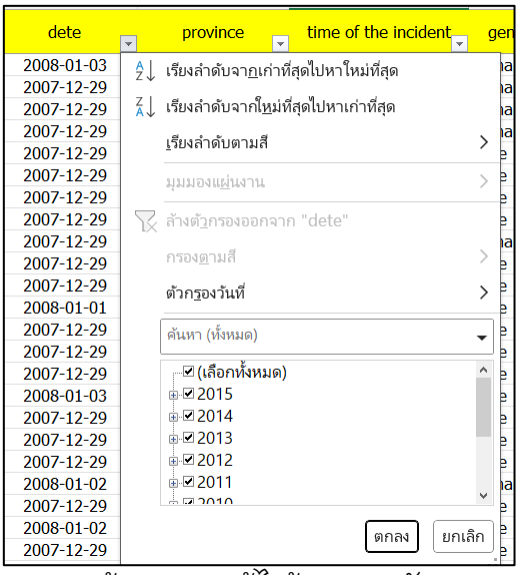
– Age (อายุ) 71–99 ปี กฎหมายไทยไม่ได้กำหนดอายุสูงสุดที่ชัดเจนสำหรับการขับขี่ของผู้สูงอายุ แต่มีการระบุข้อกำหนดเพิ่มเติมสำหรับผู้สูงวัย ผู้สูงอายุที่มีอายุ 70 ปีขึ้นไปมักถูกพิจารณาเป็นกลุ่มเสี่ยงในการขับขี่รถยนต์ เนื่องจากสมรรถภาพร่างกาย เช่น การมองเห็น การตอบสนอง และสมาธิ อาจลดลง ส่งผลต่อความปลอดภัยในการขับรถ ตามข้อมูลจาก TDRI ระบุว่าผู้สูงอายุควรได้รับการทดสอบสมรรถภาพใหม่เพื่อประเมินความสามารถในการขับขี่อย่างปลอดภัย โดยเฉพาะในกลุ่มที่มีใบขับขี่ตลอดชีพ ซึ่งอาจไม่ได้มีการตรวจสอบสมรรถภาพอย่างต่อเนื่องหลังอายุ 70 ปี (ภัทรพิศาล, 2564) พบทั้งหมด 695 แถว

– Age (อายุ) 0 ปี ข้อมูลอายุที่มีค่า 0 ปี ถือเป็นข้อมูลที่ไม่สมเหตุสมผล เนื่องจากบุคคลที่มีอายุ 0 ปี (แรกเกิด) ยังไม่สามารถเดินเท้าหรือเดินทางได้ด้วยตนเอง พบในคอลัมน์ Status (สถานะ) พบทั้งหมด 6 แถว

– Drinking alcohol (การดื่มสุรา) กฎหมายเกี่ยวกับอายุผู้ดื่มสุราบุคคลที่สามารถซื้อและดื่มสุราได้ต้องมีAge (อายุ) ไม่ต่ำกว่า 20 ปีบริบูรณ์อ้างอิงจากพระราชบัญญัติควบคุมเครื่องดื่มแอลกอฮอล์ พ.ศ. 2551 (มาตรา 29) พบทั้งหมด 10,089 แถว หลังจากที่มีการทำความสะอาดข้อมูลแล้วทำให้เหลือข้อมูลทั้งหมด 159,288 แถว

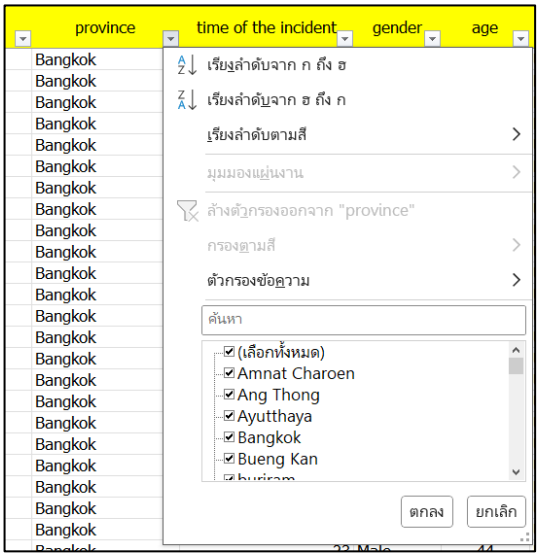
4.1.3.3 การแปลงข้อมูล (Data Transformation)ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) ในข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558 จากการวิเคราะห์พบว่าที่ได้มาเป็นตัวอักษรและตัวเลขซึ่งทำให้ไม่รองรับต่อการนำไปวิเคราะห์ ดังนั้นผู้วิเคราะห์ข้อมูลจึงได้ทำการแปลงค่าจากข้อมูลจากตัวอักษรภาษาไทยเป็นภาษาอังกฤษและตัวเลขทั้งหมด ทางผู้วิเคราะห์ข้อมูลได้ดำเนินการดังนี้

1) ปรับรูปแบบวันที่ ปรับปรุงข้อมูลวันที่ให้เป็นรูปแบบที่สอดคล้องกับการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่และเหมาะสมสำหรับการวิเคราะห์และการประมวลผล โดยรวมข้อมูลจากคอลัมน์ Date (วันที่เกิดเหตุ) และ year (ปี) เข้าด้วยกันและจัดให้อยู่ในตัวแปรวันที่ในรูปแบบมาตรฐาน YYYY–MM–DD (ปี/เดือน/วัน) ดังแสดงในภาพ 4.5



ภาพที่ 4.5 แสดงข้อมูลการแก้ไขด้วยการปรับรูปแบบวันที่

2) ข้อมูลของจังหวัด (Province) ทั้ง 77 จังหวัด ข้อมูลยังเป็นภาษาไทยจึงได้ทำการแปลงค่าให้เป็นภาษาอังกฤษ ดังแสดงในภาพ 4.6

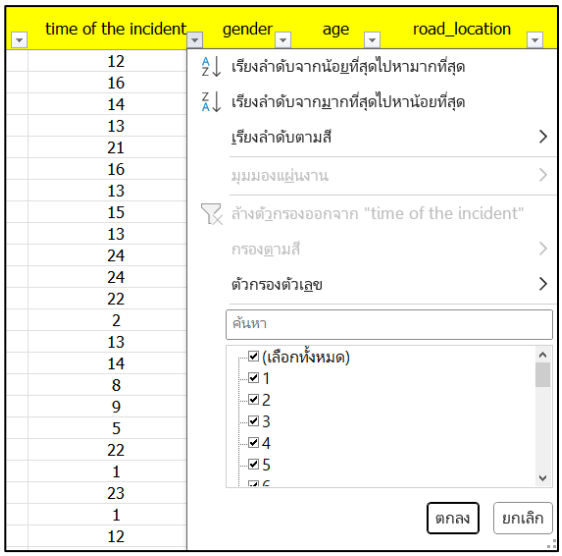


ภาพที่ 4.6 แสดงข้อมูลการแทนค่าด้วยจังหวัด 77 จังหวัดให้เป็นภาษาอังกฤษ

3) ข้อมูลเวลาเกิดเหตุ (time of the incident) ข้อมูลเป็นเวลาของประเทศไทย จึงได้ทำการแปลงค่าให้เป็นตัวเลข ดังแสดงในตาราง 4.1 และภาพ 4.7

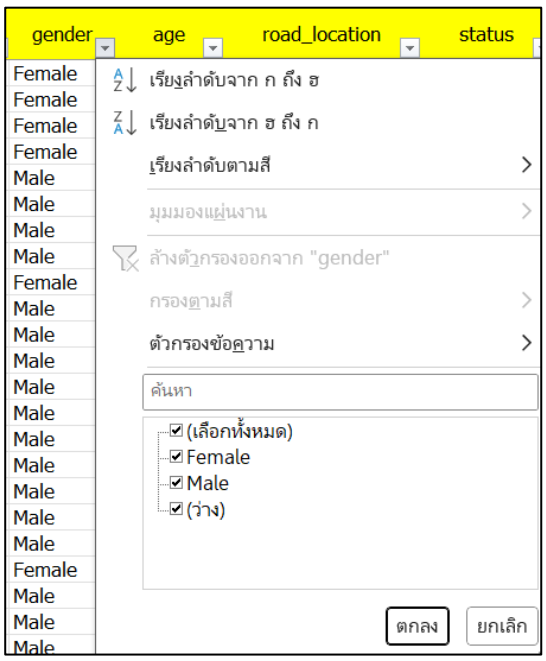
ตารางที่ 4.1 ตารางแปลงค่าข้อมูลเวลาเกิดเหตุ

เวลาเกิดเหตุ (time of the incident)			
01:01–02:00 น.	1	13:01–14:00 น.	13
02:01–03:00 น.	2	14:01–15:00 น.	14
03:01–04:00 น.	3	15:01–16:00 น.	15
04:01–05:00 น.	4	16:01–17:00 น.	16
05:01–06:00 น.	5	17:01–18:00 น.	17
06:01–07:00 น.	6	18:01–19:00 น.	18
07:01–08:00 น.	7	19:01–20:00 น.	19
08:01–09:00 น.	8	20:01–21:00 น.	20
09:01–10:00 น.	9	21:01–22:00 น.	21
10:01–11:00 น.	10	22:01–23:00 น.	22
11:01–12:00 น.	11	23:01–24:00 น.	23
12:01–13:00 น.	12	24:01–01:00 น.	24



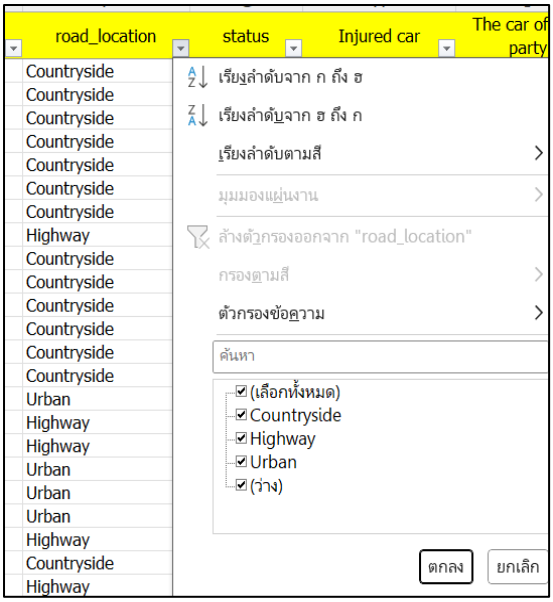
ภาพที่ 4.7 แสดงข้อมูลการแทนค่าข้อมูลเวลาเกิดเหตุให้เป็นตัวเลข

4) ข้อมูลของเพศ (Gender) ข้อมูลยังเป็นภาษาไทยจึงได้ทำการแปลงค่าให้เป็นภาษาอังกฤษ ดังแสดงในภาพ 4.8



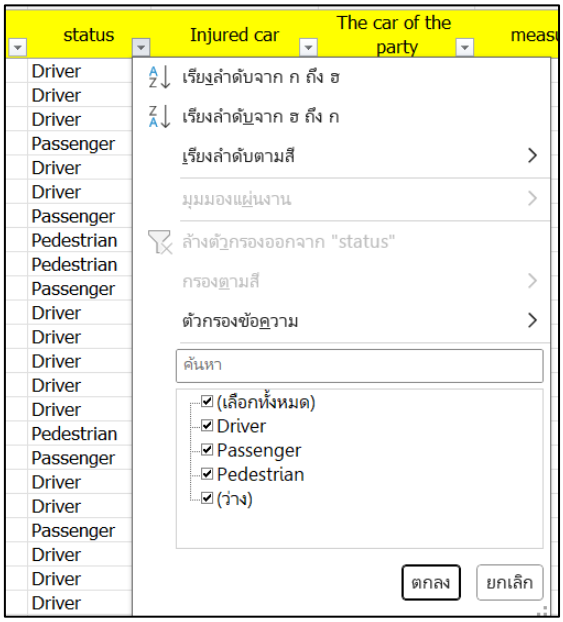
ภาพที่ 4.8 แสดงข้อมูลการแทนค่าข้อมูลของเพศให้เป็นภาษาอังกฤษ

5) ข้อมูลของถนนที่เกิดเหตุ (road_location) ข้อมูลยังเป็นภาษาไทยจึงได้ทำการแปลงค่าให้เป็นภาษาอังกฤษ ดังแสดงในภาพ 4.9



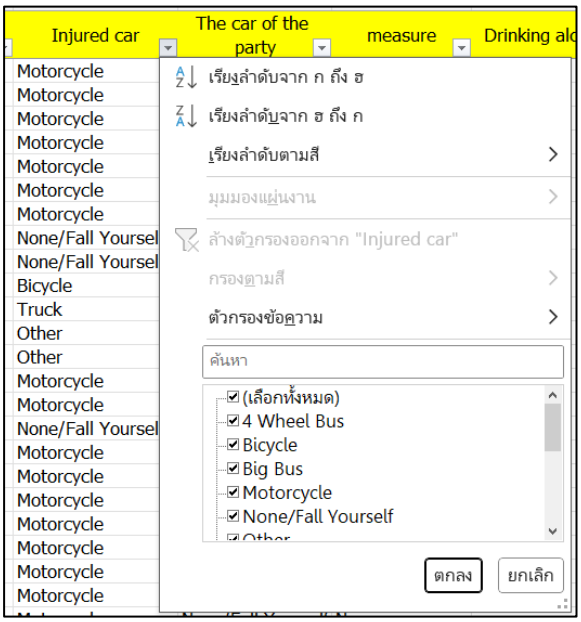
ภาพที่ 4.9 แสดงข้อมูลการแทนค่าข้อมูลของถนนที่เกิดเหตุให้เป็นภาษาอังกฤษ

6) ข้อมูลของสถานะ (Status) ข้อมูลยังเป็นภาษาไทยจึงได้ทำการแปลงค่าให้เป็นตัวเลข ดังแสดงในภาพ 4.10



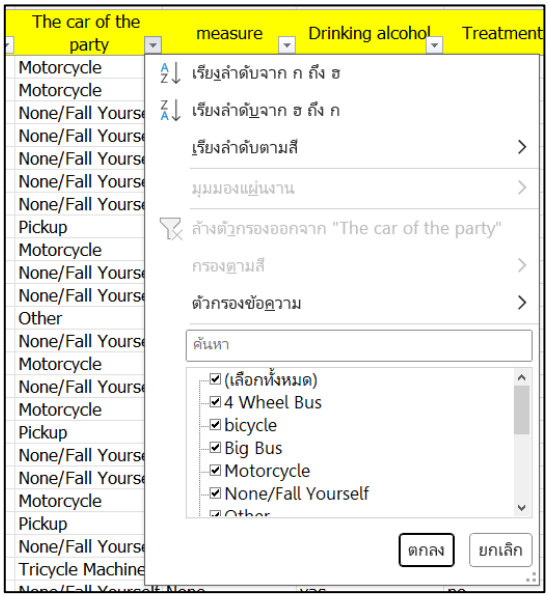
ภาพที่ 4.10 แสดงข้อมูลการแทนค่าข้อมูลของสถานะให้เป็นภาษาอังกฤษ

7) ข้อมูลของรถผู้บาดเจ็บ (Injured car) ข้อมูลยังเป็นภาษาไทยจึงได้ทำการแปลงค่าให้เป็นภาษาอังกฤษ ดังแสดงในภาพ 4.11



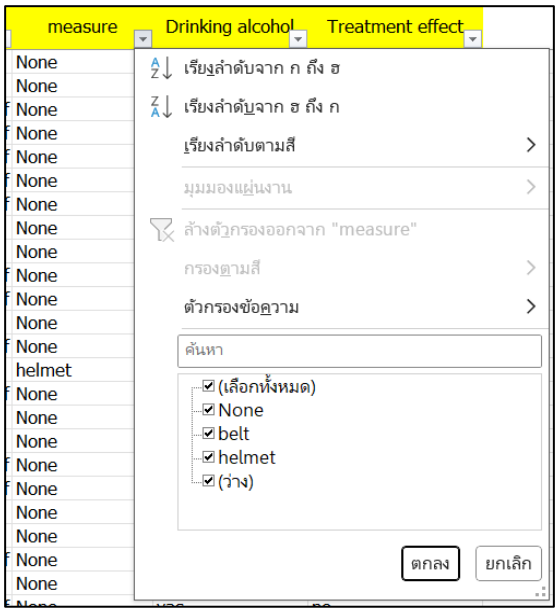
ภาพที่ 4.11 แสดงข้อมูลการแทนค่าข้อมูลของรถผู้บาดเจ็บให้เป็นภาษาอังกฤษ

8) ข้อมูลในส่วนของรถคู่กรณี (The car of the party) ข้อมูลยังเป็นภาษาไทย จึงได้ทำการ แปลงค่าให้เป็นภาษาอังกฤษ ดังแสดงในภาพ 4.12



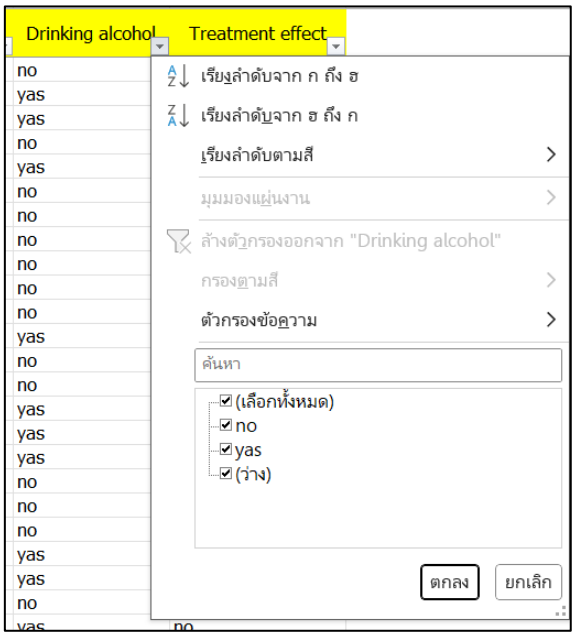
ภาพที่ 4.12 แสดงข้อมูลการแทนค่าข้อมูลของรถคู่กรณีให้เป็นภาษาอังกฤษ

9) ข้อมูลในส่วนของมาตรการ (measure) ข้อมูลยังเป็นภาษาไทยจึงได้ทำการแปลงค่าให้เป็นภาษาอังกฤษ ดังแสดงในตาราง 4.13



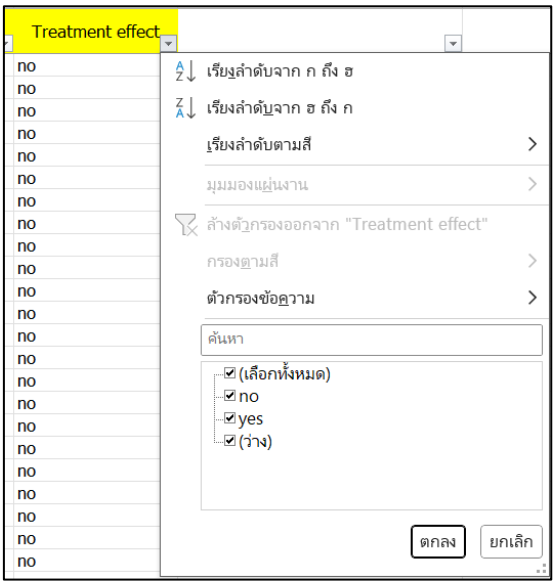
ภาพที่ 4.13 แสดงข้อมูลการแทนค่าข้อมูลของมาตรการให้เป็นภาษาอังกฤษ

10) ข้อมูลในส่วนของการดื่มสุรา (Drinking alcohol) ข้อมูลยังเป็นภาษาไทย จึงได้ทำการแปลงค่าให้เป็นภาษาอังกฤษ โดยแทนค่า ดื่ม เป็น yes และ ไม่ดื่ม เป็น no ดังแสดงในภาพ 4.14



ภาพที่ 4.14 แสดงข้อมูลการแทนค่าข้อมูลของการดื่มสุราให้เป็นภาษาอังกฤษ

11) ข้อมูลในส่วนของผลการรักษา (Treatment effect) ข้อมูลยังเป็นภาษาไทย จึงได้ทำการแปลงค่าให้เป็นภาษาอังกฤษ โดยแทนค่า เสียชีวิตเป็น yes และ ทุเลา/หาย เป็น no ดังแสดงในภาพ 4.15

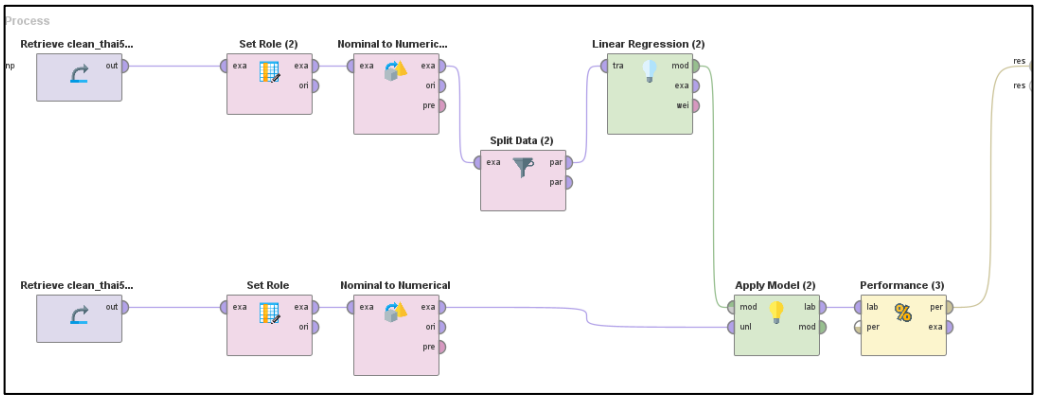


ภาพที่ 4.15 แสดงข้อมูลการแทนค่าข้อมูลของการผลการรักษาให้เป็นภาษาอังกฤษ

4.1.4 ผลการสร้างโมเดล (Modeling)

4.1.4.1 ผลการสร้างโมเดลการทำนายโดยใช้โมเดล Linear regression

โมเดล Linear regression จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

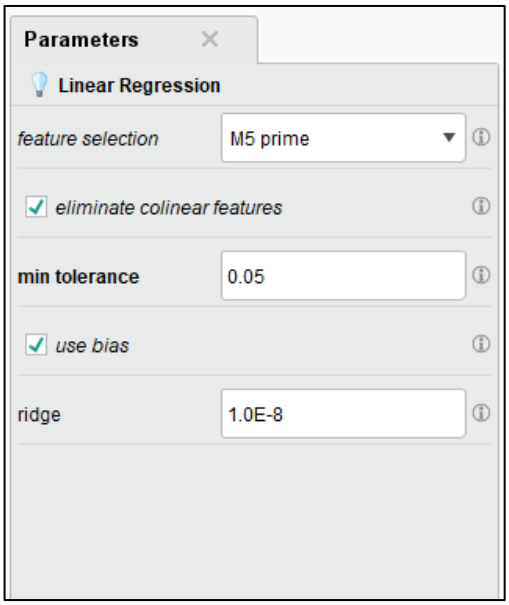


ภาพที่ 4.16 แสดงกระบวนการประมวลผลและสร้างโมเดลเทคนิคการถดถอยเชิงเส้น (linear regression)

1) ทำการดึงข้อมูลที่ได้โหลดไว้แล้วด้วยตัวดำเนินการ retrieve
2) ใช้ Parameter Nominal to Numerical ประเภท unique integers

3) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%

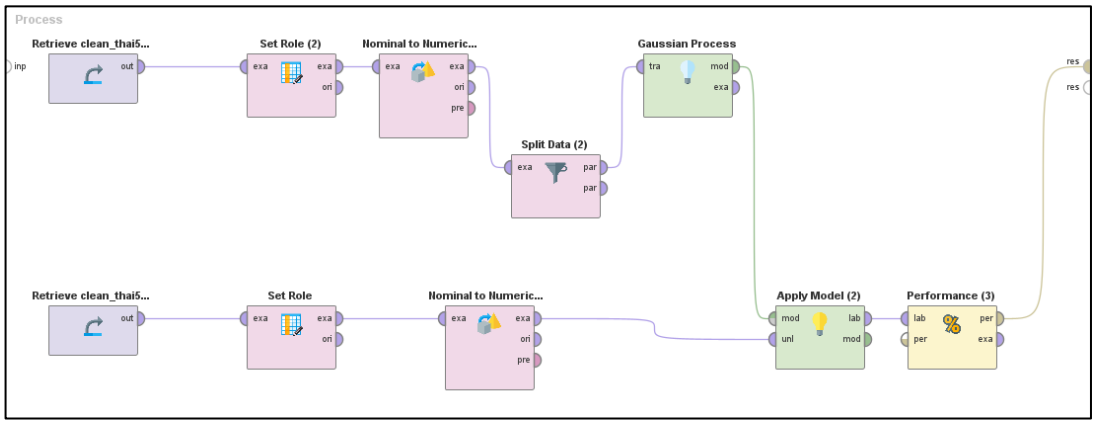
4) ทำการสร้างโมเดล linear regression โดยทำการตั้งค่าตัวแปรดังภาพที่ 4.17



ภาพที่ 4.17 การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล linear regression

4.1.4.2 ผลการสร้างโมเดลการทำนายโดยใช้โมเดล Gussian Process

โมเดล Gussian Process จะทำการสร้างและวัดผลในโปรแกรม RapidMiner



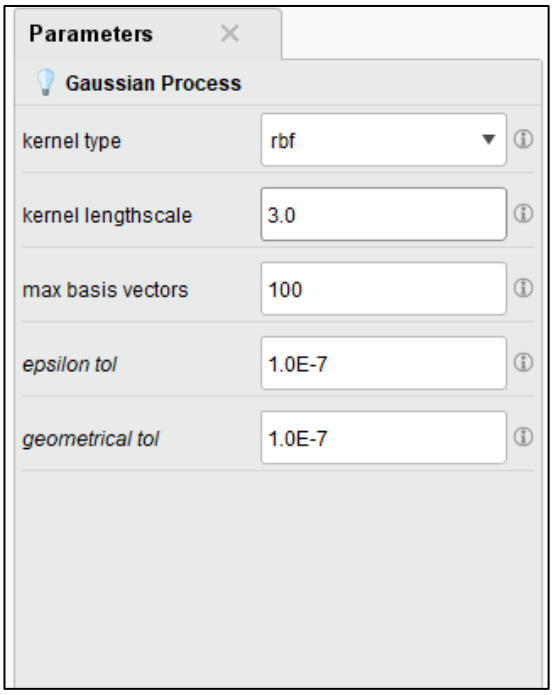
ภาพที่ 4.18 แสดงกระบวนการประมวลผลและสร้างโมเดลเทคนิคเกาส์เซียน (Gussian Process)

1) ทำการดึงข้อมูลที่ได้โหลดไว้แล้วด้วยตัวดำเนินการ retrieve

2) ใช้ Parameter Nominal to Numerical ประเภท unique integers

3) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุด ข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%

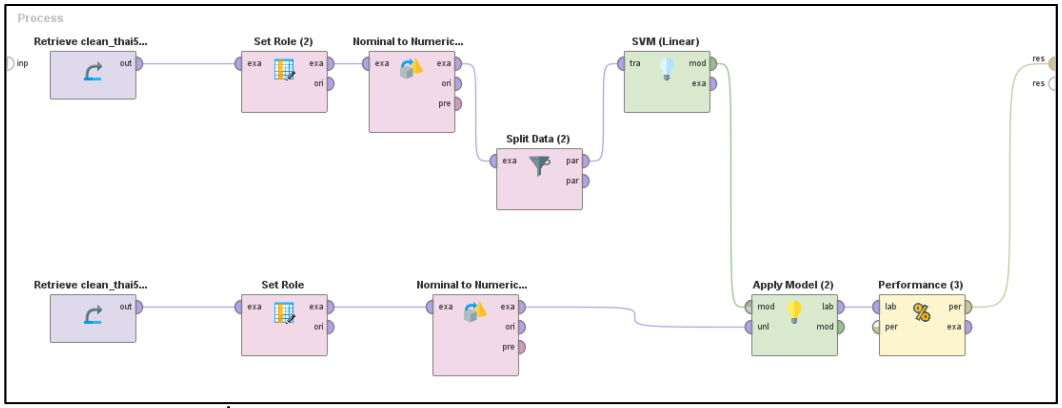
4) ทำการสร้างโมเดล Gussian Process โดยทำการตั้งค่าตัวแปรดังภาพที่ 4.1



ภาพที่ 4.19 การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล Gussian Process

4.1.4.3 ผลการสร้างโมเดลการทำนายโดยใช้โมเดล Support Vector Machine Regression

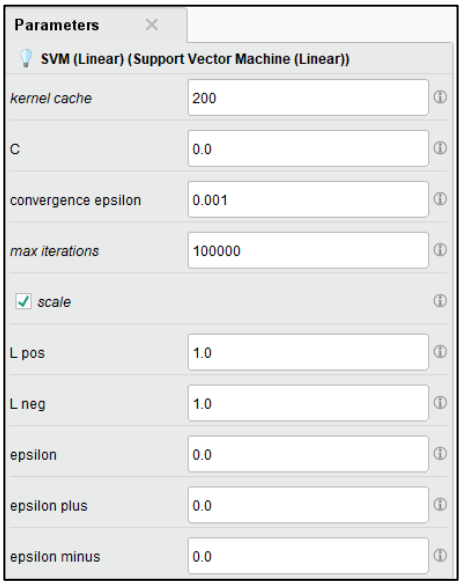
โมเดล Support Vector Machine Regression จะทำการสร้างและวัดผลในโปรแกรม RapidMiner



ภาพที่ 4.20 แสดงกระบวนการประมวลผลและสร้างโมเดล

เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression)

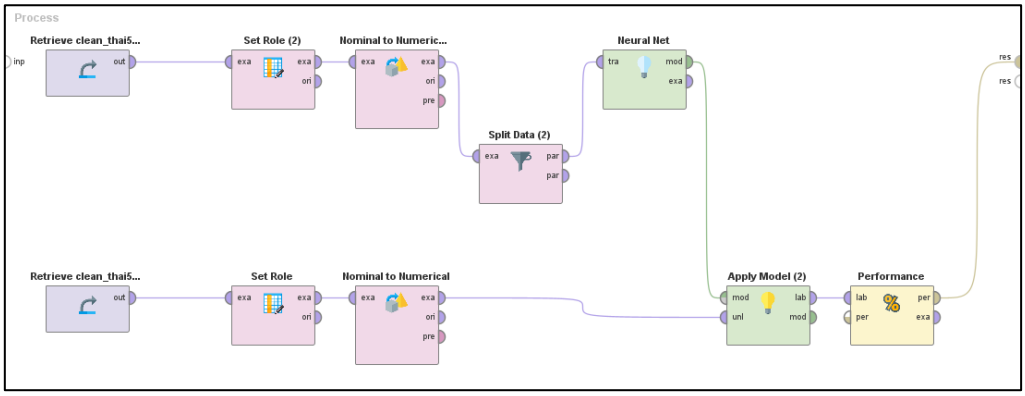
1) ทำการดึงข้อมูลที่ได้โหลดไว้แล้วด้วยตัวดำเนินการ retrieve
2) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%
3) ทำการสร้างโมเดล Support Vector Machine Regression โดยทำการตั้งค่าตัวแปรดังภาพที่ 4.20



ภาพที่ 4.21 การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล SVM

4.4.4 ผลการสร้างโมเดลการทำนายโดยใช้โมเดล Artificial Neural Networks

โมเดล Support Vector Machine Regression จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

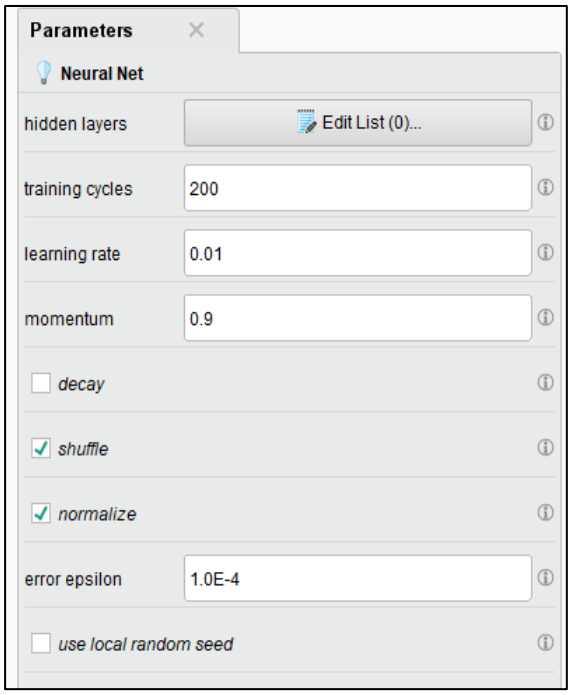


ภาพที่ 3.22 แสดงกระบวนการประมวลผลและสร้างโมเดลเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks)

1) ทำการดึงข้อมูลที่ได้โหลดไว้แล้วด้วยตัวดำเนินการ retrieve
2) ใช้ Parameter Nominal to Numerical ประเภท unique integers

3) แบ่งข้อมูลออกเป็น ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30%

4) ทำการสร้างโมเดล Artificial Neural Networks โดยทำการตั้งค่าตัวแปรดังภาพที่ 4.22



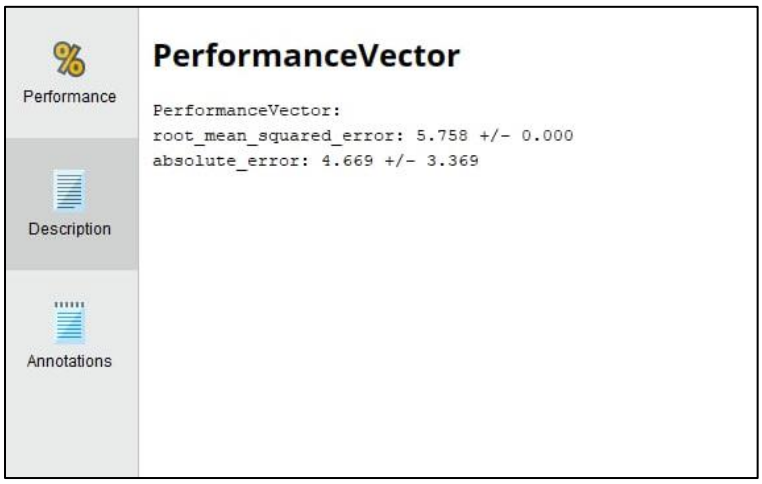
ภาพที่ 4.23 การตั้งค่า Parameter เพื่อทำการวัดผลโมเดล ANN

4.1.5 ผลการวัดประสิทธิภาพของโมเดล (Evaluation)

4.1.5.1 ผลการวัดประสิทธิภาพของโมเดล

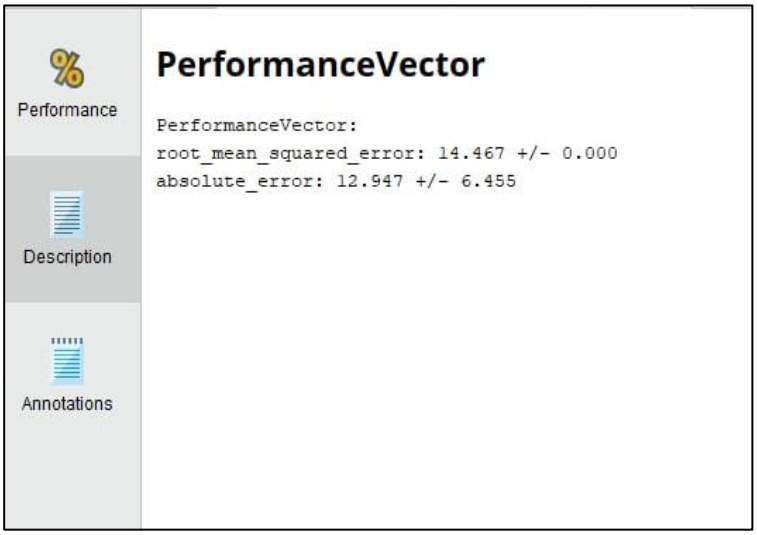
การวัดประสิทธิภาพโดยคำนวณค่า RMSE และ MAE ของโมเดลการถดถอยเชิงเส้น (linear regression) โมเดลเกาส์เซียน (Gussian Process) โมเดลซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression) และโมเดลโครงข่ายประสาทเทียม (Artificial Neural Networks) มีการวัดค่าประสิทธิภาพของแบบจำลอง ด้วยค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ในการแสดงประสิทธิภาพในการทำพยากรณ์ของแต่ละเทคนิค แล้วทำการเลือกเทคนิคที่เหมาะสมกับข้อมูลมาทำการการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ซึ่งส่งผลให้การพยากรณ์มีความแม่นยำสูงสุด

1) ผลการวัดประสิทธิภาพของโมเดล Linear Regression พบว่าค่าของ Mean Absolute Error (MAE) เท่ากับ 4.669 ,Root Mean Square Error (RMSE) เท่ากับ 5.758



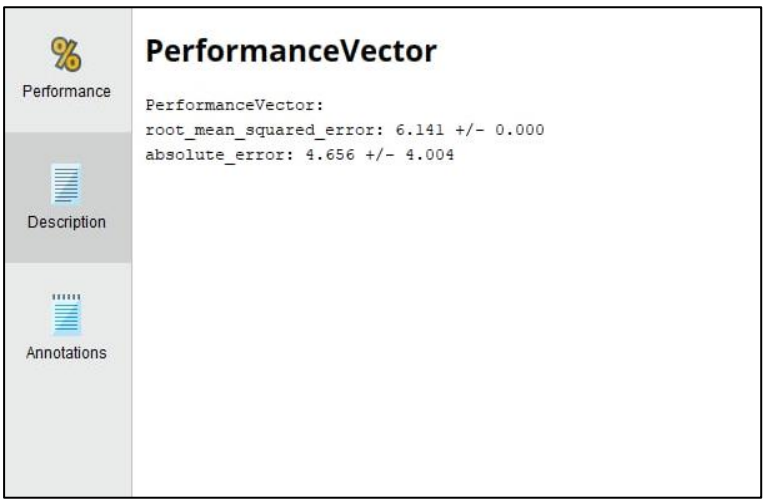
ภาพที่ 4.24 ผลการวัดประสิทธิภาพของโมเดล Linear Regression

3) ผลการวัดประสิทธิภาพของโมเดล Gussian Process พบว่าค่าของ Mean Absolute Error(MAE) เท่ากับ 12.947 ,Root Mean Square Error (RMSE) เท่ากับ 14.467



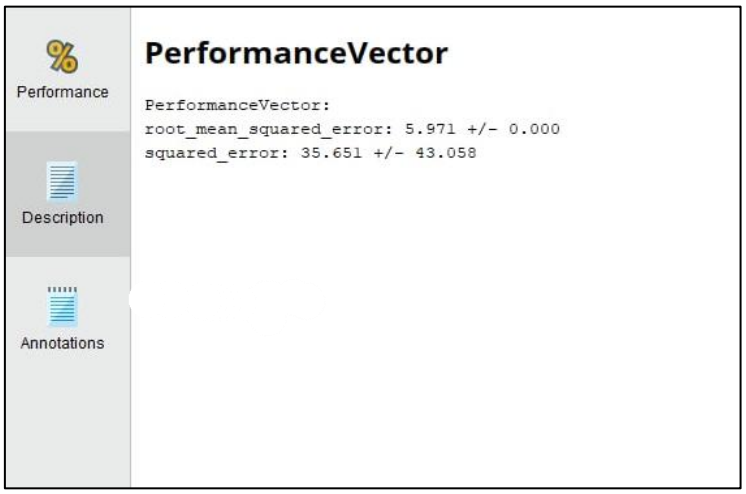
ภาพที่ 4.25 ผลการวัดประสิทธิภาพของโมเดล Gussian Process

3) ผลการวัดประสิทธิภาพของโมเดล Support Vector Machine Regression พบว่าค่าของ Mean Absolute Error(MAE) เท่ากับ 4.656 ,Root Mean Square Error (RMSE) เท่ากับ 6.141



ภาพที่ 4.26 ผลการวัดประสิทธิภาพของโมเดล Support Vector Machine Regression

4) ผลการวัดประสิทธิภาพของโมเดล Artificial Neural Networks พบว่าค่าของ Mean Absolute Error (MAE) เท่ากับ 35.651 ,Root Mean Square Error (RMSE) เท่ากับ 5.971



ภาพที่ 4.27 ผลการวัดประสิทธิภาพของโมเดล Artificial Neural Networks

4.5.2 ผลการเปรียบเทียบประสิทธิภาพของโมเดล

สรุปผลการวิเคราะห์ประสิทธิภาพโมเดลจำนวน 4 โมเดล ได้แก่ โมเดลการถดถอยเชิงเส้น (linear regression) โมเดลเกาส์เซียน (Gussian Process) โมเดลซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression) และโมเดลโครงข่ายประสาทเทียม (Artificial Neural Networks) ใช้วิธีการประเมินผลประสิทธิภาพของโมเดลด้วยตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดลได้ผลการประเมินประสิทธิภาพ ดังแสดงในตาราง 4.2

ตารางที่ 4.2 ผลตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดล

Model	MAE	RMSE
Linear Regression	4.669	5.758
Gussian Process	12.947	14.467
ANN	35.651	5.971
SVMR	4.656	6.141

จากตารางที่ 4.2 พบวาโมเดลที่มีคา MAE นอยที่สุด คือ SVMR ซึ่งมีค่าเท่ากับ 4.656 ในขณะที่โมเดล Linear Regression, Gussian Process, และ ANN มี MAE อยู่ที่ 4.669, 12.947 และ 35.651 ตามลำดับ แตถาเปรียบเทียบประสิทธิภาพของโมเดลดวยคา RMSE พบวา โมเดล Linear Regression มีคา RMSE น้อยที่สุด คือ 5.758 ในขณะที่โมเดล ANN, SVMR และ Gussian Process อยู่ที่ 5.971, 6.141 และ 14.467 ตามลำดับ

เมื่อลองใชโมเดล SVMR และโมเดล Linear Regression โดยใชขอมูล ช่วงเทศกาลปีใหม่ปี 2558 เปนขอมูลเดือน ธันวาคม ปี พ.ศ.2557 ถึง เดือน มกราคมปี พ.ศ.2558 ในการทดสอบซึ่งมีขอมูลจำนวน 18,185 คอลัมน์ ใหการคำนวณคา RMSE และ MAE นอยที่สุด โดยทำกาพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ พิเชษฐ ศิริรัตนไพศาลกุล (2022) แสดงดังตารางที่ 4.3

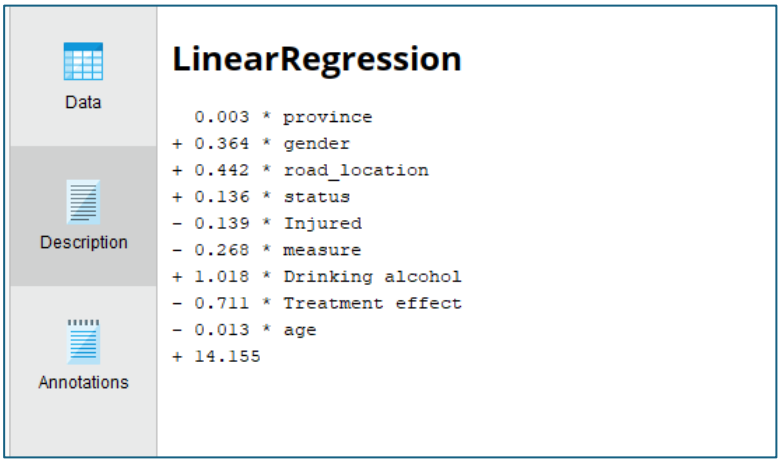
ตารางที่ 4.3 ตารางแสดงคา MAE และ RMSE และเปรียบเทียบกันระหวาง SVMR และ Linear Regression เมื่อใชขอมูลช่วงเทศกาลปีใหม่ปี 2558 ล่าสุด ในการทดสอบ

Model	MAE	RMSE
Linear Regression	3.716	5.142
SVMR	3.908	5.343

จากตารางที่ 4.3 จะเห็นไดวา โมเดล Linear Regression มีประสิทธิภาพดีกวาเนื่องจากคา MAE เทากับ 3.716 มีคานอยกวาโมเดล SVMR และ ค่า RMSE ของโมเดล Linear Regression มีคานอยกวาโมเดล SVMR เท่ากับ 5.142 เชนเดียวกัน จาก

ผลการวิจัยที่ได้สอดคล้องกับผลการศึกษาของ พิชญากร เลค (2565) ได้ศึกษาการเปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องจักร เพื่อทำนายการเกิดอุบัติเหตุ โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น แบบจำลองโครงข่าย ประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และ ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลในจังหวัดที่มีจำนวนการเกิดอุบัติเหตุมากที่สุด ได้แก่ จังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ จากการทดลองพบว่าแบบจำลองที่มีความเหมาะสมในการพยากรณ์การเกิดอุบัติเหตุในประเทศไทย คือแบบจำลองการถดถอยเชิงเส้น มีความเหมาะสมมากที่สุดกับชุดข้อมูลจังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมาและจังหวัดสมุทรปราการ มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับร้อยละ 16.29 19.29 และร้อยละ 19.64

ผลการเปรียบเทียบโมเดล พบว่า โมเดล Linear Regression เป็นโมเดลที่ให้ผลลัพธ์ที่แม่นยำที่สุด เมื่อเปรียบเทียบกับโมเดลพยากรณ์อื่น ๆ ทั้ง 4 โมเดล โดยพิจารณาจากค่าประสิทธิภาพของโมเดล ซึ่งแสดงให้ เห็นว่า Linear Regression สามารถพยากรณ์แนวโน้มของข้อมูลได้อย่างแม่นยำและมีความคลาดเคลื่อนต่ำที่สุด จึงได้สมการดังนี้



ภาพที่ 4.28 แสดงสมการของ Linear Regression ในโปรแกรม ripid miner

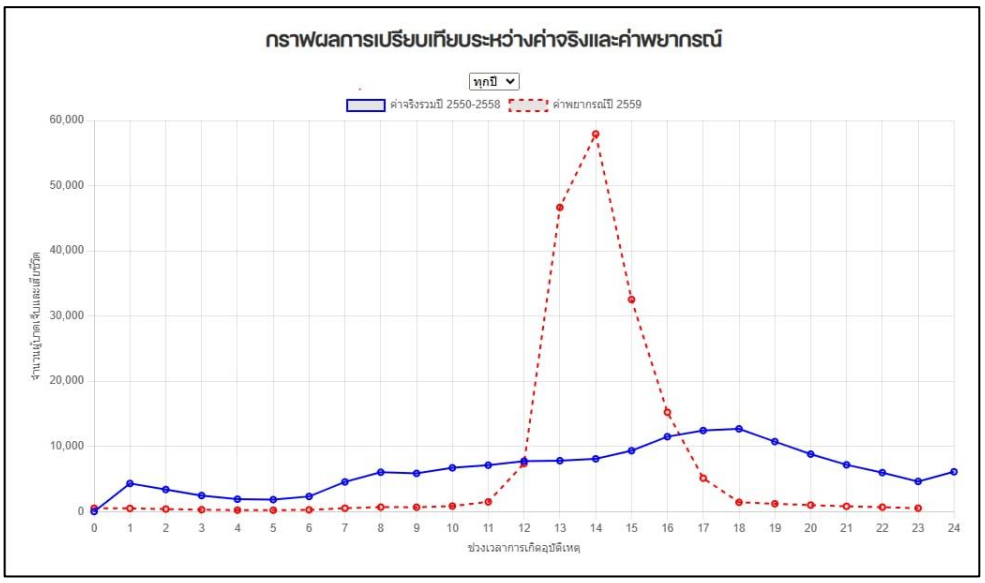
จากภาพที่ 4.1 ซึ่งในสมการเป็นวิธีการพยากรณ์โดยอิงจากตัวแปรอิสระ (Independent Variables) เพื่อคาดการณ์ค่าของตัวแปรตาม (Dependent Variable) ผ่านสมการเชิงเส้นที่กำหนดโดยน้ำหนักหรือค่าสัมประสิทธิ์ (Coefficients) ของแต่ละตัวแปรในสมการ โดยรูปแบบของสมการ มีดังนี้

Y = 14.155 + 0.003 * province + 0.364 * gender + 0.442 * road_location + 0.136 * status – 0.139 * Injured car – 0.268 * measure + 1.018 * Drinking alcohol – 0.711 * Treatment effect – 0.013 * age

ในสมการถดถอยเชิงเส้นนี้ Y แสดงค่าผลลัพธ์ที่คาดการณ์ โดยมีค่าคงที่ (Constant) เท่ากับ 14.155 ซึ่งเป็นจุดเริ่มต้นเมื่อทุกตัวแปรอิสระมีค่าเป็นศูนย์ ตัวแปรอิสระที่ส่งผลต่อค่าของ Y ประกอบด้วย

1. province ตัวแปรบ่งชี้จังหวัดที่มีค่าผลกระทบต่อ Y เท่ากับ 0.003 ต่อ
2. gender ตัวแปรที่แสดงเพศของบุคคลที่มีค่าผลกระทบต่อ Y เท่ากับ 0.364
3. road_location ตัวแปรที่ระบุถนนที่เกิดเหตุที่มีค่าผลกระทบต่อ Y เท่ากับ 0.442
4. status ตัวแปรที่ระบุสถานะของบุคคลที่มีค่าผลกระทบต่อ Y เท่ากับ 0.136
5. Injured car ตัวแปรที่ระบุประเภทของยานพาหนะที่ได้รับบาดเจ็บที่มีค่าผลกระทบต่อ Y เท่ากับ –0.139 ซึ่งเป็นค่าลบที่แสดงผลลด
6. measure ตัวแปรที่แสดงถึงมาตรการความปลอดภัยที่มีผลกระทบต่อ Y เท่ากับ –0.268 ซึ่งแสดงถึงผลลด
7. Drinking alcohol ตัวแปรที่ระบุว่ามีการดื่มแอลกอฮอล์หรือไม่ ที่มีค่าผลกระทบต่อ Y เท่ากับ 1.018 แสดงถึงความสัมพันธ์เชิงบวกที่สำคัญ
8. Treatment effect ตัวแปรที่ระบุถึงผลของการรักษาว่าได้รับบาดเจ็บหรือเสียชีวิต โดยค่าผลกระทบต่อ Y เท่ากับ –0.711 ซึ่งเป็นค่าลบที่แสดงว่าการรักษามีผลลดต่อ Y
9. age ตัวแปรที่ระบุอายุของบุคคล โดยค่าผลกระทบต่อ Y เท่ากับ –0.013 ซึ่งแสดงถึงผลลดเล็กน้อยจากอายุที่เพิ่มขึ้น

4.1.6 ผลการนำโมเดลไปใช้งานจริง (Deployment) ผลการนำโมเดล Linear Regression ที่ผ่านการพัฒนาและผลของการเปรียบเทียบมา แสดงในรูปแบบ กราฟเส้น (Line Chart) บนเว็บไซต์ โดยเขียนภาษา PHP และ HTML ในการทำกราฟนำข้อมูลของปี 2007–2015 มาพยากรณ์ปี 2016



ภาพที่ 4.29 แสดงกราฟผลการเปรียบเทียบระหว่างค่าจริงของปี 2007–2015 และค่าพยากรณ์

จากกราฟผลผลการเปรียบเทียบระหว่างค่าจริงของปี 2007–2015 และค่าพยากรณ์ ที่ได้ในโมเดล Linear Regression ของจำนวนผู้เสียชีวิตและผู้บาดเจ็บของช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ พบว่าความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุและจำนวนผู้เสียชีวิต ค่าจริงของช่วงเวลาที่เกิดอุบัติเหตุ มีจำนวนผู้เสียชีวิตและผู้บาดเจ็บมากที่สุดคือช่วง 18:01–19:00 น. และลดลงในช่วงเวลาถัดมา แนวโน้มโดยรวมแสดงให้เห็นช่วงเย็นถึงค่ำ (17:00–20:00 น.) เป็นช่วงที่มีจำนวนอุบัติเหตุร้ายแรงสูงสุด ซึ่งมีความสัมพันธ์กับช่วงที่เริ่มมีการเดินทางเพิ่มขึ้นจากการออกจากที่ทำงานหรือเดินทางกลับภูมิลำเนา ผู้ขับขี่อาจประสบปัญหาการจราจรหนาแน่นและความเร่งรีบในช่วงเทศกาลปีใหม่ ซึ่งเพิ่มความเสี่ยงต่ออุบัติเหตุ และช่วงเวลาที่มีจำนวนผู้เสียชีวิตและบาดเจ็บต่ำที่สุดคือช่วง 05:01–06:00 ซึ่งเป็นช่วงที่การสัญจรเริ่มกลับมาเป็นปกติ

ในส่วนของเส้นพยากรณ์ของปี 2016 ค่าพยากรณ์ของช่วงเวลาที่เกิดอุบัติเหตุและมีจำนวนผู้เสียชีวิตและบาดเจ็บ พบว่ามีแนวโน้มเพิ่มขึ้นอย่างชัดเจนคือช่วง 14:01–15:00 น. และมีค่าลดลงในช่วงเวลาอื่น ซึ่งแตกต่างจากค่าจริงที่แสดงให้เห็นว่าช่วง 18:01–19:00 น. มีจำนวนผู้เสียชีวิตและผู้บาดเจ็บที่สูง แต่ช่วงเวลาที่จำนวนผู้บาดเจ็บและเสียชีวิตสูงสุดในค่าจริงและค่าพยากรณ์มีความคล้ายคลึงกัน คือในช่วงเวลา 14:00 – 15:00 น. ค่าพยากรณ์ปี 2559 มีแนวโน้มสูงกว่าค่าจริงรวมในช่วงเวลาเดียวกัน แสดงถึงความน่าจะเป็นในการเพิ่มจำนวนผู้เกิดอุบัติเหตุในอนาคตและ

ช่วงเวลาตั้งแต่ 00:00 – 05:00 น. มีจำนวนผู้เกิดอุบัติเหตุต่ำที่สุดทั้งในค่าจริงและค่าพยากรณ์ ซึ่งหมายความว่า โมเดลยังสามารถพยากรณ์แนวโน้มโดยรวมได้ดีในระดับหนึ่ง แม้ไม่สามารถจับค่าที่พุ่งสูงสุดได้ สะท้อนถึงความคลาดเคลื่อนของโมเดล Linear Regression ในการจับแนวโน้มของช่วงเวลาที่มีความเสี่ยงสูงสุด ซึ่งอาจเกิดจากความไม่สัมพันธ์กันของข้อมูลในบางช่วงเวลา เนื่องจากอาจจะต้องหาปัจจัยเสริมในการเกิดอุบัติเหตุเพิ่มขึ้นเพื่อพัฒนาโมเดลที่สะท้อนสถานการณ์จริงได้แม่นยำยิ่งขึ้น

4.2 ผลการวิเคราะห์และออกแบบระบบ

4.2.1 ผลการวิเคราะห์และออกแบบฐานข้อมูล

ระบบจัดการข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ได้มีระบบมีฟังก์ชันหลัก ได้แก่การแสดงจำนวนข้อมูลทั้งหมด การเพิ่มข้อมูล แก้ไข ลบ ค้นหา การกรองข้อมูลปีและจังหวัด นอกจากนี้ ยังมีระบบล็อกอินเพื่อกำหนดสิทธิ์การเข้าถึงของผู้ใช้งาน ซึ่งช่วยให้สามารถจัดการข้อมูลได้อย่างเป็นระบบและปลอดภัย ในส่วนฐานข้อมูลของระบบนี้ ต้องสามารถจัดเก็บข้อมูลสำคัญเกี่ยวกับการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ เช่น วันเดือนปีที่เกิดอุบัติเหตุ อายุ เพศ ช่วงเวลาการเกิดอุบัติเหตุ ถนนที่เกิดเหตุ รถผู้บาดเจ็บ รถคู่กรณี สถานะ

ข้อมูลพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ระบบต้องมีการจัดเก็บข้อมูลผู้ใช้งานที่สามารถล็อกอินเข้าสู่ระบบเพื่อกำหนดสิทธิ์ในการเข้าถึงและจัดการข้อมูล โดยสิทธิ์การใช้งานของระบบแบ่งออกเป็น 2 ระดับหลัก ได้แก่ ผู้ดูแลระบบ (Admin) สามารถเผยแพร่ข้อมูลสารสนเทศเกี่ยวกับอุบัติเหตุ สามารถอัปโหลดข้อมูลเป็นไฟล์ CSV สามารถเพิ่ม ลบ และแก้ไขข้อมูลเกี่ยวกับข้อมูลอุบัติเหตุ ผู้ใช้งานทั่วไป (User) สามารถดูแนวโน้มและการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่จากการพยากรณ์ในลักษณะแผนภูมิ สามารถดูผลความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่จากการวิเคราะห์ข้อมูลในลักษณะกราฟแท่ง กราฟวงกลม สามารถดูข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในแต่ละปี สามารถดูแผนที่ประเภทยานพาหนะที่เกิดอุบัติเหตุในแต่ละพื้นที่

4.2.2 การออกแบบเว็บไซต์

1) หน้าแรกของเว็บไซต์แสดงเมนูต่างๆของหน้าเว็บและเนื้อหาที่เกี่ยวกับข้อมูลความรู้เกี่ยวกับการป้องกันอุบัติเหตุ

ภาพที่ 4.30 แสดงโครงร่างการออกแบบหน้าแรกของเว็บไซต์

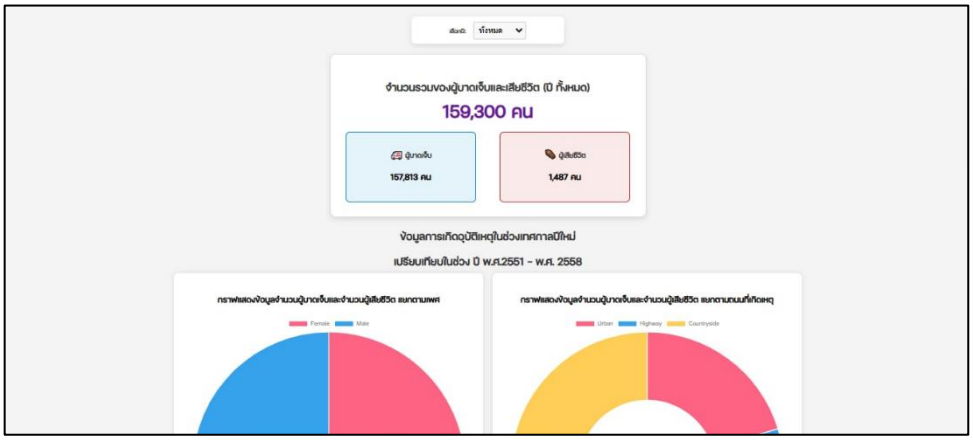
จากภาพที่ 4.30 แถบเมนูสามารถใช้งานได้ตามความต้องการของการใช้งาน ซึ่งเป็นหน้า สำหรับผู้ใช้งานทั่วไปสามารถเลือกดูและศึกษาค้นคว้าเข้าไปใช้งานในส่วนของเว็บไซต์ได้

2) หน้าแสดงข้อมูลที่มาของการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558

ภาพที่ 4.31 แสดงหน้าข้อมูลที่มาและความสำคัญ

จากภาพที่ 4.31 หน้าที่แสดงการอธิบายรายละเอียดที่และความสำคัญ วัตถุประสงค์และประโยชน์ที่ได้รับ ของการวิเคราะห์ข้อมูล การเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558

3) หน้าแสดงผลสถานการณ์ผู้บาดเจ็บและเสียชีวิตของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558



ภาพที่ 4.32 แสดงผลสถานการณ์ผู้บาดเจ็บและเสียชีวิตของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558

จากภาพที่ 4.31 หน้าที่แสดงการจำนวนผู้บาดเจ็บและเสียชีวิตในแต่ละปี และแสดงกราฟวงกลม กราฟโดนัท กราฟเส้น กราฟแท่ง และแผนที่

4) หน้าแสดงผล Dashboard ของข้อมูลการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ. 2558

ภาพที่ 4.33 แสดงผล Dashboard ของข้อมูลการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่

5) หน้าแสดงผลข้อมูลผู้จัดทำ

ภาพที่ 4.34 แสดงผลข้อมูลผู้จัดทำ

บทที่ 5

สรุปผลการดำเนินงาน และข้อเสนอแนะ

จากผลการเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ ทางผู้วิเคราะห์ได้ทำการสรุปผลการทำ โครงงาน ข้อจำกัดของเว็บไซต์ ปัญหาและอุปสรรคต่าง ๆ ของการทำโครงงาน และ ข้อเสนอแนะเพื่อพัฒนาโครงงาน ดังนี้

5.1 บทสรุปผลโครงงาน

ผู้วิเคราะห์ข้อมูลดำเนินการรวบรวมข้อมูลการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่เป็นข้อมูลที่ได้จากได้รวบรวมข้อมูลจากเว็บไซต์ ศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th/ 2567) ของสถาบันการแพทย์ฉุกเฉินแห่งชาติ และได้ทำการวิเคราะห์ข้อมูลด้วย โดยการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining) สำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปี โดยทำการสร้างโมเดล 4 แบบ คือ 1.เทคนิคการถดถอยเชิงเส้น (linear regression) 2.เทคนิคเกาส์เซียน (Gussian Process) 3. เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine Regression) 4. เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks) โดยแบ่งข้อมูลออกเป็นชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training data) 70% และ ชุดข้อมูลที่ใช้ในการทดสอบ (Testing data) 30% โดยใช้ค่า RMSE และ MAE เป็นตัวเปรียบเทียบ ได้ผลสรุปว่า โมเดลที่มีคา MAE นอยที่สุด คือ SVMR ซึ่งมีค่าเท่ากับ 4.656 แตถาเปรียบเทียบประสิทธิภาพของโมเดลด้วยคา RMSE พบวา โมเดล Linear Regression มีคา RMSE น้อยที่สุด คือ 5.758 ผู้วิเคราะห์จึงลองใชโมเดล SVMR และโมเดล Linear Regression ในการเปรียบเทียบโมเดลขึ้นมาใหม่ โดยใชขอมูลช่วงเทศกาลปีใหม่ปี 2558 ในการคำนวณคา RMSE และ MAE นอยที่สุด ผลพบว่าโมเดล Linear Regression มีประสิทธิภาพดีกวาเนื่องจากคา MAE เทากับ 3.716 มีคานอยกวาโมเดล SVMR และ ค่า RMSE ของโมเดล Linear Regression มีคานอยกวาโมเดล SVMR เท่ากับ 5.142 เชนเดียวกัน

ทางผู้วิเคราะห์ข้อมูลได้นำผลลัพธ์จากการวิเคราะห์มาสร้างเป็นสารสนเทศเพื่อจัดแสดงบนเว็บไซต์ สำหรับการเผยแพร่ข้อมูลการพยากรณ์แนวโน้มและการวิเคราะห์ความสัมพันธ์ของช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ โดยใช้เทคโนโลยีและชุดคำสั่ง HTML, CSS, และ PHP ในการพัฒนาเว็บไซต์ เพื่อเผยแพร่ข้อมูลที่เกี่ยวข้องให้ผู้ที่สนใจสามารถ

เข้าถึงและเข้าใจผลการวิเคราะห์ได้อย่างชัดเจน อีกทั้งยังเป็นประโยชน์ต่อหน่วยงานที่เกี่ยวข้องในการนำข้อมูลดังกล่าวไปใช้ประกอบการวางแผนและพัฒนากลยุทธ์ที่มีประสิทธิภาพในการลดอุบัติเหตุในช่วงเทศกาลปีใหม่

5.2 ข้อจำกัดเว็บไซต์

5.2.1 ผู้ใช้ไม่สามารถอัปโหลดข้อมูลการทำนายได้ด้วยตัวเอง

5.2.2 การแสดงผลข้อมูลที่ใช้การสร้างโมเดล ไม่สามารถทำการแก้ไข เปรียบเทียบ ได้โดยตรง

5.3 ปัญหาและอุปสรรคของโครงงาน

5.3.1 ไม่มีความรู้เกี่ยวกับภาษา JavaScript ทำให้ต้องใช้เวลาในการเรียนรู้ และ พัฒนาเว็บไซต์

5.3.2 ข้อมูลต้องแปลงเป็นตัวเลขจึงทำให้ยากต่อการนำสมการไปเขียนบนเว็บไซต์

5.3.3 แหล่งข้อมูลจากองค์กรภายในประเทศที่มีความน่าเชื่อถือ ไม่เปิดให้เข้าถึงได้จากเว็บไซต์ที่ไม่ได้ลงทะเบียนโดยตรง ทำให้ต้องทำการปรับแก้การเทรนโมเดลเพื่อให้สามารถทำการทำนายได้โดย ใช้ข้อมูลที่มาจากแหล่งอื่น

5.4 ข้อเสนอแนะ

การพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ที่จัดทำเสร็จสิ้นนี้แม้จะสามารถแสดงผลข้อมูลได้ถูกต้องตรง ตามวัตถุประสงค์ และขอบเขตที่ตั้งไว้แต่ก็ยังมีข้อจำกัดหลายประการ ซึ่งหากจะพัฒนาให้ เว็บไซต์เผยแพร่ข้อมูลนี้แสดงผลข้อมูล และทำงานได้อย่างมีประสิทธิภาพมากขึ้นผู้พัฒนาควร จะปรับปรุงในส่วนต่าง ๆ ดังต่อไปนี้

5.4.1 ควรเพิ่มปัจจัยเสริมในสมการพยากรณ์เพื่อให้โมเดลสามารถสะท้อนสถานการณ์จริงได้อย่างแม่นยำ ควรพิจารณาปัจจัยเสริมที่อาจส่งผลต่อการเกิดอุบัติเหตุ เช่น สภาพอากาศ (ฝนตก หมอก) ปริมาณการจราจรในช่วงเวลาวันหยุดนักขัตฤกษ์หรือเทศกาล การเพิ่มตัวแปรเหล่านี้อาจช่วยลดความคลาดเคลื่อนของโมเดลในช่วงเวลาที่มีการพยากรณ์ผิดพลาด

5.4.2 ควรปรับปรุงโมเดล Linear Regression แม้ Linear Regression จะให้ผลการพยากรณ์ที่ดี แต่ยังมีความคลาดเคลื่อนในบางช่วงเวลา การใช้โมเดลที่คำนึงถึงความสัมพันธ์ไม่เชิงเส้น (Non-linear Relationships) เช่น Polynomial Regression อาจช่วยเพิ่มความแม่นยำในการจับแนวโน้มที่ซับซ้อน

5.4.3 ข้อจำกัดของงานวิจัยและการศึกษาต่อในอนาคต ข้อมูลที่ใช้ในการวิเคราะห์อาจมีข้อจำกัดด้านความครอบคลุม (เฉพาะจังหวัดใหญ่) หรือความล้าสมัย การศึกษาข้อมูลที่อัปเดตและครอบคลุมทั่วประเทศจะช่วยเพิ่มความแม่นยำและความน่าเชื่อถือ ควรศึกษาการเปรียบเทียบระหว่างโมเดลอื่น ๆ เพิ่มเติม เช่น Deep Learning หรือ Time Series Analysis โดยละเอียด

5.4.4 การทดสอบโมเดลเพิ่มเติมในสถานการณ์ต่าง ๆ ควรนำโมเดลที่พัฒนาขึ้นไปทดลองในช่วงเทศกาลอื่น หรือสถานการณ์ที่แตกต่าง เพื่อประเมินความสามารถในการพยากรณ์ของโมเดลในบริบทอื่น ๆ และเพิ่มความน่าเชื่อถือของผลการวิจัย

5.4.5 ควรจะเจาะลึกในแต่ละกราฟและหาสาเหตุของการเกิดอุบัติเหตุในแต่ละกราฟของหน้าเว็บไซต์

5.4.6 เลือกใช้รูปแบบแผนที่ที่แสดงบนเว็บไซต์ให้เหมาะสมกับข้อมูลและดูง่ายสำหรับที่เข้าชมเว็บไซต์

บรรณานุกรม

จรัสสรี รุ่งรัตนาอุบล. (2566). เหมืองข้อมูล (Data Mining). สำนักพิมพ์มหาวิทยาลัยนเรศวร. เข้าถึงได้จาก nupress.grad.nu.ac.th: https://www.nupress.grad.nu.ac.th/data-mining/#1

ณภัทร ภัทรพิศาล. (4 9 2564). ใบขับขี่ตลอดชีพ-ปัญหาที่ยังหลงเหลือในประเทศไทยกับความเสี่ยงภัยของผู้ขับขี่ในวัยผู้สูงอายุ. เข้าถึงได้จาก tdri: https://tdri.or.th/2020/09/elderly-drivers-with-lifetime-driving-licenses/

ธเนศ เสถียรนาม,วิชุดา เสถียรนาม ศิรวิทย์ ชาวระ. (2566). ปัจจัยที่มีอิทธิพลต่อความรุนแรงในการเกิดอุบัติเหตุย้อนศรบนทางหลวงของประเทศไทย. เข้าถึงได้จาก conference.thaince.org: https://conference.thaince.org/index.php/ncce28/article/view/2077

นรินทร์ จิวิตัน. (2565). thaijo.org. เข้าถึงได้จาก https://ph01.tci-thaijo.org/index.php/IT_Journal/article/view/248868/168735

บริษัท เฟิร์สคราฟท์ ดิจิทัล โซลูชั่น จำกัด. (2563). Data Visualization คืออะไร? . เข้าถึงได้จาก stcraft: https://1stcraft.com/what-is-data-visualization/

บริษัท วันบีลีฟ จำกัด. (2566). หลักการ ออกแบบเว็บ ขั้นพื้นฐาน พร้อมองค์ประกอบ. เข้าถึงได้จาก 1belief: https://www.1belief.com/article/website-design/

ปทิตญา บุญรักษา. (2561). การพยากรณ์การเกิดอุบัติเหตุบนท้องถนนโดยใช้เทคนิคอนุกรมเวลา. เข้าถึงได้จาก http://202.28.34.124/dspace/bitstream/123456789/232/1/60011284505.pdf

พิชญากร เลค. (2564). การเปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องจักร เพื่อทำนายการเกิดอุบัติเหตุ โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล. เข้าถึงได้จาก e-research.siam.edu: https://e-research.siam.edu/kb/comparing-machine-learning-algorithm/

พิเชษฐ์ ศิริรัตนไพศาลกุล. (2565). การเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ราคาทองคำด้วยเทคนิคการขุดข้อมูล. เข้าถึงได้จาก thaijo: https://so03.tci-thaijo.org/index.php/msj/article/view/254062

เพ็ญพักตร์ พรชะตา. (2565). การมีส่วนร่วมของภาคีเครือข่ายในการป้องกันอุบัติเหตุชุมชน ในเขตอำเภอวังสามหมอ จังหวัดอุดรธานี. เข้าถึงได้จาก trsl.thairoads.org: https://trsl.thairoads.org/PrintDetail.aspx?id=1968

ภูมินทร์ สุขโข. (2566). การวิเคราะห์ปัจจัยที่มีอิทธิพลต่อความรุนแรงของอุบัติเหตุรถจักรยานยนต์บนทางหลวงในประเทศไทย. เข้าถึงได้จาก thaijo.org: https://ph02.tci-thaijo.org/index.php/gskku/article/view/246638/168885

bbc. (22 12 2561). อุบัติเหตุบนถนน : ไทยครองแชมป์อัตราการตายบนถนนสูงสุดในเอเชียตะวันออกเฉียงใต้. เข้าถึงได้จาก bbc: https://www.bbc.com/thai/international-46545106

blendata. (2567). รู้จัก Data Analytics ศาสตร์แห่งการวิเคราะห์ และรูปแบบการวิเคราะห์ข้อมูลที่ช่วยสร้างประโยชน์ ให้กับธุรกิจ. เข้าถึงได้จาก blendata.c: https://shorturl.asia/KhU1N

Center for Integrative Environmental and Energy Research. (2566). โครงข่ายประสาทเทียม (Artificial Neural Networks : ANN). เข้าถึงได้จาก weather4tha: https://weather4thai.kmitl.ac.th/ANN.php

doa. (2563). การวิเคราะห์ข้อมูลทางสถิติ. เข้าถึงได้จาก doa: https://www.doa.go.th/share/showthread.php?tid=2410

Eakasit Pacharawongsakda. (2558). กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM และตัวอย่างการประยุกต์ใช้ทางด้านการศึกษา. เข้าถึงได้จาก linkedin: https://th.linkedin.com/pulse/%E0%B8%81%E0%B8%A3%E0%B8%B0%E0%B8%9A%E0%B8%A7%E0%B8%99%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%A7%E0%B9%80%E0%B8%84%E0%B8%A3%E0%B8%B2%E0%B8%B0%E0%B8%AB%E0%B8%82%E0%B9%89%E0%B8%AD%E0%B8%A1%E0%B8%B9%E0%B8%A5%E0%B8%94%E0%B9%89%E0%B8%A7

KongRuksiam Studio. (2565). Machine Learning(EP.3)- การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression). เข้าถึงได้จาก kongruksiam: https://kongruksiam.medium.com/%E0%B8%AA%E0%B8%A3%E0%B8%B8%E0%

ThaiHealth Official. (2555). เทศกาลไทยกับภัยอุบัติเหตุ. เข้าถึงได้จาก thaihealth.or.th: https://www.thaihealth.or.th/%e0%b9%80%e0%b8%97%e0%b8%a8%e0%b8%81%e0%b8%b2%e0%b8%a5%e0%b9%84%e0%b8%97%e0%b8%a2%e0%b8

%81%e0%b8%b1%e0%b8%9a%e0%b8%a0%e0%b8%b1%e0%b8%a2%e0%b8%ad%e0%b8%b8%e0%b8%9a%e0%b8%b1%e0%b8%95%e0%b8%b4%e0%b9%80/

ภาคผนวก

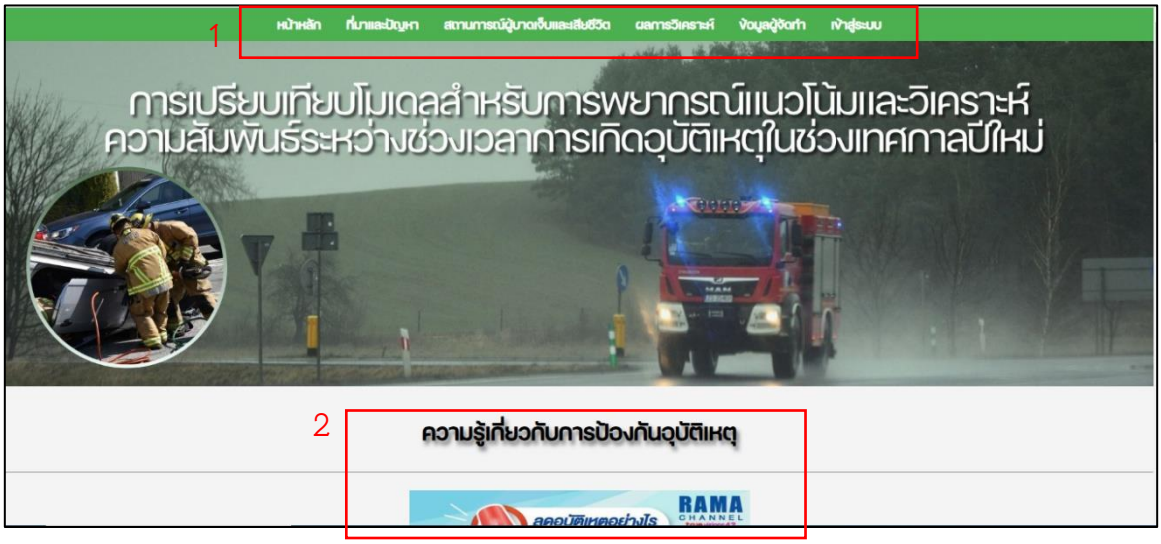
ภาคผนวก ก
คู่มือการใช้งาน

ภาคผนวก ก
คู่มือการใช้งาน

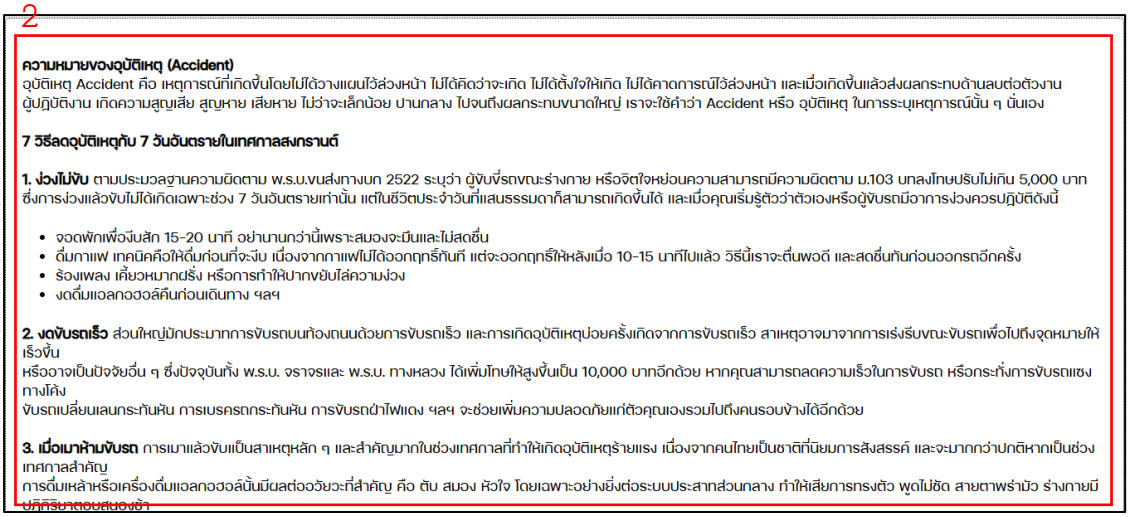
จากการดำเนินงานการเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ โดยเว็บไซต์มีส่วนใช้งานทั้งหมด 6 ส่วน ดังนี้ ส่วนของหน้าหลักของเว็บไซต์ ส่วนที่มาของการวิเคราะห์ปัญหา ส่วนของสถานการณ์ผู้บาดเจ็บและเสียชีวิต ส่วนของผลการวิเคราะห์ ส่วนของผู้จัดทำ และส่วนของการเข้าสู่ระบบของแอดมิน

คู่มือการใช้งานเว็บไซต์

1) หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์การเปรียบเทียบโมเดลสำหรับการพยากรณ์แนวโน้มและวิเคราะห์ความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่



ภาพที่ ก.1 แสดงหน้าแรกของเว็บไซต์



ภาพที่ ก.2 แสดงหน้าแรกของเว็บไซต์ ส่วนของความรู้เกี่ยวกับการป้องกันอุบัติเหตุ

หมายเลข 1 เป็นส่วนของแถบเมนูการใช้งานต่าง ๆ ภายในเว็บไซต์ ดังนี้

1.1 ส่วนของความรู้เกี่ยวกับการป้องกันอุบัติเหตุ

1.2 ส่วนของที่มาของการวิเคราะห์ข้อมูล

1.3 ส่วนของสถานการณ์ผู้บาดเจ็บและเสียชีวิต

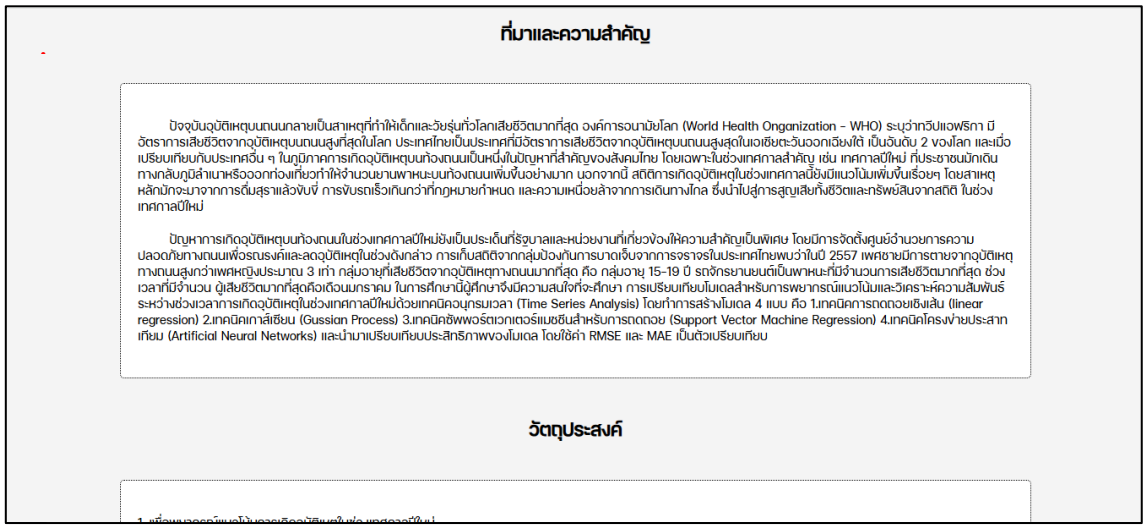
1.4 ส่วนของผลการวิเคราะห์

1.5 ส่วนของข้อมูลผู้จัดทำ

1.6 ส่วนของการเข้าสู่ระบบของแอดมิน

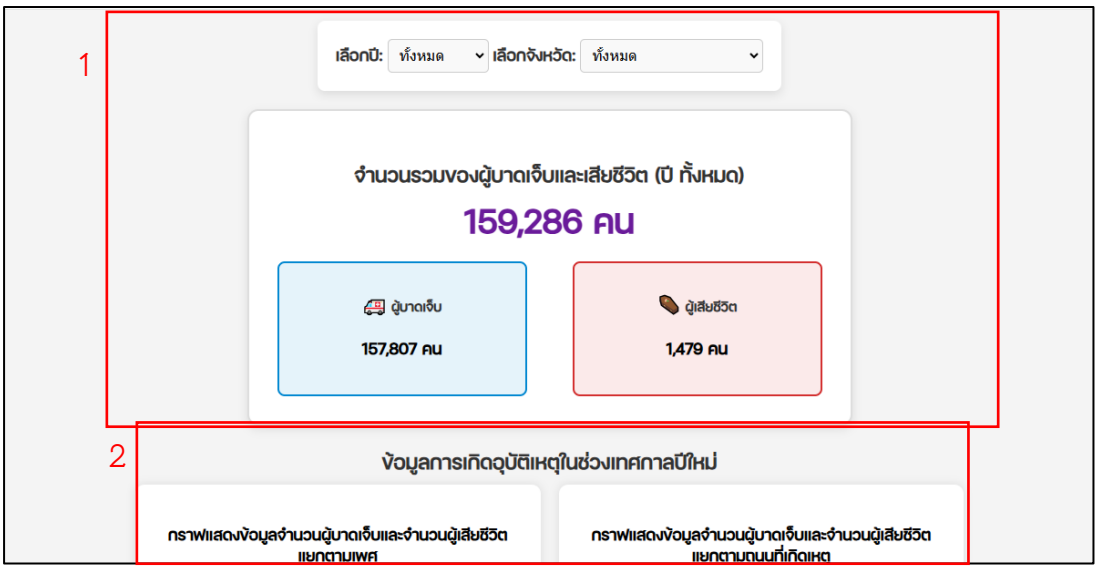
หมายเลข 2 เป็นส่วนของความรู้เกี่ยวกับการป้องกันอุบัติเหตุ

2) หน้าแสดงส่วนของข้อมูลที่มาและความสำคัญของโครงงาน



ภาพที่ ก.3 ที่มาและความสำคัญของโครงงาน

3) หน้าแสดงผลสรุปข้อมูลสถานการณ์ผู้บาดเจ็บและเสียชีวิต พร้อมทั้งอธิบายกราฟในแต่ละส่วนของแผนภาพนำเสนอ

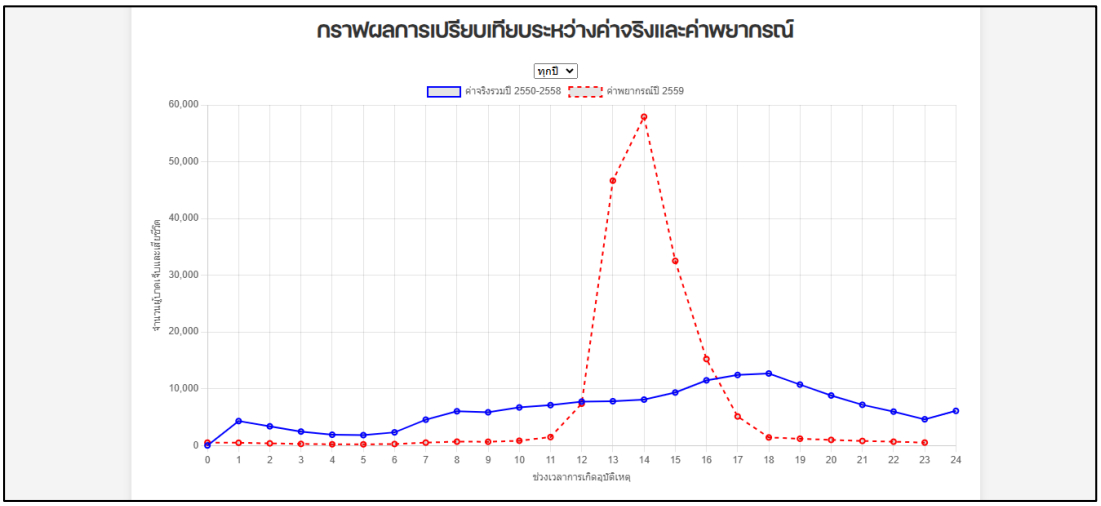


ภาพที่ ก.4 หน้าแสดงผลกราฟสรุปข้อมูลสถานการณ์ผู้บาดเจ็บและเสียชีวิต

หมายเลข 1 เป็นส่วนของแสดงจำนวนของผู้บาดเจ็บและเสียชีวิต

หมายเลข 2 เป็นส่วนของกราฟต่างๆ เช่น กราฟวงกลม กราฟแท่ง กราฟเส้น แผนที่

4) แสดงส่วนของการวิเคราะห์ข้อมูลผลความสัมพันธ์ระหว่างช่วงเวลาการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่จากการวิเคราะห์ข้อมูลในลักษณะแผนภูมิ

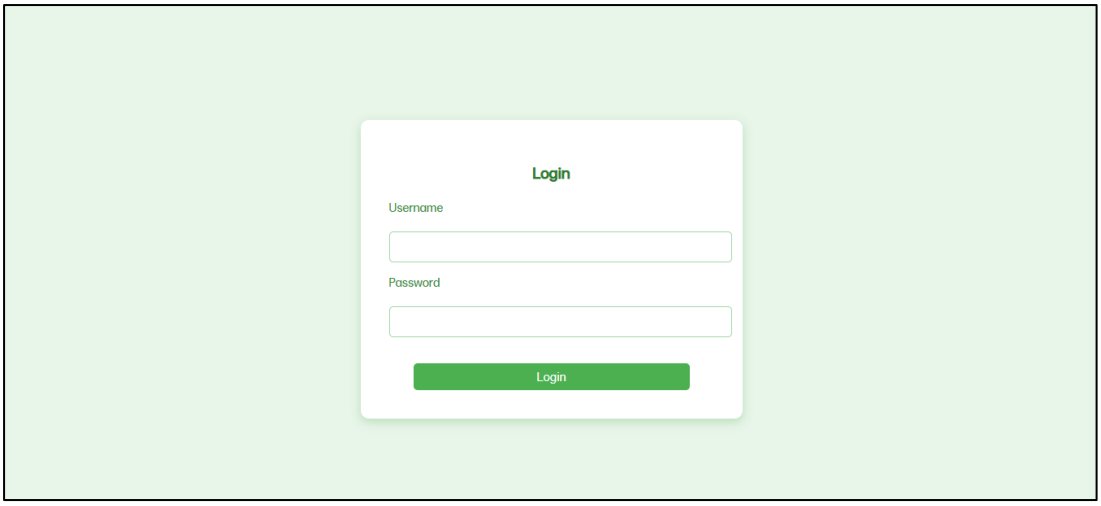


ภาพที่ ก.5 แสดงส่วนของผลการวิเคราะห์ข้อมูล

5) หน้าแสดงส่วนของข้อมูลผู้จัดทำเป็นหน้าแสดงประวัติผู้จัดทำเว็บไซต์

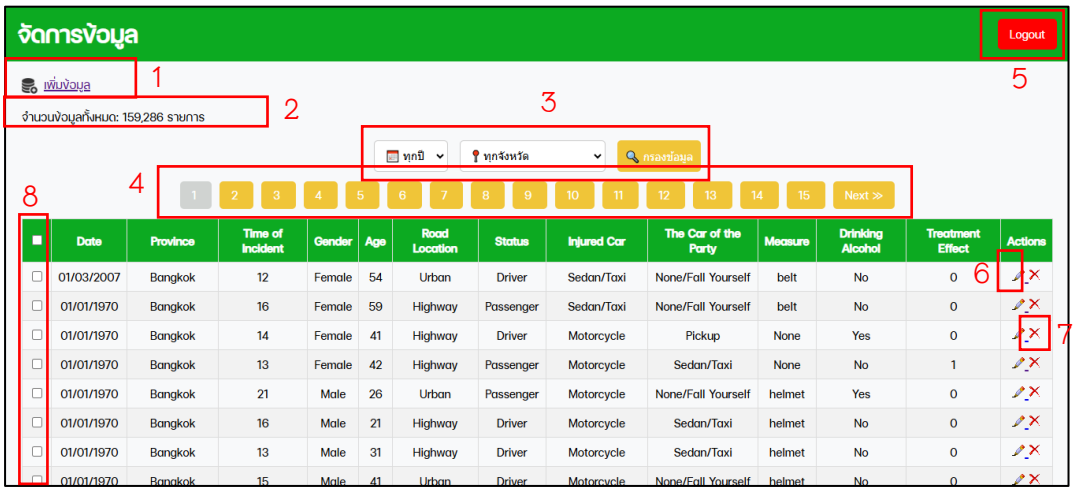
ภาพที่ ก.6 แสดงหน้าข้อมูลผู้จัดทำ

6) หน้าแสดงส่วนของการเข้าสู่ระบบเพื่อไปจัดการกับข้อมูลภายในเว็บไซต์ โดยจำกัดการ เข้าถึงข้อมูลเฉพาะแอดมินเท่านั้น



ภาพที่ ก.7 แสดงหน้าเข้าสู่ระบบ

7) หน้าแสดงส่วนของการจัดการกับข้อมูลภายในเว็บไซต์ โดยแอดมินสามารถแก้ไขเพิ่ม หรือลบข้อมูลภายในเว็บไซต์นี้ได้



ภาพที่ ก.8 แสดงหน้าการจัดการข้อมูลภายในเว็บไซต์

หมายเลข 1 ปุ่มเพิ่มไฟล์ข้อมูล

หมายเลข 2 แสดงจำนวนข้อมูลทั้งหมด

หมายเลข 3 ปุ่มสำหรับการกรองข้อมูล

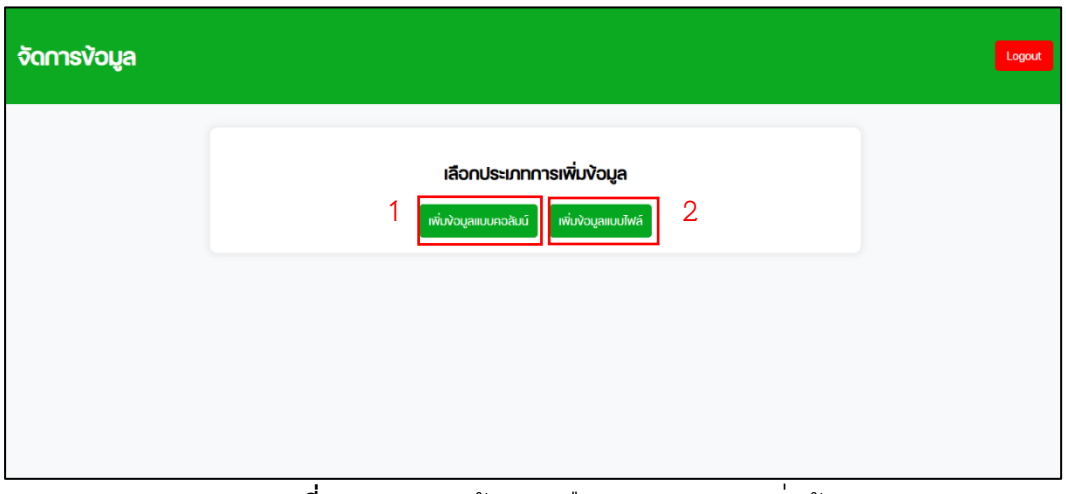
หมายเลข 4 ปุ่มเลขแบ่งหน้าข้อมูล

หมายเลข 5 ปุ่มออกจากระบบ

หมายเลข 6 ปุ่มแก้ไขข้อมูล

หมายเลข 7 ปุ่มลบข้อมูลทีละรายการ

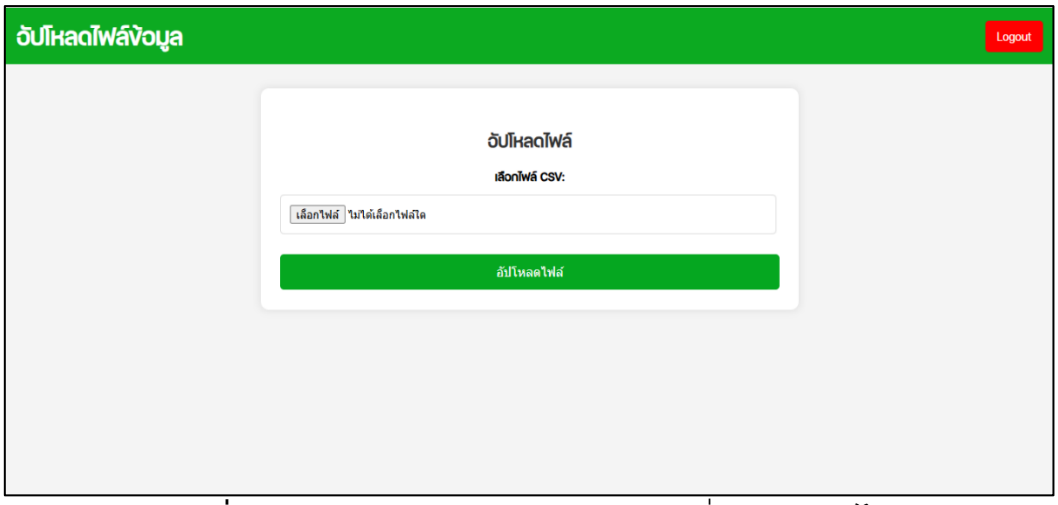
หมายเลข 8 ปุ่มลบข้อมูลทีละหลายรายการ



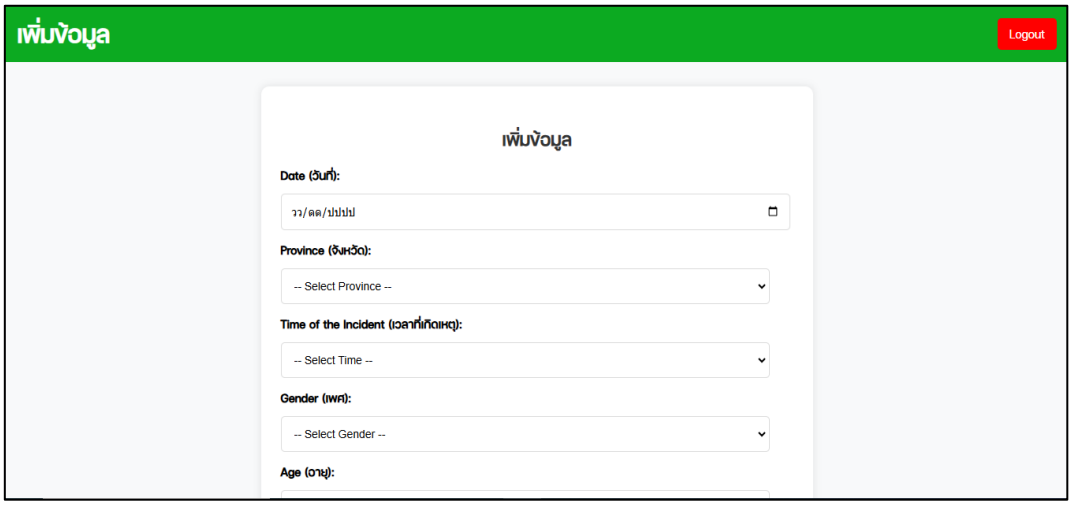
ภาพที่ ก.9 แสดงหน้าการเลือกประเภทการเพิ่มข้อมูล

หมายเลข 1 ปุ่มเพิ่มข้อมูลแบบคอลัมน์

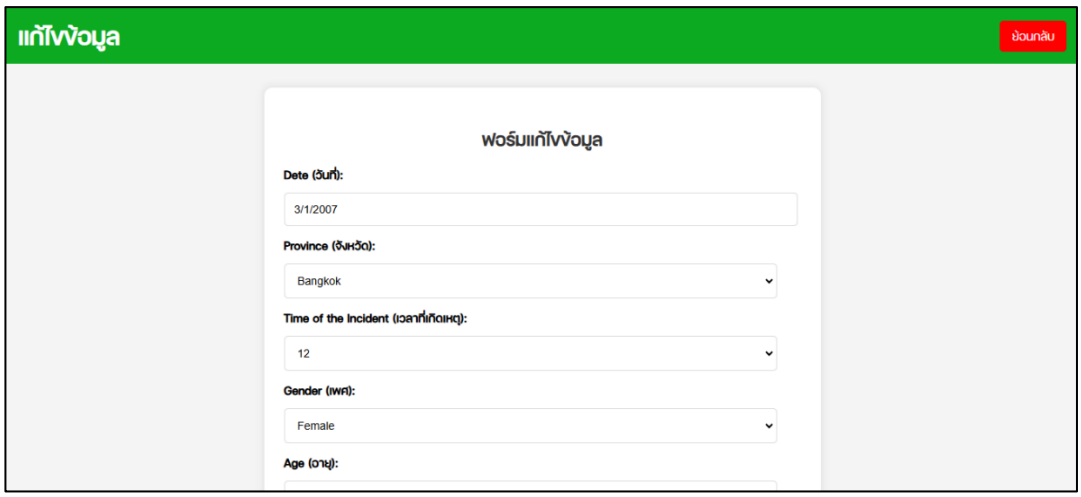
หมายเลข 2 ปุ่มเพิ่มข้อมูลแบบไฟล์



ภาพที่ ก.10 แสดงหน้าการเลือกประเภทการเพิ่มข้อมูลแบบไฟล์



ภาพที่ ก.11 แสดงหน้าการเลือกประเภทการเพิ่มข้อมูลแบบคอลัมน์



ภาพที่ ก.12 แสดงหน้าฟอร์มการแก้ไขข้อมูล

ภาคผนวก ข
ประวัติผู้จัดทำ

ประวัติผู้จัดทำ

ชื่อ-นามสกุล นางสาวปรีชญา อินต๊ะชุ่ม

วันเดือนปีเกิด 29 มิถุนายน 2545

ภูมิลำเนา 24/1 หมู่ 12 ตำบลชมพู อำเภอเมือง จังหวัดลำปาง

E-mail preechaya193@gmail.com

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น โรงเรียนลำปางกัลยาณี จังหวัดลำปาง สำเร็จการศึกษาปี 2560
- ระดับประกาศนียบัตรวิชาชีพ สาขาวิชาอิเล็กทรอนิกส์ วิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2564
- ระดับประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาเทคโนโลยีคอมพิวเตอร์ วิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2566

ประวัติผู้จัดทำ

ชื่อ-นามสกุล นางสาวกฤตยา คงดี

วันเดือนปีเกิด 22 มกราคม 2546

ภูมิลำเนา 5210054 ม.11 ต.ผักขวง อ.ทองแสนขัน จ. อุตรดิตถ์ 53230

E-mail Gittaya1978@gmail.com

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น โรงเรียนบ้านปางวุ้น จังหวัดอุตรดิตถ์ สำเร็จการศึกษาปี 2560
- ระดับประกาศนียบัตรวิชาชีพ สาขาวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษาอุตรดิตถ์ จังหวัดอุตรดิตถ์ สำเร็จการศึกษาปี 2564
- ระดับประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาเทคโนโลยีธุรกิจดิจิทัล วิทยาลัยอาชีวศึกษาอุตรดิตถ์ จังหวัดอุตรดิตถ์ สำเร็จการศึกษาปี 2566