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 มาตรฐานสำหรับการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

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

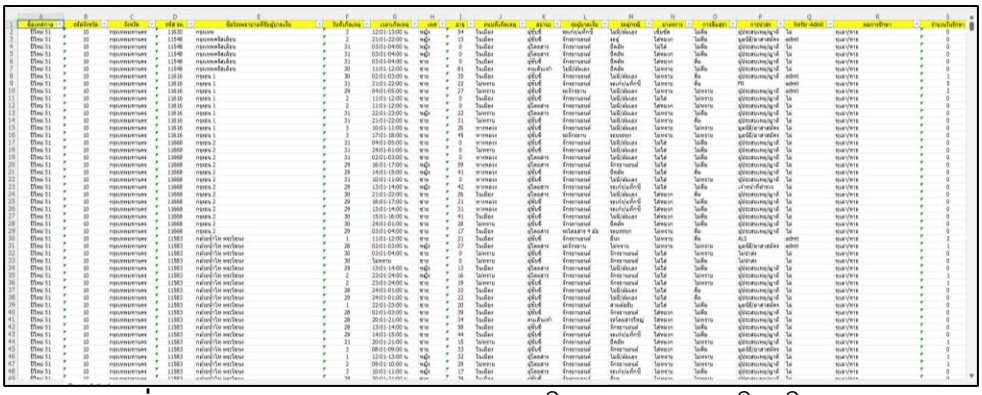
3.1.2 การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

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

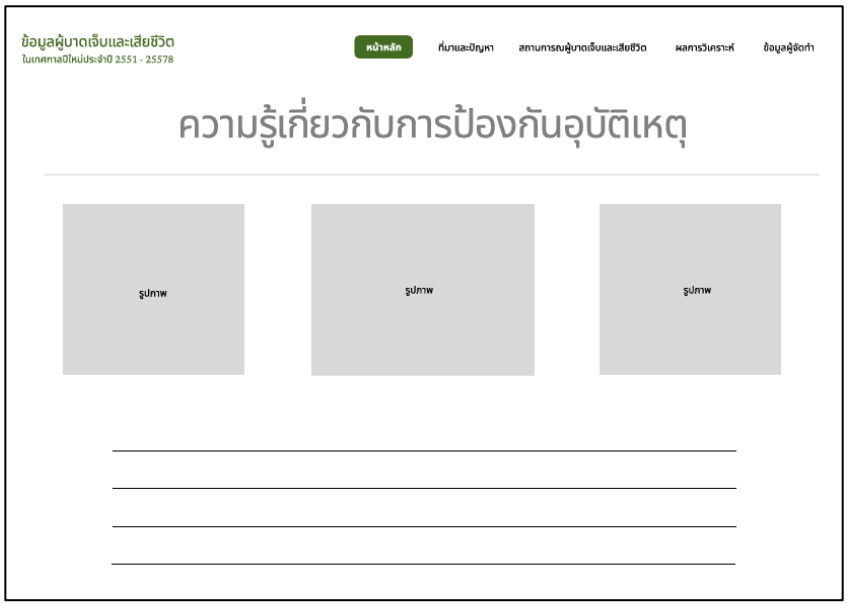
3.2.1 การวิเคราะห์และออกแบบฐานข้อมูล

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

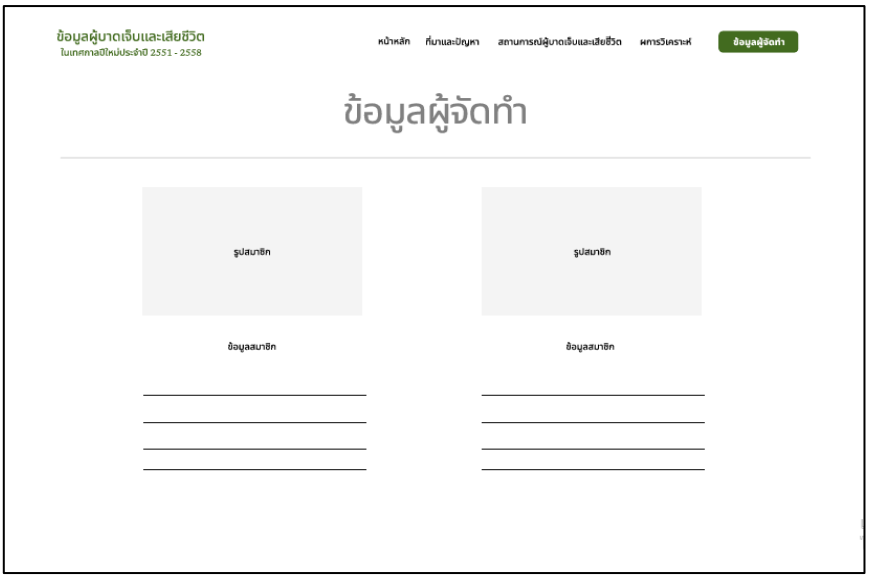
3) หน้าแสดงผลสถานการณ์ผู้บาดเจ็บและเสียชีวิตของการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ในช่วง พ.ศ.2551 – พ.ศ.2558

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