

การพยากรณ์ความต้องการโซ่อุปทานของเศษวัสดุทางการเกษตรสำหรับโรงไฟฟ้าชีวมวลด้วย การเรียนรู้ของเครื่อง

จิตาภา ขาญเจริญ¹ และ กรกฎ ไยบัวเทศ ทิพย์วงศ์^{*2}

ภาควิชาวิศวกรรมอุตสาหกรรม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่ เชียงใหม่ 50200

กลุ่มวิจัยการจัดการโซ่อุปทานและวิศวกรรม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่ เชียงใหม่ 50200

Received: 03 March 2024; Revised: 08 May 2024; Accepted: 23 May 2024

บทคัดย่อ

งานวิจัยนี้ได้ศึกษาและวิเคราะห์โซ่อุปทานของชีวมวลที่เกิดจากเศษวัสดุทางการเกษตรของพืชเศรษฐกิจจำนวน 5 ชนิด ในพื้นที่ภาคเหนือจำนวน 17 จังหวัด เพื่อคำนึงถึงปริมาณการเกิดของเศษวัสดุทางการเกษตรการเก็บเกี่ยวพืชผลที่สามารถนำไปสู่ความต้องการนำไปใช้ในการผลิตไฟฟ้าในรูปแบบชีวมวลได้ จึงนำข้อมูลในอดีตมาประยุกต์ใช้กับแบบจำลองการเรียนรู้ของเครื่อง 5 แบบจำลอง เพื่อหาแบบจำลองที่เหมาะสมที่สุดสำหรับชุดข้อมูลที่มีปัจจัย 11 ปัจจัย โดยแบ่งข้อมูลออกเป็นชุดฝึก 80 เปอร์เซ็นต์ และชุดทดสอบ 20 เปอร์เซ็นต์ แบบจำลองที่ดีที่สุดสำหรับชุดข้อมูลนี้เพื่อการพยากรณ์ผลผลิตทางการเกษตรคือ Random Forest Regression โดยใช้ค่า $n_neighbors = 1$ โดยมีค่า MAPE = 27.511 เปอร์เซ็นต์ ที่ได้จากการประเมินผลความแม่นยำของแบบจำลอง จึงสรุปได้ว่า สำหรับชุดข้อมูลและปัจจัยที่กำหนด สามารถใช้การเรียนรู้ของเครื่องในการพยากรณ์ผลผลิตทางการเกษตรได้ ซึ่งเป็นแนวทางในการนำค่าพยากรณ์ที่ได้จากแบบจำลองมาใช้เพื่อคำนวณหาเศษวัสดุทางการเกษตรที่เกิดจากการเพาะปลูกพืชต่าง ๆ และนำข้อมูลเหล่านี้ไปใช้ในการคำนวณค่าพลังงานชีวมวลที่ได้จากเศษวัสดุทางการเกษตรสำหรับโรงไฟฟ้าชีวมวลได้

คำสำคัญ: พลังงานชีวมวล, เศษวัสดุทางการเกษตร, การพยากรณ์, การเรียนรู้ของเครื่อง

* Corresponding author. E-mail: korrakot@eng.cmu.ac.th

¹ นักศึกษาระดับบัณฑิตศึกษา ภาควิชาวิศวกรรมอุตสาหกรรม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่

² ผู้ช่วยศาสตราจารย์ ภาควิชาวิศวกรรมอุตสาหกรรม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่

Forecasting the Supply Chain Demand of Agricultural Residues for Biomass Power Plants Using Machine Learning

Jidapa Chanjaroen¹ and Korrakot Yaibuathet Tippayawong^{*2}

Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University,
Chiang Mai 50200

Supply Chain and Engineering Management Research Unit, Faculty of Engineering,
Chiang Mai University, Chiang Mai 50200

Received: 03 March 2024; Revised: 08 May 2024; Accepted: 23 May 2024

Abstract

This research investigated the supply chain of biomass derived from agricultural residues of five crop types in the Northern region of 17 provinces in Thailand, with the aim of evaluating the potential quantity of agricultural residues available for electricity generation through biomass. Historical data was utilized, and 5 machine learning models were employed to determine the optimal model for the dataset, which included 11 factors. The dataset was divided into 80% for training the models and 20% for testing. The Random Forest Regression model with $n_neighbors = 1$ emerged as the most effective for predicting production from this dataset, achieving a Mean Absolute Percentage Error (MAPE) of 27.511%. The results indicate that machine learning techniques can be utilized to forecast agricultural production effectively. This approach enables the estimation of agricultural residues from various crops, thus facilitating the calculation of biomass energy derived from agricultural residues for biomass power generation. The implications of these findings are substantial, as they demonstrate the potential of machine learning in the utilization of biomass resources for renewable energy production.

Keywords: biomass, agricultural residues, forecasting, machine learning

* Corresponding author. E-mail: korrakot@eng.cmu.ac.th

1 Graduate Student in Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University

2 Assistant Professor in Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University

1. บทนำ

ชีวมวลเป็นแหล่งพลังงานทดแทนที่มีอย่างจำกัด และปริมาณของชีวมวลมีความไม่แน่นอน ซึ่งขึ้นอยู่กับสภาพแวดล้อมและความต้องการของแต่ละพื้นที่ รวมถึงต้นทุนในการขนส่งเชื้อเพลิงไปยังโรงไฟฟ้าชีวมวลที่ค่อนข้างสูง ทำให้การจัดการด้านพลังงานเป็นเรื่องยาก และส่งผลให้การจัดการโซ่อุปทานชีวมวลขาดประสิทธิภาพ ซึ่งมาจากปัจจัยต่างๆ เช่น ความพร้อมในการใช้เชื้อเพลิงชีวมวล คุณภาพของชีวมวล ค่าขนส่ง ความต้องการตลาด เป็นต้น จึงทำให้เกิดวิธีการที่จะเข้าไปช่วยแก้ปัญหาด้านการจัดการ เพื่อให้เกิดการดำเนินงานที่ยั่งยืนภายในโซ่อุปทานที่เชื่อมต่อกันตั้งแต่แหล่งวัตถุดิบจนถึงปลายทางในการนำไปใช้งาน

การเรียนรู้ของเครื่อง (Machine Learning) เป็นรูปแบบหนึ่งในปัญญาประดิษฐ์ (Artificial Intelligence) ที่ช่วยในการเพิ่มขีดความสามารถในความแม่นยำของการวิเคราะห์ หรือการสร้างแบบจำลอง การพยากรณ์ โดยใช้เทคนิคของโปรแกรมคอมพิวเตอร์ ที่อาศัยอัลกอริทึม (Algorithm) หรือแบบจำลอง (Model) จากการเรียนรู้ผ่านข้อมูล ซึ่งโปรแกรม Python เป็นภาษาโปรแกรมหนึ่งที่ถูกนำมาใช้ในปัญญาประดิษฐ์ ที่มีความสามารถในการใช้เพื่อแก้ปัญหาของฟังก์ชันเชิงอนุพันธ์และความน่าจะเป็น และการหาความสัมพันธ์ของปัจจัยต่างๆ โดยใส่ข้อมูลที่ส่งผลการพยากรณ์ และพิจารณาปัจจัยที่จะช่วยให้การพยากรณ์ในอนาคตมีความแม่นยำขึ้น สำหรับโปรแกรม Python จะมีไลบรารีที่เรียกว่า Scikit-learn ที่มีฟังก์ชันและอัลกอริทึมที่จำเป็นและสามารถเรียกใช้งานได้ซึ่งเป็นประโยชน์ต่อการนำมาประยุกต์ใช้และเป็นพื้นฐานสำหรับการทำงาน

การประเมินศักยภาพเกษตรชีวมวลทางการเกษตรในเชื้อเพลิงในโรงไฟฟ้าชีวมวลอาจประเมินได้ยาก เนื่องจากปริมาณชีวมวลที่ไม่พอเพียงในบางพื้นที่ หรือความไม่แน่นอนของปริมาณชีวมวลในการนำไปผลิตไฟฟ้า อีกทั้งการเกษตรในพื้นที่ภาคเหนือยังขาดการจัดการพื้นที่เพาะปลูก และความคุ้มค่าในการนำวัตถุดิบมาเป็นเชื้อเพลิง จึงทำให้บางพื้นที่ที่มีปริมาณเชื้อเพลิงไม่เพียงพอต่อความต้องการ ปัญหานี้แสดงให้เห็นว่า มีความเกี่ยวข้องกับโซ่อุปทานของชีวมวล ในด้านผลผลิตและเก็บเกี่ยว การจัดเก็บ การขนส่ง จนไปถึงการแปลงเป็นพลังงาน

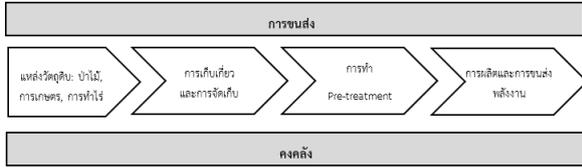
จากศึกษาข้อมูลของสำนักงานเศรษฐกิจการเกษตร (Office of Agricultural Economics: AOE) แล้วพบว่าพืชเศรษฐกิจในภาคเหนือ ได้แก่ ข้าวนาปี ข้าวนาปรัง อ้อย โรงงาน มันปะหลัง และข้าวโพดเลี้ยงสัตว์ [1] ซึ่งจากกระทรวงพลังงานได้มีการให้ข้อมูลเกี่ยวกับการนำเศษวัสดุทางการเกษตรจากพืชเศรษฐกิจเหล่านี้ ไปเป็นชีวมวลใช้ในการผลิตไฟฟ้า โดยการทราบถึงปริมาณเศษวัสดุทางการเกษตร สามารถหาได้จากผลผลิตที่เกิดขึ้นต่อปีคำนวณร่วมกับสัดส่วนการเกิดเศษวัสดุเหลือใช้ต่อผลของแต่ละชนิดพืช [2], [3] นอกจากนี้กระทรวงพลังงาน มีข้อมูลโรงไฟฟ้าชีวมวลในภาคเหนือ 46 โรงในการผลิตไฟฟ้าที่จ่ายไฟเข้าระบบแล้ว (Commercial Operation Date: COD) ทำให้ทราบกำลังการผลิตไฟฟ้าจากชีวมวลได้จากที่สามารถนำมาสอดคล้องกันเพื่อทราบความต้องการของโซ่อุปทานเศษวัสดุทางการเกษตรสำหรับโรงไฟฟ้าชีวมวลได้

ดังนั้นในงานวิจัยนี้ได้ทำการศึกษาและวิเคราะห์องค์ประกอบโซ่อุปทานการผลิตไฟฟ้าด้วยโรงไฟฟ้าชีวมวล เพื่อนำไปพยากรณ์ความต้องการเศษวัสดุทางการเกษตรที่เกิดขึ้น โดยการใช้การเรียนรู้ของเครื่อง ซึ่งสิ่งหนึ่งที่เป็นส่วนในการพิจารณาของการผลิตของโรงไฟฟ้าชีวมวล มักจะขึ้นอยู่กับ การตัดสินใจเกี่ยวกับปริมาณของทรัพยากรและสถานที่ในการจัดตั้งโรงงาน จึงเป็นผลให้งานวิจัยนี้ได้มุ่งเน้นในการศึกษาปริมาณของผลผลิตของพืชเศรษฐกิจในพื้นที่ภาคเหนือจำนวน 17 จังหวัด เพื่อคำนึงถึงปริมาณการเกิดของเศษวัสดุทางการเกษตรจากการเก็บเกี่ยวพืชผลที่สามารถนำไปสู่ความต้องการนำไปใช้ในการผลิตไฟฟ้าในรูปแบบชีวมวลได้

2. หลักการและทฤษฎีที่เกี่ยวข้อง

พลังงานชีวมวล เป็นพลังงานที่ได้มาจากเชื้อเพลิงชีวมวลที่ผ่านกระบวนการแปรสภาพเพื่อผลิตเป็นไฟฟ้า โดยใช้ความร้อนจากเผาไหม้ของโรงไฟฟ้าชีวมวล ซึ่งชีวมวลบางส่วนมาจากเศษวัสดุทางการเกษตรที่มาจากกิจกรรมการเพาะปลูกและเก็บเกี่ยว [4] งานวิจัยในอดีตได้แสดงให้เห็นว่าหนึ่งในอิทธิพลที่มีผลต่อโซ่อุปทานของความพร้อมของทรัพยากรชีวมวล คือ เศษวัสดุทางการเกษตรที่ใช้ประโยชน์ได้ ซึ่งพื้นที่การเพาะปลูกและผลผลิตทางการเกษตรเป็น

ปัจจัยสำคัญต่อทรัพยากรชีวมวล [5], [6] เป็นผลให้กิจกรรมต่างๆ มีความสัมพันธ์กันเป็นโซ่ปทานชีวมวล ดังรูปที่ 1 [7]



รูปที่ 1 โซ่ปทานของชีวมวล

การประเมินศักยภาพของพลังงานชีวมวลที่เกิดขึ้นจากเศษวัสดุทางการเกษตร สามารถประมาณการได้จากสัดส่วนของการเกิดเศษวัสดุทางการเกษตรของผลผลิตทางการเกษตรที่จะนำไปสู่การคำนวณร่วมกับค่าความร้อนที่เกิดขึ้นจนเป็นค่าพลังงาน [8] จึงเป็นผลให้เกิดการประยุกต์ใช้กับการพยากรณ์เศษวัสดุทางการเกษตรที่เกิดขึ้นด้วยผลผลิตทางการเกษตร โดยการใช้สมการที่ (1) ในการแปลงผลผลิตเป็นเศษวัสดุทางการเกษตร และสมการที่ (2) ในการแปลงเศษวัสดุทางการเกษตรเป็นค่าความร้อนของพลังงานชีวมวล [3], [9] โดยที่มีค่าสัดส่วนที่ใช้สำหรับการคำนวณดังตารางที่ 1 และตารางที่ 2 [10]

$$TBT = \sum_{i=1}^n CP(i) \times RPR(i) \times \left[\frac{100-M(i)}{100} \right] \quad (1)$$

โดยที่ TBT คือ ศักยภาพชีวมวลเชิงทฤษฎี (ตัน)
 CP(i) คือ ปริมาณผลผลิตทางการเกษตร (ตัน/ปี)
 RPR(i) คือ สัดส่วนการเกิดเศษวัสดุทางการเกษตร
 M(i) คือ สัดส่วนความชื้น
 N คือ จำนวนปี

$$TBP = \sum_{i=1}^n TBT(i) \times LHV \quad (2)$$

โดยที่ TBT(i) คือ ศักยภาพพลังงานเชิงทฤษฎี (ตัน)
 LHV คือ ค่าสัดส่วนความร้อนต่ำ (MJ/kg)
 n คือ จำนวนปี

ตารางที่ 1 ค่าสัดส่วนของศักยภาพของเศษวัสดุทางการเกษตร

ประเภท	เศษวัสดุ	ค่าความชื้น (%)	ค่าสัดส่วนการเกิดเศษ
อ้อยโรงงาน	ชานอ้อย	50	0.250
	ยอดและใบ	50	0.302
ข้าว	แกลบ	8.83	0.230
	ฟางข้าว	8.17	0.447
มันสำปะหลัง	กากมัน	-	0.088
ข้าวโพด	ชังข้าวโพด	8.65	0.250

ตารางที่ 2 ค่าสัดส่วนของศักยภาพพลังงานชีวมวล

ประเภท	เศษวัสดุทางการเกษตร	ค่าความร้อนต่ำ (MJ/kg)
อ้อยโรงงาน	ชานอ้อย	6.43
	ยอดและใบ	6.82
ข้าว	แกลบ	12.85
	ฟางข้าว	8.83
มันสำปะหลัง	กากมัน	16.99
ข้าวโพด	ชังข้าวโพด	16.63

ในการประยุกต์ใช้การเรียนรู้ของเครื่องสำหรับการพยากรณ์เป็นการคาดคะเนจากชุดตัวเลข ซึ่งเรียกว่า การถดถอย (Regression) ได้มีรูปแบบแบบจำลองที่หลากหลาย ได้แก่ Linear Regression, Support Vector Machines, Decision Trees และ Random Forests เป็นต้น [11], [12] ซึ่งเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นหนึ่งในรูปแบบการเรียนรู้ของเครื่องที่เรียนรู้ผ่านข้อมูลที่มีอยู่หรือข้อมูลในอดีตด้วยตัวเอง เมื่อก้าวถึงการประยุกต์ใช้กับการพยากรณ์ที่นำข้อมูลในอดีตมาใช้ในการประมวลผลผ่านแบบจำลองนั้นแบบจำลองจะทำการฝึกฝนข้อมูล (Training Data) ในขณะที่การทำนายตัวแปรตามต่อเนื่องจากชุดของตัวแปรอิสระโดยใช้การวิเคราะห์การถดถอย [13] ดังนั้นในงานวิจัยที่เกี่ยวข้องจึงมีการกำหนดข้อมูลนำเข้า (Input) เพื่อให้แบบจำลองได้เรียนรู้ข้อมูลและพยากรณ์เป็นข้อมูลผลลัพธ์ (Output) ดังนั้นในการประเมินศักยภาพของเศษวัสดุทางการเกษตรที่ได้มาจากการพยากรณ์ผลผลิตทางการเกษตรได้มีการกำหนดปัจจัย (Feature) ที่เป็นข้อมูลนำเข้าที่เกี่ยวข้อง เพื่อพยากรณ์เป็นผลลัพธ์เป็นปริมาณผลผลิต เช่น

อุณหภูมิ, ปริมาณฝน, พื้นที่เพาะปลูก, การใช้ปุ๋ย, การใช้อยาฆ่าแมลง เป็นต้น สิ่งเหล่านี้ถือว่าเป็นปัจจัยสำหรับชุดข้อมูลที่จะนำไปใช้ในแบบจำลอง [14]-[16]

ก่อนที่จะมีการนำปัจจัยต่างๆ ไปใช้ในการแบบจำลอง ปัจจัยเหล่านี้จะถูกรวบรวมความสัมพันธ์ระหว่างกัน โดยค่าสหสัมพันธ์ (Correlation Coefficient) มีค่าอยู่ในช่วง -1 ถึง 1 โดยที่ค่าสหสัมพันธ์เป็นบวกแสดงถึงความสัมพันธ์ที่เชื่อมโยงกันในทิศทางเดียวกัน ดังนั้น โดยปกติแล้วค่าสหสัมพันธ์มากกว่าค่าที่กำหนดไว้ เช่น ค่าอยู่ที่ 0.8 จะแสดงถึงความสัมพันธ์ที่มีความเข้มงวดมาก ซึ่งเมื่อปัจจัยที่มีผลความสัมพันธ์ระหว่างกันอาจช่วยให้สามารถทำนายผลได้ดีขึ้น [22] ดังนั้นจึงทำการเลือกปัจจัยที่มีค่าสหสัมพันธ์ที่มีความสัมพันธ์กันสูง ไม่ว่าจะเป็นในทิศทางเดียวกัน หรือทิศทางข้ามกัน มาใช้กับแบบจำลองนั้นให้เป็นข้อมูลนำเข้า จากนั้นแบบจำลองประมวลผลด้วยเรียนรู้ปัจจัยที่กำหนดไว้ จนได้แบบจำลองที่เหมาะสมกับชุดข้อมูลเหล่านั้น ซึ่งแบบจำลองที่เป็นที่นิยมสำหรับข้อมูลการถดถอย 5 แบบจำลองที่มีลักษณะการประมวลผลสำหรับข้อมูลที่ไม่มีความซับซ้อนมาก [11], [18], [19] ดังนี้

- Lasso Regression เป็นแบบจำลองที่ใช้กับข้อมูลที่เป็นการถดถอย เพื่อทำการพยากรณ์ที่มีความแม่นยำมากขึ้น แบบจำลองนี้ใช้การย่อขนาดของข้อมูล โดยที่ค่าข้อมูลจะถูกย่อให้ใกล้เคียงกับจุดกลางอย่างค่าเฉลี่ย ซึ่งเหมาะกับข้อมูลที่ไม่ซับซ้อนและมีปัจจัยน้อย

- Support Vector Regression เป็นแบบจำลองการถดถอยที่มีลักษณะการประมวลผลเช่นเดียวกับ Support Vector Machine ที่ใช้สำหรับการแยกแยะข้อมูล (Classification) ซึ่งแบบจำลองมีความยืดหยุ่นในการกำหนดจำนวนข้อผิดพลาดที่ยอมรับได้ในแบบจำลอง และจะค้นหาเส้นที่เหมาะสม เพื่อให้พอดีกับข้อมูล

- Random Forest Regression ทำงานโดยการสร้างต้นไม้ตัดสินใจจำนวนมากในขณะฝึกฝนข้อมูล และส่งออกคลาสที่เป็นการพยากรณ์เฉลี่ยสำหรับการถดถอยของต้นไม้แต่ละต้นซึ่งรวมผลลัพธ์จากการพยากรณ์หลายรายการ โดยรวมต้นไม้ตัดสินใจจำนวนมากด้วยการปรับแต่งข้อมูลที่มีประโยชน์

- Gradient Boosting Regression เป็นแบบจำลองที่ตัวทำนายจะทำการแก้ไขความคลาดเคลื่อน (Error) ของการทำนายก่อนหน้าได้ เป็นวิธีที่ใช้ในการลดความผิดพลาดในแบบจำลองและสามารถทำนายปัจจัยที่กำหนดไว้ ซึ่งแบบจำลองนี้มีความสามารถในการปรับปรุงฟังก์ชันสูญเสีย (Loss Function) ที่แตกต่างกันได้ เช่น Mean Squared Error เป็นต้น ซึ่งจะประมวลผลเสร็จสิ้นจนกว่าจะพบเกณฑ์ที่เหมาะสม

- K-nearest Neighbor เป็นแบบจำลองที่ไม่มีตัวแปรโดยประมาณการความสัมพันธ์ระหว่างตัวแปรอิสระกับผลลัพธ์ที่ต่อเนื่องโดยการหาค่าเฉลี่ยของการสังเกตในเซตรอบตัว ขนาดของเซตรอบตัวจำเป็นต้องถูกกำหนดโดยวิเคราะห์เชิงลึกหรือสามารถเลือกโดยใช้ cross-validation เพื่อเลือกขนาดที่ลดค่าความผิดพลาดเฉลี่ยของตัวแบบลงให้น้อยที่สุด

ในการประเมินแบบจำลองมีวิธีการประเมินความแม่นยำของผลการพยากรณ์เพื่อพิจารณาถึงความสามารถของแบบจำลองหลากหลายวิธีการด้วยสมการที่เกี่ยวข้องกับการถดถอยต่างๆ ยกตัวอย่างเช่น ความคลาดเคลื่อนของ Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scale Error (MASE) [20] แต่ถึงอย่างไรก็ตาม งานวิจัยนี้มุ่งเน้นที่จะเสนอความสามารถทางความแม่นยำของแบบจำลองดังสมการที่ (3) และ สมการที่ (4) เพื่อพิจารณาได้อย่างง่าย

$$R^2 = 1 - \left(\frac{\sum_i^n (Y_i - \hat{Y}_i)^2}{\sum_i^n Y_i - \bar{Y}_i} \right)^2 \quad (3)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (4)$$

โดยที่	Y_i	คือ ค่าจริง
	\hat{Y}_i	คือ ค่าพยากรณ์
	Y	คือ ค่าเฉลี่ยตัวอย่างของ Y
	n	คือ จำนวนค่าในการพยากรณ์

นอกจากนี้ การเลือกแบบจำลองเป็นขั้นตอนที่สำคัญอีกขั้นตอนหนึ่งเช่นกัน นั่นคือการปรับแต่งไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning) คือค่าตัวแปร

ต่าง ๆ ของแบบจำลองที่ผู้ใช้งานสามารถกำหนดเองได้ ก่อนที่แบบจำลองจะทำการเรียนรู้ผ่านเทคนิคต่าง ๆ ในการปรับแต่ง โดยมีวิธีการที่หลากหลาย เช่น Randomized Search และ GridSearchCV เป็นต้น [21] ซึ่งในงานวิจัยนี้ เลือก GridSearchCV เป็นเทคนิคที่ใช้ในการเรียนรู้ของ เครื่องที่ส่วนหนึ่งของไลบรารี sci-kit-learn ใน Python สำหรับการปรับแต่งไฮเปอร์พารามิเตอร์ มันเป็นวิธีการที่ ประมวลผลผ่านการรวมกันของการปรับค่าตัวแปรหลาย รูปแบบ โดยทำ cross-validation เพื่อหาว่าการปรับค่าใด ให้ประสิทธิภาพดีที่สุด และนิยมใช้สำหรับการปรับแต่ง แบบจำลอง เมื่อประมวลผลผ่านแบบจำลองแล้วจะได้ แบบจำลองที่เหมาะสมที่สุดสำหรับชุดข้อมูลนั้น [18]

จากทฤษฎีที่กล่าวมา จะเป็นแนวทางในการประยุกต์ใช้การเรียนรู้ของเครื่อง เพื่อเลือกแบบจำลองที่เหมาะสมต่อใช้ในการพยากรณ์ผลผลิตทางการเกษตร และ แสดงผลของการเกิดเศษวัสดุทางการเกษตรในภาคพื้น ที่ภาคเหนือ และนำผลลัพธ์นี้ไปคำนวณหาค่าพลังงานชีวมวลที่ เกิดขึ้น ซึ่งเป็นประโยชน์ต่อการจัดการของโซ่อุปทานของเศษ วัสดุทางการเกษตรในอนาคต

3. ระเบียบวิธีวิจัย

ในงานวิจัยนี้มีจุดประสงค์ในการพยากรณ์ผลผลิต ทางการเกษตรเพื่อที่จะสามารถคำนวณหาการเกิดชีวมวลใน การนำไปใช้ในโรงไฟฟ้าชีวมวลของพื้นที่ภาคเหนือโดยใช้การ เรียนรู้ของเครื่องในการสร้างแบบจำลองในการพยากรณ์ จึงมี ระเบียบวิธีวิจัย ดังนี้

3.1 การวิเคราะห์โซ่อุปทานของเศษวัสดุทางการเกษตร สำหรับการผลิตไฟฟ้าของพลังงานชีวมวลเพื่อรวบรวม ข้อมูล

เมื่อการวิเคราะห์โซ่อุปทานของเศษวัสดุทาง การเกษตรที่ได้จากการศึกษาทฤษฎีและวรรณกรรมเกี่ยวข้อง แล้ว ทำให้สรุปข้อมูลที่จะใช้ในการประมวลผลแบบจำลองได้ เบื้องต้น จึงทำการกำหนดและสำรวจข้อมูลที่มีความพร้อมใน การใช้งาน ความถูกต้อง และครบถ้วน เพื่อให้การประมวลผล ผ่านแบบจำลองมีความแม่นยำ ทั้งนี้ จากการสำรวจข้อมูล จึง ได้กำหนดปัจจัยที่จะใช้ในการพยากรณ์ ได้แก่ ข้อมูลผลผลิต

ของพืชทางเศรษฐกิจ จำนวน 5 ชนิด ได้แก่ ข้าวโพดเลี้ยง สัตว์ ข้าวนาปี ข้าวนาปัง มันสำปะหลัง และอ้อยโรงงาน นอกจากนี้มีข้อมูลที่เกี่ยวข้องกับผลผลิตทางการเกษตร เหล่านี้ ได้แก่ พื้นที่เพาะปลูก พื้นที่เก็บเกี่ยว การใช้ปุ๋ย จำนวนคร่าวเรือนเพาะปลูก จากสำนักงานเศรษฐกิจ การเกษตร สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย และกรมส่งเสริมการเกษตร รวมถึงข้อมูลสภาพภูมิอากาศ แต่ ละจังหวัดในพื้นที่ภาคเหนือ 17 จังหวัด ซึ่งข้อมูลเหล่านี้ ได้ ทำการรวบรวมข้อมูลเป็นรายปี ในช่วงปี พ.ศ. 2553 ถึง พ.ศ. 2563 ของแต่ละจังหวัดในพื้นที่ภาคเหนือของประเทศไทย

3.2 การเตรียมข้อมูลเพื่อใช้ในการพยากรณ์ด้วยการเรียนรู้ ของเครื่อง

หลังจากการรวบรวมข้อมูลจากแหล่งข้อมูลต่างๆ แล้ว ขั้นตอนนี้ คือการเตรียมข้อมูลเพื่อใช้สำหรับแบบจำลอง ของการเรียนรู้ของเครื่อง โดยการจัดการข้อมูลให้อยู่ในข้อมูล 1 ชุดข้อมูล หรือกล่าวคือเป็นการรวบรวมข้อมูลทั้งหมดให้อยู่ ในโปรแกรม Microsoft Excel โดยให้ข้อมูลต่างๆ อิงตาม ช่วงปี และจังหวัด เพื่อข้อมูลให้มีความพร้อมในการ ประมวลผลมากที่สุด และกำจัดข้อผิดพลาดของข้อมูลที่มา จากหลากหลายฐานข้อมูล นอกจากนี้ จะทำการตรวจสอบ ความสัมพันธ์ระหว่างปัจจัย เพื่อพิจารณาในเรื่องของการ เลือกปัจจัยที่มีความสัมพันธ์ระหว่างกันสูง ปัจจัยเหล่านั้นจะ ถูกนำมาเป็นปัจจัยในการประมวลผลของการพยากรณ์ ซึ่งค่า สหสัมพันธ์ จะถูกนำมาสร้างเป็นภาพตารางเมตริกซ์ สหสัมพันธ์ หรือ Correlation Matrix ซึ่ง จะแสดงค่า สหสัมพันธ์ระหว่างปัจจัยในแต่ละคู่ปัจจัย ว่าปัจจัยใดมีผลต่อ การพยากรณ์ผลลัพธ์ ซึ่งขั้นตอนการเตรียมข้อมูลนี้เป็น ขั้นตอนที่มีความสำคัญต่อการประมวลผลของแบบจำลอง โดยสรุปเป็นขั้นตอนย่อย ดังรูปที่ 2



รูปที่ 2 ขั้นตอนการเตรียมข้อมูลเพื่อการประมวลผลของแบบจำลอง

3.3 การเลือกแบบจำลอง (Model Selection) สำหรับการพยากรณ์

เมื่อเตรียมชุดข้อมูลสำหรับการพยากรณ์เรียบร้อยแล้ว ชุดข้อมูลจะถูกนำไปในโปรแกรม Spyder เพื่อเขียนโปรแกรมภาษา Python ที่ใช้ในการประมวลผล ซึ่งการเลือกแบบจำลองได้คำนึงถึงรูปแบบการกระจายตัวของข้อมูลด้วยเช่นกัน และประกอบกับการทบทวนวรรณกรรมที่เกี่ยวข้อง ดังนั้นจึงเลือกแบบจำลองที่นำมาพิจารณาถึงความแม่นยำของการพยากรณ์ 5 แบบจำลอง ได้แก่ LASSO Regression, Support Vector Regression, Random Forest Regression, K-Nearest Neighbors และ Gradient Boosting Regression ซึ่งจากแบบจำลองเหล่านี้ จะประมวลผลผ่านข้อมูลที่ได้รับการฝึกข้อมูลและการทดสอบข้อมูล (Training and Testing Data) โดยแบ่งข้อมูลเพื่อฝึก 80 เปอร์เซ็นต์ และข้อมูลสำหรับการทดสอบ 20 เปอร์เซ็นต์ เมื่อทำการประมวลผลในแต่ละแบบจำลองแล้ว จะทำการเลือกแบบจำลองที่มีค่าความแม่นยำสูงสุด เพื่อที่จะให้แบบจำลองนั้นเป็นแบบจำลองสำหรับการพยากรณ์ข้อมูลจึงประเมินผลด้วยสมการ R^2 ที่มีค่าของการพิจารณาที่ 0 ถึง 1 ซึ่งค่าที่เข้าใกล้ 1 แสดงว่าค่าความผันแปรของปัจจัยตอบสนองที่สามารถอธิบายได้อยู่ในตัวแบบเชิงเส้น และต่อไปจึงทำการกำหนดค่าของตัวแปรของปัจจัย เพื่อให้แบบจำลองหาค่าตัวแปรที่ทำให้แบบจำลองมีความแม่นยำ

มากที่สุด โดยทำการกำหนดค่าตัวแปร ดังตารางที่ 3 ด้วยวิธีการ GridSearchCV โดยกำหนดค่า $c = 10$ จากนั้นแบบจำลองจะทำการประมวลผลและหาค่าตัวแปรที่เหมาะสมกับชุดข้อมูล โดยจะให้คะแนนที่ดีที่สุด และตัวแปรของแบบจำลองที่ดีที่สุด

ตารางที่ 3 ค่าตัวแปรที่กำหนดของแต่ละแบบจำลอง

แบบจำลอง	ค่าตัวแปรที่กำหนด
Lasso Regression	model = Lasso (alpha=0.1) alpha = [1, 5, 10, 50, 100]
Support Vector Regression	model = SVR(gamma='auto') C = [1, 10, 20, 50, 100] Kernel = ['rbf', 'linear']
Random Forest	model= RandomForestRegressor() n_estimators = [1, 5, 10, 50, 100]
Gradient Boosting Regression	model= GradientBoostingRegressor() n_estimator = [1, 5, 10, 50, 100] max_depth = [1, 3, 5] min_samples_split = [2, 5, 10] learning_rate = [0.01, 0.1, 1]
K-nearest Neighbor	model = KNeighborsRegressor() n_neighbors: [1, 5, 10, 50, 100]

3.4 การพยากรณ์ข้อมูล

หลังจากการเลือกแบบจำลองและประเมินความแม่นยำจาก 5 แบบจำลองแล้วนั้น จะได้แบบจำลองที่ใช้สำหรับพยากรณ์ผลผลิตทางการเกษตร จึงสามารถนำแบบจำลองที่มีความแม่นยำนั้นมาใช้ในการพยากรณ์กับข้อมูลทดสอบ และตรวจสอบความแม่นยำของแบบจำลองอีกครั้งเพื่อเลือกแบบจำลองที่เหมาะสมที่สุด 1 แบบจำลองสำหรับชุดข้อมูล

4. ผลการดำเนินงานวิจัย

ในผลการดำเนินงานวิจัย เป็นผลมาจากการศึกษาทฤษฎีและวรรณกรรมที่เกี่ยวข้อง และการทดลองประมวลผลผ่านแบบจำลองโดยการเขียนโปรแกรม โดยใช้ข้อมูลตามกรอบของงานวิจัย เพื่อที่จะสามารถประยุกต์ใช้แบบจำลองในการพยากรณ์ผลผลิตทางการเกษตร โดยมีผลการวิจัย ดังนี้

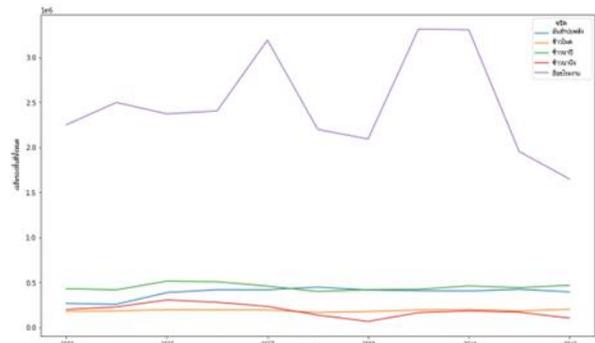
4.1 ผลการรวบรวมข้อมูลจากการวิเคราะห์โซ่อุปทานของ เศรษฐกิจทางการเกษตรสำหรับการผลิตไฟฟ้าของพลังงาน ชีวมวล

จากการศึกษาและวิเคราะห์โซ่อุปทานของเศรษฐกิจ
ทางการเกษตร รวมถึงการทบทวนวรรณกรรมที่เกี่ยวข้องกับ
การพยากรณ์ทำให้ทราบปัจจัยต่างๆ เพื่อนำไปใช้ในการ
พยากรณ์ด้วยแบบจำลอง ของการเรียนรู้ของเครื่อง พบว่า
ข้อมูลที่ใช้ในการประมวลผลผ่านแบบจำลอง มีจำนวน
11 ข้อมูล และได้กำหนดอักษรย่อเพื่อเป็นตัวแปรสำหรับ
แบบจำลองซึ่งจะช่วยให้การนำไปใช้กับโปรแกรม
Python ดังตารางที่ 4 และแสดงตัวอย่างข้อมูลดิบในการ
รวบรวมข้อมูลปัจจัยที่เกี่ยวข้อง ดังตารางที่ 5

ตารางที่ 4 ปัจจัยที่ใช้ในการพยากรณ์

ที่	ตัวแปร	อักษรย่อ	หน่วย
1	พื้นที่	Area	จังหวัด
2	พืชเศรษฐกิจ	Type	ชนิด
3	พื้นที่เพาะปลูก	PA	ไร่
4	พื้นที่เก็บเกี่ยว	HA	ไร่
5	ครวเรือนเพาะปลูก	HH	ครวเรือน
6	การใช้ปุ๋ยเคมี	FV	ตัน
7	อุณหภูมิสูงสุด	MaxTemp	องศาเซลเซียส
8	อุณหภูมิต่ำสุด	MinTemp	องศาเซลเซียส
9	อุณหภูมิเฉลี่ย	AvgTemp	องศาเซลเซียส
10	ปริมาณน้ำฝนเฉลี่ย	AvgRainfall	มิลลิเมตร
11	ปริมาณผลผลิต	Production	ตัน

ผลการรวบรวมข้อมูลต่างๆ ทางสถิติจาก
แหล่งข้อมูลของหน่วยงานที่เกี่ยวข้องด้วยโปรแกรม
Microsoft Excel และคำนวณข้อมูลทางสถิติโดยหาค่าที่มาก
ที่สุด และค่าที่น้อยที่สุด รวมถึงค่าเฉลี่ยของแต่ละข้อมูลใน
พื้นที่ภาคเหนือในช่วง พ.ศ. 2553 ถึง พ.ศ. 2563 ของพืช
ทั้งหมด ดังรูปที่ 3 เพื่อแสดงถึงภาพรวมข้อมูลที่รวบรวม ดัง
ตารางที่ 6



รูปที่ 3 กราฟของปริมาณเฉลี่ยของผลผลิตทางการเกษตรใน
พื้นที่ภาคเหนือในช่วงปี พ.ศ. 2553 ถึง 2563

นอกจากนี้ ปริมาณของพืชเฉลี่ยแต่ละชนิด ได้แก่
มันสำปะหลัง ข้าวโพดเลี้ยงสัตว์ ข้าวนาปี ข้าวนาปรัง และ
อ้อยโรงงานในช่วงปีดังกล่าว พบได้ว่า ปริมาณเฉลี่ยของอ้อย
โรงงานมีมากที่สุดของพืชเศรษฐกิจในภาคเหนือ ในส่วนของ
ปริมาณของพืชชนิดอื่นๆ มีปริมาณเฉลี่ยที่ใกล้เคียงกัน

ตารางที่ 5 ตัวอย่างการรวบรวมข้อมูลดิบของข้าวโพดเลี้ยงสัตว์ของจังหวัดเชียงรายในช่วง ปี พ.ศ. 2553 - 2563

ปี	พื้นที่เพาะปลูก (ไร่)	พื้นที่เก็บเกี่ยว (ไร่)	ครวเรือนที่ เพาะปลูก (ครวเรือน)	การใช้ปุ๋ยเคมี (ตัน)	อุณหภูมิสูงสุด (องศาเซลเซียส)	อุณหภูมิต่ำสุด (องศาเซลเซียส)	อุณหภูมิเฉลี่ย (องศาเซลเซียส)	ปริมาณน้ำฝน (มิลลิเมตร)	ปริมาณผลผลิต (ตัน)
2553	456,731	449,237	48,228	18,400.550	34.885	16.914	25.920	4.492	305,981
2554	482,469	475,073	47,551	18,822.440	34.885	16.914	25.920	4.492	333,211
2555	469,852	454,987	48,515	18,141.570	34.242	17.725	26.016	5.054	321,497
2556	489,947	478,393	51,561	19,299.120	34.175	16.863	25.629	5.452	331,456
2557	518,019	510,458	55,558	20,704.530	34.575	16.429	25.733	4.708	349,447
2558	406,040	397,570	35,516	15,750.720	35.063	17.404	25.898	3.603	263,293
2559	405,848	405,843	32,690	15,963.220	35.125	16.988	25.951	4.569	283,593
2560	393,695	393,304	35,279	15,564.090	34.892	17.063	25.807	5.952	289,421
2561	386,954	379,420	34,092	17,396.360	34.100	17.333	25.597	5.133	282,535
2562	331,783	327,119	29,026	16,124.000	35.904	16.646	26.475	2.498	223,230
2563	274,790	271,580	24,648	16,151.000	35.888	15.779	26.174	3.455	190,664

ตารางที่ 6 ข้อมูลสถิติของแต่ละข้อมูลของพื้นที่ภาคเหนือ

รายการ	ค่าสูงสุด	ค่าต่ำสุด	ค่าเฉลี่ย
1. พื้นที่เพาะปลูก	2,541,989	57	363,955.67
2. พื้นที่เก็บเกี่ยว	2,456,481	0	350,032.09
3. คริวเรือน เพาะปลูก	123,421	2	20,833.92
4. การใช้ปุ๋ยเคมี	105,947	2	16,726.26
5. อุณหภูมิสูงสุด	29.98	30.96	36.35
6. อุณหภูมิต่ำสุด	24.66	15.57	19.47
7. อุณหภูมิเฉลี่ย	38.29	20.09	27.86
8. ปริมาณน้ำฝน	5.95	0	3.82
9. ปริมาณผลผลิต	9,473,172	0	60,8726.77



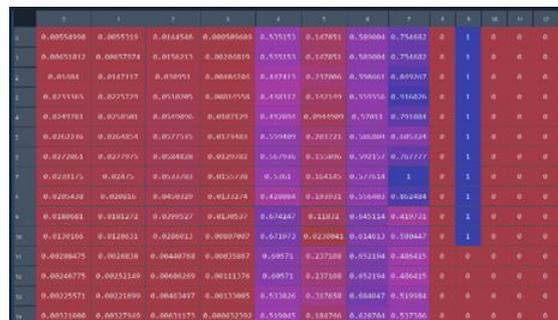
รูปที่ 4 แผนภูมิความร้อนของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแต่ละตัวแปร

4.2 ผลการเตรียมข้อมูลเพื่อใช้ในการพยากรณ์ด้วยการเรียนรู้ของเครื่อง

การเตรียมข้อมูล คือขั้นตอนที่สำคัญขั้นตอนหนึ่ง ซึ่งจากการรวบรวมข้อมูลและจัดการข้อมูลต่างๆ ให้เป็น 1 ชุดข้อมูล ซึ่งวิธีนี้เรียกว่าการผนวกข้อมูลที่อยู่ในรูปแบบตารางในโปรแกรม Microsoft Excel โดยแกนนอน (แถว) เป็นพีชแต่ละชนิดของแต่ละจังหวัดในระยะเวลา 11 ปี และ แกนตั้ง (คอลัมน์) เป็นปัจจัยที่ถูกกำหนดไว้ ทำให้ข้อมูลมี 858 แถว และ 11 คอลัมน์ เมื่อผนวกข้อมูลแล้ว ต่อไปคือนำข้อมูลเข้าไปแกรม Spyder ซึ่งหากพบข้อมูลที่ขาดหายอย่างผิดปกติ ข้อมูลนั้นจะถูกแทนด้วยค่าเฉลี่ย เพื่อลดความลำเอียงของการประมวลผลของแบบจำลอง นอกจากนี้สำหรับภารตรวจสอบความสัมพันธ์ระหว่างปัจจัยได้ใช้แผนภูมิความร้อน (Heatmap) ในแสดงผล ดังรูปที่ 4 จะเห็นได้ว่าความสัมพันธ์ของปัจจัยที่เป็นผลผลิตทางการเกษตร หรือ Production ไม่พบตัวแปรใดที่มีความสัมพันธ์ระหว่างกัน ดังนั้นจึงนำทุกปัจจัยที่คัดเลือกมาทดลองใช้ในการประมวลผลกับแบบจำลองทั้งหมด

ขั้นตอนของการจัดการข้อมูลต่อไปคือ การแปลงข้อมูลด้วยวิธีการ One-Hot Encoding เป็นวิธีการแปลงข้อมูลตัวอักษรให้อยู่ในรูปแบบตัวเลขเพื่อโปรแกรมสามารถประมวลผลได้อย่างถูกต้อง ดังนั้น ข้อมูลที่เป็นตัวอักษรที่อยู่ในชุดข้อมูล ได้แก่ จังหวัด และชนิดของพืช ให้อยู่ในรูปแบบของ 0 และ 1 โดยชื่อจังหวัด และชื่อของชนิดพืชถูกแปลงเป็นหัวตาราง

หากข้อมูลในแถวไม่มีข้อมูลในตำแหน่งใด จะปรากฏเลข 0 และในทางกลับกัน หากข้อมูลในแถวมีข้อมูลในตำแหน่งใด จะปรากฏเลข 1 ดังนั้นทำให้ 1 ชุดข้อมูลในปัจจุบันมีจำนวนของคอลัมน์เพิ่มขึ้นเป็นทั้งหมด 27 คอลัมน์ ในขั้น ตอนสุดท้ายของการเตรียมข้อมูลที่ใช้สำหรับแบบจำลองข้อมูล คือการปรับขนาดทุกข้อมูลให้อยู่ในฐานเดียวกัน เพื่อให้ข้อมูลมีการกระจายตัวแบบปกติด้วยวิธีการ Standardization ของเทคนิค MinMaxScaler ที่ทำให้ข้อมูลทั้งหมดอยู่ในช่วง 0 ถึง 1 ผลการแปลงข้อมูลเป็นดังรูปที่ 5 ซึ่งเป็นตัวอย่างข้อมูลที่ถูกปรับข้อมูลแล้วด้วยการเขียนโปรแกรม Python จะทำให้ข้อมูลทั้งหมดที่อยู่ในชุดข้อมูลที่มีหน่วยและขนาดที่แตกต่างถูกเปลี่ยนให้อยู่ในฐานเดียวกัน วิธีนี้จะช่วยลดความลำเอียงและความแปรปรวนของข้อมูลเมื่อแบบจำลองทำการเรียนรู้ และมีความพร้อมในการใช้ฝึกกับแบบจำลองที่กำหนดไว้ เมื่อเตรียมข้อมูลในขั้นตอนนี้เสร็จแล้วจึงนำข้อมูลเหล่านี้ไปฝึกผ่านแบบจำลองต่าง ๆ เพื่อที่จะเลือกแบบจำลองที่มีความเหมาะสมต่อการพยากรณ์



รูปที่ 5 ตัวอย่างส่วนหนึ่งของข้อมูลที่ถูกปรับขนาดแล้วด้วยการเขียนโปรแกรม

4.3 ผลการเลือกแบบจำลอง (Model Selection)

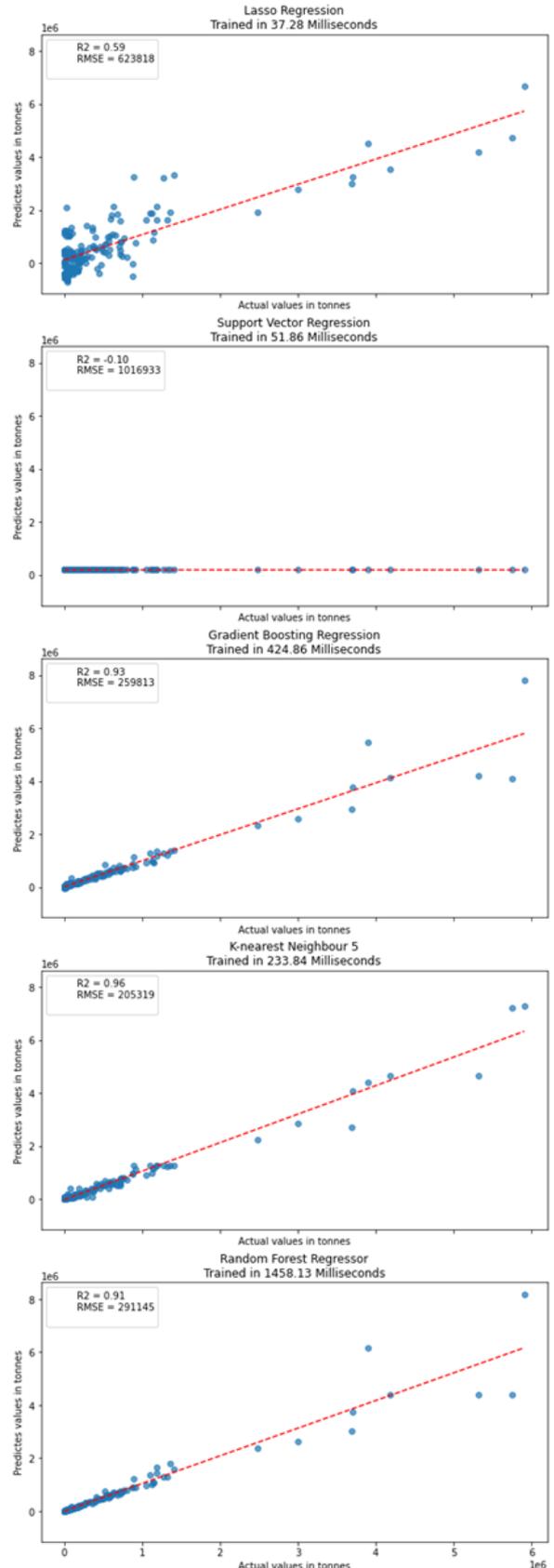
สำหรับการพยากรณ์

สำหรับการพิจารณาความเหมาะสมของแบบจำลองที่กำหนดไว้ทั้ง 5 แบบจำลอง ประมวลผลจากการฝึกข้อมูลและได้ผลลัพธ์ที่จะพิจารณาถึงความแม่นยำเป็นค่า R^2 ดังรูปที่ 6 และสรุปผลของประมวลผล ดังตารางที่ 6

ตารางที่ 6 ผลลัพธ์การฝึกข้อมูลของแต่ละแบบจำลอง

แบบจำลอง	ค่า R^2	ระยะเวลา (ms)
Lasso Regression	0.59	37.28
Support Vector Regression	-0.10	51.86
Gradient Boosting Regression	0.93	424.86
K-nearest Neighbor	0.96	233.84
Random Forest Regression	0.91	1,458.13

จากการประมวลผลของแต่ละแบบจำลอง พบว่ามีแบบจำลองที่ไม่เหมาะสมต่อชุดข้อมูล คือ Support Vector Regression เนื่องจากแบบจำลองนี้มีค่า R^2 เป็นค่าลบแสดงว่าแบบจำลองนี้ไม่สามารถปรับตัวให้เข้ากับแนวโน้มของข้อมูลได้ มีความเป็นไปได้ว่าแบบจำลองมีการเรียนรู้เกินไป (over-fitting) ทำให้แบบจำลองไม่สามารถจำแนกแยกได้ถูกต้อง รวมถึงการประมวลผลของแบบจำลอง Lasso Regression ที่มีค่า R^2 ที่ไม่มากนัก แสดงถึงความคาดเคลื่อนของแบบจำลองที่มากพอสมควร ดังนั้น ค่า R^2 ที่ต่ำเป็นการแสดงให้เห็นว่าความแม่นยำไม่มากเช่นกัน ดังนั้นจึงคัดเลือกแบบจำลอง 3 แบบจำลองที่เหลือ ได้แก่ Gradient Boosting Regression, K-nearest Neighbor และ Random Forest Regression และทำการหาค่าตัวแปรที่ดีที่สุดด้วยการปรับแต่งไฮเปอร์พารามิเตอร์ด้วยวิธีการ GridSearchCV ด้วยการกำหนดค่าตัวแปรของแบบจำลอง ได้ผลลัพธ์ดังตารางที่ 7 พบว่าแบบจำลองที่ได้รับการปรับแต่งไฮเปอร์พารามิเตอร์แล้วนั้น ทำให้ได้ค่าเฉลี่ยของการ cross-validation ของตัวแปร หรือนั่นคือค่าคะแนนที่ดีที่สุด (Best Score) ที่ใช้ในการทำนายที่ดีที่สุดที่แต่ละแบบจำลองประมวลผลได้ ซึ่งแบบจำลองที่มีความแม่นยำในการฝึกข้อมูลตามตัวแปรที่กำหนด ได้แก่ Random Forest Regression และ Gradient Boosting Regression โดยใช้ค่าตัวแปรที่แบบจำลองประมวลผลมาใช้ในการพยากรณ์ต่อไป



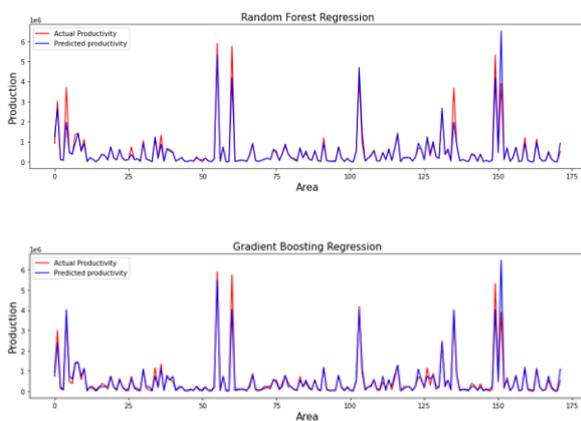
รูปที่ 6 กราฟเชิงเส้นของการประมวลผลของแต่ละแบบจำลอง

ตารางที่ 7 ค่าตัวแปรที่เหมาะสมที่สุดของแต่ละแบบจำลอง

แบบจำลอง	คะแนนที่ดีที่สุด	ค่าตัวแปรที่ดีที่สุด
Random Forest Regression	0.883	{'n_neighbors': 1}
Gradient Boosting Regression	0.821	{'learning_rate': 1, 'max_depth': 5, 'min_samples_split': 2, 'n_estimators': 1, 'random_state': 0}
K-nearest Neighbor	-0.243	{'n_neighbors': 100}

4.4 ผลการพยากรณ์ข้อมูล

เมื่อได้ผลจากการประมวลผลผ่านแบบจำลองและค่าตัวแปรเหมาะสมสำหรับชุดข้อมูลของพืชแล้ว จึงได้นำแบบจำลองนั้นๆ เมื่อเทียบระหว่างการประมวลผลกับข้อมูลที่ถูกรวบรวมและแบ่งไว้เพื่อการทดสอบ (20 เปอร์เซ็นต์ของข้อมูลทั้งหมด) จึงได้ผลลัพธ์เป็นกราฟในโปรแกรม Python ที่ประมวลผลของแบบจำลอง ซึ่งได้กราฟเส้นของการเทียบผลการพยากรณ์แทนด้วยเส้นสีน้ำเงิน และข้อมูลจริงแทนด้วยเส้นสีแดง ดังรูปที่ 7 และผลความแม่นยำของแบบจำลองจากการพยากรณ์ข้อมูลด้วยค่า MAPE ดังตารางที่ 8 ของแบบจำลอง Random Forest Regression และ Gradient Boosting Regression



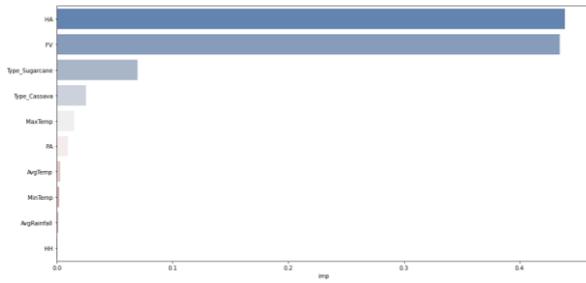
รูปที่ 7 การประมวลผลของแบบจำลองเทียบกันระหว่างข้อมูลจริงและข้อมูลพยากรณ์ของผลผลิตทางการเกษตร

ตารางที่ 8 ผลความแม่นยำของแบบจำลองจากการพยากรณ์

แบบจำลอง	ค่า MAPE (%)
Random Forest Regression	27.511
Gradient Boosting Regression	574.922

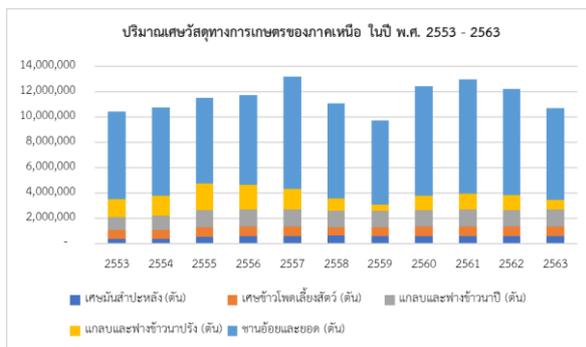
จากผลลัพธ์การประมวลผลของการพยากรณ์ผลผลิตการเกษตร และพิจารณาความแม่นยำของแบบจำลองด้วยค่า MAPE จะเห็นได้ว่าแบบจำลองของ Random Forest Regression มีค่า MAPE หรือเปอร์เซ็นต์ข้อผิดพลาดเฉลี่ย อยู่ที่ 27.511 เปอร์เซ็นต์ ในขณะที่แบบจำลอง Gradient Boosting Regression มีค่า MAPE ที่สูงถึง 574.922 เปอร์เซ็นต์ เนื่องจากค่าที่พยากรณ์ที่ได้มีค่าสูงกว่าค่าจริงอย่างมาก แสดงให้เห็นว่าแบบจำลองนี้ไม่เหมาะสมสำหรับการพยากรณ์ชุดข้อมูลนี้ ดังนั้นจึงสรุปได้ว่าในการพยากรณ์ผลผลิตทางการเกษตรด้วยข้อมูลในอดีตที่เกี่ยวข้อง 11 ปัจจัยนี้ได้ ผ่านการเรียนรู้ของเครื่อง และได้แบบจำลอง Random Forest Regression โดยกำหนดตัวแปรในแบบจำลอง n_neighbors = 1

นอกจากนี้ หลังจากการได้แบบจำลองที่เหมาะสมสำหรับการพยากรณ์ จึงได้ทำการพิจารณาถึงปัจจัยที่มีนัยยะสำคัญต่อการพยากรณ์ผลผลิตทางการเกษตรเพื่อเกิดการประยุกต์ใช้และสามารถนำไปใช้ประโยชน์สำหรับชีวมวลในอนาคตได้ ด้วยการพิจารณาปัจจัยที่มีความสำคัญที่ส่งผลต่อการพยากรณ์ 10 ลำดับแรก ดังรูปที่ 8 โดยปัจจัยของพื้นที่การเก็บเกี่ยวและการใช้ปุ๋ยเคมีมีผลต่อการพยากรณ์มากที่สุด ดังนั้นหากคำนึงถึง 2 ปัจจัยนี้ก็จะสามารถทำให้ทราบถึงผลผลิตทางการเกษตรได้ รองมาคือ ปริมาณผลผลิตของอ้อยโรงงาน และถัดไปเป็นปริมาณผลผลิตของมันสำปะหลัง แสดงให้เห็นว่า ปริมาณของพืชทั้งสองชนิดมีจำนวนมากที่สุดในชุดข้อมูล ทำให้มีผลมากที่สุดที่แบบจำลอง ซึ่งสะท้อนให้เห็นว่า ผลผลิตที่มีผลต่อการเกิดเศษวัสดุทางการเกษตรที่จะนำไปใช้เพื่อเป็นชีวมวลสำหรับการผลิตไฟฟ้า คืออ้อยโรงงาน ในทางกลับกันปัจจัยที่เป็นสภาพภูมิอากาศและจำนวนคริวเรือนที่เพาะปลูกแทบจะไม่มีผลต่อการพยากรณ์เลย



รูปที่ 8 ตัวแปรสำคัญที่มีผลต่อการพยากรณ์ผลผลิตทางการเกษตร

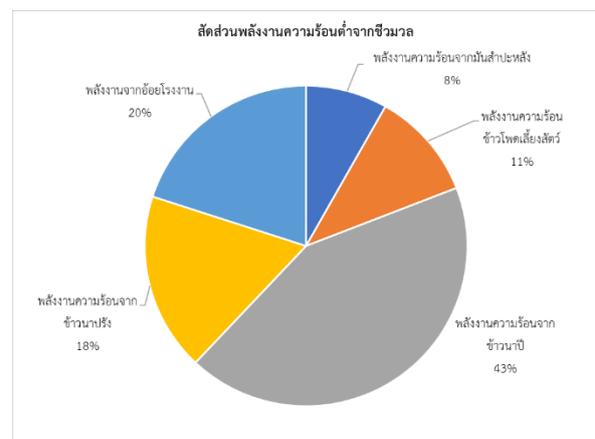
อีกทั้ง เมื่อนำแบบจำลองไปใช้ในการพยากรณ์ประยุกต์ใช้กับข้อมูลในอดีตทั้งหมด และได้นำค่าพยากรณ์ออกมาคำนวณเพื่อ ปริมาณการเกิดเศษวัสดุการเกษตรในพื้นที่ภาคเหนือ พบว่าอ้อยโรงงานมีการเกิดเศษวัสดุทางการเกษตรมากที่สุด ดังรูปที่ 9 ซึ่งสอดคล้องกับปัจจัยที่มีความสำคัญต่อแบบจำลอง และทำการสรุปเป็นค่าเฉลี่ยการเกิดเศษวัสดุทางการเกษตรต่อปีของพืชแต่ละชนิดในพื้นที่ภาคเหนือของประเทศไทย ดังตารางที่ 9 แต่ถึงอย่างไรก็ตาม การเกิดเศษวัสดุทางการเกษตรของมันสำปะหลังมีไม่มากนัก ถึงแม้ผลผลิตมีปริมาณมาก เนื่องจากมีค่าสัดส่วนการเกิดเศษวัสดุทางการเกษตรที่น้อย ดังนั้น จึงทำการสรุปสัดส่วนพลังงานค่าความร้อนต่ำที่เกิดขึ้นเมื่อนำเศษวัสดุทางการเกษตรเหล่านี้ไปเป็นแหล่งพลังงานชีวมวล เพื่อแสดงถึงความพร้อมของทรัพยากรพลังงาน ดังรูปที่ 10



รูปที่ 9 ปริมาณการเกิดเศษวัสดุทางการเกษตรของภาคเหนือ

ตารางที่ 9 ค่าเฉลี่ยของเศษวัสดุทางการเกษตรต่อปีของภาคเหนือ

ชนิดของการเกิดเศษวัสดุทางการเกษตร	ปริมาณเฉลี่ย (ตัน/ปี)
เศษมันสำปะหลัง	541,061
เศษข้าวโพด	737,275
แกลบและฟางข้าวมาปี	1,278,335
แกลบและฟางข้าวมาปี	1,315,150
ชานอ้อยและยอด	7,661,348



รูปที่ 10 สัดส่วนพลังงานความร้อนต่ำจากชีวมวล

5. สรุปผลการดำเนินงานและแนวทางในการดำเนินงานวิจัยในอนาคต

งานวิจัยนี้ได้ศึกษาและวิเคราะห์โซ่อุปทานของพลังงานชีวมวลสำหรับโรงผลิตไฟฟ้าในภาคเหนือของประเทศไทย ด้วยการเริ่มศึกษากิจกรรมชีวมวลตั้งแต่ต้นน้ำจนถึงปลายน้ำ ซึ่งพบว่าส่วนหนึ่งของชีวมวลนั้นมาจากการเกิดเศษวัสดุทางการเกษตร และได้ถูกนำไปใช้เป็นทรัพยากรของการผลิตพลังงานชีวมวล เมื่อวิเคราะห์การเกิดเศษวัสดุทางการเกษตรนั้นได้เกิดจากกิจกรรมการเพาะปลูกและการเก็บเกี่ยวเพื่อให้ได้ซึ่งผลผลิตทางการเกษตร ทำให้ได้ผลวิเคราะห์ว่า ผลผลิตทางการเกษตรมีความเชื่อมโยงกันเป็นโซ่อุปทานของชีวมวล ดังนั้นการทราบปริมาณผลผลิตทางการเกษตรช่วยให้ทราบปริมาณการเกิดเศษวัสดุทางการเกษตรได้เช่นกัน

ในการประยุกต์ใช้การเรียนรู้ของเครื่อง หรือ Machine Learning งานวิจัยนี้ได้ใช้การเขียนภาษา Python ด้วยเขียนโปรแกรม Spyder รวมถึงการกำหนดแบบจำลองที่นำมาประมวลผลข้อมูล ซึ่งแบบจำลองที่เหมาะสมสำหรับชุดข้อมูลในอดีตที่รวบรวมมานั้นคือแบบจำลอง Random Forest Regression ที่ให้ค่า MAPE เท่ากับ 27.511 โดยมีตัวแปรใช้ตัวแปร 11 ตัวแปร ได้แก่ พื้นที่ (จังหวัด), ชนิดของพืช, พื้นที่เพาะปลูก (ไร่), พื้นที่เก็บเกี่ยว (ไร่), จำนวนครัวเรือนเพาะปลูก (ครัวเรือน), จำนวนการใช้ปุ๋ยเคมี (ตัน), อุณหภูมิสูงสุด (องศาเซลเซียส), อุณหภูมิต่ำสุด (องศาเซลเซียส), อุณหภูมิเฉลี่ย (องศาเซลเซียส), ปริมาณน้ำฝน (มิลลิเมตร), และปริมาณผลผลิตทางการเกษตร (ตัน) โดยที่ตัวแปรที่มีความสำคัญมากที่สุด คือพื้นที่เก็บเกี่ยวและการใช้ปุ๋ยเคมี เมื่อทราบถึงแบบจำลองที่จะใช้ในการพยากรณ์แล้วนั้น ทำให้สามารถนำแบบจำลองนี้ไปประยุกต์ใช้ในหาพลังงานชีวมวลที่จะเกิดจากเศษวัสดุได้ โดยการนำแบบจำลองนี้ไปพยากรณ์ผลผลิตที่ผ่านมาและนำไปคำนวณกับค่าสัดส่วนที่ทำให้เกิดเศษวัสดุของพืชแต่ละชนิดได้ จนนำไปสู่การคำนวณของค่าพลังงานความร้อนต่ำของชีวมวลที่จะเกิดขึ้นจากเศษวัสดุทางการเกษตร ซึ่งจะแสดงให้เห็นถึงความพร้อมของทรัพยากรที่ใช้ในการผลิตไฟฟ้าด้วยพลังงานชีวมวลของพื้นที่ภาคเหนือที่สะท้อนถึงความพร้อมใช้ทรัพยากรได้ ซึ่งจากการพิจารณาผลจากแบบจำลองพบว่า อ้อยโรงงานเป็นพืชที่มีความสำคัญต่อการเกิดเศษวัสดุทางการเกษตร แต่ถึงอย่างไรก็ตาม ชีวมวลจากเศษวัสดุทางการเกษตรที่ทำให้เกิดพลังงานมากที่สุดเป็นพืชประเภทข้าว ซึ่งผลลัพธ์ที่ได้จะเป็นข้อมูลที่มีประโยชน์ในการวางแผนและการจัดทำยุทธศาสตร์ในการใช้พลังงานชีวมวล เพื่อตอบสนองต่อความต้องการในการผลิตไฟฟ้า และช่วยเพิ่มประสิทธิภาพในการใช้ทรัพยากรทางการเกษตรในมุมมองของการผลิตพลังงานทดแทน ทั้งนี้ แบบจำลองของการเรียนรู้ของเครื่องสามารถทำการพัฒนาแบบจำลองได้อย่างต่อเนื่อง จึงเป็นแผนในการวิจัยในอนาคต ในการหาแบบจำลองที่เหมาะสมของพืชแต่ละชนิด เพื่อเพิ่มความแม่นยำของแบบจำลองในการพยากรณ์ และเพื่อเห็นแนวโน้มของพืชแต่ละชนิดได้ดีมากขึ้น

เอกสารอ้างอิง

- [1] กรมพัฒนาพลังงานทดแทนและอนุรักษ์พลังงาน, คู่มือการพัฒนาและการลงทุนพลังงานชีวมวล, พิมพ์ครั้งที่ 1., กรุงเทพมหานคร: เอเปิล คอนซัลแตนท์, 2554.
- [2] กองนโยบายและแผนพัฒนาการเกษตร, ภาวะเศรษฐกิจทางการเกษตร ปี 2563 และแนวโน้มปี 2564, กรุงเทพมหานคร: สำนักงานเศรษฐกิจการเกษตร, 2563.
- [3] พุฒิชชาติ คิดหาทอง, วีรินทร์ หวังจิรนิรันดร และ อัจฉริยา สุริยะวงศ์, “การศึกษาศักยภาพเชิงพื้นที่ของชีวมวลสำหรับผลิตไฟฟ้าของประเทศไทย,” *วารสารวิจัยพลังงาน.*, ปีที่ 11, ฉบับที่ 1, น. 63-76, 2557.
- [4] A. Kantasa-Ard, M. Nouri, A. Bekrar, A. Ait el Cadi and Y. Sallez, “Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand,” *International Journal of Production Research.*, vol. 59, no. 24, pp. 7491-7515, 2020.
- [5] กรมพัฒนาพลังงานทดแทนและอนุรักษ์พลังงาน, คู่มือการพัฒนาและการลงทุนพลังงานชีวมวล, พิมพ์ครั้งที่ 1., กรุงเทพมหานคร: เอเปิล คอนซัลแตนท์, 2554.
- [6] B. Dangprok, K. Y. Tippayawong and N. Tippayawong, “Development of a cost optimization model for power generation from agricultural residual biomass in Thailand,” *Energy Reports.*, vol. 9, pp. 55-62, 2023.
- [7] A. Welfle, P. Gilbert and P. Thornley, “Increasing biomass resource availability

- through supply chain analysis,” *Biomass and Bioenergy.*, vol. 70, pp. 249-266, 2014.
- [8] E. Iakovou, A. Karagiannidis, D. Vlachos, A. Toka and A. Malamakis, “Waste biomass-to-energy supply chain management: a critical synthesis,” *Waste Management.*, vol. 30, no. 10, pp. 1860-1870, 2010.
- [9] K. Y. Tippayawong, N. Chaidi, T. Ngamlertsappakit and N. Tippayawong, “Demand and Cost Analysis of Agricultural Residues Utilized as Biorenewable Fuels for Power Generation,” in *International Conference on Power and Energy Systems Engineering*, Japan, 2020.
- [10] A. O. Avcioglu, M. A. Dayioglu and U. Türker, “Assessment of the energy potential of agricultural biomass residues in Turkey,” *Renewable Energy.*, vol. 138, pp. 610-619, 2019.
- [11] B. Sajjakulnukit, R. Yingyuad, V. Maneekhao, V. Pongnarintasut, S. C. Bhattacharya and P. A. Salam, “Assessment of sustainable energy potential of non-plantation biomass resources in Thailand,” *Biomass and Bioenergy.*, vol. 29, no. 3, pp. 214-224, 2005.
- [12] กอบเกียรติ สระอุบล, เรียงรู้ Data Science และ AI: Machine Learning ด้วย Python, พิมพ์ครั้งที่ 1., กรุงเทพมหานคร: ด่านสุทธาการพิมพ์, 2563.
- [13] T. V. Klompenburga, A. Kassahuna and C. Catal, “Crop yield prediction using machine learning: a systematic literature review,” *Computers and Electronics in Agriculture.*, vol. 177, pp. 1-18, 2020.
- [14] S. Makkar, G. N. R. Devi and V. K. Solanki, “Applications of machine learning techniques in supply chain optimization,” in *ICICCT 2019–System Reliability, Quality Control, Safety, Maintenance and Management: Applications to Electrical, Electronics and Computer Science and Engineering*, Singapore, Springer, 2020, pp. 861-869.
- [15] A. Nigam, S. Garg, A. Agrawal and P. Agrawal, “Crop Yield Prediction Using Machine Learning Algorithms,” in *Fifth International Conference on Image Information Processing (ICIIP)*, India, 2019.
- [16] L. S. Cedric, W. Y. H. Adoni, R. Aworka, J. T. Zoueu, F. K. Mutombo, M. Krichen and C. L. M. Kimpolo, “Crops yield prediction based on machine learning models: case of west african countries,” *Smart Agricultural Technology.*, vol. 2, 2022.
- [17] M. Shahhosseini, G. Hu and S. V. Archontoulis, “Forecasting corn yield with machine learning ensembles,” *Frontiers in Plant Science.*, vol. 11, no. 1120, pp. 1–16, 2020.
- [18] T. Onsree and N. Tippayawong, “Machine learning application to predict yields of solid products from biomass torrefaction,” *Renewable Energy.*, vol. 167, no. 2, pp. 425-432, 2021.
- [19] B. Panigrahi, K. C. R. Kathala and M. Sujatha, “A machine learning-based comparative

approach to predict the crop yield using supervised learning with regression models,” *Procedia Computer Science.*, vol. 218, pp. 2684-2693, 2023.

- [20] A. Chris, *Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning*, Sebastopol: O’Reilly Media, 2018.
- [21] T. Katongtung, T. Onsree, K. Y. Tippayawong and N. Tippayawong, “Prediction of biocrude oil yields from hydrothermal liquefaction using a gradient tree boosting machine approach with principal component analysis,” *Energy Reports.*, vol. 9, pp. 215-222, 2023.
- [22] M. Laurence, *AI and Machine Learning for Coders: A Programmer’s Guide to Artificial Intelligence*, Sebastopol: O’Reilly Media, 2021.