

การเปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องจักร เพื่อทำนายการเกิดอุบัติเหตุ โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล

Comparing Machine Learning Algorithm for Predicting Road Accidents Using Time Series Data Mining Techniques

ผศ.ดร.พิชญากกร เลค¹

Asst.Prof. Pitchayakorn Lake, Ph.D.¹

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ เปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องจักร เพื่อทำนายการเกิดอุบัติเหตุ โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลในจังหวัดที่มีจำนวนการเกิดอุบัติเหตุมากที่สุด ได้แก่ จังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ โดยรวบรวมข้อมูลตั้งแต่ปี พ.ศ. 2558-2562 จำนวน 60 เดือน จากการทดลองพบว่าแบบจำลองที่มีความเหมาะสมในการพยากรณ์การเกิดอุบัติเหตุในประเทศไทย ดังต่อไปนี้แบบจำลองการถดถอยเชิงเส้น มีความเหมาะสมมากที่สุดกับชุดข้อมูลจังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับร้อยละ 16.29 19.29 และร้อยละ 19.64

คำสำคัญ : การพยากรณ์, การวิเคราะห์อนุกรมเวลา, เทคนิคเหมืองข้อมูล, การเกิดอุบัติเหตุ

Abstract

The purpose of this research is to compare machine learning algorithm for predicting road accidents using time series data mining techniques. There are three techniques such as Linear Regression, Multi-Layer Perceptron and Support Vector Machine for Regression. The data used for the study collected from the provinces, which cause the highest road accidents are Bangkok, Nakhon Ratchasima and Samut Prakarn. The data was collected from 2015-2019 AD. totally 60 months. This research found that the suitable forecasting model for number of road accidents in Thailand as followed: The forecasting model using Linear Regression was the most suitable are Bangkok, Nakhon Ratchasima and Samut Prakarn, which had the rate of MMRE (Mean Magnitude of Relative Error) with the percentage of 16.29, 19.29, and 19.64, respectively.

Keywords : Predicting, Time Series Analysis, Data Mining Techniques, Accidents

¹ ภาควิชาธุรกิจดิจิทัล คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยสยาม กรุงเทพฯ 10160 ประเทศไทย

Department of Business Digital, Faculty of Information Technology, Siam University, Bangkok 10160, Thailand.

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันนี้ ปัญหาด้านอุบัติเหตุบนถนนนั้นมีความโน้มเอียงเพิ่มขึ้น แม้จะมีการดำเนินการเพื่อแก้ปัญหาการเกิดอุบัติเหตุ แต่อัตรารับบาดเจ็บและผู้เสียชีวิตและผู้ที่ได้รับบาดเจ็บเพิ่มขึ้นอย่างต่อเนื่อง จากสถิติที่ผ่านมา ส่วนใหญ่อุบัติเหตุบนท้องถนนเกิดจากพฤติกรรมของผู้ที่ขับขี่ยานพาหนะ สัดส่วนผู้เสียชีวิตด้วยอุบัติเหตุทางถนนทั่วโลกมีผู้เสียชีวิตจากรถยนต์มากที่สุด ร้อยละ 29 รถจักรยานยนต์ ร้อยละ 28 และผู้ขับขี่จักรยานและผู้เดินเท้า ร้อยละ 26 สถาบันวิจัยทีดีอาร์ไอ คำนวณมูลค่าความสูญเสียจากการเสียชีวิตและบาดเจ็บสาหัสจากอุบัติเหตุจราจรคิดเป็นมูลค่าทางเศรษฐกิจประมาณ 5 แสนล้านบาทต่อปี

ผู้วิจัยได้พัฒนางานวิจัยนี้เพื่อสร้างแบบจำลองการพยากรณ์การเกิดอุบัติเหตุในประเทศไทย โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล เพื่อสร้างแบบจำลอง (Model) ในการทำนายแนวโน้มของจำนวนอุบัติเหตุ โดยใช้จำนวนการเกิดอุบัติเหตุของปีที่ผ่านมา ชุดข้อมูลของผู้วิจัยนำมาใช้เป็นชุดข้อมูลของจังหวัดที่มีจำนวนครั้งในการเกิดอุบัติเหตุมากที่สุด ได้แก่ จังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ ประจำปี พ.ศ. 2558 – 2562

1.2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ข้อมูลอนุกรมเวลา (Time Series Data)

Hoshmand (2009) กล่าวว่าข้อมูลอนุกรมเวลา คือ ชุดข้อมูลที่รวบรวมและจัดเก็บตามลำดับต่อเนื่องกันภายใต้การเพิ่มขึ้นของเวลา โดยข้อมูลดังกล่าวจะถูกเก็บรวบรวมอย่างต่อเนื่องในระยะเวลาติดต่อกันตามช่วงเวลาที่ต้องการ เช่น ข้อมูลราคาทองคำ ข้อมูลปริมาณน้ำในเขื่อน ซึ่งจะถูกบันทึกข้อมูลเป็นวัน ในบางกรณีการจัดเก็บข้อมูลอาจมีลักษณะการจัดเก็บแบบเป็นช่วงเวลาต่อเนื่องกัน เพื่อสร้างแบบจำลองในการทำนายปริมาณที่จะเกิดขึ้นในอนาคต (Time Series Model) โดยการวิเคราะห์อนุกรมเวลา (Time Series Analysis)

การวิเคราะห์การถดถอย (Linear Regression)

เป็นวิธีการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ ซึ่งเทคนิคนี้จะอาศัยความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรเพื่อใช้ทำนาย (Kavitha, Varuna, & Ramya, 2016) โดยสมการ Linear Regression จะแสดงให้เห็นค่าความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระในลักษณะของข้อมูลเชิงปริมาณ โดยสามารถเขียนได้ดังนี้

$$Y = a + bx_1 + cx_2 + \dots \quad (1)$$

โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอน หลายชั้น (Multi-Layer Perceptron: MLP)

โครงข่ายประสาทเทียมแบบหลายชั้น ที่มีลักษณะการเชื่อมต่อแบบโยนไปข้างหน้าแบบทั่วถึง (Fully connected feed-forward nets) มีจำนวนชั้นตั้งแต่หนึ่งชั้นขึ้นไป (Lin et al. 2015; Kahani et al. 2018) ถูกพัฒนาขึ้นมาเพื่อแก้จุดอ่อนของโครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Perceptron) ให้ความสามารถในการคำนวณที่สูงขึ้น โครงข่ายประสาทเทียมแบบหลายชั้นประกอบด้วย ชั้นข้อมูลเข้า (Input Layer) ตามจำนวนแอททริบิวต์ของชุดข้อมูล ชั้นซ่อน (Hidden Layer) ความเหมาะสมของการกำหนดขึ้นอยู่กับทดสอบประสิทธิภาพ และชั้นผลลัพธ์ (Output Layer)

ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression)

ซัพพอร์ตเวกเตอร์แมชชีน หรือ SVM เป็นอัลกอริทึมที่ใช้ในการจำแนกกลุ่มของข้อมูลด้วยวิธีการหาระนาบการตัดสินใจ (Decision Hyperplane) หรือไฮเปอร์เพลนที่เหมาะสม สำหรับการแบ่งข้อมูล 2 ส่วนจากกัน SVM ยังถูกนำมาประยุกต์ใช้ในการสร้างสมการในการประมาณการหาค่าฟังก์ชันเชิงเส้น $f(x)$ ในการใช้ระนาบตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย หรือ SVR จะเป็นการนำข้อมูลปัจจุบัน และข้อมูลในอดีตจำนวนหนึ่ง มาทำการเรียนรู้ (Training) เพื่อให้ทราบรูปแบบสำหรับคาดการณ์ผลซึ่งจะเกิดขึ้นในอนาคต ในงานวิจัยนี้ผู้วิจัยใช้วิธีสร้างสมการค่าฟังก์ชันเชิงเส้นด้วยวิธี Sequential Minimal Optimization for SVM Regression (SMOreg) (Shevade *et al.*, 2000; Smola & Schölkopf, 2004)

รากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE)

เป็นค่าที่ใช้ในการวัดขนาดของความคลาดเคลื่อนของการพยากรณ์ โดยค่าดังกล่าวได้จากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) ซึ่งเป็นการนำผลต่างของค่าจริงและค่าที่ได้จากการพยากรณ์มายกกำลังสอง ค่า RMSE สามารถคำนวณได้จากสมการที่ (2)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2)$$

เมื่อ Y_i คือ ค่าประมาณการจากแบบจำลองการพยากรณ์ \hat{Y}_i คือ ค่าจริงที่ใช้ในการสร้างแบบจำลอง n คือ จำนวนข้อมูลในชุดข้อมูล

ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE)

เป็นค่าเฉลี่ยของความแตกต่างสัมบูรณ์ระหว่างค่าพยากรณ์และค่าจริง หากค่า MAE มีค่าน้อย แสดงว่าแบบจำลองสามารถประมาณค่าประมาณได้ใกล้เคียงกับค่าจริง ค่า MAE สามารถคำนวณได้จากสมการที่ (3)

$$MAE = \frac{1}{N} \sum_{i=1}^N (T_i - F_i) \quad (3)$$

เมื่อ T_i คือ ค่าจริง F_i คือ ค่าพยากรณ์ N คือ จำนวนข้อมูลในชุดข้อมูล

วีรศักดิ์ ฟองเงิน และคณะ (2560) การพยากรณ์ปริมาณน้ำในเขื่อนกัวลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล ได้นำข้อมูลที่มีผลต่อการเปลี่ยนแปลงระดับน้ำ เช่น ปริมาณน้ำไหลเข้าเขื่อน ปริมาณน้ำในเขื่อน เป็นต้น ผลการวิจัยพบว่า วิธีแบบจำลองต้นไม้เอนิมไฟว์พี วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีวิเคราะห์การถดถอย และวิธีโครงข่ายประสาทเทียม มีค่าความคลาดเคลื่อนเท่ากับ 10.56, 10.84, 11.12 และ 12.53

สุภาพร บรรดาศักดิ์ และคณะ (2559) พยากรณ์การใช้ลูกกอล์ฟในสนามฝึกซ้อมกอล์ฟโดยใช้ตัวแปรอนุกรมเวลา เพื่อใช้จัดเตรียมจำนวนลูกกอล์ฟที่ใช้ในวันถัดไป พบว่าแบบจำลองการถดถอยเชิงเส้นมีค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ประมาณ 34.32027 มีค่าเฉลี่ยความคลาดเคลื่อนน้อยกว่าเมื่อเปรียบเทียบกับการถดถอยเวกเตอร์ (SMOreg: Sequential Minimal Optimization Regression)

วัตถุประสงค์

1. เพื่อเปรียบเทียบประสิทธิภาพและเลือกเทคนิคที่เหมาะสมในการวิเคราะห์อนุกรมเวลาของอัตราการเกิดอุบัติเหตุในประเทศไทย
2. เพื่อสร้างรูปแบบการพยากรณ์อัตราการเกิดอุบัติเหตุด้วยวิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล

อุปกรณ์และวิธีการ

1. การเตรียมข้อมูล (Data Preparation) รวบรวมข้อมูลการเกิดอุบัติเหตุในประเทศไทย ผู้วิจัยนำข้อมูลมาจากภาครัฐ ในระหว่างปี พ.ศ. 2558 – 2562 อยู่ในรูปแบบของไฟล์คอมพิวเตอร์ (.pdf) แปลงไฟล์ข้อมูลในรูปแบบการใช้งานไฟล์ (.csv) จำนวน 89,674 รายการ ประกอบด้วย 14 แอททริบิวต์ ได้แก่ วันที่ อุบัติเหตุ (ครั้ง) จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ขับรถเร็วเกินกำหนด หนักใน อุปกรณ์รถบกพร่อง พลิกคว่ำ ชนสิ่งกีดขวาง ชนท้าย รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ รถบรรทุกมากกว่า 10 ล้อ (รถพ่วง) รถปิคอัพบรรทุก 4 ล้อ และจังหวัด โดยผู้วิจัยตัดส่วนที่ไม่จำเป็นออก คงเหลือ 8 แอททริบิวต์ ได้แก่ วันที่ อุบัติเหตุ (ครั้ง) จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ รถบรรทุกมากกว่า 10 ล้อ (รถพ่วง) รถปิคอัพบรรทุก 4 ล้อ และจังหวัด เพื่อนำเข้าสู่โปรแกรม WEKA ผู้วิจัยเลือกจังหวัดที่มีจำนวนครั้งในการเกิดอุบัติเหตุมากที่สุด ได้แก่ จังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ โดยแบ่งข้อมูลออกเป็น 2 ส่วน ตั้งแต่เดือนมกราคม ถึง เดือนธันวาคม คือ 1) ชุดข้อมูลการเรียนรู้ (Training Data Set) ใช้ข้อมูลในปี พ.ศ. 2558-2561 จำนวน 69,712 รายการ ใช้วิธีการวัดรากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย สำหรับการเปรียบเทียบประสิทธิภาพของแบบจำลอง และ 2) ชุดข้อมูลทดสอบ (Testing Data Set) ใช้ข้อมูลในปี พ.ศ. 2562 จำนวน 19,962 รายการ โดยใช้วิธีการประมาณความแม่นยำในการพยากรณ์ด้วยค่าความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error: MRE) สำหรับทดสอบประสิทธิภาพของแบบจำลอง โดยแยกชุดข้อมูลจำนวนการเกิดอุบัติเหตุในประเทศไทยเป็นรายเดือนและใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error: MMRE) เพื่อใช้ในการทดสอบประสิทธิภาพโดยรวมของแบบจำลองการพยากรณ์ (Weiss & Indurkha, 1998)

2. การวิเคราะห์ข้อมูล (Data Analysis) ผู้วิจัยเลือกใช้โปรแกรม WEKA Version 3.8.4 เพื่อทำการวิเคราะห์และสร้างแบบจำลองการพยากรณ์การเกิดอุบัติเหตุในประเทศไทย โดยวิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล โดยวิธีที่ผู้วิจัยใช้จะได้ผลลัพธ์ในรูปแบบของโมเดล ซึ่งถือเป็นลักษณะของการแทนความรู้ (Knowledge Representation) แบบหนึ่ง สำหรับรูปแบบการพยากรณ์จะมีลักษณะเป็นแบบ Sliding Window ซึ่งเป็นการจัดเรียงข้อมูลแบบอนุกรมเวลา โดยสร้างชุดข้อมูลสำหรับการสร้างแบบจำลองการพยากรณ์ โดยผู้วิจัยนำข้อมูลปี พ.ศ. 2558-2561 เป็นชุดข้อมูลทดสอบแบ่งเป็น ชุดข้อมูลย้อนหลัง (Lagged) ดังนี้ 1) 3 เดือน 2) 6 เดือน 3) 9 เดือน 4) 12 เดือน และ 5) 15 เดือน โดยผลลัพธ์ในการทดสอบจะใช้ข้อมูลทั้ง 3 ชุด จะเป็นตัวชี้วัดว่าอัลกอริทึมใดเหมาะสมกับการใช้ในการสร้างแบบจำลองให้กับจำนวนครั้งของการเกิดอุบัติเหตุ สำหรับปี พ.ศ. 2562 เป็นค่าเป้าหมาย (Target) ในการพยากรณ์ประสิทธิภาพ ประสิทธิภาพของการพยากรณ์ในแต่ละอัลกอริทึมจะเป็นดัชนีชี้วัดว่าอัลกอริทึมใดเหมาะสมกับการพยากรณ์

ผลและวิจารณ์

ผู้วิจัยได้ดำเนินการวิจัยโดยใช้กระบวนการ CRIS-DM เพื่อนำองค์ความรู้ร่วมกับการใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล 6 ขั้นตอนดังนี้ 1) Business Understanding เข้าใจปัญหาจำนวนการเกิดอุบัติเหตุในประเทศไทยเพิ่มขึ้น 2) Data Understanding กำหนดจำนวนแอททริบิวต์ที่ใช้ในการวิเคราะห์ข้อมูล 3) Data Preparation นำข้อมูลบางส่วนที่ไม่จำเป็นออก และแปลงผลข้อมูลเพื่อใช้ในโปรแกรม WEKA 4) Modeling ผู้วิจัยเลือกใช้ Time Series Model และเลือกเทคนิคที่เหมาะสมในการพยากรณ์จำนวนครั้งในการเกิดอุบัติเหตุในแต่ละจังหวัด 5) Evaluation ผู้วิจัยนำผลที่ได้มาเปรียบเทียบประสิทธิภาพในการหาค่าความคลาดเคลื่อนสัมพัทธ์จากชุดข้อมูลทดสอบ 6) Deployment ผลลัพธ์ที่ได้ในการวิจัยนี้ ถือเป็นองค์ความรู้ต่อกับผู้วิจัยเพื่อลดการเกิดอุบัติเหตุ (Shearer, 2000) ผลการทดลองได้ทำการวัดประสิทธิภาพแบบจำลอง เพื่อใช้ในการทำนายจำนวนครั้งของการเกิดอุบัติเหตุในประเทศไทย ตามวิธีดำเนินการวิจัยโดยใช้ 3 เทคนิค กับชุดข้อมูลฝึกฝนจำนวน 3 ชุด ตามขั้นตอนวิธีวิจัย ผลการทดสอบสามารถแบ่งออกเป็น 2 ส่วน ดังนี้

1. การเปรียบเทียบประสิทธิภาพของแบบจำลองทำนายข้อมูล

จากชุดข้อมูลที่ผู้วิจัยเลือกใช้จำนวนอุบัติเหตุในประเทศไทย จำนวน 3 แบบ โดยผู้วิจัยนำข้อมูลในปี พ.ศ. 2558-2561 เป็นชุดข้อมูลทดสอบ โดยแบ่งเป็นชุดข้อมูลย้อนหลัง ดังนี้ 1) 3 เดือน 2) 6 เดือน 3) 9 เดือน 4) 12 เดือน และ 5) 15 เดือน นำมาสร้างแบบจำลองในการทำนายจำนวนการเกิดอุบัติเหตุในประเทศไทย เพื่อวิเคราะห์ประสิทธิภาพที่ได้จากเทคนิคเหมืองข้อมูลทั้ง 3 แบบ โดยวิธีการเปรียบเทียบความคลาดเคลื่อนด้วยวิธี MAE และ RMSE จากผลการทดลองสามารถแสดงได้ดังตารางที่ 1

ตารางที่ 1 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายโดยใช้ชุดข้อมูลย้อนหลัง

ชุดข้อมูล	Times Series Data Mining Techniques						
	Quarter Lagged	Linear Regression		Multilayer Perceptron		SMOreg	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
DS#1 กรุงเทพมหานคร	Lag 3	25.3595	31.4837	24.1875	31.7481	23.1060	32.3655
	Lag 6	23.9642	31.2691	23.1981	30.2928	22.6707	32.1750
	Lag 9	23.2909	30.7326	14.6770	18.2800	21.3510	31.7133
	Lag 12	22.3543	29.5351	23.6549	26.9246	20.5477	31.9538
	Lag 15	21.2826	28.9280	5.0467	5.6356	16.1304	28.9812
DS#2 นครราชสีมา	Lag 3	16.5761	23.0349	16.3677	23.2069	15.3345	23.4981
	Lag 6	16.4682	22.3860	11.4835	15.1834	14.1193	24.2295
	Lag 9	14.5284	21.1858	10.0655	12.8791	11.9825	21.3645
	Lag 12	12.1538	15.9008	6.3449	9.7336	9.4031	14.8956
	Lag 15	9.3619	12.6083	16.1353	17.5767	9.0526	14.0644
DS#3 สมุทรปราการ	Lag 3	11.6171	14.2276	20.0634	23.3269	11.0212	15.3485
	Lag 6	10.5827	12.4649	14.0605	16.9874	9.6779	13.8763
	Lag 9	9.8043	12.1207	9.1037	10.7375	8.0492	12.2675
	Lag 12	8.6799	10.6902	3.8470	4.1481	6.7224	10.1305
	Lag 15	12.0781	15.0877	4.0715	5.0058	9.8990	17.5671

จากตารางที่ 1 แสดงผลการทำนายของแต่ละเทคนิคกับการใช้ชุดข้อมูลโดยมีจำนวนวันย้อนหลังที่แตกต่างกันเมื่อใช้ค่า MAE และ RMSE เป็นดัชนีชี้วัดประสิทธิภาพ จากการทดลองแสดงให้เห็นว่า จังหวัดกรุงเทพมหานคร (DS#1) และจังหวัดนครราชสีมา (DS#2) เมื่อสร้างแบบจำลองด้วยเทคนิคการถดถอยเชิงเส้น

และแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ด้วยข้อมูลย้อนหลัง 15 เดือน มีประสิทธิภาพสูงสุด เมื่อสร้างแบบจำลองโครงข่ายประสาทเทียม จังหวัดกรุงเทพมหานคร (DS#1) ด้วยข้อมูลย้อนหลัง 15 เดือน มีประสิทธิภาพสูงสุด จังหวัดนครราชสีมา (DS#2) ด้วยข้อมูลย้อนหลัง 12 เดือน มีประสิทธิภาพสูงสุด จังหวัดสมุทรปราการ (DS#3) เมื่อสร้างแบบจำลองด้วยเทคนิคการถดถอยเชิงเส้น แบบจำลองโครงข่ายประสาทเทียม และแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ด้วยข้อมูลย้อนหลัง 12 เดือน มีประสิทธิภาพสูงสุด

2. การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายในแต่ละเดือน

จากการนำแบบจำลองการพยากรณ์การเกิดอุบัติเหตุในประเทศไทยปี พ.ศ. 2562 โดยแยกจำนวนอุบัติเหตุในแต่ละจังหวัดที่เกิดอุบัติเหตุมากที่สุด 3 จังหวัด ได้แก่ จังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ จำนวน 12 เดือน โดยวิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล โดยใช้อัลกอริทึม 3 แบบ และคำนวณหาค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ แสดงผลดังตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายการเกิดอุบัติเหตุในจังหวัดกรุงเทพมหานคร

ชุดข้อมูล			Time Series Data Mining Techniques					
ชนิด	ข้อมูล 2562		LinearRegression		MultilayerPerceptron		SMOreg	
			Lag 15		Lag 15		Lag 15	
	Month	Actual	Estimate	MRE	Estimate	MRE	Estimate	MRE
DS#1 กรุงเทพมหานคร	มกราคม	179	167.2591	0.0656	199.4760	0.1144	152.7153	0.1468
	กุมภาพันธ์	158	138.4960	0.1234	162.7119	0.0298	132.2732	0.1628
	มีนาคม	156	141.5145	0.0929	113.2348	0.2741	168.7776	0.0819
	เมษายน	171	150.1414	0.1220	232.4329	0.3593	134.2215	0.2151
	พฤษภาคม	133	123.4796	0.0716	-17.7209	1.1332	76.6325	0.4238
	มิถุนายน	152	113.8075	0.2513	66.6452	0.5615	93.0401	0.3879
	กรกฎาคม	169	134.5043	0.2041	231.1574	0.3678	152.3655	0.0984
	สิงหาคม	115	141.2482	0.2282	226.6213	0.9706	172.6098	0.5010
	กันยายน	180	133.6049	0.2578	8.3023	0.9539	169.9119	0.0560
	ตุลาคม	176	138.8137	0.2113	272.3725	0.5476	175.4370	0.0032
	พฤศจิกายน	113	144.6901	0.2804	229.2433	1.0287	153.4049	0.3576
	ธันวาคม	141	134.4474	0.0465	52.7832	0.6257	127.9575	0.0925
	MMRE		16.29 %		58.055 %		21.059 %	

จากตารางที่ 2 พบว่าการสร้างแบบจำลองเทคนิคการถดถอยเชิงเส้น ให้ค่าประสิทธิภาพสูงสุด ได้แก่ ทำนายการเกิดอุบัติเหตุในจังหวัดกรุงเทพมหานคร มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับร้อยละ 16.29

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายการเกิดอุบัติเหตุในจังหวัดนครราชสีมา

ชุดข้อมูล			Time Series Data Mining Techniques					
ชนิด	ข้อมูล 2562		LinearRegression		MultilayerPerceptron		SMOreg	
			Lag 15		Lag 12		Lag 15	
	Month	Actual	Estimate	MRE	Estimate	MRE	Estimate	MRE
DS# 2 นครราชสีมา	มกราคม	114	93.6736	0.1783	79.8611	0.2995	78.5660	0.3108
	กุมภาพันธ์	82	92.0591	0.1227	70.8074	0.1365	113.2822	0.3815
	มีนาคม	81	65.0593	0.1968	28.5700	0.6473	83.2252	0.0275
	เมษายน	137	158.0456	0.1536	130.9330	0.0443	152.3285	0.1119
	พฤษภาคม	55	69.3462	0.2608	60.2133	0.0948	88.2911	0.6053
	มิถุนายน	52	25.6725	0.5063	118.5349	1.2795	59.8094	0.1502
	กรกฎาคม	61	73.1044	0.1984	113.3657	0.8585	100.1869	0.6424
	สิงหาคม	72	51.4639	0.2852	57.5267	0.2010	77.3357	0.0741
	กันยายน	36	42.4853	0.1801	149.1951	3.1443	67.9678	0.8880
	ตุลาคม	52	50.0123	0.0382	15.5568	0.7008	81.8541	0.5741
	พฤศจิกายน	51	45.8690	0.1006	131.0356	1.5693	73.0842	0.4330
	ธันวาคม	84	91.8988	0.0940	20.4183	0.7569	120.9176	0.4395
	MMRE		19.29 %		81.11 %		38.65 %	

จากตารางที่ 3 พบว่าการสร้างแบบจำลองเทคนิคการถดถอยเชิงเส้น ให้ค่าประสิทธิภาพสูงสุด ได้แก่ จังหวัดนครราชสีมา มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับร้อยละ 19.29

ตารางที่ 4 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายการเกิดอุบัติเหตุในจังหวัดสมุทรปราการ

ชุดข้อมูล			Time Series Data Mining Techniques					
ชนิด	ข้อมูล 2562		LinearRegression		MultilayerPerceptron		SMOreg	
			Lag 12		Lag 12		Lag 12	
	Month	Actual	Estimate	MRE	Estimate	MRE	Estimate	MRE
DS#3 สมุทรปราการ	มกราคม	45	51.1582	0.1368	60.1019	0.3356	52.1808	0.1596
	กุมภาพันธ์	45	45.9277	0.0206	80.1233	0.7805	55.6324	0.2363
	มีนาคม	48	54.5255	0.1359	56.4569	0.1762	56.9455	0.1864
	เมษายน	53	56.2212	0.0608	64.5971	0.2188	65.9188	0.2438
	พฤษภาคม	57	54.8048	0.0385	52.4557	0.0797	51.7606	0.0919
	มิถุนายน	60	59.1507	0.0142	67.7978	0.1300	67.3527	0.1225
	กรกฎาคม	71	59.9341	0.1559	39.2373	0.4474	50.5784	0.2876
	สิงหาคม	27	62.4871	1.3143	6.2812	0.7674	61.8422	1.2905
	กันยายน	60	61.2553	0.0209	-25.7349	1.4289	58.2862	0.0286
	ตุลาคม	72	51.9534	0.2784	2.7144	0.9623	54.3256	0.2455
	พฤศจิกายน	46	50.3218	0.0940	41.3944	0.1001	47.5394	0.0335
	ธันวาคม	63	57.5669	0.0862	96.3043	0.5286	60.906	0.0332
	MMRE		19.64 %		49.63 %		24.66 %	

จากตารางที่ 4 พบว่าการสร้างแบบจำลองเทคนิคการถดถอยเชิงเส้น ให้ค่าประสิทธิภาพสูงสุด ได้แก่ จังหวัดสมุทรปราการ มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับร้อยละ 19.64

สรุป

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพ และเลือกเทคนิคที่เหมาะสม ในการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมือนข้อมูลของการเกิดอุบัติเหตุในประเทศไทย โดยใช้เทคนิคเหมือนข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้นและซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยแบ่งชุดข้อมูลออกเป็น 2 ส่วน คือ ชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบ จากนั้นผู้วิจัยทำการทดสอบเพื่อหาค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ โดยเลือกชุดข้อมูลย้อนหลัง ที่มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ที่น้อยที่สุด เพื่อเลือกเทคนิคที่เหมาะสมและแม่นยำในการพยากรณ์อุบัติเหตุ อย่างไรก็ตามควรคำนึงถึงปัจจัยอื่นมาใช้ในการพยากรณ์ ได้แก่ บริเวณที่เกิดอุบัติเหตุ ทางตรง ทางโค้ง ทางแยก เส้นทางที่เกิดอุบัติเหตุ เช่น ทางหลวงแผ่นดินหมายเลข 1 และช่วงเวลาที่เกิดอุบัติเหตุ เพื่อเป็นแนวทางในการป้องกัน และลดการเกิดอุบัติเหตุ

เอกสารอ้างอิง

- กระทรวงคมนาคม. #อัตราการเสียชีวิตจากการบาดเจ็บทางถนน กระทรวงคมนาคม, เข้าถึงเมื่อ 12 ธันวาคม 2563, สืบค้นจาก <http://appgis.mot.go.th/trams>
- วีรศักดิ์ ฟองเงิน, วรปภา อารีราษฎร์ และเผด็จ พรหมสาขา ณ สกลนคร. (2560). การพยากรณ์ปริมาณน้ำในเขื่อน โดยใช้เทคนิคเหมือนข้อมูล. *วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม*, 4(1), 27-33.
- สุภาพร บรรดาศักดิ์, จุฑามาศ แป้นทอง, & นุตยา กัลลา. (2559). การพยากรณ์การใช้ลูกกอล์ฟในสนามฝึกซ้อมกอล์ฟโดยใช้ตัวแปรอนุกรมเวลา. การประชุมสวนสุนันทาวิชาการระดับชาติด้าน “การวิจัยเพื่อการพัฒนาอย่างยั่งยืน” ครั้งที่ 4. วันที่ 26 สิงหาคม 2559. อาคารศูนย์วิทยาศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา. กรุงเทพฯ. หน้า 338-346.
- สถาบันวิจัยเพื่อการพัฒนาประเทศไทย (ทีดีอาร์ไอ). #อุบัติเหตุบนถนน, เข้าถึงเมื่อวันที่ 10 ธันวาคม 2563, สืบค้นจาก https://tdri.or.th/2017/08/econ_traffic_accidents/
- Hoshmand, AR. (2009). *Business Forecasting* (2nd ed.). New York: Routledge.
- Kahani, M., Ahmadi, M.H., Tatar, A., & Sadeghzadeh, M. (2018). Development of multilayer perceptron artificial neural network (MLP-ANN) and least square support vector machine (LSSVM) models to predict Nusselt number and pressure drop of TiO₂/water nanofluid flows through non-straight pathways. *Numerical Heat Transfer, Part A: Applications*, 74(4), 1190-1206.
- Kavitha, S., Varuna, S., & Ramya, R. (2016). Comparative Analysis on Linear Regression and Support Vector Regression. *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, 1188 – 1193. doi: 10.1109/72.870050
- Lin, Y., Hong, L. Yi, L., Xinxin L., & Muhammad W.A. (2015). Biomedical Named Entity Recognition based on Deep Neutral Network. *International Journal of Hybrid Information Technology*, 8(8), 279-288.

- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5(4), 13-22.
- Shevade, S.K., Keerthi S.S., Bhattacharyya, C., & Murthy K.R.K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Trans Neural Networks*, 11(5), 1188-1193.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222.
- Weiss, S.M., & Indurkha, N. (1998). *Predictive Data Mining: A Practical Guide*. San Francisco: Morgan Kaufmann.