การพัฒนาโมเดลสำหรับการทำนายผลผลิตพืชผักสลัด กรณีศึกษา: โครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่

ธีรภพ แสงศรี[1] สุรีนาฏ มะโนลา[2*]

Received : September 19, 2024
Revised : April 4, 2025
Accepted : April 30, 2025

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดลทำนายผลผลิตของพืชผักอายุสั้นตระกูลสลัดในโครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่ เพื่อลดปัญหาความไม่สอดคล้องระหว่างปริมาณผลผลิตและความต้องการของตลาด โดยใช้กระบวนการ CRISP-DM และเทคนิคการถดถอยเชิงเส้นพหุคูณในการวิเคราะห์ข้อมูล โดยใช้ข้อมูลผลผลิตและปัจจัยแวดล้อมในช่วงปี พ.ศ. 2561 - พ.ศ. 2563 ผลการศึกษาพบว่า โมเดลที่พัฒนาขึ้นสามารถอธิบายความแปรปรวนของผลผลิตได้ 65.04% (R-squared = 0.6504) โดยมีค่าความคลาดเคลื่อนเฉลี่ยของการทำนาย (RMSE) เท่ากับ 68.73 กิโลกรัม และค่าความผิดพลาดสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 55.33 กิโลกรัม ปัจจัยที่มีนัยสำคัญทางสถิติต่อผลผลิต ได้แก่ อุณหภูมิสูงสุด อุณหภูมิต่ำสุด และจำนวนเกษตรกร ขณะที่ปัจจัยอื่น ๆ เช่น ความชื้นและความเร็วลม ไม่แสดงผลกระทบอย่างมีนัยสำคัญ โมเดลที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้เพื่อช่วยเกษตรกรและเจ้าหน้าที่ที่เกี่ยวข้องในการวางแผนการผลิตให้สอดคล้องกับความต้องการของตลาด นอกจากนี้ ควรศึกษาปัจจัยแวดล้อมเพิ่มเติมที่อาจมีผลต่อผลผลิต และเปรียบเทียบประสิทธิภาพของโมเดลพยากรณ์หลายรูปแบบ เพื่อเพิ่มความแม่นยำในการคาดการณ์ อีกทั้งควรขยายการศึกษาสู่พืชหรือพื้นที่เพาะปลูกอื่น ๆ เพื่อให้โมเดลมีความยืดหยุ่นและสามารถนำไปใช้งานได้อย่างแพร่หลาย

คำสำคัญ: กระบวนการทำเหมืองข้อมูล การถดถอยเชิงเส้นพหุคูณ การพยากรณ์ผลผลิต โครงการพัฒนาพื้นที่สูงแบบโครงการหลวง

[1] อาจารย์ประจำหลักสูตรระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา อีเมล: tees@rmutl.ac.th

[2] อาจารย์ประจำหลักสูตรระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา อีเมล: sureenat@rmutl.ac.th

[*] ผู้นิพนธ์หลัก อีเมล: sureenat@rmutl.ac.th

DEVELOPMENT OF A PREDICTIVE MODEL FOR SHORT-CYCLE SALAD VEGETABLE YIELD: CASE STUDY OF HIGHLAND RESEARCH AND DEVELOPMENT INSTITUTE (HRDI) IN PA PAE, MAE TAENG, CHIANG MAI

Theeraphop Saengsri[1] Sureenat Manola[2*]

Abstract

The current research was aimed to develop a predictive model for short-cycle salad crops in the Highland Research and Development Institute (HRDI), Pa Pae, Mae Taeng District, Chiang Mai. In addition, it was aimed to decrease the inconsistent problem between production volume and market demand by utilizing the CRISP-DM process and multiple linear regression techniques for data analysis. Based on production and environmental data between 2018 to 2020, the research results were indicated that the developed model displayed 65.04% of the variation in yield (R-squared = 0.6504), with a Root Mean Squared Error (RMSE) of 68.73 kilograms and a Mean Absolute Error (MAE) of 55.33 kilograms. Statistically significant factors affecting yield included maximum temperature, minimum temperature, and the number of farmers. While humidity and wind speed showed no significant impact. The model could assist farmers and relevant officials in planning production to better align with market demand. Further studies should explore additional environmental factors affecting crop yield and compare predictive models to improve forecasting accuracy. Expanding the study to other crops or cultivating areas would enhance the model's adaptability and broader agricultural applications.

Keywords: Data Mining Process, Multiple Linear Regression, Yield Prediction, Highland Research and Development Institute (HRDI)

[1] Lecturer, Department of Business Information System, Faculty of Faculty of Business Administration and Liberal Arts, Rajamangala University of Technology, e-mail: tees@rmutl.ac.th

[2] Lecturer, Department of Business Information System, Faculty of Faculty of Business Administration and Liberal Arts, Rajamangala University of Technology, e-mail: sureenat@rmutl.ac.th

[*] Corresponding author, e-mail: sureenat@rmutl.ac.th

บทนำ

ตำบลป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่ เป็นพื้นที่เกษตรกรรมที่เกษตรกรส่วนใหญ่ปลูกชาเมี่ยงและมะแขว่นเป็นพืชเศรษฐกิจหลัก อย่างไรก็ตาม ความต้องการของตลาดที่เปลี่ยนแปลงไปส่งผลกระทบต่อรายได้ของเกษตรกร ส่งผลให้มีการก่อตั้ง โครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ เพื่อส่งเสริมอาชีพและพัฒนาการเกษตรในพื้นที่ โดยเน้นการผลิตพืชที่เหมาะสมกับสภาพภูมิอากาศ เช่น ไม้ผลและพืชผักอายุสั้น (สถาบันวิจัยและพัฒนาพื้นที่สูง (สวพส.), 2563) ปัจจุบัน พืชผักสลัด เช่น คะน้าฮ่องกงและผักกาดหอม ได้รับการส่งเสริมให้ปลูกเป็นรายได้เสริม แต่ความแปรปรวนของปัจจัยแวดล้อมส่งผลให้ผลผลิตไม่สอดคล้องกับความต้องการของตลาด ซึ่งสอดคล้องกับผลการศึกษาที่ระบุว่าปัจจัยทางสิ่งแวดล้อม เช่น อุณหภูมิ ความชื้น และปริมาณฝน มีผลกระทบต่อผลผลิตพืชอย่างมีนัยสำคัญ (Joshua et al., 2022)

การพยากรณ์ผลผลิตทางการเกษตร เป็นแนวทางที่ช่วยเพิ่มประสิทธิภาพการวางแผนการผลิตและการตลาด โดยงานวิจัยหลายฉบับได้ศึกษาแบบจำลองเชิงสถิติและกระบวนการทำเหมืองข้อมูลเพื่อพยากรณ์ปริมาณผลผลิต เช่น การศึกษาในงานวิจัยซึ่งใช้แบบจำลองการถดถอยเชิงเส้นพหุคูณเป็นแบบจำลองอ้างอิงในการพยากรณ์ผลผลิตข้าว ในการพยากรณ์พบว่าแบบจำลองการถดถอยเชิงเส้นพหุคูณมีความแม่นยำ 82% และสามารถใช้เป็นพื้นฐานในการพัฒนาโมเดลที่แม่นยำยิ่งขึ้น (Joshua et al., 2022) ในขณะที่การศึกษางานวิจัยที่มีการเปรียบเทียบแบบจำลองการถดถอยเชิงเส้นพหุคูณกับแบบจำลองการเรียนรู้ของเครื่อง เช่น Random Forest และ Neural Networks และพบว่าแบบจำลองการเรียนรู้ของเครื่องสามารถเพิ่มความแม่นยำของการพยากรณ์ได้มากกว่าวิธีการทางสถิติแบบดั้งเดิม (Khade et al., 2023)

การศึกษาเกี่ยวกับการพยากรณ์ผลผลิตยังพบว่า ปัจจัยด้านสภาพแวดล้อมและปัจจัยทางเศรษฐกิจเป็นองค์ประกอบสำคัญที่ส่งผลต่อผลผลิตพืช โดยมีงานวิจัยที่แสดงให้เห็นว่าแบบจำลองการถดถอยเชิงเส้นพหุคูณเป็นเครื่องมือที่สามารถใช้ในการพยากรณ์ผลผลิตได้อย่างมีประสิทธิภาพ ในขณะแบบจำลองการเรียนรู้ของเครื่องเหมาะสมสำหรับชุดข้อมูลขนาดใหญ่และซับซ้อน (Joshua et al., 2022)

จากปัญหาดังกล่าว งานวิจัยนี้มุ่งเน้นการพัฒนาแบบจำลองพยากรณ์ผลผลิตพืชผักสลัด โดยใช้กระบวนการทำเหมืองข้อมูลร่วมกับแบบจำลองการถดถอยเชิงเส้นพหุคูณ เพื่อช่วยให้การวางแผนการผลิตและการตลาดมีความสอดคล้องกับความต้องการของตลาดมากขึ้น ตลอดจนเป็นแนวทางในการสนับสนุนเกษตรกรในโครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่ นอกจากนี้ งานวิจัยยังสามารถพัฒนาต่อยอดได้โดยการเพิ่มตัวแปรที่เกี่ยวข้อง หรือปรับใช้เทคนิคการพยากรณ์ที่เหมาะสมยิ่งขึ้น เพื่อปรับปรุงความแม่นยำของแบบจำลองให้สามารถนำไปใช้ในทางปฏิบัติได้อย่างมีประสิทธิภาพ

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาโมเดลการทำนายผลผลิตของพืชผักอายุสั้นตระกูลสลัด ในโครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่

2. เพื่อประเมินความถูกต้องและประสิทธิภาพของโมเดล ในการทำนายผลผลิต

วิธีดำเนินการวิจัย

ภาพที่ 1 แสดงกรอบแนวคิดของงานวิจัย

จากภาพที่ 1 เริ่มจากปัญหาความไม่สอดคล้องระหว่างผลผลิตและความต้องการของตลาด ซึ่งนำไปสู่วัตถุประสงค์ในการพัฒนาและประเมินประสิทธิภาพโมเดลทำนายผลผลิต เพื่อแก้ไขปัญหาดังกล่าว วิธีการดำเนินงานวิจัยประกอบด้วยการเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล การพัฒนาโมเดลในการทำนายผลผลิต และการประเมินประสิทธิภาพโมเดล ซึ่งคาดว่าจะนำไปสู่ผลลัพธ์ที่สำคัญ ได้แก่ โมเดลการทำนายผลผลิต ค่าความถูกต้องของโมเดล

ผู้วิจัยได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM ซึ่งเป็นแนวทางที่เป็นระบบในการทำเหมืองข้อมูลและวิเคราะห์ข้อมูลที่มีความซับซ้อน (วีรศักดิ์ ฟองเงิน และคณะ, 2560) โดยใช้การพยากรณ์แบบการจำลองการถดถอยพหุคูณ เพื่อหาความสัมพันธ์ระหว่างปัจจัยต่าง ๆ ที่มีผลต่อผลลัพธ์ ซึ่งเป็นเทคนิคที่ได้รับการยอมรับอย่างกว้างขวางในการศึกษาปัจจัยที่มีอิทธิพลต่อตัวแปรตาม (พัชรา ศรีพระบุ, 2565) โดยมีขั้นตอนต่าง ๆ ดังนี้

1. ทำความเข้าใจในธุรกิจโดยการปรึกษาหารือ (Business Understanding) จากการวิเคราะห์ความต้องการของผู้มีส่วนได้เสีย พบว่าปัญหาหลักในการบริหารจัดการผลผลิตพืชผักสลัดในโครงการหลวงป่าแป๋ คือ ความคลาดเคลื่อนสูงระหว่างปริมาณผลผลิตที่คาดการณ์และปริมาณผลผลิตจริง ซึ่งเกิดจากการพึ่งพาองค์ความรู้และประสบการณ์ของเจ้าหน้าที่และเกษตรกรในการวางแผนการผลิต เพื่อแก้ไขปัญหานี้ ผู้วิจัยจึงนำเสนอการพัฒนาแบบจำลองการพยากรณ์ผลผลิตโดยใช้เทคนิคการถดถอยเชิงเส้นพหุคูณ เพื่อเพิ่มประสิทธิภาพในการคาดการณ์ผลผลิต

2. ทำความเข้าใจข้อมูล (Data Understanding) ได้ดำเนินการรวบรวมข้อมูลที่เกี่ยวข้องกับการผลิตพืชผักสลัดจากสองแหล่งหลัก ได้แก่ ข้อมูลแผนการปลูกพืชผักสลัดรายสัปดาห์ย้อนหลัง 3 ปี (พ.ศ. 2561 - 2563) จากโครงการหลวงป่าแป๋ และข้อมูลสภาพอากาศย้อนหลังที่ได้จากเว็บไซต์ของสถาบันวิจัยและพัฒนาพื้นที่สูง (สวพส.)

จากการวิเคราะห์ข้อมูลเบื้องต้นพบว่า ข้อมูลที่ได้รับจากภายในองค์กรในแต่ละตารางเป็นข้อมูลแผนการผลิตรายปี โดยในหนึ่งตารางมีทั้งหมด 8 คอลัมน์ ประกอบด้วย (1) ลำดับที่ชุดการปลูกผักคะน้าฮ่องกง mk แผน (400 กก.), (2) ชื่อเกษตรกร, (3) จำนวนต้นกล้า/สัปดาห์, (4) จำนวนต้นพื้นที่ปลูก (ตร.ม.), (5) วันที่เพาะกล้า, (6) วันที่ย้ายปลูก, (7) วันที่เก็บเกี่ยว, และ (8) น้ำหนักหลังแต่ง (กก.)/สัปดาห์

ในส่วนของข้อมูลสภาพอากาศย้อนหลังที่ได้รับจากภายนอกองค์กรนั้น จัดเก็บแบบรายเดือนย้อนหลัง 3 ปี แต่ละคอลัมน์จะระบุวันที่ (Date), อุณหภูมิเฉลี่ย (°C), อุณหภูมิสูงสุด (°C), อุณหภูมิต่ำสุด (°C), ความชื้นเฉลี่ย (%), ความชื้นสูงสุด (%), ความชื้นต่ำสุด (%), ปริมาณฝน (มม.) และความเร็วลม (กม./ชม.)

3. เตรียมข้อมูล (Data Preparation) กระบวนการเตรียมข้อมูลประกอบด้วยขั้นตอนดังต่อไปนี้

3.1 การทำความสะอาดข้อมูล (Data Cleaning) ผู้วิจัยได้จัดการข้อมูลสูญหาย (Missing Values) ในข้อมูลสภาพอากาศย้อนหลังรายวันด้วยวิธีการ “เติมข้อมูลสูญหาย” (Imputation) แบบอนุกรมเวลา เพื่อรักษาความต่อเนื่องของข้อมูล โดยกำหนดทิศทางการเติม (Limit Direction) ให้เหมาะสมกับตำแหน่งที่เกิดข้อมูลสูญหาย ได้แก่ Forward, Backward และ Both อันขึ้นอยู่กับช่วงที่เกิดการขาดหายของข้อมูล (ต้น กลาง หรือปลายชุดข้อมูล) นอกจากนี้ได้ทำการตรวจสอบและจัดการข้อมูลผิดปกติ (Outliers) ด้วยการวิเคราะห์การกระจายตัวผ่าน Boxplot หาก Outlier เป็นค่าที่เกิดขึ้นตามธรรมชาติ เช่น อุณหภูมิที่ต่ำมากในฤดูหนาว จึงคงไว้เพื่อสะท้อนความเป็นจริงของข้อมูล แต่หากพบว่าเป็นค่าที่เกิดจากความผิดพลาดในการบันทึกอาจพิจารณาตัดทิ้งหรือปรับปรุงให้เหมาะสมตามบริบทของข้อมูลต่อไป

3.2 การแปลงข้อมูล (Data Transformation) ข้อมูลสภาพอากาศมีการจัดเก็บในรูปแบบรายวัน ในขณะที่ข้อมูลผลผลิตคะน้าฮ่องกงเป็น “รายชุดการเพาะปลูก” (Crop) ที่ใช้เวลาประมาณ 43-45 วันต่อรอบการปลูก ผู้วิจัยจึงทำการสรุปข้อมูลรายวันให้อยู่ในระดับ “รายชุดการเพาะปลูก” โดยคำนวณค่าสถิติเบื้องต้น ได้แก่ ค่าเฉลี่ย (Mean) ค่าต่ำสุด (Min) ค่าสูงสุด (Max) และผลรวม (Sum) เพื่อสะท้อนสภาพอากาศตลอดช่วงอายุของพืชในแต่ละรอบการปลูก ทั้งนี้ยังรวมถึงการตัดคอลัมน์ที่ข้อมูลขาดหายมากเกินไป (เช่น ปริมาณน้ำฝนในบางปี) เพื่อให้ข้อมูลโดยรวมมีความสม่ำเสมอและเหมาะสมกับการวิเคราะห์

3.3 การสรุปข้อมูลสำหรับใช้ในแบบจำลอง (Data Aggregation and Merging) ผู้วิจัยได้รวมข้อมูลสภาพอากาศรายชุดการเพาะปลูกเข้ากับข้อมูลปริมาณผลผลิตที่บันทึกในตารางแผนการเพาะปลูกและผลผลิตจริง โดยมีการระบุช่วงวันที่เริ่มต้นการเพาะกล้าและวันที่เก็บเกี่ยว เพื่อให้ชุดข้อมูลสุดท้ายมีตัวแปรทั้งตัวแปรอิสระ (Independent Variables) และตัวแปรตาม (Dependent Variable) โดยในขั้นตอนนี้จะสอดคล้องกับรูปแบบการวิเคราะห์ถดถอยเชิงเส้นพหุคูณ

จากผลการวิเคราะห์แผนภูมิ Boxplot พบว่า ตัวแปร อุณหภูมิเฉลี่ย อุณหภูมิสูงสุด อุณหภูมิต่ำสุด ความชื้นเฉลี่ย ความชื้นสูงสุด ความชื้นต่ำสุด ความเร็วลมต่ำสุด จำนวนเกษตรกร และน้ำหนักหลังตัดแต่งไม่

ปรากฏค่าสิ่งผิดปกติ (Outlier) อย่างไรก็ตาม มีการตรวจพบ Outlier ในตัวแปรอุณหภูมิเฉลี่ยช่วง 15-20 องศาเซลเซียส ซึ่งเมื่อผู้วิจัยย้อนกลับไปพิจารณาข้อมูลต้นทางแล้วพบว่าเป็นสภาพอากาศจริงที่เกิดขึ้นในฤดูหนาวของทุกปี จึงสามารถยอมรับได้ในเชิงบริบทด้านเกษตรกรรม เช่นเดียวกับ Outlier ในตัวแปรอุณหภูมิสูงสุดช่วง 25-30 องศาเซลเซียส ก็เกิดขึ้นในฤดูหนาวเช่นเดียวกัน ในกรณีตัวแปรความชื้นต่ำสุด (54-56%) พบว่าอยู่ในช่วงฤดูฝนถึงฤดูหนาวของปี 2561 และ 2563 จึงสอดคล้องกับสภาพอากาศจริงและยอมรับได้ สุดท้าย ตัวแปรความชื้นสูงสุด (99.0–99.8%) ที่ตรวจพบมีความเป็นไปได้ว่าเป็นผลจากฤดูกาลทางธรรมชาติ ด้วยเหตุนี้ ผู้วิจัยจึงเลือกคงข้อมูล Outlier ไว้ เนื่องจากอยู่ในเกณฑ์ตามฤดูกาลและแบบจำลองการถดถอยเชิงเส้นพหุคูณจะมีการคัดเลือกตัวแปรที่มีความสัมพันธ์กันอีกครั้งในขั้นตอนถัดไป

```
In [52]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 66 entries, 0 to 65
Data columns (total 11 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   temp_avg   66 non-null     float64
 1   temp_max   66 non-null     float64
 2   temp_min   66 non-null     float64
 3   humin_avg  66 non-null     float64
 4   humin_max  66 non-null     float64
 5   humin_min  66 non-null     float64
 6   win_avg    66 non-null     float64
 7   win_max    66 non-null     float64
 8   win_min    66 non-null     int64
 9   farmer     66 non-null     int64
 10  weight     66 non-null     float64
dtypes: float64(9), int64(2)
memory usage: 5.8 KB
```

ภาพที่ 2 แสดงตัวแปรทั้งหมดในชุดข้อมูล

4. การสร้างแบบจำลองในการทำนายผลผลิต (Modeling) ในการสร้างสมการถดถอยเชิงเส้นพหุคูณเพื่อพยากรณ์ปริมาณผลผลิต จำเป็นต้องคัดเลือกตัวแปรอิสระที่เกี่ยวข้องและหลีกเลี่ยงปัญหาการที่ตัวแปรมีความสัมพันธ์กันมากเกินไป (Multicollinearity) ซึ่งอาจทำให้ผลการวิเคราะห์คลาดเคลื่อน และลดความแม่นยำของแบบจำลอง ดังนั้น ผู้วิจัยจึงดำเนินการคัดเลือกตัวแปรตามลำดับขั้นดังนี้

4.1 กำหนดรายการตัวแปรเริ่มต้น ในขั้นต้น ได้ระบุตัวแปรอิสระที่อาจมีความเกี่ยวข้องกับปริมาณผลผลิต จำนวน 10 ตัว ได้แก่ temp_avg (อุณหภูมิเฉลี่ย), temp_max (อุณหภูมิสูงสุด), temp_min (อุณหภูมิต่ำสุด), humin_avg (ความชื้นเฉลี่ย), humin_max (ความชื้นสูงสุด), humin_min (ความชื้นต่ำสุด), win_avg (ความเร็วลมเฉลี่ย), win_max (ความเร็วลมสูงสุด), win_min (ความเร็วลมต่ำสุด) และ farmer (จำนวนเกษตรกร) โดยกำหนดตัวแปรตามเป็น weight (น้ำหนักผลผลิต หน่วยกิโลกรัม)

4.2 ตรวจสอบความสัมพันธ์กับตัวแปรตาม เริ่มจากวิเคราะห์กราฟกระจาย (Scatter Plot) และตารางค่าสหสัมพันธ์ (Correlation Matrix) เพื่อตรวจสอบว่าตัวแปรอิสระแต่ละตัวมีความสัมพันธ์เชิงเส้นกับ weight มากน้อยเพียงใด รวมถึงดูว่าบางตัวแปรมีความสัมพันธ์ซ้อนกันเองในระดับสูง (เช่น เกิน 0.70) หรือไม่ จากการประเมินพบว่าตัวแปร win_min (ความเร็วลมต่ำสุด) แทบไม่แสดงความสัมพันธ์อย่างมีนัยสำคัญ จึงตัดออกตั้งแต่ขั้นนี้

4.3 ประเมิน Multicollinearity ด้วยค่า VIF (Variance Inflation Factor) ตัวแปรที่มีความสัมพันธ์กันเองสูงอาจก่อให้เกิด Multicollinearity ซึ่งเป็นปัจจัยที่ทำให้โมเดลถดถอยมีค่าพารามิเตอร์ที่ไม่น่าเชื่อถือ ผู้วิจัยจึงคำนวณค่า VIF ของตัวแปรอิสระที่เหลือ พบว่าตัวแปร temp_avg (อุณหภูมิเฉลี่ย) มีค่า VIF สูงที่สุดในกลุ่มอุณหภูมิ (ซ้ำซ้อนกับ temp_max และ temp_min) ขณะที่ win_max (ความเร็วลมสูงสุด) ก็มีค่า VIF สูงในกลุ่มลมเฉลี่ย (win_avg) จึงตัด temp_avg และ win_max ออกจากโมเดลเพื่อลดความซ้ำซ้อน

ในการพยากรณ์ผลผลิต ผู้วิจัยใช้เทคนิคการถดถอยเชิงเส้นพหุคูณ เพื่อสร้างแบบจำลองทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างปัจจัยต่าง ๆ กับผลผลิต สมการถดถอยเชิงเส้นพหุคูณที่ใช้มีรูปแบบดังนี้

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7$$

โดยที่

y คือ ปริมาณผลผลิตของพืชผักสลัด (กิโลกรัม) ซึ่งเป็นตัวแปรตามที่ต้องการทำนาย

x_1 คือ อุณหภูมิสูงสุด (°C) หรือ temp_max

x_2 คือ อุณหภูมิต่ำสุด (°C) หรือ temp_min

x_3 คือ ความชื้นเฉลี่ย (%) หรือ humin_avg

x_4 คือ ความชื้นสูงสุด (%) หรือ humin_max

x_5 คือ ความชื้นต่ำสุด (%) หรือ humin_min

x_6 คือ ความเร็วลมเฉลี่ย (กม./ชม.) หรือ win_avg

x_7 คือ จำนวนเกษตรกร (คน) หรือ farmer

b_0 คือ ค่าคงที่ของสมการ หรือ จุดตัดแกน y

b_i คือ สัมประสิทธิ์การถดถอย (Regression Coefficients) ซึ่งแสดงถึงอัตราการเปลี่ยนแปลงของตัวแปรตาม (y) เมื่อตัวแปรอิสระแต่ละตัว (x_i) เปลี่ยนแปลงไป 1 หน่วย ในขณะที่ตัวแปรอิสระอื่นๆ คงที่

ใช้ไลบรารี Pandas เพื่ออ่านและจัดการข้อมูล และ ไลบรารี Scikit-learn เพื่อสร้างและฝึกฝนแบบจำลองการถดถอยเชิงเส้นพหุคูณ

Out[104]:

	crop	years	temp_avg	temp_max	temp_min	humin_avg	humin_max	humin_min	rain	win_avg	win_max	win_min	farmer	weight
0	1	2561	23.862143	32.88	19.88	94.595238	100.0	51.86	122.0	0.003571	0.03	0	5	408.5
1	2	2561	23.752791	32.88	19.88	94.677209	100.0	51.86	127.5	0.003721	0.03	0	6	411.0
2	3	2561	23.717674	34.63	19.88	95.063256	100.0	46.85	114.5	0.003023	0.03	0	5	429.5
3	4	2561	23.554651	34.63	19.88	96.246744	100.0	46.85	127.0	0.003256	0.04	0	5	555.0
4	5	2561	23.480233	34.63	19.88	96.703488	100.0	46.85	140.0	0.003023	0.04	0	5	297.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
61	18	2563	20.521556	31.94	12.50	91.535333	100.0	55.62	0.0	0.000000	0.00	0	5	293.0
62	19	2563	20.038409	31.94	12.50	91.395909	100.0	55.62	0.0	0.000000	0.00	0	6	354.0
63	20	2563	19.508864	31.94	12.06	90.898636	100.0	55.62	0.0	0.000000	0.00	0	5	241.0
64	21	2563	18.623636	31.94	10.94	90.398409	99.7	55.62	0.0	0.000000	0.00	0	6	293.0
65	22	2563	17.853409	30.75	10.94	90.018636	99.7	55.62	0.0	0.000000	0.00	0	4	175.0

66 rows × 14 columns

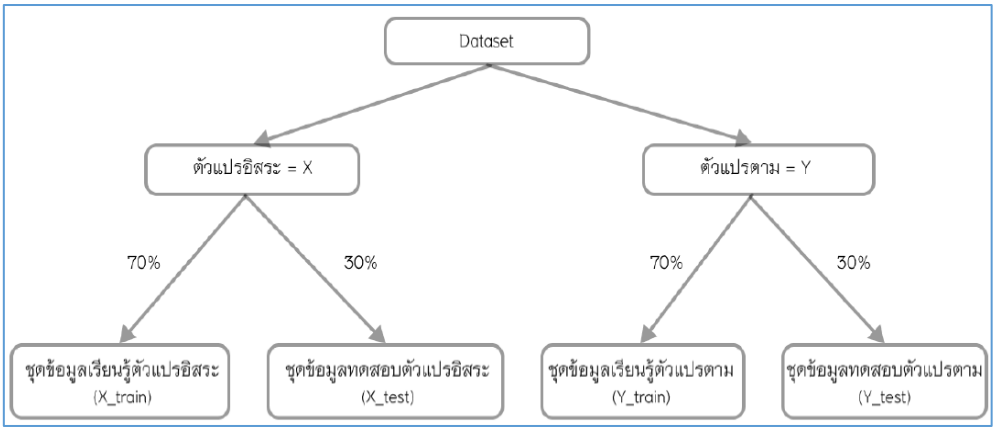
ภาพที่ 3 แสดงผลลัพธ์การนำเข้าข้อมูลโดยการใช้ไลบรารี่ Pandas และ Scikit-learn

การกำหนดตัวแปร ตัวแปรอิสระ (x) ประกอบด้วย อุณหภูมิสูงสุด อุณหภูมิต่ำสุด ความชื้นเฉลี่ย ความชื้นสูงสุด ความชื้นต่ำสุด ความเร็วลมเฉลี่ย และจำนวนเกษตรกร ในขณะที่ตัวแปรตาม (y) คือ ปริมาณผลผลิต (กิโลกรัม)

เพื่อประเมินประสิทธิภาพของแบบจำลองในการทำนายผลผลิต จึงดำเนินการแบ่งชุดข้อมูลออกเป็นสองส่วน ได้แก่ ชุดฝึกอบรม (Training set) และชุดทดสอบ (Testing set) โดยใช้ฟังก์ชัน train_test_split() ในสัดส่วน 70:30 ตามลำดับ การแบ่งข้อมูลนี้ช่วยให้สามารถประเมินว่าโมเดลสามารถทำนายผลผลิตได้ดีเพียงใดกับข้อมูลที่โมเดลไม่เคยพบเห็นมาก่อน และยังช่วยป้องกันปัญหา overfitting ซึ่งโมเดลอาจเรียนรู้ข้อมูลการฝึกอบรมได้ดีเกินไป แต่ไม่สามารถทำนายข้อมูลใหม่ได้แม่นยำ

```
X_train, X_test = train_test_split(X, test_size=0.3, random_state=1)
Y_train, Y_test = train_test_split(Y, test_size=0.3, random_state=1)
```

ภาพที่ 4 แสดงคำสั่งเพื่อวัดประสิทธิภาพในการทำนายของแบบจำลองและข้อมูลโดยใช้ฟังก์ชัน train_test_split()



ภาพที่ 5 แสดงการแบ่งชุดข้อมูล (Split Data)

แบบจำลองการถดถอยเชิงเส้นถูกสร้างขึ้นโดยใช้ฟังก์ชัน LinearRegression() จากไลบรารี scikit-learn และถูกเก็บไว้ในตัวแปร kale_predict_model จากนั้น โมเดลนี้ถูกฝึกฝน (fit) ด้วยชุดข้อมูลฝึกอบรม (X_train และ Y_train) โดยอัตโนมัติจะเลือกสร้างแบบจำลองการถดถอยเชิงเส้นพหุคูณเมื่อตัวแปรอิสระมีหลายตัว หรือแบบจำลองการถดถอยเชิงเส้นอย่างง่ายเมื่อมีตัวแปรอิสระเพียงตัวเดียว

```
In [194]: kale_predict_model = LinearRegression()
          kale_predict_model.fit(X_train,Y_train)

Out[194]: LinearRegression()
```

ภาพที่ 6 แสดงคำสั่งการเรียกใช้ฟังก์ชันสร้างแบบจำลองเชิงเส้น

เพื่อทำความเข้าใจผลลัพธ์ของโมเดลการถดถอยเชิงเส้นพหุคูณ จะทำการตรวจสอบค่าคงที่ของสมการ และค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัว ซึ่งค่าเหล่านี้จะบ่งบอกถึงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม และสามารถนำไปใช้ในการตีความและทำนายผลผลิตได้

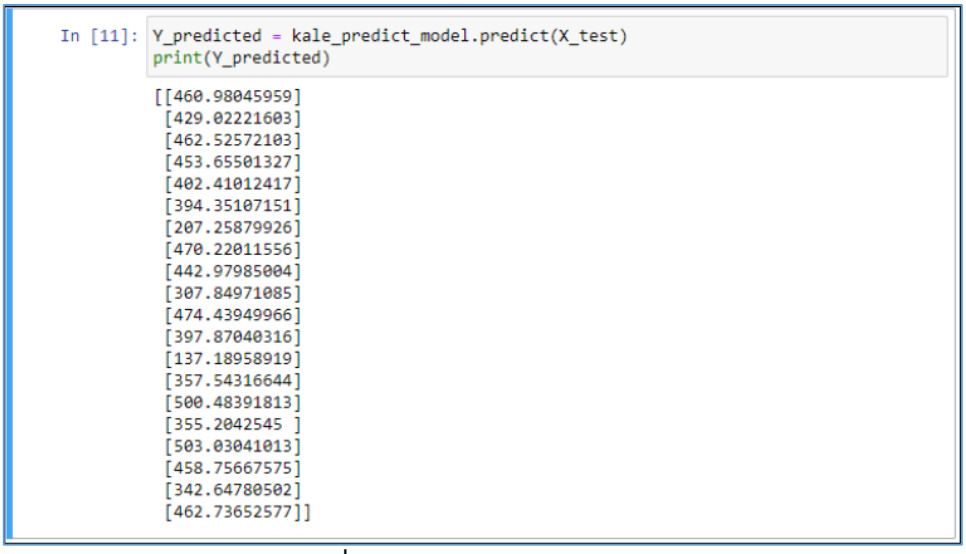
```
In [252]: intercept = kale_predict_model.intercept_[0]
          coefficent = kale_predict_model.coef_[0][0]
          print("The intercept for kale_predict model is {}".format(float(intercept)))
          print("\n")

          for coef in zip(X.columns, kale_predict_model.coef_[0]):
              print("The Coefficient for {} is {}".format(coef[0],float(coef[1])))

          The intercept for kale_predict model is 1199.8432203691455


          The Coefficient for temp_max is 17.29023494226685
          The Coefficient for temp_min is 11.345801828326929
          The Coefficient for humin_avg is 5.411927031563421
          The Coefficient for humin_max is -22.224222957788218
          The Coefficient for humin_min is -3.077415823805867
          The Coefficient for win_avg is -5604.151546967964
          The Coefficient for farmer is 59.71260491747818
```

ภาพที่ 7 แสดงค่าจุดตัด (Intercept) และค่าสัมประสิทธิ์การถดถอย (Regression Coefficient)

จากค่าที่ได้จึงแทนค่าในรูปสมการได้ดังนี้

$$y = 1199.84 + (17.29x_1) + (11.35x_2) + (5.41x_3) + (-22.22x_4) + (-3.08x_5) + (-5604.15x_6) + (59.71x_7)$$

จากนั้น ดำเนินการทำนายผลผลิตโดยใช้ชุดข้อมูลทดสอบ (X_test) และเก็บผลลัพธ์การทำนายไว้ในตัวแปร Y_predicted



```
In [11]: Y_predicted = kale_predict_model.predict(X_test)
         print(Y_predicted)

         [[460.98045959]
          [429.02221603]
          [462.52572103]
          [453.65501327]
          [402.41012417]
          [394.35107151]
          [207.25879926]
          [470.22011556]
          [442.97985004]
          [307.84971085]
          [474.43949966]
          [397.87040316]
          [137.18958919]
          [357.54316644]
          [500.48391813]
          [355.2042545 ]
          [503.03041013]
          [458.75667575]
          [342.64780502]
          [462.73652577]]
```

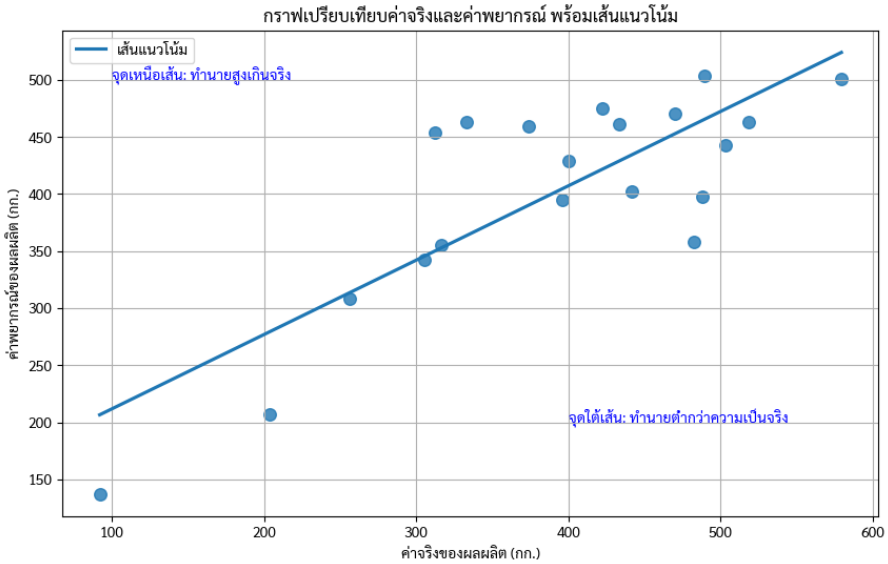
ภาพที่ 8 แสดงชุดข้อมูลพยากรณ์ผลผลิต

5. ประเมินประสิทธิภาพของแบบจำลอง (Evaluation) เพื่อวัดประสิทธิภาพการพยากรณ์ของแบบจำลอง ผู้วิจัยเลือกใช้ชุดข้อมูลทดสอบ โดยใช้ทั้งชุดข้อมูลทดสอบตัวแปรตามและชุดข้อมูลทดสอบตัวแปรอิสระในการทดสอบตามขั้นตอนต่าง ๆ ดังนี้

5.1 มาตราค่าผิดพลาด (Measures of Error) ระหว่างค่าที่เกิดขึ้นจริงกับค่าพยากรณ์ มีค่าที่ต้องการทราบ คือ ค่ารากที่สองเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error (RMSE)) โดยใช้ฟังก์ชัน mean_squared_error() พบว่าค่า RMSE มีค่าเท่ากับ 68.73 กิโลกรัม ซึ่งบ่งชี้ถึงความคลาดเคลื่อนเฉลี่ยของผลผลิตที่โมเดลทำนายเทียบกับค่าจริง อย่างไรก็ตาม เนื่องจากชุดข้อมูลมีค่าผิดปกติ ซึ่ง ค่า RMSE มีความอ่อนไหวต่อค่าผิดปกติดังกล่าว จึงได้ทำการประเมินความคลาดเคลื่อนเพิ่มเติมโดยใช้ค่าเฉลี่ยความผิดพลาดสัมบูรณ์ (Mean Absolute Error (MAE)) โดยใช้ฟังก์ชัน mean_absolute_error() ซึ่งมีค่าเท่ากับ 55.33 กิโลกรัม ซึ่งให้ค่าความคลาดเคลื่อนที่ทนทานต่อค่าผิดปกติได้ดีกว่า

5.2 เพื่อประเมินความสามารถในการอธิบายของโมเดล ผู้วิจัยได้คำนวณค่าสัมประสิทธิ์การตัดสินใจ (R-squared) โดยใช้ฟังก์ชัน r2_score() ซึ่งได้ผลลัพธ์เท่ากับ 0.6504 แสดงให้เห็นว่าตัวแปรอิสระทั้งหมดที่รวมอยู่ในโมเดลสามารถอธิบายความแปรปรวนของตัวแปรตาม (ปริมาณผลผลิต) ได้ 65.04% ซึ่งบ่งชี้ว่าโมเดลมีความสามารถในการอธิบายความสัมพันธ์ระหว่างตัวแปรอิสระและผลผลิตได้ในระดับปานกลาง

เพื่อประเมินแนวโน้มของผลการพยากรณ์ ได้มีการสร้าง Scatter Plot เพื่อเปรียบเทียบค่าจริงของผลผลิตกับค่าที่ได้จากโมเดล โดยกราฟนี้ใช้ในการวิเคราะห์แนวโน้มและความแม่นยำของการพยากรณ์ว่ามีความสอดคล้องกับค่าจริงหรือมีความคลาดเคลื่อนมากน้อยเพียงใด ซึ่งช่วยในการประเมินประสิทธิภาพของโมเดลในเชิงลึก



ภาพที่ 9 แสดงความสัมพันธ์ระหว่างค่าจริงและค่าพยากรณ์ของผลผลิต พร้อมเส้นแนวโน้มการทำนายของแบบจำลอง

จากภาพพบว่าจุดข้อมูลส่วนใหญ่อยู่ใกล้เส้นแนวโน้ม แสดงว่าโมเดลมีความสามารถในการทำนายผลผลิตอยู่ในระดับที่น่าพอใจ อย่างไรก็ตาม พบว่ามีจุดข้อมูลบางส่วนกระจายอยู่เหนือเส้นแนวโน้ม แสดงถึงการทำนายผลผลิตที่สูงเกินจริง (Overestimate) และบางจุดอยู่ใต้เส้นแนวโน้ม แสดงถึงการทำนายผลผลิตต่ำกว่าความเป็นจริง (Underestimate) ซึ่งสะท้อนให้เห็นว่าโมเดลยังคงมีข้อจำกัดบางประการที่ควรปรับปรุงเพิ่มเติมในอนาคต เพื่อให้การพยากรณ์แม่นยำยิ่งขึ้นและสามารถนำไปใช้งานจริงได้อย่างมีประสิทธิภาพมากขึ้น

ตารางที่ 1 แสดงค่าประสิทธิภาพของแบบจำลอง

ตัวชี้วัด	ค่าที่ได้	หน่วย	ความหมาย
RMSE	68.73	กก.	ค่าคลาดเคลื่อนเฉลี่ยในการทำนาย
MAE	55.33	กก.	ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย
R-squared	0.6504	-	ความสามารถของโมเดลในการอธิบายความแปรปรวน

จากตารางที่ 1 แสดงค่าประสิทธิภาพของแบบจำลองการถดถอยเชิงเส้นพหุคูณในการพยากรณ์ผลผลิตพืชผักสลัดจากชุดข้อมูลทดสอบ ซึ่งประกอบด้วยค่ารากที่สองเฉลี่ยความผิดพลาดกำลังสอง (RMSE) เท่ากับ 68.73 กิโลกรัม ค่าความผิดพลาดสัมบูรณ์เฉลี่ยเท่ากับ 55.33 กิโลกรัม และค่าสัมประสิทธิ์การตัดสินใจ (R-squared) เท่ากับ 0.6504

เตรียมสร้างตารางสรุปชุดข้อมูล (Summary of the Model Output) โดยการเรียกใช้ฟังก์ชัน add_constant() และกำหนดข้อมูลค่าคงที่ที่ต้องป้อนคือ ชุดข้อมูลตัวแปรอิสระ (X) จากนั้นกำหนดรูปแบบสมการโดยใช้ฟังก์ชัน OLS() และให้การเรียนรู้ของเครื่องสร้างแบบจำลองโดยใช้ฟังก์ชัน fit()

5.3 ช่วงความเชื่อมั่น (Confidence Intervals (CI)) โดยค่า CI จะแสดงให้เห็นถึงช่วงความเชื่อมั่นในระดับ 95% ซึ่งเป็นค่าตั้งต้นของฟังก์ชันนี้ แต่ค่า CI ยังสามารถบอกในนัยสำคัญทางสถิติได้อีกด้วย ซึ่งหากช่วงของ CI คร่อม Null Hypothesis หรือ Null Value ในที่นี้คือ 0 นั่นหมายถึง ตัวแปรนั้นไม่มีนัยสำคัญทางสถิติ ในทางกลับกันหากช่วง CI ไม่คร่อม 0 นั่นหมายถึง มีนัยสำคัญทางสถิติ โดยใช้คำสั่ง conf_int()

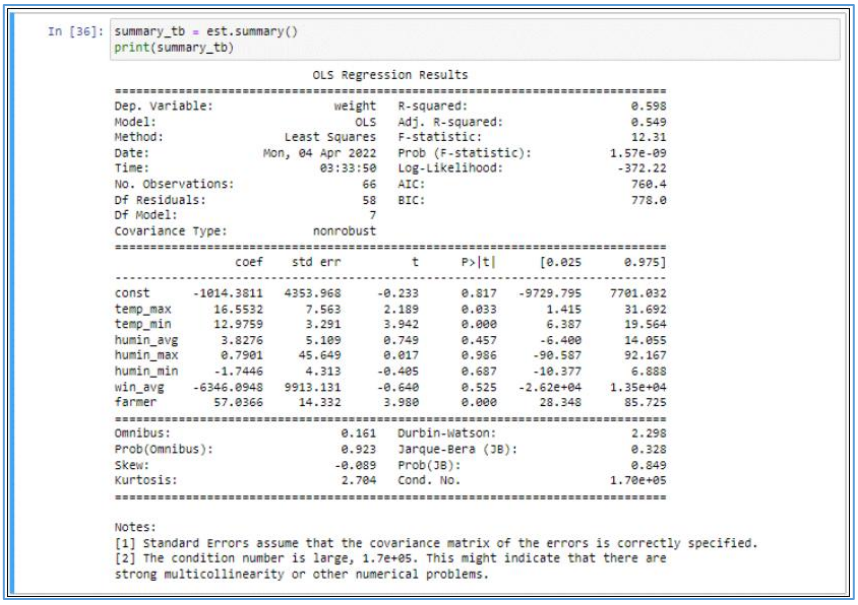
```
In [33]: linear_multi_Reg.conf_int()
Out[33]:
```

	0	1
const	-9729.794572	7701.032467
temp_max	1.414521	31.691787
temp_min	6.387384	19.564376
humin_avg	-6.399668	14.054779
humin_max	-90.586898	92.167147
humin_min	-10.377135	6.888001
win_avg	-26189.382950	13497.193439
farmer	28.348151	85.725092

ภาพที่ 10 แสดงผลลัพธ์ของช่วงความเชื่อมั่น (Confidence Intervals (CI))

จากภาพ แสดงค่าช่วงความเชื่อมั่นพบว่า ช่วงข้อมูล อุณหภูมิสูงสุด (1.414521) , อุณหภูมิต่ำสุด (6.387384) และจำนวนเกษตรกร (28.348151) อยู่ในช่วงไม่คร่อมเลขศูนย์ หรือกล่าวคือมีนัยสำคัญทางสถิติ ในทางกลับกัน ช่วงข้อมูล ความชื้นเฉลี่ย (-6.399668) , ความชื้นสูงสุด (-90.586898) , ความชื้นต่ำสุด (-10.377135) และความเร็วลมเฉลี่ย (-26189.382950) อยู่ในช่วงคร่อม 0 หรือกล่าวคือ ไม่มีนัยสำคัญทางสถิติ

5.4 ค่า P-Value จากตารางสรุปชุดข้อมูล เพื่อเปรียบเทียบนัยสำคัญทางสถิติ กับค่าของช่วงความเชื่อมั่นว่าสอดคล้องไปในทิศทางเดียวกันหรือไม่ ซึ่งหากค่า P-Value > 0.05 ถือว่าไม่มีนัยสำคัญทางสถิติ โดยใช้คำสั่ง summary()



```
In [36]: summary_tb = est.summary()
         print(summary_tb)

                            OLS Regression Results
==============================================================================
Dep. Variable:                 weight   R-squared:                       0.598
Model:                            OLS   Adj. R-squared:                  0.549
Method:                 Least Squares   F-statistic:                     12.31
Date:                Mon, 04 Apr 2022   Prob (F-statistic):           1.57e-09
Time:                        03:33:50   Log-Likelihood:                -372.22
No. Observations:                  66   AIC:                             760.4
Df Residuals:                      58   BIC:                             778.0
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -1014.3811   4353.968     -0.233      0.817   -9729.795    7701.032
temp_max      16.5532      7.563      2.189      0.033       1.415      31.692
temp_min      12.9759      3.291      3.942      0.000       6.387      19.564
humin_avg      3.8276      5.109      0.749      0.457      -6.400      14.055
humin_max      0.7901     45.649      0.017      0.986     -90.587      92.167
humin_min     -1.7446      4.313     -0.405      0.687     -10.377       6.888
win_avg    -6346.0948   9913.131     -0.640      0.525   -2.62e+04    1.35e+04
farmer        57.0366     14.332      3.980      0.000      28.348      85.725
==============================================================================
Omnibus:                        0.161   Durbin-Watson:                   2.298
Prob(Omnibus):                  0.923   Jarque-Bera (JB):                0.328
Skew:                          -0.089   Prob(JB):                        0.849
Kurtosis:                       2.704   Cond. No.                     1.70e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.7e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

ภาพที่ 11 แสดงคำสั่งตารางสรุปชุดข้อมูล (Summary of the Model Output)

จากภาพพบว่า ค่า P-Value ของข้อมูล อุณหภูมิสูงสุด (0.033) , อุณหภูมิต่ำสุด (0.000) และจำนวนเกษตรกร (0.000) ไม่เกิน 0.05 หรือกล่าวคือมีนัยสำคัญทางสถิติ ในทางกลับกัน ช่วงข้อมูล ความชื้นเฉลี่ย (0.457) , ความชื้นสูงสุด (0.986) , ความชื้นต่ำสุด (0.687) และความเร็วลมเฉลี่ย (0.525) เกิน 0.05 หรือกล่าวคือ ไม่มีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับค่าช่วงความเชื่อมั่นในขั้นตอนก่อนหน้า

ผลการวิจัยและอภิปรายผล

1. งานวิจัยนี้ประสบความสำเร็จในการพัฒนาโมเดลทำนายผลผลิตพืชผักสลัดในโครงการหลวงป่าแป๋ โดยใช้ตัวแบบการถดถอยเชิงเส้นพหุคูณ ซึ่งประกอบด้วยตัวแปรอิสระ 7 ตัว ได้แก่ อุณหภูมิสูงสุด อุณหภูมิต่ำสุด ความชื้นเฉลี่ย ความชื้นสูงสุด ความชื้นต่ำสุด ความเร็วลมเฉลี่ย และจำนวนเกษตรกร โดยมีน้ำหนักผลผลิตเป็นตัวแปรตาม

2. ผลการประเมินประสิทธิภาพของโมเดลที่สร้างขึ้นเพื่อตรวจสอบความถูกต้องและประสิทธิภาพของโมเดลในการทำนายผลผลิต ได้ผลลัพธ์ต่าง ๆ ดังต่อไปนี้

2.1 ค่า RMSE และ MAE ที่ได้มีค่าเท่ากับ 68.73 กิโลกรัม และ 55.33 กิโลกรัม ตามลำดับ ซึ่งบ่งชี้ว่าโมเดลมีความสามารถในการทำนายผลผลิตได้ในระดับหนึ่ง โดยมีค่าความคลาดเคลื่อนที่ยอมรับได้ ทั้งนี้สอดคล้องกับผลการศึกษาก่อนหน้า ซึ่งระบุว่าแบบจำลองการถดถอยพหุคูณสามารถให้ความแม่นยำในการพยากรณ์อยู่ในระดับที่น่าพอใจ (ศศิธร คงอุดมทรัพย์ และพงศิษฏ์ ทวิชพงศ์ธร, 2564) อย่างไรก็ตาม เมื่อเปรียบเทียบค่าคลาดเคลื่อนกับผลผลิตเฉลี่ยของพืชผักสลัด ซึ่งโดยทั่วไปอยู่ที่ประมาณ 400–500 กิโลกรัมต่อรอบการผลิต พบว่าค่า RMSE และ MAE ดังกล่าวอยู่ในระดับปานกลางถึงสูง สะท้อนให้เห็นว่าโมเดลยังมี

ข้อจำกัดในด้านความแม่นยำ โดยเฉพาะเมื่อต้องการนำไปใช้ในการวางแผนการผลิตที่ต้องการความถูกต้องแม่นยำสูง จึงควรพัฒนาโมเดลเพิ่มเติมเพื่อเพิ่มความแม่นยำในอนาคต

2.2 ค่า R-squared ที่ 0.6504 แสดงให้เห็นว่าโมเดลสามารถอธิบายความแปรปรวนของผลผลิตได้ 65.04% ซึ่งบ่งชี้ถึงความสัมพันธ์ในระดับปานกลางระหว่างตัวแปรอิสระที่เลือกและผลผลิต อย่างไรก็ตาม ค่า R-squared ที่ไม่สูงมากนัก อาจชี้ให้เห็นว่ายังมีปัจจัยอื่น ๆ นอกเหนือจากตัวแปรที่เลือกไว้ในโมเดล ที่มีอิทธิพลต่อผลผลิตพืชผักสลัด

ผลการวิเคราะห์ทางสถิติพบว่า อุณหภูมิสูงสุด อุณหภูมิต่ำสุด และจำนวนเกษตรกร มีความสัมพันธ์อย่างมีนัยสำคัญกับผลผลิต (P-Value < 0.05) ในขณะที่ ความชื้นเฉลี่ย ความชื้นสูงสุด ความชื้นต่ำสุด และความเร็วลมเฉลี่ย ไม่พบความสัมพันธ์ที่มีนัยสำคัญทางสถิติ ผลการวิจัยนี้สอดคล้องกับงานวิจัยที่พบว่าอุณหภูมิมีผลต่อผลผลิตข้าวหอมมะลิ อย่างไรก็ตาม การที่ความชื้นไม่มีความสัมพันธ์อย่างมีนัยสำคัญกับผลผลิต อาจเป็นผลมาจากสภาพแวดล้อมเฉพาะของพื้นที่ศึกษา หรืออาจมีปัจจัยอื่น ๆ ที่เกี่ยวข้องกับความชื้นที่ไม่ได้นำมาพิจารณาในโมเดลนี้ (ปิยะพร แซ่ลิ้ม และคณะ, 2562) ซึ่งสอดคล้องกับข้อเสนอแนะของงานวิจัยที่เน้นย้ำถึงความสำคัญของการพิจารณาปัจจัยทางสิ่งแวดล้อมที่มีความเฉพาะเจาะจงต่อพื้นที่ในการศึกษา (อธิศา ปานคง และเพ็ญนภา ปาละปิน, 2561)

2.4 การวิเคราะห์ความคลาดเคลื่อนของโมเดล (Error Analysis) เมื่อพิจารณาค่าคลาดเคลื่อนของการทำนายเทียบกับค่าจริง พบว่าการกระจายตัวส่วนใหญ่มีลักษณะสมมาตรรอบค่าเฉลี่ย จึงไม่ปรากฏ Bias อย่างชัดเจน อย่างไรก็ตาม จากกราฟ Scatter Plot ที่แสดงความสัมพันธ์ของค่าจริงกับค่าพยากรณ์ พบว่าจุดข้อมูลบางส่วนมีการทำนายสูงกว่าความเป็นจริง (Overestimate) และต่ำกว่าความเป็นจริง (Underestimate) โดยเฉพาะอย่างยิ่งกลุ่มข้อมูลที่มีผลผลิตสูงมาก แบบจำลองมีแนวโน้มทำนายต่ำกว่าค่าจริง อย่างชัดเจน ซึ่งสะท้อนว่าปัจจัยเฉพาะทางของแต่ละแปลงหรือมาตรการจัดการอื่น ๆ อาจมีบทบาทสำคัญต่อปริมาณผลผลิต

3. งานวิจัยนี้มีความสอดคล้องกับงานวิจัยที่ศึกษาการประยุกต์ใช้การเรียนรู้ของเครื่องในการพยากรณ์ปริมาณผลผลิตพืชเศรษฐกิจในพื้นที่ระเบียงเศรษฐกิจภาคกลาง - ภาคตะวันตก โดยงานวิจัยดังกล่าวใช้เทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์ผลผลิตทางการเกษตร (สมฤดี กลิ่นหอม และคณะ, 2566) และพบว่าประสิทธิภาพของอัลกอริทึมที่ใช้ในการทำนายอาจแตกต่างกันไปขึ้นอยู่กับชนิดของพืชและปัจจัยอื่น ๆ ที่เกี่ยวข้อง อย่างไรก็ตาม งานวิจัยนี้มีความแตกต่างจากงานวิจัยดังกล่าวในด้านชนิดของพืชที่ศึกษา ตัวแปรที่ใช้ และพื้นที่ศึกษา ซึ่งอาจส่งผลต่อความแตกต่างของผลการวิจัย

สรุป

งานวิจัยนี้ประสบความสำเร็จในการพัฒนาโมเดลพยากรณ์ผลผลิตพืชผักสลัดในโครงการหลวงป่าแป๋ โดยนำเทคนิคการทำเหมืองข้อมูลตามแนวทาง CRISP-DM มาประยุกต์ร่วมกับการถดถอยเชิงเส้นพหุคูณ ผลลัพธ์ที่ได้แสดงให้เห็นว่าแบบจำลองสามารถอธิบายความแปรปรวนของผลผลิตได้ถึง 65.04% โดยมีค่า RMSE เท่ากับ 68.73 กิโลกรัม และ MAE เท่ากับ 55.33 กิโลกรัม ซึ่งชี้ให้เห็นว่าแม้จะมีข้อจำกัดบางประการ แต่โมเดลนี้มี

ศักยภาพเพียงพอที่จะนำไปใช้งานจริงและสามารถเป็นเครื่องมือในการวางแผนการผลิตและการตลาดให้สอดคล้องกับความต้องการของตลาดได้

ผลการศึกษายังพบว่าปัจจัยสำคัญที่มีผลต่อการพยากรณ์ ได้แก่ อุณหภูมิสูงสุด อุณหภูมิต่ำสุด และจำนวนเกษตรกร ในขณะที่ปัจจัยอื่น ๆ เช่น ความชื้นและความเร็วลมไม่แสดงความสัมพันธ์อย่างมีนัยสำคัญ ส่งผลให้เห็นถึงความจำเป็นในการศึกษาและนำปัจจัยเพิ่มเติมมาใช้เพื่อปรับปรุงประสิทธิภาพของโมเดลในอนาคต

นอกจากนี้ มีงานวิจัยที่ยืนยันว่า การประยุกต์ใช้เทคนิค Machine Learning ในการพยากรณ์ผลผลิตสามารถเพิ่มความแม่นยำและรองรับความซับซ้อนของข้อมูลในสภาพแวดล้อมที่แตกต่างกันได้ (Reddy & Kumar, 2021; Venugopal et al., 2021) ซึ่งเป็นข้อพิสูจน์เพิ่มเติมว่าแนวทางที่ใช้ในงานวิจัยนี้มีความน่าเชื่อถือและมีศักยภาพในการพัฒนาต่อยอดเป็นระบบสนับสนุนการตัดสินใจ (DSS) ที่สามารถขยายการใช้งานไปยังพื้นที่ที่มีลักษณะทางภูมิศาสตร์หรือระบบการผลิตที่คล้ายคลึงกัน เพื่อสนับสนุนการพัฒนาเกษตรกรรมอย่างยั่งยืนในอนาคต

ข้อเสนอแนะ

ข้อเสนอแนะเกี่ยวกับงานวิจัย

แม้ว่าโมเดลที่พัฒนาขึ้นจะแสดงให้เห็นถึงประสิทธิภาพในการทำนายผลผลิตพืชผักสลัดได้ในระดับหนึ่ง แต่ยังคงมีโอกาสในการปรับปรุงเพื่อเพิ่มความแม่นยำยิ่งขึ้น การพิจารณาเพิ่มเติมปัจจัยอื่น ๆ ที่อาจส่งผลต่อผลผลิต เช่น ปริมาณน้ำฝน ชนิดของปุ๋ย และการระบาดของโรคและแมลงศัตรูพืช อาจช่วยเพิ่มความสามารถในการอธิบายของโมเดล นอกจากนี้ การสำรวจเทคนิคการสร้างโมเดลที่ซับซ้อนมากขึ้น เช่น Support Vector Regression, Random Forest หรือ Neural Network อาจนำไปสู่การพยากรณ์ที่แม่นยำยิ่งขึ้น เพื่อให้มั่นใจในความสามารถในการนำโมเดลไปใช้ในสถานการณ์จริง การตรวจสอบความถูกต้องของโมเดลกับข้อมูลใหม่ที่ไม่ได้ใช้ในการสร้างโมเดลเป็นสิ่งจำเป็น ในขั้นตอนสุดท้าย ควรพัฒนาต่อยอดโมเดลให้สามารถใช้งานได้ง่ายและสะดวกสำหรับเกษตรกรและผู้ที่เกี่ยวข้อง เช่น ผ่านแอปพลิเคชันบนมือถือ หรือระบบเว็บไซต์ที่ใช้งานง่าย ซึ่งจะช่วยให้สามารถนำผลการวิจัยไปใช้ประโยชน์ได้จริงในวงกว้าง

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

เพื่อเพิ่มประสิทธิภาพและความสามารถในการนำโมเดลไปใช้งานจริงในสถานการณ์ที่หลากหลาย ควรมีการเก็บข้อมูลปัจจัยต่าง ๆ และผลผลิตในระยะยาว เพื่อให้โมเดลสามารถเรียนรู้และปรับตัวเข้ากับความผันผวนของสภาพแวดล้อมและปัจจัยอื่น ๆ ได้ดียิ่งขึ้น นอกจากนี้ ยังควรสำรวจปัจจัยอื่น ๆ ที่อาจส่งผลต่อผลผลิต เช่น ชนิดของดิน เทคนิคการเพาะปลูก และปัจจัยทางเศรษฐกิจและสังคม เพื่อเพิ่มความครอบคลุมและความแม่นยำของโมเดล อีกทั้งการเปรียบเทียบประสิทธิภาพของโมเดลที่พัฒนาขึ้นกับโมเดลอื่น ๆ หรือวิธีการพยากรณ์แบบดั้งเดิมจะช่วยในการประเมินและเลือกโมเดลที่เหมาะสมที่สุดสำหรับการทำนายผลผลิตพืชผักสลัด มีงานวิจัยที่ได้นำเสนอแนวทางการเปรียบเทียบโมเดลการเรียนรู้แบบมีการกำกับโดยใช้เทคนิค Regression Models ที่สามารถนำมาประยุกต์ใช้ร่วมกับงานวิจัยในอนาคตเพื่อเพิ่มความแม่นยำของโมเดลโดยการรวมเอา

เทคนิค Machine Learning ที่สามารถจับความซับซ้อนและลักษณะความไม่เชิงเส้นของข้อมูลเข้ามาใช้เพิ่มเติม (Panigrahi et al., 2023; Reddy & Kumar, 2021; Venugopal et al., 2021) สุดท้าย การขยายผลการศึกษาไปยังพืชชนิดอื่น ๆ หรือพื้นที่เพาะปลูกอื่น ๆ จะเป็นการตรวจสอบความสามารถในการประยุกต์ใช้งานในบริบทที่หลากหลายของโมเดลและเปิดโอกาสในการนำไปประยุกต์ใช้ในวงกว้างมากขึ้น ซึ่งจะส่งผลดีต่อภาคการเกษตรในภาพรวม

กิตติกรรมประกาศ

ขอขอบพระคุณ โครงการพัฒนาพื้นที่สูงแบบโครงการหลวงป่าแป๋ อำเภอแม่แตง จังหวัดเชียงใหม่ ที่ได้อนุเคราะห์ข้อมูลอันเป็นประโยชน์อย่างยิ่งต่องานวิจัยนี้ ข้อมูลที่ได้รับจากโครงการฯ มีส่วนสำคัญในการวิเคราะห์และสร้างโมเดลการทำนายผลผลิตพืชผักสลัด ซึ่งจะเป็นประโยชน์ต่อการพัฒนาการเกษตรในพื้นที่ต่อไป

เอกสารอ้างอิง

ปิยะพร แซ่ลิ้ม, วิวรรณ กาญจนวจี, พัชนา สุวรรณแสน, และณพฐ์ โสภีพันธ์. (2562). การพยากรณ์ผลผลิตข้าวหอมมะลิ ในจังหวัดนครราชสีมา. **วารสารวิจัยวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครราชสีมา**, 4(2), 25-37.

พัชรา ศรีพระบุ. (2565). การวิเคราะห์การถดถอยพหุคูณแบบขั้นตอนของปัจจัยพยากรณ์ที่ส่งผลต่อการตัดสินใจเลือกใช้บริการท่าเรือสินค้าทั่วไป. **วารสารวิศวกรรมศาสตร์และนวัตกรรม**, 15(3), 154-163.

วีรศักดิ์ ฟองเงิน, วรปภา อารีราษฎร์, และเผด็จ พรหมสาขา ณ สกลนคร. (2560). การพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม โดยใช้เทคนิคเหมืองข้อมูล. **วารสารวิทยาการจัดการสมัยใหม่**, 10(2), 121-131.

ศศิธร คงอุดมทรัพย์, และพงศิษฏ์ ทวิชพงศ์ธร. (2564). ความแม่นยำของการใช้แบบจำลองการถดถอยพหุคูณในการวิเคราะห์ฐานข้อมูลที่ได้จากการสุ่มแบบมีเงื่อนไขเพื่อการคาดการณ์ผลการเรียนของนักเรียนนายเรือ. **วารสารวิชาการโรงเรียนนายเรือด้านวิทยาศาสตร์และเทคโนโลยี**, 4(1), 76-88.

สถาบันวิจัยและพัฒนาพื้นที่สูง (สวพส.). (2563). **วัตถุประสงค์ ภารกิจหลัก อำนาจหน้าที่**. สืบค้นจาก https://www.hrdi.or.th/About/Introduce

สมฤดี กลิ่นหอม, สุพิชญาไพจิตต์, ปัณณวุฒิ ปิ่นสวาสดิ์, และเยาวลักษณ์ เกิดปั้น. (2566). ก ารประยุกต์การเรียนรู้ของเครื่องในการพยากรณ์ปริมาณผลผลิตพืชเศรษฐกิจในโซ่อุปทานภาคเกษตรกรรม พื้นที่ระเบียงเศรษฐกิจภาคกลาง –ภาคตะวันตก. **วารสารวิทยาศาสตร์และเทคโนโลยีนอร์ทเทิร์น**, 4(3), 63-76.

อธิศา ปานคง, และเพ็ญนภา ปาละปิน. (2561). **ปัจจัยที่เป็นอุปสรรคในการปลูกพืชหมุนเวียนในพื้นที่เกษตรกรรม อำเภอบ้านดุง จังหวัดอุดรธานี**. การประชุมวิชาการและนำเสนอผลงานระดับชาติด้านมนุษยศาสตร์และสังคมศาสตร์ ครั้งที่ 1 ประจำปีการศึกษา 2561 (น. 383-390). มหาวิทยาลัยราชภัฏสวนสุนันทา.

Joshua, S. V., Priyadharson, A. S. M., Kannadasan, R., Khan, A. A., Lawanont, W., Khan, F. A., Rehman, A. U., & Ali, M. J. (2022). Crop yield prediction using machine learning approaches on a wide spectrum. **Computers, Materials & Continua**, 72(3), 5663-5679. https://doi.org/10.32604/cmc.2022.027178

Khade, S., Malaye, V., Mhase, S., Shitole, A., & Shirsath, S. (2023). Crop yield prediction using machine learning. **International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)**, 3(1), 202-204. https://doi.org/10.48175/ijarsct-8584

Panigrahi, B., Kathala, K. C. R., & Sujatha, M. (2023). A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models. **Procedia Computer Science,** 218, 2684-2693. https://doi.org/10.1016/j.procs.2023.01.241

Reddy, D. J., & Kumar, M. R. (2021). Crop yield prediction using machine learning algorithms. **In Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021)** (pp. 1466-1470). https://doi.org/10.1109/ICICCS51141.2021.9432236

Venugopal, A., S, A., Mani, J., Mathew, R., & Williams, V. (2021). Crop yield prediction using machine learning algorithms. **International Journal of Engineering Research & Technology (IJERT)**, 9(13), 87-91.