

พยากรณ์ผลผลิตข้าวในประเทศไทยโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร

รัตนา คนไว^{1*}, อาภัสรา พิมพ์อักษร¹, เมธิส ประกอบ¹ และกัลยณัฐ กุหลาบเพ็ชรทอง¹

¹สาขาวิทยาการคอมพิวเตอร์และนวัตกรรมข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนสุนันทา

*s65122250043@ssru.ac.th

บทคัดย่อ

บทความนี้นำเสนอการศึกษาเชิงปริมาณเพื่อวิเคราะห์และพยากรณ์ผลผลิตข้าวในประเทศไทยโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร (Machine Learning) สองรูปแบบ ได้แก่ การถดถอยเชิงเส้น (Linear Regression) และการถดถอยแบบสุ่มป่า (Random Forest Regression) ข้อมูลที่ใช้ในการวิเคราะห์ครอบคลุมช่วงปี พ.ศ. 2565 และ 2566 ซึ่งประกอบด้วยตัวแปรที่เกี่ยวข้องกับผลผลิตข้าว เช่น ปีการผลิต ปริมาณน้ำฝนสูงสุดและต่ำสุด คุณสมบัติน้ำ อุณหภูมิสูงสุดและต่ำสุด รวมถึงผลผลิตข้าว

การพัฒนาแบบจำลองมีวัตถุประสงค์เพื่อพยากรณ์ผลผลิตข้าว โดยการประเมินสมรรถนะของแบบจำลองดำเนินการโดยพิจารณาค่าสัมประสิทธิ์การตัดสินใจ (R^2) และค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAE) ผลการทดลองบ่งชี้ว่าแบบจำลองการถดถอยแบบสุ่มป่า มีความสามารถในการพยากรณ์ผลผลิตข้าวได้ดีกว่าแบบจำลองการถดถอยเชิงเส้น โดยมีค่า R^2 ที่ 38.19% ซึ่งสูงกว่า 30.45% ของแบบจำลองการถดถอยเชิงเส้น

อย่างไรก็ตาม ผลการประเมินยังชี้ให้เห็นว่าแบบจำลองทั้งสองยังคงมีความคลาดเคลื่อนในการพยากรณ์ในระดับหนึ่ง การศึกษาแนะนำแนวทางในการพัฒนาแบบจำลองในอนาคต โดยเน้นการปรับแต่งพารามิเตอร์ของแบบจำลอง การคัดเลือกคุณลักษณะที่มีผลต่อผลผลิตข้าว รวมถึงการเพิ่มคุณภาพและปริมาณของข้อมูลที่ใช้ในการฝึกแบบจำลอง โดยเฉพาะการขยายขอบเขตข้อมูลให้ครอบคลุมช่วงเวลาที่ยาวนานขึ้นและตัวแปรที่หลากหลายมากขึ้น เพื่อเพิ่มความแม่นยำและความน่าเชื่อถือในการพยากรณ์ผลผลิตข้าว ซึ่งจะมีประโยชน์ในการวางแผนและกำหนดนโยบายภาคการเกษตรของประเทศไทย

คำสำคัญ: ข้าว พยากรณ์ การถดถอยเชิงเส้น การถดถอยแบบสุ่มป่า การเรียนรู้ของเครื่องจักร

Forecasting Rice Yield in Thailand Using Machine Learning Techniques

Ruttana Khonwai ^{1*}, Apatsara Pimoakson¹, Methis Prakob¹
and Kunyanuth Kularbphetong¹

¹Computer Science and Data Innovation Program, Faculty of Science and Technology

Suan Sunandha Rajabhat University

*s65122250043@ssru.ac.th

Abstract

Linear regression and random forest regression (RFR), two machine learning approaches, are used in this study's quantitative analysis to predict rice yield in Thailand. The analysis utilizes data from 2022 and 2023, encompassing variables pertinent to rice production, including production year, maximum and minimum rainfall, soil characteristics, maximum and minimum temperatures, and rice output. The models were developed with the intention of predicting rice yield, and their performance was assessed using the Mean Absolute Error (MAE) and coefficient of determination (R^2). Experimental results demonstrate that the Random Forest Regression model surpasses the Linear Regression model, attaining a R^2 value of 38.19%, beyond the 30.45% achieved by the Linear Regression model.

Nonetheless, the evaluation results indicate that both models continue to have a degree of predictive error. The research indicates prospective enhancements in model construction, emphasizing parameter optimization, the selection of pertinent factors influencing rice yield, and the augmentation of both the quality and quantity of data utilized for model training. Expanding the data scope to encompass a longer time frame and incorporate a broader range of variables would enhance forecast accuracy and reliability. This would enhance planning and policy development in Thailand's agriculture industry.

Keywords: Rice, Forecasting, Linear regression, random forest regression (RFR), machine learning

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ข้าวจัดเป็นพืชเศรษฐกิจที่มีความสำคัญอย่างยิ่งต่อภาคการเกษตรและขับเคลื่อนเศรษฐกิจของหลายประเทศทั่วโลก โดยเฉพาะประเทศไทยซึ่งเป็นหนึ่งในผู้ผลิตและส่งออกข้าวรายสำคัญ [1,3] อย่างไรก็ตาม การเพาะปลูกข้าวนอกเขตชลประทานต้องเผชิญกับความท้าทายหลากหลายประการ อาทิ การขาดแคลนน้ำ ความผันผวนของสภาพภูมิอากาศ และปัจจัยทางกายภาพของดิน [2,4] ซึ่งปัจจัยเหล่านี้ล้วนส่งผลกระทบต่อผลผลิตข้าว และนำไปสู่ความไม่แน่นอนของผลผลิต รวมถึงรายได้ของเกษตรกรผู้เพาะปลูก

ด้วยความก้าวหน้าของเทคโนโลยี การเรียนรู้ของเครื่องจักร ได้เปิดโอกาสให้สามารถนำข้อมูลทางสิ่งแวดล้อมและปัจจัยทางการเกษตรมาวิเคราะห์เพื่อพัฒนาระบบการพยากรณ์ผลผลิตข้าวที่มีประสิทธิภาพสูงขึ้น [1,3,5] แบบจำลองทางสถิติ

และปัญญาประดิษฐ์ เช่น การถดถอยเชิงเส้น และการถดถอยป่าไม้สุ่ม มีศักยภาพในการช่วยวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยแวดล้อมกับผลผลิตข้าวได้อย่างแม่นยำยิ่งขึ้น [3,5] ส่งผลให้สามารถพัฒนาและวางกลยุทธ์ในการบริหารจัดการการเพาะปลูกได้อย่างมีประสิทธิภาพ ลดความเสี่ยงจากความแปรปรวนของสภาพอากาศ และเพิ่มผลผลิตได้ในที่สุด [1,5]

การศึกษานี้จึงมุ่งเน้นไปที่การพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องจักรเพื่อพยากรณ์ผลผลิตข้าว โดยใช้ข้อมูลปัจจัยทางสิ่งแวดล้อมที่สำคัญ เช่น อุณหภูมิ ปริมาณน้ำฝน คุณสมบัติของดิน และปริมาณผลผลิตข้าว ซึ่งเป็นข้อมูลของปีพุทธศักราช 2565 และ 2566 [2,4] ผลลัพธ์ของการศึกษานี้จะเป็นประโยชน์อย่างยิ่งต่อเกษตรกรและผู้กำหนดนโยบายด้านการเกษตร ในการวางแผนการเพาะปลูก ที่เหมาะสม ช่วยเพิ่มผลผลิต และลดความเสี่ยงที่อาจเกิดขึ้นจากปัจจัยที่ควบคุมไม่ได้ [1,3,5] อันจะนำไปสู่การพัฒนาภาคการเกษตรอย่างยั่งยืน

1.2 วัตถุประสงค์

1. เพื่อพัฒนาและทดสอบโมเดลการทำนายผลผลิตข้าวโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร ได้แก่ Linear Regression และ Random Forest Regression
2. เพื่อเปรียบเทียบประสิทธิภาพของโมเดลที่ใช้ในการทำนายผลผลิตข้าว โดยใช้ตัวชี้วัดทางสถิติเพื่อประเมินความแม่นยำของแต่ละโมเดล

1.3 ขอบเขตการวิจัย

1. ปัจจัยที่ศึกษา คือการศึกษาจะพิจารณาปัจจัยทางสภาพแวดล้อมที่ส่งผลต่อผลผลิตข้าว โดยข้อมูลที่ใช้จะครอบคลุมช่วงปีพุทธศักราช 2565 และ 2566 ซึ่งประกอบด้วย อุณหภูมิสูงสุดและต่ำสุด, ปริมาณน้ำฝนสูงสุด และต่ำสุด คุณสมบัติของดิน และปริมาณผลผลิตข้าว
2. การศึกษานี้ประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องจักร เพื่อพัฒนาแบบจำลองสำหรับการพยากรณ์ผลผลิตข้าว โดยพิจารณาแบบจำลอง 2 รูปแบบ ได้แก่ การถดถอยเชิงเส้น และ การถดถอยแบบสุ่มป่าจากนั้นจะเปรียบเทียบสมรรถนะของแบบจำลองทั้งสองเพื่อระบุแบบจำลองที่มีความแม่นยำสูงสุดในการพยากรณ์ผลผลิตข้าว

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. เพิ่มประสิทธิภาพการผลิตข้าว ช่วยให้เกษตรกรสามารถวางแผนและใช้ทรัพยากรอย่างมีประสิทธิภาพ เพื่อเพิ่มผลผลิต
2. ลดความเสี่ยงจากสภาพแวดล้อม ช่วยให้เตรียมพร้อมรับมือกับความไม่แน่นอนของสภาพอากาศและปัจจัยที่ควบคุมไม่ได้
3. สนับสนุนการวิจัยและเทคโนโลยีเกษตรเป็นแนวทางในการพัฒนาเทคโนโลยีและเครื่องมือช่วยการเพาะปลูกในอนาคต

2. วิธีวิจัย/วิธีการศึกษา

2.1 ประชากร/กลุ่มเป้าหมาย

2.1.1 ประชากร

ประชากร ในที่นี้คือ ข้อมูลผลผลิตข้าวและข้อมูลสิ่งแวดล้อมทั้งหมดที่เกี่ยวข้องจากพื้นที่เพาะปลูกข้าวทั่วประเทศไทย ซึ่งเป็นข้อมูลของปีพุทธศักราช 2565 และ 2566

2.1.2 กลุ่มเป้าหมาย

กลุ่มเป้าหมาย หลักของการศึกษานี้คือผู้ที่คาดว่าจะได้รับประโยชน์จากผลการวิจัยโดยตรง ได้แก่

1. เกษตรกร
2. ผู้กำหนดนโยบายด้านการเกษตร

2.2 เครื่องมือและการหาคุณภาพเครื่องมือ

2.2.1 เครื่องมือ

เครื่องมือที่ใช้ในการวิจัย คือ โปรแกรมและภาษาคอมพิวเตอร์ พร้อมทั้งไลบรารีที่สนับสนุนการทำงานด้านวิทยาการข้อมูล และการสร้างแบบจำลอง เพื่อพยากรณ์ผลผลิตข้าวในประเทศไทยโดยใช้ เทคนิคการเรียนรู้ของเครื่องจักรได้แก่

1. ไพทอน (Python) เป็นภาษาโปรแกรมหลักที่ใช้ในการวิเคราะห์ข้อมูลและพัฒนาแบบจำลอง เนื่องจากมีไลบรารีที่หลากหลายและแข็งแกร่งสำหรับการเรียนรู้ของเครื่อง [1,3,5]

2. ไลบรารี Python สำหรับวิทยาการข้อมูลและการเรียนรู้ของเครื่องหลัก ๆ ได้แก่

2.1 Pandas ใช้สำหรับการจัดการและปรับแต่งข้อมูล (Data Manipulation and Preprocessing) เช่น การนำเข้าข้อมูลจากไฟล์, การรวมข้อมูล, การจัดการกับข้อมูลที่ขาดหายไป (Missing Values), และการจัดโครงสร้างข้อมูล [1,3,5]

2.2 NumPy ใช้สำหรับการคำนวณเชิงตัวเลขที่มีประสิทธิภาพ โดยเฉพาะการทำงานกับอาร์เรย์ (Arrays) และเมทริกซ์ (Matrices) [1,3,5]

2.3 Scikit-learn (sklearn) เป็นไลบรารีหลักสำหรับสร้างและฝึกแบบจำลองการเรียนรู้ของเครื่อง โดยเฉพาะแบบจำลองการถดถอยเชิงเส้น และการถดถอยแบบสุ่มป่า ซึ่งเป็นแบบจำลองที่ใช้ในวิจัยนี้ รวมถึงการแบ่งชุดข้อมูลสำหรับการฝึกและการทดสอบ (Train-Test Split) และการคำนวณเมตริกสำหรับการประเมินสมรรถนะของแบบจำลอง [1,3,5]

2.4 Matplotlib และ Seaborn ใช้สำหรับการสร้างภาพข้อมูล (Data Visualization) เพื่อทำความเข้าใจความสัมพันธ์ของตัวแปรและนำเสนอผลลัพธ์ [1,5]

3. เครื่องมือในการวิเคราะห์ข้อมูล (Optional) ใช้ Google Colab เพื่อเขียนโค้ด วิเคราะห์ และนำเสนอผลลัพธ์ [1,5]

4. โปรแกรมสเปรดชีต ใช้ Microsoft Excel สำหรับการจัดเก็บและตรวจสอบข้อมูลเบื้องต้น [2,4]

2.2.2 การหาคุณภาพเครื่องมือ

การหาคุณภาพของเครื่องมือ แบ่งเป็น 2 ส่วน ได้แก่

ส่วนที่ 1 การตรวจสอบคุณภาพข้อมูล ประกอบด้วย

1. ความสมบูรณ์ของข้อมูล คือตรวจสอบว่ามีข้อมูลขาดหายไปหรือไม่ (Missing Values) เช่น ปีการผลิต, ปริมาณน้ำฝน, อุณหภูมิ, คุณสมบัติน้ำ, ผลผลิตข้าว หากมี อาจต้องมีการเติมเต็มข้อมูล (Imputation) หรือลบข้อมูลที่ไม่มีสมบูรณ์ออก [1,3,5]

2. ความถูกต้องของข้อมูล คือตรวจสอบความผิดปกติของข้อมูล (Outliers) หรือค่าที่ไม่สมเหตุสมผล เช่น อุณหภูมิติดลบมากผิดปกติ หรือผลผลิตข้าวที่สูงหรือต่ำเกินจริง [5,6]

3. ความสอดคล้องของข้อมูล คือตรวจสอบว่าข้อมูลในแต่ละคอลัมน์มีรูปแบบที่สอดคล้องกันหรือไม่ [3,5]

4. ความซ้ำซ้อนของข้อมูล คือตรวจสอบว่ามีข้อมูลที่ซ้ำกันมากเกินไปหรือไม่

ถ้ามีมากเกินไปหรือเกินกำหนดควรกำจัดตามความเหมาะสม เพื่อป้องกันความ

คลาดเคลื่อนของข้อมูล [1,3,5]

5. แหล่งที่มาของข้อมูล คือตรวจสอบความน่าเชื่อถือของแหล่งข้อมูล

รวบรวมมา เพื่อให้มั่นใจว่าข้อมูลเป็นทางการและเชื่อถือได้ [2,4]

ส่วนที่ 2 การประเมินสมรรถนะของแบบจำลอง

การประเมินสมรรถนะของแบบจำลอง การถดถอยเชิงเส้น และ การถดถอยแบบสุ่มป่ามีเกณฑ์การ

ประเมิน ดังนี้

1. ค่าสัมประสิทธิ์การตัดสินใจ (R-squared หรือ R^2) ใช้ในการวัดความสามารถของแบบจำลองในการอธิบายความแปรปรวนของผลผลิตข้าว ค่า R^2 ที่สูงขึ้น (ใกล้ 1 หรือ 100%) บ่งชี้ว่าแบบจำลองสามารถอธิบายการเปลี่ยนแปลงของผลผลิตได้ดีขึ้น [3,5,8]

2. ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) ใช้ในการวัดขนาดความคลาดเคลื่อนเฉลี่ยของการพยากรณ์ ค่า MAE ที่ต่ำลงบ่งชี้ว่าแบบจำลองมีความแม่นยำในการพยากรณ์มากขึ้น [3,5,8]

2.3 การเก็บรวบรวมข้อมูล

2.3.1 แหล่งข้อมูลทุติยภูมิ

เก็บรวบรวมข้อมูลที่เกี่ยวข้องกับความรู้ทั่วไปซึ่งเป็นข้อมูลที่ถูกรวบรวมและเผยแพร่โดยองค์กรหรือหน่วยงานอื่น เพื่อนำมาใช้เป็นฐานข้อมูลสำหรับสร้างและทดสอบแบบจำลองการเรียนรู้ของเครื่องจักร ข้อมูลที่เกี่ยวข้องกับผลผลิตข้าวและปัจจัยสิ่งแวดล้อมได้แก่ ข้อมูลผลผลิตข้าวของปีพุทธศักราช 2565 และ 2566 ซึ่งได้รับการรวบรวมจาก สำนักงานเศรษฐกิจการเกษตร (OAE) [2] นอกจากนี้ ข้อมูลอุณหภูมิสูงสุดและต่ำสุดรวมถึงปริมาณน้ำฝนสูงสุดและต่ำสุด ได้มาจากสถาบันสารสนเทศทรัพยากรน้ำ (สสน.) [2] ขณะที่ข้อมูลคุณสมบัติของดิน อาทิ ค่า pH และปริมาณธาตุอาหารหลัก ได้รวบรวมจาก กรมพัฒนาที่ดิน [4] พร้อมทั้งศึกษาพิจารณาควบคู่ไปพร้อมกับงานวิจัยที่เกี่ยวข้อง เช่น การพยากรณ์ผลผลิตข้าวด้วย Machine Learning [6,7] เอกสารทางวิชาการ, บทความวิจัย และบทความออนไลน์ที่น่าเชื่อถือ เพื่อประกอบการวิเคราะห์ผลการวิจัยในการเก็บรวบรวมข้อมูล

2.4 การวิเคราะห์ข้อมูล

2.4.1 การเตรียมข้อมูล (Data Preprocessing)

การเตรียมข้อมูลเพื่อจัดระเบียบและปรับปรุงคุณภาพของชุดข้อมูลที่ใช้ในการวิเคราะห์ เริ่มจากการรวมข้อมูล (Data Integration) ซึ่งเป็นการผนวกข้อมูลผลผลิตข้าว ข้อมูลสภาพแวดล้อม (อุณหภูมิและปริมาณน้ำฝน) และข้อมูลคุณสมบัติของดิน เข้าด้วยกันตามช่วงเวลา (ปีพุทธศักราช 2565 และ 2566) และตำแหน่งพื้นที่ที่สัมพันธ์กัน เพื่อให้ได้ชุดข้อมูลที่ครบถ้วนสำหรับการวิเคราะห์ [1,3]

ถัดมาคือ การทำความสะอาดข้อมูล (Data Cleaning) ซึ่งเป็นการตรวจสอบและจัดการกับความผิดปกติของข้อมูล อาทิ ข้อมูลที่ขาดหายไป (Missing Values) หรือค่าผิดปกติ (Outliers) เพื่อรับรอง ความถูกต้องและน่าเชื่อถือของชุดข้อมูล [1,3]

สุดท้ายคือ การแปลงข้อมูล (Data Transformation/Scaling) ซึ่งอาจรวมถึงการปรับมาตราส่วนของข้อมูล เพื่อให้ตัวแปรแต่ละตัวมีอิทธิพลต่อแบบจำลองอย่างเหมาะสม [1,3]

2.4.2 การแบ่งชุดข้อมูล (Data Splitting)

ชุดข้อมูลที่เตรียมไว้จะถูกแบ่งออกเป็น 2 ส่วนหลัก ได้แก่ ชุดข้อมูลสำหรับฝึก (Training Set) และชุดข้อมูลสำหรับทดสอบ (Testing Set) โดยในวิจัยนี้แบ่งชุดข้อมูลออกเป็น 80:20 เพื่อให้แบบจำลองได้เรียนรู้จากชุดข้อมูลฝึก และประเมินความสามารถในการพยากรณ์ข้อมูลที่ไม่เคยเห็นจากชุดข้อมูลทดสอบ [1,3,5]

2.4.3 การสร้างและฝึกแบบจำลอง (Model Development and Training)

นำชุดข้อมูลสำหรับฝึกไปใช้ในการสร้างและฝึกแบบจำลองการเรียนรู้ของเครื่องจักรในวิจัยนี้เลือกใช้แบบจำลอง 2 รูปแบบ ได้แก่ การถดถอยเชิงเส้น และการถดถอยแบบสุ่มป่า ในระหว่างกระบวนการฝึก อาจมีการปรับแต่งพารามิเตอร์ (Hyperparameter Tuning) ของแต่ละแบบจำลอง เพื่อเพิ่มประสิทธิภาพในการเรียนรู้และลดความคลาดเคลื่อนในการพยากรณ์ [1,3,5,8]

2.4.4 การประเมินสมรรถนะของแบบจำลอง (Model Performance Evaluation)

หลังจากแบบจำลองได้รับการฝึกฝนแล้ว จะถูกนำไปพยากรณ์ผลผลิตข้าวจากชุดข้อมูลสำหรับทดสอบ จากนั้นจึงประเมินสมรรถนะของแต่ละแบบจำลองโดยใช้ ค่าสัมประสิทธิ์การตัดสินใจ (R-squared หรือ R^2) ซึ่งใช้วัดความสามารถของแบบจำลองในการอธิบายความแปรปรวนของผลผลิตข้าว โดยค่า R^2 ที่สูงขึ้นบ่งชี้ว่าแบบจำลองอธิบายข้อมูลได้ดี [1,3,5,] และค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) ซึ่งใช้วัดขนาดความคลาดเคลื่อนเฉลี่ยของการพยากรณ์ โดยค่า MAE ที่ต่ำลงแสดงถึงความแม่นยำที่สูงขึ้นของแบบจำลอง [8]

2.4.5 การเปรียบเทียบและอภิปรายผล

นำผลการประเมินสมรรถนะของแบบจำลองทั้งสองมาเปรียบเทียบกัน เพื่อระบุว่าแบบจำลองใดมีความสามารถในการพยากรณ์ผลผลิตข้าวได้ดีกว่ากัน [1,3,5] พร้อมทั้งอภิปรายผลการวิเคราะห์ โดยเชื่อมโยงผลลัพธ์ที่ได้กับวัตถุประสงค์การวิจัย ทฤษฎีที่เกี่ยวข้อง และงานวิจัยอื่น ๆ [9]

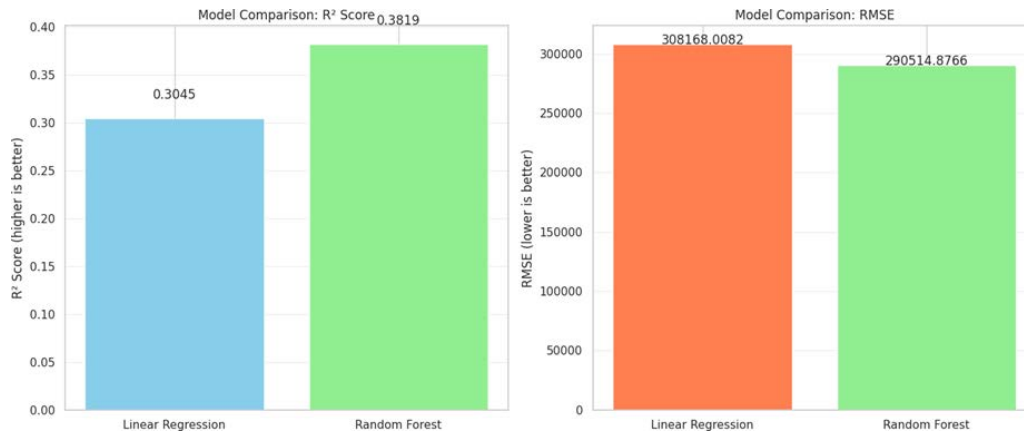
นอกจากนี้ยังวิเคราะห์ปัจจัยที่มีอิทธิพลต่อประสิทธิภาพของแบบจำลอง เช่น คุณสมบัติของชุดข้อมูล ปริมาณ และคุณภาพของข้อมูล การเลือกตัวแปรสำคัญ (Feature Selection) และการปรับแต่งพารามิเตอร์ของแบบจำลอง [3,5,8] รวมถึงเสนอแนวทางในการพัฒนาต่อยอด เช่น การใช้โมเดล Deep Learning, การรวมข้อมูลจากแหล่งใหม่ หรือการปรับปรุงเทคนิคการประเมินสมรรถนะ [10,11]

3. ผลการวิจัยและอภิปรายผล

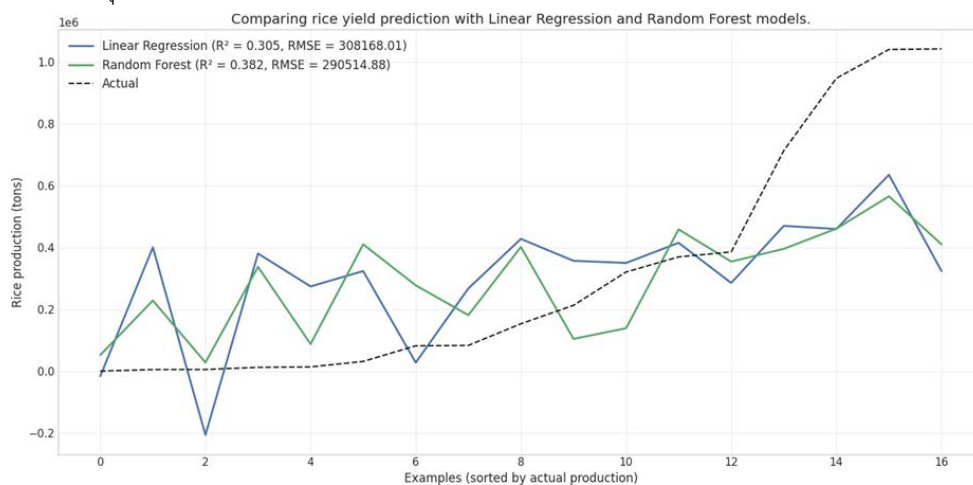
การศึกษานี้ได้ทำการประเมินสมรรถนะของแบบจำลองการเรียนรู้ของเครื่องจักรสองรูปแบบ คือ การถดถอยเชิงเส้น และการถดถอยแบบสุ่มป่า ในการพยากรณ์ผลผลิตข้าว ผลการทดสอบแสดงในตารางที่ 1

ตารางที่ 1 ผลการประเมินสมรรถนะของแบบจำลองระหว่าง การถดถอยเชิงเส้น และ การถดถอยแบบสุ่มป่า

แบบจำลอง	RMSE	R^2
การถดถอยเชิงเส้น	308,168.0082	0.3045
การถดถอยแบบสุ่มป่า (แบบจำลองการถดถอยแบบสุ่มป่า)	290,514.8766	0.3819



ภาพที่ 1 กราฟแท่ง (Bar Graph) ความแตกต่างระหว่าง R2 score และ RMSE ของแบบจำลองระหว่าง การถดถอยเชิงเส้น และการถดถอยแบบสุ่มป่า



ภาพที่ 2 กราฟเส้น (Line Graph) สรุปการเปรียบเทียบความแตกต่าง ระหว่าง Random Forest Regressor และ Linear Regressor

จากผลการประเมินด้วยค่า R^2 (Coefficient of Determination) ซึ่งเป็นตัวบ่งชี้ความสามารถของแบบจำลองในการอธิบายความแปรปรวนของผลลัพธ์ พบว่า แบบจำลอง การถดถอยเชิงเส้น มีค่า R^2 เท่ากับ 0.3045 ซึ่งหมายความว่าแบบจำลองสามารถอธิบายความแปรปรวนของผลผลิตข้าวได้ประมาณ 30.45% เท่านั้น แสดงให้เห็นว่าแบบจำลองเชิงเส้นตรงอาจไม่สามารถจับความสัมพันธ์ที่ซับซ้อนในข้อมูลได้เพียงพอ

แบบจำลอง การถดถอยแบบสุ่มป่า มีค่า R^2 เท่ากับ 0.3819 ซึ่งสูงกว่าแบบจำลอง การถดถอยเชิงเส้น อย่างชัดเจน (0.3819 กับ 0.3045) บ่งชี้ว่า แบบจำลองการถดถอยแบบสุ่มป่า มีความสามารถในการพยากรณ์ได้ดีกว่า เนื่องจากสามารถจัดการกับความสัมพันธ์แบบไม่เชิงเส้นได้ดี อย่างไรก็ตาม ค่า R^2 ที่ 38.19% ยังอยู่ในระดับปานกลาง ซึ่งชี้ให้เห็นว่ายังมีความแปรปรวนของผลผลิตข้าวที่ยังไม่สามารถอธิบายได้ด้วยชุดตัวแปรและแบบจำลองปัจจุบัน นอกจากนี้ ค่า RMSE ที่ 290,514.8766 ยังบ่งชี้ถึงขนาดความคลาดเคลื่อนเฉลี่ยในการพยากรณ์ที่ค่อนข้างสูง

4. สรุปผล

จากการวิเคราะห์และประเมินผลแบบจำลองพยากรณ์ผลผลิตข้าวในประเทศไทยโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร สามารถสรุปผลการวิจัยได้ว่า แบบจำลอง การถดถอยแบบสุ่มป่า มีประสิทธิภาพในการพยากรณ์สูงกว่าแบบ

จำลอง การถดถอยเชิงเส้น โดยพิจารณาจากค่า R^2 ที่แบบจำลองการถดถอยแบบสุ่มป่า สามารถอธิบายความแปรปรวนของผลผลิตข้าวได้ 38.19% เมื่อเทียบกับ การถดถอยเชิงเส้น ที่ 30.45%

อย่างไรก็ตาม ค่าความแม่นยำของแบบจำลองทั้งสองยังคงอยู่ในระดับปานกลาง ซึ่งบ่งชี้ว่าแบบจำลองปัจจุบันยังมีข้อจำกัดในการพยากรณ์ และอาจสะท้อนถึงความซับซ้อนโดยธรรมชาติของปัจจัยที่ส่งผลต่อผลผลิตข้าว หรืออาจยังมีตัวแปรสำคัญอื่น ๆ ที่ไม่ได้ถูกนำมาพิจารณาในการวิเคราะห์ รวมถึงข้อจำกัดด้านคุณภาพและปริมาณของข้อมูลที่ใช้ในการฝึกแบบจำลองในช่วงปีพุทธศักราช 2565-2566

เพื่อเพิ่มประสิทธิภาพและความแม่นยำของแบบจำลองในการพยากรณ์ผลผลิตข้าวในอนาคต จึงควรดำเนินการปรับแต่งพารามิเตอร์ (Hyperparameter Tuning) ของแบบจำลอง และ ประยุกต์ใช้วิธีการคัดเลือกคุณลักษณะ (Feature Selection) เพื่อคัดเลือกตัวแปรที่มีอิทธิพลต่อการพยากรณ์มากที่สุด นอกจากนี้ การสำรวจและเพิ่มปริมาณ รวมถึงคุณภาพของข้อมูล ที่ใช้ในการสร้างแบบจำลอง อาทิ การขยายขอบเขตข้อมูลให้ครอบคลุมช่วงเวลาที่ยาวนานขึ้น และการพิจารณาตัวแปรที่หลากหลายมากขึ้น จะเป็นแนวทางสำคัญในการยกระดับความน่าเชื่อถือและความแม่นยำของแบบจำลองพยากรณ์ผลผลิตข้าวต่อไป ซึ่งจะนำไปสู่การสนับสนุนการวางแผนและกำหนดนโยบายภาคการเกษตรของประเทศไทยให้มีประสิทธิภาพยิ่งขึ้น

5. เอกสารอ้างอิง (References)

- [1] Cheikh Abdelkader Ahmed. & Mohamedou El Ghotob Cheikh Tourad. (2024). REVOLUTIONIZING RICE YIELD PREDICTION: A DATA- DRIVEN APPROACH IN MAURITANIA. University of Nouakchott Al Aasriya. https://www.researchgate.net/publication/379830551_REVOLUTIONIZING_RICE_YIELD_PREDICTION_A_DATA-_DRIVEN_APPROACH_IN_MAUROITANIA.
- [2] Hydro - Informatics Institute. (2022, May 9). Monthly provincial rainfall distribution. <https://data.hii.or.th/sv/dataset/spatial-rain/resource/120aba77-b0a7-44c0-8043-91124dac45ed>.
- [3] Joseph Ingio. & Aamo Iorliam. (2024). Comparing Machine Learning Algorithms for Rice Yield Prediction in Adamawa and Cross Rivers States of Nigeria. American University of Nigeria and Benue State University. https://www.researchgate.net/publication/384430919_Comparing_Machine_Learning_Algorithms_for_Rice_Yield_Prediction_in_Adamawa_and_Cross_Rivers_States_of_Nigeria.
- [4] Land Development Department. (2023, October 31). Integrated Data Management System for Soil Health Card Farmers. <https://lddcatalog.ddd.go.th/it/dataset/osd02/resource/9f552569-f025-4e97-abe1-ec44fa3ae31f>.
- [5] Lingqing Zhan 8 andYongkun He 1,2,.(2024). Rice Yield Estimation Using Machine Learning and Feature Selection in Hilly and Mountainous Chongqing, China,Li Fan 1,2,Shibo Fang 3ORCID,Jinlong Fan 4,5,6,Yan Wang 7. <https://www.mdpi.com/2077-0472/14/9/1615>.
- [6] Md Didarul Islam, Liping Di, Faisal Mueen Qamer, Sravan Shrestha, Liying Guo, Li Lin, Timothy J. Mayer, andAparna R. Phalke. (2023). Rapid Rice Yield Estimation Using Integrated Remote Sensing and Meteorological Data and Machine Learning. <https://www.mdpi.com/2072-4292/15/9/2374>.
- [7] Sonam Sah, Dipanwita Haldar, RN Singh, B. Das & Ajeet Singh Nain. (2024). Rice yield prediction through integration of biophysical parameters with SAR and optical remote sensing data using machine learning models. <https://www.nature.com/articles/s41598-024-72624-4>.
- [8] Ratchaphum Jaikla; Sansanee Auephanwiriyaikul; Attachai Jintrawet. (2021). Rice yield prediction using a



- Support Vector Regression method. <https://ieeexplore.ieee.org/document/4600365>.
- [9] Joginder Kumar a, Monika Devi a, Deepika Verma b, D.P. Malik c, Ajay Sharma a. (2021). Pre-harvest forecast of rice yield based on meteorological parameters using discriminant function analysis. <https://www.sciencedirect.com/science/article/pii/S266615432100096X>.
- [10] Zheng Chu, Jiong Y. (2020). An end-to-end model for rice yield prediction using deep learning fusion. <https://www.sciencedirect.com/science/article/abs/pii/S0168169920300314>.
- [11] Xiao Han, Fangbiao Liu, Xiaoliang He, Fenglou Ling. (2022). Research on Rice Yield Prediction Model Based on Deep Learning. <https://onlinelibrary.wiley.com/doi/10.1155/2022/1922561>.