

ตัวแบบผสมอาร์มา-ซัพพอร์ตเวกเตอร์แมชชีนเน้นองค์ประกอบเชิงบวก สำหรับการพยากรณ์ปริมาณฝุ่นละอองขนาดไม่เกิน 10 ไมครอน

Hybrid ARIMA-SVM Model Emphasizing on Additive Components for PM₁₀ Forecasting

ธรณินทร์ สัจจวิริยทรัพย์*

สาขาวิศวกรรมโลจิสติกส์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย

แขวงรัชดาภิเษก เขตดินแดง กรุงเทพมหานคร 10400

Thorarin Sujjaviriyasup*

Logistics Engineering, School of Engineering, University of the Thai Chamber of Commerce,

Ratchadaphisek, Din Daeng, Bangkok 10400

บทคัดย่อ

สำหรับงานวิจัยนี้ ตัวแบบผสมระหว่างตัวแบบอาร์มาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนพัฒนาขึ้นเพื่อพยากรณ์ปริมาณฝุ่นละอองที่มีขนาดไม่เกิน 10 ไมครอน ในเขตพื้นที่ตำบลหน้าพระลาน อำเภอเฉลิมพระเกียรติ จังหวัดสระบุรี ซึ่งตั้งอยู่บนสมมติฐานของข้อมูลอนุกรมเวลาของปริมาณฝุ่นละอองขนาดเล็ก มีองค์ประกอบทั้งแบบเชิงเส้นตรงและแบบไม่เป็นเชิงเส้นตรง ที่มีความสัมพันธ์ระหว่าง 2 องค์ประกอบ เป็นแบบเชิงบวก โดยอาศัยคุณลักษณะเด่นในการพยากรณ์ข้อมูลอนุกรมเวลาที่มีความสัมพันธ์เชิงเส้นตรงของตัวแบบอาร์มาในการอธิบายองค์ประกอบเชิงเส้นตรงของข้อมูลอนุกรมเวลา รวมทั้งคุณลักษณะเด่นของตัวแบบซัพพอร์ตเวกเตอร์แมชชีนที่สามารถสร้างฟังก์ชันการทำนายที่ซับซ้อน และพยากรณ์ข้อมูลอนุกรมเวลาที่มีความสัมพันธ์แบบไม่เป็นเชิงเส้นตรงได้ดีในการอธิบายองค์ประกอบที่ไม่เป็นเชิงเส้นตรงของข้อมูลอนุกรมเวลา นอกจากนี้ตัวแบบผสมที่พัฒนาขึ้นได้นำมาเปรียบเทียบกับประสิทธิภาพการพยากรณ์กับตัวแบบพยากรณ์เชิงเดี่ยวทั้งตัวแบบอาร์มาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนด้วยเกณฑ์การประเมินประสิทธิภาพการพยากรณ์ทั้งหมด 6 เกณฑ์ จากผลการวิจัยบ่งชี้ว่าตัวแบบซัพพอร์ตเวกเตอร์แมชชีนที่สร้างจากความสัมพันธ์แบบไม่เป็นเชิงเส้นตรงด้วยฟังก์ชันเรเดียลเบสิคให้ผลการพยากรณ์ที่ดีกว่าตัวแบบอาร์มาเพียงเล็กน้อย แต่ตัวแบบผสมกลับให้ผลการพยากรณ์ที่มีความถูกต้องเพิ่มมากขึ้นกว่าตัวแบบเชิงเดี่ยวทั้งสองตัวแบบ ซึ่งแสดงให้เห็นว่าข้อมูลอนุกรมเวลาของปริมาณฝุ่นละอองขนาดเล็กอาจไม่มีความสัมพันธ์แบบเชิงเส้นตรงหรือแบบไม่เป็นเชิงเส้นตรงอย่างชัดเจน การเลือกใช้ตัวแบบผสมจะสามารถช่วยลดความเสี่ยงของการเลือกใช้ตัวแบบที่ไม่เหมาะสมกับชุดข้อมูลลงได้ ดังนั้นตัวแบบผสมระหว่างตัวแบบอาร์มาและตัวแบบซัพพอร์ต

*ผู้รับผิดชอบบทความ : thorarin_suj@utcc.ac.th

เวกเตอร์แมชชีนจึงเป็นตัวแทนที่มีความเหมาะสมที่จะใช้สำหรับการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในเขตพื้นที่ตำบลหน้าพระลาน เพื่อใช้เป็นสารสนเทศที่เป็นประโยชน์ต่อการเฝ้าระวังและวางแผนเพื่อลดผลกระทบจากมลพิษทางฝุ่นละอองลงอย่างมีประสิทธิภาพ ซึ่งไม่เพียงแต่เป็นงานที่ทำหายแต่ยังเป็นปัญหาสาธารณสุขที่สำคัญสำหรับสุขภาพของประชาชนอีกด้วย รวมทั้งยังเป็นตัวแทนที่อาจมีประโยชน์ในการประยุกต์ใช้สำหรับการพยากรณ์ปริมาณฝุ่นละอองในเขตพื้นที่อื่น ๆ ได้

คำสำคัญ : ตัวแบบผสม; การพยากรณ์มลพิษทางอากาศ; อารีมา; ซัพพอร์ตเวกเตอร์แมชชีน; ฝุ่นละออง

Abstract

In this research, a hybrid model of ARIMA and SVM is developed and proposed for PM₁₀ forecasting in Na Phra Lan sub-district, Chaloe Phra Kiat district, Saraburi province based on an assumption of additive components. In this regard, the ARIMA model, which has dominated in linear forecasting problems, is used to predict linear components of the PM₁₀ time series. Whilst, the SVM model is a well-known approach with superior performance of nonlinear forecasting model, which is used to formulate nonlinear and complex function in order to predict nonlinear components of the PM₁₀ time series. In addition, the developed model is compared to both ARIMA model and SVM model based on six accuracy measures. The empirical results revealed that the SVM model based on radial kernel function slightly outperforms ARIMA model. In contrast, the developed hybrid model can provide lower errors than both ARIMA and SVM models. The comparison between ARIMA model and SVM model indicated that the PM₁₀ time series may not be explicit linear or nonlinear patterns. Subsequently, the developed hybrid model can reduce a risk of using improper forecasting model. Consequently, the hybrid ARIMA-SVM model can be a promising tool for PM₁₀ forecasting in other area in order to provide useful information to monitor and make a critical decision on air quality that is not only a very challenging task but also a public concerned problem for human health.

Keywords: combined model; air pollution forecasting; ARIMA; support vector machine; particulate matter

1. บทนำ

การเติบโตทางเศรษฐกิจของประเทศไทยในช่วงหลายสิบปีที่ผ่านมา ส่งผลให้เกิดการขยายตัวของภาคอุตสาหกรรมและการจราจรที่คับคั่งเพิ่มมากขึ้น ซึ่งก่อให้เกิดมลพิษทางสิ่งแวดล้อมอย่างหลีกเลี่ยง

ไม่ได้ มลพิษทางอากาศซึ่งมีสาเหตุจากฝุ่นละอองขนาดเล็ก (ขนาดเล็กกว่า 10 ไมครอน) ถือเป็นหนึ่งในปัญหาด้านสิ่งแวดล้อมที่สำคัญและมีผลกระทบต่อคุณภาพชีวิตของประชาชนในพื้นที่ [1,2] เนื่องจากฝุ่นละอองที่มีขนาดเล็กจะลอยฟุ้งอยู่ในอากาศเป็นเวลานาน จึง

สามารถปนเปื้อนกับอากาศและเข้าสู่ระบบหายใจของประชาชนในพื้นที่ ส่งผลให้เกิดโรคทางเดินหายใจและโรคอื่น ๆ ตามมา

จังหวัดสระบุรีเป็นหนึ่งพื้นที่ซึ่งประกอบไปด้วยโรงงานอุตสาหกรรมและการจราจรที่คับคั่ง [1,2] จึงก่อให้เกิดมลพิษทางอากาศจากฝุ่นละอองขนาดเล็ก และมีค่าปริมาณฝุ่นละอองที่เกินเกณฑ์มาตรฐานมากที่สุดของประเทศในปี พ.ศ. 2558 โดยเฉพาะอย่างยิ่งในเขตพื้นที่ตำบลหน้าพระลาน อำเภอเฉลิมพระเกียรติ ซึ่งถือเป็นพื้นที่เฝ้าระวังด้านมลพิษทางอากาศจากฝุ่นละอองขนาดเล็ก

เพื่อเฝ้าระวังและลดผลกระทบของมลพิษอันเนื่องมาจากฝุ่นละอองดังที่กล่าว [3-6] การบริหารจัดการที่เหมาะสมจึงเป็นขั้นตอนการดำเนินการที่สำคัญ ซึ่งต้องอาศัยการวางแผนและจัดเตรียมอุปกรณ์ที่เหมาะสมและเพียงพอกับปริมาณฝุ่นละอองในแต่ละสถานการณ์ แต่ปริมาณฝุ่นละอองขนาดเล็กในแต่ละเดือนมีปริมาณที่ไม่แน่นอนและยากต่อการคาดการณ์ ด้วยวิธีการอย่างง่าย จึงส่งผลให้การพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กถือเป็นสารสนเทศที่สำคัญ เพื่อประกอบการตัดสินใจและวางแผนในการควบคุมและลดผลกระทบอย่างมีประสิทธิภาพมากที่สุด

เทคนิคการพยากรณ์อนุกรมเวลาเป็นหนึ่งในเครื่องมือที่ได้รับความนิยมและใช้ในการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กอย่างแพร่หลายในช่วงหลายปีที่ผ่านมา ซึ่งสามารถแบ่งออกได้เป็น 3 กลุ่มหลัก คือ ตัวแบบพยากรณ์ทางสถิติ ตัวแบบการเรียนรู้ของเครื่อง (machine learning) และตัวแบบผสม

ตัวแบบอาร์มีมาเป็นหนึ่งในตัวแบบพยากรณ์ทางสถิติที่ได้รับการยอมรับถึงประสิทธิภาพการพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์เชิงเส้นตรง จึงนิยมประยุกต์ใช้ในการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กอย่างกว้างขวาง [7,8] อย่างไรก็ตาม จุดด้อยของตัว

แบบอาร์มีมาที่สร้างจากความสัมพันธ์เชิงเส้นตรงอาจไม่สามารถอธิบายข้อมูลอนุกรมเวลาที่มีคุณลักษณะความสัมพันธ์แบบไม่เป็นเชิงเส้นตรงได้ นอกจากนี้วิธีการวิเคราะห์คุณลักษณะของข้อมูลอนุกรมเวลายังไม่สามารถระบุความสัมพันธ์ที่ชัดเจนได้ จึงส่งผลให้ตัวแบบอาร์มีมายังคงได้รับความสนใจและประยุกต์ใช้ในการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในปัจจุบัน แต่จากข้อกำหนดด้านความสัมพันธ์เชิงเส้นตรงรวมทั้งสมมติฐานในการสร้างตัวแบบทางสถิติที่ค่อนข้างมีมาก จึงได้มีการพัฒนาตัวแบบการเรียนรู้ของเครื่อง ซึ่งมียุคคุณลักษณะที่โดดเด่นในการพยากรณ์ข้อมูลอนุกรมเวลาที่มีความสัมพันธ์ไม่เป็นเชิงเส้นตรงและเป็นตัวแบบที่สร้างความสัมพันธ์จากคุณลักษณะของข้อมูล (data-driven model) จึงไม่จำเป็นต้องพิจารณาข้อกำหนดในการสร้างตัวแบบมากเหมือนกับตัวแบบอาร์มีมา

ตัวแบบซัพพอร์ตเวกเตอร์แมชชีนเป็นหนึ่งในตัวแบบพยากรณ์ทางการเรียนรู้ของเครื่อง ซึ่งใช้หลักการลดความเสี่ยงเชิงโครงสร้างให้น้อยที่สุด (principle of structural risk minimization) จากหลักการนี้ตัวแบบซัพพอร์ตเวกเตอร์ที่ได้จะเป็นตัวแบบที่เหมาะสมที่สุดและให้ผลการพยากรณ์เพียงค่าเดียว ส่งผลให้ตัวแบบซัพพอร์ตเวกเตอร์ได้รับความนิยมและประยุกต์ใช้อย่างแพร่หลาย [9-12] อย่างไรก็ตาม ประสิทธิภาพของตัวแบบซัพพอร์ตเวกเตอร์แมชชีนจะขึ้นอยู่กับทางเลือกค่าพารามิเตอร์ที่เหมาะสมสำหรับการสร้างตัวแบบ ดังนั้นการเลือกค่าพารามิเตอร์ที่เหมาะสมจึงเป็นหนึ่งในประเด็นที่สำคัญสำหรับการสร้างตัวแบบซัพพอร์ตเวกเตอร์แมชชีน โดยหนึ่งในวิธีการหาค่าพารามิเตอร์ที่เหมาะสมสำหรับตัวแบบซัพพอร์ตเวกเตอร์แมชชีนด้วยการอาศัยการค้นหาแบบกริด (grid search)

แต่จากสถานการณ์ของความซับซ้อนของข้อมูล

ในปัจจุบัน ตัวแบบการพยากรณ์เชิงเดี่ยวอาจไม่สามารถอธิบายพฤติกรรมของข้อมูลที่มีความซับซ้อนได้ครบถ้วน รวมทั้งอาจเกิดความเสี่ยงสูงในการพยากรณ์ที่ผิดพลาด หากข้อมูลไม่มีคุณลักษณะความสัมพันธ์เป็นแบบเชิงเส้นตรงหรือไม่เป็นเชิงเส้นตรงเพียงอย่างเดียวอย่างชัดเจน นอกจากนี้ยังไม่มีการวิเคราะห์ลักษณะความสัมพันธ์ของข้อมูลอนุกรมเวลายังไม่สามารถระบุได้อย่างชัดเจน จากที่กล่าวมานั้น ในการพัฒนาตัวแบบพยากรณ์ที่มีคุณลักษณะที่สามารถอธิบายข้อมูลอนุกรมเวลาที่มีองค์ประกอบทั้งแบบเชิงเส้นตรงและไม่เป็นเชิงเส้นตรง จึงเป็นแนวคิดที่น่าสนใจ [13-18] โดยตั้งอยู่บนสมมติฐานของความสัมพันธ์ระหว่างองค์ประกอบเชิงเส้นตรงและไม่เป็นเชิงเส้นตรงเป็นความสัมพันธ์เชิงบวก ซึ่งเป็นหนึ่งในวิธีการสร้างตัวแบบพยากรณ์แบบผสม เพื่อให้สามารถอธิบายคุณลักษณะของข้อมูลที่มีความซับซ้อนในปัจจุบันรวมทั้งลดโอกาสเสี่ยงของการใช้ตัวแบบที่ไม่เหมาะสม

งานวิจัยนี้ ตัวแบบผสมระหว่างตัวแบบอาร์มาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนได้พัฒนาขึ้น โดยมีวัตถุประสงค์ที่จะใช้ความสามารถของตัวแบบอาร์มาซึ่งมีความโดดเด่นในการพยากรณ์ข้อมูลที่มีความสัมพันธ์เชิงเส้นตรงสำหรับองค์ประกอบเชิงเส้นตรงของข้อมูลอนุกรมเวลา และใช้ความสามารถของตัวแบบซัพพอร์ตเวกเตอร์แมชชีน ซึ่งมีความโดดเด่นในการสร้างความสัมพันธ์ไม่เป็นเชิงเส้นตรงและเป็นความสัมพันธ์ที่ซับซ้อนสำหรับองค์ประกอบไม่เป็นเชิงเส้นตรงของข้อมูลอนุกรมเวลา โดยตั้งอยู่บนสมมติฐานของความสัมพัทธ์ระหว่างองค์ประกอบเชิงเส้นตรงและไม่เป็นเชิงเส้นตรงแบบเชิงบวก เพื่อพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในเขตพื้นที่ตำบลหน้าพระลาน อำเภอเฉลิมพระเกียรติ จังหวัดสระบุรี ซึ่งเป็นพื้นที่เฝ้าระวังด้านมลพิษทางฝุ่นละอองและเปรียบเทียบประสิทธิภาพความถูกต้องของการพยากรณ์กับทั้งตัวแบบอาร์

มาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนด้วยเกณฑ์การประเมินทั้งหมด 6 เกณฑ์ ได้แก่ ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ (mean absolute error) ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (root mean squared error) ค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์ (mean absolute percentage error) ค่ารากที่สองของค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนกำลังสอง (root mean square percentage error) ค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์แบบสมมาตร (systematic mean absolute percentage error) และค่าสัมประสิทธิ์การตัดสินใจ (R-squared) เพื่อค้นหาตัวแบบพยากรณ์ที่เหมาะสมที่สุดและสามารถนำเสนอสารสนเทศที่ได้จากตัวแบบพยากรณ์ประกอบการตัดสินใจและวางแผนสำหรับการเฝ้าระวังและลดผลกระทบที่จะเกิดกับประชาชนในพื้นที่ให้น้อยที่สุดอย่างมีประสิทธิภาพ

2. ข้อมูลและวิธีการวิจัย

2.1 ข้อมูลอนุกรมเวลาของปริมาณฝุ่นละอองขนาดไม่เกิน 10 ไมครอน ในเขตตำบลหน้าพระลาน

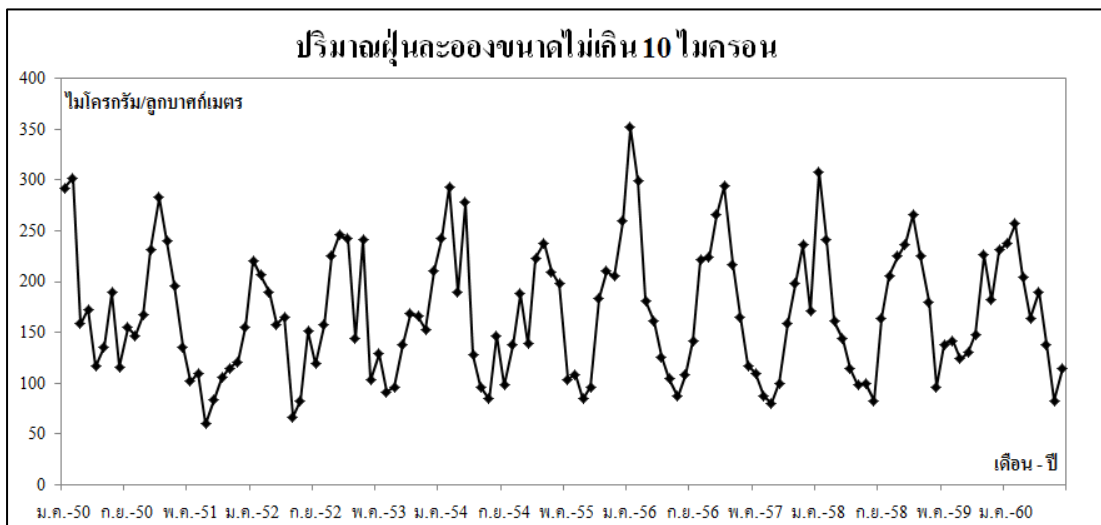
ข้อมูลอนุกรมเวลาที่ใช้ในการวิจัยครั้งนี้ [19] เป็นข้อมูลฝุ่นละอองขนาดเล็กกรายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2550 ถึงเดือนสิงหาคม พ.ศ. 2560 ซึ่งเป็นค่าสูงสุดของค่าเฉลี่ยราย 24 ชั่วโมง จากข้อมูลเว็บไซต์ของสำนักจัดการคุณภาพอากาศและเสียง กรมควบคุมมลพิษ จำนวน 128 ค่า โดยข้อมูลทั้งหมดเสนอในรูปที่ 1

2.2 ตัวแบบอาร์มา

ตัวแบบอาร์มาเป็นตัวแบบอนุกรมเวลาที่มีความโดดเด่นในการพยากรณ์ข้อมูลที่มีความสัมพันธ์เป็นแบบเชิงเส้นตรง ซึ่งเป็นรูปแบบทั่วไปของตัวแบบอาร์มา (ARMA) ในกรณีที่ข้อมูลอนุกรมเวลาไม่อยู่ใน

สภาวะคงที่ (non-stationary state) โดยมีสมการทางคณิตศาสตร์ของตัวแบบอาร์มาแบบทั้งที่มีอิทธิพลของ

ฤดูกาลและไม่มีอิทธิพลของฤดูกาล ดังแสดงในสมการที่ 1 และ 2 ตามลำดับ



รูปที่ 1 ค่าปริมาณฝุ่นละอองสูงสุดของค่าเฉลี่ย 24 ชั่วโมง ตั้งแต่เดือนมกราคม พ.ศ. 2550 ถึงเดือนสิงหาคม พ.ศ. 2560

$$\Phi_p(B^d) \phi_p(B) \nabla_s^d \nabla_t^d Y_t = \mu + \theta_q(B) \Theta_Q(B^d) \varepsilon_t \quad (1)$$

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) (1 - B)^d (Y_t - \mu) = \left(1 - \sum_{j=1}^q \theta_j B^j\right) \varepsilon_t \quad (2)$$

เมื่อ Y_t และ ε_t คือ ค่าจริงและค่าความคลาดเคลื่อน ณ เวลาที่ t ตามลำดับ; B และ μ คือ ตัวแปลงค่าอนุกรมเวลาย้อนกลับ (backward shift operator) และค่าเฉลี่ย ตามลำดับ; d และ D คือ ลำดับของผลต่างและลำดับของผลต่างฤดูกาล ตามลำดับ; p และ q คือ ลำดับของอาร์มาแบบไม่มีฤดูกาล; P และ Q คือ ลำดับของอาร์มาแบบมีฤดูกาล

สำหรับงานวิจัยนี้ ตัวแบบอาร์มาที่เหมาะสมที่สุดจะเลือกจากเกณฑ์อะไคอินฟอร์มชันคริติที่เรียกว่าแบบมีขนาดตัวอย่างแบบจำกัด (AICc) ที่ให้ค่าเกณฑ์อะไคอินฟอร์มชันคริติที่เรียกว่าแบบมีขนาดตัวอย่างแบบจำกัดที่ต่ำที่สุด ซึ่งเขียนขึ้นจากโปรแกรมภาษา R ด้วยไลบรารี “forecast” ของโปรแกรมภาษา R ที่ได้

พัฒนาขึ้น [20] เพื่อช่วยในการหาตัวแบบอาร์มาทั้งแบบที่มีอิทธิพลของฤดูกาลและไม่มีอิทธิพลของฤดูกาลที่เหมาะสมที่สุดแบบอัตโนมัติ ซึ่งได้อ้างอิงและประยุกต์ใช้งานในงานวิจัยที่ได้เผยแพร่ในวารสารที่เป็นที่ยอมรับต่าง ๆ [21-24] แต่เนื่องจากการหาตัวแบบอาร์มาที่เหมาะสมจะขึ้นอยู่กับจำนวนข้อมูลอนุกรมเวลาที่มีการเปลี่ยนแปลงในแต่ละช่วงเวลา จึงส่งผลให้ลำดับของตัวแบบอาร์มามีการปรับเปลี่ยนไปด้วย ดังนั้นจึงใช้สัญลักษณ์ ARIMA (ตัวแบบอาร์มาแบบไม่ระบุลำดับ) ในส่วนของผลการวิจัยเพื่อความกระชับในการนำเสนอ

2.3 ตัวแบบซัพพอร์ตเวกเตอร์แมชชีน

ตัวแบบซัพพอร์ตเวกเตอร์แมชชีนเป็นหนึ่งในตัวแบบการเรียนรู้ของเครื่องที่นำเสนอโดย Vapnik ซึ่งอาศัยหลักการเรียนรู้ทางสถิติและการลดความเสี่ยงเชิงโครงสร้างที่น้อยที่สุดในการจำแนกกลุ่ม นอกจากนี้

ตัวแบบซัพพอร์ตเวกเตอร์แมชชีนยังพัฒนาต่อเพื่อใช้ในการวิจัยหลายลักษณะ เช่น การพยากรณ์อนุกรมเวลา หรือสมการถดถอย สมการทางคณิตศาสตร์ของตัวแบบซัพพอร์ตเวกเตอร์แมชชีนสำหรับการพยากรณ์เชิงเส้นตรงดังสมการที่ 3

$$f(x) = (w, x) + b \quad x \in \mathcal{R}^d, \quad b \in \mathcal{R} \quad (3)$$

เมื่อ $f(x)$, w และ b คือ ค่าตัวแปรตาม เวกเตอร์ถ่วงน้ำหนัก (weight vector) และค่าเอนเอียง (bias) ตามลำดับ; (\cdot, \cdot) คือ การคูณแบบดอท (dot product) และ x คือ เวกเตอร์ของชุดข้อมูลป้อนเข้า (input data) ซึ่งมีขนาดเท่ากับ d ค่า โดยความเรียบของสมการที่ 3 สามารถทำให้เรียบที่สุดจากการใช้หลักการหาค่าความเหมาะสมที่สุด ดังนี้

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^T (\xi_i + \xi_i^*) \quad (4)$$

สมการข้อช่วย

$$y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i$$

$$(w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i$$

เมื่อ $\xi_i, \xi_i^*, C \geq 0$

โดยการเลือกค่า C จะมีผลต่อการพิจารณา ระหว่างค่าความเรียบและค่าของการเบี่ยงเบนที่ออกนอกขอบเขตที่กำหนด ดังนั้นการเลือกค่า C ที่เหมาะสมจะส่งผลให้ประสิทธิภาพการพยากรณ์ของซัพพอร์ตเวกเตอร์แมชชีนดีที่สุด และตัวแปร ξ_i และ ξ_i^* คือ ระยะห่างของจุดที่อยู่นอกขอบเขตที่กำหนด ซึ่งใช้เพื่อให้ครอบคลุมกรณีของค่าที่อยู่นอกขอบเขตที่กำหนด การหาค่าที่เหมาะสมที่สุดมีความซับซ้อนจึงต้องอาศัยวิธีการของลากรางจ์ (Lagrange multiplier) ในการหาค่าที่เหมาะสมในรูปแบบคู่ควบ (dual formation) ซึ่งซัพพอร์ตเวกเตอร์แมชชีนที่มีฟังก์ชันการทำนายเชิงเส้นตรง แสดงในรูปแบบของสมการที่ 5

$$f(x) = \sum_{i=1}^T (\alpha_i - \alpha_i^*) (x, x_i) + b \quad (5)$$

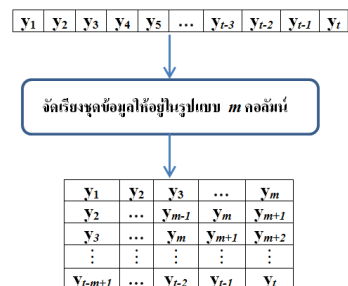
เมื่อ α_i และ α_i^* คือ ค่าของลากรางจ์ และ b คือ ค่าเอนเอียง

จากสมการที่ 5 ความสัมพันธ์มีลักษณะเป็นเชิงเส้นตรง แต่ปัญหาการพยากรณ์ส่วนมากมักพบลักษณะแบบไม่เป็นเชิงเส้นตรง การแปลงข้อมูลจากข้อมูลแบบไม่เป็นเชิงเส้นตรงไปสู่ข้อมูลที่เป็นเชิงเส้นตรงด้วยการใช้ฟังก์ชันเคอร์เนล (Kernel function) โดยฟังก์ชันเคอร์เนลสำหรับข้อมูลที่ไม่เป็นเชิงเส้นตรงคือ ฟังก์ชันเรเดียลเบสิก (radial basis function) ซึ่งเป็นฟังก์ชันที่มีประสิทธิภาพและนิยมใช้ในการพยากรณ์ข้อมูลที่ไม่เป็นเชิงเส้นตรง โดยมีสมการดังนี้

$$K(x, x_i) = \exp(-\sigma \|x - x_i\|^2) \quad (6)$$

เมื่อ σ คือ ค่าพารามิเตอร์ของฟังก์ชันเรเดียลเบสิก

สำหรับตัวแบบซัพพอร์ตเวกเตอร์แมชชีนที่ใช้ในงานวิจัยนี้ เขียนขึ้นจากโปรแกรมภาษา R ด้วยไลบรารี “kernlab” ของโปรแกรมภาษา R ที่ได้พัฒนาขึ้น ซึ่งมีการนำไปใช้ในงานวิจัยอย่างแพร่หลาย [25-27] โดยการจัดเรียงข้อมูลอนุกรมเวลาให้อยู่ในรูปแบบ m คอลัมน์ ดังรูปที่ 2 ซึ่งใช้ข้อมูลย้อนหลังตั้งแต่คอลัมน์ที่ 1 ถึงคอลัมน์ที่ $m-1$ เป็นข้อมูลป้อนเข้า และคอลัมน์ที่ m ซึ่งเป็นคอลัมน์สุดท้ายเป็นข้อมูลเป้าหมาย (target data) เพื่อการเรียนรู้ของซัพพอร์ตเวกเตอร์แมชชีนในการสร้างตัวแบบพยากรณ์



รูปที่ 2 การจัดรูปแบบข้อมูลสำหรับการพยากรณ์ด้วยตัวแบบซัพพอร์ตเวกเตอร์แมชชีน

ดังนั้นจึงใช้สัญลักษณ์ SVM(m) แทนตัวแบบซัพพอร์ตเวกเตอร์แมชชีนที่จัดเรียงใหม่ m คอลัมน์ใน ส่วนของผลการวิจัย เพื่อให้การนำเสนอผลอยู่ในรูปแบบที่กระชับ โดยจำนวนคอลัมน์มีค่าตั้งแต่ 2-14 คอลัมน์ ที่ใช้ในการพัฒนาตัวแบบพยากรณ์เพื่อหาตัวแบบที่เหมาะสมที่สุด นอกจากนี้การปรับค่าพารามิเตอร์ของซัพพอร์ตเวกเตอร์แมชชีนทำโดยการเพิ่มค่าที่ละ 0.001 จากค่า 0.001 ถึง 300 เพื่อให้ได้ตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์

2.4 ตัวแบบผสมระหว่างอาร์มีมาและซัพพอร์ตเวกเตอร์แมชชีน

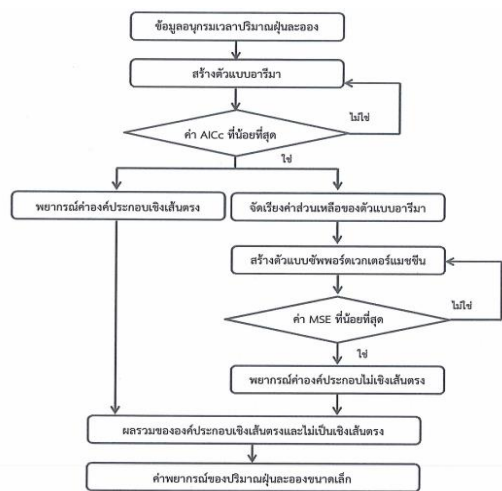
ความต้องการผลการพยากรณ์ที่มีความคลาดเคลื่อนที่ลดน้อยลง การพึ่งพาเทคนิคการพยากรณ์เชิงเดี่ยวอาจไม่สามารถอธิบายหรือพยากรณ์ข้อมูลที่มีความซับซ้อนในปัจจุบันได้ดีเพียงพอ จึงมีการนำเสนอแนวความคิดการผสมตัวแบบพยากรณ์อย่างน้อย 2 ตัวแบบ โดยอาศัยความโดดเด่นของแต่ละตัวแบบเพื่อให้ค่าการพยากรณ์ที่เหมาะสมขึ้น หนึ่งในตัวแบบผสมที่นำเสนอ คือ ตัวแบบผสมระหว่างตัวแบบอาร์มีมาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีน ซึ่งตั้งอยู่บนสมมติฐานของข้อมูลอนุกรมเวลาที่มีองค์ประกอบของข้อมูลเชิงเส้นตรงและไม่เป็นเชิงเส้นตรง โดยมีความสัมพันธ์ระหว่างองค์ประกอบของข้อมูลแบบเชิงบวก ซึ่งสามารถอธิบายด้วยสมการทางคณิตศาสตร์ ดังสมการที่ 7

$$Y_t = L_t + N_t + \varepsilon_t \tag{7}$$

เมื่อ Y_t และ ε_t คือ ค่าจริงและค่าความคลาดเคลื่อน ณ เวลาที่ t ตามลำดับ; L_t และ N_t คือ ข้อมูลองค์ประกอบเชิงเส้นตรงและไม่เป็นเชิงเส้นตรง ณ เวลาที่ t ตามลำดับ

จากความสัมพันธ์ในสมการที่ 7 ส่งผลให้การเลือกใช้ตัวแบบพยากรณ์ที่เหมาะสมกับลักษณะขององค์ประกอบของข้อมูลจะช่วยให้ผลการพยากรณ์ค่า

อนุกรมเวลาจริงมีความคลาดเคลื่อนลดน้อยลง แต่ในทางกลับกันหากใช้เพียงตัวแบบพยากรณ์เชิงเดี่ยวเพียงลักษณะใดลักษณะหนึ่ง จะส่งผลให้ข้อมูลในส่วนที่เหลือรวมเข้าในพจน์ของค่าความคลาดเคลื่อน ทำให้ค่าความคลาดเคลื่อนของการพยากรณ์ที่ได้มีความคลาดเคลื่อนสูงขึ้น เพื่ออธิบายการสร้างตัวแบบพยากรณ์แบบผสมในภาพรวม ขั้นตอนการดำเนินการได้แสดงในรูปที่ 3



รูปที่ 3 ขั้นตอนการสร้างตัวแบบผสมระหว่างตัวแบบอาร์มีมาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีน

นอกจากนี้ขั้นตอนการพัฒนาตัวแบบผสมระหว่างตัวแบบอาร์มีมาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนสามารถอธิบายได้ ดังนี้

- (1) ตัวแบบอาร์มีมาที่เหมาะสมที่สุดประยุกต์ใช้ในการอธิบายองค์ประกอบเชิงเส้นตรงของข้อมูลอนุกรมเวลา ด้วยการพยากรณ์ค่าองค์ประกอบเชิงเส้นตรงและนำผลการพยากรณ์ที่ได้หักลบออกจากข้อมูลอนุกรมเวลาจริง เพื่อให้เหลือเพียงค่าองค์ประกอบไม่เป็นเชิงเส้นตรงและค่าความคลาดเคลื่อนซึ่งสามารถอธิบายด้วยสมการทางคณิตศาสตร์ได้ดังสมการที่ 8 และ 9

$$\hat{L}_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_1) \quad (8)$$

$$Y_t - \hat{L}_t = N_t + \varepsilon_t \quad (9)$$

(2) ตัวแบบซัพพอร์ตเวกเตอร์แมชชีนที่เหมาะสมที่สุดถูกใช้ในการพยากรณ์ค่าองค์ประกอบไม่เป็นเชิงเส้นตรง เพื่ออธิบายองค์ประกอบไม่เป็นเชิงเส้นตรงของข้อมูลอนุกรมเวลาและเหลือเพียงค่าความคลาดเคลื่อนจริงเท่านั้น โดยอธิบายด้วยสมการทางคณิตศาสตร์ได้ดังสมการที่ 10 และ 11

$$\hat{N}_t = f((Y_{t-1} - \hat{L}_{t-1}), (Y_{t-2} - \hat{L}_{t-2}), (Y_{t-3} - \hat{L}_{t-3}), \dots, (Y_1 - \hat{L}_1)) \quad (10)$$

$$N_t - \hat{N}_t = \varepsilon_t \quad (11)$$

(3) ผลการพยากรณ์ทั้งองค์ประกอบเชิงเส้นตรงและไม่เป็นเชิงเส้นตรงของข้อมูลอนุกรมเวลาที่ได้จากข้อ 1 และ 2 สามารถสร้างผลการพยากรณ์เพื่อที่จะอธิบายข้อมูลอนุกรมเวลา โดยอาศัยความสัมพันธ์เชิงบวกได้ดังสมการที่ 12

$$\hat{Y}_t = \hat{L}_t + \hat{N}_t \quad (12)$$

3. การวัดประสิทธิภาพของตัวแบบพยากรณ์

เนื่องจากวิธีการวิเคราะห์ลักษณะความสัมพันธ์ของข้อมูลอนุกรมเวลายังไม่สามารถระบุได้อย่างชัดเจนรวมทั้งยังไม่มีตัวแบบพยากรณ์ใดที่มีประสิทธิภาพดีในการพยากรณ์ทุกสถานการณ์ จึงจำเป็นที่จะต้องนำตัวแบบพยากรณ์ต่าง ๆ เพื่อสร้างตัวแบบสำหรับพยากรณ์ค่าในอนาคต ซึ่งต้องอาศัยการแบ่งชุดข้อมูลสำหรับทดสอบและเกณฑ์การประเมินประสิทธิภาพ เพื่อพิจารณาความถูกต้องของผลการพยากรณ์และเลือกใช้ตัวแบบพยากรณ์ที่เหมาะสมที่สุดในการหาสารสนเทศที่เป็นประโยชน์ต่อการวางแผนและการตัดสินใจที่เหมาะสม

3.1 การแบ่งชุดข้อมูลเพื่อทดสอบตัวแบบพยากรณ์ (cross-validation)

เพื่อให้เกิดความเชื่อมั่นต่อประสิทธิภาพในการพยากรณ์ค่าในอนาคต การแบ่งชุดข้อมูลเพื่อ

ทดสอบประสิทธิภาพการพยากรณ์ของแต่ละตัวแบบจึงเป็นหนึ่งในขั้นตอนที่สำคัญ โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลสำหรับการสร้างตัวแบบพยากรณ์ และข้อมูลสำหรับการทดสอบตัวแบบพยากรณ์ ในสัดส่วนร้อยละ 70 และ 30 ตามลำดับ ซึ่งข้อมูลในชุดทดสอบตัวแบบพยากรณ์จะปิดไว้และไม่เกี่ยวข้องกันชุดข้อมูลในการสร้างตัวแบบ จนกระทั่งข้อมูลชุดนั้นได้เฉลยเพื่อเปรียบเทียบกับผลการพยากรณ์ที่ได้จากตัวแบบพยากรณ์ที่เหมาะสมสำหรับการประเมินประสิทธิภาพ จึงนำข้อมูลที่เฉลยเข้ามารวมกับชุดข้อมูลสำหรับการสร้างตัวแบบพยากรณ์เดิมเพื่อเป็นชุดข้อมูลสำหรับการสร้างตัวแบบพยากรณ์ตัวใหม่และใช้ในการพยากรณ์ค่าในอนาคต 1 ค่าล่วงหน้าต่อไป โดยดำเนินการดังที่กล่าวมาข้างต้นจะกระทำพยากรณ์ครบทุกค่าในชุดข้อมูลสำหรับการทดสอบตัวแบบพยากรณ์ เพื่อนำค่าพยากรณ์และค่าจริงที่ได้ทั้งหมดสำหรับการประเมินประสิทธิภาพการพยากรณ์ด้วยเกณฑ์ต่าง ๆ

3.2 การประเมินประสิทธิภาพของตัวแบบพยากรณ์

สำหรับการประเมินประสิทธิภาพการพยากรณ์ของตัวแบบต่าง ๆ เกณฑ์การประเมินความถูกต้องของการพยากรณ์ต่าง ๆ จึงนำมาใช้สำหรับวัดประสิทธิภาพ ซึ่งเกณฑ์ที่ใช้ในการประเมินในงานวิจัยนี้มี 6 เกณฑ์ ดังนี้

(1) ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์

$$MAE = \text{mean}(|Y_t - \hat{Y}_t|) \quad (13)$$

(2) ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

$$RMSE = \sqrt{\text{mean}((Y_t - \hat{Y}_t)^2)} \quad (14)$$

(3) ค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์

$$MAPE = \text{mean}\left(\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\right) \quad (15)$$

(4) ค่ารากที่สองของค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนกำลังสอง

$$RMSPE = \sqrt{\text{mean} \left(\left(\frac{Y_i - \hat{Y}_i}{Y_i} \times 100 \right)^2 \right)} \quad (16)$$

(5) ค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์แบบสมมาตร

$$sMAPE = \text{mean} \left(2 \times \left| \frac{Y_i - \hat{Y}_i}{Y_i + \hat{Y}_i} \right| \times 100 \right) \quad (17)$$

(6) ค่าสัมประสิทธิ์การตัดสินใจ

$$r^2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (18)$$

เมื่อ x_i และ y_i คือ ค่าใด ๆ ของตัวแปรแต่ละตัว; \bar{x} และ \bar{y} คือ ค่าเฉลี่ยของแต่ละตัวแปร

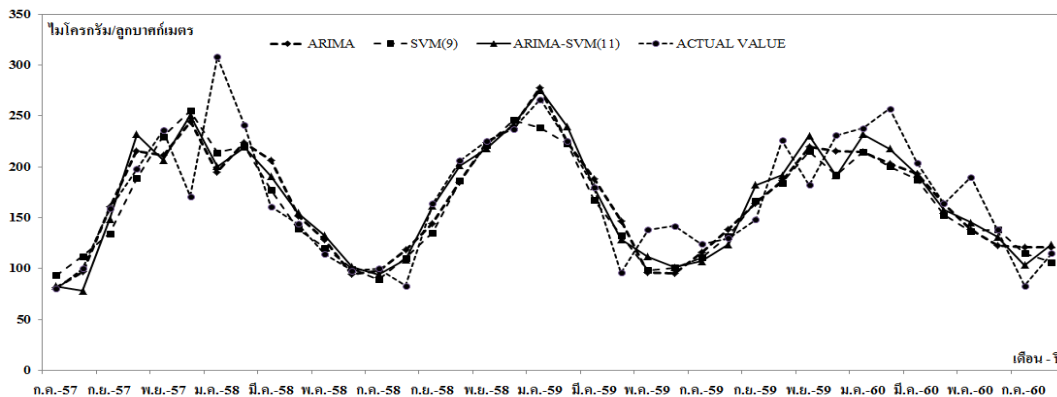
4. ผลการวิจัยและการอภิปรายผล

ผลการพยากรณ์ด้วยตัวแบบพยากรณ์ต่าง ๆ ที่ได้ นำเสนอในงานวิจัยนี้ เพื่อคาดการณ์ปริมาณฝุ่น

ละอองขนาดเล็กกว่า 10 ไมครอน ในเขตพื้นที่ตำบลหน้าพระลาน อำเภอเฉลิมพระเกียรติ จังหวัดสระบุรี นำเสนอในรูปแบบที่ 4 และผลของค่าการพยากรณ์จากตัวแบบต่าง ๆ แสดงในตารางที่ 1

ผลการพยากรณ์ของแต่ละตัวแบบสามารถพิจารณาประสิทธิภาพของแต่ละตัวแบบพยากรณ์ โดยอาศัยเกณฑ์การพิจารณา 6 เกณฑ์ ได้ดังนี้

ในตารางที่ 2 พบว่าตัวแบบซัพพอร์ตเวกเตอร์แมชชีนที่สร้างขึ้นจากความสัมพันธ์แบบไม่เป็นเชิงเส้นเส้นตรงด้วยฟังก์ชันเรเดียลเบสิคมีประสิทธิภาพดีกว่าตัวแบบอาร์มาที่สร้างขึ้นจากความสัมพันธ์แบบเชิงเส้นตรงเล็กน้อย โดยตัวแบบซัพพอร์ตเวกเตอร์แมชชีนให้ผลการพยากรณ์ที่ดีกว่าตัวแบบอาร์มาในเกณฑ์ประเมิน 4 เกณฑ์ ได้แก่ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์ ค่ารากที่สองของค่าเฉลี่ย



รูปที่ 4 ผลการพยากรณ์ของตัวแบบพยากรณ์ต่าง ๆ

ตารางที่ 2 ผลสรุปของการประเมินประสิทธิภาพของการพยากรณ์

Model	MAE	RMSE	MAPE	RMSPE	sMPAE	R-Squared
ARIMA	23.472	33.344	14.50 %	20.46	14.25 %	0.679
ARIMA-SVM(11)	22.843	31.498	13.78 %	17.78	13.69 %	0.715
SVM(9)	23.745	31.820	14.49 %	18.82	14.53 %	0.722

ตารางที่ 1 ผลของค่าการพยากรณ์จากตัวแบบต่าง ๆ

เดือน/ปี	ปริมาณ ฝุ่นละออง ขนาดเล็ก	ค่าพยากรณ์			เดือน/ปี	ปริมาณ ฝุ่นละออง ขนาดเล็ก	ค่าพยากรณ์		
		ARIMA	SVM(9)	ARIMA- SVM(11)			ARIMA	SVM(9)	ARIMA- SVM(11)
ก.ค. 57	80	81.591	93.565	82.470	ก.พ. 59	225	224.442	222.831	239.150
ส.ค. 57	100	96.841	112.161	78.219	มี.ค. 59	180	188.456	167.729	180.501
ก.ย. 57	159	160.863	134.568	148.715	เม.ย. 59	96	146.464	132.696	128.493
ต.ค. 57	198	215.028	188.796	231.765	พ.ค. 59	138	96.526	98.282	111.695
พ.ย. 57	236	211.030	229.578	206.428	มิ.ย. 59	142	94.896	101.011	101.323
ธ.ค. 57	171	244.165	255.156	250.327	ก.ค. 59	124	115.974	110.656	107.064
ม.ค. 58	308	194.460	213.811	200.500	ส.ค. 59	130	138.051	131.749	123.480
ก.พ. 58	241	223.567	220.053	219.189	ก.ย. 59	148	164.330	166.165	181.913
มี.ค. 58	161	206.122	177.526	190.919	ต.ค. 59	226	186.369	184.050	191.577
เม.ย. 58	144	151.396	139.094	154.923	พ.ย. 59	182	219.723	215.297	230.425
พ.ค. 58	114	128.163	120.161	132.485	ธ.ค. 59	231	215.091	192.015	191.082
มิ.ย. 58	98	94.522	99.499	101.697	ม.ค. 60	238	214.691	214.814	231.713
ก.ค. 58	100	97.597	89.450	94.316	ก.พ. 60	257	203.056	200.500	217.522
ส.ค. 58	83	118.632	110.607	108.649	มี.ค. 60	204	193.022	187.601	192.706
ก.ย. 58	164	144.073	135.367	161.541	เม.ย. 60	164	164.461	152.633	157.634
ต.ค. 58	206	184.573	186.334	201.140	พ.ค. 60	190	138.259	136.331	145.204
พ.ย. 58	225	223.713	221.124	217.697	มิ.ย. 60	138	122.499	138.477	130.962
ธ.ค. 58	237	240.145	245.841	241.470	ก.ค. 60	83	121.286	115.020	103.511
ม.ค. 59	266	277.677	238.600	275.333	ส.ค.-60	115	120.577	106.394	123.720

ของเปอร์เซ็นต์ของความคลาดเคลื่อนกำลังสอง และค่าสัมประสิทธิ์การตัดสินใจ แต่ในทางกลับกัน ตัวแบบอาร์มีมีประสิทธิภาพที่ดีกว่าตัวแบบซัพพอร์ตเวกเตอร์แมชชีนในเกณฑ์ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์และค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์แบบสมมาตร ซึ่งบ่งชี้ว่าข้อมูลอนุกรมเวลาของปริมาณฝุ่นละอองขนาดเล็กอาจไม่มีความสัมพันธ์เชิงเส้นตรงหรือแบบไม่เป็นเชิงเส้นตรงอย่างชัดเจน จึงสอดคล้องกับผลการประเมินประสิทธิภาพของตัวแบบผสมอาร์มี-ซัพพอร์ตเวกเตอร์แมชชีน ซึ่งสร้างจากสมมติฐานของค่า

องค์ประกอบของข้อมูลอนุกรมเวลาที่ประกอบไปด้วยองค์ประกอบเชิงเส้นตรงและไม่เป็นเชิงเส้นตรง โดยมีความสัมพันธ์ระหว่างองค์ประกอบทั้ง 2 แบบเชิงบวกซึ่งให้ค่าความถูกต้องของการพยากรณ์ที่เพิ่มมากขึ้น ยกเว้นเกณฑ์ค่าสัมประสิทธิ์การตัดสินใจที่ให้ค่าที่สูงกว่าตัวแบบอาร์มีแต่กลับมีค่าน้อยกว่าตัวแบบซัพพอร์ตเวกเตอร์แมชชีนเล็กน้อย จากเกณฑ์การประเมินประสิทธิภาพทั้งหมดจึงสนับสนุนการใช้ตัวแบบผสมเพื่อเพิ่มความถูกต้องของผลการพยากรณ์และลดความเสี่ยงการเลือกใช้ตัวแบบเชิงเดี่ยวที่อาจไม่เหมาะสมกับคุณลักษณะของข้อมูลอนุกรมเวลา นอกจากนี้ การเพิ่ม

ความถูกต้องในการพยากรณ์ความเสี่ยงจากอันตรายที่มีผลกระทบต่อสุขภาพของประชากรสามารถก่อให้เกิดประโยชน์อย่างมากในการป้องกันและลดความสูญเสียจากปัญหาฝุ่นละอองขนาดไม่เกิน 10 ไมครอน

5. สรุปผลการวิจัย

ผลการวิเคราะห์ประสิทธิภาพของตัวแบบพยากรณ์ทั้งหมด ทั้งตัวแบบอาร์มาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความคลาดเคลื่อนค่อนข้างใกล้เคียงกัน ซึ่งสามารถใช้ตัวแบบทั้งสองในการพยากรณ์ได้ในกรณีที่ต้องการใช้ตัวแบบพยากรณ์เชิงเดี่ยว อย่างไรก็ตาม ตัวแบบซัพพอร์ตเวกเตอร์แมชชีนควรเป็นตัวเลือกแรกในการใช้พยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในเขตพื้นที่ตำบลหน้าพระลานสำหรับตัวแบบพยากรณ์เชิงเดี่ยว นอกจากนี้เพื่อลดความคลาดเคลื่อนในน้อยลง การเลือกใช้ตัวแบบพยากรณ์ผสมระหว่างตัวแบบอาร์มาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนสามารถให้ค่าความคลาดเคลื่อนที่ลดลงเมื่อเทียบกับตัวแบบพยากรณ์เชิงเดี่ยว ซึ่งสามารถลดความเสี่ยงของการเลือกใช้ตัวแบบเชิงเดี่ยวที่ไม่เหมาะสมกับคุณลักษณะของมลอนุกรมเวลาได้ นอกจากนี้ตัวแบบผสมที่พัฒนาขึ้นสามารถเป็นตัวแบบที่น่าสนใจสำหรับการประยุกต์ใช้พยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในพื้นที่อื่น ๆ เพื่อนำสารสนเทศที่ได้สำหรับการเฝ้าระวังและวางแผนการลดผลกระทบที่จะเกิดขึ้นได้อย่างมีประสิทธิภาพ

ส่วนของการวิจัยต่อไป การพัฒนาตัวแบบผสมระหว่างตัวแบบอาร์มาและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนที่เน้นการค้นหาคุณลักษณะที่เด่นชัดขององค์ประกอบที่ไม่เป็นเชิงเส้นตรง เพื่อให้ตัวแบบซัพพอร์ตเวกเตอร์แมชชีนสามารถเรียนรู้และจดจำรูปแบบของการเปลี่ยนแปลงได้ดียิ่งขึ้น จึงน่าจะส่งผลให้ประสิทธิภาพของตัวแบบผสมที่พัฒนาเพิ่มเติมมี

ความแม่นยำมากยิ่งขึ้น เนื่องจากตัวแบบซัพพอร์ตเวกเตอร์แมชชีนสร้างตัวแบบพยากรณ์จากการเรียนรู้คุณลักษณะของข้อมูล การสกัดสารสนเทศที่ชัดเจนมากยิ่งขึ้น ย่อมส่งผลต่อประสิทธิภาพที่ดียิ่งขึ้น

6. รายการอ้างอิง

- [1] แนวทางการเฝ้าระวังพื้นที่เสี่ยงจากมลพิษทางอากาศกรณีฝุ่นละอองขนาดเล็ก, กรมอนามัย และกรมควบคุมโรค กระทรวงสาธารณสุข, แหล่งที่มา : [http://hia.anamai.moph.go.th/download/hia/manual/book/book44 .pdf](http://hia.anamai.moph.go.th/download/hia/manual/book/book44.pdf), 1 ตุลาคม 2560.
- [2] สถานการณ์มลพิษประเทศไทย ปี 2558 รอบ 6 เดือน, กรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม, แหล่งที่มา : <http://www.oic.go.th/FILEWEB/CABINFOCENTER/3/DRAWER056/GENERAL/DATA0000/00000425.PDF>, 1 ตุลาคม 2560.
- [3] Güler, N. and İşçi, Ö.G., 2016, The regional prediction model of PM 10 concentrations for Turkey, *Atmospheric Res.* 180: 64-77.
- [4] Song, Y., Qin, S., Qu, J. and Liu, F., 2015, The forecasting research of early warning systems for atmospheric pollutants: A case in Yangtze River Delta region, *Atmospheric Environ.* 118: 58-69.
- [5] Yang, Z. and Wang, J., 2017, A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction, *Environ. Res.* 158: 105-117.
- [6] Xu, Y., Yang, W. and Wang, J., 2017, Air quality early-warning system for cities in

- China, *Atmospheric Environ.* 148: 239-257.
- [7] Zafra, C., Ángel, Y. and Torres, E., 2017, ARIMA analysis of the effect of land surface coverage on PM₁₀ concentrations in a high-altitude megacity, *Atmospheric Pollut. Res.* 8: 660-668.
- [8] Jian, L., Zhao, Y., Zhu, Y.P., Zhang, M.B. and Bertolatti, D., 2012, An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China, *Sci. Total Environ.* 426: 336-345.
- [9] Dumitrache, R.C., Iriza, A., Maco, B.A., Barbu, C.D., Hirtl, M., Mantovani, S., Nicola, O., Irimescu, A., Craciunescu, V., Ristea, A. and Diamandi, A., 2016, Study on the influence of ground and satellite observations on the numerical air-quality for PM₁₀ over Romanian territory, *Atmospheric Environ.* 143: 278-289.
- [10] Nieto, P.G., Combarro, E.F., del Coz Díaz, J.J. and Montañés, E., 2013, A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study, *Appl. Math. Comput.* 219: 8923-8937.
- [11] Sánchez, A.S., Nieto, P.G., Fernández, P.R., del Coz Díaz, J.J. and Iglesias-Rodríguez, F.J., 2011, Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain), *Math. Comput. Model.* 54: 1453-1466.
- [12] Hirtl, M., Mantovani, S., Krüger, B.C., Triebnig, G., Flandorfer, C., Bottoni, M. and Cavicchi, M., 2014, Improvement of air quality forecasts with satellite and ground based particulate matter observations, *Atmospheric Environ.* 84: 20-27.
- [13] Wang, P., Zhang, H., Qin, Z. and Zhang, G., 2017, A novel hybrid-Garch model based on ARIMA and SVM for PM_{2.5} concentrations forecasting, *Atmospheric Pollut. Res.* 8: 850-860.
- [14] Khashei, M. and Bijari, M., 2011, A new hybrid methodology for nonlinear time series forecasting, *Model. Simulat. Eng.* 2011: 15.
- [15] Khashei, M. and Bijari, M., 2011, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Appl. Soft Comput.* 11: 2664-2675.
- [16] Zhu, B. and Wei, Y., 2013, Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology, *Omega* 41: 517-524.
- [17] Guo, Y., Wang, G., Zhang, X. and Deng, W., 2014, An improved hybrid ARIMA and support vector machine model for water quality prediction, pp. 411-422, *International Conference on Rough Sets and Knowledge Technology*, Springer, Cham.
- [18] de Oliveira, J.F. and Ludermir, T.B., 2016, A hybrid evolutionary decomposition

- system for time series forecasting, Neurocomputing 180: 27-34.
- [19] สรุปข้อมูลคุณภาพอากาศ พ.ศ. 2557-2559, สำนักจัดการคุณภาพอากาศและเสียง กรมควบคุมมลพิษ, แหล่งที่มา : <http://aqnis.pcd.go.th>, 1 ตุลาคม 2560.
- [20] Khandakar, Y. and Hyndman, R.J., 2008, Automatic time series forecasting: the forecast Package for R, J. Stat. Software 27: 1-22.
- [21] Sujjaviriyasup, T., 2017, Forecasting Thailand's monthly export quantity of rubber compound using support vector machine model, Srinakharinwirot Sci. J. 33(1): 205-220. (In Thai)
- [22] Muller, J. and Bogenberger, K., 2015, Time series analysis of booking data of a free-floating carsharing system in Berlin, Transp. Res. Proc. 10: 345-354.
- [23] Kokic, P., Crimp, S. and Howden, M., 2014, A probabilistic analysis of human influence on recent record global mean temperature changes, Climate Risk Manage. 3: 1-12.
- [24] Ghodhi, Z., Silva, S.E. and Hassani, H., 2015, Bicoid Signal Extraction with a Selection of Parametric and Nonparametric Signal Processing Techniques, Genomics Proteomics Bioinformatics 13: 183-191.
- [25] Haworth, J., Shawe-Taylor, J., Cheng, T. and Wang, J., 2014, Local online kernel ridge regression for forecasting of urban travel times, Transp. Res. C 46: 151-178.
- [26] Götz, M., Richerzhagen, M., Bodenstein, C., Cavallaro, G., Glock, P., Riedel, M. and Benediktsson, J.A., 2015, On scalable data mining techniques for earth science, Proc. Comput. Sci. 51: 2188-2197.
- [27] Appelhans, T., Mwangomo, E., Hardy, D.R., Hemp, A. and Nauss, T., 2015, Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt.Kilimanjaro, Tanzania, Spatial Stat. 14: 91-113.