

แบบจำลองความสัมพันธ์ระหว่างค่าสุดขีดของปริมาณน้ำฝนกับผลผลิตข้าวโดยใช้การเรียนรู้ของเครื่องในกลุ่มจังหวัดภาคเหนือตอนล่าง

Modeling the Relationship between Extreme Values of Rainfall and Rice Yields using Machine Learning in the Lower Northern Provinces

ศรัณญา ทองสุข^{1*} และณัฐินี ดีแท้¹Saranya Thongsook^{1*} and Natthinee Deetae¹

Received date: 30 พ.ย. 66 Revised date: 10 ก.ย. 67 Accepted date: 2 ต.ค. 67

DOI: <https://doi.org/10.55003/kmaj.2025.06.16.001>

บทคัดย่อ

งานวิจัยนี้นำเสนอการพัฒนาแบบจำลองความสัมพันธ์ระหว่างค่าสุดขีดของปริมาณน้ำฝนกับผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่าง ข้อมูลผลผลิตข้าวนาปีรายปีและปริมาณน้ำฝนรวมรายปีในช่วงปี 2533 ถึง ปี 2562 ของ 6 จังหวัดภาคเหนือตอนล่าง ซึ่งประกอบด้วย นครสวรรค์ พิจิตร พิษณุโลก กำแพงเพชร เพชรบูรณ์ และสุโขทัย ถูกนำมาใช้สำหรับพัฒนาแบบจำลองความสัมพันธ์สำหรับ 6 จังหวัดภาคเหนือตอนล่างดังกล่าว โดยใช้วิธีการเรียนรู้ของเครื่อง 3 วิธี ประกอบด้วย วิธีการวิเคราะห์การถดถอยพหุคูณ วิธีป่าสุ่ม และวิธีซัพพอร์ตเวกเตอร์แมชชีน สำหรับประสิทธิภาพของแบบจำลองแต่ละแบบในแต่ละจังหวัดถูกประเมินจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) และค่าอาร์กำลังสอง (R-squared) ผลการศึกษาพบว่าแบบจำลองที่ถูกสร้างขึ้นโดยวิธีป่าสุ่มและมีตัวแปรอิสระเป็นค่าสุดขีดของปริมาณน้ำฝนเป็นแบบจำลองที่มีประสิทธิภาพเหนือกว่าแบบจำลองที่ถูกสร้างขึ้นโดยวิธีการวิเคราะห์การถดถอยพหุคูณและวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยเฉพาะในจังหวัดนครสวรรค์, จังหวัดพิจิตร, จังหวัดพิษณุโลกและจังหวัดกำแพงเพชร โดยพิจารณาจากค่าอาร์กำลังสองที่มีค่าสูงกว่าวิธีอื่น ๆ และค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าต่ำกว่าวิธีอื่น ๆ นอกจากนี้พบว่าการใช้ค่าสุดขีดของปริมาณน้ำฝน เป็นตัวแปรนำเข้าสำหรับการสร้างแบบจำลองให้ประสิทธิภาพของแบบจำลองที่ดีกว่าการใช้ปริมาณน้ำฝนรวมรายปี ซึ่งค่าสุดขีดของปริมาณน้ำฝนสามารถอธิบายความผันแปรของผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่างดังกล่าวได้ถึง 64-79%

คำสำคัญ: ค่าสุดขีดของปริมาณน้ำฝน การเรียนรู้ของเครื่อง ผลผลิตข้าว จังหวัดภาคเหนือตอนล่าง

Abstract

This research presents the development of a model for the relationship between extreme values of rainfall and rice yield in the lower northern provinces. Data on annual rice yield and total annual rainfall from 1990 to 2019 of six lower northern provinces which included Nakhonsawan, Phichit, Phitsanulok, Kamphaengphet, Phetchabun and Sukhotha were used to develop a relationship model for the six lower northern provinces. Three machine learning methods were used for the development of a model. The methods comprised multiple regression analysis, random forest and support vector machine. In this research, the performance of each model in each province was evaluated from the mean absolute error (MAE), the root mean square error (RMSE), the mean square error (MSE) and R-squared value. The results showed that a model created using the random forest method, with extreme rainfall values as independent variable, outperformed the models built using the multiple regression analysis and support vector matching methods. This was particularly true in the provinces of Nakhon Sawan, Phichit, Phitsanulok, and Kamphaeng Phet, where it had both higher R-squared values and lower root mean squared error values. In addition, it was found that using the extreme values of rainfall as an input variable for modeling, the model performance is better than using total annual rainfall. The extreme values of rainfall can explain 64-79% of the variation in rice yield in the lower northern provinces.

Keywords: extreme values of rainfall, machine learning, rice yield, the lower northern provinces

¹ สาขาวิชาวิทยาการข้อมูลและสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม จังหวัดพิษณุโลก 65000

¹ Department of Data Science and Statistics, Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok 65000

*corresponding author e-mail: saranya.tho@psru.ac.th

คำนำ

การประมาณผลผลิตทางการเกษตรก่อนการเก็บเกี่ยวถือเป็นประเด็นสำคัญและจำเป็นสำหรับภาคการเกษตร เนื่องจากการเปลี่ยนแปลงของผลผลิตทางการเกษตรในแต่ละปีจะมีอิทธิพลต่อธุรกิจระหว่างประเทศ การจัดหาอาหารและราคาในตลาดโลก นอกจากนี้การคาดการณ์ผลผลิตทางการเกษตรล่วงหน้าถือเป็นการให้ข้อมูลที่จำเป็นประโยชน์แก่ผู้ที่เกี่ยวข้องสำหรับการวางแผนการใช้ที่ดินและนโยบายเศรษฐกิจอย่างมีประสิทธิภาพอีกด้วย ทั้งนี้หากการคาดการณ์ตามสภาพอากาศมีความแม่นยำมากขึ้นเกษตรกรจะได้รับการแจ้งเตือนล่วงหน้าอย่างดีเพื่อลดการสูญเสียครั้งใหญ่ ประเทศไทยเป็นประเทศเกษตรกรรมที่มีข้าวเป็นผลผลิตหลักทางการเกษตร ซึ่งแหล่งปลูกข้าวในประเทศไทยพบได้กระจายทั่วทุกภาคของประเทศไทย ข้าวจึงเป็นสินค้าส่งออกที่สำคัญสร้างรายได้ให้กับประเทศ ดังนั้นการศึกษาวិธีการประมาณผลผลิตข้าวจึงเป็นสิ่งสำคัญและจำเป็นกับประเทศไทยอย่างยิ่ง

สภาพภูมิอากาศเป็นหนึ่งในปัจจัยหลักที่สำคัญสำหรับประมาณผลผลิตทางการเกษตร ดังนั้นจึงมีงานวิจัยสำหรับทำการศึกษาเพื่อสร้างแบบจำลองความสัมพันธ์ระหว่างปัจจัยทางภูมิอากาศและผลผลิตทางการเกษตรชนิดต่าง ๆ โดยงานวิจัยต่าง ๆ ดังกล่าวใช้ปัจจัยทางภูมิอากาศหลาย ๆ ปัจจัยสำหรับการประยุกต์วิธีการทางสถิติและเทคนิคการเรียนรู้ของเครื่องเพื่อให้ได้แบบจำลองความสัมพันธ์ที่มีความถูกต้องและแม่นยำมากยิ่งขึ้น ปัจจุบันการเปลี่ยนแปลงของสภาพภูมิอากาศส่งผลกระทบต่อทั้งทางตรงและทางอ้อมต่อการเกษตรอย่างหลีกเลี่ยงไม่ได้ ในปี 2561 งานวิจัยของเครน-ดรอเอสช์ (Crane-Droesch, 2018) ชี้ให้เห็นว่าการเปลี่ยนแปลงสภาพภูมิอากาศมีผลกระทบเชิงลบอย่างมากต่อผลผลิตข้าวโพด อีกทั้งงานวิจัยของเฉินและคณะ (Chen et al., 2019) ชี้ให้เห็นว่าค่าสุดขีดของสภาพภูมิอากาศทำให้ผลผลิตมวลรวมลดลง 2.8% ต่อปีในประเทศจีนตลอดช่วงสามทศวรรษที่ผ่านมา ดังนั้นจึงมีงานวิจัยในต่างประเทศมากมายที่ทำการศึกษผลกระทบหรือความสัมพันธ์ของค่าสุดขีดของข้อมูลสภาพภูมิอากาศกับปริมาณผลผลิตทางการเกษตร อริดอสและคณะ (Alidoost et al., 2019) ศึกษาการสร้างแบบจำลองระหว่างดัชนีค่าสุดขีดของสภาพภูมิอากาศกับปริมาณผลผลิต ต้นทุนและราคาของน้ำมันในประเทศเนเธอร์แลนด์โดยวิธีการประยุกต์โคปูลาพบว่าวิธีการประยุกต์ใช้โคปูลาเป็นวิธีที่เหมาะสมสำหรับสร้างแบบจำลองผลกระทบของดัชนีค่าสุดขีดของสภาพภูมิอากาศที่มีต่อตัวแปรตามโดยเฉพาะในกรณีที่มีตัวแปรตามหลายตัว วอกเกิลและคณะ (Vogel et al., 2019) ศึกษาผลกระทบของค่าสุดขีดของสภาพภูมิอากาศที่มีต่อความผิดปกติของผลผลิตของข้าวโพด ถั่วเหลือง ข้าว และข้าวสาลีฤดูใบไม้ผลิโดยใช้ข้อมูลผลผลิตในประเทศต่างๆ โดยใช้วิธีป่าสุ่ม (Random Forest Regression) พบว่าค่าสภาพภูมิอากาศเฉลี่ยและค่าสุดขีดของสภาพภูมิอากาศสามารถอธิบายความผันแปรของความผิดปกติของผลผลิตได้ 20% - 49% โดย 18% - 43% ของความผันแปรที่อธิบายมีสาเหตุมาจากค่าสุดขีดของสภาพภูมิอากาศโดยขึ้นอยู่กับชนิดของผลผลิต คอนดურიและคณะ (Konduri et al., 2020) ศึกษาความสัมพันธ์และแบบจำลองการพยากรณ์ของดัชนีค่าสุดขีดของสภาพภูมิอากาศกับปริมาณผลผลิตข้าวโพดโดยใช้วิธีการถดถอยแบบส่วนประกอบหลัก (Principal Component Regression) วิธีการถดถอยแบบบริดจ์ (Ridge Regression) วิธีซัพพอร์ตเวกเตอร์รีเกรชัน (Support Vector Regression : SVR) และวิธีป่าสุ่ม (Random Forest Regression) พบว่าวิธีป่าสุ่มเป็นวิธีที่ดีที่สุดและให้ค่าความคลาดเคลื่อนน้อยที่สุดสำหรับการพยากรณ์ในข้อมูลชุดทดสอบในการสร้างแบบจำลองการพยากรณ์ปริมาณผลผลิตข้าวโพดกับดัชนีค่าสุดขีดของสภาพภูมิอากาศ แฟนและคณะ (Fan et al., 2020) ศึกษาความสัมพันธ์ระหว่างค่าสุดขีดของปริมาณน้ำฝนกับผลผลิตข้าวและผลผลิตข้าวโพดของมณฑลเสฉวนประเทศจีนโดยประยุกต์ใช้ค่า Z-index ของปาล์มเมอร์และวิธีการเรียนรู้ของเครื่อง (Machine Learning) พบว่าค่าผลรวมของ Z-index สามารถใช้จับเหตุการณ์ที่มีความแห้งแล้งและมีความชื้นมากๆ ได้อย่างแม่นยำ และวิธีป่าสุ่มเป็นวิธีที่ดีสำหรับการสร้างแบบจำลองในการประเมินผลกระทบของค่าสุดขีดของสภาพภูมิอากาศที่มีต่อผลผลิตพืช จากงานวิจัยดังกล่าวข้างต้นแสดงให้เห็นว่าเหตุการณ์ที่สภาพภูมิอากาศมีการเปลี่ยนแปลงมีผลกระทบกับปริมาณผลผลิตทางการเกษตรชนิดต่างๆ จึงคาดการณ์ได้ว่าผลกระทบของค่าสุดขีดของสภาพภูมิอากาศที่มีต่อระบบเกษตรกรรมจะเพิ่มขึ้นในอนาคต (Harrison et al., 2016) ซึ่งสามค่าสุดขีดของสภาพภูมิอากาศที่พบว่ามีผลกระทบต่อผลผลิตทางการเกษตรที่สำคัญคือ ค่าสุดขีดของความแห้งแล้ง ค่าสุดขีดของอุณหภูมิต่ำ และค่าสุดขีดของปริมาณน้ำฝน (Rotter et al., 2018)

สำหรับในประเทศไทยงานวิจัยที่เกี่ยวข้องกับศึกษาผลกระทบของค่าสุดขีดของสภาพภูมิอากาศที่มีต่อปริมาณผลผลิตทางการเกษตรยังมีไม่มากนัก ด้วยเหตุผลดังกล่าวคณะผู้วิจัยจึงมีความสนใจศึกษาผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศโดยเฉพาะค่าสุดขีดของปริมาณน้ำฝนที่มีต่อผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่าง โดยมีวัตถุประสงค์เพื่อศึกษาแบบจำลองความสัมพันธ์ระหว่างค่าสุดขีดของปริมาณน้ำฝนกับผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่างและเพื่อหาเครื่องมือที่มีประสิทธิภาพและมีความน่าเชื่อถือสำหรับแบบจำลองดังกล่าว

วิธีการศึกษา

งานวิจัยนี้ทำการรวบรวมข้อมูลทุติยภูมิรายเดือนของผลผลิตข้าวนาปีสำหรับ 6 จังหวัดภาคเหนือตอนล่าง ประกอบด้วย นครสวรรค์ พิจิตร พิษณุโลก กำแพงเพชร เพชรบูรณ์ และสุโขทัยซึ่งเป็นจังหวัดที่มีพื้นที่การปลูกข้าวนาปีมากกว่าร้อยละ 10 ของพื้นที่การปลูกข้าวของจังหวัดภาคเหนือตอนล่างจากสำนักงานเศรษฐกิจการเกษตร และข้อมูลทุติยภูมิรายเดือนของปริมาณน้ำฝนจากกรมอุตุนิยมวิทยา โดยใช้ข้อมูลตั้งแต่ปี 2533 ถึง ปี 2562 รวมทั้งสิ้น 30 ปี ซึ่งข้อมูลทุติยภูมิจากทั้งสองแหล่งข้อมูลเป็นข้อมูลแบบมีโครงสร้าง โดยมีขั้นตอนการวิเคราะห์ข้อมูลตามรายละเอียดดังนี้

กำหนดตัวแปรสำหรับงานวิจัย

ตัวแปรสำหรับงานวิจัยนี้ประกอบด้วย ตัวแปรนำเข้า (ตัวแปรอิสระ) คือ ค่าปริมาณน้ำฝนรวมรายปีและค่าสุดขีดของปริมาณน้ำฝน และตัวแปรผลลัพธ์ (ตัวแปรตาม) คือ ผลผลิตข้าวนาปีของพื้นที่เป้าหมายในแต่ละปี สำหรับค่าปริมาณน้ำฝนจากกรมอุตุนิยมวิทยาจะถูกทำความสะอาดข้อมูล (Cleaning data) โดยการตัดค่าที่เป็น NA ซึ่งหมายถึงการวัดค่าปริมาณน้ำฝนไม่ได้ในแต่ละวันออกก่อนแล้วจึงนำค่าที่เหลือมาคำนวณค่าสุดขีดของปริมาณน้ำฝนต่อไป สำหรับการคำนวณค่าสุดขีดของปริมาณน้ำฝนในงานวิจัยนี้ใช้การประยุกต์ใช้วิธีดัชนีภัยแล้งรุนแรงของพาลเมอร์ (Palmer Z-index) เช่นเดียวกับงานวิจัยของฟิงและคณะ (Feng et al., 2019) ซึ่งจะได้ค่าดัชนีดัชนีอุทกภัยหรือภัยแล้ง (Z-index) โดยคำนวณได้จากสมการที่ (1)-(3)

$$Z_i = \frac{6}{C_s} \left(\frac{C_s}{6} \phi_i + 1 \right)^{\frac{1}{3}} - \frac{6}{C_s} + \frac{C_s}{6} \quad (1)$$

ซึ่ง C_s และ ϕ_i คือ ค่าสัมประสิทธิ์ของความเบ้และตัวแปรปริมาณน้ำฝนที่ถูกทำให้เป็นปกติตามลำดับ โดย i คือตัวแปรปริมาณน้ำฝนที่ถูกทำให้เป็นปกติในเดือนที่ i จากค่า X_i สำหรับค่าทั้งสองค่าสามารถหาได้จากสมการที่ (2) และ (3)

$$C_s = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n\sigma^3} \quad (2)$$

$$\phi_i = \frac{X_i - \bar{X}}{\sigma} \quad (3)$$

ซึ่ง X_i , $\bar{X}$ และ σ คือ ผลรวมของปริมาณน้ำฝนรายเดือนหรือผลรวมของปริมาณน้ำฝนรายเดือน ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของผลรวมของปริมาณน้ำฝนรายเดือนหรือผลรวมของปริมาณน้ำฝนรายเดือนตามลำดับ

เมื่คำนวณค่าดัชนีอุทกภัยหรือภัยแล้ง (Z-index) เรียบร้อยแล้วคณะผู้วิจัยจะได้ค่าสุดขีดของปริมาณน้ำฝนโดยคำนวณจากผลรวมของค่า Z-index ที่ตกอยู่ในทุกกลุ่มยกเว้นกลุ่มปกติ ซึ่งพิจารณาจากค่า Z-index ดังแสดงใน Table 1

การแบ่งข้อมูล (split data)

สำหรับการแบ่งข้อมูลจะทำการแบ่งข้อมูลออกเป็น 2 ส่วน ส่วนแรกใช้สำหรับสร้างสมการ (train data) และส่วนที่สองใช้สำหรับทดสอบสมการ (test data) สำหรับการศึกษาแบ่งข้อมูลสำหรับสร้างแบบจำลอง (train data) คิดเป็น 80% ของข้อมูลทั้งหมด (ตามขนาดของชุดข้อมูล) ตามลำดับ และข้อมูลทดสอบสำหรับทดสอบแบบจำลอง (test data) คิดเป็น 20% ของข้อมูลทั้งหมดตามลำดับ ทั้งนี้เนื่องจากข้อจำกัดของจำนวนหน่วยในวารสารและจำนวนข้อมูลที่แตกต่างกันในแต่ละจังหวัดจึงได้ใช้การแบ่งข้อมูลสำหรับสร้างแบบจำลองสำหรับทดสอบแบบจำลองเป็น 80-20%

Table 1 The z-index of precipitation and temperature extremes

Category	Value Range	Quantile	Category	Value Range	Quantile
Extreme dry	$Z \leq -1.65$	less than 5%	Moderate wet	$0.52 \leq Z \leq 1.04$	70% - 85%
Severe dry	$-1.65 \leq Z \leq -1.04$	5% - 15%	Severe wet	$1.04 \leq Z \leq 1.65$	85% - 95%
Moderate dry	$-1.04 \leq Z \leq -0.52$	15% - 30%	Extreme wet	$Z \geq 1.65$	greater than 95%
Normal	$-0.52 \leq Z \leq 0.52$	30% - 70%			

สร้างแบบจำลองความสัมพันธ์ โดยใช้อัลกอริทึมการเรียนรู้ของเครื่อง (Machine learning algorithms)

งานวิจัยนี้เลือกใช้ทั้งวิธีตัวแบบเชิงเส้นซึ่งได้แก่ การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression : MLR) และตัวแบบไม่เชิงเส้นซึ่งได้แก่ วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และวิธีป่าสุ่ม (Random Forest: RF) สำหรับทำนายผลผลิตข้าวปลอดภัยในจังหวัดพิษณุโลกโดยใช้โปรแกรมอาร์ (R Program) เวอร์ชัน 4.3.1 ซึ่งแต่ละวิธีมีรายละเอียดโดยย่อดังนี้

การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression: MLR) การวิเคราะห์การถดถอยเชิงพหุคูณเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ หลายตัวกับตัวแปรตาม 1 ตัว เพื่อศึกษาว่ามีตัวแปรอิสระตัวใดบ้างที่ร่วมกันทำนายหรือพยากรณ์ หรืออธิบายการผันแปรของตัวแปรตามได้ โดยเขียนความสัมพันธ์ในรูปแบบของสมการที่ (4)

$$Y_i = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e \quad (4)$$

โดยที่ Y_i คือ ตัวแปรตาม (Dependent Variable) $X_1, \dots, X_p$ คือ ตัวแปรอิสระ (Independent Variable), β_0 คือ จุดตัดแกน Y หรือค่าของ Y เมื่อ $X = 0$, $\beta_1, \dots, \beta_p$ คือ ความชัน (Slope) ของเส้นตรงเป็นค่าที่แสดงถึงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนไป 1 หน่วย, e คือ ความคลาดเคลื่อนอย่างสุ่ม (Random Error) เมื่อ $i = 1, \dots, p$

วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) สำหรับการวิเคราะห์การถดถอย เป็นแบบจำลองการเรียนรู้ของเครื่องที่สามารถวิเคราะห์ข้อมูลแบบเชิงเส้น (Linear) และไม่เชิงเส้น (Nonlinear) ทั้งแบบการจำแนกประเภท (Classification) และการวิเคราะห์การถดถอย (Regression Analysis) ซึ่งเหมาะกับข้อมูลที่มีขนาดเล็กไปจนถึงขนาดกลาง SVM สำหรับการวิเคราะห์การถดถอยกำหนดฟังก์ชันวัตถุประสงค์ (Objective Function) ได้ดังสมการที่ (5)

$$\text{ฟังก์ชันวัตถุประสงค์} : \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\epsilon_i + \epsilon_i^*) \quad (5)$$

โดยมีเงื่อนไข : $y_i - \langle w, x_i \rangle - b \leq \epsilon + \epsilon_i$, $\langle w, x_i \rangle + b - y_i \leq \epsilon + \epsilon_i$, $\epsilon_i, \epsilon_i^* \geq 0$

เมื่อ w คือ เวกเตอร์น้ำหนัก, b คือ ค่าสัมประสิทธิ์, ϵ_i คือ ตัวแปรส่วนขาด

โดยเป้าหมายในการฝึกแบบจำลองคือ การหาฟังก์ชันการตัดสินใจที่ y_i ไม่เกินค่า ϵ สำหรับทุก ๆ x_i ดังแสดงในสมการที่ (6)

$$|e|_\epsilon = \begin{cases} 0 & \text{if } |e| \leq \epsilon \\ |e| - \epsilon & \text{otherwise} \end{cases} \quad (6)$$

วิธีป่าสุ่ม (Random Forest: RF) สำหรับการวิเคราะห์การถดถอย เป็นหนึ่งในอัลกอริทึมพื้นฐานของการเรียนรู้ของเครื่อง (Machine Learning) และถือได้ว่าเป็นเครื่องมือทางสถิติที่มีประสิทธิภาพสำหรับการทำนายอีกวิธีหนึ่ง ซึ่งวิธีป่าสุ่มสามารถใช้ได้ทั้งในงานการจำแนกประเภทและการวิเคราะห์การถดถอย ซึ่งเกิดจากการทำงานร่วมกันของวิธีการจำแนกประเภทแบบพิเศษและวิธีการเรกเรชันทรี (Regression Trees : CART) วิธีป่าสุ่มเป็นวิธีการที่ใช้การคาดการณ์โดยการรวมการคาดการณ์จากต้นไม้การตัดสินใจหลายๆ แบบโดยใช้เทคนิคการรวบรวม Bootstrap หรือที่เรียกว่า Bagging สำหรับการทำนายขั้นสุดท้ายของวิธีป่าสุ่มทำโดยการหาค่าเฉลี่ยจากต้นไม้แต่ละต้นวิธีนี้เป็นวิธีที่เหมาะสมกับข้อมูลเชิงปริมาณและเหมาะสมสำหรับแบบจำลองที่ไม่เป็นเชิงเส้นสำหรับค่าพารามิเตอร์สำหรับวิธีป่าสุ่ม เช่น จำนวนต้นไม้ ความลึกสูงสุดของต้นไม้ จำนวนสูงสุดในการแบ่งกลุ่มแต่ละครั้ง และจำนวนตัวอย่างที่น้อยที่สุดในการแบ่งกลุ่มแต่ละครั้งจะถูกกำหนดโดยการค้นหาแบบกริด ซึ่งในงานวิจัยนี้เลือกใช้วิธีป่าสุ่มสำหรับการวิเคราะห์การถดถอยผ่านแพ็คเกจชื่อว่า randomForest ในโปรแกรมอาร์โดยกำหนดค่าพารามิเตอร์ให้เท่ากับค่าเริ่มต้น

การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)

การตรวจสอบประสิทธิภาพของแบบจำลองโดยวิธี Train/Test split จากนั้นจะคำนวณค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE), ค่าอาร์กำลังสอง (R-squared) และอาร์กำลังสองที่ปรับค่าแล้ว (Adjusted R-squared) เพื่อใช้สำหรับพิจารณาประสิทธิภาพของแบบจำลองที่ได้จากแบบจำลองทั้ง 3 วิธี ซึ่งสามารถสรุปเป็นกรอบแนวคิดสำหรับการวิเคราะห์ข้อมูลเพื่อศึกษาความสัมพันธ์ระหว่างค่าสุดขีดของปริมาณน้ำฝนกับผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่าง แสดงไว้ดัง Figure 1

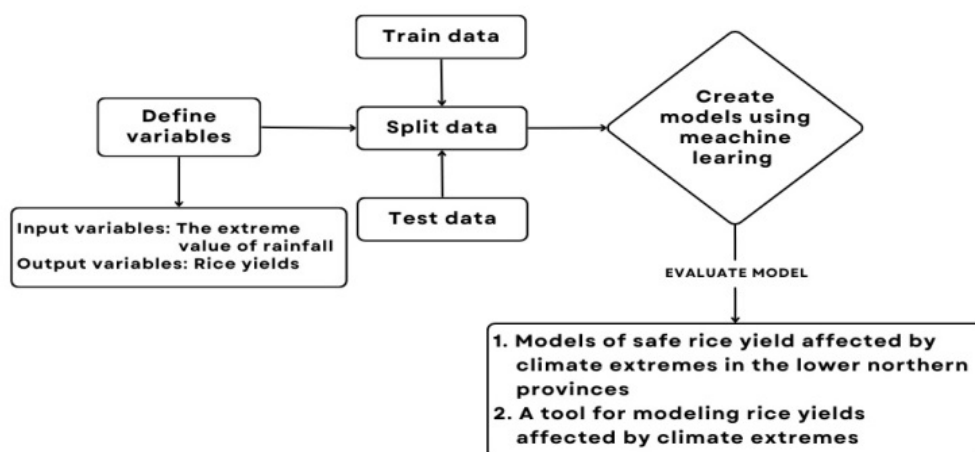


Figure 1 Research conceptual framework

สำหรับการวิจัยนี้จะใช้ค่าขีดสุดของปริมาณน้ำฝนและค่าขีดสุดของอุณหภูมิเป็นตัวแปรนำเข้า (ตัวแปรอิสระ) ในขณะที่ผลผลิตข้าวปลอดภัยของแต่ละปีถูกใช้เป็นตัวแปรผลลัพธ์ (ตัวแปรตาม) และเพื่อการตรวจสอบความถูกต้องและความน่าเชื่อถือของแบบจำลองชุดข้อมูลจะถูกแบ่งออกเป็นสองส่วนแบบสุ่ม โดยส่วนหนึ่งสำหรับสร้างแบบจำลอง (Training) และอีกส่วนหนึ่งสำหรับทดสอบแบบจำลอง (Testing) ในอัตราส่วน 80-20% ของชุดข้อมูลทั้งหมดตามลำดับ และเมื่อได้แบบจำลองแล้วจะทำการตรวจสอบประสิทธิภาพของแบบจำลองที่ได้โดยคำนวณค่า r-squared และรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error (RMSE))

ผลการศึกษาและวิจารณ์

จากการวิเคราะห์ความสัมพันธ์ระหว่างค่าขีดสุดของปริมาณน้ำฝนกับผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่างโดยใช้วิธีการเรียนรู้ด้วยเครื่อง 3 วิธี ได้แก่ วิธีการถดถอยเชิงเส้นพหุคูณ Multiple Linear Regression : MLR), ป่าสุ่ม (Random Forest : RF) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) และใช้ตัวแปรนำเข้า(ตัวแปรอิสระ) คือ ค่าขีดสุดของปริมาณน้ำฝน (zindex_prec) และค่าปริมาณน้ำฝนรวมรายปี (Precipitation) ซึ่งได้แบบจำลอง 6 แบบจำลอง สำหรับการวิเคราะห์ดังนี้ แบบจำลองที่ 1 (MLE+zindex_prec) คือ แบบจำลองที่ถูกสร้างขึ้นโดยวิธีการวิเคราะห์การถดถอยพหุคูณ (MLR) โดยมีตัวแปรอิสระเป็นค่าขีดสุดของปริมาณน้ำฝน (zindex_prec), แบบจำลองที่ 2 (RF+zindex_prec) คือ แบบจำลองที่ถูกสร้างขึ้นโดยวิธีป่าสุ่ม (RF) โดยมีตัวแปรอิสระเป็นค่าขีดสุดของปริมาณน้ำฝน (zindex_prec), แบบจำลองที่ 3 (SVM+zindex_prec) คือ แบบจำลองที่ถูกสร้างขึ้นโดยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยมีตัวแปรอิสระเป็นค่าขีดสุดของปริมาณน้ำฝน (zindex_prec), แบบจำลองที่ 4 (MLE+ Precipitation) คือ แบบจำลองที่ถูกสร้างขึ้นโดยวิธีการวิเคราะห์การถดถอยพหุคูณ (MLR) โดยมีตัวแปรอิสระเป็นค่าปริมาณน้ำฝนรวมรายปี (Precipitation), แบบจำลองที่ 5 (RF+ Precipitation) คือ แบบจำลองที่ถูกสร้างขึ้นโดยวิธีป่าสุ่ม (RF) โดยมีตัวแปรอิสระเป็นค่าปริมาณน้ำฝนรวมรายปี (Precipitation) และแบบจำลองที่ 6 (SVM+ Precipitation) คือ แบบจำลองที่ถูกสร้างขึ้นโดยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยมีตัวแปรอิสระเป็นค่าปริมาณน้ำฝนรวมรายปี (Precipitation)

Table 2 แสดงค่าอาร์กกำลังสองของแบบจำลองที่มีค่าขีดสุดของปริมาณน้ำฝน (zindex_prec) และปริมาณน้ำฝนรวมรายปี (Precipitation) เป็นตัวแปรอิสระสำหรับชุดข้อมูลสร้างแบบจำลอง-ทดสอบแบบจำลอง (Training- Testing) 80-20% ผลการวิจัยพบว่า แบบจำลองที่ 2 ซึ่งมีตัวแปรอิสระคือ ค่าขีดสุดของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีป่าสุ่มเป็นแบบจำลองที่มีค่าอาร์กกำลังสองสูงสุดในจังหวัดพิจิตร, พิษณุโลก, กำแพงเพชรและสุโขทัย โดยมีค่าเท่ากับ 0.62, 0.73, 0.64 และ 0.77 ตามลำดับ สำหรับจังหวัดนครสวรรค์และเพชรบูรณ์พบว่า แบบจำลองที่ 5 ซึ่งมีตัวแปรอิสระคือ ปริมาณน้ำฝนรวมรายปี (Precipitation) และถูกสร้างโดยวิธีป่าสุ่มเป็นแบบจำลองที่มีค่าอาร์กกำลังสองสูงสุด โดยมีค่าเท่ากับ 0.82 และ 0.77 ตามลำดับ

Table 2 The R-squared of models for the z-index of rainfall and the total annual rainfall

Model	provinces					
	Nakhonsawan	Phichit	Phitsanulok	Kampaengphet	Phetchabun	Sukhothai
1. MLE + zindex_prec	0.6394	0.0559	0.0611	0.0340	0.0177	0.0011
2. RF + zindex_prec	0.7852	0.6196	0.7260	0.6396	0.6110	0.7741
3. SVM + zindex_prec	0.5382	0.1551	0.3382	-0.0203	0.0450	0.3144
4. MLE + Precipitation	0.6675	0.1031	0.2498	0.0345	0.0006	0.0153
5. RF + Precipitation	0.8209	0.5947	0.6543	0.6280	0.7691	0.7313
6. SVM + Precipitation	0.6160	0.2405	0.2469	-0.0151	0.2166	0.0224

zindex_prec is z-index of rainfall and Precipitation is the total annual rainfall

Table 3 แสดงค่าอาร์กำลังสองที่ปรับค่าแล้วของแบบจำลองที่มีค่าสถิติของปริมาณน้ำฝน (zindex_prec) และปริมาณน้ำฝนรวมรายปี (Precipitation) เป็นตัวแปรอิสระสำหรับชุดข้อมูลสร้างแบบจำลอง-ทดสอบแบบจำลอง (Training- Testing) 80-20% ผลการวิจัยพบว่า แบบจำลองที่ 2 ซึ่งมีตัวแปรอิสระคือ ค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีป่าสุ่มเป็นแบบจำลองที่มีค่าอาร์กำลังสองที่ปรับค่าแล้วสูงสุดในจังหวัดพิจิตร, พิษณุโลก, กำแพงเพชรและสุโขทัย โดยมีค่าเท่ากับ 0.60, 0.71, 0.62 และ 0.76 ตามลำดับ สำหรับจังหวัดนครสวรรค์และเพชรบูรณ์พบว่า แบบจำลองที่ 5 ซึ่งมีตัวแปรอิสระคือ ปริมาณน้ำฝนรวมรายปี (Precipitation) และแบบจำลองถูกสร้างโดยวิธีป่าสุ่มเป็นแบบจำลองที่มีค่าอาร์กำลังสองที่ปรับค่าแล้วสูงสุด โดยมีค่าเท่ากับ 0.80 และ 0.76 ตามลำดับ

Table 3 The Adjusted R-squared of models for the z-index of rainfall and the total annual rainfall

Model	provinces					
	Nakhonsawan	Phichit	Phitsanulok	Kampaengphet	Phetchabun	Sukhothai
1. MLE + zindex_prec	0.5943	-0.0031	0.0117	-0.0197	-0.0269	-0.0443
2. RF + zindex_prec	0.7584	0.5958	0.7116	0.6196	0.5934	0.7638
3. SVM + zindex_prec	0.4804	0.1023	0.3034	-0.0770	0.0016	0.2832
4. MLE + Precipitation	0.6260	0.0471	0.2103	-0.0191	-0.0448	-0.0276
5. RF + Precipitation	0.7985	0.5694	0.6361	0.6074	0.7586	0.7196
6. SVM + Precipitation	0.5680	0.1930	0.2073	-0.0715	0.1810	-0.0201

zindex_prec is z-index of rainfall and Precipitation is the total annual rainfall

Table 4 Efficiency of models for the z-index of rainfall and the total annual rainfall

Provinces	Models	Efficiency of models			Provinces	Models	Efficiency of models		
		MAE	MSE ($\times 10^5$)	RMSE			MAE	MSE ($\times 10^5$)	RMSE
Nakhonsawan	1	164.07	377.49	194.29	Kampaengphet	1	162.99	290.80	170.53
	2	160.18	341.69	184.85		2	160.33	314.02	177.21
	3	165.69	377.19	194.21		3	124.72	180.69	134.42
	4	194.59	463.55	215.30		4	163.24	291.78	170.82
	5	209.86	504.87	224.07		5	162.23	324.02	180.01
	6	190.26	447.17	211.46		6	125.64	183.00	135.28
Phichit	1	58.14	76.51	87.47	Phetchabun	1	41.90	22.85	47.80
	2	100.07	105.16	102.55		2	114.21	169.74	130.28
	3	94.69	146.13	120.88		3	44.14	22.38	47.31
	4	72.24	123.73	111.24		4	43.93	26.49	51.35
	5	171.21	435.77	208.75		5	45.47	33.57	57.94
	6	102.19	207.37	144.01		6	30.71	14.92	38.62
Phitsanulok	1	120.86	158.31	125.82	Sukhothai	1	96.34	140.10	118.36
	2	101.13	156.27	125.01		2	151.73	370.53	192.49
	3	101.19	124.14	111.42		3	97.23	104.60	102.27
	4	136.84	214.40	146.43		4	103.48	145.20	120.50
	5	136.63	223.57	149.52		5	134.55	242.37	152.59
	6	133.47	200.63	141.64		6	98.29	141.11	118.20

zindex_prec is z-index of rainfall and Precipitation is the total annual rainfall

สำหรับค่าประสิทธิภาพของแบบจำลองโดยพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของแบบจำลองที่มีค่าสถิติของปริมาณน้ำฝน (zindex_prec) และปริมาณน้ำฝนรวมรายปี (Precipitation) เป็นตัวแปรอิสระสำหรับชุดข้อมูลสร้างแบบจำลอง-ทดสอบแบบจำลอง (Training- Testing) 80-20% ตามลำดับ ดังแสดงไว้ใน Table 4 ผลการวิจัยพบว่า แบบจำลองที่ 2 ของจังหวัดนครสวรรค์ซึ่งมีตัวแปรอิสระคือ ค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีปัสเมเป็นแบบจำลองที่มีค่าประสิทธิภาพของแบบจำลองทั้ง 3 ค่าต่ำสุด โดยมีค่าเท่ากับ 160.18 ตัน/ไร่, 341.69×10^5 ตัน/ไร่ และ 184.85 ตัน/ไร่ ตามลำดับ สำหรับจังหวัดพิจิตรพบว่า แบบจำลองที่ 1 ซึ่งมีตัวแปรอิสระคือ ค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีการวิเคราะห์การถดถอยพหุคูณเป็นแบบจำลองที่มีค่าประสิทธิภาพของแบบจำลองทั้ง 3 ค่าต่ำสุด โดยมีค่าเท่ากับ 58.14 ตัน/ไร่, 76.51×10^5 ตัน/ไร่ และ 87.47 ตัน/ไร่ ตามลำดับ สำหรับจังหวัดพิษณุโลกพบว่า แบบจำลองที่ 3 ซึ่งมีตัวแปรอิสระคือ ค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีซัพพอร์ตเวกเตอร์แมทชีนเป็นแบบจำลองที่มีค่าประสิทธิภาพของแบบจำลองทั้ง 2 ค่าต่ำสุด โดยมีค่าเท่ากับ 124.14×10^5 ตัน/ไร่ และ 111.42 ตัน/ไร่ ตามลำดับ สำหรับจังหวัดกำแพงเพชรพบว่า แบบจำลองที่ 3 ซึ่งมีตัวแปรอิสระคือ ค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีซัพพอร์ตเวกเตอร์แมทชีนเป็นแบบจำลองที่มีค่าประสิทธิภาพของแบบจำลองทั้ง 3 ค่าต่ำสุด โดยมีค่าเท่ากับ 124.72 ตัน/ไร่, 180.69×10^5 ตัน/ไร่ และ 134.42 ตัน/ไร่ ตามลำดับ สำหรับจังหวัดเพชรบูรณ์พบว่า แบบจำลองที่ 6 ซึ่งมีตัวแปรอิสระคือ ปริมาณน้ำฝนรวมรายปี (Precipitation) และถูกสร้างโดยวิธีซัพพอร์ตเวกเตอร์แมทชีนเป็นแบบจำลองที่มีค่าประสิทธิภาพของแบบจำลองทั้ง 3 ค่าต่ำสุด โดยมีค่าเท่ากับ 30.71 ตัน/ไร่, 14.92×10^5 ตัน/ไร่ และ 38.62 ตัน/ไร่ ตามลำดับ และสำหรับจังหวัดสุโขทัยพบว่า แบบจำลองที่ 3 ซึ่งมีตัวแปรอิสระคือ ค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีซัพพอร์ตเวกเตอร์แมทชีนเป็นแบบจำลองที่มีค่าประสิทธิภาพของแบบจำลองทั้ง 2 ค่าต่ำสุด โดยมีค่าเท่ากับ 104.60×10^5 ตัน/ไร่, และ 102.27 ตัน/ไร่ ตามลำดับ

สำหรับการพิจารณาความสัมพันธ์ระหว่างค่าสถิติของปริมาณน้ำฝนกับผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่าง,หาเครื่องมือที่มีประสิทธิภาพและมีความน่าเชื่อถือสำหรับแบบจำลองผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่างที่ได้รับผลกระทบจากค่าสถิติของปริมาณน้ำฝน จะพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ซึ่งควรมีค่าต่ำที่สุดเมื่อเทียบกับแบบจำลองอื่นๆ และค่าอาร์กำลังสอง (R-squared) ซึ่งควรมีค่ามากกว่า 0.5 ในเวลาเดียวกัน สำหรับแบบจำลองทั้ง 6 แบบจำลอง ของแต่ละจังหวัด ดังแสดงใน Figure 2-7

Figure 2 แสดงค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าอาร์กำลังสองสำหรับแบบจำลองทั้ง 6 แบบจำลองของจังหวัดนครสวรรค์ พบว่าแบบจำลองที่มีอาร์กำลังสองสูงสุดและในขณะเดียวกันก็มีค่าประสิทธิภาพทั้ง 3 ค่าต่ำสุด คือ แบบจำลองที่ 2 ซึ่งมีตัวแปรอิสระคือค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีป่าสุ่ม โดยมีค่าเท่ากับ 160.18 ตัน/ไร่, 184.85 ตัน/ไร่, 341.69×10^5 ตัน/ไร่ และ 0.79 ตามลำดับ

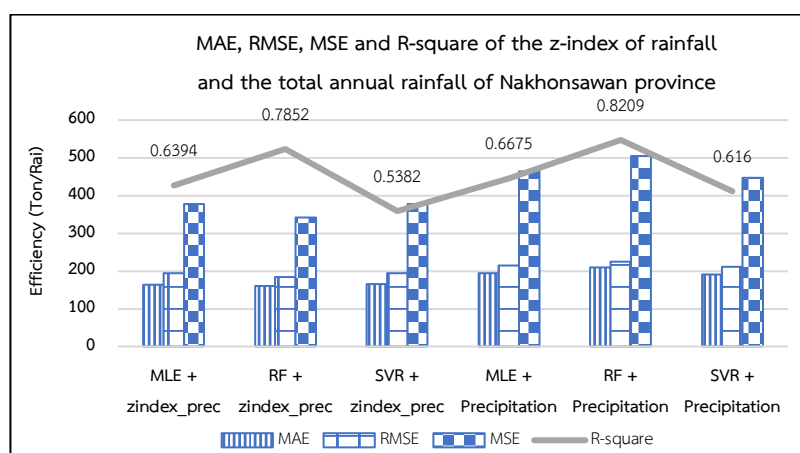


Figure 2 Efficiency of models for Nakhonsawan province

Figure 3 แสดงค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าอาร์กำลังสองสำหรับแบบจำลองทั้ง 6 แบบจำลองของจังหวัดพิจิตร พบว่าแบบจำลองที่มีอาร์กำลังสองมากกว่า 0.50 คือ แบบจำลองที่ 2 และแบบจำลองที่ 5 แต่อย่างไรก็ตามเมื่อพิจารณาค่าประสิทธิภาพทั้ง 3 ค่ารวมด้วย ซึ่งค่าประสิทธิภาพควรมีต่ำสุดพบว่า แบบจำลองที่ 2 ซึ่งมีตัวแปรอิสระคือ ค่าสถิติของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีป่าสุ่ม โดยมีค่าเท่ากับ 100.07 ตัน/ไร่, 102.55 ตัน/ไร่, 105.16×10^5 ตัน/ไร่ และ 0.62 ตามลำดับ

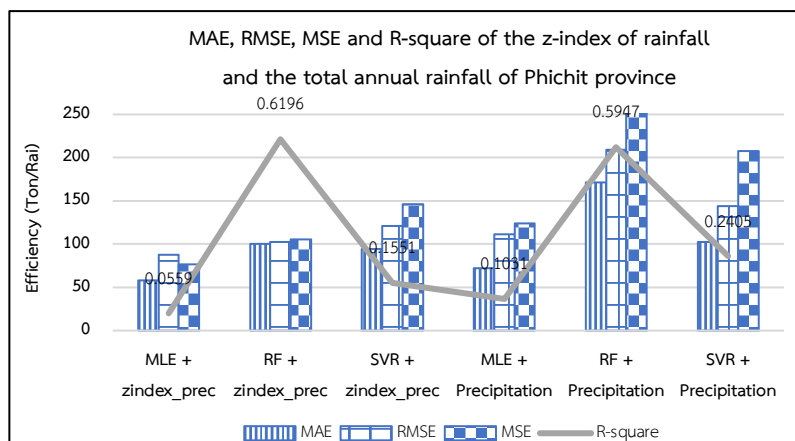


Figure 3 Efficiency of models for Phichit province

Figure 4 แสดงค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าอาร์กำลังสองสำหรับแบบจำลองทั้ง 6 แบบจำลองของจังหวัดพิษณุโลก พบว่าแบบจำลองที่มีอาร์กำลังสองมากกว่า 0.50 คือ แบบจำลองที่ 2 และแบบจำลองที่ 5 แต่อย่างไรก็ตามเมื่อพิจารณาค่าประสิทธิภาพทั้ง 3 ค่ารวมด้วย ซึ่งค่าประสิทธิภาพควรมีต่ำสุดพบว่า แบบจำลองที่ 2 ซึ่งมีตัวแปรอิสระคือ ค่าสุดขีดของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีป่าสุ่ม โดยมีค่าเท่ากับ 101.13 ตัน/ไร่, 125.01 ตัน/ไร่, 156.27×10^5 ตัน/ไร่ และ 0.73 ตามลำดับ

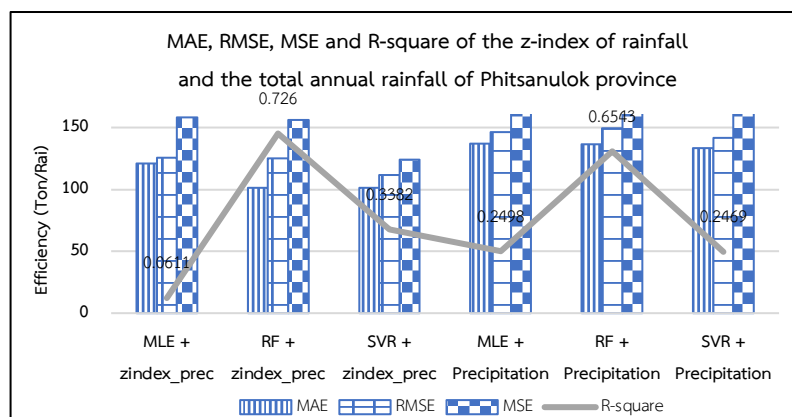


Figure 4 Efficiency of models for Phitsanulok province

Figure 5 แสดงค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าอาร์กำลังสองสำหรับแบบจำลองทั้ง 6 แบบจำลองของจังหวัดกำแพงเพชร พบว่าแบบจำลองที่มีอาร์กำลังสองมากกว่า 0.50 คือ แบบจำลองที่ 2 และแบบจำลองที่ 5 แต่อย่างไรก็ตามเมื่อพิจารณาค่าประสิทธิภาพทั้ง 3 ค่ารวมด้วย ซึ่งค่าประสิทธิภาพควรมีต่ำสุดพบว่า แบบจำลองที่ 2 ซึ่งมีตัวแปรอิสระคือ ค่าสุดขีดของปริมาณน้ำฝน (zindex_prec) และถูกสร้างโดยวิธีป่าสุ่ม โดยมีค่าเท่ากับ 160.33 ตัน/ไร่, 177.21 ตัน/ไร่, 314.02×10^5 ตัน/ไร่ และ 0.64 ตามลำดับ

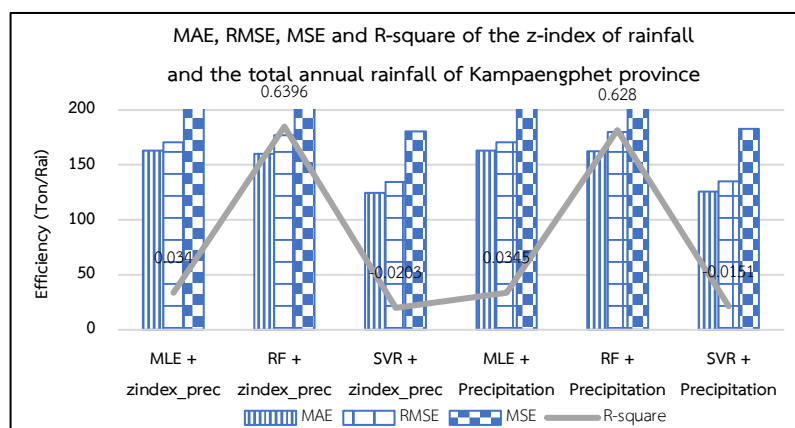


Figure 5 Efficiency of models for Kampaengphet province

Figure 6 แสดงค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าอาร์กำลังสองสำหรับแบบจำลองทั้ง 6 แบบจำลองของจังหวัดเพชรบูรณ์ พบว่าแบบจำลองที่มีอาร์กำลังสองมากกว่า 0.50 คือ แบบจำลองที่ 2 และแบบจำลองที่ 5 แต่อย่างไรก็ตามเมื่อพิจารณาค่าประสิทธิภาพทั้ง 3 ค่ารวมด้วย ซึ่งค่าประสิทธิภาพควรมีต่ำสุดพบว่า แบบจำลองที่ 5 ซึ่งมีตัวแปรอิสระคือ ปริมาณน้ำฝนรวมรายปี (Precipitation) และถูกสร้างโดยวิธีป่าสุ่ม โดยมีค่าเท่ากับ 45.47 ตัน/ไร่, 57.94 ตัน/ไร่, 33.57×10^5 ตัน/ไร่ และ 0.77 ตามลำดับ

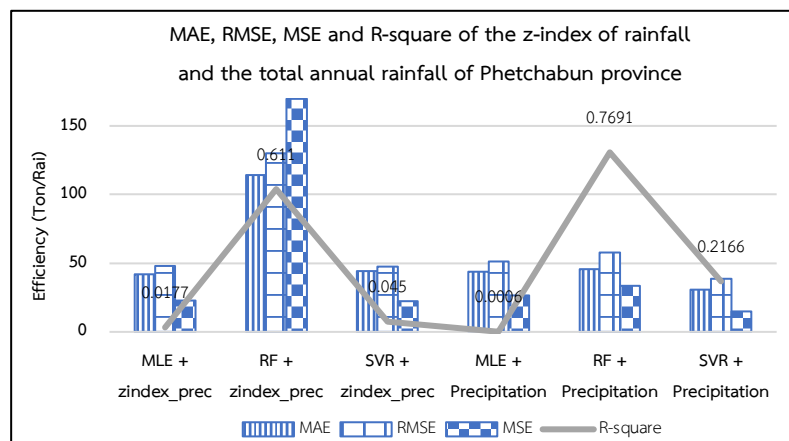


Figure 6 Efficiency of models for Phetchabun province

Figure 7 แสดงค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE), ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าอาร์กำลังสองสำหรับแบบจำลองทั้ง 6 แบบจำลองของจังหวัดสุโขทัย พบว่าแบบจำลองที่มีอาร์กำลังสองมากกว่า 0.50 คือ แบบจำลองที่ 2 และแบบจำลองที่ 5 แต่อย่างไรก็ตามเมื่อพิจารณาค่าประสิทธิภาพทั้ง 3 ค่ารวมด้วย ซึ่งค่าประสิทธิภาพควรมีต่ำสุดพบว่า แบบจำลองที่ 5 ซึ่งมีตัวแปรอิสระคือ ปริมาณน้ำฝนรวมรายปี (Precipitation) และถูกสร้างโดยวิธีป่าสุ่ม โดยมีค่าเท่ากับ 134.55 ตัน/ไร่, 152.59 ตัน/ไร่, 242.37×10^5 ตัน/ไร่ และ 0.73 ตามลำดับ

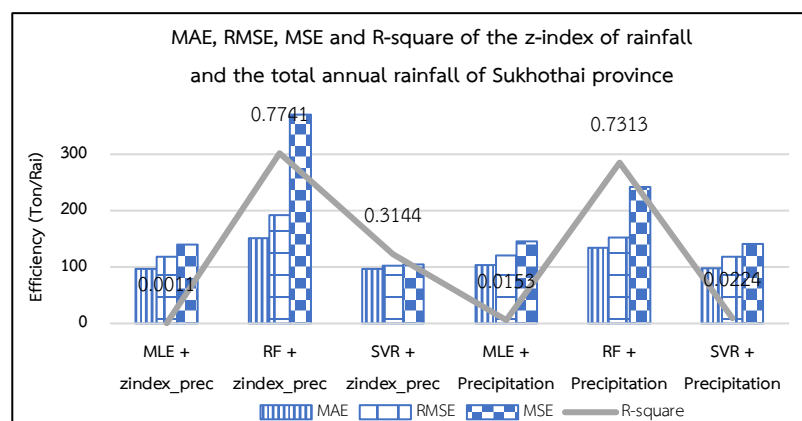


Figure 7 Efficiency of models for Sukhothai province

สรุปผลการศึกษา

จากการศึกษาการพัฒนาแบบจำลองความสัมพันธ์ระหว่างค่าสุดขีดของปริมาณน้ำฝนกับผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่างแสดงให้เห็นว่า แบบจำลองที่ถูกสร้างขึ้นโดยวิธีป่าสุ่มและมีตัวแปรอิสระเป็นค่าสุดขีดของปริมาณน้ำฝน (RF+zindex_prec) เป็นแบบจำลองที่มีประสิทธิภาพเหนือกว่าแบบจำลองที่ถูกสร้างขึ้นโดยวิธีการวิเคราะห์การถดถอยพหุคูณ (MLE) และวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยเฉพาะในจังหวัดนครสวรรค์, จังหวัดพิจิตร, จังหวัดพิษณุโลกและจังหวัดกำแพงเพชร ทั้งนี้เนื่องจากค่าอาร์กำลังสองที่ค่อนข้างสูงและมีค่าความคลาดเคลื่อนอย่างน้อย 2 วิธี ต่ำที่สุด นอกจากนี้ผลการวิจัยแสดงให้เห็นว่าการใช้ค่าสุดขีดของปริมาณน้ำฝน (zindex_prec) เป็นตัวแปรนำเข้าสำหรับการสร้างแบบจำลองให้ประสิทธิภาพของแบบจำลองที่ดีกว่าการใช้ปริมาณน้ำฝนรวมรายปี (Precipitation) เนื่องจากค่าสุดขีดของปริมาณน้ำฝนสามารถอธิบายได้ทั้งความแห้งแล้งและน้ำท่วม ในขณะที่ปริมาณฝนสะสมอาจละเลยเหตุการณ์บางอย่างเนื่องจากการกระจายตัวของฝนไม่สม่ำเสมอ (Fan et al., 2020) ซึ่งสอดคล้องกับงานวิจัยของ Jeong et al. (2016) ที่แสดงให้เห็นว่าวิธีป่าสุ่มเป็นวิธีการเรียนรู้ของเครื่องที่มีประสิทธิภาพและมีประโยชน์สำหรับการคาดการณ์ผลผลิตพืชผลในระดับภูมิภาคและระดับโลก เนื่องจากมีความแม่นยำและเที่ยงตรงสูง ใช้งานง่าย และมีประโยชน์ในการวิเคราะห์ข้อมูล อีกทั้งสอดคล้องกับงานวิจัยของแฟนและคณะ (Fan et al., 2020) ซึ่งแสดงให้เห็นว่าแบบจำลองที่สร้างด้วยวิธีป่าสุ่มและมีตัวแปรอิสระเป็นค่าสุดขีดของปริมาณน้ำฝนเป็นวิธีที่ดีสำหรับการสร้างแบบจำลองในการประเมินผลกระทบของค่าสุดขีดของสภาพภูมิอากาศที่มีต่อผลผลิตพืช สำหรับงานวิจัยนี้ค่าสุดขีดของปริมาณน้ำฝนสามารถอธิบายความผันแปรของผลผลิตข้าวในกลุ่มจังหวัดภาคเหนือตอนล่างสำหรับบางจังหวัดได้ถึง 64-79% แต่อย่างไรก็ตามพบว่าการใช้ปริมาณน้ำฝนรวมรายปี (Precipitation) เป็นตัวแปรนำเข้าสำหรับการสร้างแบบจำลองให้ประสิทธิภาพของแบบจำลองที่ดีกว่าการใช้ค่าสุดขีดของปริมาณน้ำฝน (zindex_prec) ในบางจังหวัดได้แก่ จังหวัดเพชรบูรณ์และจังหวัดสุโขทัย รายละเอียดดังสรุปไว้ใน Table 5

ดังนั้นจึงกล่าวได้ว่าแบบจำลองที่ถูกสร้างขึ้นโดยวิธีป่าสุ่มและมีตัวแปรอิสระเป็นค่าสุดขีดของปริมาณน้ำฝนเป็นวิธีที่ง่ายและสามารถใช้งานได้จริงสำหรับผู้บริหารเพื่อการตัดสินใจและปฏิบัติงานในระดับจังหวัดเพื่อทำความเข้าใจความสัมพันธ์ระหว่างผลผลิตข้าวกับค่าสุดขีดของปริมาณน้ำฝนและเพื่อใช้เป็นเครื่องมือสำหรับเฝ้าระวังและติดตามปริมาณผลผลิตข้าวตลอดช่วงระยะเวลาการผลิตสำหรับผู้ใช้งาน, ผู้สนใจหรือหน่วยงานที่ต้องการทราบปริมาณผลผลิตข้าวในพื้นที่ต่างๆ ในช่วงระยะเวลาการผลิตในปัจจุบันและในอนาคตต่อไป นอกจากนี้ข้อมูลที่ได้ดังกล่าวจะเป็นประโยชน์ทางอ้อมต่อหน่วยงานภาครัฐซึ่งเป็นผู้กำหนดนโยบายต่าง ๆ ทางภาคการเกษตรและเป็นประโยชน์ต่อเกษตรกรผู้ผลิตข้าวในแง่ของการตัดสินใจ แต่อย่างไรก็ตามการใช้ค่าสุดขีดของปริมาณน้ำฝนสำหรับการสร้างแบบจำลองผลผลิตข้าวอาจไม่เหมาะสมสำหรับบางจังหวัด ดังนั้นงานวิจัยในอนาคตควรทำการศึกษาหาวิธีวัดค่าสุดขีดของปริมาณน้ำฝนด้วยวิธีอื่น ๆ เพื่อใช้สำหรับการศึกษาความสัมพันธ์ระหว่างผลผลิตข้าวกับค่าสุดขีดของปริมาณน้ำฝนด้วยวิธีอื่นๆ เพื่อให้เหมาะสมสำหรับทุก ๆ จังหวัดต่อไป

Table 5 The summary of the best method for the relationship between the extreme values of rainfall and the rice yields in the lower northern provinces

Provinces	Method	R-squared		Efficiency (Ton/Rai) for zp			Efficiency (Ton/Rai) for pr		
		zp	pr	MAE	RMSE	MSE (x10 ⁵)	MAE	RMSE	MSE (x10 ⁵)
Nakhonsawan	RF	0.79	0.82	160.18	184.85	341.69	209.86	224.07	504.87
Phichit	RF	0.62	0.60	100.07	102.55	105.16	171.21	208.75	435.77
Phitsanulok	RF	0.73	0.65	101.13	125.01	156.27	136.63	149.52	223.57
Kamphaengphet	RF	0.64	0.63	160.33	177.21	314.02	162.23	180.01	324.02
Phetchabun	RF	0.61	0.77	114.21	169.74	130.28	45.47	57.94	33.57
Sukhothai	RF	0.77	0.73	151.73	370.53	192.49	134.55	152.59	242.37

zp is the model that zindex_prec is independent variable, pr is the model that precipitation is independent variable

ผลประโยชน์ทับซ้อน

ผู้เขียนขอประกาศว่าบทความนี้ไม่มีผลประโยชน์ทับซ้อน

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงโดยได้รับความอนุเคราะห์และเงินสนับสนุนจากสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) งบประมาณแผ่นดิน (แบบปกติ) ประจำปีงบประมาณ 2565 และสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏพิบูลสงคราม

การมีส่วนร่วมในการเขียนบทความของผู้เขียน

ความคิดริเริ่ม สมมติฐาน: ศรัณญา ทองสุข. วิธีการเก็บข้อมูล การจัดเก็บข้อมูล การวิเคราะห์ข้อมูล การแปลผล: ศรัณญา ทองสุข, ญัฐินี ดีแท้. การสรุปผล การวิพากษ์วิจารณ์ผล การแสดงการเปรียบเทียบข้อสรุป: ศรัณญา ทองสุข. การมีส่วนร่วมในการเขียนต้นฉบับบทความ: ศรัณญา ทองสุข.

เอกสารอ้างอิง

- Alidoost, F., Su, Z., & Stein, A. (2019). Evaluating the effects of climate extremes on crop yield, production and price using multivariate distributions: A new copula application. **Weather and Climate Extremes**, 26(1), 1-9.
- Chen, W. Z., Zhu, D., Huang, C. J., & Ciais, P. (2019). Negative extreme events in gross primary productivity and their drivers in China during the past three decades. **Agricultural and Forest Meteorology**, 275(1), 47-58.
- Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. **Environmental Research Letter**, 13(1), 114003.
- Fan, J., Jintrawet, A., & Sangchyoswat, C. (2020). The relationships between extreme precipitation and rice and maize yields using machine learning in Sichuan province, China. **Current Applied Science and Technology**, 20(3), 453-469.
- Feng, S. F., Hao, Z. C., Zhang, X., & Hao, F. H. (2019). Probabilistic evaluation of the impact of compound dry-hot events on global maize yields. **Science of the Total Environment**, 689(1), 1228-1234.
- Harrison, M. T., Cullen, B. R., & Rawnsley, R. P. (2016). Modelling the sensitivity of agricultural systems to climate change and extreme climatic events. **Agricultural Systems**, 148(1), 135-148.
- Konduri, V. S., Vandal, T. J., Ganguly, S., & Ganguly, A. R. (2020). Data science for weather impacts on crop yield. **Frontiers in Sustainable Food System**, 52(4), 1-11.
- Jeong, J. H., Resop, J. P., Mueller, N. D., & Fleisher, D. H. (2016). Random forests for global and regional crop yield predictions. **Plos One**, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
- Rotter, R. P., Appiah, M., Fichtler, E., & Kersebaum, K. C. (2018). Linking modelling and experimentation to better capture crop impacts of agroclimatic extremes-A review. **Field Crops Research**, 221(1), 142-156.
- Vogel, E., Donat, M. G., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., Meinshausen, N., & Frieler, K. (2019). The effects of climate extremes on global agricultural yields. **Environmental Research Letter**, 14(5), 054010.