

การพยากรณ์ผลผลิตข้าวนาปรังในภาคตะวันออกเฉียงเหนือของประเทศไทย

ศิรา จิตต์ใจฉ่ำ¹ เกษมสันต์ ชัยรัตน์^{2,*} ธัญญวรรณ ส่งเสริม³ และ กัลยณัฐ กุหลาบเพชรทอง⁴

¹นักศึกษาปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์และนวัตกรรมข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนสุนันทา

²นักศึกษาปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์และนวัตกรรมข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนสุนันทา

³นักศึกษาปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์และนวัตกรรมข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนสุนันทา

⁴รองศาสตราจารย์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนสุนันทา

*Corresponding author. E-mail: s64122250054@ssru.ac.th

บทคัดย่อ

การวิจัยครั้งนี้เป็นการศึกษาการพยากรณ์ผลผลิตข้าวนาปรังในภาคตะวันออกเฉียงเหนือของประเทศไทย มีวัตถุประสงค์เพื่อหาวิธีการที่เหมาะสมสำหรับการพยากรณ์ผลผลิตข้าวนาปรัง เทคนิคการพยากรณ์ที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ K-Nearest Neighbors Regression (KNN-R) และ Multiple Linear Regression (MLR) และใช้ค่าสัมประสิทธิ์การตัดสินใจ (R2) ค่าความคลาดเคลื่อน (MSE) และรากของค่าความคลาดเคลื่อนกำลังสอง (RMSE) เป็นเกณฑ์ในการเปรียบเทียบวิธีการพยากรณ์ทั้ง 2 วิธี ในการวิจัยครั้งนี้ ข้อมูลที่นำมาศึกษานั้น เป็นข้อมูลทุติยภูมิรายปี จากแหล่งข้อมูล 2 แหล่ง คือ สำนักงานเศรษฐกิจการเกษตร และสถาบันนวัตกรรมและกรรมาภิบาลข้อมูล

ผลเปรียบเทียบตัวแบบพยากรณ์ที่ได้จากเทคนิคการพยากรณ์ที่ใช้ในการวิเคราะห์ข้อมูลทั้ง 2 วิธีพิจารณาจากค่าสัมประสิทธิ์การตัดสินใจ (R2) ค่าความคลาดเคลื่อน (MSE) และรากของค่าความคลาดเคลื่อนกำลังสอง (RMSE) สัดส่วนของข้อมูลที่ใช้คือ 80-20 หมายถึง ข้อมูลสำหรับการเรียนรู้ 80% และการทดสอบ 20% พบว่า พบว่า เทคนิคที่มีประสิทธิภาพดีที่สุดในการพยากรณ์ผลผลิตข้าวนาปรังโดยใช้ปริมาณน้ำฝน คือ เทคนิคแบบ K-Nearest Neighbors Regression โดยมีค่าสัมประสิทธิ์การตัดสินใจมากที่สุด (0.85) ค่าความคลาดเคลื่อน (46,260,190.62) และรากของค่าความคลาดเคลื่อนกำลังสอง (6,801.48) น้อยที่สุด ส่วนตัวแบบที่มีประสิทธิภาพในการพยากรณ์ผลผลิตข้าวนาปรังโดยใช้ปริมาณน้ำได้ถูกต้องแม่นยำน้อยที่สุด คือ เทคนิคแบบ Multiple Linear Regression โดยมีค่าสัมประสิทธิ์การตัดสินใจน้อยที่สุด (0.68) ค่าความคลาดเคลื่อน (98,104,943.63) และรากของค่าความคลาดเคลื่อนกำลังสอง (9,904.79) มากที่สุด

คำสำคัญ: การพยากรณ์, ผลผลิตข้าวนาปรัง, ภาคตะวันออกเฉียงเหนือ

Forecasting Glutinous Rice Production in the Northeastern Region of Thailand

Sira Jitjaicham¹ Kasemsan Chairat^{2,*} Tanyawan Songserm³ and Kunyanuth Kularbphetong⁴

¹Undergraduate student, Department of Computer Science and Data Innovation, Faculty of Science and Technology, Suan Sunandha Rajabhat University

²Undergraduate student, Department of Computer Science and Data Innovation, Faculty of Science and Technology, Suan Sunandha Rajabhat University

³Undergraduate student, Department of Computer Science and Data Innovation, Faculty of Science and Technology, Suan Sunandha Rajabhat University

⁴Associate Professor, Faculty of Science and Technology, Suan Sunandha Rajabhat University

*Corresponding author. E-mail: s64122250054@ssru.ac.th

Abstract

This research study examines the forecasting of glutinous rice production in the Northeastern region of Thailand. The objectives are to find suitable methods for forecasting glutinous rice production, using the K-Nearest Neighbors Regression (KNN-R) and Multiple Linear Regression (MLR) techniques. Decision coefficient (R^2), mean squared error (MSE), and root mean square error (RMSE) are used as criteria for comparing the two forecasting methods. The data used in this study are annual geographical data from two sources: the Department of Agricultural Economics and the Institute of Innovation and Data Ethics.

Comparing the forecasting models obtained from the two data analysis techniques using R^2 , MSE, and RMSE, with an 80-20 data ratio (meaning 80% for training and 20% for testing), it was found that the most efficient technique for forecasting glutinous rice production using rainfall quantity is the K-Nearest Neighbors Regression technique, with the highest decision coefficient (0.85), the lowest MSE (46,260,190.62), and the lowest RMSE (6,801.48). On the other hand, the model with the least accuracy in forecasting glutinous rice production using water quantity is the Multiple Linear Regression technique, with the lowest decision coefficient (0.68), the highest MSE (98,104,943.63), and the highest RMSE (9,904.79).

Keywords: Forecasting, Glutinous Rice Production, Northeastern Region

บทนำ

การผลิตข้าวนาปรังเป็นกิจกรรมสำคัญทางเศรษฐกิจในหลายภูมิภาคทั่วโลก อย่างไรก็ตาม, ผลผลิตข้าวนาปรังมีการพบเจอกับความผันผวนของสภาพอากาศและปริมาณน้ำฝนที่มีผลกระทบต่อผลผลิตข้าว ปริมาณน้ำฝนที่เข้ามาในพื้นที่นาปรังมีผลแก่การเจริญเติบโตของข้าว การสูญเสียที่เกิดขึ้นจากน้ำฝนที่น้อยหรือมากเกินไปสามารถก่อให้เกิดเกษตรกรขาดทุนมากมาย นอกจากนี้, ปริมาณน้ำฝนมีความผันผวนในระยะเวลาที่สั้นๆ ทำให้การวางแผนการผลิตข้าวนาปรังเป็นเรื่องยากและท้าทายมากยิ่งขึ้น เพื่อแก้ไขปัญหาและช่วยเกษตรกรในการจัดการแปรปรวนการเกษตรและการวางแผนการผลิตข้าวอย่างมีประสิทธิภาพในสภาวะที่มีความผันผวนของปริมาณน้ำฝนในพื้นที่นาปรัง, งานวิจัยนี้มุ่งสู่การพยากรณ์ผลผลิตข้าวนาปรังโดยใช้ปริมาณน้ำฝนเป็นตัวแปรหลัก. การวิจัยนี้นำเสนอการใช้โมเดลทางสถิติและแบบจำลองการเรียนรู้เครื่องในการทำนายผลผลิตข้าวนาปรัง ผ่านการวิเคราะห์ปริมาณน้ำฝนที่เข้ามาในระบบ, รวมถึงการเรียนรู้จากประสบการณ์ในอดีตและข้อมูลปริมาณน้ำฝนปัจจุบัน เราเชื่อว่าการใช้เทคโนโลยีทางสารสนเทศในการวิเคราะห์ข้อมูลและการสื่อสารกับเกษตรกรจะช่วยให้การพยากรณ์นั้นมีประสิทธิภาพมากยิ่งขึ้น

การวิจัยนี้เป็นการผสมผสานระหว่างการพยากรณ์อากาศและเทคโนโลยีทางสารสนเทศเพื่อสร้างระบบพยากรณ์ที่ทันสมัยและมีความแม่นยำ นอกจากนี้, งานวิจัยนี้ยังเน้นการสร้างแนวทางการจัดการที่ดีขึ้นสำหรับเกษตรกรในการปรับตัวเองต่อสภาวะที่มีการเปลี่ยนแปลงที่ไม่ค่อยแน่นอนในสภาพภูมิธรรมชาติและเศรษฐกิจ การวิจัยนี้หวังว่าจะช่วยเพิ่มผลผลิตข้าวนาปรังและเสริมความยั่งยืนในการเกษตรในพื้นที่นาปรังที่มีอัตราการใช้น้ำเพียงพอและความขาดแคลนของน้ำเป็นปัญหาหลักในการผลิตข้าวนาปรัง

วัตถุประสงค์ของการวิจัย

1. การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษา และพัฒนาระบบการคาดเดาผลผลิตข้าวนาปรังที่มีความแม่นยำ โดยใช้ข้อมูลจากพื้นที่เพาะปลูก, ผลผลิตต่อไร่ และปริมาณน้ำฝน ในการพยากรณ์
2. ค้นหาวิธีการที่เหมาะสมสำหรับการพยากรณ์ผลผลิตข้าวนาปรัง

สมมติฐานของการวิจัย

ปัจจัยที่เกี่ยวข้องกับผลผลิตข้าวนาปรังในภาคตะวันออกเฉียงเหนือของประเทศไทย ในช่วงที่ผ่านมา พ.ศ. 2564 - 2565 (ค.ศ. 2021 - 2022)

ขอบเขตของการวิจัย

การวิจัยครั้งนี้ เป็นการพยากรณ์ผลผลิตข้าวนาปรังในบริเวณภาคตะวันออกเฉียงเหนือของประเทศไทย โดยเก็บรวบรวมข้อมูลตัวแปรแต่ละตัวของจังหวัดในภาคตะวันออกเฉียงเหนือ 20 จังหวัด ได้แก่ จังหวัด บึงกาฬ หนองคาย นครพนม สกลนคร อุดรธานี หนองบัวลำภู เลย มุกดาหาร กาฬสินธุ์ ขอนแก่น อำนาจเจริญ ยโสธร ร้อยเอ็ด มหาสารคาม ชัยภูมิ นครราชสีมา บุรีรัมย์ สุรินทร์ ศรีสะเกษ และอุบลราชธานี ตั้งแต่ปี พ.ศ. 2564 ถึงปี พ.ศ.2565 ดังนี้

1. ผลผลิตข้าวนาปรังในบริเวณภาคตะวันออกเฉียงเหนือ
2. ปริมาณน้ำฝนต่ำสุด, ปริมาณน้ำฝนสูงสุด, และปริมาณน้ำฝนโดยเฉลี่ย

วิธีดำเนินการวิจัย

1. รวบรวมข้อมูลที่ใช้ในการดำเนินการวิจัยจากแหล่งต่าง ๆ ดังนี้ สำนักงานสถิติแห่งชาติจากกรมอุตุนิยมวิทยา และ สำนักงานเศรษฐกิจการเกษตร โดยเลือกใช้ข้อมูลในช่วงปี พ.ศ. 2564 – 2565 (ค.ศ. 2021 – 2022)

- ข้อมูลปริมาณข้าวนาปรัง

ตารางที่ 1 ข้อมูลปริมาณข้าวนาปรัง

Feature	ความหมาย
ProvinceName	จังหวัด
Prod	ผลผลิต (ตัน)
Prop	สัดส่วน
HarvArea	เนื้อที่เก็บเกี่ยว (ไร่)
ProdPerHarvArea	ผลผลิตต่อเนื้อที่เก็บเกี่ยว (กก.)

- ข้อมูลปริมาณน้ำฝนของประเทศไทย ในแต่ละจังหวัด

ตารางที่ 2 ข้อมูลปริมาณน้ำฝน

Feature	ความหมาย
ProvinceID	รหัสจังหวัด
ProvinceName	ชื่อจังหวัด(ไทย)
ProvinceName 2	ชื่อจังหวัด(อังกฤษ)
MinRain	ปริมาณน้ำฝนต่ำสุดต่อเดือน
MaxRain	ปริมาณน้ำฝนสูงสุดต่อเดือน
AvgRain	ปริมาณน้ำฝนเฉลี่ยต่อเดือน
region	ภูมิภาคของประเทศไทย
Year	ปีที่บ้านทีกข้อมูลปริมาณน้ำฝน
Month	ลำดับที่ของเดือน
Date	วันที่บ้านทีกข้อมูลปริมาณน้ำฝน
Month_TH	ชื่อเดือน(ไทย)

2. ตรวจสอบ และทำความสะอาดข้อมูลโดยตรวจสอบความถูกต้องสมบูรณ์ และดำเนินการแปลงข้อมูลที่เหมาะสม
3. ทำความสะอาดข้อมูล เพื่อแก้ไขหากข้อมูลที่นำมาใช้มีความไม่ถูกต้องสมบูรณ์
4. แบ่งชุดข้อมูลออกเป็นชุดข้อมูลสำหรับการฝึกอบรม และชุดข้อมูลสำหรับการทดสอบ
 - ชุดข้อมูลฝึกอบรม 80%
 - ชุดข้อมูลทดสอบ 20%
5. สร้าง และฝึกฝนโมเดลระหว่าง K-Nearest Neighbors Regression กับ Model Multiple Linear Regression
6. เปรียบเทียบความแม่นยำระหว่าง KNN-R กับ MLR

ผลการวิจัย

ในการวิจัย “การพยากรณ์ผลผลิตข้าวนาปรังในภาคตะวันออกเฉียงเหนือของประเทศไทย” มีผลการวิจัยดังนี้

1. ผลลัพธ์ของการพยากรณ์

- K-Nearest Neighbors Regression (KNN-R)

ค่า R^2 : 0.85

ค่า MSE: 46,260,190.62

ค่า RMSE: 6,801.48

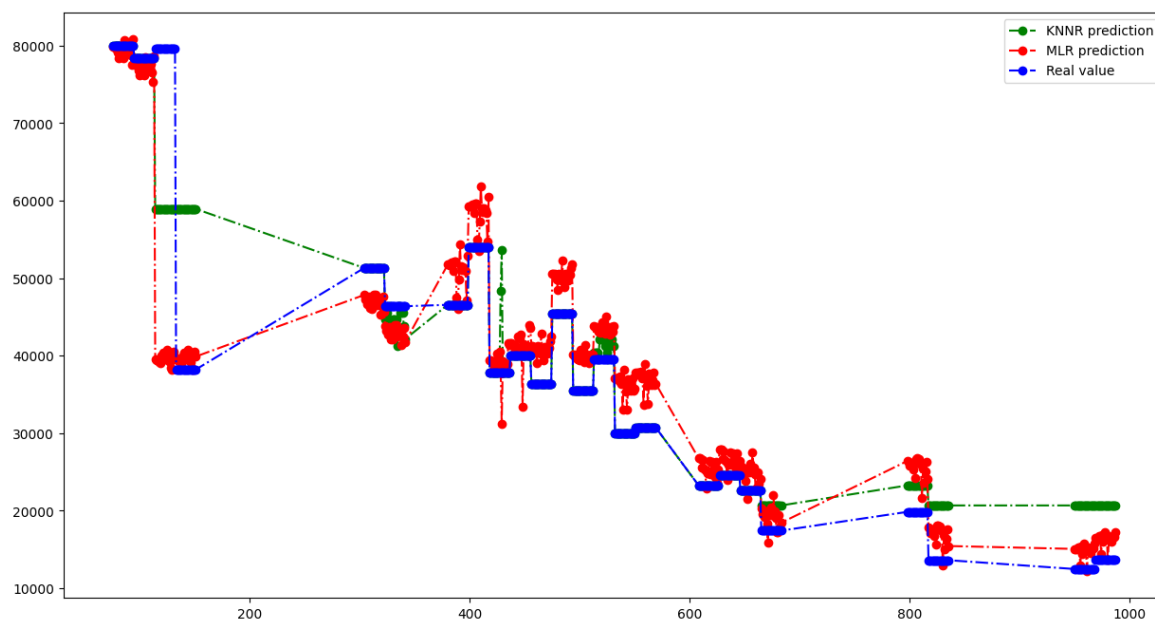
- Multiple Linear Regression (MLR)

ค่า R^2 : 0.68

ค่า MSE: 98,104,943.63

ค่า RMSE: 9,904.79

2. ผลการประเมินประสิทธิภาพระหว่างเทคนิคการพยากรณ์แบบ KNN-R และ MLR แสดงผล ดังนี้



ภาพที่ 1 ผลการเปรียบเทียบประสิทธิภาพของเทคนิค KNN-R และ MLR และค่าจริง (Real Value)

ตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพของเทคนิค KNN-R และ MLR

เทคนิค	R^2	MSE	RMSE
K-Nearest Neighbors Regression (KNN-R)	0.85	46,260,190.62	6,801.48
Multiple Linear Regression (MLR)	0.68	98,104,943.63	9,904.79

จากตารางที่ 1 พบว่า เทคนิคที่มีประสิทธิภาพดีที่สุดในการพยากรณ์ผลผลิตข้าวนาปรังโดยใช้ปริมาณน้ำฝน คือ เทคนิคแบบ K-Nearest Neighbors Regression โดยมีค่าสัมประสิทธิ์การตัดสินใจมากที่สุด (0.85) ค่าความคลาดเคลื่อน (46,260,190.62) และรากของค่าความคลาดเคลื่อนกำลังสอง (6,801.48) น้อยที่สุด

ส่วนเทคนิคที่มีประสิทธิภาพในการพยากรณ์ผลผลิตข้าวนาปรังโดยใช้ปริมาณน้ำได้ถูกต้อง แม่นยำน้อยที่สุด คือ เทคนิคแบบ Multiple Linear Regression โดยมีค่าสัมประสิทธิ์การตัดสินใจน้อยที่สุด (0.68) ค่าความคลาดเคลื่อน (98,104,943.63) และรากของค่าความคลาดเคลื่อนกำลังสอง (9,904.79) มากที่สุด

สรุปผลการวิจัย

จากตัวชี้วัดต่างๆ ที่กำหนดไว้ ทำให้พบว่าเทคนิค K-Nearest Neighbors Regression (KNN-R) มีประสิทธิภาพในการพยากรณ์ที่แม่นยำกว่า Multiple Linear Regression (MLR) ดังนี้

1. KNN-R มีค่าสัมประสิทธิ์การตัดสินใจ (R^2) ที่มากกว่า MLR กล่าวคือ KNN-R สามารถตัดสินใจได้ดีกว่า MLR
2. KNN-R มีค่าความคลาดเคลื่อน (MSE) ที่น้อยกว่า MLR กล่าวคือ KNN-R มีประสิทธิภาพในการพยากรณ์ได้ดีกว่า MLR
3. KNN-R มีรากของค่าความคลาดเคลื่อนกำลังสอง (RMSE) ที่น้อยกว่า MLR กล่าวคือ KNN-R มีความถูกต้องแม่นยำมากกว่า MLR

ผลลัพธ์ที่ได้แสดงให้เห็นว่า K-Nearest Neighbors Regression (KNN-R) มีประสิทธิภาพในการทำนายที่ดีกว่า Multiple Linear Regression (MLR) โดยมีค่า R^2 สูงกว่า (0.85 เมื่อเปรียบเทียบกับ 0.68) ค่า MSE น้อยกว่า (46,260,190.62 เมื่อเปรียบเทียบกับ 98,104,943.63) และค่า RMSE น้อยกว่า (6,801.48 เมื่อเปรียบเทียบกับ 9,904.79) ซึ่งหมายความว่า KNN-R มีความเป็นไปได้ที่จะให้คำทำนายที่เที่ยงตรงและเป็นประโยชน์มากกว่า MLR

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ รศ. กัลยณัฐ กุหลาบเพชรทอง อาจารย์ที่ปรึกษางานวิจัย ที่ให้คำปรึกษาในการเรียนการค้นคว้าวิจัย ตลอดจนการตรวจแก้ไขงานวิจัยจนกระทั่งเสร็จสมบูรณ์

ขอกราบขอบพระคุณสำนักงานสถิติแห่งชาติจากกรมอุตุนิยมวิทยา และสำนักงานเศรษฐกิจการเกษตร ที่เป็นแหล่งข้อมูลในการทำการศึกษการวิจัยในครั้งนี้

สุดท้ายนี้ความดีหรือประโยชน์อันใดที่เกิดจากการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยขอมอบแต่บิดา มารดา ครู อาจารย์ และผู้มีพระคุณทุกท่าน ที่ได้ให้ความกรุณาและให้กำลังใจมาโดยตลอด

เกษมสันต์ ชัยรัตน์
กุมภาพันธ์ 2567

เอกสารอ้างอิง

- สำนักงานเศรษฐกิจการเกษตร. (2564). *สินค้าเกษตรข้าวนาปรัง*. วันที่ค้นข้อมูล 19 ตุลาคม 2566, เข้าถึงได้จาก <https://mis-app.oae.go.th/product/ข้าวนาปรัง>
- สถาบันนวัตกรรมและยุทธศาสตร์ข้อมูล. (2564). *ข้อมูลปริมาณน้ำฝนในประเทศไทย*. วันที่ค้นหาข้อมูล 15 ตุลาคม 2566, เข้าถึงได้จาก <https://digi.data.go.th/showcase/thailand-rainfall-data/>
- ปรีชา กาพย์เพชร และเกริก ปั่นแห่งเพชร. (2555). ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการผลิตอ้อยในภาคตะวันออกเฉียงเหนือของประเทศไทย : พื้นที่ศึกษา จังหวัดกาฬสินธุ์. *แก่นเกษตร 40 ฉบับพิเศษ 3*, 83-91.
- วิภาพร ฉิมณรงค์. (2552). ความสัมพันธ์ระหว่างข้อมูลสภาพภูมิอากาศกับผลผลิตอ้อย: กรณีศึกษาพื้นที่ปลูกอ้อยโรงงานน้ำตาลรวมเกษตรกรรมอุตสาหกรรม จำกัด. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์สิ่งแวดล้อม บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น.
- นรวัฒน์ เหลืองทอง. (2558). การเลือกตัวแบบพยากรณ์ผลผลิตการเกษตรที่เหมาะสม. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิศวกรรมอุตสาหการ บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์.
- ศุขมา โชคเพิ่มพูน, เตือนใจ น้อยพา และกนกทิพย์ โคตรสำราญ. (2561). การทำนายปริมาณสารไลโคปีนในผักข้าวด้วยการวิเคราะห์ภาพสี RGB ร่วมกับวิธีการโครงข่ายประสาทเทียม. *คณะวิทยาศาสตร์และวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตเฉลิมพระเกียรติ*.
- ศรียักษ์ ศรีทองชัย. (2566). การเปรียบเทียบวิธีการประมาณค่าสูญหายในการพยากรณ์ความเข้มข้นของ PM2.5 ด้วยโครงข่ายประสาทเทียม LSTM. *ภาควิชาวิศวกรรมอุตสาหการและการจัดการ คณะ วิศวกรรมศาสตร์และเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์*.
- วีระยุทธ พิมพ์ภรณ์. (2561). การพยากรณ์ผลสัมฤทธิ์ทางการศึกษาด้วยวิธีการจัดการเรียนการสอนแบบผสมผสานโดยใช้เทคนิคเหมืองข้อมูล. *วารสารศรีปทุมปริทัศน์*, 10(15), 8. เข้าถึงได้จาก <https://www.lib.cmru.ac.th/web62/core/FILE/1565321805.pdf>
- ศรัญญา ทองสุข, โสภณา สำราญ และณัฐินี ดีแท้. (2566). แบบจำลองความสัมพันธ์ค่าสุดขีดของปริมาณ น้ำฝนและอุณหภูมิกับผลผลิตข้าวปลอดภัยในจังหวัดพิษณุโลก. *Life Sciences and Environment Journal*, 24(2), 533. เข้าถึงได้จาก <https://doi.org/10.14456/lsej.2023.39>
- กนกพร ภาศิฉาย, นิโรจน์ สิ้นณรงค์, กฤตวิทย์ อัจฉริยะพานิชกุล และพัชรินทร์ สุภาพันธ์. (2563). ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อผลผลิตข้าวในพื้นที่ภาคกลางของประเทศไทย. *วารสารเศรษฐศาสตร์และกลยุทธ์การจัดการ*, 7(2), 17-18. เข้าถึงได้จาก <https://kuojs.lib.ku.ac.th/index.php/jems/article/view/2825/1888>
- อดิศักดิ์ พาศิริราษฎร์, สิทธิรัตน์ เพิ่มहरษา, ธนพล พิมพ์อุบล และเปรม รังสิวณิพงษ์. (2563). การศึกษาผลกระทบของปรากฏการณ์เอลนีโญและลานีญาที่มีต่อปริมาณน้ำฝนและผลผลิตของข้าวในลุ่มน้ำยม. *ภาควิชาวิศวกรรมทรัพยากรน้ำ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ*.