



FEAT JOURNAL

FARM ENGINEERING AND AUTOMATION TECHNOLOGY JOURNAL

วารสารวิศวกรรมฟาร์มและเทคโนโลยีการควบคุมอัตโนมัติ

การเปรียบเทียบแบบจำลองที่ใช้ในการพยากรณ์ค่าความหวานของอ้อย

A Comparative Study of Prediction Model

Case Study Commercial Cane Sugar Prediction

ลลิตา ขำทอง และ ศิโรรัตน์ พัฒนไพโรจน์¹⁾Lalita Khamthong and Sirorat Pattanapairoj¹⁾¹⁾สาขาวิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น

Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand

Received: 26 March 2024

Revised: 17 April 2024

Accepted: 17 April 2024

Available online: 26 June 2024

บทคัดย่อ

ค่าความหวานของอ้อยเป็นหนึ่งในปัจจัยหลักที่ส่งผลกระทบต่อปัญหาด้านปริมาณการผลิตน้ำตาลของอุตสาหกรรมน้ำตาลและเกี่ยวข้องกับระบบการซื้อขายอ้อยของเกษตรกร บทความนี้จึงมีจุดประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อค่าความหวานของอ้อยโดยใช้วิธีการถดถอยแบบขั้นบันได (Stepwise Regression) ซึ่งการศึกษานี้แบ่งปัจจัยที่ศึกษาออกเป็น 2 กรณี คือ กรณี 1 ศึกษาปัจจัยที่เกี่ยวกับการปลูกอ้อย และกรณี 2 ศึกษาปัจจัยที่เกี่ยวกับการปลูกอ้อยร่วมกับปัจจัยที่เกี่ยวกับการดูแลระหว่างการปลูกอ้อย จากนั้นนำปัจจัยที่มีผลต่อค่าความหวานของอ้อยมาสร้างแบบจำลองที่ใช้ในการพยากรณ์ค่าความหวานของอ้อย โดยใช้วิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) วิธีต้นไม้ตัดสินใจ (Decision Tree) และวิธีการสุ่มป่าไม้ (Random Forest) ร่วมกับแบ่งข้อมูลแบบตรวจสอบไขว้ (K-fold Cross-Validation) ผลการศึกษพบว่าแบบจำลองที่ใช้ในการพยากรณ์ค่าความหวานของอ้อยที่มีประสิทธิภาพสูงสุด ได้แก่ แบบจำลองที่ถูกสร้างจากวิธีการสุ่มป่าไม้ รองลงมาคือวิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียม ตามลำดับ

คำสำคัญ: ค่าความหวานของอ้อย การถดถอยแบบขั้นบันได โครงขายประสาทเทียม ต้นไม้ตัดสินใจ การสุ่มป่าไม้

Abstract

Commercial Cane Sugar (CCS) is major concern in the sugar industry, impacting both production quantity and the trading systems of sugarcane farmers. The purpose of this article is to examine the factors related to CCS using the Stepwise Regression method. This research comprises two case studies: Case 1 focused on 13 factors related to sugarcane planting, while Case 2 investigates 17 factors associated with both planting and cultivation care. The study then incorporates various factors found to impact CCS levels to construct three forecasting models: Artificial Neural Network, Decision Tree, and Random Forest, together with a K-fold Cross-Validation. The findings reveal that Random Forest is the most effective CCS forecasting model, followed by Decision Tree and Artificial Neural Network.

Keywords: CCS: Stepwise Regression: Artificial Neural Network: Decision Tree: Random Forest

*ติดต่อ: Lalita.kt@kkumail.com

1. บทนำ

ประเทศไทยคือหนึ่งในประเทศที่มีอุตสาหกรรม การเกษตรเป็นตัวขับเคลื่อนเศรษฐกิจ เนื่องจากตั้งอยู่ในเขตรมรสุมเอเชียตะวันออกเฉียงใต้ มีสภาพภูมิประเทศ ทรัพยากร สิ่งแวดล้อม และภูมิอากาศที่เอื้ออำนวยต่อการทำเกษตร โดยอุตสาหกรรม การเกษตรถือว่าเป็นธุรกิจทางการเกษตรที่เกี่ยวข้องกับการผลิตและการแปรรูปที่ต้องผ่านกระบวนการทางเครื่องจักรด้านอุตสาหกรรมให้ออกมาเป็นสินค้า หรือผลิตภัณฑ์สำหรับจัดจำหน่าย ซึ่งประเทศไทยมีการส่งออกสินค้าอุตสาหกรรมเกษตรไปยังหลาย ประเทศในภูมิภาคต่างๆ ไม่ว่าจะเป็นสหรัฐอเมริกา

จีน อินเดีย และจากสถิติมูลค่าการส่งออกสินค้า อุตสาหกรรมเกษตรของประเทศไทยในช่วง ทศวรรษที่ผ่านมา พบว่า สินค้าส่งออกอันดับต้นๆ ของ ประเทศไทย ได้แก่ อาหารทะเลแปรรูป ผลิตภัณฑ์ข้าว สาลี น้ำตาลทราย อาหารสัตว์เลี้ยง และผลไม้กระป๋อง ซึ่งจะเห็นว่าน้ำตาลทรายเป็นหนึ่งในสินค้า อุตสาหกรรมเกษตรที่สำคัญต่อการส่งออกสินค้า ของประเทศไทย ทั้งนี้สำหรับอุตสาหกรรมน้ำตาลจะ ใช้อ้อยเป็นวัตถุดิบหลักในการผลิต

อ้อยถือเป็นหนึ่งในพืชเศรษฐกิจที่สำคัญของ ประเทศไทย ในปี พ.ศ. 2565 ประเทศไทยมีพื้นที่ปลูก อ้อยรวม 47 จังหวัด รวมทั้งสิ้น 11.39 ล้านไร่ โดย กระจายอยู่ตามภาคเหนือ ภาคกลาง และภาค

ตะวันออกเฉียงเหนือ มีปริมาณอ้อยเข้าหีบรวม 92.1 ล้านตัน ดังนั้นอ้อยเข้าหีบเฉลี่ยอยู่ที่ 8.08 ตันต่อไร่ และสามารถผลิตน้ำตาลทรายได้เพียง 10.02 ล้านตัน [1] ซึ่งปัจจัยที่ส่งผลต่อปริมาณน้ำตาลทรายที่ผลิตจากอ้อยมีหลายปัจจัย หนึ่งในนั้นคือค่าความหวานของอ้อย (Commercial Cane Sugar: CCS) และค่าความหวานของอ้อยถือว่าเป็นปัจจัยหลักที่ส่งผลต่อปัญหาด้านปริมาณการผลิตน้ำตาลในแต่ละปีของอุตสาหกรรมน้ำตาลในปัจจุบัน จากสารานุกรมเยาวชน เล่มที่ 5 เรื่องอ้อย ระบุความหมายของค่าความหวานของอ้อยไว้ว่า ความหวานของอ้อย คือ ปริมาณร้อยละของน้ำตาลซูโครสที่มีอยู่ในอ้อย ซึ่งสามารถสกัดออกมาได้ในรูปของน้ำตาลทรายขาวบริสุทธิ์ ตัวอย่างเช่น ถ้าอ้อยมีค่าความหวานเท่ากับ 10 ซี.ซี.เอส หมายถึง อ้อยที่มีน้ำหนัก 1 ตัน (1,000 กิโลกรัม) จะสามารถนำมาผลิตเป็นน้ำตาลซูโครสได้สูงสุด 0.1 ตัน (100 กิโลกรัม) [2] ค่าความหวานของอ้อยจะถูกวัดค่าในช่วงฤดูกาลการผลิตน้ำตาลหรือฤดูกาลหีบอ้อย โดยฤดูกาลหีบอ้อยจะใช้ระยะเวลาประมาณ 3 ถึง 4 เดือน คือ ในช่วงเดือนธันวาคมถึงเดือนมีนาคมของทุกปี หรือขึ้นอยู่กับปริมาณอ้อยที่นำมาส่งโรงงานในปีนั้นๆ และนอกจากค่าความหวานของอ้อยจะส่งผลต่อปริมาณผลผลิตน้ำตาลทรายแล้วยังเกี่ยวข้องกับระบบการซื้อขายอ้อย นั่นคือ รายได้ของเกษตรกรที่ได้จากการขายอ้อยจะคิดคำนวณตามค่าความหวาน โดยสำหรับการผลิต พ.ศ. 2565-2566 มีอัตราราคาอ้อยขั้นต่ำอยู่ที่ 1,080 บาทต่อหนึ่งตันอ้อย ณ ค่าความหวาน 10 ซี.ซี.เอส และมีอัตราราคาขึ้นลงเท่ากับ 64.80 บาทต่อค่าความหวาน 1 ซี.ซี.เอส กล่าวคือ หากอ้อยมีความหวาน 11 ซี.ซี.เอส เกษตรกร

จะสามารถขายอ้อยได้ในราคา 1,144.80 บาท แต่ถ้าหากอ้อยมีค่าความหวาน 9 ซี.ซี.เอส เกษตรกรจะขายอ้อยได้ในราคาเพียง 1,015.30 บาท

จากที่กล่าวมาผู้วิจัยได้เล็งเห็นความสำคัญของค่าความหวานของอ้อย จึงได้ทำการศึกษาปัจจัยที่มีผลต่อค่าความหวานของอ้อยและสร้างแบบจำลองการพยากรณ์ค่าความหวานของอ้อย ซึ่งผลการวิจัยนี้จะส่งผลดีต่อโรงงานน้ำตาลและเกษตรกร โดยโรงงานน้ำตาลสามารถคาดการณ์ผลผลิตน้ำตาลได้ล่วงหน้าด้วยแบบจำลองที่มีประสิทธิภาพ ส่วนเกษตรกรสามารถนำไปวางแผนการปลูกอ้อยเพื่อเพิ่มรายได้ในฤดูกาลผลิตต่อไป

ซึ่งแบบจำลองการพยากรณ์มีประสิทธิภาพจะสามารถให้คำตอบที่มีความแม่นยำและใกล้เคียงค่าข้อมูลจริง สำหรับการสร้างแบบจำลองการพยากรณ์มีหลากหลายวิธี ตัวอย่างเช่น Attapol (2014) [3] ได้สร้างแบบจำลองการพยากรณ์ผลผลิตข้าวจากเทคนิคพยากรณ์ 3 วิธี คือ วิธีโครงข่ายประสาทเทียม การวิเคราะห์ถดถอยพหุคูณ และวิธีซัพพอร์ตเวกเตอร์เกรสชัน Duongduen et al. (2020) [4] ได้ศึกษาวิธีสุ่มป่าไม้ วิธีโครงข่ายประสาทเทียม วิธีต้นไม้สุ่ม วิธี K-nearest neighbor และวิธี K-star สำหรับพยากรณ์ปริมาณการจัดเก็บวัสดุอะไหล่ในคลัง Nieto et al. (2024) [5] ศึกษาการพยากรณ์จุดวิกฤตของอุณหภูมิวัสดุตัวนำยิ่งยวดด้วยวิธีการสุ่มป่าไม้ โครงข่ายประสาทเทียม ต้นไม้ตัดสินใจแบบ M5 และการถดถอยแบบพหุตัวแปร โดยสำหรับการศึกษาที่ผู้วิจัยใช้วิธีโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจแบบ M5 และวิธีการสุ่มป่าไม้ เนื่องจากศึกษาการสร้างแบบจำลองการพยากรณ์ที่มี

ปัจจัยส่งออกหนึ่งตัวแต่มีหลายปัจจัยนำเข้า และ ข้อมูลที่ศึกษาไม่จำเป็นต้องมีความสัมพันธ์กันแบบ เชิงเส้น

2. ทฤษฎีที่เกี่ยวข้อง

2.1. การถดถอยแบบขั้นบันได (Stepwise Regression)

วิธีการถดถอยแบบขั้นบันได เป็นวิธีที่ดำเนินการ เลือกแบบไปข้างหน้าในการคัดเลือกปัจจัยเข้าไปใน ตัวแบบถดถอยและใช้วิธีกำจัดแบบถดถอยหลังในการ คัดเลือกปัจจัยออก โดยมีหลักการ คือ ปัจจัยอิสระแรก ที่ถูกพิจารณาเข้าไปในตัวแบบถดถอยจะมีความสัมพันธ์กับปัจจัยตามมากที่สุด และมีความสามารถในการพยากรณ์ค่าปัจจัยตาม จากนั้น จะใช้วิธีกำจัดแบบถดถอยหลังในการคัดเลือกปัจจัย อิสระออกจากตัวแบบถดถอย หากค่าปัจจัยอิสระที่อยู่ ก่อนมีความสำคัญในการพยากรณ์ค่าปัจจัยตาม จะทำวิธีดำเนินการเลือกแบบไปข้างหน้าซ้ำ ซึ่งวิธี เลือกปัจจัยอิสระด้วยวิธีการถดถอยแบบขั้นบันไดจะทำ 2 วิธีนี้สลับกันจนกระทั่งได้ตัวแบบถดถอยที่ เหมาะสม และวิธีการถดถอยแบบขั้นบันไดจะสิ้นสุด การดำเนินการก็ต่อเมื่อไม่สามารถคัดเลือกปัจจัย อิสระที่เหลืออยู่เข้ามาในตัวแบบถดถอยได้ [6]

2.2. วิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

เป็นวิธีที่สร้างแบบจำลองทางคณิตศาสตร์ที่ถูก พัฒนาในการประมวลผลให้มีการทำงานคล้ายกับ โครงข่ายประสาทในสมองของมนุษย์ ซึ่งสามารถ เรียนรู้ ผิดพลาด และนำการเรียนรู้ที่ได้ไปใช้ประโยชน์ใน การแก้ปัญหา โดยพื้นฐานของโครงข่ายประสาทเทียม ประกอบด้วย 3 ชั้น คือ ชั้นข้อมูลนำเข้า ชั้นซ่อน และ

ชั้นข้อมูลส่งออก ซึ่งสามารถแบ่งประเภทของโครงข่าย ประสาทเทียมได้เป็น 2 ประเภท ได้แก่ โครงข่าย ประสาทเทียมแบบชั้นเดียว และโครงข่ายประสาท เทียมแบบหลายชั้น [7]

โดยสำหรับการศึกษานี้จะใช้โครงข่ายประสาทเทียม แบบแพร่กระจายย้อนกลับ (Back Propagation Neural Network: BPNN) ซึ่งเป็นโครงข่ายประสาท เทียมแบบหลายชั้นที่สามารถเรียนรู้และปรับปรุงแก้ไข ได้ด้วยตนเอง เนื่องจากการประมวลผลของแต่ละชั้น ของโครงข่ายประสาทเทียมจะเชื่อมโยงกันอย่างทั่วถึง หรือสามารถย้อนกลับได้ โดยหากการสร้าง แบบจำลองโครงข่ายประสาทเทียมให้คำตอบที่ผิด โครงข่ายจะปรับค่าน้ำหนักจนกว่าค่าความผิดพลาด ไม่เกิดเกณฑ์ที่กำหนด ซึ่งคำตอบที่จะได้รับในครั้ง ถัดไปจะมีความแม่นยำเพิ่มขึ้น

2.3. วิธีต้นไม้ตัดสินใจแบบ M5

วิธีต้นไม้ตัดสินใจ (Decision Tree) เป็นวิธีที่สร้าง แบบจำลองหรือพยากรณ์ล่วงหน้า ที่มีลักษณะ โครงสร้างคล้ายกับต้นไม้กลับหัว คือ มีรากของต้นไม้ อยู่ส่วนบนและใบอยู่ส่วนล่าง สำหรับวิธีต้นไม้ ตัดสินใจแบบ M5 ถูกพัฒนาโดย Quinlan ปี พ.ศ. 2535 ซึ่งเป็นวิธีที่ได้การยอมรับและนำไปใช้ในการ พยากรณ์อย่างกว้างขวาง วิธี M5 เป็นวิธีที่ใช้หลักการ การดัดแปลงฟังก์ชันที่ไม่เชิงเส้น (Nonlinear Function) โดยแบ่งให้เป็นส่วนย่อยหลาย ๆ ส่วน แล้ว นำแต่ละส่วนย่อยมาคำนวณหาความสัมพันธ์โดยใช้ ฟังก์ชันถดถอยเชิงเส้น โดยในการสร้างแบบจำลอง ด้วยวิธี M5 จะประกอบด้วย 3 ขั้นตอน คือ ขั้นตอน การสร้างต้นไม้ ขั้นตอนการตัดกิ่ง และขั้นตอนการทำ ให้เรียบ [8]

2.4. วิธีการสุ่มป่าไม้ (Random Forest)

เป็นหนึ่งในวิธีที่ได้รับความนิยมสำหรับการแก้ปัญหาการจำแนกและการวิเคราะห์ถดถอย โดยวิธีการสุ่มป่าไม้เป็นวิธีที่ได้รับการต่อยอดมาจากวิธีต้นไม้ตัดสินใจ แต่ต่างกันที่วิธีการสุ่มป่าไม้เป็นการเพิ่มจำนวนต้นไม้เป็นหลายๆ ต้น เพื่อเพิ่มประสิทธิภาพการพยากรณ์และป้องกันการเกิดปัญหาการเรียนรู้เกินความเป็นจริง (Over Fitting) ซึ่งวิธีการสุ่มป่าไม้มีหลักการทำงาน คือ จะแบ่งข้อมูลออกเป็นต้นไม้ตัดสินใจหลายๆ ต้น โดยจะสร้างจากการสุ่มข้อมูลตัวอย่างแบบเลือกแล้วใส่กลับ ทำให้ต้นไม้ตัดสินใจแต่ละต้นจะได้รับคุณลักษณะและข้อมูลที่ไม่เหมือนกันทั้งหมด เพื่อให้ต้นไม้ตัดสินใจมีความหลากหลายและมีเป็นอิสระต่อกันมากขึ้น สำหรับค่าพยากรณ์ที่เหมาะสมของการแก้ปัญหาการวิเคราะห์ถดถอย จะคำนวณจากค่าเฉลี่ยของค่าพยากรณ์ของต้นไม้ตัดสินใจแต่ละต้น [9]

3. วิธีการการวิจัย

การวิจัยนี้ได้ศึกษาหาปัจจัยที่มีผลต่อค่าความหวานของอ้อย โดยใช้วิธีการถดถอยแบบขั้นบันได (Stepwise Regression) จากนั้นนำปัจจัยที่มีผลต่อค่าความหวานของอ้อยมาใช้ในการสร้างแบบจำลองที่ใช้ในการพยากรณ์ค่าความหวานของอ้อย โดยมีรายละเอียดขั้นตอน ดังนี้

3.1. การเตรียมข้อมูล

การศึกษาปัจจัยที่มีผลต่อค่าความหวานของอ้อยมีหลากหลายงานวิจัย ตัวอย่างเช่น Prasit et al. (2013) [10] ได้สร้างแบบจำลองทางคณิตศาสตร์สำหรับการพยากรณ์ค่าความหวานของอ้อยในฤดูกาลหีบอ้อย

โดยเก็บข้อมูลในรูปของอุณหภูมิและค่าความชื้นสะสม Hongthong et al. (2017) [11] ได้ศึกษาปัจจัยสาเหตุของความผันแปรรายวันของค่าความหวานของอ้อยในช่วงฤดูกาลหีบอ้อย โดยได้ศึกษาปัจจัยเกี่ยวกับอ้อยจำนวน 12 ปัจจัย Pelioia et al. (2019) [12] ได้สร้างต้นไม้ตัดสินใจสำหรับการพยากรณ์ผลผลิตของโรงงานน้ำตาลโดยมีปัจจัยที่ใช้ในการสร้างต้นไม้จำนวน 68 ปัจจัย ซึ่งผู้วิจัยได้ศึกษาปัจจัยที่ถูกเลือกใช้ในหลายงานวิจัย ปัจจัยที่สามารถเก็บได้ตามข้อกำหนดและปัจจัยที่คาดว่าจะมีผลต่อค่าความหวานของอ้อย

โดยข้อมูลที่ใช้ในงานวิจัยถูกรวบรวมจากโรงงานน้ำตาลแห่งหนึ่งในจังหวัดสิงห์บุรี และเกษตรกรที่ส่งอ้อยให้โรงงานน้ำตาล ในช่วงฤดูกาลผลิตปี พ.ศ. 2565-2566 ซึ่งแบ่งกลุ่มปัจจัยที่จะถูกนำมาพิจารณาในการค้นหาปัจจัยที่มีผลต่อค่าความหวานของอ้อยออกเป็น 2 กรณี ได้แก่

กรณีที่ 1 พิจารณากลุ่มปัจจัยที่เกี่ยวกับการปลูกอ้อยจำนวน 13 ปัจจัย ประกอบด้วย สายพันธุ์อ้อย ประเภทของอ้อย ประเภทของดิน ประเภทของปุ๋ย ระยะเวลาปลูกอ้อย ลักษณะในการตัดอ้อย ปริมาณน้ำฝนสะสม ปริมาณน้ำฝนมากที่สุด จำนวนวันที่ฝนตกตลอดปี ประเภทของแหล่งน้ำที่ใช้ การใช้ประโยชน์จากที่ดินก่อนการปลูกอ้อย เวลาที่ใช้เดินทางจากไร่ถึงโรงงาน และค่าความหวานของอ้อย และกรณีที่ 2 พิจารณากลุ่มปัจจัยที่เกี่ยวกับการปลูกอ้อย ร่วมกับปัจจัยที่เกี่ยวข้องกับการดูแลระหว่างการปลูกอ้อยอีกจำนวน 4 ปัจจัย ได้แก่ จำนวนครั้งในการใส่ปุ๋ยตลอดการปลูก ระยะห่างในการใส่ปุ๋ย จำนวนครั้งในการกำจัดวัชพืช และจำนวนครั้งในการให้น้ำตลอดการปลูก รวมเป็น 17 ปัจจัย

ข้อมูลปัจจัยผ่านการคัดกรองและกำจัดข้อมูลที่ไม่สมบูรณ์หรือขาดหายทิ้ง ออกจากข้อมูลทั้งหมดด้วยวิธีตัดแถวข้อมูลที่ไม่สมบูรณ์ (List wise Deletion) และแปลงข้อมูลให้อยู่ในรูปมาตรฐานด้วยวิธี Min-Max Normalization เพื่อลดความเอนเอียงของข้อมูล

3.2. การวิเคราะห์หาปัจจัยที่มีผลต่อค่าความหวานของอ้อย

การวิเคราะห์หาปัจจัยที่มีผลต่อค่าความหวานของอ้อย จะถูกแบ่งออกเป็น 2 กรณีดังที่ได้กล่าวมาในหัวข้อ 3.1. สำหรับการคัดเลือกปัจจัยงานวิจัยนี้ได้ใช้การคัดเลือกปัจจัยด้วยวิธีการถดถอยแบบขั้นบันได ซึ่งปัจจัยที่ถูกวิเคราะห์ว่ามีผลต่อค่าความหวานของอ้อยจะถูกนำไปใช้ในการสร้างแบบจำลองที่ใช้ในการพยากรณ์ค่าความหวานต่อไป

3.3. การสร้างแบบจำลองการพยากรณ์ค่าความหวานของอ้อย

ในการสร้างแบบจำลองที่ใช้ในการพยากรณ์ค่าความหวานของอ้อยจะใช้ข้อมูลปัจจัยที่ได้จากการคัดกรองด้วยวิธีการถดถอยแบบขั้นบันไดเป็นข้อมูลนำเข้า (Input) ให้กับแบบจำลองที่สร้างขึ้นจากวิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) วิธีต้นไม้ตัดสินใจ (Decision Tree) และวิธีการสุ่มป่าไม้ (Random Forest) โดยข้อมูลนำเข้านั้นจะถูกแบ่งออกเป็นชุดข้อมูลสำหรับแบบจำลองได้เรียนรู้เพื่อค้นหาความสัมพันธ์ระหว่างปัจจัยที่มีผลต่อค่าความหวานของอ้อยกับค่าความหวานของอ้อย และชุดข้อมูลสำหรับทดสอบประสิทธิภาพในการพยากรณ์ค่าความหวานของอ้อยของแบบจำลองที่สร้างขึ้น งานวิจัยนี้ใช้วิธีการแบ่งข้อมูลแบบตรวจสอบ

ไขว้ (K-fold Cross Validation) กล่าวคือ ข้อมูลทั้งหมดจะถูกแบ่งออกเป็น K กลุ่ม โดยข้อมูลทุกกลุ่มจะถูกสลับเป็นทั้งชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ เช่น หากใช้ $K=3$ ข้อมูลทั้งหมดจะถูกแบ่งออกเป็น 3 กลุ่ม ในครั้งแรกข้อมูลในกลุ่มที่ 1 และ 2 จะถูกใช้เป็นข้อมูลสำหรับให้แบบจำลองได้เรียนรู้เพื่อหาความสัมพันธ์ระหว่างปัจจัยที่มีผลต่อค่าความหวานของอ้อยกับค่าความหวานของอ้อย และข้อมูลกลุ่มที่ 3 จะถูกนำไปใช้เป็นชุดข้อมูลสำหรับทดสอบประสิทธิภาพในการพยากรณ์ค่าความหวานของอ้อย ในครั้งที่สอง ข้อมูลกลุ่มที่ 1 และ 3 จะถูกใช้เป็นข้อมูลสำหรับให้แบบจำลองได้เรียนรู้เพื่อหาความสัมพันธ์ระหว่างปัจจัยที่มีผลต่อค่าความหวานของอ้อยกับค่าความหวานของอ้อย และข้อมูลกลุ่มที่ 2 จะถูกนำไปใช้เป็นชุดข้อมูลสำหรับทดสอบประสิทธิภาพในการพยากรณ์ค่าความหวานของอ้อย

และในครั้งที่สาม ข้อมูลกลุ่มที่ 2 และ 3 จะถูกใช้เป็นข้อมูลสำหรับให้แบบจำลองได้เรียนรู้เพื่อหาความสัมพันธ์ระหว่างปัจจัยที่มีผลต่อค่าความหวานของอ้อยกับค่าความหวานของอ้อย และข้อมูลกลุ่มที่ 1 จะถูกนำไปใช้เป็นชุดข้อมูลสำหรับทดสอบประสิทธิภาพในการพยากรณ์ค่าความหวานของอ้อย เป็นต้น

ในงานวิจัยนี้ได้กำหนดให้ K มีค่าเท่ากับ 2 ถึง 5 และทำการสุ่มแบ่งข้อมูลในแต่ละ K จำนวน 20 ครั้ง เพื่อลดความเอนเอียงในการสุ่มข้อมูล

3.4. การเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์

การเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์สามารถใช้ได้หลายวิธี โดยงานวิจัยนี้จะใช้

วิธีรากที่สองของค่าความคลาดเคลื่อนเฉลี่ย (Root Mean Square Error: RMSE) [12] ซึ่งเป็นวิธีที่ได้รับ ความนิยมและแปลผลง่าย RMSE สามารถคำนวณได้ จากสมการ

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i' - Y_i)^2} \quad (1)$$

โดยที่ Y_i' คือ ค่าข้อมูลจากการพยากรณ์

Y_i คือ ค่าข้อมูลจริง

n คือ จำนวนข้อมูลทั้งหมด

4. ผลการวิจัย

4.1. ผลการคัดเลือกปัจจัยที่มีผลต่อค่าความหวาน ของอ้อย

ผลการคัดเลือกปัจจัยที่มีผลต่อค่าความหวานของ อ้อย โดยใช้วิธีการถดถอยแบบขั้นบันไดผ่านโปรแกรม สำเร็จรูป IBM SPSS Statistics v.23 แบ่งออกเป็น 2 กรณี ดังนี้

กรณีที่ 1 เมื่อพิจารณาปัจจัยนำเข้า 12 ปัจจัย และ ปัจจัยส่งออก 1 ปัจจัย จากข้อมูลจำนวน 383 ข้อมูล พบว่า ปัจจัยที่มีผลต่อค่าความหวานของอ้อย ได้แก่ ระยะเวลาปลูกอ้อย ประเภทของอ้อย ลักษณะในการ ตัดอ้อย ปริมาณน้ำฝนมากที่สุด และการใช้ประโยชน์ จากที่ดินก่อนการปลูกอ้อย รวม 5 ปัจจัย

กรณีที่ 2 เมื่อพิจารณาปัจจัยนำเข้า 16 ปัจจัย และ ปัจจัยส่งออก 1 ปัจจัย จากข้อมูลจำนวน 103 ข้อมูล พบว่า ปัจจัยที่มีผลต่อค่าความหวานของอ้อย ได้แก่ ระยะเวลาปลูกอ้อย ประเภทของอ้อย จำนวนวันที่ฝน ตก ระยะห่างในการใส่ปุ๋ยตลอดการปลูก และปริมาณ น้ำฝนสะสม รวม 5 ปัจจัย แสดงในตารางที่ 1

ตารางที่ 1

ปัจจัยที่มีผลต่อค่าความหวานของอ้อย

ปัจจัยที่ศึกษา	ปัจจัยที่มีผลต่อค่าความหวานของอ้อย	
	กรณี 1	กรณี 2
1. สายพันธุ์อ้อย		
2. ประเภทของอ้อย	✓	✓
3. ประเภทของดิน		
4. ประเภทของปุ๋ย		
5. ระยะเวลาปลูก	✓	✓
6. ลักษณะในการตัดอ้อย	✓	
7. ปริมาณน้ำฝนสะสม		✓
8. ปริมาณน้ำฝนมากที่สุด	✓	
9. จำนวนวันที่ฝนตก		✓
10. ประเภทของแหล่งน้ำที่ใช้		
11. การใช้ประโยชน์จากที่ดิน	✓	
12. เวลาที่ใช้ในการเดินทาง		
13. จำนวนครั้งในการใส่ปุ๋ย		
ปัจจัยที่ศึกษา	ปัจจัยที่มีผลต่อค่าความหวานของอ้อย	
	กรณี 1	กรณี 2
14. ระยะห่างในการใส่ปุ๋ย		✓
15. จำนวนครั้งในการกำจัดวัชพืช		
16. จำนวนครั้งในการให้น้ำตลอดการปลูก		

4.2. ประสิทธิภาพแบบจำลองการพยากรณ์ค่า

ความหวานของอ้อย

ตารางที่ 2

ค่า RMSE ของแบบจำลองที่สร้างขึ้นจากการใช้ปัจจัยกรณี 1 เป็นข้อมูลนำเข้า

K-fold	แบบจำลองค่าความหวานของอ้อย		
	โครงข่ายประสาทเทียม	ต้นไม้ตัดสินใจ	การสุ่มป่าไม้
K – fold 2	1.2945	1.0805	1.0788
K – fold 3	1.1885	1.0740	1.0738
K – fold 4	1.1648	1.0689	1.0683
K – fold 5	1.1801	1.0996	1.0779

ตารางที่ 3

ค่า RMSE ของแบบจำลองที่สร้างขึ้นจากการใช้ปัจจัยกรณี 2 เป็นข้อมูลนำเข้า

K-fold	แบบจำลองค่าความหวานของอ้อย		
	โครงข่ายประสาทเทียม	ต้นไม้ตัดสินใจ	การสุ่มป่าไม้
K – fold 2	1.5986	1.3881	0.9059
K – fold 3	1.3384	0.9249	0.8727
K – fold 4	1.3230	0.9188	0.8613
K – fold 5	1.2003	0.9377	0.8507

สำหรับการสร้างแบบจำลองค่าความหวานของอ้อย จะสร้างผ่านโปรแกรมสำเร็จรูป WEKA 3.8.3 และการกำหนดพารามิเตอร์ของวิธีโครงข่ายประสาทเทียมจะกำหนดอัตราการเรียนรู้ (Learning Rate) เท่ากับ 0.1 ค่าโมเมนตัม (Momentum) เท่ากับ 0.1 และรอบการเรียนรู้ (Learning Cycle) เท่ากับ 5,000 รอบ

จากตารางที่ 2 และ 3 แสดงให้เห็นว่า แบบจำลองที่สร้างขึ้นจากการใช้ปัจจัยที่ได้จากวิธีการถดถอยแบบขั้นบันไดร่วมกับวิธีการสุ่มป่าไม้มีความแม่นยำในการพยากรณ์ค่าความหวานของอ้อยมากกว่าวิธีต้นไม้

ตัดสินใจ และวิธีโครงข่ายประสาทเทียม ในทุกระดับ K เนื่องจากเป็นวิธีที่ให้ค่า RMSE ต่ำสุด

5. อภิปรายและสรุปผลการวิจัย

บทความวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อค่าความหวานของอ้อยโดยใช้วิธีการถดถอยแบบขั้นบันได และสร้างแบบจำลองการพยากรณ์ค่าความหวานของอ้อยด้วยวิธีโครงข่ายประสาทเทียมวิธีต้นไม้ตัดสินใจ และวิธีการสุ่มป่าไม้ โดยแบ่งกลุ่มปัจจัยที่ศึกษาออกเป็น 2 กรณี คือ กรณีที่ 1 กลุ่มปัจจัยที่เกี่ยวข้องกับการปลูกอ้อย จำนวน 13 ปัจจัย และกรณีที่ 2 กลุ่มปัจจัยที่เกี่ยวข้องกับการปลูกอ้อยร่วมกับปัจจัยที่เกี่ยวข้องกับการดูแลระหว่างการปลูกอ้อย จำนวน 17 ปัจจัย ผลการศึกษาปัจจัยที่มีผลต่อค่าความหวานของอ้อย พบว่า ปัจจัยที่มีผลต่อค่าความหวานของอ้อยของกรณีที่ 1 ประกอบด้วย ระยะเวลาปลูก ประเภทของอ้อย ลักษณะในการตัดอ้อย ปริมาณน้ำฝนมากที่สุด และการใช้ประโยชน์จากที่ดินก่อนการปลูกอ้อย รวม 5 ปัจจัย ส่วนปัจจัยที่มีผลต่อค่าความหวานของอ้อยกรณีที่ 2 ประกอบด้วย ระยะเวลาปลูก ประเภทของอ้อย จำนวนวันที่ฝนตก ระยะห่างในการใส่ปุ๋ยตลอดการปลูก และปริมาณน้ำฝนสะสม รวม 5 ปัจจัย ซึ่งจากปัจจัยที่มีผลต่อค่าความหวานของอ้อยกรณีที่ 1 และกรณีที่ 2 พบว่า ปัจจัยที่เกี่ยวข้องกับปริมาณน้ำฝนสอดคล้องกับงานวิจัยของ Prasit et al. ที่ได้สร้างแบบจำลองทางคณิตศาสตร์สำหรับการพยากรณ์ค่าความหวานของอ้อยในช่วงฤดูกาลหีบอ้อยให้อยู่ในรูปของอุณหภูมิและปริมาณ ส่วนปัจจัยระยะเวลาปลูกอ้อยสอดคล้องกับผลการศึกษาปัจจัยสาเหตุความผันแปรของความ

หวานอ้อยในฤดูกาลเก็บอ้อยของ Hongthong et al. สำหรับการสร้างแบบจำลองการพยากรณ์วิธีการสุ่มป่าไม้มีประสิทธิภาพในการพยากรณ์ค่าความหวานของอ้อยมากกว่าวิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียม เนื่องจากเป็นวิธีที่เพิ่มความแม่นยำ (Accuracy) จากการสร้างต้นไม้หลายๆ ต้น และมีการป้องกันการเกิดปัญหาการเรียนรู้ดีเกินความเป็นจริง (Over Fitting) ซึ่งสอดคล้องกับงานวิจัยของกฤษฎาณ [13] ที่ได้ทำการเปรียบเทียบแบบจำลองสำหรับจำแนกการใช้ประโยชน์ที่ดินด้วยวิธีโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ วิธีการสุ่มป่าไม้ และวิธีซัพพอร์ตเวกเตอร์ ซึ่งผลการวิจัยนี้สามารถนำไปประยุกต์ใช้ทั้งโรงงานน้ำตาลและเกษตรกร โดยโรงงานน้ำตาลสามารถคาดการณ์ผลผลิตน้ำตาลได้ล่วงหน้าด้วยแบบจำลองที่มีประสิทธิภาพ ส่วนเกษตรกรสามารถนำไปวางแผนการปลูกอ้อยเพื่อเพิ่มรายได้ในฤดูกาลผลิตต่อไปได้ อย่างไรก็ตาม การศึกษานี้เป็นการศึกษาที่มีข้อจำกัดในการเก็บรวบรวมข้อมูล และศึกษาการสร้างแบบจำลองการพยากรณ์เพียง 3 วิธีเท่านั้น ซึ่งในอนาคตสามารถปรับเปลี่ยนเพิ่มเติม ปัจจัยอื่นๆ ที่เกี่ยวกับการปลูกอ้อย หรือศึกษาการสร้างแบบจำลองการพยากรณ์ให้เหมาะสมกับข้อมูลเพื่อเพิ่มประสิทธิภาพในการพยากรณ์มากขึ้น

6. เอกสารอ้างอิง

- [1] สำนักงานคณะกรรมการอ้อยและน้ำตาล. รายงานสถานการณ์การปลูกอ้อยปีการผลิต 65/66 (สื่อออนไลน์). [เข้าถึงเมื่อวันที่ 24 สิงหาคม 2566]. เข้าถึงได้จาก : <http://www.ocsb.go.th/upload/journal/fileupload/13813-1585.pdf>
- [2] เกษม สุขสถาน. อ้อย (สื่อออนไลน์). [เข้าถึงเมื่อวันที่ 10 สิงหาคม 2566]. เข้าถึงได้จาก : <http://saranukromthai.or.th/sub/book/book.php?book=5&page=main>
- [3] อัดพล คุณเลิศ. การศึกษาแบบจำลองการพยากรณ์ปริมาณผลผลิตข้าว [วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต]. ขอนแก่น: มหาวิทยาลัยขอนแก่น; 2557.
- [4] ดวงเดือน อ่องรุ่งเรือง, ศุภศิริ ไทยวิรัช, วินทนี ประจวบศุภกิจ. ตัวแบบการพยากรณ์ปริมาณการจัดเก็บวัสดุอะไหล่ในคลังโดยใช้เทคนิคการเรียนรู้ของเครื่อง: กรณีศึกษาบริษัทผลิตปูนซีเมนต์. PSRU Journal of Science and Technology. 2020; 5(1): 81-92.
- [5] Nieto PJ, Gonzala EG, Garcia LA, Prado LA, Sanchez AB. Predicting the critical superconducting temperature using the random forest, MLP neural network, M5 model tree and multivariate linear regression. Alexandria Engineering Journal. 2024; 86: 144-156.
- [6] ธิดาเดียว มยุรีสุวรรณ. การคัดเลือกตัวแปรและการสร้างตัวแบบ. โรงพิมพ์มหาวิทยาลัยขอนแก่น: ขอนแก่น; 2558.
- [7] ธนาวุฒิ ประกอบผล. โครงข่ายประสาทเทียม (Artificial Neural Network). วารสารมชกวิชาการ. 2552; 12: 73-87.

- [8] Ditthakit P, Chinnarasri C. Estimation of pan coefficient using M5 model tree. *American Journal of Environmental Sciences*. 2012; 8: 95-103.
- [9] สุภาภรณ์ พัฒนาวาส์ปราการ. การวิเคราะห์เทคนิคการจำแนกประเภทข้อมูลกรณีศึกษา การทำนายระดับชั้นผู้รับเหมาก่อสร้างสำหรับโครงการก่อสร้างของภาครัฐ [วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต]. กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์; 2563.
- [10] Somjinda P, Trakootingcharoen P, Sangkhasila K. Mathematical Model for Predicting the Effects of Air Temperature and Heat Degree Days. *Journal of Science and Technology Kasetsart University*. 2013; 2: 31-55.
- [11] Hongthong P, Patanothai A. Causal Factors of Daily Variation in Commercial Cane Sugar Value of Processed sugarcane Throughout the Milling Season in Northeast Thailand. *Thai Journal of Agricultural Science*. 2017; 50: 49-65.
- [12] Peloia PR, Bocca FF, Rodrigues LHA. Identification of patterns for increasing production with decision trees in sugarcane mill data. *Science Agricola*. 2019; 76(4): 281-289.
- [13] กฤษฎาณ อินทร์ตัน. การจำแนกการใช้ประโยชน์ที่ดินในจังหวัดนครนายกด้วยอัลกอริทึมการเรียนรู้เครื่องและภาพถ่ายจากดาวเทียม Sentinel-2. *วารสารวิทยาศาสตร์บูรพา*. 2565; 27: 1153-1171.
- [14] ขนิษฐา กุลนาวิน, วันเพ็ญ โพธิ์เกษม และ ศันสนีย์ เลี้ยงพานิชย์. การออกแบบแบบจำลองการพยากรณ์ผลผลิตมันสำปะหลังในเขตพื้นที่จังหวัดนครราชสีมาด้วยเทคนิคเมื่องข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี*. 2565; 1: 1-16.